

(12) FASCÍCULO DE PATENTE DE INVENÇÃO

(22) Data de pedido: 2007.03.05	(73) Titular(es): CALIFORNIA INSTITUTE OF TECHNOLOGY 1200 EAST CALIFORNIA BLVD. MS 201-85 PASADENA CA 91125 US
(30) Prioridade(s): 2006.03.03 US 779375 P 2006.03.03 US 779376 P	
(43) Data de publicação do pedido: 2008.12.10	(72) Inventor(es): PIN WANG INCHAN KWON SOOJIN SON YI TANG DAVID A. TIRRELL US US US US US
(45) Data e BPI da concessão: 2014.06.25 183/2014	(74) Mandatário: NUNO MIGUEL OLIVEIRA LOURENÇO RUA CASTILHO, Nº 50 - 9º 1269-163 LISBOA PT

(54) Epígrafe: **INCORPORAÇÃO SÍTIO-ESPECÍFICA DE AMINOÁCIDOS EM MOLÉCULAS**

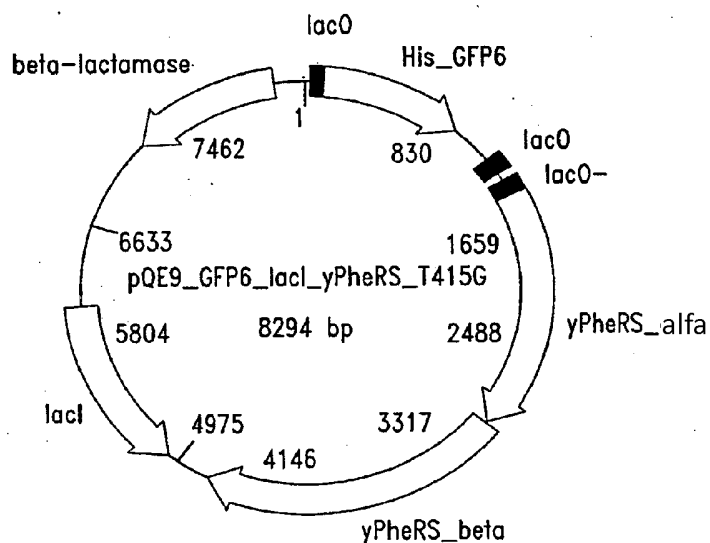
(57) Resumo:

A INVENÇÃO FORNECE CERTAS FORMAS DE REALIZAÇÃO RELACIONADAS COM MÉTODOS E COMPOSIÇÕES PARA INCORPORAR AMINOÁCIDOS NÃO-NATURAIS NUM POLIPÉPTIDO OU PROTEÍNA UTILIZANDO UMA AMINOACIL-TRNA SINTETASE MUTANTE OU MODIFICADA PARA TRANSFERIR O AMINOÁCIDO NÃO-NATURAL PARA O TRNA CORRESPONDENTE. EM CERTAS FORMAS DE REALIZAÇÃO, O TRNA É TAMBÉM MODIFICADO DE MODO A QUE O COMPLEXO FORME EMPARELHAMENTOS DE BASE DE WATSON-CRICK PERFEITOS COM UM CODÃO QUE NORMALMENTE FORMA EMPARELHAMENTO DE BASES WOBBLE COM PARES DE TRNA NÃO MODIFICADO / AMINOACIL-TRNA SINTETASE.

RESUMO

"INCORPORAÇÃO SÍTIO-ESPECÍFICA DE AMINOÁCIDOS EM MOLÉCULAS"

A invenção fornece certas formas de realização relacionadas com métodos e composições para incorporar aminoácidos não-naturais num polipéptido ou proteína utilizando uma aminoacil-tRNA sintetase mutante ou modificada para transferir o aminoácido não-natural para o tRNA correspondente. Em certas formas de realização, o tRNA é também modificado de modo a que o complexo forme emparelhamentos de base de Watson-Crick perfeitos com um codão que normalmente forma emparelhamento de bases *wobble* com pares de tRNA não modificado / aminoacil-tRNA sintetase.



DESCRIÇÃO

"INCORPORAÇÃO SÍTIO-ESPECÍFICA DE AMINOÁCIDOS EM MOLÉCULAS"

DECLARAÇÃO DE INTERESSE DO GOVERNO

Esta invenção foi feita com apoio do governo federal sob a concessão número GM62523, atribuída pelo NIH. O governo dos Estados Unidos tem certos direitos da invenção.

ANTECEDENTES DA TÉCNICA DA INVENÇÃO

Engenharia de proteínas é uma ferramenta preponderante para a modificação das propriedades estruturais, catalíticas e de ligação de proteínas naturais e para a concepção *de novo* de proteínas artificiais. Engenharia de proteínas baseia-se num mecanismo de reconhecimento eficiente para incorporar aminoácidos mutantes nas sequências de proteína desejadas. Embora este processo seja muito útil para a concepção de novas macromoléculas com controlo preciso da composição e arquitetura, uma grande limitação é que a mutagénese é restrita para os 20 aminoácidos naturais. No entanto, está a tornar-se cada vez mais claro que a incorporação de aminoácidos antinaturais pode ampliar o alcance e o impacto de métodos de engenharia de proteínas.

Aminoácidos não-naturais carregando uma grande variedade de novos grupos funcionais têm sido globalmente substituídos para substituição de resíduos específicos ou incorporação em proteínas recombinantes. Assimilação biossintética de aminoácidos não-canónicos em proteínas tem sido alcançada em grande parte explorando a capacidade do aparelho de síntese de tipo selvagem para utilizar os análogos de

aminoácidos naturais (Budisa 1995, Eur. J. Biochem 230: 788-796; Deming 1997, J. Macromol. Sci. Pure Appl. Chem A34: 2143-2150; Duewel 1997, Biochemistry 36: 3404-3416; van Hest and Tirrell 1998, FEBS Lett 428(1-2): 68-70; Sharma et al., 2000, FEBS Lett 467(1): 37-40). No entanto, há situações em que a substituição de local único ou incorporação por aminoácidos não-naturais é necessária. Tal metodologia permitiria a personalização de uma proteína (o tamanho, acidez, nucleofilicidade, formação de pontes de hidrogênio ou propriedades hidrofóbicas, etc. de aminoácidos) para cumprir uma propriedade estrutural ou funcional específica de interesse. A capacidade de incorporar sítio-especificamente tais análogos de aminoácidos em proteínas iria expandir a nossa capacidade de racional e sistematicamente manipular as estruturas das proteínas, tanto para investigar a função da proteína como para criar proteínas com novas propriedades. Por exemplo, a capacidade de sintetizar grandes quantidades de proteínas contendo átomos pesados facilitaria a determinação da estrutura da proteína, e a capacidade de substituir sítio-especificamente fluoróforos ou grupos foto-cliváveis nas proteínas em células vivas iria fornecer ferramentas poderosas para estudar funções de proteínas *in vivo*.

Nos últimos anos, vários laboratórios investigaram uma expansão no número de aminoácidos geneticamente codificados, usando um tRNA supressor *nonsense* ou supressor de *frame-shift* para incorporar aminoácidos não-canônicos em proteínas em resposta a codões âmbar ou de quatro bases, respetivamente (Bain et al., J. Am. Chem. Soc. 111: 8013, 1989; Noren et al., Science 244: 182, 1989; Furter, Protein Sci. 7: 419, 1998; Wang et al., Proc. Natl. Acad. Sci. U.S.A., 100: 56, 2003; Hohsaka et al., FEBS Lett. 344: 171:

1994; Kowal and Oliver, *Nucleic Acids Res.* 25: 4685, 1997). Tais métodos inserem aminoácidos não-canônicos em posições de codão que terminarão normalmente a síntese do péptido de tipo selvagem (por exemplo, um codão stop ou uma mutação *frame-shift*). Estes métodos funcionaram bem para inserção de local único de aminoácidos novos. No entanto, a sua utilidade na substituição ou incorporação específica de posição multi-local (versus específica de resíduo) é limitada por eficiências de supressão modestas (20-60%) (Anderson et al., *J. Am. Chem. Soc.* 124: 9674, 2002; Bain et al., *Nature* 356: 537, 1992; Hohsaka et al., *Nucleic Acids Res.* 29: 3646, 2001). Isto é assim parcialmente porque uma eficiência elevada de supressão de codão stop irá interferir com a terminação normal da tradução de algumas proteínas não-visadas no organismo. Por outro lado, uma eficiência de supressão baixa provavelmente será insuficiente para suprimir mais do que um local de mutação *nonsense* ou *frame-shift* na proteína alvo, de tal modo que se torna cada vez mais difícil ou impraticável sintetizar uma proteína alvo completa incorporando cada vez mais aminoácidos não-canônicos.

Eficiente incorporação em múltiplos locais tem sido alcançada pela substituição de aminoácidos naturais em estirpes auxotróficas de *Escherichia coli*, por exemplo, usando aminoacil-tRNA sintetases com especificidade de substrato relaxada ou atividade de edição alterada (Wilson and Hatfield, *Biochim. Biophys. Acta* 781: 205, 1984; Kast and Hennecke, *J. Mol. Biol.* 222: 99, 1991; Ibba et al., *Biochemistry* 33: 7107, 1994; Sharma et al., *FEBS Lett.* 467: 37, 2000; Tang and Tirrell, *Biochemistry* 41: 10635, 2002; Datta et al., *J. Am. Chem. Soc.* 124: 5652, 2002; Doring et al., *Science* 292: 501, 2001). Embora este método forneça

eficiente incorporação de análogos em múltiplos locais, sofre da limitação de que o aminoácido novo deve “partilhar” codões com um dos aminoácidos naturais. Assim, para qualquer posição de determinado codão onde aminoácidos tanto naturais como novos podem ser inseridos, além de uma probabilidade de incorporação, há relativamente pouco controlo sobre qual aminoácido vai acabar por ser inserido. Isto pode ser indesejável, uma vez que para uma enzima ou proteína construída por engenharia, incorporação não-canónica de aminoácido num local não intencional pode inesperadamente comprometer a função da proteína, enquanto falhando a incorporação do aminoácido não-canónico no local designado irá impedir alcançar o objetivo do projeto.

Em geral, métodos de substituição em múltiplos locais são relativamente simples de realizar, mas todos os locais correspondentes a um determinado aminoácido natural em toda a proteína são substituídos. A extensão da incorporação do aminoácido natural e não-natural também pode variar. Além disso, a incorporação de análogos em múltiplos locais muitas vezes resulta em toxicidade quando as células são utilizadas, o que torna difícil estudar a proteína mutante em células vivas. A presente invenção supera estes obstáculos permitindo mutação sítio-específica de aminoácidos em proteínas.

Certas formas de realização divulgadas neste documento fornecem uma nova técnica para a incorporação de aminoácidos de substituição, incluindo aminoácidos naturais, ou aminoácidos não-padrão ou não-canónicos, em proteínas, que se baseia em quebrar a degeneração do código genético. Especificamente, certas formas de realização neste documento permitem a substituição ou incorporação

específica de posição de alta fidelidade de aminoácidos não-naturais em proteínas.

US2004/0053390 (Datta et al.) divulga métodos, reagentes, e ferramentas computacionais para desenhar análogos não-naturais de substratos para enzimas, especialmente para desenhar análogos de aminoácidos antinaturais para aminoacil-tRNA sintetases, tais como a Phe tRNA sintetase.

BREVE SUMÁRIO DA INVENÇÃO

Certas formas de realização divulgadas neste documento fornecem composições de componentes usados na maquinaria de biossíntese de proteínas, que incluem moléculas mutantes externas de aminoacil-tRNA sintetase (AARS) opcionalmente emparelhadas com moléculas mutantes externas de aminoacil-tRNA.

São também fornecidos métodos para gerar e selecionar tRNAs mutantes externos, aminoacil-tRNA sintetases mutantes externas, e pares dos mesmos que são capazes de incorporar aminoácidos, incluindo aminoácidos não-naturais, em polipéptidos ou proteínas. Certas composições de formas de realização específicas incluem pares de novo tRNA mutante externo ou aminoacil-tRNA sintetase mutante externa. As moléculas de novo tRNA mutante externo, moléculas AARS ou pares AARS-tRNA podem ser usados para incorporar um aminoácido antinatural num polipeptídeo *in vitro* e *in vivo*. Outras formas de realização da invenção incluem selecionar pares mutantes externos.

Composições da presente invenção incluem uma aminoacil-tRNA sintetase mutante externa, onde a tRNA sintetase mutante

externa preferencialmente aminoacila um tRNA mutante externo com um aminoácido antinatural, opcionalmente, *in vivo*. Numa forma de realização um ácido nucleico ou polinucleótido codificando uma sintetase mutante externa é fornecido, ou uma sequência complementar de ácidos nucleicos dos mesmos.

Portanto, certas formas de realização incluem uma composição compreendendo um primeiro vetor contendo um polinucleótido codificando uma fenilalanina-tRNA aminoacil sintetase (PheRS) modificada de levedura, em que o polinucleótido referido é mutado nos códons codificando o aminoácido em posições de sequência selecionadas do grupo consistindo dos números de posição de sequência aminoacídica (i) 412 e 415; (ii) 415 e 418; (iii) 415 e 437; (iv) 412, 415 e 437; (v) 415, 418 e 437; (vi) 412, 415 e 418; e (vii) 412, 415, 418 e 437 do polipéptido codificado pela SEQ ID NO:3 e em que a sintetase modificada referida é capaz de carregar uma molécula de tRNA com um aminoácido não-natural.

Pelo menos uma forma de realização compreende um segundo vetor contendo um polinucleótido codificando uma molécula de tRNA. Em pelo menos uma forma de realização, tais primeiro e segundo vetores são o mesmo vetor. Noutras formas de realização, tais primeiro e segundo vetores são vetores diferentes.

Em pelo menos uma forma de realização, o tRNA é endógeno, e em pelo menos uma forma de realização, o tRNA é modificado. Em pelo menos uma forma de realização, o tRNA é modificado de tal modo que contém um anticodão mutante que emparelha com um codão degenerado *wobble* correspondente com uma

afinidade maior do que a afinidade do tRNA natural. Em certas formas de realização, a AARS e o tRNA são dos mesmos ou diferentes organismos. Em pelo menos uma forma de realização, o aminoácido não-natural é selecionado do grupo consistindo de: azidonorleucina, 3-(1-naftil)alanina, 3-(2-naftil)alanina, p-etinil-fenilalanina, p-propargli-óxi-fenilalanina, m-etinil-fenilalanina, 6-etinil-triptofano, 5-etinil-triptofano, ácido (R)-2-amino-3-(4-etinil-1H-pirol-3-il)propânico, p-bromofenilalanina, p-idiofenilalanina, p-azidofenilalanina, 3-(6-cloroindolil)alanina, 3-(6-bromoindolil)alanina, 3-(5-bromoindolil)alanina, azidohomoalanina, e p-clorofenilalanina.

Outra forma de realização compreende um polipéptido compreendendo uma fenilalanina-tRNA sintetase (PheRS) modificada de levedura, em que a sintetase modificada referida é mutada nas posições de sequência aminoacídica selecionadas do grupo consistindo dos números de posições de sequência aminoacídica (i) 412 e 415; (ii) 415 e 418; (iii) 415 e 437; (iv) 412, 415 e 437; (v) 415, 418 e 437; (vi) 412, 415 e 418; e (vii) 412, 415, 418 e 437 do polipéptido codificado pela SEQ ID NO:3.

Certas formas de realização incluem um sistema de tradução compreendendo um polinucleótido codificando uma fenilalanina-tRNA sintetase modificada de levedura, em que o polinucleótido da sintetase modificada referida é mutada em um ou mais codões codificando o aminoácido em posições de sequência selecionadas do grupo consistindo dos números de posições de sequência aminoacídica (i) 412 e 415; (ii) 415 e 418; (iii) 415 e 437; (iv) 412, 415 e 437; (v) 415, 418 e 437; (vi) 412, 415 e 418; e (vii) 412, 415, 418 e 437

do polipéptido codificado pela SEQ ID NO:3 e em que a sintetase modificada referida é capaz de carregar uma molécula de tRNA com um aminoácido não-natural. Em pelo menos uma forma de realização, o sistema compreende uma célula hospedeira. Em pelo menos uma forma de realização, a tRNA aminoacil sintetase modificada é derivada de um organismo diferente da célula hospedeira. Noutra forma de realização, o sistema de tradução compreende um polinucleótido codificando uma molécula de tRNA modificada.

Em certas formas de realização, a molécula de tRNA modificada é derivada de um organismo diferente da célula hospedeira. Em certas formas de realização, a molécula de tRNA modificada é derivada de uma célula eucariótica e a célula hospedeira é uma célula procariótica. Ainda noutras certas formas de realização, a célula é uma auxotrófica.

Em certas formas de realização, o sistema de tradução compreende adicionalmente um meio de cultura contendo um ou mais aminoácidos não-naturais. Noutras formas de realização, o um ou mais aminoácidos não-naturais referidos são selecionados do grupo consistindo de: azidonorleucina, 3-(1-naftil)alanina, 3-(2-naftil)alanina, p-etinil-fenilalanina, p-propargli-óxi-fenilalanina, m-etinil-fenilalanina, 6-etinil-triptofano, 5-etinil-triptofano, ácido (R)-2-amino-3-(4-etinil-1H-pirol-3-il)propânico, p-bromofenilalanina, p-idiofenilalanina, p-azidofenilalanina, 3-(6-cloroindolil)alanina, 3-(6-bromoindolil)alanina, 3-(5-bromoindolil)alanina, azidohomoalanina, e p-clorofenilalanina.

Outras formas de realização referem-se a um método para incorporar um aminoácido não-natural num polipéptido alvo

em uma ou mais posições especificadas, o método compreendendo as etapas de: fornecer um sistema de tradução contendo pelo menos um aminoácido não-natural; fornecer ao sistema de tradução uma ou mais fenilalanil-tRNA sintetase modificada de levedura conforme definido acima, fornecer ao sistema de tradução um polinucleótido codificando um polipéptido alvo de interesse, e, permitir a tradução do polipéptido alvo de interesse, deste modo incorporando pelo menos um aminoácido não-natural no polipéptido alvo, em que as etapas são efetuadas em qualquer ordem.

Em certas formas de realização, tal sistema de tradução compreende uma célula. Em algumas formas de realização, o aminoácido não-natural é fornecido por contactar o sistema de tradução referido com uma solução contendo o aminoácido não-natural. Em pelo menos uma forma de realização, a constante de especificidade (k_{cat}/K_M) para a ativação do aminoácido não-natural referido pela AARS modificada referida é pelo menos 5 vezes maior do que aquela para o aminoácido natural referido. Em certas formas de realização, a AARS modificada carrega incorretamente um tRNA a uma taxa de não mais de 1%, 2%, 3%, 4%, 5%, 6%, 7% ou 8%. Ainda noutras formas de realização, o tRNA é um tRNA modificado. Em certas formas de realização, o primeiro polinucleótido referido ou o segundo polinucleótido referido compreende adicionalmente uma sequência de um promotor constitutivamente ativo ou indutível que controla a expressão do tRNA ou PheRS. Em pelo menos uma forma de realização, o método compreende adicionalmente a etapa de triagem de células que contêm uma PheRS modificada. Noutra forma de realização, o método compreende adicionalmente a etapa de verificar a incorporação do aminoácido não-

natural. Ainda outras formas de realização compreendem um polipéptido produzido pelo método divulgado.

Certas formas de realização divulgadas neste documento incluem um método para incorporar pelo menos um aminoácido não-natural num polipéptido alvo em um ou mais locais especificados, o método compreendendo fornecer um sistema de tradução contendo pelo menos um aminoácido não-natural; fornecer ao sistema de tradução um ou mais PheRS modificadas; fornecer ao sistema de tradução um polinucleótido codificando um polipéptido alvo de interesse; e que permite a tradução de interesse, deste modo incorporando pelo menos um aminoácido não-natural no polipéptido alvo. Certas formas de realização incluem um polipéptido produzido por este método.

BREVE DESCRIÇÃO DAS VÁRIAS PERSPETIVAS DOS DESENHOS

A Figura 1 apresenta um alinhamento de sequências das variantes de PheRS de sequências conservadas de *Thermus thermophilus*, *Escherichia coli*, e *Saccharomyces cerevisiae*.

A Figura 2 apresenta vários aminoácidos exemplares (naturais ou não-naturais) usados para algumas formas de realização divulgadas neste documento.

A Figura 3 apresenta a sequência aminoacídica de um polipéptido exemplar usado para algumas formas de realização divulgadas neste documento, dihidrofolato redutase (DHFR). Quatro fragmentos proteolíticos de péptidos (denominados Péptido A, Péptido B, Péptido C e Péptido D) foram usados para análises de MALDI e

cromatografia líquida-espetrometria de massa/espetrometria de massa (LC-MS/MS) como sublinhado.

A Figura 4 apresenta um MALDI-MS de fragmentos proteolíticos de péptidos derivados de mDHFR expressa em meio suplementado com (a) aminoácido 1 (3 mM); (b) aminoácido 7 (3 mM) e 1 (0,03 mM); (c) aminoácido 2 (3 mM) e 1 (0,03); (d) aminoácido 2 (3 mM) e 1 (0,03 mM). Triptofano não é suplementado durante a indução, exceto que 1 mM de triptofano é suplementado no meio em (c). O Péptido B, contendo um codão de Phe, é o controle. O Péptido A contém um codão âmbar (Z), o aminoácido que é atribuído com base nas unidades de massa para o péptido A em condições de expressão diferentes. Devido à incorporação de lisina com o Péptido A (c), lisina C-terminal foi clivada para produzir um Péptido A mais curto (NGDLPWPPLRNEK) (SEQ ID NO: 4).

Figura 5 Espectro de massa sequencial do Péptido A (NGDLPWPPLRNEZK) (SEQ ID NO: 5) derivado de DHFR expressa em meio suplementado com aminoácido 7 (3 mM) e aminoácido 1 (0,03 mM). A sequência parcial de WPPLRNE (SEQ ID NO: 6) e resíduo Z (correspondendo ao aminoácido 7) do Péptido A pode ser atribuída a partir das séries de ião y e b, respetivamente.

A Figura 6 apresenta MALDI-MS de fragmentos proteolíticos de péptidos derivados de mDHFR expressa em meio suplementado com (a) aminoácido 2 (3 mM) e aminoácido 6 (3 mM) e aminoácido 1 (0,03 mM) e aminoácido 13 (0,2 mM); (b) aminoácido 2 (3 mM) e aminoácido 6 (0,01 mM) e aminoácido 1 (0,03 mM) e aminoácido 13 (0,2 mM); (c) aminoácido 3 (3 mM) e

aminoácido 6 (0,01 mM) e aminoácido 1 (0,03 mM) e aminoácido 13 (0,01 mM). O Péptido C, com um codão de Phe, é o controlo.

A Figura 7 apresenta resultados de MALDI-MS de fragmentos proteolíticos de péptidos derivados de mDHFR expressa em meio suplementado com (a) aminoácido 9 (3 mM) e aminoácido 6 (0,03 mM) e aminoácido 1 (0,01 mM); (b) aminoácido 10 (3 mM) e aminoácido 6 (0,03 mM) e aminoácido 1 (0,03 mM); (c) aminoácido 11 (3 mM) e aminoácido 6 (0,01 mM) e aminoácido 1 (0,03 mM). O Péptido D foi o controlo.

A Figura 8 apresenta a aminoacilação de $\text{tRNA}^{\text{Phe}}_{\text{CUA}}$ (quadrado) e $\text{tRNA}^{\text{Phe}}_{\text{CUA\_UG}}$ (círculo) com lisina por eLysS.

A Figura 9 apresenta cromatogramas de LC-MS de digestões trípticas dos polipéptidos de mDHFR.

A Figura 10 apresenta taxas de transferência ATP-PPi para fenilalanina, triptofano e p-bromofenilalanina por PheRS de tipo selvagem de levedura e PheRS T415G mutante externa de levedura ou PheRS T415A mutante externa de levedura.

A Figura 11 apresenta aminoacilação de fenilalanina e triptofano por aminoacil-tRNA sintetase de tipo selvagem ou tRNA sintetase mutante externa (T415G ou T415A).

A Figura 12 apresenta incorporação de lisina (quadrado aberto), triptofano (linhas enviesadas) ou p-bromofenilalanina (quadriculado).

A Figura 13 apresenta espectros de massa de p-etinilfenilalanina incorporada num polipéptido usando uma tRNA sintetase modificada (T415G) e tRNA^{Phe} com um supressor âmbar numa célula hospedeira.

A Figura 14 apresenta um FACS de proteína verde fluorescente (GFP) da incorporação de aminoácidos usando codão âmbar de supressão numa proteína de teste.

A Figura 15 ilustra um mapeamento de plasmídeo exemplar de uma aminoacil-tRNA sintetase mutante (T415G).

A Figura 16 ilustra mutações exemplares efetuadas numa fenilalanina-tRNA sintetase de levedura.

DESCRIÇÃO PORMENORIZADA DA INVENÇÃO

As proteínas são as vias de virtualmente todos os processos biológicos, da fotossíntese e visão à transdução de sinal e resposta imune. Modificar proteínas ou polipéptidos para incluir aminoácidos não-naturais tem um grande potencial para o uso na terapêutica humana, agricultura, biocombustíveis e outras áreas.

Aminoacil-tRNA sintetases catalisam a reação de aminoacilação para a incorporação de aminoácidos em proteínas através das correspondentes moléculas de RNA de transferência. A manipulação precisa de atividade de sintetase pode alterar a especificidade de aminoacilação para ligar estavelmente aminoácidos não-canónicos ao tRNA pretendido. Depois, através da interação codão-anticodão entre RNA mensageiro (mRNA) e o tRNA, os análogos de aminoácidos podem ser entregues a uma cadeia de

polipéptidos crescente. Portanto, a incorporação de aminoácidos não-naturais em proteínas baseia-se na manipulação de especificidade de aminoácidos de aminoacil-tRNA sintetases (AARS).

Aminoacil-tRNA sintetases funcionam para transformar as sequências do código genético em proteínas biologicamente funcionais através de uma reação de aminoacilação de duas etapas. Como uma etapa inicial, o aminoácido cognato é ativado pela AARS na presença de ATP para formar o aminoácido adenilato; posteriormente AARS catalisa a reação de esterificação para juntar o aminoácido ao 2'- ou 3'-OH do ribonucleótido terminal do seu tRNA cognato. Assim que a reação de aminoacilação ocorre, o aminoácido é dirigido para a cadeia de polipéptido crescente pelo tRNA carregado.

Certas formas de realização divulgadas neste documento referem-se a moléculas PheRS mutantes ou modificadas que foram mutadas ou modificadas de tal modo que as enzimas são capazes de carregar uma molécula de tRNA com um aminoácido de substituição, preferencialmente um ácido aminado não-natural devido à interrupção entre a sintetase e o correspondente aminoácido natural.

Por exemplo, a interrupção pode ser devida a interferir com o emparelhamento de bases de Watson-Crick, interferir com emparelhamento de bases wobble, ou a criação de novo wobble ou outro emparelhamento de base.

Algumas formas de realização referem-se a um polinucleótido codificando um tRNA mutante ou modificado de um tRNA para um aminoácido natural, em que o aminoácido natural é codificado por um ou mais códons *wobble* degenerados, o tRNA

modificado compreende uma sequência de anticodão modificada que forma emparelhamento de Watson-Crick com um dos codões degenerados *wobble*. Preferencialmente, o tRNA modificado não é ou é apenas ineficientemente carregado por uma aminoacil-tRNA sintetase (AARS) endógena para o aminoácido natural. Numa forma de realização, múltiplas moléculas de PheRS modificadas podem ser usadas com uma ou mais moléculas de tRNA. Numa forma de realização, uma ou mais moléculas de PheRS modificadas ou mutadas podem ser usadas com um ou mais moléculas de tRNA nativo, enquanto noutra forma de realização uma PheRS modificada ou mutada pode ser usada com uma molécula de tRNA modificada ou mutada. Em certas formas de realização, um ou mais pares de moléculas de tRNA/PheRS modificadas podem ser utilizadas. Em certas formas de realização, pares heterólogos podem ser usados. Em certas formas de realização, uma ou mais PheRS e/ou tRNA modificadas ou mutadas podem ser derivadas do mesmo ou diferente organismo.

Em algumas formas de realização exemplares, uma PheRS particular pode utilizar vários métodos para a incorporação de um aminoácido de substituição (incluindo um aminoácido não-natural) num polipéptido ou proteína. Por exemplo, uma única PheRS pode utilizar um codão *nonsense* (tal como um codão stop âmbar) para a incorporação de um aminoácido de substituição (tal como um aminoácido não-natural) num local específico no polipéptido. Além ou em vez disso, um codão *wobble* (tal como UUU) poderia ser usado para a incorporação do aminoácido de substituição no local do codão *wobble*.

Noutras formas de realização exemplares, múltiplos aminoácidos de substituição (tal como dois aminoácidos não-naturais diferentes) podem ser incorporados num polipéptido

ou proteína através do uso de vários métodos. Por exemplo, um aminoácido não-natural pode ser incorporado num local *wobble*, enquanto um aminoácido não-natural diferente pode ser incorporado num codão stop âmbar. Em algumas formas de realização, incorporação de múltiplos aminoácidos de substituição (incluindo aminoácidos não-naturais) inclui utilizar uma PheRS modificada para incorporar múltiplos análogos de fenilalanina diferentes, tais como bromofenilalanina e/ou p-iodofenilalanina, no mesmo polipéptido ou proteína.

Uma abordagem semelhante envolve o uso de uma sintetase heteróloga e um tRNA iniciador mutante do mesmo organismo ou de um organismo relacionado como uma molécula de tRNA. (Ver, por exemplo, Kowal, et al., PNAS, 98, 2268 (2001)).

Em algumas formas de realização, a RS modificada ou mutada interage com a substituição de aminoácido desejado (quer seja um aminoácido natural ou não-natural) com uma especificidade de ligação alterada e/ou evento catalítico alterado da enzima para a substituição de aminoácido quando comparado à enzima RS tipo selvagem ou aminoácido de tipo selvagem correspondente.

Em cinética enzimática, K_{cat} é uma constante de velocidade de primeira ordem correspondendo à etapa ou etapas mais lentas na via catalítica geral. A K_{cat} representa o número máximo de moléculas de substrato que podem ser convertidas em produto por molécula de enzima por unidade de tempo (o que ocorre se a enzima está "saturada" com substrato), e portanto é frequentemente referida como o número de *turnover*. O K_m é uma constante de dissociação aparente e está relacionada com a afinidade da enzima para o

substrato; é o produto de todas as constantes de dissociação e equilíbrio antes da primeira etapa irreversível na via. Frequentemente, é uma medida próxima da constante de dissociação enzima-substrato. O k_{cat}/K_m é uma constante de velocidade de segunda ordem que se refere à enzima livre (não complexo enzima-substrato) e é também uma medida da eficiência global de catálise da enzima e é também referida como a constante de especificidade.

Em certas formas de realização, a sintetase mutante externa tem propriedades enzimáticas melhoradas ou reforçadas, p.ex., a K_m é superior ou inferior, a k_{cat} é superior ou inferior, o valor de k_{cat}/K_m é superior ou inferior ou semelhante, para o aminoácido antinatural em comparação com um aminoácido natural, p.ex., um dos 20 aminoácidos conhecidos. A K_m da AARS mutante ou modificada é preferencialmente igual ou inferior para o aminoácido não-natural do que para o correspondente aminoácido natural de tipo selvagem.

Em certas formas de realização, os valores de k_{cat}/K_m da variante RS podem variar entre 3 ordens de grandeza, 5 ordens de grandeza, 10 ordens de grandeza, 25 ordens de grandeza, 50 ordens de grandeza, 100 ordens de grandeza, 150 ordens de grandeza, 200 ordens de grandeza, 250 ordens de grandeza, 300 ordens de grandeza, 350 ordens de grandeza, 385 ordens de grandeza, 400 ordens de grandeza superior do que para o aminoácido natural.

Em certas formas de realização, o tRNA modificado interage com o codão degenerado *wobble* com uma afinidade a 37°C de pelo menos cerca de 1,0 kcal/mole, 1,5 kcal/mole, 2,0 kcal/mole, 2,5 kcal/mole, 3,0 kcal/mole, 3,5 kcal/mole, 4,0

kcal/mole, 4,5 kcal/mole, 5,0 kcal/mole ou maior (ou qualquer valor entre estes) favoravelmente relativamente à interação entre a sua versão não modificada e o codão degenerado *wobble*.

Por exemplo, fenilalanina (Phe) é codificada por dois codões, UUC e UUU. Ambos os codões são lidos por um único tRNA, que é equipado com a sequência anticodão GAA. O codão UUC é portanto reconhecido através de emparelhamento de bases de Watson-Crick padrão entre codão e anticodão; UUU é lido através de uma par de bases *wobble* G-U na primeira posição do anticodão (Crick, J. Mol. Biol. 19: 548, 1966; Soll and RajBhandary, J. Mol. Biol. 29:113, 1967). Desnaturação térmica de cadeias duplas de RNA produziu estimativas das energias livres de Gibbs de derretimento dos pares de bases G-U, G-C, A-U e A C como 4,1, 6,3, 6,5 e 2,6 kcal/mole, respetivamente, a 37°C. Assim, o par de bases *wobble*, G-U, é menos estável do que o par de bases de Watson-Crick, A-U. Um tRNA^{Phe} modificado, equipado com o anticodão AAA (tRNA^{Phe}_{AAA}) foi projetado para ler o codão UUU e foi previsto ler tais codões mais rápido do que o tRNA^{Phe}_{AAA} de tipo selvagem.

Em algumas formas de realização, a cavidade de ligação da RS é modificada de modo a que a RS modificada exiba uma preferência para o aminoácido não-natural relativamente ao aminoácido natural correspondente. Em formas de realização preferidas, a RS é modificada em um ou mais codões necessários para o contato estrutural entre a RS e o aminoácido sendo transferido para o tRNA. Em certas formas de realização, o um ou mais codões selecionados para mutação ou modificação são selecionados por meio de modelação de computador. Enquanto qualquer RS pode ser

modificada, a presente invenção refere-se a fenilalanina-tRNA sintetase (PheRS). A RS modificada é de *Saccharomyces cerevisiae*, ou outra célula de levedura. Em certas formas de realização em que a RS é uma PheRS, a enzima tem uma ou mais mutações pontuais como definido acima, em particular mutação pontual (N412G), (T415G), (T415A), (S418C), ou (S437F) na subunidade alfa da enzima. As mutações pontuais (por exemplo, a mutação de T para G ou A na posição 415, ou mutação de S para C na posição 418, ou a mutação de N para G na posição 412, ou a mutação de S para F na posição 437) estão localizadas na região da cavidade de ligação da aminoacil-tRNA sintetase (RS).

Em algumas formas de realização exemplares, valores de K_m típicos para diferentes análogos com AARS podem variar entre aproximadamente 15 microM, 20 microM, 30 microM, 50 microM, 75 microM, 100 microM, 150 microM, 200 microM, 300 microM, 400 microM, 440 microM, 500 microM, 1000 microM, 1500 microM, 2000 microM, 3000 microM, 4000 microM, 5000 microM, 6000 microM, ou superior ou qualquer valor entre estes.

Da mesma forma, valores da K_{cat} da PheRS mutante são preferencialmente iguais ou maiores para o análogo de aminoácido do que para o aminoácido natural. Por exemplo, valores K_{cat} para diferentes análogos com a correspondente PheRS podem variar entre aproximadamente $0,002 \text{ sec}^{-1}$, $0,0018 \text{ sec}^{-1}$, $0,0015 \text{ sec}^{-1}$, $0,014 \text{ sec}^{-1}$, $0,1 \text{ sec}^{-1}$, $0,3 \text{ sec}^{-1}$, 1 sec^{-1} , 3 sec^{-1} , 5 sec^{-1} , 8 sec^{-1} , 10 sec^{-1} , $13,3 \text{ sec}^{-1}$, 15 sec^{-1} , ou superiores.

Portanto, o k_{cat}/K_m da PheRS mutante é idealmente igual a ou superior para o análogo do aminoácido do que para o

aminoácido natural de tipo selvagem. Valores k_{cat}/K_m típicos podem variar entre aproximadamente $0.0001 \text{ M}^{-1} \text{ s}^{-1}$, $0.0003 \text{ M}^{-1} \text{ s}^{-1}$, $0,005 \text{ M}^{-1} \text{ s}^{-1}$, $0,05 \text{ M}^{-1} \text{ s}^{-1}$, $0,5 \text{ M}^{-1} \text{ s}^{-1}$, $0,547 \text{ M}^{-1} \text{ s}^{-1}$, $1 \text{ M}^{-1} \text{ s}^{-1}$, $5 \text{ M}^{-1} \text{ s}^{-1}$, $10 \text{ M}^{-1} \text{ s}^{-1}$, $20 \text{ M}^{-1} \text{ s}^{-1}$, $30 \text{ M}^{-1} \text{ s}^{-1}$, $32 \text{ M}^{-1} \text{ s}^{-1}$, $500 \text{ M}^{-1} \text{ s}^{-1}$, $600 \text{ M}^{-1} \text{ s}^{-1}$, $1000 \text{ M}^{-1} \text{ s}^{-1}$, $5000 \text{ M}^{-1} \text{ s}^{-1}$, $11000 \text{ M}^{-1} \text{ s}^{-1}$.

Em certas formas de realização, as mutações pontuais da PheRS podem ser alteradas para qualquer aminoácido, dependendo das características do aminoácido não-natural desejado para a incorporação na proteína/polipéptido de teste. Em certas formas de realização, um aminoácido na cavidade de ligação da PheRS pode ser mutado para um códon para um aminoácido com uma cadeia lateral pequena, um aminoácido com uma cadeia lateral alifática, um aminoácido cíclico, um aminoácido com cadeias laterais contendo hidroxilo ou enxofre, um aminoácido aromático, um aminoácido básico, um aminoácido ácido (ou amida). Seleção de aminoácidos para mutar a PheRS num ponto específico é rotina, dependendo do resultado desejado e aminoácido não-natural desejado a ser incorporado no polipéptido/proteína alvo ou de teste. Por exemplo, se o objetivo é alargar a cavidade de ligação da molécula de PheRS, então um aminoácido com uma cadeia lateral alifática, ou uma cadeia lateral pequena, poderia ser escolhida para mutar a PheRS. Noutros casos, se é desejada uma cavidade de ligação que engloba uma cavidade carregada, então um aminoácido básico ou ácido pode ser selecionado para mutação pontual da PheRS.

Em algumas formas de realização, a RS modificada pode ser usada num sistema de tradução, incluindo uma célula hospedeira auxotrófica ou célula hospedeira prototrófica juntamente com um tRNA supressor a fim de permitir a

atribuição de um codão stop (tais como um codão *nonsense* âmbar, ocre ou opala, um codão stop que não está presente num organismo particular, qualquer codão *nonsense*, um codão de quatro ou cinco pares de base, ou outro aminoácido natural que não está presente em níveis significativos na proteína, tal como metionina) para incorporar um outro aminoácido, incluindo um aminoácido análogo. Portanto, as enzimas RS podem ser "reprogramadas" para a especificidade de substrato promíscua para facilitar a incorporação de um aminoácido não-natural num polipéptido de forma sítio-específica. Em formas de realização particulares, qualquer aminoácido aromático não-natural pode ser utilizado com a PheRS modificada. Esta reprogramação permite a incorporação de alta fidelidade de um aminoácido, incluindo aminoácidos não-naturais, em polipéptidos ou proteínas com ou sem o uso de células hospedeiras auxotróficas.

Reprogramar uma enzima AARS pode envolver análise estrutural ou bioquímica, incluindo de modelação de computador ou alinhamento de sequências. Uma vez que existem informações de sequência disponíveis para muitas moléculas AARS, por exemplo em GenBank, comparar os alinhamentos de sequências é um procedimento de rotina assim que a sequência particular da região de interesse é determinada.

O uso de células hospedeiras auxotróficas pode aumentar o nível de incorporação de aminoácidos não-naturais, ou diminuir o nível de incorporação incorreta de outro aminoácido no lugar do aminoácido não-natural desejado. Por exemplo, após melhorar a reatividade celular de aminoacilação através da expressão de PheRS de tipo selvagem no hospedeiro, nós surpreendentemente descobrimos

que alguns dos aminoácidos análogos lentos também poderiam ser introduzidos em proteínas, mesmo na ausência de uma célula hospedeira auxotrófica.

Em certas formas de realização, a expressão de uma ou mais moléculas de PheRS modificadas ou mutantes, uma ou mais moléculas de tRNA modificadas ou mutantes, ou ambas, pode ser regulada por um promotor constitutivo ou indutível ou outro sistema de expressão indutível.

Certas formas de realização divulgadas neste documento referem-se a permitir inserção sítio-seletiva de um ou mais ácidos aminados não-naturais em qualquer posição desejada de qualquer proteína, (ii) são aplicáveis tanto a células procariotas como eucariotas, e permite estudos *in vivo* de proteínas mutantes em adição à produção de grandes quantidades de proteínas mutantes purificadas. Além disso, certas formas de realização referem-se à adaptação para incorporar um qualquer de uma grande variedade de aminoácidos antinaturais em proteínas *in vivo*. Portanto, numa sequência específica do polipéptido um número de diferentes inserções sítio-seletivas de aminoácidos antinaturais é possível. Tais inserções são opcionalmente todas do mesmo tipo (p.ex., múltiplos exemplos de um tipo de aminoácido antinatural inserido em múltiplos pontos num polipéptido) ou são opcionalmente de diversos tipos (p.ex., diferentes tipos de aminoácido antinaturais são inseridos em múltiplos pontos num polipéptido).

Um resultado surpreendente divulgado neste documento mostra que o re-desenho do sítio sintético de uma enzima PheRS pode expandir a capacidade de introduzir aminoácidos de substituição (incluindo aminoácidos não-naturais) em

polipéptidos ou proteínas. Em algumas formas de realização, as composições e os métodos de modificar ou mutar PheRS podem incluir opcionalmente alterar a função *editing* da PheRS modificada ou mutante. Em algumas formas de realização, a capacidade editora ou de *proofreading* da PheRS modificada ou mutante é aproximadamente igual à da PheRS de tipo selvagem (inalterada). Em algumas formas de realização, a capacidade editora ou de *proofreading* da PheRS modificada ou mutante é reduzida. Ainda em algumas formas de realização, a capacidade de *editing* ou de *proofreading* da PheRS modificada ou mutante é eliminada. Em algumas certas formas de realização, a alteração da função de *editing* ou *proofreading* da PheRS é inerente à modificação da PheRS a fim de acomodar um aminoácido de substituição (por exemplo, modificação da cavidade de ligação da PheRS). Noutras formas de realização, a alteração da função de *editing* ou de *proofreading* pode ser realizada em adição à modificação da PheRS a fim de acomodar um aminoácido de substituição. Ainda noutras formas de realização, a função de *editing* ou *proofreading* da PheRS é inalterada. Portanto, além de modificação ou alteração da cavidade de ligação de uma PheRS, o domínio de *editing* ou *proofreading* da PheRS modificada pode também ser alterada para permitir maior especificidade de aminoacilar o aminoácido de substituição (incluindo aminoácido não-natural) a um tRNA (quer seja um tRNA endógeno quer seja mutante externo), enquanto hidrolisando opcionalmente o(s) aminoácido(s)-adenilato de tipo selvagem que pode(m) formar-se e resultando em maior fidelidade ou especificidade de incorporação do aminoácido de substituição (incluindo um aminoácido não-natural) no polipéptido ou proteína.

Em algumas formas de realização, as taxas de incorporação de um aminoácido não-natural foram aproximadamente 65% ou maior, 70% ou maior, 75% ou maior, 80% ou maior, 85% ou maior, 90% ou maior, 91% ou maior, 92% ou maior, 93% ou maior, 94% ou maior, 95% ou maior, 96% ou maior, 97% ou maior, maior, ou 99% ou 98% ou maior utilizando uma RS modificada.

Conforme divulgado neste documento, nomeadamente, a estrutura cristalina da RS exata a ser modificada, ou a estrutura cristalina de uma RS homóloga pode ser usada para modelação molecular da interação enzima-aminoácido a fim de determinar os pontos de contacto entre a RS e o correspondente aminoácido natural, e/ou os pontos de contacto entre a RS e o aminoácido não-natural selecionado desejado ser incorporado num polipéptido. Por exemplo, em certas formas de realização neste documento, a estrutura cristalina da PheRS de *Thermus thermophilus* complexada com fenilalanina foi usada para o projeto de modelação molecular de uma PheRS de *Saccharomyces cerevisiae*, devido à identidade de sequência de cerca de 40% na região de sítio ativo das sintetases. Mutação da Treonina na posição 415 para Glicina ou Alanina (T415G ou T415A, respetivamente) aumentou o sítio ativo e possibilitou o alojamento de análogos de fenilalanina maiores. Mutações como esta que interrompem o emparelhamento de bases de Watson-Crick com o aminoácido natural designado para uma determinada RS permitem maior especificidade para a incorporação de um aminoácido não-natural e diminuição da incorporação incorreta de outro aminoácido. Conforme estabelecido nos Exemplos e Figuras, a fenilalanina-aminoacil-tRNA sintetase (PheRS) (T415A) de levedura revelou uma preferência de 5 ordens de grandeza maior para

bromofenilalanina do que para fenilalanina natural. Portanto, é possível alterar uma molécula de aminoacil-tRNA sintetase (RS) para preferencialmente incorporar um aminoácido desejado 2 ordens de grandeza, 3 ordens de grandeza, 4 ordens de grandeza, 5 ordens de grandeza, 6 ordens de grandeza, 7 ordens de grandeza, 8 ordens de grandeza, 9 ordens de grandeza, 10 ordens de grandeza, ou mais, dependendo das propriedades da RS particular e aminoácido desejado (incluindo aminoácido não-natural).

Numa forma de realização particular, foi descoberto que o espaço em torno da posição para do anel arilo de Phe ligada poderia ser ligeiramente reduzido para excluir outros aminoácidos aromáticos, tais como Trp, enquanto ainda acomodando um aminoácido não-natural, tais como pBrF. Assim, uma forma de realização divulga a incorporação de um grupo funcional brometo de arilo num polipéptido numa posição programada fornecendo uma ligação quimio-seletiva através de acoplamento catalisado por paládio com parceiros de reação etino ou acetileno. Enquanto tais RSs reprogramadas ou modificadas podem ser usadas num qualquer número de células, incluindo células hospedeiras auxotróficas, o elevado nível de eficiência da incorporação do aminoácido desejado (ou análogo) das RSs modificadas de certas formas de realização, assim como altos rendimentos de produção de proteína, tornam o uso de células hospedeiras auxotróficas desnecessário.

Como a região do sítio ativo para quase todas as sintetases de aminoácidos é conhecida ou prontamente deduzida, uma tal técnica exemplar pode ser aplicada a outras AARSS num esforço para reprogramar a especificidade do aminoácido de

um aminoácido natural para um aminoácido não-natural com uma expectativa de sucesso e sem experimentação indevida.

Como um exemplo ilustrativo, a treonina na posição 415 na PheRS de levedura é o equivalente a treonina 251 na PheRS de *E. coli*. Assim, a mutação da PheRS de levedura (T415G) permitiu a ativação de uma variedade de análogos de Phe. (Ver Exemplos, neste documento). Mais mutações pontuais e/ou uso de uma célula hospedeira auxotrófica permitiram a diminuição da incorporação incorreta na variante PheRS T415G de levedura.

Noutra forma de realização particular divulgada neste documento, uma RNA de transferência mutante de levedura ($\text{ytRNA}^{\text{Phe}}_{\text{CUA}}$) do qual um emparelhamento de bases de Watson-Crick entre a posição do aminoácido 30 e posição do aminoácido 40, foi interrompida, foi carregado com p-bromo-fenilalanina (pBrF) por uma fenilalanina de levedura co-expressa. Em certas formas de realização, a cavidade de ligação de aminoácido da AARS constitui aproximadamente 200, aproximadamente 100, aproximadamente 75, aproximadamente 50, aproximadamente 25, aproximadamente 10, aproximadamente 5 ou mais ou menos aminoácidos. Em algumas formas de realização, os aminoácidos a serem mutados no sítio ativo ou de ligação são trechos contíguos de aminoácidos. Noutras formas de realização, os aminoácidos a serem mutados estão localizados próximos uns dos outros mas não são contíguos.

Em certas formas de realização, o aminoácido natural é codificado por dois ou mais códigos genéticos (portanto codificado por códigos genéticos degenerados). Na maioria, se não todos os casos, isto inclui 18 dos 20 aminoácidos

naturais, exceto Met e Trp. Nestas circunstâncias, para reconhecer todos os códigos genéticos degenerados para o aminoácido natural, o *loop* anticodão do(s) tRNA(s) de tipo selvagem baseia-se tanto em emparelhamento de bases *wobble* e emparelhamento de bases de Watson-Crick puro. O tRNA modificado em questão contém pelo menos uma modificação no seu *loop* de anticodão, de tal modo que o *loop* anticodão modificado forma agora emparelhamento de bases de Watson-Crick a um dos códigos genéticos degenerados, que o tRNA anteriormente se liga somente através de emparelhamento de bases *wobble*.

Uma vez que o emparelhamento de bases de Watson-Crick é invariavelmente mais forte e mais estável do que emparelhamento de bases *wobble*, o tRNA modificado em questão ligar-se-á preferencialmente a um código genético de emparelhamento de bases *wobble* anterior (agora através de emparelhamento de bases de Watson-Crick), em vez de emparelhamento de bases de Watson-Crick anterior (agora através de emparelhamento de bases *wobble*). Deste modo um análogo pode ser incorporado no codão em questão, se o tRNA modificado estiver carregado com um análogo de um aminoácido natural, que pode ou não ser o mesmo que o aminoácido natural codificado pelo codão em questão.

Portanto em certas formas de realização, se é desejável incorporar certos análogos de aminoácidos em codões para Met ou Trp, um tRNA para um aminoácido natural (p.ex., um tRNA Met, um tRNA Trp, ou mesmo um tRNA Phe, etc) pode ser modificado para reconhecer o codão Met ou Trp. Sob este tipo de situação única, tanto o tRNA modificado e o tRNA natural competem para ligar o mesmo (único) código genético através de emparelhamento de bases de Watson-Crick. Alguns,

mas não todos tais códons aceitarão os seus aminoácidos naturais, enquanto outros podem aceitar análogos de aminoácidos carregados pelo tRNA modificado. Outros fatores, tais como a abundância do aminoácido natural vs aquela dos análogos, podem afetar o resultado final. (Ver exemplos divulgados neste documento).

Em certas formas de realização preferidas, o tRNA modificado não é carregado ou é apenas ineficientemente carregado por uma aminoacil-tRNA sintetase (AARS) endógena para qualquer aminoácido natural, de tal modo que o tRNA modificado em grande parte (se não exclusivamente) carrega um análogo de aminoácido, mas não um aminoácido natural. Embora um tRNA modificado em questão ainda possa ser útil se puder ser carregado pela AARS endógena com um aminoácido natural.

Em certas formas de realização, o tRNA modificado carregado com um análogo de aminoácido tem uma forma geral e tamanho tais que o análogo-tRNA é um complexo aceitável pelo ribossoma, ou seja, o complexo tRNA-análogo pode ser aceite pelos ribossomas procariotas ou eucariotas num sistema de tradução *in vivo* ou *in vitro*.

Preferencialmente, a PheRS modificada especificamente ou preferencialmente carrega o análogo para o tRNA modificado relativamente a qualquer aminoácido natural. Numa forma de realização preferida, a constante de especificidade para a ativação do análogo pela AARS modificada (definido como k_{cat}/K_M) é igual ou maior do que pelo menos cerca de duas ordens de grandeza maior do que para o aminoácido natural, preferencialmente cerca 3 ordens de grandeza, 4 ordens de grandeza, 5 ordens de grandeza, 6 ordens de

grandeza, 7 ordens de grandeza, 8 ordens de grandeza, 9 ordens de grandeza, 10 ordens de grandeza ou mais do que isto para o aminoácido natural.

Em certas formas de realização, o tRNA modificado compreende adicionalmente uma mutação no quarto local estendido do anticodão para aumentar a eficiência de tradução.

O uso de codões estendidos é baseado em supressão *frame-shift* da tradução. Quatro codões base têm potencial para a inserção de múltiplos aminoácidos não-naturais na mesma proteína. Por exemplo, o quadruplete UAGA pode ser decodificado por um tRNA^{Leu} com um anticodão UCUA com uma eficiência de 13 a 26%. (Ver, por exemplo, Moore, et al., J. Mol. Biol., 298: 195 (2000)). O uso de codões estendidos apenas em si tem potenciais problemas, tais como a leitura correta do quadro de leitura das três primeiras bases como um triplete no codão estendido compete com a supressão *frame-shift* global. Em alguns casos, codões estendidos baseados em codões raros ou codões *nonsense* podem reduzir leitura *missense* e supressão *frame-shift* noutros locais indesejados. Estes problemas podem ser superados, no entanto, com o uso de um codão/anticodão estendido e uma AARS e/ou tRNA modificada/o conforme indicado em algumas formas de realização divulgadas neste documento.

Portanto, para resumir, codões específicos são reservados para uso em métodos divulgados neste documento pela PheRS mutante ou modificada e/ou tRNA mutante ou modificado para incorporação de um aminoácido de substituição (incluindo um aminoácido natural ou não-natural). Tais métodos podem incluir o uso de decodificação âmbar (ocre, opala, ou

outro tRNA supressor) que lê codões stop (TAG), descodificação enviesada que explora tRNAs não utilizados responsáveis pelo enviesamento de codão, decodificação *wobble*, que cria novos tRNAs que leem codões *wobble*, e codões estendidos (4-5 bases ou mais) que usam os tRNAs mutantes "supressores" que usam anticodões de 4 ou 5 bases (ou mais).

Pelo menos uma outra forma de realização fornece um método para incorporar um análogo de aminoácido numa proteína alvo em uma ou mais posições especificadas, o método compreendendo: (1) fornecer a um ambiente um primeiro polinucleótido em questão para um tRNA modificado, ou um tRNA modificado em questão; (2) proporcionar ao ambiente um primeiro polinucleótido em questão codificando uma PheRS modificada, no qual a PheRS modificada é capaz de carregar o tRNA modificado com o análogo; (3) proporcionar ao ambiente o análogo; (4) fornecer um molde de polinucleótido codificando a proteína alvo, em que o codão no molde de polinucleótido para a posição especificada apenas forma emparelhamento de bases de Watson-Crick com o tRNA modificado; e (5) permitir a tradução do molde de polinucleótido prosseguir, assim incorporando o análogo na proteína alvo na posição especificada, em que as etapas (1)-(4) são efetuadas em qualquer ordem.

Em certas formas de realização, o método compreende adicionalmente verificar a incorporação do análogo por, por exemplo, espectrometria de massa, sequenciação de proteínas, marcação de aminoácidos tais como por fluorescência, radioatividade, etc., ELISA, ou outra seleção por anticorpos, ensaios funcionais ou seleções, ou outros métodos.

Em certas formas de realização, o método incorpora o análogo na posição com uma eficiência pelo menos 50%, ou 60%, 70%, 80%, 90%, 95%, 99% ou quase 100%.

Definições

Antes de descrever certas formas de realização em detalhe, deve ser entendido que esta invenção não está limitada a composições ou sistemas biológicos particulares, que podem, é claro, variar. Deve também ser entendido que a terminologia usada neste documento tem o propósito de descrever formas de realização ilustrativas particulares apenas, e não se destina a ser limitativa. Como usado nesta descrição e reivindicações anexadas, as formas singular "um", "uma", "a" e "o" incluem referentes a plural a menos que o conteúdo claramente indique de outra forma. Portanto, por exemplo, a referência a "uma molécula" opcionalmente inclui uma combinação de duas ou mais tais moléculas, e semelhantes.

Salvo especificamente definido abaixo, os termos usados nesta descrição geralmente têm os seus significados comuns na arte, no contexto geral desta invenção e no contexto específico onde cada termo é usado. Certos termos são discutidos abaixo ou noutro lugar na descrição, para fornecer orientações adicionais ao utilizador descrevendo as composições e os métodos da invenção e como produzi-los e usá-los. O alcance e o significado de qualquer uso de um termo será aparente a partir do contexto específico em que o termo é usado.

"Cerca" e "aproximadamente" geralmente designam um grau aceitável de erro para a quantidade medida dada a natureza

ou a precisão das medições. Graus exemplares típicos de erro estão dentro de 20 por cento (%), preferencialmente dentro de 10%, e mais preferencialmente dentro de 5% de um determinado valor ou intervalo de valores. Alternativamente, e particularmente em sistemas biológicos, os termos "cerca" e "aproximadamente" podem significar valores que estão dentro de uma ordem de magnitude, preferencialmente dentro de 5 ordens de grandeza e mais preferencialmente 2 ordens de grandeza relativamente a um determinado valor. Quantidades numéricas apresentadas neste documento são aproximadas salvo indicação em contrário, o que significa que o termo "cerca" ou "aproximadamente" pode ser inferido quando não é expressamente declarado.

"Análogo de aminoácido", "aminoácido não-canônico", ou "aminoácido não-padrão", "aminoácido não-natural", "aminoácido antinatural" e semelhantes podem todos ser usados de forma intercambiável, e pretende-se incluir todos os compostos de tipo aminoácido que são semelhantes em estrutura e/ou de forma geral a um ou mais dos 20 L-aminoácidos comumente encontrados em proteínas naturais (Ala ou A, Cys ou C, Asp ou D, Glu ou E, Phe ou F, Gly ou G, His ou H, Ile ou I, Lys ou K, Leu ou L, Met ou M, Asn ou N, Pro ou P, Gin ou Q, Arg ou R, Ser ou S, Thr ou T, Val ou V, Trp ou W, Tyr ou Y, como definido e listado em WIPO Standard ST.25 (1998), Anexo 2, Tabela 3). Análogo de aminoácido também pode ser aminoácidos naturais com cadeias laterais ou cadeia principal modificadas. Aminoácidos podem também ser aminoácidos naturais em configuração D-, em vez de L-. Preferencialmente, estes análogos normalmente não são "substratos" para as aminoacil-tRNA sintetases (AARSs) devido à especificidade normalmente elevada das AARSs. Embora ocasionalmente, certos análogos com estruturas ou

formas suficientemente próximas dos aminoácidos naturais podem ser incorretamente incorporados em proteínas pelas AARSs, especialmente AARSs modificadas com especificidade de substrato relaxada. Numa forma de realização preferida, os análogos partilham estruturas de cadeia principal, e/ou mesmo a maioria das estruturas de cadeia lateral de um ou mais aminoácidos naturais, com a(s) única(s) diferença(s) sendo que contêm um ou mais grupos modificados na molécula. Tais modificações podem incluir, sem limitação, a substituição de um átomo (tal como N) por um átomo relacionado (tal como S), adição de um grupo (tal como grupo metilo, ou hidroxilo, etc.) ou um átomo (tal como o Cl ou Br, etc), deleção de um grupo (supra), a substituição de uma ligação covalente (ligação simples por ligação dupla, etc.), ou combinações das mesmas. Análogos de aminoácidos podem incluir α -hidroxiácidos e β -aminoácidos e também podem ser referidos como "aminoácidos modificados", ou "substratos de AARS antinaturais".

Os análogos de aminoácidos podem ser naturais ou não-naturais (p.ex., sintetizados). Como será apreciado pelos peritos na arte, qualquer estrutura para a qual um conjunto de rotâmeros é conhecida ou pode ser gerada pode ser usada como um análogo de aminoácido. As cadeias laterais podem ser de configuração (R) ou (S) (ou configuração D- ou L-). Numa forma de realização preferida, os aminoácidos são de configuração (S) ou L-.

Preferencialmente, a forma global e tamanho dos análogos de aminoácidos são de tal forma que estes, ao serem carregados em tRNAs (naturais ou modificados ou re-desenhados) pela AARS (natural ou re-desenhada), o análogo-tRNA é um complexo aceite pelo ribossoma, ou seja, o complexo

análogo-tRNA pode ser aceite pelos ribossomas procariotas ou eucariotas num sistema de tradução *in vivo* ou *in vitro*.

"Resíduos *achor*" são posições de resíduo na AARS que mantêm interações críticas entre a AARS e a cadeia principal do aminoácido natural.

"Cadeia principal", ou "molde" inclui os átomos da cadeia principal e quaisquer cadeias laterais fixas (tais como as cadeias laterais de resíduo âncora) da proteína (p.ex., AARS). Para efeitos de cálculo, a cadeia principal de um análogo é tratada como parte da cadeia principal da AARS.

"Estrutura de cadeia principal da proteína" ou equivalentes gramaticais neste documento significa as coordenadas tridimensionais que definem a estrutura tridimensional de uma proteína particular. As estruturas que compreendem uma estrutura de cadeia principal de proteína (de uma proteína natural) são o nitrogénio, o carbono de carbonilo, o α -carbono e o oxigénio do carbonilo, juntamente com a direção do vetor do α -carbono para o β -carbono.

A estrutura de cadeia principal da proteína que é o dado de entrada num computador para predição computacional de estrutura molecular ou de interação, pode incluir ou as coordenadas tanto para a cadeia principal como para as cadeias laterais de aminoácidos, ou apenas a cadeia principal, ou seja, com as coordenadas para as cadeias laterais de aminoácidos removidas. Se o anterior for efetuado, os átomos de cadeia lateral de cada aminoácido da estrutura da proteína podem ser "retirados" ou removidos da estrutura de uma proteína, como é conhecido na arte, deixando apenas as coordenadas para os átomos da "cadeia

principal" (o nitrogénio, carbono e oxigénio de carbonilo, e o α -carbono, e os átomos de hidrogénio ligados ao nitrogénio e α -carbono).

Opcionalmente, a estrutura da cadeia principal da proteína pode ser alterada antes da análise descrita abaixo. Nesta forma de realização, a representação da estrutura da cadeia principal inicial da proteína é reduzida a uma descrição do arranjo espacial dos seus elementos estruturais secundários. A posição relativa dos elementos estruturais secundários é definida por um conjunto de parâmetros chamados parâmetros de estrutura super-secundários. A estes parâmetros são atribuídos valores que podem ser sistematicamente ou aleatoriamente variados para alterar a disposição dos elementos da estrutura secundária introduzir flexibilidade da cadeia principal explícita. As coordenadas atómicas da cadeia principal são depois alteradas para refletir os parâmetros estruturais super-secundários alterados, e estas novas coordenadas são inseridas no sistema para uso na subsequente automatização do desenho da proteína. Para obter detalhes, ver U.S. Pat. No. 6,269,312, todo o conteúdo incorporado neste documento por referência.

"Energia conformacional" geralmente refere-se à energia associada a uma "conformação" particular, ou estrutura tridimensional, de uma macromolécula, tal como a energia associada com a conformação de uma proteína particular. Interações que tendem a estabilizar uma proteína têm energias que são representadas como valores de energia negativa, enquanto as interações que desestabilizam uma proteína têm valores de energia positiva. Portanto, a energia conformacional para qualquer proteína estável é quantitativamente representada por um valor de energia

negativa conformacional. Geralmente, a energia conformacional para uma proteína particular estará relacionada com a estabilidade da proteína. Em particular, moléculas que têm uma energia conformacional inferior (ou seja, mais negativa) são tipicamente mais estáveis, p.ex., a temperaturas mais altas (ou seja, possuem maior "estabilidade térmica"). Neste sentido, a energia conformacional de uma proteína pode também ser referida como a "energia de estabilização".

Normalmente, a energia conformacional é calculada usando um "campo de forças" de energia que calcula ou estima a contribuição da energia de várias interações que dependem da conformação de uma molécula. O campo de forças é compreendido de termos que incluem a energia conformacional da cadeia principal do alfa-carbono, interações cadeia lateral - cadeia principal, e interações cadeia lateral - cadeia lateral. Normalmente, as interações com a cadeia principal ou lateral incluem termos para rotação de ligação, torção de ligação e comprimento de ligação. As interações cadeia principal-cadeia lateral e cadeia lateral-cadeia lateral incluem termos de interações de van der Waals, pontes de hidrogénio, eletrostática e solvatação. Interações eletrostáticas podem incluir interações coulombianas, interações dipolo e interações quadrupolo). Outros termos semelhantes também podem ser incluídos. Campos de forças que podem ser usados para determinar a energia conformacional para um polímero são bem conhecidos na arte e incluem os campos de forças CHARMM (ver, Brooks et al, J. Comp. Chem. 1983,4:187-217; MacKerell et al., in The Encyclopedia of Computational Chemistry, Vol. 1:271-277, John Wiley & Sons, Chichester, 1998), AMBER (ver, Cornell et al., J. Amer. Chem. Soc.

1995, 117:5179; Woods et al., J. Phys. Chem. 1995, 99:3832-3846; Weiner et al., J. Comp. Chem. 1986, 7:230; e Weiner et al., J. Amer. Chem. Soc. 1984, 106:765) e DREIDING (Mayo et al., J. Phys. Chem. 1990, 94:-8897), entre outros.

Numa implementação preferida, os termos de pontes de hidrogénio e de eletrostática são os conforme descritos em Dahiyat & Mayo, (Science 1997 278:82). O campo de forças também pode ser descrito para incluir termos conformacionais atômicos (ângulos de ligação, comprimentos de ligação, torções), como em outras referências. Ver p.ex., Nielsen, et al. Prot. Eng., 12: 657662(1999); Stikoff, et al., Biophys. J., 67: 2251-2260 (1994); Hendsch, et al., Prot. Sci., 3: 211-226 (1994); Schneider, et al., J. Am. Chem. Soc., 119: 5742-5743 (1997); Sidelar, et al., Prot. Sci., 7: 1898-1914 (1998). Termos de solvatação podem também ser incluídos. Ver p.ex., Jackson, et al., Biochemistry, 32: 11259-11269 (1993); Eisenberg, et al., Nature, 319: 199-203 (1986); Street A G and Mayo S L; Folding & Design, 3: 253-258 (1998); Eisenberg and Wesson, Prot. Sci., 1: 227-235 (1992); Gordon & Mayo, supra.

"Resíduos acoplados" são resíduos numa molécula que interagem, através de qualquer mecanismo. A interação entre os dois resíduos é portanto referida como uma "interação de acoplamento". Resíduos acoplados geralmente contribuem para a capacidade de polímero através da interação de acoplamento. Normalmente, a interação de acoplamento é uma interação física ou química, tal como uma interação eletrostática, uma interação de van der Waals, uma interação de pontes de hidrogénio ou uma combinação destas. Como resultado da interação de acoplamento, alterando a identidade de qualquer resíduo afetará a "capacidade" da

molécula, particularmente se a mudança interrompe a interação de acoplamento entre os dois resíduos. Interação de acoplamento também pode ser descrita por um parâmetro de distância entre resíduos numa molécula. Se os resíduos estão dentro de uma certa distância limite, são considerados interagirem.

“Capacidade” é usada para denotar o nível ou grau a que uma determinada propriedade ou uma combinação particular de propriedades para uma molécula, por exemplo, uma proteína, são otimizadas. Em certas formas de realização da invenção, a capacidade de uma proteína de preferência é determinada pelas propriedades que um utilizador deseja melhorar. Portanto, por exemplo, a capacidade de uma proteína pode referir-se à estabilidade térmica, atividade catalítica, afinidade de ligação, solubilidade (p.ex., num solvente aquoso ou orgânico) da proteína, e afins. Outros exemplos de propriedades de capacidade incluem enantiosseletividade, atividade para substratos antinaturais, e mecanismos catalíticos alternativos. Interações de acoplamento podem ser modeladas como uma forma de avaliar ou prever a capacidade (estabilidade). Capacidade pode ser determinada ou avaliada experimentalmente ou teoricamente, p.ex., computacionalmente.

Preferencialmente, a capacidade é quantificada de modo a que cada molécula, p.ex., cada aminoácido terá um determinado “valor de capacidade”. Por exemplo, a capacidade de uma proteína pode ser a taxa à qual a proteína catalisa uma reação química específica, ou afinidade de ligação da proteína por um ligando. Numa forma de realização particularmente preferida, a capacidade de uma proteína refere-se à energia conformacional do polímero

e é calculada, p.ex., usando qualquer método conhecido na arte. Ver, p.ex., Brooks, et al., J. Comp. Chem., 4: 187-217 (1983); Mayo, et al., J. Phys. Chem., 94: 8897-8909 (1990); Pabo, et al., Biochemistry, 25: 5987-5991 (1986), Lazar, et al., Prot. Sci., 6: 1167-1178 (1997); Lee, et al., Nature, 352: 448-451 (1991); Colombo, et al., J. Am. Chem. Soc., 121: 6895-6903 (1999); Weiner, et al., J. Am. Chem. Soc., 106: 765-784 (1984): Geralmente, a capacidade de uma proteína é quantificada de modo a que o valor de capacidade aumente à medida que a propriedade ou combinação de propriedades é otimizada. Por exemplo, em formas de realização onde a estabilidade térmica de uma proteína vai ser otimizada a energia (de conformação) é preferencialmente diminuída), o valor de capacidade pode ser a energia negativa conformacional; ou seja, $F=-E$.

A "contribuição de capacidade" de um resíduo de proteína refere-se ao nível ou medida $f(i_a)$ a que o resíduo i_a , possuindo uma identidade, contribui para a capacidade total da proteína. Assim, por exemplo, se alterar ou mutar um resíduo de aminoácido particular diminuir consideravelmente a capacidade da proteína, esse resíduo é dito ter uma contribuição de capacidade elevada para o polímero. Em contraste, normalmente alguns resíduos i_a numa proteína podem possuir uma variedade de possíveis identidades a sem afetar a capacidade da proteína. Tais resíduos, portanto, têm uma baixa contribuição para a capacidade de proteína. "Dead-end elimination" (DEE) é um algoritmo de pesquisa determinístico que visa eliminar sistematicamente rotâmeros maus e combinações de rotâmeros até que reste uma única solução. Por exemplo, resíduos de aminoácidos podem ser modelados como rotâmeros que interagem com uma cadeia principal fixa. A base teórica para DEE prevê que, se a

pesquisa de DEE convergir, a solução é a conformação de energia mínima global (GMEC) com nenhuma incerteza (Desmet et al., 1992).

Dead-end elimination baseia-se no seguinte conceito. Considerem-se dois rotâmeros, i_r e i_t , no resíduo i , e o conjunto de todas as outras configurações do rotâmero $\{S\}$ em todos os resíduos excluindo i (do qual o rotâmero j_s é um membro). Se a energia emparelhada contribuída entre i_r e j_s é maior do que a energia emparelhada entre i_t e j_s para todos $\{S\}$, então o rotâmero i_r não pode existir na conformação global de energia mínima, e pode ser eliminado. Esta noção é expressa matematicamente pela desigualdade.

$$E(i_r) + \sum_{j \neq i}^N E(i_r, j_s) > E(i_t) + \sum_{j \neq i}^N E(i_t, j_s) \{ S \} \quad (\text{Equation A})$$

Se esta expressão for verdadeira, o rotâmero individual i_r pode ser eliminado (Desmet et al., 1992).

Desta forma, a equação A não é computacionalmente informatizada porque, para fazer uma eliminação, é necessário que todo o espaço da sequência (rotâmero) seja enumerado. Para simplificar o problema, limites implicados pela equação A podem ser utilizados:

$$E(i_r) + \sum_{j \neq i}^N \min(s) E(i_r, j_s) > E(i_t) + \sum_{j \neq i}^N \max(s) E(i_t, j_s) \{ S \} \quad (\text{Equation B})$$

Usando um argumento análogo, a equação B pode ser estendida para a eliminação de pares de rotâmeros inconsistentes com a GMEC. Isto é feito determinando-se que um par de rotâmeros i_r no resíduo i e j_s no resíduo j , contribuem sempre energias mais elevadas do que rotâmeros i_u e j_v com

todas as combinações possíveis de rotâmeros $\{L\}$. Semelhante à equação B, o limite estrito desta instrução é dado por:

$$\varepsilon(i_r, j_s) + \sum_{k \neq i, j}^N \min(t) \varepsilon(i_r, j_s, k_t) > \varepsilon(i_u, j_v) + \sum_{k \neq i, j}^N \max(t) \varepsilon(i_u, j_v, k_t) \quad (\text{Equation C})$$

onde ε é a energia combinada para os pares de rotâmeros

$$\varepsilon(i_r, j_s) = E(i_r) + E(j_s) + E(i_r, j_s) \quad (\text{Equation D}),$$

e

$$\varepsilon(i_r, j_s, k_t) = E(i_r, k_t) + E(j_s, k_t) \quad (\text{Equation E}).$$

Isto leva à eliminação de duplas do par de rotâmeros i_r e j_s , mas não elimina os rotâmeros individuais completamente uma vez que qualquer um pode existir independentemente na GMEC. A etapa de eliminação de duplas reduz o número de pares possíveis (reduz S) que necessitam ser avaliadas no lado direito da Equação 6, permitindo mais rotâmeros serem individualmente eliminados.

Os critérios de individuais e duplas, apresentados por Desmet *et al.* não descobrem as condições especiais que levam à determinação de mais rotâmeros *dead-ending*. Por exemplo, é possível que a contribuição da energia do rotâmero i_t seja sempre menor do que i_r sem o máximo de i_t ser abaixo do mínimo de i_r . Para resolver este problema, Goldstein 1994 apresentou uma modificação dos critérios que determina se os perfis de energia de dois rotâmeros se cruzam. Se não o fizerem, o rotâmero de energia mais elevada pode ser determinado para ser *dead-ending*. O cálculo de duplas usa significativamente mais tempo computacional do que o cálculo de individuais. Para acelerar o processo, outros métodos computacionais foram

desenvolvidos para prever os cálculos de duplas que serão os mais produtivos (Gordon & Mayo, 1998). Estes tipos de modificações, referidos coletivamente como duplas rápidas, melhoraram significativamente a velocidade e a eficácia da DEE.

Várias outras modificações também melhoram a DEE. Rotâmeros de múltiplos resíduos podem ser combinados nos assim denominados super-rotâmeros para solicitar mais eliminações (Desmet et al., 1994; Goldstein, 1994). Isto tem a vantagem de eliminar múltiplos rotâmeros numa única etapa. Além disso, verificou-se que "dividir" o espaço conformacional entre rotâmeros melhora a eficiência de DEE (Pierce et al., 2000). A divisão lida com o seguinte caso especial. Considere-se o rotâmero i_r . Se um rotâmero i_{t1} contribui uma energia menor do que i_r para uma parte do espaço conformacional, e um rotâmero i_{t2} possui uma energia menor do que i_r para a fração restante, então i_r pode ser eliminado. Este caso não seria detetado pelos critérios Desmet ou Goldstein menos sensíveis. Em formas de realização conforme descrito neste documento, todos os melhoramentos descritos para DEE foram usados.

Para discussão adicional destes métodos ver, Goldstein, *Biophysical Journal* 66, 1335-1340 (1994); Desmet, et al., *Nature* 356, 539-542 (1992); Desmet, et al., *The Protein Folding Problem and Tertiary Structure Prediction* (Jr., K. M. & Grand, S. L., eds.), pp. 307-337 (Birkhauser, Boston, 1994); De Maeyer, et al., *Folding & Design* 2, 53-66 (1997), Gordon, and Mayo, *J. Comp. Chem.* 19, 1505-1514 (1998); Pierce, et al., *J. Comp. Chem.* 21, 999-1009 (2000).

Outro cálculo, apelidado de SCREAM (Side-Chain Rotamer Energy Analysis Method), pode ser usado. SCREAM permite que a análise do mecanismo de discriminação contra aminoácidos não-cognatos, calculando as energias relativas de ligação dos 20 aminoácidos naturais a uma AARS particular. (Ver, por exemplo, McClendon, et al., Prot. Eng. Design & Select. 19: 195-203 (2006)).

Como um primeiro passo, o espectro de energia do rotâmero é calculado para um único aminoácido numa cadeia principal desocupada, com nenhuma outra cadeia lateral movível. De seguida, começando com os rotâmeros mais baixos a partir da cadeia principal desocupada, preencher as cadeias laterais mas eliminar confrontos. Por exemplo, colocar cadeias laterais em cada local, estimando as energias de excitações baixas a partir do espectro da cadeia principal desocupada e calcular interações dos pares, eliminando as configurações que apresentam confrontos. Portanto, $E_{tot}(A,B) = E_{self}(A) + E_{self}(B) + E_{int}(A,B)$ and $E_{tot}(A,B,C) = E_{self}(A) + E_{self}(B) + E_{self}(C) + E_{int}(A,B) + E_{int}(A,C) + E_{int}(B,C) = E_{self}(AB) + E_{self}(C) + E_{int}(A,C) + E_{int}(B,C) = E_{self}(A,B) + E_{self}(C) + E_{int}(AB,C)$ estabiliza a relação recursiva. De seguida, todas as excitações de baixa energia das cadeias laterais devem ser analisadas até a distribuição de energia das energias do rotâmero na cadeia principal desocupada cessar de aumentar. Resumidamente, a energia do estado fundamental para todos os resíduos é avaliada, seguida por um conjunto de rotâmeros com a menor energia de soma linear, e finalmente a próxima energia de soma linear menor e assim por diante. Além disso, interações eletrostáticas devem ser abordadas uma vez que as cargas polarizam o ambiente para proteger as cargas e reduzir a interação de aminoácidos desejada. Uma vez que

métodos de modelação de dinâmica molecular geralmente não levam em conta a polarização, o enviesamento é a favor de pontes salinas. A fim de superar isto, os resíduos são neutralizados e os parâmetros avaliados novamente de acordo com a parametrização de DREIDING.

Para parametrização de DREIDING, as interações carga perdida-carregado e carregado-dipolo são compensadas introduzindo o termo de ponte de hidrogénio. Isto pode ser feito em conjunto com outros programas, incluindo CHARMM, conforme descrito neste documento. Assim, usando uma estrutura cristalográfica de uma AARS particular, ou AARS homóloga de outro organismo, podemos usar um programa tal como SCREAM, e HierDock, ou outros, para prever a conformação de ligação e energia de ligação de cada um dos 20 aminoácidos naturais no sítio de ligação no melhor modo de ligação e o modo de ativação, ordenando os cálculos de acordo com que aminoácidos competem para a ligação a uma AARS particular.

Em particular, ligação seletiva é a primeira execução, que fornece o aminoácido e a molécula de ATP para ligar ao sítio ativo da AARS. Isto às vezes leva a uma mudança conformacional. De seguida, ativação seletiva da AARS para catalisar a formação de uma ligação covalente entre o aminoácido e o ATP, forma um aminoacil-adenilato complexado com a AARS e remove pirofosfato inorgânico. Em terceiro lugar, se ativação incorreta de um aminoácido não-cognato ocorre, a AARS pode de modo hidrolítico clivar o complexo aminoacil-adenilato (como *proofreading* de pré-transferência). Finalmente, se um aminoacil-adenilato não-cognato sobreviveu, a AARS pode de modo hidrolítico clivar o complexo aminoacil-tRNA (*proofreading* pós-transferência).

Assim, tais programas computacionais permitem previsão fiável da probabilidade dos aminoácidos naturais que competem para se ligarem e aminoacilação por enzimas AARS de tipo selvagem ou mutantes. Utilizando múltiplos programas (como SCREAM e HierDock juntos) reduz a taxa de incorporação incorreta e permite maior previsibilidade na seleção de localizações de aminoácidos que ligam o aminoácido especificamente.

Ainda outros programas de modelação incluem SCAP (Side Chain Amino Acid Prediction Program) (Xiang and Honig, J. Mol. Biol. 311, 421-430 (2001)), e SCWRL (Side Chain Replacement With a Rotamer Library), o que é útil para a adição de cadeias laterais a uma cadeia principal de uma proteína baseada na biblioteca de rotâmeros dependentes da cadeia principal. A biblioteca SCWRL fornece listas de valores χ_1 - χ_2 - χ_3 - χ_4 e as suas probabilidades relativas para resíduos a dados valores ψ - ϕ , e explora estas conformações para minimizar confrontos cadeia lateral-cadeia principal e confrontos cadeia lateral-cadeia lateral. (Ver, por exemplo, Bower, et al., J. Mol. Biol., 267, 1268-1282 (1997)).

A previsibilidade computacional é devida, em grande parte, a utilizar sequências conhecidas de ácido nucleico e/ou aminoácidos de enzimas AARS. Por exemplo, o domínio catalítico está conservado em todos os membros de uma classe particular de enzima AARS. (O'Donoghue and Luthey-Schulten, Microbiol. And Mol. Biol. Rev.: 550-573 (2003); Diaz-Lazcoz, et al., Mol. Biol. Evol. 15(11): 1548-1561 (1998); Wang, et al., Chem. Commun. 1-11 (2002)).

“Sistema de expressão” significa uma célula hospedeira, ou componentes celulares e vetor compatível sob condições apropriadas, p.ex., para a expressão de uma proteína codificada por DNA exógeno transportado pelo vetor e introduzido na célula hospedeira. Sistemas de expressão comuns incluem células hospedeiras de *E. coli* e vetores plasmídicos, células hospedeiras de inseto tais como células Sf9, Hi5 ou S2 e vetores de Baculovirus, células de *Drosophila* (células Schneider) e sistemas de expressão, e células hospedeiras e vetores de mamíferos.

“Célula hospedeira” significa qualquer célula de qualquer organismo que é selecionada, modificada, transformada, crescida ou usada ou manipulada de alguma forma para a produção de uma substância pela célula. Uma célula hospedeira pode ser auxotrófica, que é incapaz de sintetizar pelo menos um composto orgânico particular exigido para a sua manutenção ou crescimento e tem de obter o composto a partir de outra fonte, como o seu ambiente ou meios de cultura. Além disso, uma célula hospedeira auxotrófica pode ter níveis de auxotrofia simples, dupla, tripla, quádrupla ou mais, de tal modo que é incapaz de sintetizar um, dois, três, quatro ou mais compostos orgânicos necessários para o seu crescimento ou manutenção, respetivamente. Por exemplo, uma célula hospedeira pode ser uma que é manipulada para expressar um gene específico, uma sequência de DNA ou RNA, uma proteína ou uma enzima. Células hospedeiras podem ser cultivadas *in vitro* ou uma ou mais células num animal não-humano (p.ex., um animal geneticamente modificado ou um animal transitoriamente transfetado).

Certas formas de realização divulgadas neste documento expressamente utilizam apenas um sistema de expressão ou tradução sem células e não uma célula hospedeira. Certas outras formas de realização expressamente utilizam apenas uma célula hospedeira auxotrófica. Ainda certas outras formas de realização expressamente utilizam apenas uma célula hospedeira não-auxotrófica, ou uma célula hospedeira prototrófica.

Similaridade da sequência pode ser relevante para certas formas de realização uma vez que podem incluir etapas de comparar sequências umas às outras, incluindo a sequência de tipo selvagem a uma ou mais mutantes. Tais comparações normalmente compõem alinhamentos de sequências de polímero, p.ex., usando programas de alinhamento de sequências e/ou algoritmos que são conhecidos na arte (por exemplo, BLAST, FASTA e MEGALIGN, para citar alguns). O artesão hábil pode prontamente apreciar que, em tais alinhamentos, onde uma mutação contém uma inserção ou deleção de resíduo, o alinhamento de sequências introduzirá um "intervalo" (normalmente representado por um traço, "-", ou "Δ") na sequência de polímero não contendo o resíduo inserido ou deletado.

"Homólogo", em todas as suas formas gramaticais e variações de ortografia, refere-se à relação entre duas moléculas (p.ex., proteínas, tRNAs, ácidos nucleicos) que possuem uma "origem evolutiva comum", incluindo proteínas de superfamílias na mesma espécie de organismo, bem como proteínas homólogas de diferentes espécies de organismos. Tais proteínas (e os seus ácidos nucleicos que as codificam) tem homologia de sequência e/ou estrutural, como refletido pela sua similaridade de sequência, quer em

termos de percentagem de identidade ou pela presença de resíduos específicos ou motivos e posições conservadas. Moléculas homólogas frequentemente também partilham funções similares ou mesmo idênticas.

Em alguns aspetos, homólogos pode incluir uma sequência que é pelo menos 50% homóloga, mas que apresenta uma estrutura homóloga em três dimensões, ou seja, inclui uma carga superficial substancialmente semelhante ou a apresentação de grupos hidrofóbicos. Uma vez que pontes de hidrogénio, forças de van der Waals e interações hidrofóbicas podem funcionar para ligar um aminoácido à cavidade de ligação de uma AARS, manipulação de uma estrutura da AARS também pode alterar uma ou mais destas forças.

Portanto, como usado neste documento, proteínas e/ou sequências de proteínas são "homólogas" quando são derivadas, natural ou artificialmente, de uma proteína ancestral comum ou sequência da proteína. Da mesma forma, os ácidos nucleicos e/ou sequências de ácidos nucleicos são homólogos quando são derivados, natural ou artificialmente, de um ancestral comum do ácido nucleico ou sequência de ácido nucleico. Por exemplo, qualquer ácido nucleico natural que pode ser modificado por qualquer método de mutagénese disponível para incluir um ou mais codões seletores. Quando expresso, este ácido nucleico mutagenizado codifica um polipéptido compreendendo um ou mais aminoácidos antinaturais. O processo de mutação pode, claro, adicionalmente alterar um ou mais codões padrão, deste modo alterando também um ou mais aminoácidos padrão na proteína mutante resultante. Homologia é geralmente inferida a partir da similaridade da sequência entre dois ou mais ácidos nucleicos ou proteínas (ou sequências dos

mesmos). A percentagem exata de similaridade entre as sequências que é útil no estabelecimento de homologia varia com o ácido nucleico e proteína em questão, mas tão pouco como 25% de similaridade da sequência é rotineiramente usada para estabelecer homologia. Níveis mais elevados de similaridade de sequência, p.ex., 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95% ou 99% ou mais também podem ser usados para estabelecer homologia. Métodos para a determinação de percentagens de similaridade de sequência (p.ex., BLASTP e BLASTN usando parâmetros por defeito) são descritos neste documento e estão geralmente disponíveis.

O termo "similaridade de sequência", em todas as suas formas gramaticais, refere-se ao grau de identidade ou correspondência entre sequências de ácidos nucleicos ou de aminoácidos que podem ou não podem partilhar uma origem evolutiva comum (ver, Reeck et al., supra). No entanto, no uso comum e no pedido iminente, o termo "homólogo", quando modificado com um advérbio como "altamente", pode referir-se a similaridade de sequência e pode ou não dizer respeito a uma origem evolutiva comum.

Uma molécula de ácido nucleico é "hibridizável" com uma outra molécula de ácido nucleico, tais como um cDNA, DNA genómico, ou RNA, quando uma forma de cadeia simples da molécula de ácido nucleico pode emparelhar com a outra molécula de ácido nucleico sob as condições adequadas de temperatura e força iónica da solução (ver Sambrook et al., Molecular Cloning: A Laboratory Manual, Second Edition (1989) Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.). As condições de temperatura e força iónica determinam a "estringência" da hibridação. Para a pesquisa preliminar de ácidos nucleicos homólogos, condições de

baixa estringência de hibridação, correspondente a uma T_m (temperatura de fusão) de 55°C, podem ser usadas, p.ex., 5×SSC, 0,1% SDS, 0,25% de leite, e sem formamida; ou 30% formamida, 5×SSC, 0,5% SDS). Condições moderadas de estringência correspondem a uma T_m mais elevada, p.ex., 40% formamida, com 5x ou 6×SSC. Condições de hibridação de elevada estringência correspondem à T_m mais elevada p.ex., 50% formamida, 5x ou 6×SSC. SSC é um 0,15 M NaCl, 0,015 M Na-citrato. Hibridação requer que os dois ácidos nucleicos contenham trechos complementares de sequências génicas ou aminoacídicas, embora dependendo da estringência da hibridação, emparelhamentos incorretos entre bases são possíveis.

A estringência adequada para hibridação de ácidos nucleicos depende do comprimento de ácidos nucleicos e no grau de complementação, variáveis conhecidas na arte. Quanto maior o grau de similaridade ou homologia entre duas sequências de nucleótidos, maior é o valor de T_m para híbridos de ácidos nucleicos com essas sequências. A relativa estabilidade (correspondente a maior T_m) da hibridação de ácidos nucleicos diminui na seguinte ordem: RNA:RNA, DNA:RNA, DNA:DNA. Para os híbridos de mais de 100 nucleótidos de comprimento, equações para o cálculo da T_m foram derivadas (ver Sambrook et al., supra, 9.50-9.51). Por hibridação com ácidos nucleicos menores, ou seja, oligonucleótidos, a posição de desemparelhamento torna-se mais importante, e o comprimento do oligonucleótido determina a sua especificidade (ver Sambrook et al., supra, 11.7-11.8). Um comprimento mínimo para um ácido nucleico hibridizável é pelo menos cerca de 10 nucleótidos; preferencialmente pelo menos cerca de 15 nucleótidos; e

mais preferencialmente o comprimento é pelo menos cerca de 20 nucleótidos.

Salvo especificado, o termo "condições padrão de hibridação" refere-se a uma T_m de cerca de 55°C, e utiliza as condições estabelecidas acima. Numa forma de realização preferida, a T_m é 60°C; numa forma de realização mais preferida, a T_m é 65°C. Numa forma de realização específica, "estringência elevada" refere-se a hibridação e/ou condições de lavagem a 68°C em 0,2×SSC, a 42°C em 50% formamida, 4×SSC, ou em condições que possibilitam os níveis de hibridação equivalentes aos observados sob cada uma daquelas duas condições.

Condições adequadas de hibridação para oligonucleótidos (p.ex., para sondas ou iniciadores de oligonucleótidos) são tipicamente algo diferentes do que para os ácidos nucleicos completos (p.ex., cDNA completo), por causa da baixa temperatura de fusão dos oligonucleótidos. Uma vez que a temperatura de fusão dos oligonucleótidos dependerá do comprimento das sequências dos oligonucleótidos envolvidos, temperaturas de hibridação adequadas irão variar dependendo das moléculas de oligonucleótidos usadas. Temperaturas exemplares podem ser 37°C (para oligonucleótidos de 14 bases), 48°C (para oligonucleótidos de 17 bases), 55°C (para oligonucleótidos de 20 bases) e 60°C (para oligonucleótidos de 23 bases). Condições exemplares de hibridação adequadas para oligonucleótidos incluem lavar em 6×SSC/0,05% pirofosfato de sódio, ou outras condições que possibilitam níveis equivalentes de hibridação.

"Polipéptido", "péptido" ou "proteína" são usados de forma intercambiável para descrever uma cadeia de aminoácidos que

são ligados entre si por ligações químicas denominadas "ligações peptídicas". Uma proteína ou um polipéptido, incluindo uma enzima, pode ser "nativa" ou "de tipo selvagem", ou seja, que ocorre na natureza; ou pode ser "mutante", "variante" ou "modificada", significando que foi feita, alterada, derivada, ou é de alguma forma diferente ou alterada a partir de uma proteína nativa ou de outro mutante.

"Rotâmero" é definido como um conjunto de possíveis conformers para cada cadeia lateral de aminoácidos ou análogos. Ver, See Ponder, et al., Acad. Press Inc. (London) Ltd. pp. 775-791 (1987); Dunbrack, et al., Struc. Biol. 1(5):334-340 (1994); Desmet, et al., Nature 356:539-542 (1992). Uma "biblioteca de rotâmeros" é uma coleção de um conjunto de conformações de rotâmeros possíveis/admissíveis para um determinado conjunto de aminoácidos ou análogos. Existem dois tipos gerais de bibliotecas de rotâmeros: "dependente de cadeia principal" e "independente de cadeia principal". Uma biblioteca de rotâmeros dependente de cadeia principal permite rotâmeros diferentes dependendo da posição do resíduo na cadeia principal; assim, por exemplo, certos rotâmeros leucina são permitidos se a posição está numa hélice α , e diferentes rotâmeros leucina são permitidos se a posição não está numa α -hélice. Uma biblioteca de rotâmeros independente de cadeia principal utiliza todos os rotâmeros de um aminoácido em cada posição. Em geral, uma biblioteca independente de cadeia principal é preferida na consideração de resíduos de núcleo, já que a flexibilidade no núcleo é importante. No entanto, bibliotecas independentes de cadeia principal são computacionalmente mais caras, e assim para as posições de superfície e

periferia, uma biblioteca dependente da cadeia principal é preferida. No entanto, qualquer tipo de biblioteca pode ser usado em qualquer posição.

“Posição variável de resíduo” neste documento significa uma posição do aminoácido da proteína a ser projetado que não está fixa no método de projeto como um resíduo específico ou rotâmero, geralmente o resíduo de tipo selvagem ou rotâmero. Note-se que mesmo que uma posição seja escolhida como uma posição variável, é possível que certos métodos divulgados neste documento irão otimizar a sequência de forma a selecionar o resíduo de tipo selvagem na posição variável. Isto geralmente ocorre com mais frequência para resíduos de núcleo e menos regularmente para resíduos de superfície. Além disso, é possível fixar resíduos como aminoácidos de tipo não-selvagem também.

“Posição fixa de resíduo” significa o resíduo identificado na estrutura tridimensional como estando numa conformação definida. Em algumas formas de realização, uma posição fixa é deixada na sua conformação original (que pode ou não pode correlacionar-se com um rotâmero específico da biblioteca de rotâmeros que está a ser usada). Alternativamente, os resíduos podem ser fixados como um resíduo de tipo não-selvagem dependendo das necessidades do *design*; por exemplo, quando técnicas de mutagénese sítio-dirigidas conhecidas demonstraram que um determinado resíduo é desejável (por exemplo, para eliminar um local proteolítico ou alterar a especificidade de substrato de uma AARS), o resíduo pode ser fixado como um aminoácido particular. Resíduos que podem ser fixados incluem, mas não estão limitados a, resíduos estruturalmente ou biologicamente funcionais, por exemplo, os resíduos âncora.

Em certas formas de realização, uma posição fixa pode ser "flutuante"; o aminoácido ou análogo nessa posição é fixo, mas diferentes rotâmeros desse aminoácido ou análogo são testados. Nesta forma de realização, os resíduos variáveis podem ser pelo menos um, ou qualquer valor entre 0,1% a 99,9% do número total de resíduos. Assim, por exemplo, pode ser possível mudar apenas alguns (ou um) resíduos, ou a maioria dos resíduos, com todas as possibilidades no meio.

Como usado neste documento, o termo "mutante externo" refere-se a uma molécula modificada (p.ex., um tRNA mutante externo e/ou uma aminoacil-tRNA sintetase mutante externa) que exhibe uma eficiência reduzida (em comparação com o tipo selvagem ou endógena) para aminoacilação com o aminoácido de tipo selvagem correspondente. "Mutante externo" refere-se a incapacidade ou eficiência reduzida, p.ex., menos de 20% eficiente, menos de 10% eficiente, menos de 5% eficiente, ou, por exemplo, menos de 1% eficiente, de um tRNA e/ou RS para funcionar com o correspondente aminoácido natural no sistema de tradução de interesse. Por exemplo, uma RS mutante externa num sistema de tradução de interesse aminoacila qualquer tRNA endógeno de um sistema de tradução de interesse com o aminoácido de tipo selvagem a eficiência reduzida ou mesmo zero, quando comparado com a aminoacilação de um tRNA endógeno pela RS endógena.

Deve ser notado, no entanto, que uma RS mutante externa aminoacila um tRNA endógeno com um aminoácido de substituição (quer seja natural ou não-natural) com uma maior eficiência em comparação com a capacidade da RS endógena para aminoacilar um tRNA endógeno com um aminoácido de substituição. Da mesma forma, um tRNA mutante

externo funciona a uma eficiência mais elevada para o aminoácido de substituição (quer seja não-natural ou outro aminoácido natural) do que para o aminoácido de tipo selvagem correspondente.

"Codão degenerado *wobble*" refere-se a um codão codificando um aminoácido natural, cujo codão, quando presente no mRNA, é reconhecido por um anticodão natural de tRNA através de pelo menos um emparelhamento de bases não-Watson-Crick, ou emparelhamento de bases *wobble* (p.ex., emparelhamento de bases A-C ou G-U). Emparelhamento de bases de Watson-Crick refere-se a G-C ou A-U (RNA ou híbrido DNA/RNA) ou emparelhamento de bases A-T (DNA). Quando usado no contexto emparelhamento de bases de codão de mRNA - anticodão de tRNA, emparelhamento de bases de Watson-Crick significa todos os emparelhamentos de bases codão-anticodão são mediados através de emparelhamentos G-C ou A-U.

O termo "preferencialmente aminoacila" refere-se a uma eficiência, p.ex., cerca de 20%, cerca de 30%, cerca de 40%, cerca de 50%, cerca de 60%, cerca de 70%, cerca de 75%, cerca de 85%, cerca de 90%, cerca de 95%, cerca de 99% ou mais eficiente. A eficiência pode ser medida pela qual uma aminoacil-tRNA sintetase modificada ou mutante externa aminoacila um tRNA com um aminoácido de substituição, quer seja um aminoácido antinatural ou outro aminoácido natural, quando comparado com o aminoácido natural correspondente atribuído ao tRNA particular, AARS ou ambos. O termo "preferencialmente aminoacila" pode adicionalmente referir-se à eficiência com que a aminoacil-tRNA sintetase modificada ou mutante externa aminoacila ou carrega um tRNA com qualquer aminoácido que não seja o aminoácido natural correspondente atribuído ao tRNA particular, AARS ou ambos.

O termo "preferencialmente aminoacila" pode adicionalmente referir-se à eficiência com que a aminoacil-tRNA sintetase modificada ou mutante externa aminoacila um tRNA com um aminoácido não-natural em comparação com a AARS natural ou não modificada.

Deve ser notado que a eficiência da aminoacilação do tRNA pela AARS pode ser correlacionada com a eficiência da especificidade, ou fidelidade de incorporação do aminoácido não-natural na proteína ou polipéptido alvo. Isto é devido à função da maquinaria de síntese de proteínas na medida em que assim que um tRNA é aminoacilado com um aminoácido (quer seja o aminoácido de tipo selvagem, ou um aminoácido não-natural), o tRNA carregado é libertado da enzima AARS e o aminoácido é incorporado no polipéptido alvo. Quando a capacidade de *proofreading* da AARS é alterada, a enzima permitirá o aminoácido de substituição carregar o tRNA e ser libertado para a incorporação na proteína alvo. Portanto, a eficiência de aminoacilação pela AARS correlaciona-se diretamente com a fidelidade ou especificidade de incorporação do aminoácido não-natural no polipéptido alvo.

O aminoácido de substituição (quer seja não-natural ou natural) é então incorporado numa cadeia polipeptídica crescente com alta fidelidade, p.ex., a uma eficiência maior do que cerca de 20%, 30%, 40%, 50%, 60%, 75%, 80%, 90%, 95%, ou cerca de 99% para um codão particular.

O termo "complementar" refere-se aos componentes de um par mutante externo, o tRNA mutante externo e sintetase mutante externa que podem funcionar em conjunto, p.ex., a sintetase mutante externa aminoacila o tRNA mutante externo.

O termo "derivado de" refere-se a um componente que é isolado a partir de um organismo ou isolado e modificado, ou gerado, p.ex., quimicamente sintetizado, usando informações do componente do organismo.

O termo "sistema de tradução" refere-se aos componentes necessários para incorporar um aminoácido natural ou antinatural numa cadeia polipeptídica crescente (proteína). Por exemplo, os componentes podem incluir ribossomas, tRNA(s), sintetase(es), mRNA e semelhantes. Os componentes divulgados neste documento podem ser adicionados a um sistema de tradução, *in vivo* ou *in vitro*. Um sistema de tradução *in vivo* pode ser uma célula (célula eucariótica ou procariótica). Um sistema de tradução *in vitro* pode ser um sistema sem células, tal como um reconstituído com componentes de diferentes organismos (purificados ou produzidos de forma recombinante). Em certas formas de realização, o sistema de tradução não compreende uma célula. Em certas formas de realização, o sistema de tradução não compreende uma célula auxotrófica. Se o sistema de tradução não compreende uma célula auxotrófica, pode compreender outra célula ou componentes celulares.

O termo "RS inativa" refere-se a uma sintetase que foi mutada para que não possa aminoacilar o seu tRNA cognato com qualquer aminoácido, quer seja natural ou não-natural. O termo "RS modificada" refere-se a uma sintetase que foi mutada para que não possa aminoacilar o seu tRNA cognato com o aminoácido natural correspondente, mas pode ser capaz de aminoacilar o seu tRNA cognato com outro aminoácido, preferencialmente um aminoácido não-natural.

O termo "agente seletivo" refere-se a um agente que quando presente permite uma seleção de determinados componentes de uma população, p.ex., um antibiótico, comprimento de onda de luz, um anticorpo, um nutriente ou semelhantes. O agente de seleção pode ser variado, p.ex., tais como a concentração, intensidade, etc.

O termo "marcador de seleção positiva" refere-se a um marcador que quando presente, p.ex., expresso, ativado ou algo semelhante, resulta na identificação de um organismo com o marcador de seleção positiva relativamente àqueles sem o marcador de seleção positiva.

O termo "marcador de seleção negativa" refere-se a um marcador que quando presente, p.ex., expresso, ativado ou algo semelhante, permite identificação de um organismo que não possui a propriedade desejada (p.ex., quando comparado com um organismo que possui a propriedade desejada).

O termo "repórter" refere-se a um componente que pode ser usado para selecionar componentes descritos na divulgação. Por exemplo, um repórter pode incluir uma proteína verde fluorescente, uma proteína luciferase de pirilampo, ou genes tais como β -gal/lacZ (β -galactosidase), Adh (desidrogenase de álcool) ou semelhantes.

O termo "reconhecido não eficientemente" refere-se a uma eficiência, p.ex., menos de cerca de 10%, menos de cerca de 5%, ou menos de cerca de 1%, à qual uma RS de um organismo aminoacila um tRNA mutante externo. Em certas formas de realização, a RS pode ser do mesmo ou de um organismo diferente do que o tRNA mutante externo. Em algumas formas de realização, a RS foi modificada para aminoacilar um tRNA

com um aminoácido particular, preferencialmente um aminoácido não-natural.

O termo "eucariota" refere-se a organismos pertencentes ao domínio filogenético Eucarya tais como animais (p.ex., mamíferos, insetos, répteis, aves, etc.), ciliados, plantas, fungos (p.ex., leveduras, etc.), flagelados, microsporidia, protistas, etc. Adicionalmente, o termo "procariota" refere-se a organismos não-eucarióticos pertencentes aos domínios filogenéticos Eubacteria (p.ex., *Escherichia coli*, *Thermus thermophilus*, etc.) e Archaea (p.ex., *Methanococcus jannaschii*, *Methanobacterium thermoautotrophicum*, *Halobacterium* tais como *Haloferax volcanii* e espécies NRC-1 de *Halobacterium*, *A. fulgidus*, *P. firiokus*, *P. horikoshii*, *A. pernix*, etc.)

O Código Genético, Células Hospedeiras, e os Codões Degenerados

O código genético padrão que a maioria das células usa está listado abaixo.

O Meio do Código Genético

Primeira	U	C	A	G	Última
	Phe	Ser	Tyr	Cys	U
	Phe	Ser	Tyr	Cys	C
U	Leu	Ser	Stop		
(Ocre)	Stop				
(Opala)	A				
	Leu	Ser	Stop		
(Ambâr)	Trp	G			
	Leu	Pro	His	Arg	U
	Leu	Pro	His	Arg	C

Primeira	U	C	A	G	Última
C	Leu	Pro	Gin	Arg	A
	Leu	Pro	Gin	Arg	G
	Ile	Thr	Asn	Ser	U
A	lie	Thr	Asn	Ser	C
	Ile	Thr	Lys	Arg	A
	Met	Thr	Lys	Arg	G
	Val	Ala	Asp	Gly	U
G	Val	Ala	Asp	Gly	C
	Val	Ala	Glu	Gly	A
	Val	Ala	Glu	Gly	G

O código genético é degenerado, na medida em que a maquinaria biossintética de proteínas utiliza 61 codões de sentido de mRNA para direcionar a polimerização baseada em modelo dos 20 monómeros de aminoácidos naturais. (Crick et al., Nature 192: 1227, 1961). Apenas dois aminoácidos, ou seja, metionina e triptofano, são codificados por tripletos de mRNA únicos.

O código genético padrão aplica-se à maioria dos, mas não todos, casos. Exceções foram encontradas no DNA mitocondrial de muitos organismos e no DNA nuclear de alguns organismos inferiores. Alguns exemplos são dados na tabela seguinte.

Exemplos de códigos genéticos não-padrão.

Mitocôndria	Vertebrados	UGC→ Trp; AGA, AGG → STOP
	Invertebrados	UGA→ Trp; AGA, AGG → Ser
	Leveduras	UGC→ Trp; CUN → Thr
	Protista	UGA→ Trp;
Núcleo	Bactéria	GUG, UUG, AUU, CUG → iniciação
	Leveduras	CUG → Ser
	Ciliados	UAA, UAG → Gln
*Células vegetais usam o código genético padrão tanto nas mitocôndrias como no núcleo.		

O NCBI (National Center for Biotechnology Information) mantém uma lista detalhada do código genético padrão, e códigos genéticos usados em vários organismos, incluindo o código mitocondrial de vertebrados; o código mitocondrial de levedura; o código mitocondrial do fungo do bolor, protozoários e celenterados e o código de micoplasma / spiroplasma; o código mitocondrial de invertebrados; o código nuclear de ciliados, dasycladacean e hexamita; o código mitocondrial de equinodermos e platelmintos; o código nuclear de euplotid; o código de plastídeo bacteriano e vegetal; o código nuclear alternativo de levedura; o código mitocondrial de ascídias; o código mitocondrial alternativo de platelmintos; código nuclear de blepharisma; código mitocondrial de clorofíceas; código mitocondrial trematode; código mitocondrial de *Scenedesmus obliquus*; código mitocondrial de *thraustochytrium* (todos incorporados neste documento por referência). Estes são principalmente baseados nos artigos de revisão de Osawa et al., Microbiol. Rev. 56: 229-264, 1992, e Jukes and Osawa, Comp. Biochem. Physiol. 106B: 489-494, 1993.

Células Hospedeiras

Alguns métodos divulgados neste documento podem ser praticados dentro de uma célula, o que permite os níveis de produção de proteínas serem feitos para fins práticos. Em formas de realização preferenciais, as células usadas são de células que podem ser crescidas em cultura (ou seja, células que podem ser crescidas sob condições laboratoriais). Células adequadas incluem células de mamíferos (humanos ou mamíferos não-humanos), células bacterianas e células de inseto, etc.

Um exemplo inclui tecnologia PFENEX™, que é uma linha celular usando a linha celular baseada em *Pseudomonas fluorescens* que aumenta a expressão celular mantendo certas características de solubilidade e atividade devido ao uso de diferentes vias no metabolismo de determinados açúcares em comparação com *E. coli*.

Além disso, outras linhas de células hospedeiras auxotróficas incluem a estirpe auxotrófica de Phe (AF) baseada em K10, estirpes auxotróficas duplas de Phe/Trp (AFW), estirpes auxotróficas triplas de Phe/Trp/Lys (AFWK), um auxotrófico de Met (M15MA em fundo M15), assim como estirpe AF baseada em DH10B.

Células que podem ser usadas para a prática de certas formas de realização divulgadas neste documento incluem células hospedeiras auxotróficas (quer seja sejam procarióticas ou eucarióticas). Células auxotróficas podem apresentar níveis de auxotrofia simples, duplos, triplos, quádruplos ou maiores (cada nível de auxotrofia indicando

um composto orgânico particular que o organismo é incapaz de sintetizar e tem de ser fornecido à célula). Certas formas de realização divulgadas neste documento expressamente não utilizam uma célula hospedeira auxotrófica. Na medida em que uma célula hospedeira auxotrófica não é usada, outra célula ou componentes celulares podem ainda ser usados para praticar certas formas de realização divulgadas neste documento. Outras formas de realização podem usar uma, duas, três ou mais células hospedeiras auxotróficas diferentes que podem ser dos mesmos ou diferentes estirpes ou organismos.

Células hospedeiras são geneticamente modificadas (p.ex., transformadas, transduzidas ou transfectadas) com os vetores desta divulgação, que podem ser, por exemplo, um vetor de clonagem ou um vetor de expressão. O vetor pode ser, por exemplo, sob a forma de um plasmídeo, uma bactéria, um vírus, um polinucleótido nu ou um polinucleótido conjugado. Os vetores são introduzidos em células e/ou microorganismos por métodos padrão incluindo eletroporação (From et al., PNAS. USA 82, 5824 (1985)), infecção por vetores virais, penetração por balística de alta velocidade com pequenas partículas com o ácido nucleico dentro da matriz de pequenas esferas ou partículas, ou na superfície (Klein et al., Nature 327, 70-73 (1987)). Berger, Sambrook, e Ausubel fornecem uma variedade de métodos de transformação adequados.

As células hospedeiras modificadas por engenharia podem ser cultivadas em meios nutritivos convencionais modificados conforme apropriado para atividades tais como, por exemplo, as etapas de seleção, ativar os promotores ou selecionar

transformantes. Estas células podem opcionalmente ser cultivadas em organismos transgênicos.

Certas formas de realização divulgadas neste documento incluem adicionalmente métodos de seleção de AARSs modificadas e/ou tRNAs modificados. Por exemplo, numa forma de realização, uma biblioteca de PheRS de levedura é submetida a seleção de triagem dupla a fim de detetar altos níveis de incorporação de um aminoácido não-natural ou incorporação incorreta de aminoácidos naturais além de Phe vai levar a graves enrolamentos incorretos ou desenrolamentos de GFP. As células da biblioteca de PheRS de levedura estão portanto sujeitas a seleção de elevado processamento baseada na análise de citometria de fluxo (FACS). Primeiro, as células de biblioteca de PheRS de levedura são expressas na presença de 2Nal e células de fluorescência baixa indicando maior incorporação de 2Nal ou de outros aminoácidos naturais são recolhidas por FACS. De seguida, as células de biblioteca de PheRS de levedura são expressas sem 2Nal. Células brilhantes são recolhidas de forma a eliminar as variantes de PheRS de levedura que podem incorporar incorretamente outros aminoácidos naturais. Numa forma de realização particular, dois ciclos de seleção renderam uma PheRS mutante de levedura com mutações em N412G, S418C, T415G e S437F, que tinha baixa fluorescência na presença de 2Nal e elevada fluorescência na ausência de 2Nal. Esta técnica permite a incorporação de 2Nal no codão UUU, aumentando para cerca de 90%.

Outras referências úteis, p.ex., para o isolamento e cultura de células (p.ex., para o subsequente isolamento de ácidos nucleicos) incluem Freshney (1994) *Culture of Animal Cells, a Manual of Basic Technique*, third edition, Wiley-

Liss, New York, e as referências citadas nesse documento; Payne et al. (1992) *Plant Cell and Tissue Culture in Liquid Systems* John Wiley & Sons, Inc. New York, N.Y.; Gamborg and Phillips (eds.) (1995) *Plant Cell, Tissue and Organ Culture; Fundamental Methods Springer Lab Manual*, Springer-Verlag (Berlin Heidelberg New York) and Atlas and Parks (eds.) *The Handbook of Microbiological Media* (1993) CRC Press, Boca Raton, Fla.

Vários métodos conhecidos para introduzir ácidos nucleicos alvo em células bacterianas estão disponíveis, qualquer dos quais pode ser usado em certas formas de realização divulgadas neste documento. Estes incluem: fusão das células recetoras com protoplastos bacterianas contendo o DNA, eletroporação, bombardeamento de projéteis e infecção com vetores virais, etc. Células bacterianas podem ser usadas para amplificar o número de plasmídeos contendo construções de DNA de certas formas de realização divulgadas neste documento. As bactérias são crescidas até à fase exponencial e os plasmídeos dentro das bactérias podem ser isolados por uma variedade de métodos conhecidos na arte (ver, por exemplo, Sambrook). Além disso, uma infinidade de *kits* estão disponíveis comercialmente para a purificação de plasmídeos de bactérias, (ver, p.ex., EASYPREP™, FLEXIPREP™, ambos da Pharmacia Biotech; STRATACLEAN™, da STRATAGENE; e, QIAPREP™ da Qiagen). Os plasmídeos isolados e purificados são então adicionalmente manipulados para produzir outros plasmídeos, usados para transfetar células ou incorporados em vetores relacionados para infetar organismos.

Vetores típicos contêm terminadores de transcrição e de tradução, sequências de iniciação de transcrição e

tradução, e promotores úteis para regulação da expressão do ácido nucleico alvo específico. Os vetores opcionalmente incluem cassetes de expressão génica contendo pelo menos uma sequência de terminador independente, sequências permitindo a replicação da cassette em eucariotas, ou procariotas, ou ambos, (p.ex., vetores transportadores) e marcadores de seleção para sistemas tanto procariotas como eucariotas. Vetores são adequados para replicação e integração em procariotas, eucariotas, ou preferencialmente ambos. Ver Gilman & Smith, *Gene* 8:81 (1979); Roberts, et al., *Nature*, 328:731 (1987); Schneider, B., et al., *Protein Expr. Purif.* 6435:10 (1995); Ausubel, Sambrook, Berger (todos supra). Um catálogo de Bactérias e Bacteriófagos úteis para clonagem é fornecido, p.ex., pela ATCC, p.ex., *The ATCC Catalogue of Bacteria and Bacteriophage* (1992) Gherna et al. (eds.), publicado pela ATCC. Procedimentos básicos adicionais para sequenciação, clonagem e outros aspetos da biologia molecular e considerações teóricas subjacentes também são encontrados em Watson et al. (1992) *Recombinant DNA 2nd Edition* Scientific American Books, NY.

Seleção de Codões Degenerados

Conforme descrito acima, todos os aminoácidos, com exceção da metionina e triptofano são codificados por mais de um codão. De acordo com alguns métodos divulgados neste documento, um codão no genoma que normalmente é usado para codificar um aminoácido natural é reprogramado, em parte pela maquinaria de transcrição ou de tradução, para em vez daquele, codificar um análogo de aminoácido. Um análogo de aminoácido pode ser um análogo de aminoácido natural ou canónico. Numa forma de realização preferida, o análogo de aminoácido não é um aminoácido canonicamente codificado.

A estabilidade termodinâmica de um par codão-anticodão pode ser prevista ou determinada experimentalmente. De acordo com algumas formas de realização, é preferível que o tRNA mutante externo interaja com o codão degenerado com uma afinidade (a 37°C) pelo menos cerca de 1,0 kcal/mol mais fortemente, mais preferencialmente 1,5 kcal/mole mais fortemente e ainda mais preferencialmente mais de 2,0 kcal/mol mais fortemente do que um tRNA natural na célula reconheceria a mesma sequência. Estes valores são conhecidos por um perito na arte e podem ser determinados por experiências de desnaturação térmica (ver, p.ex., Meroueh and Chow, *Nucleic Acids Res.* 27: 1118, 1999).

A tabela seguinte lista algumas das sequências anticodão conhecidas para *E. coli*. Em geral, para qualquer organismo, a sequência anticodão do tRNA pode ser rotineiramente determinada usando tecnologias reconhecidas na arte. Por exemplo, qualquer gene de tRNA pode ser amplificado por, por exemplo, PCR. Sequenciação pode ser realizada para determinar as sequências exatas do *loop* de anticodão. Alternativamente, ensaio de ligação bioquímico pode ser usado para determinar a afinidade de ligação de um tRNA purificado a um dos 2-6 codões possíveis. O codão que liga o tRNA com a mais alta especificidade / afinidade presumivelmente tem emparelhamentos puros de Watson-Crick em todas as três posições do codão, deste modo determinando a sequência do *loop* de anticodão.

Em geral, a base *wobble* no *loop* de anticodão tende a ser G ou U (em vez de A ou C).

Os Codões Degenerados para *E. coli*

Aminoácido	Anticodão	Emparelhamento na 3ª base	Codão	Aminoácido	Anticodão	Emparelhamento de bases	Codão
Ala	GGC	W/C ¹	GCC	His	GUG	W/C	CAC
		Wobble ²	GCU			Wobble	CAU
	UGC	W/C	GCA	Ile	GAU	W/C	AUC
		Wobble	GCG			Wobble	AUU, AUA
Asp	GUC	W/C	GAC	Leu	GAG	W/C	CUC, CUA, CUG, UUC, UUG
		Wobble	GAU			Wobble	CUU
Asn	GUU	W/C	AAC	Lys	UUU	W/C	AAA
		Wobble	AAU			Wobble	AAG
Cys	GCA	W/C	UGC	Phe	GAA	W/C	UUC
		Wobble	UGU			Wobble	UUU
Glu	UUC	W/C	GAA	Ser	GGA	W/C	UUC, AGU
		Wobble	GAG			Wobble	UCU, AGC, UCA, UCG
Gly	GCC	W/C	GGC, GGA, GGG	Tyr	GUA	W/C	UAC
		Wobble	GGU			Wobble	UAU
Met		W/C	AUG	Thr		W/C	ACC, ACA, ACG
Gln		W/C	CAA, CAG			Wobble	ACU
Arg		W/C	AGA, AGG, CGU, CGG	Pro		W/C	CCC, CCA, CCG
		Wobble	CGC, CGA	Trp		Wobble	CCU
					W/C	UGG	
STOP		W/C	UGA, UAA	Val		W/C	GUC, GUA
		Wobble	UAG			Wobble	GUU, GUG
¹ Emparelhamento de bases de Watson-Crick							
² Emparelhamento de bases Wobble							

Quando um único tRNA reconhece um codão através de uma interação complementar perfeita entre o anticodão do tRNA e um codão, é denominado emparelhamento de bases de Watson-

Crick. Quando um único tRNA reconhece um segundo codão, degenerado, é denominado emparelhamento de bases *wobble* ou outros não-padrão. Em certas formas de realização divulgadas neste documento, um novo tRNA pode ser construído possuindo uma sequência anticodão que é perfeitamente complementar com um codão degenerado ou um codão para um aminoácido não-natural, utilizando assim emparelhamento de bases *wobble* ou de Watson-Crick. Da mesma forma, uma nova AARS pode ser construída que utiliza um aminoácido de substituição (que não o de tipo selvagem - pode ser outro aminoácido natural ou um aminoácido não-natural) para aminoacilar o tRNA correspondente. Isto pode ser em adição a ou em vez de modificar uma molécula de tRNA para incorporação de um aminoácido de substituição.

A AARS modificada pode ser alterada de modo a que a eficiência de ligação para o aminoácido não-natural, ou outro aminoácido natural selecionado, é maior do que a eficiência de ligação da AARS modificada para o correspondente aminoácido natural. Desta forma, uma AARS modificada pode preferencialmente ligar um aminoácido não-natural a fim de carregar um tRNA mesmo na presença do aminoácido natural que corresponde à AARS no seu estado inalterado. Esta "reprogramação" de uma aminoacil-tRNA sintetase permite a incorporação de um aminoácido não-natural num polipéptido com níveis inferiores de incorporação incorreta de outros aminoácidos no local desejado.

A "reprogramação" pode adicionalmente permitir o uso da sintetase modificada ou mutante externa com altos níveis de incorporação em células hospedeiras padrão, sem a necessidade de células hospedeiras auxotróficas, e com ou

sem esgotar o meio do aminoácido natural correspondente. Portanto, enquanto certas formas de realização divulgadas neste documento podem ser praticadas usando uma célula hospedeira auxotrófica, certas outras formas de realização podem ser praticadas sem o uso de uma célula hospedeira auxotrófica. No caso de não se usar uma célula hospedeira auxotrófica para praticar certas formas de realização, outra célula hospedeira pode ser usada, componentes celulares podem ser usados, ou um sistema inteiramente sem células pode ser usado.

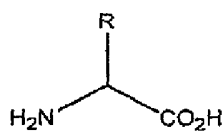
Quando a célula tem múltiplas moléculas de tRNA para um aminoácido particular, e um tRNA possui uma sequência anticodão que é perfeitamente complementar ao codão degenerado selecionado, o gene que codifica o tRNA pode ser desabilitado através de quaisquer meios disponíveis a um perito na arte, incluindo, por exemplo, mutagénese sítio-dirigida ou deleção do gene ou da sequência do promotor do gene. Expressão do gene pode também ser desativada através de quaisquer técnicas antissentido ou de RNA de interferência.

Aminoácidos Antinaturais ou Não-Naturais

O primeiro passo no processo de engenharia de proteínas é geralmente selecionar um conjunto de aminoácidos antinaturais ou não-naturais que possuem as propriedades químicas desejadas. A seleção de aminoácidos não-naturais depende de propriedades químicas pré-determinadas que um gostaria de ter, e as modificações que um gostaria de fazer na proteína alvo. Aminoácidos antinaturais, uma vez selecionados, podem ser adquiridos a partir de fornecedores, ou quimicamente sintetizados.

Uma grande variedade de aminoácidos antinaturais ou não-naturais pode ser usada nos métodos divulgados neste documento. O aminoácido antinatural pode ser escolhido com base em características desejadas do aminoácido antinatural, p.ex., função do aminoácido antinatural, tais como modificar propriedades biológicas de proteínas tais como toxicidade, biodistribuição, imunogenicidade ou meia-vida, propriedades estruturais, propriedades espectroscópicas, propriedades químicas e/ou fotoquímicas, propriedades catalíticas, capacidade de reagir com outras moléculas (covalentemente ou não covalentemente), ou semelhantes.

Como usado neste documento um "aminoácido antinatural" refere-se a qualquer aminoácido, aminoácido modificado ou análogo de aminoácido diferentes de selenocisteína e os seguintes vinte alfa-aminoácidos geneticamente codificados: alanina, arginina, asparagina, ácido aspártico, cisteína, glutamina, ácido glutâmico, glicina, histidina, isoleucina, leucina, lisina, metionina, fenilalanina, prolina, serina, treonina, triptofano, tirosina, valina. A estrutura genérica de um alfa-aminoácido é ilustrada pela fórmula I:



Formula I

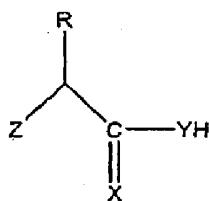
Um aminoácido antinatural é normalmente qualquer estrutura possuindo a fórmula I na qual o grupo R é qualquer substituinte diferente dos usados nos vinte aminoácidos naturais. Ver, p.ex., qualquer texto de bioquímica tal como

Biochemistry by L. Stryer, 3rd ed. 1988, Freeman and Company, New York, para estruturas dos vinte aminoácidos naturais. Observe-se que os aminoácidos antinaturais divulgados neste documento podem ser compostos naturais que não sejam os vinte alfa-aminoácidos acima. Porque os aminoácidos antinaturais divulgados neste documento normalmente diferem dos aminoácidos naturais na cadeia lateral apenas, os aminoácidos antinaturais formam ligações amida com outros aminoácidos, p.ex., naturais ou antinaturais, da mesma forma em que são formados em proteínas naturais. No entanto, os aminoácidos antinaturais possuem grupos de cadeia lateral que os distinguem dos aminoácidos naturais. Por exemplo, R na fórmula I opcionalmente compreende um grupo alquil-, aril-, halogeneto de arilo, halogeneto de vinilo, halogeneto de alquilo, acetilo, cetona, aziridina, nitrilo, nitro, halogeneto, acil-, ceto-, azido-, hidroxil-, hidrazina, ciano-, halo-, hidrazida, alceno, alcino, éter, tioéter, epóxido, sulfona, ácido borónico, éster boronato, borano, ácido fenilborónico, tiol, seleno-, sulfonil-, borato, boronato, fosfo, fosfono, fosfina, heterocíclico-, piridil, naftil, benzofenona, um anel restrito tal como um ciclooctina; tioéster, enona, imina, aldeído, éster, tioácido, hidroxilamina, amino, ácido carboxílico, ácido carboxílico alfa-ceto, ácidos e amidas alfa ou beta insaturados, amida glioxil, ou organossilano, ou semelhantes ou qualquer combinação dos mesmos.

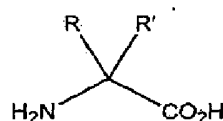
Substituições de arilo podem ocorrer em várias posições, p.ex., orto, meta, para e com um ou mais grupos funcionais colocados no anel arilo. Outros aminoácidos antinaturais de interesse incluem, mas não estão limitados a, aminoácidos compreendendo um elemento de reticulação foto-ativável,

aminoácidos marcados com *spin*, aminoácidos marcados com corante, aminoácidos fluorescentes, aminoácidos de ligação a metais, aminoácidos contendo metais, aminoácidos radioativos, aminoácidos com novos grupos funcionais, aminoácidos com alterada hidrofiliicidade, hidrofobicidade, polaridade ou capacidade de estabelecer pontes de hidrogénio, aminoácidos que covalentemente ou não covalentemente interagem com outras moléculas, aminoácidos *photocaged* e/ou fotoisomerizáveis, aminoácidos compreendendo biotina ou um análogo de biotina, aminoácidos glicosilados tal como uma serina com açúcar substituído, outros aminoácidos modificados com carboidratos, aminoácidos contendo ceto, aminoácidos compreendendo polietilenoglicol ou um poliéter, um poliálcool, ou um polissacárido, aminoácidos que podem sofrer metátese, aminoácidos que podem sofrer cicloadições, aminoácidos com substituições de átomos pesados, aminoácidos quimicamente cliváveis e/ou foto-cliváveis, aminoácidos com cadeias laterais alongadas em comparação com aminoácidos naturais, p.ex., poliéteres ou hidrocarbonetos de cadeia longa, p.ex., maior do que cerca de 5 ou maior do que cerca de 10 carbonos, aminoácidos ligados a carbono contendo açúcar, aminoácidos ativos em redox, aminoácidos contendo amino tioácido, aminoácidos contendo um grupo de medicamento, e aminoácidos compreendendo um ou mais grupos tóxicos.

Em adição a aminoácidos antinaturais que contêm novas cadeias laterais, aminoácidos antinaturais também opcionalmente compreendem estruturas de cadeia principal modificada, p.ex., conforme ilustrado pelas estruturas da Fórmula II e III:



Formula II



Formula III

em que Z tipicamente compreende OH, NH₂, SH, NH₂O-, NH-R', R'NH-, R'S-, ou S-R'-; X e Y, que podem ser os mesmos ou diferentes, tipicamente compreendem S, N, ou O, e R e R', que opcionalmente são os mesmos ou diferentes, são tipicamente selecionados da mesma lista de componentes para o grupo R acima descrito para os aminoácidos antinaturais possuindo Fórmula I assim como hidrogénio ou (CH₂)_x ou as cadeias laterais de aminoácidos naturais. Por exemplo, aminoácidos antinaturais divulgados neste documento opcionalmente incluem substituições no grupo amina ou carboxilo conforme ilustrado por fórmulas II e III. Aminoácidos antinaturais deste tipo incluem, mas não estão limitados a, ácidos α-hidroxi, α-tioácidos α-aminotiocarboxilados, ou aminoácidos α-α-di-substituídos, com cadeias laterais correspondendo p.ex. aos vinte aminoácidos naturais ou a cadeias laterais antinaturais. Também incluem mas não estão limitados a β-aminoácidos ou γ-aminoácidos, tais como β-alanina substituída e ácido butírico γ-amino. Além disso, substituições ou modificações no α-carbono opcionalmente incluem isómeros L ou D, tais como o D-glutamato, D-alanina, D-metil-O-tirosina, ácido aminobutírico e semelhantes. Outras alternativas estruturais incluem aminoácidos cíclicos, tais como análogos de prolina, assim como análogos de prolina em anel de 3-, 4-, 6-, 7-, 8- e 9- membros. Alguns aminoácidos não-naturais, tais como halogenetos de arilo (p-bromo-

fenilalanina, p-iodofenilalanina, fornecem reações versáteis de acoplamento cruzado catalisadas por paládio com reações de etino ou acetileno que permitem a formação de ligações carbono-carbono, carbono-nitrogénio e carbono-oxigénio entre halogenetos de arilo e uma grande variedade de parceiros de acoplamento.

Por exemplo, muitos aminoácidos antinaturais são baseados em aminoácidos naturais, tais como a tirosina, fenilalanina, glutamina, e semelhantes. Análogos de tirosina incluem tirosinas para-substituídas, tirosinas orto-substituídas, e tirosinas meta-substituídas, em que a tirosina substituída compreende um grupo acetilo, um grupo benzoílo, um grupo amino, uma hidrazina, um hidroxiamina, um grupo tiol, um grupo carboxilo, um grupo isopropilo, um grupo metilo, um hidrocarboneto C₆-C₂₀ de cadeia linear ou ramificada, um hidrocarboneto saturado ou insaturado, um grupo O-metil, um grupo poliéter, um grupo nitro, ou semelhantes. Além disso, múltiplos anéis de arilo substituído também são contemplados. Análogos de glutamina incluem, mas não estão limitados a, derivados α -hidroxi, derivados β -substituídos, derivados cíclicos, e derivados de glutamina com amida substituída. Análogos de fenilalanina exemplares incluem, mas não estão limitados a, fenilalaninas meta-substituídas, em que o substituinte compreende um grupo hidroxilo, um grupo metóxi, um grupo metilo, um grupo alilo, um grupo acetilo, ou semelhantes.

Exemplos específicos de aminoácidos antinaturais incluem, mas não estão limitados a, formas o, m e/ou p de aminoácidos ou análogos de aminoácidos (aminoácidos não-naturais), incluindo homoalilglicina, cis- ou trans-crotilglicina, ácido 6,6,6-trifluoro-2-aminohexanóico,

ácido 2-aminoheptanóico, norvalina, norleucina, O-metil-L-tirosina, o-, m- ou p-metil-fenilalanina, O-4-alil-L-tirosina, uma 4-propil-L-tirosina, uma tri-O-acetil-GlcNAc β -serina, um L-Dopa, uma fenilalanina fluorada, uma isopropil-L-fenilalanina, uma p-azidofenilalanina, uma p-acil-L-fenilalanina, uma p-benzoil-L-fenilalanina, uma L-fosfoserina, uma fosfonoserina, uma fosfonotirosina, uma p-iodo-fenilalanina, o-, m-, ou p-bromofenilalanina, 2-,3-, ou 4-piridilalanina, p-idiofenilalanina, ácido diaminobutírico, ácido aminobutírico, benzofuranilalanina, 3-bromo-tirosina, 3-(6-cloroindolil)alanina, 3-(6-bromoindolil)alanina, 3-(5-bromonindolil)alanina, p-clorofenilalanina, p-etinil-fenilalanina, p-propargli-óxi-fenilalanina, m-etinil-fenilalanina, 6-etinil-triptofano, 5-etinil-triptofano, ácido (R)-2-amino-3-(4-etinil-1H-pirol-3-il)propanóico, azidonorleucina, azidohomoalanina, p-acetilfenilalanina, p-amino-L-fenilalanina, homopropargliglicina, p-etil-fenilalanina, p-etinil-fenilalanina, p-propargli-óxi-fenilalanina, isopropil-L-fenilalanina, uma 3-(2-naftil)alanina, 3-(1-naftil)alanina, 3-idio-tirosina, O-propargil-tirosina, homoglutamina, uma O-4-alil-L-tirosina, uma 4-propil-L-tirosina, uma 3-nitro-L-tirosina, uma tri-O-acetil-GlcNAc β -serina, uma L-Dopa, uma fenilalanina fluorada, uma isopropil-L-fenilalanina, uma p-azido-L-fenilalanina, uma p-acil-L-fenilalanina, uma p-acetil-L-fenilalanina, uma m-acetil-L-fenilalanina, selenometionina, telurometionina, selenocisteína, uma fenilalanina alcina, uma O-alil-L-tirosina, uma O-(2-propinil)-L-tirosina, uma p-etiltiocarbonil-L-fenilalanina, uma p-(3-oxobutanoil)-L-fenilalanina, uma p-benzoil-L-fenilalanina, uma L-fosfoserina, uma fosfonoserina, uma fosfonotirosina, homopropargliglicina, azidohomoalanina, uma p-iodo-fenilalanina, uma p-bromo-L-fenilalanina,

dihidroxi-fenilalanina, dihidroxil-L-fenilalanina, uma p-nitro-L-fenilalanina, uma m-metóxi-L-fenilalanina, uma p-iodo-fenilalanina, uma p-bromofenilalanina, uma p-amino-L-fenilalanina, e uma isopropil-L-fenilalanina, trifluoroleucina, norleucina, 4-, 5-, ou 6-fluorotriptofano, 4-aminotriptofano, 5-hidroxitriptofano, biocitina, ácido aminooxiacético, m-hidroxifenilalanina, m-alil fenilalanina, grupo m-metoxifenilalanina, β -GlcNAc-serina, α -GalNAc-treonina, p-acetoacetilfenilalanina, para-halo-fenilalanina, seleno-metionina, etionina, S-nitroso-homocisteína, thia-prolina, 3-tienilo-alanina, homo-alil-glicina, trifluoroisoleucina, ácido trans e cis-2-amino-4-hexenóico, 2-butinil-glicina, alil-glicina, para-azido-fenilalanina, para-ciano-fenilalanina, para-etinil-fenilalanina, hexafluoroleucina, 1,2,4-triazolo-3-alanina, 2-fluoro-histidina, histidina L-metil, 3-metil-L-histidina, β -2-tienil-L-alanina, β -(2-tiazolil)-DL-alanina, homopropargliglicina (HPG) e azidohomoalanina (AHA) e semelhantes. As estruturas de uma variedade de aminoácidos antinaturais não limitantes são fornecidas nas figuras, p.ex., FIGS. 29, 30, e 31 de US 2003/0108885 A1, cujo conteúdo completo está incorporado neste documento por referência.

Análogos de tirosina incluem tirosinas para-substituídas, tirosinas orto-substituídas, e tirosinas meta-substituídas, em que a tirosina substituída compreende um grupo acetilo, um grupo benzoílo, um grupo amino, uma hidrazina, uma hidroxiamina, um grupo tiol, um grupo carboxi, um grupo isopropil, um grupo metilo, um hidrocarboneto C6-C20 de cadeia linear ou ramificada, um hidrocarboneto saturado ou insaturado, um grupo O-metil, um grupo poliéter, um grupo nitro, ou semelhantes. Além disso, múltiplos anéis de arilo

substituído também são contemplados. Os análogos de glutamina da invenção incluem, mas não estão limitados a, derivados α -hidroxi-, derivados β -substituídos, derivados cíclicos, e derivados de glutamina com amida substituída. Análogos de fenilalanina exemplares incluem, mas não estão limitados a, fenilalaninas meta-substituídas, em que o substituinte compreende um grupo hidroxi, um grupo metóxi, um grupo metilo, um grupo alilo, um grupo acetilo, ou semelhantes.

Adicionalmente, outros exemplos opcionalmente incluem (mas não estão limitados a) um análogo antinatural de um aminoácido tirosina; um análogo antinatural de um aminoácido glutamina; um análogo antinatural de um aminoácido fenilalanina; um análogo antinatural de um aminoácido serina; um análogo antinatural de um aminoácido treonina; um alquil, aril, acil, azido, ciano, halo, hidrazina, hidrazida, hidroxilo, alceno, alcino, éter, tiol, sulfonil, seleno, éster, tioácido, borato, boronato, fosfo, fosfono, fosfina, heterocíclicos, enona, imina, aldeído, hidroxilamina, ceto, ou aminoácido com amino substituído, ou qualquer combinação dos mesmos; um aminoácido com um elemento de reticulação foto-ativável; um aminoácido marcado com *spin*; um aminoácido fluorescente; um aminoácido com um novo grupo funcional; um aminoácido que covalentemente ou não covalentemente interage com outra molécula; um aminoácido de ligação a metais; um aminoácido contendo metais; um aminoácido radioativo; um aminoácido *photocaged*; um aminoácido fotoisomerizável; um aminoácido contendo biotina ou análogo de biotina; um aminoácido glicosilado ou com carboidrato modificado; um aminoácido contendo ceto; um aminoácido compreendendo polietilenoglicol; um aminoácido compreendendo poliéter; um

aminoácido substituído com átomo pesado; um aminoácido clivável quimicamente ou foto-clivável; um aminoácido com uma cadeia lateral alongada; um aminoácido contendo um grupo tóxico; um aminoácido substituído com açúcar, p.ex., uma serina substituída com açúcar ou semelhantes; um aminoácido contendo açúcar ligado a carbono; um aminoácido ativo em redox; um aminoácido contendo α -hidroxi; um aminoácido contendo tio-aminoácido; um aminoácido α,α di-substituído; um β -aminoácido; e um aminoácido cíclico.

Normalmente, os aminoácidos antinaturais utilizados neste documento para certas formas de realização podem ser selecionados ou projetados para fornecer características adicionais indisponíveis nos vinte aminoácidos naturais. Por exemplo, aminoácidos naturais são opcionalmente projetados ou selecionados para modificar as propriedades biológicas de uma proteína, p.ex., em que são incorporados. Por exemplo, as propriedades seguintes são opcionalmente modificadas por inclusão de um aminoácido natural numa proteína: toxicidade, biodistribuição, solubilidade, estabilidade, p.ex., térmica, hidrolítica e oxidativa, resistência a degradação enzimática, e semelhantes, facilidade de purificação e de processamento, propriedades estruturais, propriedades espectroscópicas, propriedades químicas e/ou fotoquímicas, atividade catalítica, potencial redox, meia-vida, capacidade de reagir com outras moléculas, p.ex., covalentemente ou não covalentemente, e semelhantes.

Outros exemplos de análogos de aminoácidos incluem (mas não estão limitados a) um análogo antinatural de um aminoácido tirosina; um análogo antinatural de um aminoácido glutamina; um análogo antinatural de um aminoácido

fenilalanina; um análogo antinatural de um aminoácido serina; um análogo antinatural de um aminoácido treonina; um alquil, aril, acil, azido, ciano, halo, hidrazina, hidrazida, hidroxilo, alceno, alcino, éter, tiol, sulfonil, seleno, éster, tioácido, borato, boronato, fosfo, fosfono, fosfina, heterocíclicos, enona, imina, aldeído, hidroxilamina, ceto, ou aminoácido com amina substituída, ou qualquer combinação dos mesmos; um aminoácido com um elemento de reticulação foto-ativável; um aminoácido marcado com *spin*; um aminoácido fluorescente; um aminoácido com um novo grupo funcional; um aminoácido que covalentemente ou não covalentemente interage com outra molécula; um aminoácido de ligação a metais; um aminoácido contendo metais; um aminoácido radioativo; um aminoácido *photocaged*; um aminoácido fotoisomerizável; um aminoácido contendo biotina ou análogo de biotina; um aminoácido glicosilado ou com carboidrato modificado; um aminoácido contendo ceto; um aminoácido compreendendo polietilenoglicol; um aminoácido compreendendo poliéter; um aminoácido substituído com átomo pesado; um aminoácido clivável quimicamente ou foto-clivável; um aminoácido com uma cadeia lateral alongada; um aminoácido contendo um grupo tóxico; um aminoácido substituído com açúcar, p.ex., uma serina substituída com açúcar ou semelhantes; um aminoácido contendo açúcar ligado a carbono; um aminoácido ativo em redox; um aminoácido contendo α -hidroxi; um aminoácido contendo tio-aminoácido; um aminoácido α,α di-substituído; um β -aminoácido; e um aminoácido cíclico diferente de prolina.

Aminoacil-tRNA Sintetases

A aminoacil-tRNA sintetase (usado de modo intercambiável neste documento com AARS, RS ou "sintetase") usada em certos métodos divulgados neste documento pode ser uma sintetase natural derivada de um organismo, quer seja o mesmo (homóloga) ou diferente (heteróloga), uma sintetase mutante ou modificada, ou uma sintetase projetada.

A sintetase usada pode reconhecer o análogo de aminoácido (antinatural) desejado seletivamente em relação a aminoácidos relacionados disponíveis. Por exemplo, quando o análogo de aminoácido a ser usado é estruturalmente relacionado com um aminoácido natural, a sintetase deveria carregar a molécula de tRNA mutante externa com o análogo de aminoácido desejado com uma eficiência pelo menos substancialmente equivalente à de um, e mais preferencialmente pelo menos cerca de duas vezes, 3 vezes, 4 vezes, 5 vezes ou mais do que aquela do aminoácido natural. No entanto, nos casos em que não é necessário um produto de proteína bem definido, a sintetase pode ter especificidade relaxada para carregar aminoácidos. Numa tal forma de realização, uma mistura de tRNAs mutantes externos poderia ser produzida, com vários aminoácidos ou análogos.

Em certas formas de realização, é preferível que a sintetase possua atividade tanto para o análogo de aminoácido como para o aminoácido que é codificado pelo codão correspondente da molécula de tRNA.

Uma sintetase pode ser obtida por uma variedade de técnicas conhecidas de um perito na arte, incluindo combinações de técnicas como, por exemplo, métodos computacionais, métodos

de seleção, e incorporação de sintetases de outros organismos (ver abaixo).

Em certas formas de realização, sintetases podem ser usadas ou desenvolvidas que carregam eficientemente moléculas de tRNA que não são carregadas por sintetases da célula hospedeira. Por exemplo, pares apropriados podem geralmente ser desenvolvidos através da modificação de sintetases de organismos distintos da célula hospedeira. Em certas formas de realização, a sintetase pode ser desenvolvida por processos de seleção. Em certas formas de realização, a sintetase pode ser projetada usando técnicas computacionais tais como as descritas em Datta et al., J. Am. Chem. Soc. 124: 5652-5653, 2002, e na patente dos EUA No. 7,139,665, incorporados neste documento por referência.

Desenho Computacional de AARS

Especificamente, numa forma de realização, o método em questão parcialmente depende do *design* e engenharia da AARS natural para uma forma modificada que possui especificidade de substrato relaxada, de tal forma que pode usar análogos de aminoácidos não-canônicos como um substrato, e carregar um tRNA modificado (com o seu anticodão alterado) com tal aminoácido não-canônico. As secções seguintes descrevem resumidamente um método para a geração de tal AARS modificada, tal método é descrito em mais detalhe na patente dos EUA N° 7,139,665, cujo conteúdo completo é incorporado neste documento por referência.

Resumidamente, os métodos de algumas formas de realização descritos naqueles documentos relacionam-se com ferramentas computacionais para modificar a especificidade de substrato

de uma aminoacil-tRNA sintetases (AARSs) através de mutação para tornar a enzima mais eficiente ao utilizar análogos de aminoácidos em sistemas de tradução de proteínas, ou *in vitro*, em células inteiras, ou noutros sistemas de tradução. Uma característica de algumas formas de realização inclui re-desenhar sistematicamente o sítio de ligação do substrato de uma enzima AARS para facilitar o uso de substratos antinaturais no péptido ou reação de tradução de proteína que a enzima catalisa.

De acordo com um método, uma biblioteca de rotâmeros para o aminoácido artificial é construída variando os seus ângulos de torção para criar rotâmeros que caberiam na cavidade de ligação do substrato natural. A orientação geométrica da cadeia principal do análogo de aminoácido é especificada pela orientação cristalográfica da cadeia principal do substrato natural na estrutura cristalina. A estrutura cristalográfica da sintetase de aminoácidos específica de organismo pode ser usada, ou uma estrutura homóloga de outro organismo pode ser usada, dependendo da similaridade estrutural. Aminoácidos na cavidade de ligação da sintetase que interagem com a cadeia lateral do análogo são permitidas variar na identidade e conformação de rotâmero em cálculos de conceção de proteína subsequentes.

Um tal protocolo também emprega um método computacional para melhorar as interações entre o substrato e as posições de proteína. Isto é feito aumentando a escala das energias dos pares entre o substrato e os aminoácidos permitidos nas posições de *design* na proteína nos cálculos de energia. Num cálculo de otimização onde as interações da proteína-substrato são ampliadas em comparação com as interações intra-proteína, a seleção de sequência é enviesada no

sentido de selecionar aminoácidos a serem aqueles que têm uma interação favorável com o substrato.

O método descrito ajudou a construir uma nova forma modificada da fenilalanil-tRNA sintetase de *E. coli*, baseada na estrutura conhecida da PheRS relacionada de *Thermus thermophilus* (tPheRS). A nova forma modificada da fenilalanil-tRNA sintetase (ePheRS) de *E. coli* permite eficiente incorporação *in vivo* da funcionalidade de cetona aril reativa em proteínas recombinantes. Além disso, uma triptofanil-tRNA sintetase modificada foi modificada de forma semelhante e tem demonstrado a capacidade de incorporar os análogos de aminoácidos não-naturais em polipéptidos no lugar de triptofano natural. Os resultados descritos naquele documento também demonstram o poder geral de *design* computacional de proteínas no desenvolvimento de aminoacil-tRNA sintetases para ativação e carregamento de aminoácidos não-naturais.

A. Informação Estrutural e de Sequência para tRNA Sintetases Disponível

Tradução da proteína a partir de um molde de mRNA é realizada por ribossomas. Durante o processo de tradução; cada tRNA é combinado com seu aminoácido muito tempo antes de alcançar o ribossoma. A correspondência é feita por uma coleção de enzimas conhecidas como as aminoacil-tRNA sintetases (AARS). Estas enzimas carregam cada tRNA com o aminoácido apropriado, permitindo assim que cada tRNA efetue a tradução apropriada do código genético do DNA (e do mRNA transcrito a partir do DNA) para o código de aminoácidos das proteínas.

A maioria das células faz vinte aminoacil-tRNA sintetases diferentes, uma para cada tipo de aminoácido. Estas vinte enzimas são cada uma otimizadas para a função com o seu próprio aminoácido particular e o conjunto de moléculas de tRNA apropriadas para esse aminoácido. Aminoacil-tRNA sintetases têm de executar as suas tarefas com precisão elevada. Muitas destas enzimas reconhecem as suas moléculas de tRNA usando o anticodão. Estas enzimas cometem um erro em 10000. Para a maioria dos aminoácidos, este nível de precisão não é muito difícil de alcançar, já que a maioria dos aminoácidos são bastante diferentes uns dos outros.

O método em questão, uma descrição exata da cavidade de ligação da tRNA é importante para a abordagem de *design* computacional, uma vez que depende da estrutura cristalina para as descrições de cadeia principal da proteína, embora em muitos casos é perfeitamente aceitável usar a estrutura cristalina de uma proteína homóloga (por exemplo, um homólogo de uma espécie relacionada) ou mesmo um domínio conservado para substituir a descrição de estrutura cristalográfica da cavidade de ligação. A estrutura cristalina também define a orientação do substrato aminoácido natural na cavidade de ligação de uma sintetase, assim como a posição do substrato aminoácido relativamente aos resíduos da sintetase, especialmente aqueles resíduos dentro e em torno da cavidade de ligação. Para projetar a cavidade de ligação para os análogos, é preferível que estes análogos se liguem à sintetase na mesma orientação do substrato aminoácido natural, uma vez que esta orientação pode ser importante para a etapa de adenilação.

As AARSs podem ser de qualquer organismo, incluindo procariotas e eucariotas, com enzimas de bactérias, fungos,

extremófilos tais como as arqueobactérias, nemátodos, insetos, peixes, anfíbios, aves, animais (particularmente mamíferos e particularmente humanos) e plantas todos possíveis.

Conforme descrito acima, a maioria das células fazem vinte aminoacil-tRNA sintetases diferentes, uma para cada tipo de aminoácido. Algumas sintetases adequadas são conhecidas, incluindo: fenilalanil-tRNA sintetase de levedura (Kwon et al., J. Am. Chem. Soc. 125: 7512-7513, 2003); tirosil-tRNA sintetase de *Methonococcus jannaschii* (Wang et al., Science 292, 498-500, 2001); e tirosil-tRNA sintetase de levedura (Ohno et al., J. Biochem. 130, 417-423, 2001). Na verdade, as estruturas cristalinas de quase todas as 20 diferentes enzimas AARS estão atualmente disponíveis no Brookhaven Protein Data Bank (PDB, ver Bernstein et al., J. Mol. Biol. 112: 535-542, 1977). Uma lista de todas as AARSs com estruturas cristalinas resolvidas a partir de Abril de 2001 está disponível no sítio Web do PDB. Por exemplo, a estrutura cristalina da fenilalanil-tRNA sintetase de *Thermus Aquaticus* complexada com Fenilalanina tem uma resolução de 2,7 Å, e a sua ID PDB é 1B70.

A base de dados de estrutura ou Molecular Modeling DataBase (MMDB) contém dados experimentais de determinações cristalográficas e de estrutura de NMR. Os dados para MMDB são obtidos do Protein Data Bank (PDB). O NCBI (National Center for Biotechnology Information) tem dados estruturais de cruzados com informações bibliográficas, com as bases de dados de sequências, e com a taxonomia NCBI. Cn3D, o visualizador da estrutura 3D do NCBI, pode ser usado para fácil visualização interativa de estruturas moleculares de Entrez.

A base de dados Entrz 3D Domains contém domínios da proteína da Conserved Domain Database (CDD) do NCBI. Biólogos computacionais definem domínios conservados com base em padrões ou motivos de sequências recorrentes. CDD atualmente contém domínios derivados de duas coleções populares, Smart e Pfam, além de contribuições de colegas no NCBI, tais como COG. As bases de dados de fonte também fornecem descrições e *links* para citações. Uma vez que domínios conservados correspondem a unidades estruturais compactas, CDs contém *links* para estrutura 3D via Cn3D sempre que possível.

Para identificar domínios conservados numa sequência de proteína, o serviço CD-Search emprega o algoritmo BLAST de posição inversa específica. A sequência de pesquisa é comparada com uma matriz de pontuação de posição específica preparada a partir do alinhamento do domínio conservado subjacente. *Hits* podem ser exibidos como um alinhamento emparelhado da sequência de pesquisa com uma sequência de domínio representante, ou como um alinhamento múltiplo. CD-Search é agora executado por defeito em paralelo com pesquisas BLAST de proteínas. Enquanto o usuário aguarda que a fila de BLAST processe o pedido adiante, a arquitetura de domínio da pesquisa pode já ser estudada. Além disso, o CDART, Conserved Domain Architecture Retrieval Tool permite ao usuário pesquisar proteínas com arquiteturas de domínio semelhantes. CDART usa resultados CD-Search de pré-computação para rapidamente identificar proteínas com um conjunto de domínios semelhantes àquele da pesquisa. Para mais detalhes, ver Marchler-Bauer et al., Nucleic Acids Research 31: 383-387, 2003; e Marchler-Bauer et al., Nucleic Acids Research 30: 281-283, 2002.

Além disso, uma base de dados de aminoacil-tRNA sintetases conhecidas foi publicado por Maciej Szymanski, Marzanna A. Deniziak e Jan Barciszewski, em *Nucleic Acids Res.* 29:288-290, 2001 (entitulado "Aminoacyl-tRNA synthetases database"). Um sítio Web correspondente (rose.man.poznan.pl/aars/seq_main.html) fornece detalhes sobre todas as AARSs conhecidas de espécies diferentes. Por exemplo, de acordo com a base de dados, a Isoleucil-tRNA Sintetase das bactérias radiorresistentes *Deinococcus radiodurans* (Accession No. AAF10907) possui 1078 aminoácidos, e foi publicada por White et al. em *Science* 286:1571-1577(1999); a Valil-tRNA Sintetase de ratinho (*Mus Musculus*) possui 1263 aminoácidos (Accession No. AAD26531), e foi publicada por Snoek M. e van Vugt H. em *Immunogenetics* 49: 468-470(1999); e as sequências da Fenilalanil-tRNA Sintetase para humanos, *Drosophila*, *S. pombe*, *S. cerevisiae*, *Candida albicans*, *E. coli*, e várias outras bactérias incluindo *Thermus aquaticus ssp. thermophilus* estão também disponíveis. A base de dados foi atualizada em Novembro de 2006. Informação semelhante para outras AARSs recentemente identificadas pode ser obtida, por exemplo, realizando uma pesquisa BLAST usando qualquer uma das sequências conhecidas na base de dados da AARS como pesquisa contra as bases de dados públicas (como a base de dados não-redundante no NCBI, ou "nr") ou privadas proprietárias disponíveis.

Alternativamente, em certas formas de realização, se a estrutura cristalina exata de uma AARS particular não é conhecida, mas a sua sequência de proteína é similar ou homóloga a uma sequência de AARS conhecida com uma estrutura cristalina conhecida. Em tais casos, é expectável que a conformação da AARS em questão será semelhante à

estrutura cristalina conhecida da AARS homóloga. A estrutura conhecida pode, portanto, ser usada como a estrutura para a AARS de interesse, ou mais preferencialmente, pode ser usada para prever a estrutura da AARS de interesse (ou seja, em "modelação de homologia" ou "modelação molecular"). Como um exemplo particular, o Molecular Modeling Database (MMDB) descrito anteriormente (ver Wang et al., Nucl. Acids Res. 2000, 28:243-245; Marchler-Bauer et al., Nucl. Acids Res. 1999, 27:240-243) fornece motores de busca que podem ser usados para identificar proteínas ou ácidos nucleicos que são similares ou homólogos a uma sequência de proteína (referida como sequências "vizinhas" no MMDB), incluindo sequências vizinhas cujas estruturas tridimensionais são conhecidas. A base de dados fornece adicionalmente *links* para as estruturas conhecidas juntamente com ferramentas de alinhamento e visualização, como Cn3D (desenvolvido pelo NCBI), RasMol, etc., através do qual as sequências homólogas e precursoras podem ser comparadas e uma estrutura pode ser obtida para a sequência precursora com base em tais alinhamentos de sequências e estruturas conhecidas.

A sequência da AARS homóloga com estrutura 3D conhecida é preferencialmente pelo menos cerca de 60%, ou pelo menos cerca de 70%, ou pelo menos cerca de 80%, ou pelo menos cerca de 90%, ou pelo menos cerca de 95% idêntica, ou pelo menos cerca de 98% idêntica à da AARS de interesse na região do sítio ativo ou a região da cavidade de ligação do substrato aminoácido. Tal sítio ativo ou sítio de cavidade pode não ser contínuo na sequência primária de aminoácidos da AARS uma vez que aminoácidos distantes podem unir-se na estrutura 3D. Neste caso, homologia ou identidade de

sequência pode ser calculada usando-se, por exemplo, os programas padrão BLASTp do NCBI para proteínas usando condições por defeito, em regiões alinhadas juntas (sem inserções ou deleções em qualquer uma das duas sequências sendo comparadas) e incluindo resíduos conhecidos estarem envolvidos na ligação do substrato aminoácido. Por exemplo, a subunidade catalítica (alfa) da fenilalanil-tRNA sintetase de *Thermus Aquaticus* aparenta ter uma região de "inserção" do resíduo 156 a 165 quando comparado com os seus homólogos de outras espécies. Esta região pode ser desconsiderada no cálculo de identidade de sequência. Alternativamente, a AARS homóloga é preferencialmente cerca de 35%, ou 40%, ou 45%, ou 50% ou 55% globalmente idêntica à AARS de interesse. A subunidade catalítica (alfa) da fenilalanil-tRNA sintetase de *E. coli* é cerca de 45% globalmente idêntica e cerca de 80% idêntica na região do sítio ativo à fenilalanil-tRNA sintetase de *Thermus Aquaticus*. As subunidades alfa da fenilalanil-tRNA sintetase humana é cerca de 62%, 60%, 54%, 50% idênticas em geral às suas homólogas de *Drosophila*, nemátodo (*C. elegans*), planta (*Arabidopsis thaliana*), levedura (*S. cerevisiae*), respetivamente.

Nos poucos casos onde a estrutura para uma determinada sequência de AARS pode não ser conhecida ou estar disponível, é possível determinar a estrutura usando técnicas experimentais rotineiras (por exemplo, de cristalografia de raios-X e espectroscopia de Ressonância Magnética Nuclear (RMN)) e sem experimentação indevida. Ver, p.ex., NMR of Macromolecules: A Practical Approach, G. C. K. Roberts, Ed., Oxford University Press Inc., New York (1993); Ishima and Torchia, Nat. Struct. Biol. 7: 740-743, 2000; Gardner and Kay, Annu. Rev. Bioph. Biom. 27: 357-406,

1998; Kay, Biochem. Cell. Biol. 75: 1-15, 1997; Dayie et al., Annu. Rev. Phys. Chem. 47: 243-282, 1996; Wuthrich, Acta Crystallogr. D 51: 249-270, 1995; Kahn et al., J. Synchrotron Radiat. 7: 131-138, 2000; Oakley and Wilce, Clin. Exp. Pharmacol. P. 27: 145-151, 2000; Fourme et al., J. Synchrotron Radiat. 6: 834-844, 1999.

Alternativamente, a estrutura tridimensional de uma sequência de AARS pode ser calculada a partir da própria sequência e utilizando técnicas de modelação molecular *ab initio* já conhecidas na arte. Ver p.ex., Smith et al., J. Comput. Biol. 4: 217-225, 1997; Eisenhaber et al., Proteins 24: 169-179, 1996; Bohm, Biophys Chem. 59: 1-32, 1996; Fetrow and Bryant, BioTechnol. 11: 479-484, 1993; Swindells and Thornton, Curr. Opin. Biotech. 2: 512-519, 1991; Levitt et al., Annu. Rev. Biochem. 66: 549-579, 1997; Eisenhaber et al., Crit. Rev. Biochem. Mol. 30: 1-94, 1995; Xia et al., J. Mol. Biol. 300: 171-185, 2000; Jones, Curr. Opin. Struc. Biol. 10: 371-379, 2000. Estruturas tridimensionais obtidas de modelação *ab initio* são tipicamente menos fidedignas do que estruturas obtidas usando técnicas empíricas (p.ex., espectroscopia de RMN ou cristalografia de raios-X) ou semi-empíricas (p.ex., de modelação por homologia). No entanto, tais estruturas geralmente serão de qualidade suficiente, embora menos preferidas, para uso em alguns métodos divulgados neste documento.

B. Métodos para Prever a Estrutura 3D com base na Homologia de Sequência

Para proteínas AARS que não tenham sido cristalizadas ou sido o foco de outras determinações estruturais, um modelo molecular da AARS gerado por computador e o seu sítio de

ligação podem no entanto ser gerados usando qualquer uma de um número de técnicas disponíveis na arte. Por exemplo, as posições de C α -carbono da sequência de AARS alvo podem ser mapeadas num determinado padrão de coordenadas de uma enzima AARS ("AARS conhecida"), possuindo uma sequência similar e estrutura deduzida usando técnicas de modelação por homologia, e a estrutura da proteína alvo e velocidades de cada átomo calculadas a uma temperatura de simulação (T_0) em que uma simulação de encaixe com um análogo de aminoácido será determinada. Normalmente, um tal protocolo envolve principalmente a previsão de conformações de cadeia lateral na proteína AARS alvo modelada, enquanto assumindo um traço da cadeia principal retirada de uma estrutura terciária, tal como fornecida pela proteína AARS conhecida. Programas de computador para executar rotinas de minimização de energia são comumente usados para gerar modelos moleculares. Por exemplo, ambos os algoritmos CHARMM (Brooks et al. (1983) J Comput Chem 4:187-217) e AMBER (Weiner et al (1981) J. Comput. Chem. 106: 765) lidam com toda a configuração do sistema molecular, cálculo de campo de forças e análise (ver também, Eisenfield et al. (1991) Am J Physiol 261:C376-386; Lybrand (1991) J Pharm Belg 46:49-54; Froimowitz (1990) Biotechniques 8:640-644; Burbam et al. (1990) Proteins 7:99-111; Pedersen (1985) Environ Health Perspect 61:185-190; e Kini et al. (1991) J Biomol Struct Dyn 9:475-488). No centro destes programas está um conjunto de sub-rotinas que, dada a posição de cada átomo no modelo, calculam a energia potencial total do sistema e a força em cada átomo. Estes programas podem utilizar um conjunto inicial de coordenadas atômicas, os parâmetros para os vários termos da função de energia potencial, e uma descrição da topologia molecular (estrutura covalente). Características comuns de tais

métodos de modelação molecular incluem: provisões para a manipulação de pontes de hidrogénio e outras forças de limitação; o uso de condições de delimitação periódicas; e provisões para ocasionalmente ajustar posições, velocidades ou outros parâmetros a fim de manter ou mudar a temperatura, pressão, volume, forças de limitação, ou outras condições controladas externamente.

A maioria dos métodos de minimização de energia mais convencionais usam os dados de coordenadas de entrada e o facto de que a função de energia potencial é uma função diferenciável e explícita de coordenadas cartesianas para calcular a energia potencial e o seu gradiente (que fornece a força em cada átomo) para qualquer conjunto de posições atómicas. Esta informação pode ser usada para gerar um novo conjunto de coordenadas num esforço para reduzir a energia potencial total e, repetindo este processo várias vezes, para otimizar a estrutura molecular sob um determinado conjunto de condições externas. Estes métodos de minimização de energia são rotineiramente aplicados às moléculas similares a proteínas AARS em questão.

Em geral, métodos de minimização de energia podem ser realizados para uma dada temperatura, T_i , que pode ser diferente da temperatura de simulação de encaixe, T_o . Durante a minimização de energia da molécula em T_i , coordenadas e velocidades de todos os átomos do sistema são computadorizadas. Além disso, os modos normais do sistema são calculados. Será apreciado pelos peritos na arte que cada modo normal é um movimento coletivo, periódico, com todas as partes do sistema movendo-se em fase umas com as outras, e que o movimento da molécula é a sobreposição de todos os modos normais. Para uma dada temperatura, a

amplitude média quadrática de movimento num modo particular é inversamente proporcional à constante de força efetiva para esse modo. A este respeito, as vibrações de baixa frequência irão frequentemente dominar o movimento da molécula.

Após a energia do modelo molecular ter sido minimizada em T_i , o sistema é "aquecido" ou "arrefecido" para a temperatura de simulação, T_o , através da realização de uma execução de equilíbrio onde as velocidades dos átomos são dimensionadas de forma faseada até à temperatura desejada, T_o , ser alcançada. O sistema é adicionalmente equilibrado durante um período de tempo especificado até certas propriedades do sistema, tais como a energia cinética média, permanecerem constantes. As coordenadas e velocidades de cada átomo são então obtidas a partir do sistema equilibrado.

Novas rotinas de minimização de energia também podem ser realizadas. Por exemplo, uma segunda classe de métodos envolve calcular soluções aproximadas para a EOM limitada para a proteína. Estes métodos usam uma abordagem iterativa para resolver os multiplicadores de Lagrange e, tipicamente, apenas necessitam algumas iterações se as correções necessárias são pequenas. O método mais popular deste tipo, SHAKE (Ryckaert et al. (1977) J. Comput. Phys. 23:327; e Van Gunsteren et al. (1977) Mol. Phys. 34:1311) é fácil de implementar e dimensiona como $O(N)$ à medida que o número de limitações aumenta. Portanto, o método é aplicável a macromoléculas tais como proteínas AARS. Um método alternativo, RATTLE (Anderson (1983) J. Comput. Phys. 52:24) é baseado na versão de velocidade do algoritmo de Verlet. Como SHAKE, RATTLE é um algoritmo iterativo e

pode ser usado para minimizar energia no modelo de uma proteína AARS em questão.

C. Métodos Alternativos

Noutras formas de realização, em vez de manter a identidade do análogo de aminoácido constante e variando a estrutura da AARS (modelando várias estruturas mutantes diferentes), o método em questão é realizado usando o(s) modelo(s) molecular(es) para uma única AARS modificada (p.ex., em que um ou mais resíduos de aminoácido não-âncora são alterados) e testando uma variedade de diferentes análogos de aminoácidos ou potenciais fragmentos dos mesmos, para identificar análogos que são prováveis interagir com, e ser substratos para a enzima AARS modificada. Esta abordagem pode fazer uso de bibliotecas coordenadas para análogos de aminoácidos (incluindo variantes de rotâmeros) ou bibliotecas de grupos funcionais e espaçadores que podem ser unidos para formar a cadeia lateral de um análogo do aminoácido.

Usando tais abordagens conforme descrito acima, p.ex., modelação por homologia, uma coordenada definida para o sítio de ligação para a AARS modificada pode ser derivada.

Há uma variedade de métodos computacionais que podem ser prontamente adaptados para identificar a estrutura de análogos de aminoácidos que teriam propriedades estéricas e eletrônicas adequadas para interagir com o sítio de ligação do substrato de uma AARS modificada. Ver, p.ex., Cohen et al. (1990) J. Med. Cam. 33: 883-894; Kuntz et al. (1982) J. Mol. Biol 161: 269-288; DesJarlais (1988) J. Med. Cam. 31: 722-729; Bartlett et al. (1989) (Spec. Publ., Roy. Soc.

Chem.) 78: 182-196; Goodford et al. (1985) J. Med. Cam. 28: 849-857; DesJarlais et al. J. Med. Cam. 29: 2149-2153). Métodos dirigidos geralmente dividem-se em duas categorias: (1) *design* por analogia no qual estruturas 3D de moléculas conhecidas (tais como de uma base de dados cristalográficos) são encaixadas na estrutura do sítio de ligação da AARS e classificados relativamente à qualidade do encaixe; e (2) *design de novo*, em que o modelo do análogo de aminoácido é construído peça a peça no sítio de ligação da AARS. A última abordagem, em particular, pode facilitar o desenvolvimento de novas moléculas, exclusivamente projetadas para se ligarem ao sítio de ligação da AARS modificada em questão.

Numa forma de realização ilustrativa, o *design* dos análogos de aminoácidos potenciais que podem funcionar com uma AARS modificada particular começa a partir da perspetiva geral de forma complementar ao sítio de ligação do substrato da enzima, e um algoritmo de busca é utilizado que é capaz de digitalizar uma base de dados de pequenas moléculas de estrutura tridimensional conhecida para candidatos que se encaixam geometricamente no sítio de ligação do substrato. Tais bibliotecas podem ser bibliotecas gerais de moléculas pequenas, ou podem ser bibliotecas direcionadas para os análogos de aminoácidos ou pequenas moléculas que podem ser usadas para criar análogos de aminoácidos. Não é esperado que as moléculas encontradas na pesquisa de forma serão necessariamente pistas em si mesmo, uma vez que nenhuma avaliação de interação química é necessariamente efetuada durante a pesquisa inicial. Em vez disso, prevê-se que tais candidatos podem atuar como a base para *design* adicional, fornecendo esqueletos moleculares nos quais substituições atômicas apropriadas podem ser efetuadas. Claro, a

complementaridade química destas moléculas pode ser avaliada, mas é esperado que tipos de átomo serão alterados para maximizar as interações eletrostáticas, por pontes de hidrogénio e hidrofóbicas com o sítio de ligação do substrato. A maioria dos algoritmos deste tipo fornecem um método para encontrar uma grande variedade de estruturas químicas que podem ser complementares à forma do sítio de ligação do substrato da AARS.

Por exemplo, cada um de um conjunto de pequenas moléculas de uma base de dados particular, tais como Cambridge Crystallographic Data Bank (CCDB) (Allen et al. (1973) J. Chem. Doc. 13: 119), é individualmente encaixado no sítio de ligação da AARS modificada numa série de orientações geometricamente permissíveis com uso de um algoritmo de encaixe. Numa forma de realização preferida, um conjunto de algoritmos de computador chamado DOCK, pode ser usado para caracterizar a forma de invaginações e sulcos que formam o sítio de ligação. Ver, por exemplo, Kuntz et al. (1982) J. Mol. Biol 161: 269-288. O programa também pode pesquisar uma base de dados de pequenas moléculas para moldes cujas formas são complementares ao sítio de ligação particular da AARS modificada. Algoritmos exemplares que podem ser adaptados para esta finalidade são descritos em, por exemplo, DesJarlais et al. (1988) J. Med. Chem. 31:722-729.

As orientações são avaliadas relativamente à qualidade de encaixe e as melhores são mantidas para exame adicional usando programas de mecânica molecular, tais como ABER ou CHARMM. Tais algoritmos foram anteriormente bem sucedidos em encontrar uma variedade de moléculas que são complementares na forma a um sítio de ligação determinado de um recetor ou enzima, e foram demonstrados possuírem

várias características atraentes. Em primeiro lugar, tais algoritmos podem disponibilizar uma notável diversidade de arquiteturas moleculares. Em segundo lugar, as melhores estruturas, em aplicações anteriores a outras proteínas, demonstraram complementaridade de forma impressionante sobre uma área de superfície estendida. Em terceiro lugar, a abordagem global aparenta ser bastante robusta em relação a pequenas incertezas no posicionamento dos átomos candidatos.

Em certas formas de realização, o método em questão pode utilizar um algoritmo descrito por Goodford (1985, J. Med. Chem. 28:849-857) e Boobbyer et al. (1989, J. Med. Chem. 32:1083-1094). Estes documentos descrevem um programa de computador (GRID) que procura determinar regiões de elevada afinidade para diferentes grupos químicos (denominados sondas) na superfície molecular do sítio de ligação. GRID fornece portanto uma ferramenta para sugerir modificações para ligandos conhecidos que podem melhorar a ligação. Pode prever-se que alguns dos sítios discernidos por GRID como regiões de elevada afinidade correspondem a "padrões farmacofóricos" determinados inferencialmente a partir de uma série de ligandos conhecidos. Como usado neste documento, um padrão farmacofórico é um arranjo geométrico das características do análogo de aminoácido antecipado que se acredita ser importante para a ligação. Goodsell e Olson (1990, Proteins: Struct. Funct. Genet. 8:195-202) usaram o algoritmo Metropolis (emparelhamento simulado) para encaixar um único ligando conhecido numa proteína alvo, e a sua abordagem pode ser adaptada para a identificação de análogos de aminoácidos adequados para encaixar no sítio de ligação da AARS. Este algoritmo pode permitir flexibilidade torsional na cadeia lateral do aminoácido e usar mapas de

energia de interação GRID como tabelas de pesquisa rápida para a computação de energias de interação aproximadas.

Ainda uma forma de realização adicional utiliza um algoritmo de computador tal como CLIX que pesquisa tais bases de dados como CCDB para pequenas moléculas que podem ser orientadas no sítio de ligação do substrato da AARS de uma forma que é tanto estericamente aceitável como tem uma alta probabilidade de alcançar interações químicas favoráveis entre a molécula candidata e os resíduos de aminoácidos circundantes. O método baseia-se na caracterização do sítio de ligação do substrato em termos de um conjunto de posições favoráveis de ligação para diferentes grupos químicos e depois procura orientações das moléculas candidatas que causam coincidência espacial máxima de grupos químicos individuais candidatos com membros do conjunto. A disponibilidade atual de poder de computador dita que uma pesquisa baseada em computador para novos ligandos segue uma estratégia em amplitude. Uma estratégia em amplitude visa reduzir progressivamente o tamanho do espaço de pesquisa de potenciais candidatos através da aplicação de critérios cada vez mais rigorosos, em oposição a uma estratégia em amplitude em que uma análise detalhada ao máximo de um candidato é realizada antes de prosseguir para a próxima. CLIX está de acordo com esta estratégia na medida em que a sua análise de ligação é rudimentar - visa satisfazer as condições necessárias de encaixe estérico de ter grupos individuais em locais "corretos" para a ligação, sem impor a condição suficiente que interações de ligação favoráveis realmente ocorrem. Uma "lista" classificada de moléculas, nas suas orientações favorecidas, é produzida que pode então ser examinada numa base de molécula a molécula, usando computação gráfica e

técnicas mais sofisticadas de modelação molecular. CLIX também é capaz de sugerir mudanças para os grupos substituintes químicos das moléculas do candidato que podem melhorar a ligação. Novamente, a biblioteca inicial pode ser de análogos de aminoácidos ou de moléculas que podem ser usadas para gerar a cadeia lateral de um análogo de aminoácido.

Os detalhes do algoritmo de CLIX estão descritos em Lawrence et al. (1992) Proteins 12:31-41, e o algoritmo CLIX pode ser resumido como o seguinte. O programa GRID é usado para determinar posições de interação favorável distintas (denominadas sítios alvo) no sítio de ligação da proteína AARS para uma grande variedade de grupos químicos representativos. Para cada ligando candidato no CCDB uma tentativa exaustiva é efetuada para tornar coincidente, num sentido espacial do sítio de ligação da proteína, um par de grupos químicos substituintes do candidato com um par de sítios de interação favorável correspondentes proposto por GRID. Todas as combinações possíveis de pares de grupos de ligandos com pares de sítios de GRID são consideradas durante este procedimento. Após localizar tal coincidência, o programa gira o ligando candidato sobre os dois pares de grupos e verifica se há bloqueio estérico e coincidência de outros grupos atômicos candidatos com sítios alvo apropriados. Combinações de candidato/orientação específicas que são bons encaixes geométricos no sítio de ligação e mostram coincidência suficiente dos grupos atômicos com sítios de GRID são retidas.

Coerente com a estratégia em amplitude, esta abordagem envolve simplificar as suposições. Geometria de proteína rígida e de pequena molécula é mantida ao longo do

processo. Como uma primeira aproximação geometria rígida é aceitável como as coordenadas de energia minimizada do sítio de ligação da AARS modificada, descreve um mínimo de energia para a molécula, embora uma local.

Uma suposição adicional implícita em CLIX é que o ligando potencial, quando introduzido no sítio de ligação do substrato da AARS modificada, não induz a mudança na estereoquímica da proteína ou a distribuição de carga parcial e então altera a base sobre a qual os mapas de interação de energia foram calculados. Importa também salientar que os sítios de interação previstos por GRID são usados num sentido posicional e de tipo apenas, ou seja, quando um grupo atômico candidato é colocado num sítio previsto como favorável por GRID, nenhuma verificação é feita para garantir que a geometria de ligação, o estado de protonação, ou a distribuição de carga parcial favorece uma forte interação entre a proteína e esse grupo. Tal análise detalhada deve formar parte da modelação mais avançada dos candidatos identificados na lista CLIX.

Ainda outra forma de realização de um método molecular de *design* assistido por computador para a identificação de análogos de aminoácido que podem ser utilizados por uma AARS modificada pré-determinada compreende a síntese *de novo* de potenciais inibidores por conexão algorítmica de pequenos fragmentos moleculares que irá expor a desejada complementaridade estrutural e eletrostática com o sítio de ligação do substrato da enzima. Cada estágio de crescimento de ligando é avaliado de acordo com uma função de energia baseada em mecânica molecular, que considera interações de van der Waals e coulombiana, energia de deformação interna do ligando alongado e dessolvatação tanto do ligando como

da enzima. O espaço de pesquisa pode ser gerenciado pelo uso de uma árvore de dados que é mantida sob controle por corte de acordo com os critérios de ligação.

Ainda noutra forma de realização, análogos de aminoácidos potenciais podem ser determinados usando um método baseado num algoritmo de dinâmica molecular reprimida por minimização de energia para determinar posições energeticamente favoráveis de grupos funcionais no sítio de ligação do substrato de uma enzima AARS modificada. O método pode ajudar no *design* de moléculas que incorporam tais grupos funcionais por modificação do aminoácido conhecido e análogos de aminoácidos ou através de síntese *de novo*.

Por exemplo, o método de pesquisa simultânea de múltiplas cópias (MCSS) descrito por Miranker et al. (1991) Proteins 11: 29-34 pode ser adaptado para uso no método em questão. Para determinar e caracterizar mínimos locais de um grupo funcional no campo de forças da proteína, múltiplas cópias de grupos funcionais selecionados são inicialmente distribuídas num sítio de ligação de interesse na proteína AARS. Minimização de energia destas cópias por mecânica molecular ou dinâmica reprimida produz mínimos locais distintos. A vizinhança destes mínimos pode então ser explorada por uma pesquisa GRID ou por minimização limitada. Numa forma de realização, o método MCSS usa a clássica aproximação dependente de tempo de Hartee (TDH) para simultaneamente minimizar ou reprimir muitos grupos idênticos no campo de forças da proteína.

Implementação do algoritmo MCSS requer uma escolha de grupos funcionais e de um modelo de mecânica molecular para

cada um destes. Grupos devem ser simples o suficiente para ser facilmente caracterizados e manipulados (3-6 átomos, poucos ou nenhum grau(s) de liberdade diédrico(s)), mas complexos o suficiente para aproximar as interações eletrostáticas e estéricas que o grupo funcional teria na ligação do substrato ao sítio da proteína AARS. Um conjunto preferido é, por exemplo, um em que as moléculas orgânicas mais podem ser descritas como uma coleção de tais grupos (Patai's Guide to the Chemistry of Functional Groups, ed. S. Patai (New York: John Wiley, and Sons, (1989)). Isto inclui fragmentos como acetonitrilo, metanol, acetato, metil amônia, éter dimetílico, metano, e acetaldeído.

Determinação dos mínimos locais de energia no sítio de ligação requer que muitas posições iniciais sejam analisadas. Isto pode ser conseguido distribuindo, por exemplo, 1000-5000 grupos de forma aleatória dentro de uma esfera centrada no sítio de ligação; apenas o espaço não ocupado pela proteína necessita de ser considerado. Se a energia de interação de um grupo particular num determinado local com a proteína é mais positiva do que um determinado limite (por exemplo, 5,0 kcal/mole) o grupo é descartado desse sítio. Dado o conjunto de posições iniciais, todos os fragmentos são minimizados simultaneamente pelo uso da aproximação TDH (Elber et al. (1990) J. Am. Chem. Soc. 112: 9161-9175). Neste método, as forças em cada fragmento consistem das suas forças internas e aquelas devido à proteína. O elemento essencial deste método é que as interações entre os fragmentos são omitidas e as forças na proteína são normalizadas para aquelas devido a um único fragmento. Desta forma a simultânea minimização ou dinâmica de qualquer número de grupos funcionais no campo de uma única proteína pode ser efetuada.

Minimização é efetuada sucessivamente em subconjuntos de, p.ex., 100, dos grupos aleatoriamente colocados. Depois de um certo número de intervalos entre etapas, tais como 1000 intervalos, os resultados podem ser examinados para eliminar grupos que convergem para o mesmo mínimo. Este processo é repetido até a minimização ser concluída (p.ex., o gradiente RMS de 0,01 kcal/mole/Å). Portanto o conjunto de moléculas de energia minimizada resultante compreende o que equivale a um conjunto de fragmentos desconectados em três dimensões que representam cadeias laterais potenciais para análogos de aminoácidos.

O próximo passo então é ligar os pedaços com espaçadores montados a partir de pequenas entidades químicas (átomos, cadeias, ou grupos de anéis) para formar os análogos de aminoácidos, p.ex., cada um dos desconectados pode ser ligado no espaço para gerar uma única molécula usando tais programas de computador tais como, por exemplo, NEWLEAD (Tschinke et al. (1993) J. Med. Chem. 36: 3863,3870). O procedimento adotado por NEWLEAD executa a seguinte sequência de comandos (1) conectar dois grupos isolados, (2) reter as soluções intermediárias para processamento adicional, (3) repetir os passos acima para cada uma das soluções intermediárias até unidades desconectadas não serem encontradas, e (4) apresentar as soluções finais, cada uma delas é molécula única. Um tal programa pode usar, por exemplo, três tipos de espaçadores: espaçadores de biblioteca, espaçadores de átomo único, e espaçadores de anéis fusionados. Os espaçadores de biblioteca são estruturas otimizadas de pequenas moléculas tais como etileno, benzeno e metilamida. Os resultados produzidos por programas tais como NEWLEAD consistem de um conjunto de

moléculas que contém os fragmentos originais agora ligados por espaçadores. Os átomos pertencentes aos fragmentos de entrada mantêm as suas orientações originais no espaço. As moléculas são quimicamente plausíveis por causa da formação simples dos espaçadores e grupos funcionais, e energeticamente aceitáveis por causa da rejeição de soluções com violações de raios de van der Waals.

Além disso, a ordem em que as etapas deste método são efetuadas é meramente ilustrativa por natureza. Na verdade, as etapas podem ser efetuadas em qualquer ordem ou em paralelo, salvo indicado em contrário pela presente divulgação.

Além disso, os métodos divulgados neste documento podem ser realizados em *hardware*, *software* ou qualquer combinação destes, uma vez que estes termos são atualmente conhecidos na arte. Em particular, o presente método pode ser realizado por *software*, *firmware*, ou microinstrução operando num computador ou computadores de qualquer tipo. Além disso, o *software* pode incluir instruções de computador em qualquer forma (p.ex., código fonte, código objeto, código interpretado, etc.) armazenadas em qualquer suporte informático (p.ex., ROM, RAM, *media* magnética, fita perfurada ou cartão, disco compacto (CD) em qualquer forma, DVD, etc.). Além disso, tal *software* pode também estar sob a forma de um sinal de dados de computador incorporado numa onda portadora, tais como aquelas encontradas dentro das conhecidas páginas Web transferidas entre dispositivos ligados à Internet. Da mesma forma, certas formas de realização não estão limitadas a qualquer plataforma específica, salvo especificamente indicado de outra forma na presente divulgação.

Meios exemplares de *hardware* de computador adequados para a realização de certas formas de realização podem ser um servidor Silicon Graphics Power Challenge com 10 processadores R10000 funcionando em paralelo. Ambiente de desenvolvimento de *software* adequado inclui CERIUS2 por Biosym/Molecular Simulations (San Diego, CA), ou outros equivalentes.

O método computacional descrito acima tem sido usado eficientemente para modificar enzimas da maquinaria de síntese de proteínas (p.ex., AARS) para permitir a incorporação de aminoácidos antinaturais. O mesmo conjunto de ferramentas computacionais também pode ser aproveitado para projetar os produtos finais (p.ex., anticorpos monoclonais ou outros terapêuticos) em que os aminoácidos antinaturais seriam incorporados a fim de melhorar ou modificar as suas propriedades estruturais ou funcionais.

Enquanto formas de realização particulares divulgadas neste documento foram apresentadas e descritas, será evidente para os peritos na arte que alterações e modificações podem ser feitas sem se desviarem do aspeto mais amplo e, portanto, as reivindicações anexadas englobam no seu âmbito todas as tais alterações e modificações que sejam abrangidas no verdadeiro espírito desta invenção.

Adoção de AARS de Diferentes Organismos

Uma segunda estratégia para gerar um tRNA mutante externo, RS modificada ou mutante externa ou par tRNA/RS modificado envolve a importação de um tRNA e/ou sintetase de outro organismo para o sistema de tradução de interesse, tais

como *Escherichia coli*. Neste exemplo específico, as propriedades da sintetase heteróloga candidata incluem, p.ex., que não carrega o tRNA de *Escherichia coli* razoavelmente bem (preferencialmente não carrega em absoluto), e as propriedades do tRNA heterólogo candidato incluem, p.ex., que não é acilado pela sintetase de *Escherichia coli* numa medida razoável (preferencialmente não é acilado em absoluto).

Schimmel et al. relataram que a GlnRS de *Escherichia coli* (EcGlnRS) não acila o tRNAGln de *Saccharomyces cerevisiae* (EcGlnRS não possui um domínio N-terminal de ligação ao RNA possuído por GlnRS de *Saccharomyces cerevisiae* (ScGlnRS)). Ver, E. F. Whelihan and P. Schimmel, EMBO J., 16:2968 (1997). Por exemplo, o tRNAGln supressor âmbar de *Saccharomyces cerevisiae* (SctRNAGlnCUA) foi analisado para determinar se também não é um substrato para a EcGlnRS. Ensaio de aminoacilação *in vitro* mostraram que este é o caso; e estudos de supressão *in vitro* mostram que o SctRNAGlnCUA é competente na tradução. Ver, p.ex., Liu and Schultz, PNAS. U S A, 96:4780 (1999). Foi adicionalmente demonstrado que ScGlnRS não acila qualquer tRNA de *Escherichia coli*, somente o SctRNAGlnCUA *in vitro*. O grau em que ScGlnRS é capaz de aminoacilar o SctRNAGlnCUA em *Escherichia coli* também foi avaliado usando um ensaio de complementação *in vivo*. Uma mutação nonsense âmbar foi introduzida num local permissivo no gene da β -lactamase. Supressão da mutação por um tRNA supressor âmbar deve produzir β -lactamase completa e confere resistência a ampicilina à célula. Quando apenas SctRNAGlnCUA é expresso, as células exibem um IC₅₀ de 20 μ g/mL ampicilina, indicando virtualmente nenhuma acilação por sintetases endógenas de *Escherichia coli*; quando SctRNAGlnCUA é co-expresso com

ScGlnRS, as células adquirem um IC₅₀ de cerca de 500 µg/mL ampicilina, demonstrando que ScGlnRS acila SctRNAGlnCUA eficientemente em *Escherichia coli*. Ver, Liu and Schultz, PNAS, U S A, 96:4780 (1999).

Como outro exemplo, tRNA^{Asp} de *Saccharomyces cerevisiae* é conhecido ser um mutante externo para sintetases de *Escherichia coli*. Ver, p.ex., Doctor and Mudd, J. Biol. Chem., 238:3677 (1963); and, Kwok and Wong, Can. J. Biochem., 58:213 (1980). Foi demonstrado que um tRNA supressor âmbar derivado dele (SctRNA^{Asp}_{CUA}) também é um mutante externo em *Escherichia coli* usando o ensaio de β-lactamase *in vivo* descrito acima. No entanto, o anticodão de tRNA^{Asp} é um elemento de reconhecimento crítico da AspRS, ver, p.ex., Giege, et al, Biochimie, 78:605 (1996), e a mutação do anticodão para CUA resulta numa perda de afinidade do supressor para AspRS. Um mutante de AspRS E93K de *Escherichia coli* foi mostrado reconhecer SctRNA^{Asp}_{CUA} supressor âmbar de *Escherichia coli* uma ordem de magnitude melhor do que AspRS de tipo selvagem. Ver, p.ex., Martin, 'Thesis', Universite Louis Pasteur, Strasbourg, France, 1995. Foi especulado que a introdução da mutação relacionada em AspRS de *Saccharomyces cerevisiae* (E188K) pode restaurar a sua afinidade para SctRNA^{Asp}_{CUA}. Foi determinado que o mutante AspRS (E188K) de *Saccharomyces cerevisiae* não acila tRNAs de *Escherichia coli*, mas carrega SctRNA^{Asp}_{CUA} com eficiência moderada como mostrado por experiências *in vitro* de aminoacilação. Ver, p.ex., Pastrnak, et al., Helv. Chim. Acta, 83:2277 (2000).

Uma abordagem semelhante envolve o uso de uma sintetase heteróloga como a sintetase mutante externa e um tRNA iniciador mutante do mesmo organismo ou de um organismo

relacionado como o tRNA modificado. RajBhandary e colegas descobriram que um mutante âmbar de tRNA^{Met} iniciador humano é acilado por GlnRS de *Escherichia coli* e atua como um supressor âmbar em células de levedura apenas quando EcGlnRS é co-expressa. Ver, Kowal, et al., PNAS USA, 98:2268 (2001). Este par representa portanto um par mutante externo para uso em levedura. Além disso, um mutante âmbar de tRNA^{Met} iniciador de *Escherichia coli* foi encontrado que é inativo em relação a quaisquer sintetases de *Escherichia coli*. Uma TyrRS mutante de levedura foi selecionada que carrega este tRNA mutante, resultando num par mutante externo em *Escherichia coli*.

Usando os métodos divulgados neste documento, os pares e os componentes de pares desejados acima são evoluídos para gerarem tRNA e/ou RS mutante externos que possuem a característica desejada, p.ex., que pode preferencialmente aminoacilar um O-tRNA com um aminoácido antinatural.

Em certas formas de realização, o tRNA modificado e a RS modificada podem ser derivados por mutação de um tRNA e RS naturais de uma variedade de organismos. Numa forma de realização, o tRNA modificado e/ou RS modificada são derivados de pelo menos um organismo, onde o organismo é um organismo procariota, p.ex., *Methanococcus jannaschii*, *Methanobacterium thennoautotrophicum*, *Halobacterium*, *Escherichia coli*, *A. fulgidus*, *P. furiosus*, *P. horikoshii*, *A. pernix*, *T. thermophilus*, ou semelhantes. Opcionalmente, o organismo é um organismo eucariota, p.ex., plantas (p.ex., plantas complexas como monocotiledóneas ou dicotiledóneas), algas, fungos (por exemplo, leveduras, etc), animais (p.ex., mamíferos, insetos, artrópodes, etc.), insetos, protistas, ou semelhantes. Opcionalmente, o

tRNA modificado é derivado por mutação de um tRNA natural de um primeiro organismo e a RS modificada é derivada por mutação de uma RS natural de um segundo organismo. Numa forma de realização, o tRNA modificado e RS modificada podem ser derivados de um tRNA mutado e RS mutada. Em certas formas de realização, a RS modificada e/ou tRNA modificado de um primeiro organismo é fornecida a um sistema de tradução de um segundo organismo, que possui opcionalmente RS e/ou tRNA endógena não-funcional em relação aos códons reconhecidos pelo tRNA modificado ou RS modificada.

O tRNA mutante externo e/ou a tRNA sintetase mutante externa também pode opcionalmente ser isolado a partir de uma variedade de organismos. Numa forma de realização, o tRNA mutante externo e/ou sintetase mutante externa é isolado de pelo menos um organismo, onde o organismo é um organismo procariota, p.ex., *Methanococcus jannaschii*, *Methanobacterium thermoautotrophicum*, *Halobacterium*, *Escherichia coli*, *A. fulgidus*, *P. furiosus*, *P. horikoshii*, *A. pernix*, *T. thermophilus*, ou semelhantes. Opcionalmente, o organismo é um organismo eucariota, p.ex., plantas (p.ex., plantas complexas como monocotiledóneas ou dicotiledóneas), algas, fungos (p.ex., leveduras, etc), animais (p.ex., mamíferos, insetos, artrópodes, etc.), insetos, protistas, ou semelhantes. Opcionalmente, o tRNA externo é isolado de um tRNA natural de um primeiro organismo e a sintetase mutante externa é isolada de uma RS natural de um segundo organismo. Numa forma de realização, o tRNA mutante externo e/ou tRNA sintetase mutante externa pode ser isolado de uma ou mais bibliotecas (que opcionalmente compreende um ou mais tRNA e/ou RS de um ou

mais organismos (incluindo aqueles compreendendo procariotas ou eucariotas)).

Métodos para selecionar um par de tRNA mutante externo e/ou tRNA sintetase para uso em qualquer sistema de tradução são também indicados neste documento. Os métodos incluem: introduzir um gene marcador, um tRNA e/ou um aminoacil-tRNA sintetase (RS) isolado ou derivado de um primeiro organismo num primeiro conjunto de células do segundo organismo; introduzir o gene marcador e o tRNA ou RS num conjunto duplicado de células do segundo organismo; e, selecionar as células sobreviventes no primeiro conjunto que não conseguem sobreviver no conjunto duplicado de células, onde o primeiro conjunto e o conjunto duplicado de células são crescidas na presença de um agente de seleção, e onde as células sobreviventes compõem o tRNA e/ou RS mutante externo para uso num sistema de tradução. Numa forma de realização, comparar e selecionar incluem um ensaio *in vivo* de complementação. Noutra forma de realização, a concentração do agente de seleção é variada. O mesmo ensaio também pode ser realizado num sistema *in vitro* ou *in vivo* baseado no segundo organismo.

Geração de AARS por Mutagénese e Seleção/Triagem

A mutação ou modificação de uma AARS deve ser usada para incorporação de um aminoácido não-natural num polipéptido ou proteína alvo pode ser efetuada usando mutagénese dirigida assim que os aminoácidos de contacto desejados forem identificados. Identificação de aminoácidos de contacto pode ser efetuada usando qualquer método que permite a análise da estrutura da AARS, incluindo análise cristalográfica, modelação de computador, espectroscopia de

ressonância magnética nuclear (NMR), triagem de biblioteca ou uma combinação de qualquer destes ou outros métodos.

Um número de moléculas AARS foi sequenciado e fornece orientação sobre quais aminoácidos são importantes para a ligação do aminoácido com o qual carregar o tRNA correspondente. Ver, por exemplo, SEQ ID Nos 48-103.

Em certas formas de realização, a AARS capaz de carregar um tRNA mutante externo particular com um determinado aminoácido antinatural pode ser obtido por mutagenese da AARS para gerar uma biblioteca de candidatos, seguido de triagem e/ou seleção da AARS candidata capaz da sua função desejada. Tais AARSs mutantes externas e tRNAs mutantes externas podem ser usados para a produção *in vitro* / *in vivo* de proteínas desejadas com aminoácidos antinaturais modificados.

Assim, métodos para gerar componentes da maquinaria biossintética de proteínas, tais como as RSs mutantes externas, tRNAs mutantes externas, e/ou pares tRNA/RS mutantes externos que podem ser usados para incorporar um aminoácido antinatural são fornecidos em certas formas de realização divulgadas neste documento.

Numa forma de realização, métodos para produzir pelo menos uma aminoacil-tRNA sintetase mutante externa recombinante compreendem: (a) gerar uma biblioteca de RSs (opcionalmente mutantes) derivadas de pelo menos uma aminoacil-tRNA sintetase (RS) de um primeiro organismo, p.ex., um organismo eucariótico (tais como uma levedura), ou um organismo procariota, tais como *Methanococcus jannaschii*, *Methanobacterium thennoautotrophicum*, *Halobacterium*,

Escherichia coli, *A. fulgidus*, *P. furiosus*, *P. horikoshii*, *A. pemix*, *T. thermophilus*, ou semelhantes; (b) selecionar (e/ou efetuar triagem) a biblioteca de RSs (opcionalmente RSs mutantes) relativamente aos membros que aminoacilam um tRNA mutante externo na presença de um aminoácido antinatural e um aminoácido natural, deste modo proporcionando um conjunto de RSs (opcionalmente mutantes) ativas; e/ou, (c) selecionar (opcionalmente através de seleção negativa) o conjunto de RSs ativas (p.ex., RSs mutantes) que preferencialmente aminoacilam o O-tRNA na ausência do aminoácido antinatural, deste modo proporcionando pelo menos uma sintetase mutante externa recombinante, em que pelo menos uma sintetase mutante externa recombinante preferencialmente aminoacila o tRNA mutante externo com o aminoácido antinatural. Sintetases mutantes externas recombinantes produzidas pelos métodos também estão incluídos certas formas de realização divulgadas neste documento.

Numa forma de realização, a RS é uma RS inativa, que pode ter sido gerada por mutação de uma RS ativa. Por exemplo, a RS inativa pode ser gerada mutando pelo menos cerca de 1, pelo menos cerca de 2, pelo menos cerca de 3, pelo menos cerca de 4, pelo menos cerca de 5, pelo menos cerca de 6, ou pelo menos cerca de 10 ou mais aminoácidos para diferentes aminoácidos, p.ex., alanina.

Bibliotecas de RSs mutantes podem ser geradas usando várias técnicas de mutagénese conhecidas na arte. Por exemplo, as RSs mutantes podem ser geradas por mutações sítio-específicas, mutações aleatórias, mutações de recombinação geradoras de diversidade, construções quiméricas, e por

outros métodos descritos neste documento ou conhecidos na arte.

Numa forma de realização, selecionar (e/ou efetuar triagem) a biblioteca de RSs (opcionalmente RSs mutantes) relativamente aos membros que estão ativos, p.ex., que aminoacilam um tRNA mutante externo na presença de um aminoácido antinatural e um aminoácido natural, inclui: introduzir uma seleção positiva ou marcador de seleção, p.ex., um gene de resistência a antibióticos, ou semelhantes, e a biblioteca das RSs (opcionalmente mutantes) numa pluralidade de células, onde a seleção positiva e/ou marcador de seleção compreende pelo menos um codão, cuja tradução (opcionalmente condicionalmente) depende da capacidade de uma RS candidata para carregar o tRNA mutante externo (com um aminoácido natural ou antinatural); crescer a pluralidade de células na presença de um agente de seleção; identificar as células que sobrevivem (ou mostram uma resposta específica) na presença do agente de seleção e/ou de triagem traduzindo com sucesso o codão na seleção positiva ou marcador de seleção, proporcionando assim um subconjunto de células positivamente selecionadas que contém o conjunto de RSs (opcionalmente mutantes) ativas. Opcionalmente, a concentração do agente de seleção e/ou triagem pode ser variada. Em certas formas de realização, as células não contêm um tRNA, RS ou par tRNA-RS endógeno funcional que pode ajudar a traduzir o codão. O par tRNA/RS endógeno pode ser desativado por deleção de genes e/ou inibidores de RS.

Uma vez que muitos genes essenciais da célula provavelmente também contêm tal codão que depende da capacidade da sintetase mutante externa para carregar o tRNA modificado

na ausência do par RS/tRNA endógeno funcional, numa forma de realização, marcadores positivos de seleção extra não são necessários para o processo de seleção positiva - a sobrevivência da célula pode ser usada como uma leitura do processo de seleção positiva.

Num aspeto, o marcador de seleção positiva é um gene de cloranfenicol acetiltransferase (CAT). Opcionalmente, o marcador de seleção positiva é um gene de β -lactamase. Noutro aspeto o marcador de seleção positiva compreende um marcador de seleção fluorescente ou luminescente ou um marcador de seleção baseado em afinidade (p.ex., um marcador de superfície celular).

Numa forma de realização semelhante, um sistema sem células *in vitro* pode ser usado para testar a capacidade da sintetase mutante externa para carregar o tRNA modificado numa seleção positiva. Por exemplo, a capacidade do sistema *in vitro* para traduzir um gene de seleção positiva, tal como um gene marcador fluorescente, pode depender da capacidade da sintetase mutante externa para carregar o tRNA modificado para ler através de um códon do gene marcador.

Numa forma de realização, efetuar seleção ou triagem negativa do conjunto relativamente a RSs ativas (opcionalmente mutantes) que preferencialmente aminoacilam o tRNA mutante na ausência do aminoácido antinatural inclui: introduzir um marcador de seleção ou triagem negativa com o conjunto de RSs (opcionalmente mutante) ativas da seleção ou triagem positiva numa pluralidade de sistema de tradução, em que a seleção ou marcador de seleção negativo compreende pelo menos um códon (p.ex.,

codão para um gene marcador tóxico, p.ex., um gene de ribonuclease barnase), cuja tradução depende da capacidade de uma RS candidata para carregar o tRNA mutante externo (com um aminoácido natural); e, identificar o sistema de tradução que mostra uma resposta específica de seleção num primeiro meio suplementado com o aminoácido antinatural e o agente de seleção ou triagem, mas não conseguem mostrar a resposta específica num segundo meio suplementado com o aminoácido natural e o agente de seleção ou triagem, deste modo proporcionando células sobreviventes ou células selecionadas com pelo menos uma RS recombinante.

Num aspeto, a concentração do agente de seleção (e/ou triagem) é variada. Em alguns aspetos o primeiro e segundo organismos são diferentes. Assim, o primeiro ou segundo organismo opcionalmente compreende: um procariota, um eucariota, um mamífero, uma *Escherichia coli*, um fungo, uma levedura, uma arqueiabactéria, uma eubacterium, uma planta, um inseto, um protista, etc. Noutras formas de realização, o marcador de seleção compreende um marcador de seleção fluorescente ou luminescente (tais como a proteína verde fluorescente) ou um marcador de seleção baseado em afinidade.

Também, alguns aspetos incluem em que o marcador de seleção negativa compreende um gene de ribonuclease barnase (que compreende pelo menos um tal codão). Outros aspetos incluem em que o marcador de seleção opcionalmente inclui um marcador de seleção fluorescente ou luminescente ou um marcador de seleção baseado em afinidade. Nas formas de realização neste documento, as seleções e/ou triagens opcionalmente incluem variação da estringência de triagem e/ou seleção.

Num aspeto, o segundo conjunto de RS mutante derivado de pelo menos uma RS recombinante que pode ser gerado por mutagénese, p.ex., mutagénese aleatória, mutagénese sítio-específica, recombinação ou uma combinação dos mesmos.

Os métodos incorporados neste documento opcionalmente compreendem em que o aminoácido antinatural é selecionado a partir de, por exemplo: um O-metil-L-tirosina, uma L-3-(2-naftil)alanina, uma 3-metil-fenilalanina, uma O-4-alil-L-tirosina, uma 4-propil-L-tirosina, uma tri-O-acetil-GlcNAc β -serina, uma L-Dopa, uma fenilalanina fluorada, uma isopropil-L-fenilalanina, uma p-azido-L-fenilalanina, uma p-acil-L-fenilalanina, uma p-benzoil-L-fenilalanina, uma L-fosfoserina, uma fosfonoserina, uma fosfonotirosina, uma p-iodo-fenilalanina, uma p-bromofenilalanina, uma p-amino-L-fenilalanina e uma isopropil-L-fenilalanina. Uma RS recombinante produzida pelos métodos deste documento também está incluída nas formas de realização divulgadas neste documento.

Num aspeto relacionado, métodos para produzir um tRNA mutante externo recombinante incluem: (a) gerar uma biblioteca de tRNAs mutantes derivados de pelo menos um tRNA, de um primeiro organismo; (b) selecionar (p.ex., selecionar negativamente) ou efetuar uma triagem da biblioteca relativamente a tRNAs (opcionalmente mutantes) que são aminoacilados por uma aminoacil-tRNA sintetase (RS) de um segundo organismo na ausência de uma RS do primeiro organismo, proporcionando assim um conjunto de tRNAs (opcionalmente mutantes); e, (c) selecionar ou efetuar a triagem do conjunto de tRNAs (opcionalmente mutantes) de membros que são aminoacilados por uma RS mutante externa

introduzida, proporcionando assim pelo menos um tRNA recombinante; em que o pelo menos um tRNA recombinante reconhece um codão de aminoácido não-natural e não é eficientemente reconhecido pela RS do segundo organismo e é preferencialmente aminoacilado pela RS mutante externa.

Os vários métodos descritos neste documento opcionalmente compreendem em que selecionar ou efetuar triagem compreende uma ou mais seleções ou triagens positivas ou negativas, p.ex., uma mudança na permeabilidade de aminoácido, uma mudança na eficiência de tradução, e uma mudança na fidelidade de tradução. Além disso, a uma ou mais alterações é opcionalmente baseada numa mutação em um ou mais genes num organismo em que um par de tRNA mutante externo-tRNA sintetase é usado para produzir tal proteína. Selecionar e/ou efetuar triagem neste documento opcionalmente compreende em que pelo menos 2 codões em um ou mais genes de seleção ou em um ou mais genes de triagem são usados. Tais múltiplos codões são opcionalmente no mesmo gene ou em genes de triagem/seleção diferentes. Além disso, os múltiplos codões opcionais são opcionalmente codões diferentes ou compreendem o mesmo tipo de codões.

Kits são uma característica adicional de certas formas de realização descritas neste documento. Por exemplo, os *kits* podem incluir um ou mais sistemas de tradução conforme mencionado acima (p.ex., uma célula), um ou mais tRNAs (incluindo tRNAs modificados ou mutados), uma ou mais AARSs (incluindo AARSs mutadas ou modificadas), um ou mais aminoácidos antinaturais, p.ex., com material de embalagem adequado, recipientes para conter os componentes do *kit*, materiais de instrução para praticar os métodos neste documento e/ou semelhantes. Se uma ou mais AARSs e/ou um ou

mais tRNAs são fornecidos num *kit*, podem ser fornecidos como ácidos nucleicos, ou proteínas, e podem ser parte de um único vetor ou contidos em vetores separados. Da mesma forma, produtos dos sistemas de tradução (p.ex., proteínas como análogos EPO compreendendo aminoácidos antinaturais) podem ser fornecidos em forma de *kit*, p.ex., com recipientes para conter os componentes do *kit*, materiais de instrução para praticar os métodos neste documento e/ou semelhantes.

Usos Exemplares

Bem mais de 100 aminoácidos não codificados (todos aceitáveis pelos ribossomas) foram, de acordo com relatos, introduzidos em proteínas usando outros métodos (ver, por exemplo, Schultz et al., J. Am. Chem. Soc., 103: 1563-1567, 1981; Hinsberg et al., J. Am. Chem. Soc., 104: 766-773, 1982; Pollack et al., Science, 242: 1038-1040, 1988; Nowak et al., Science, 268: 439-442, 1995) todos estes análogos podem ser usados nos métodos em questão para incorporação eficiente destes análogos em produtos proteicos.

Noutra forma de realização preferida, dois ou mais análogos podem ser usados no mesmo sistema de tradução *in vitro* ou *in vivo*, cada um com seu tRNA mutante externo ou pares de sintetase mutante externa. Isto é mais facilmente alcançado quando um aminoácido natural é codificado por quatro ou mais códons (tais como seis para Leu e Arg). No entanto, para aminoácidos codificados por apenas dois códons, um pode ser reservado para o aminoácido natural, enquanto o outro "partilhado" por um ou mais análogos de aminoácidos. Estes análogos podem assemelhar-se apenas a um aminoácido natural (por exemplo, diferentes análogos de Phe), ou

assemelham-se a diferentes aminoácidos (por exemplo, análogos de Phe e Tyr).

Em certas formas de realização, um primeiro ácido nucleico codificando uma molécula de tRNA mutante externa / modificada que não é carregada eficientemente por uma aminoacil-tRNA sintetase endógena na célula 1 no sistema de tradução *in vitro* (IVT), ou o próprio tRNA mutante externo / modificado. De acordo com algumas formas de realização, um segundo ácido nucleico codificando uma aminoacil-tRNA sintetase (AARS) mutante/modificada externa é também introduzido na célula/IVT. A AARS mutante/modificada externo é capaz de carregar o tRNA mutante/modificado externo com um análogo de aminoácido escolhido. O análogo de aminoácido pode depois ser fornecido à célula para que possa ser incorporado em uma ou mais proteínas dentro da célula ou IVT.

Noutras formas de realização, o ambiente é uma célula. Uma variedade de células (ou lisados das mesmas adequados para IVT) pode ser usada em certos métodos, incluindo, por exemplo, uma célula bacteriana, uma célula fúngica, uma célula de inseto, e uma célula de mamífero (p.ex., uma célula humana ou uma célula de mamífero não-humano). Numa forma de realização, a célula é uma célula de *E. coli*, noutra forma de realização, a célula é uma célula de *Pseudomonas*.

Em certas formas de realização, o análogo do aminoácido pode ser fornecido contactando diretamente a célula ou IVT com o análogo, por exemplo, aplicando uma solução do análogo à célula na cultura, ou adicionando diretamente o análogo ao IVT. O análogo também pode ser fornecido através

da introdução de uma ou mais construções de ácido nucleico adicionais na célula / IVT, em que a construção de ácido nucleico adicional codifica uma ou mais proteínas de síntese de análogo de aminoácido que são necessárias para a síntese do análogo desejado.

Certas formas de realização envolvem adicionalmente introduzir uma construção de ácido nucleico molde na célula / IVT, o molde codificando uma proteína, em que a construção de ácido nucleico contém pelo menos uma sequência de códons degenerados. Os ácidos nucleicos introduzidos na célula / IVT podem ser introduzidos como uma construção ou uma pluralidade de construções. Em certas formas de realização, os vários ácidos nucleicos são incluídos na mesma construção. Por exemplo, os ácidos nucleicos podem ser introduzidos em quaisquer vetores adequados capazes de expressar o tRNA e/ou proteínas codificados na célula / IVT. Numa forma de realização, a primeira e segunda sequências de ácidos nucleicos são fornecidas em um ou mais plasmídeos. Noutra forma de realização, o vetor ou vetores usados são vetores virais, incluindo, por exemplo, vetores de adenovírus e lentivirais. As sequências podem ser introduzidas com uma sequência de promotor apropriado para a célula / IVT, ou múltiplas sequências que podem ser indutíveis para controlar a expressão das sequências.

Para uso *in vitro*, uma ou mais sintetases mutantes externas podem ser produzidas de forma recombinante e fornecidas a qualquer dos sistemas de tradução *in vitro* disponíveis (tais como o comercialmente disponível PROTEINscript-PRO™ baseado em lisado de germe de trigo, sistema de *E. coli* Ambion®'s para transcrição/tradução acopladas *in vitro*; ou

o Retic Lysate IVT™ Kit da Ambion®) baseado em lisado de reticulócitos de coelho. Opcionalmente, o sistema de tradução *in vitro* pode ser seletivamente empobrecido de uma ou mais AARSs naturais (por, por exemplo, imuno-depleção usando anticorpos imobilizados contra AARS natural) e/ou aminoácidos naturais de modo a que incorporação melhorada do análogo pode ser obtida. Alternativamente, ácidos nucleicos codificando as sintetases mutantes externas redesenhadas podem ser fornecidas no lugar de AARSs produzidas de forma recombinante. O sistema de tradução *in vitro* também é fornecido com os análogos para serem incorporados em produtos de proteína madura.

Embora a síntese de proteínas *in vitro* geralmente não possa ser efetuada na mesma escala como na síntese *in vivo*, métodos *in vitro* podem render centenas de microgramas de proteína purificada contendo análogos de aminoácidos. Tais proteínas foram produzidas em quantidades suficientes para a sua caracterização utilizando dicróismo circular (CD), espectrometria de ressonância magnética nuclear (NMR) e cristalografia de raios-X. Esta metodologia também pode ser usada para investigar o papel da hidrofobicidade, empacotamento, entropia de cadeia lateral e ligações por pontes de hidrogénio em determinar a estabilidade da proteína e enrolamento. Também pode ser usada para sondar o mecanismo catalítico, transdução de sinal e transferência de eletrões em proteínas. Além disso, as propriedades das proteínas podem ser modificadas usando esta metodologia. Por exemplo, proteínas *photocaged* podem ser geradas que podem ser ativadas por fotólise, e novos grupos químicos foram introduzidos em proteínas para a incorporação de local específico de sondas óticas e outras sondas espectroscópicas.

O desenvolvimento de uma abordagem geral para a incorporação de análogos de aminoácidos em proteínas *in vivo*, diretamente a partir do meio de crescimento, iria aumentar consideravelmente o poder de mutagénese de aminoácido antinatural. Por exemplo, a capacidade de sintetizar grandes quantidades de proteínas contendo átomos pesados facilitaria a determinação da estrutura da proteína, e a capacidade de substituir sítio-especificamente fluoróforos ou grupos foto-cliváveis em proteínas em células vivas iria fornecer ferramentas poderosas para estudar a função das proteínas *in vivo*. Alternativamente, um pode ser capaz de melhorar as propriedades de proteínas fornecendo blocos de construção com novos grupos funcionais, tais como um aminoácido contendo ceto.

Para uso *in vivo*, uma ou mais AARS podem ser fornecidas a uma célula hospedeira (procariótica ou eucariótica) como materiais genéticos, tais como sequências codificadoras em plasmídeos ou vetores virais, que podem opcionalmente integrar o genoma do hospedeiro e expressar constitutivamente ou de forma indutível as AARSS re-desenhadas. Uma proteína heteróloga ou endógena de interesse pode ser expressa numa tal célula hospedeira, com a presença de análogos de aminoácido fornecidos. Os produtos proteicos podem então ser purificados usando quaisquer técnicas de purificação de proteínas reconhecidas na arte, ou técnicas especialmente concebidas para a proteína de interesse.

Estes são alguns meios possíveis para gerar um transcrito que codifica um polipéptido. Em geral, quaisquer meios

conhecidos na arte para gerar transcritos podem ser empregues para sintetizar proteínas com análogos de aminoácidos. Por exemplo, qualquer sistema de transcrição *in vitro* ou sistemas de transcrição/tradução acopladas podem ser usados para gerar um transcrito de interesse, que então serve como um molde para a síntese proteica. Alternativamente, qualquer célula, célula / linha celular gerada por engenharia, ou componentes funcionais (lisados, frações de membrana, etc.) que é capaz de expressar proteínas a partir de materiais genéticos pode ser usada para gerar um transcrito. Estes meios para gerar um transcrito normalmente irão incluir tais componentes como RNA polimerase (T7, SP6, etc) e co-fatores, nucleótidos (ATP, CTP, GTP, UTP), fatores de transcrição necessários, e as condições de tamponização apropriadas, assim como pelo menos um molde de DNA adequado, mas outros componentes podem também ser adicionados para condição otimizada de reação. Um artesão hábil prontamente vislumbraria outras formas de realização semelhantes às descritas neste documento.

Grupos Químicos

Em certas formas de realização, o(s) aminoácido(s) natural(is) e/ou a molécula terapêutica compreende um grupo quimicamente reativo. O grupo pode ser fortemente eletrofílico ou nucleofílico e assim estar disponível para reagir diretamente com a molécula terapêutica ou o anticorpo ou fragmento do mesmo. Alternativamente, o grupo pode ser um eletrófilo ou nucleófilo mais fraco e portanto requerer ativação antes da conjugação com a molécula terapêutica ou o anticorpo ou fragmento do mesmo. Esta alternativa seria desejável onde é necessário atrasar a

ativação do grupo quimicamente reativo até um agente ser adicionado à molécula a fim de evitar a reação do agente com o grupo. Em qualquer cenário, o grupo é quimicamente reativo, os cenários diferem (no cenário de reagir com o anticorpo) na medida em que após a adição de um agente, o grupo é reagido diretamente com um anticorpo ou fragmento do mesmo ou é reagido primeiro com um ou mais produtos químicos para tornar o grupo capaz de reagir com um anticorpo ou fragmento do mesmo. Em certas formas de realização, o grupo quimicamente reativo inclui um grupo amina, um grupo sulfidril, um grupo hidroxilo, um grupo contendo carbonilo, ou um grupo alquilo solto.

Certas formas de realização podem utilizar química *click*, que inclui, mas não está limitada a, cicloadição 1, 3-dipolar de Huisgen, em particular a variante catalisada passo a passo por Cu(I), reação de Diels-Alder, substituição nucleofílica de especialmente pequenos anéis tensos como compostos de epóxi e aziridina, formação de química de tipo carbonilo de ureias e amidas, reações de adição a ligações duplas carbono-carbono como epoxidação e dihidroxilação.

Portanto, em adição a ou em vez de glicosilação de polipéptidos das formas de realização divulgadas neste documento, outros grupos químicos (incluindo poli(etileno) glicol) podem ser adicionados, ligados, unidos, ou de outra forma conjugados ou incorporados nos polipéptidos modificados. Adição de PEG é um processo para anexar covalentemente oligossacáridos e polímeros sintéticos tais como polietilenoglicol (PEG) sítio-especificamente a moléculas proteicas terapêuticas. Adição de PEG pode melhorar significativamente a meia-vida da proteína por

proteção do polipéptido em relação a enzimas proteolíticas e aumentando o tamanho aparente da proteína, reduzindo assim as taxas de remoção. Além disso, conjugados de PEG podem melhorar a solubilidade da proteína e ter efeitos benéficos na biodistribuição. As propriedades físicas e farmacológicas de proteínas com PEG são afetadas pelo número e o tamanho das cadeias PEG anexadas ao polipéptido, a localização dos locais PEG, e a química usada para adição de PEG.

Exemplos de conjugação de PEG com proteínas incluem reações de PEGs derivatizados de N-hidroxisuccinimidil éster com lisina, reações de adição-1,4 de PEGs derivatizados de maleimida e vinilsulfona com cisteína, e condensação de hidrazida contendo PEGs com aldeídos gerados pela oxidação de glicoproteínas. Quando mais de um sítio reativo está presente numa proteína (p.ex., múltiplos grupos amina ou tiol) ou eletrófilos reativos são usados, acoplamento não-seletivo de uma ou múltiplas moléculas de PEG pode ocorrer, levando à geração de uma mistura heterogénea que é difícil separar. A falta de seletividade e controlo posicional no acoplamento de cadeias de PEG pode levar a perdas significativas na atividade biológica e possivelmente maior imunogenicidade das proteínas conjugadas. Na verdade, perda de atividade biológica e heterogeneidade de produto têm sido os dois problemas mais comuns encontrados no desenvolvimento de fármacos proteicos de ação prolongada usando técnicas padrão de adição de PEG. Modificação de proteínas com PEGs reativos com amina geralmente resulta em drástica perda de atividade biológica devido à modificação de resíduos de lisina localizados em regiões da proteína importantes para a atividade biológica. Em certas situações, a bioatividade de hormonas de crescimento pode

ser reduzida 400 vezes ou mais. Por exemplo, a bioatividade de GCSF é reduzida 1000 vezes quando as proteínas são modificadas usando tecnologias convencionais de adição de PEG a amina (Clark et al., J. Biol. Chem. 271: 21969, 1996; Bowen et al., Exp. Hematol. 27, 425, 1999). Existe portanto uma necessidade para um método que permite o acoplamento completamente sítio-específico e irreversível de cadeias de PEG a proteínas.

Seria vantajoso usar tecnologias avançada de engenharia de proteínas para criar ação prolongada, fármacos para proteínas humanas "compatíveis com pacientes", por, por exemplo, incorporar aminoácidos antinaturais numa proteína para medicamentos, de tal modo que a proteína de engenharia para medicamentos pode obter meia-vida mais longa e/ou atividade biológica equivalente ou mesmo melhorada. Para este propósito, algumas formas de realização divulgadas neste documento podem ser usadas para superar problemas como heterogeneidade e perda de atividade inerente em técnicas padrão de adição de PEG a aminas. Incorporar aminoácidos antinaturais irá fornecer sítios exclusivos pré-determinados longe do sítio de ligação ou do sítio catalítico na proteína alvo onde moléculas PEG podem ser conjugadas sítio-especificamente. Além disso, moléculas de PEG podem ser acopladas a aminoácidos antinaturais através de técnicas diferentes de adição de PEG a aminas, assim poupando os grupos de aminas primárias de lisinas de adição de PEG indesejada. Estas técnicas podem ser usadas para melhorar a meia-vida, eficácia, e/ou segurança de bio-fármacos em todas as áreas, incluindo o campo específico do cancro, endocrinologia, doenças infecciosas, e inflamação, etc.

Como um exemplo ilustrativo, química Click ou cicloadição podem ser usadas para formar uma ligação de triazolo. Um exemplo particular de cicloadição é uma cicloadição [3+2] de Huisgen mediada por cobre (Tornoe et al., J. Org. Chem. 67: 3057, 2002; Rostovtsev et al., Angew. Chem., Int. Ed. 41: 596, 2002; and Wang et al., J. Am. Chem. Soc. 125: 3192, 2003) de uma azida e um alcino é mutante externo a todos os grupos funcionais encontrados em proteínas, e forma uma ligação estável de triazolo, esta reação pode ser usada para a adição seletiva de PEG a proteínas. Por exemplo, Deiters et al. (Bioorg. Med. Chem. Lett. 14(23): 5743-5745, 2004) relataram uma metodologia de adição de PEG geralmente aplicável baseada na incorporação sítio-específica da para-azidofenilalanina em proteínas em levedura. O grupo azido foi usado numa reação de cicloadição suave [3 + 2] com um reagente de PEG derivatizado de alcino para obter seletivamente proteína com PEG. Esta estratégia deve ser útil para a geração de proteínas seletivamente adicionadas de PEG para aplicações terapêuticas.

Em certas formas de realização, o polipéptido é uma proteína terapêutica, de diagnóstico, ou outra proteína selecionada a partir de: antitripsina Alfa-1, Angiostatina, Fator antihemolítico, anticorpos (incluindo um anticorpo ou um fragmento funcional ou derivado do mesmo selecionado a partir de: Fab, Fab', F(ab)2, Fd, Fv, ScFv, diabody, tribody, tetrabody, dímero, trímero ou minibody), moléculas angiogénicas, moléculas angiostáticas, Apolipoproteína, Apoproteína, Fator Natriurético Atrial, polipéptido Natriurético Atrial, Asparaginase, Adenosina deaminase, hirudina, fator Neurotrófico Ciliar, fator morfogénico do osso (todos e quaisquer BMPs), péptidos Atrial, quimiocinas

C-X-C (p.ex., T39765, NAP-2, ENA-78, Gro-a, Gro-b, Groc, IP-10, GCP-2, NAP-4, SDF-1, PF4, MIG), Calcitonina, quimiocinas CC (p.ex., proteína-1 de atração quimiotática de Monócito, proteína-2 de atração quimiotática de Monócito, proteína-3 de atração quimiotática de Monócito, proteína-1 alfa inflamatória de Monócito, proteína-1 beta inflamatória de Monócito, RANTES, I309, R83915, R91733, HCC1, T58847, D31065, T64262), ligando CD40, Calcitonina, ligando C-kit, colagénico, factor estimulante de colónia (CSF), péptido natriurético tipo C (CNP), fator 5a de Complemento, inibidor de Complemento, recetor 1 de Complemento, citocinas, (p.ex., Neutrophil Activating Peptide-78 epitelial, GRO α /MGSA, GRO β , GRO γ , MIP-1 α , MIP-1 δ , MCP-1), ácidos desoxirribonucleicos, Fator de Crescimento Epidérmico (EGF), Eritropoietina, toxinas A e B Esfoliantes, Fator IX, Fator VII, Fator VIII, Fator X, Fator de Crescimento Fibroblástico (FGF), fibrinogénio, fibronectina, fator estimulante de colónia de granulócitos (G-CSF), fator estimulante de colónia de macrófagos granulócitos (GM-CSF), folitropina, Glucocerebrosidase, Gonadotropina, glucagões, GLP-1, fatores de crescimento, proteínas Hedgehog (p.ex., Sonic, Indian, Desert), Hormona de Crescimento Humano, Hemoglobina, Fator de Crescimento de Hepatócitos (HGF), vírus da Hepatite, hirudina, albumina de soro Humano, Insulina, Fator de Crescimento semelhante à insulina (IGF), interferões (p.ex., IFN- α , IFN- β , IFN- γ , IFN- ϵ , IFN- ζ , η -IFN, IFN- κ , IFN- λ , IFN- τ , IFN -

- δ ,

, IFN-w), interleucinas (p.ex., ., IL-1, IL-2, IL-3, IL-4, IL-5, IL-6, IL-7, IL-8, IL-9, IL-10, IL-11, IL-12, IL-13, IL-14, IL-15, etc.), Fator de Crescimento de Queratinócitos (KGF), Lactoferrina, fator inibitório de leucemia, Luciferase, hormona Luteinizante, Neurturina, fator

inibidor de Neutrófilos (NIF), oncostatina M, proteína Osteogénica, hormona paratiroide, PD-ECSF, PDGF, hormonas peptídicas (p.ex., Fator de Crescimento Humano), Pleiotropina, Proteína A, Proteína G, fenilalanina hidroxilase, hormona paratiroideia (PTH), Prolactina, exotoxinas Pirogénicas A, B e C, Relaxina, Renina, ácidos ribonucleicos, SCF, recetor I de complemento Solúvel, I-CAM 1 Solúvel, recetores de interleucina solúvel (IL-1, IL-2, IL-3, IL-4, IL-5, IL-6, IL-7, IL-9, IL-10, IL-11, IL-12, IL-13, IL-14, 1IL-5), Recetor de TNF solúvel, Somatomedina, Somatostatina, Somatotropina, Estreptoquinase, Superantigénios, ou seja, enterotoxinas estafilocócicas (SEA, SEB, SEC1, SEC2, SEC3, SED, SEE), superóxido dismutase (SOD), toxina da síndrome do choque tóxico (TSST-1), Timosina Alfa 1, ativador tecidual do plasminogénio, fator de necrose tumoral beta (TNF-beta), recetor de fator de necrose tumoral (TNFR), fator de necrose tumoral alfa (TNF-alfa), ligando indutor de apoptose relacionado com necrose tumoral (TRIAL), Fator de Crescimento Endotelial Vascular (VEGEF), uroquinase; um modulador transcricional que modula o crescimento celular, diferenciação ou regulamento, em que o modulador transcricional é de procariotas, vírus, ou eucariotas, incluindo plantas, fungos, leveduras, insetos e animais, incluindo mamíferos; ativador de expressão selecionado de citocinas, moléculas inflamatórias, fatores de crescimento, seus recetores, produtos do oncogene, interleucinas (p.ex., IL-1, IL-2, IL-8, etc.), interferões, 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 de oncogene correspondentes, p.ex., Mos, Ras, Raf e Met; ativadores e supressores transcricionais, p.ex., p53, Tat, Fos, Myc, Jun, Myb, Rel;

recetores de hormonas esteróides, seleccionados a partir de recetores para estrogénio, progesterona, testosterona, aldosterona, LDL, ou corticosterona; ou uma enzima seleccionada a partir de: amidases, aminoácido racemases, acilases, desalogenasa, dioxigenases, diarilpropano peroxidases, epimerases, epóxido hidrolases, esterases, isomerases, cinases, isomerases de glicose, glicosidases, glicosil transferases, haloperoxidases, monooxigenases (p.ex., p450s), lipases, lenhina peroxidases, nitrilo hidratases, nitrilases, proteases, fosfatases, subtilisinas, transaminase, ou nucleases.

No caso em que a proteína ou molécula de interesse a ser modificada é um anticorpo ou fragmento de anticorpo, o(s) resíduo(s) de aminoácidos não-naturais pode(m) ser colocado(s) em qualquer localização ou posição na estrutura do anticorpo, dependendo do objetivo desejado. Por exemplo, o resíduo de aminoácido não-natural pode ser colocado na região variável Fab, a região Fc, ou noutro local que interage com a região Fc do anticorpo. Noutras formas de realização, o resíduo de aminoácido não-natural pode ser colocado na interface de ligação do anticorpo, ou na região V_H. Em certas formas de realização, o anticorpo modificado exhibe um aumento ou diminuição da sua capacidade de eliminar um ou mais alvos. Em particular, um anticorpo com maior capacidade de eliminar um ou mais alvos, ou com efeitos secundários reduzidos pode ser desejado.

Noutras formas de realização, o(s) aminoácido(s) não-natural(is) confere(m) afinidade de ligação melhorada a um recetor de Fc e/ou a C1q do sistema complemento. Em particular, um anticorpo modificado pode possuir uma afinidade e/ou especificidade alterada (p.ex., melhorada)

para um antígeno ou um parceiro de ligação de proteínas (p.ex., Clq do complemento e/ou o recetor de Fc em macrófagos, etc.). Por exemplo, a modificação de uma molécula pode aumentar ou diminuir a sua função de citotoxicidade mediada por células dependentes de anticorpo (ADCC), ou atividade de fixação do complemento. Noutros exemplos, modificação de uma molécula particular pode aumentar ou diminuir a sua capacidade de se ligar a outra molécula de estrutura contrária natural (tal como um anticorpo).

Glicosilação através de Aminoácidos Antinaturais

A modificação pós-traducional de proteínas por glicosilação pode afetar o enrolamento e estabilidade de proteínas, modificar a atividade intrínseca das proteínas, e modular as suas interações com outras biomoléculas. Ver, p.ex., Varki, *Glycobiology* 3: 97-130, 1993. Glicoproteínas naturais estão frequentemente presentes como uma população de muitas glicoformas diferentes, o que torna difícil a análise da estrutura de glicano e o estudo dos efeitos de glicosilação na estrutura e função da proteína. Portanto, os métodos para a síntese de proteínas homogeneamente glicosiladas naturais e antinaturais são necessários para a compreensão sistemática da função de glicanos, e para o desenvolvimento da terapêutica de glicoproteínas melhorada.

Uma abordagem previamente conhecida para produzir proteínas possuindo padrões de glicosilação desejados faz uso de glicosidases para converter uma glicoproteína natural heterogénea num simples núcleo homogéneo, no qual sacáridos podem então ser enxertados sequencialmente com glicosil transferases. Ver, p.ex., Witte et al., *J. Am. Chem. Soc.*

119: 2114-2118, 1997. Uma limitação desta abordagem é que os sítios de glicosilação primária são pré-determinados pela linha celular em que a proteína é expressa. Alternativamente, um glicopéptido contendo a estrutura de glicano desejada pode ser sintetizado por síntese do péptido de fase sólida. Este glicopéptido pode ser acoplado com outros fragmentos de péptidos ou de proteínas recombinantes para se obter uma glicoproteína maior por ligação química normal (ver, p.ex., Shin et al., J. Am. Chem. Soc. 121: 11684-11689, 1999), ligação de proteína expressa (ver, p.ex., Tolbert and Wong, J. Am. Chem. Soc. 122: 5421-5428, 2000), ou com proteases modificadas por engenharia (ver, p.ex., Witte et al., J. Am. Chem. Soc. 120: 1979-1989, 1998). Tanto a ligação química normal como a ligação de proteína expressa são mais eficientes com proteínas pequenas, e necessitam um resíduo de cisteína no N-terminal do glicopéptido.

Quando uma protease é usada para unir péptidos, o sítio da ligação deve ser colocado longe do local da glicosilação para bons rendimentos de acoplamento. Ver, p.ex., Witte et al., J. Am. Chem. Soc. 120: 1979-1989, 1998. Uma terceira abordagem é modificar as proteínas com sacáridos diretamente usando métodos químicos. Boa seletividade pode ser conseguida com derivados de sacárido de haloacetamida, que são acoplados ao grupo tiol da cisteína (ver, p.ex., Davis and Flitsch, Tetrahedron Lett. 32: 6793-6796, 1991; e Macmillan et al., Org. Lett. 4: 1467-1470, 2002). Mas este método pode tornar-se problemático com proteínas que possuem mais de um resíduo de cisteína.

Certas formas de realização fornecidas neste documento divulgam métodos para síntese de glicoproteínas. Estes

métodos envolvem, em algumas formas de realização, incorporar numa proteína um aminoácido antinatural que compreende um primeiro grupo reativo; e contactar a proteína com um grupo sacárido que compreende um segundo grupo reativo, em que o primeiro grupo reativo reage com o segundo grupo reativo, formando assim uma ligação covalente que anexa o grupo sacárido ao aminoácido antinatural da proteína. Glicoproteínas produzidas por estes métodos também estão incluídas em determinadas formas de realização.

O primeiro grupo reativo é, em algumas formas de realização, um grupo eletrofílico (p.ex., um grupo ceto, um grupo aldeído, e/ou semelhantes), e o segundo grupo reativo é um grupo nucleofílico. Em algumas formas de realização, o primeiro grupo reativo é um grupo nucleofílico e o segundo grupo reativo é um grupo eletrofílico (p.ex., um grupo ceto, um grupo aldeído e/ou semelhantes). Por exemplo, um grupo eletrofílico é anexado ao grupo sacárido e o grupo nucleofílico é anexado ao aminoácido antinatural. O grupo sacárido pode incluir um único grupo de carboidrato, ou o grupo sacárido pode incluir dois ou mais grupos de carboidrato.

Em algumas formas de realização, os métodos envolvem adicionalmente contactar o grupo sacárido com uma glicosil transferase, um grupo dador de açúcar, e outros reagentes necessários para a atividade de glicosil transferase por um tempo suficiente e em condições adequadas para transferir um açúcar do grupo dador de açúcar para o grupo sacárido. O produto desta reação pode, se desejado, ser contactado por pelo menos uma segunda glicosil transferase, juntamente com o grupo dador de açúcar adequado.

Em algumas formas de realização, o método compreende adicionalmente contactar o grupo sacárido com uma ou mais β 1-4N-acetilglucosaminil transferase, uma α 1,3-fucosil transferase, uma α 1,2-fucosil transferase, uma α 1,4-fucosil transferase, uma β 1-4-galactosil transferase, uma sialil transferase, e/ou semelhantes, para formar uma estrutura de oligossacárido bi-ramificada ou tri-ramificada. Numa forma de realização, o grupo sacárido compreende uma GlcNAc terminal, o grupo dador de açúcar é UDP-Gal e a glicosil transferase é uma β 1-4-galactosil transferase.

Numa forma de realização, o grupo sacárido compreende uma GlcNAc terminal, o grupo dador de açúcar é UDP-GlcNAc e a glicosil transferase é uma β 1-4N-acetilglucosaminil transferase.

Opcionalmente, alguns métodos compreendem adicionalmente contactar o produto da reação da N-acetilglucosaminil transferase com uma β 1-4-manosil transferase e GDP-manose para formar um grupo sacárido que compreende $\text{Man}\beta$ 1-4GlcNAc β 1-4GlcNAc-. Opcionalmente, o método compreende adicionalmente contactar o grupo $\text{Man}\beta$ 1-4GlcNAc β 1-4GlcNAc com uma α 1-3-manosil transferase e GDP-manose para formar um grupo sacárido que compreende $\text{Man}\alpha$ 1-3 $\text{Man}\beta$ 1-4GlcNAc β 1-4GlcNAc-. Opcionalmente, o método compreende adicionalmente contactar o grupo $\text{Man}\alpha$ 1-3 $\text{Man}\beta$ 1-4GlcNAc β 1-4GlcNAc- com uma α 1-6 manosil transferase e GDP-manose para formar um grupo sacárido que compreende $\text{Man}\alpha$ 1-6($\text{Man}\alpha$ 1-3) $\text{Man}\beta$ 1-4GlcNAc β 1-4GlcNAc-. Opcionalmente, o método compreende adicionalmente contactar o grupo $\text{Man}\alpha$ 1-6($\text{Man}\alpha$ 1-3) $\text{Man}\beta$ 1-4GlcNAc β 1-4GlcNAc- com uma β 1-2N-acetilglucosaminil transferase e UDP-GlcNAc para formar um grupo sacárido que compreende $\text{Man}\alpha$ 1-

6(GlcNAc β 1-2Man α 1-3)Man β 1-4GlcNAc β 1-4GlcNAc-.

Opcionalmente, o método compreende adicionalmente contactar o grupo Man α 1-6(GlcNAc β 1-2Man α 1-3)Man β 1-4GlcNAc β 1-4GlcNAc- com uma β 1-2N-acetilglucosaminil transferase e UDP-GlcNAc para formar um grupo sacárido que compreende GlcNAc β 1-2Man α 1-6(GlcNAc β 1-2Man α 1-3)Man β 1-4GlcNAc β 1-4GlcNAc-.

A etapa de incorporar numa proteína um aminoácido antinatural que compreende um primeiro grupo reativo, em algumas formas de realização, compreende usar um tRNA mutante externo, uma RS mutante externa ou um par tRNA/RS mutante externo. Em tais casos, o tRNA mutante externo preferencialmente reconhece um codão degenerado para o tRNA de tipo selvagem, e incorpora o aminoácido antinatural na proteína em resposta ao codão degenerado, e em que a sintetase mutante externa preferencialmente aminoacila o tRNA mutante externo com o aminoácido antinatural. Em algumas formas de realização, o aminoácido antinatural é incorporado no polipéptido *in vivo*.

Uma grande variedade de grupos reativos adequados são conhecidos por peritos na arte. Tais grupos reativos adequados podem incluir, por exemplo, amino, hidroxilo, carboxilo, carboxilato, carbonilo, alceno, alcino, aldeído, éster, éter (p.ex., tio-éter), amida, amina, nitrilo, vinilo, sulfeto, sulfonil, fosforil, ou grupos quimicamente reativos de forma semelhante. Grupos reativos apropriados adicionais incluem, mas não são limitado a, maleimida, N-hidroxisuccinimida, sulfo-N-hidroxisuccinimida, ácido nitrilotriacético, hidroxilo ativado, haloacetil (p.ex., bromoacetil, iodoacetil), carboxilo ativado, hidrazida, epóxi, aziridina, sulfonilcloreto, trifluorometildiaziridina, piridildisulfeto, N-acil-

imidazole, imidazolecarbamato, vinilsulfona, succinimidilcarbonato, arilazida, anidrido, diazoacetato, benzofenona, isotiocianato, isocianato, imidoéster, fluorobenzeno, biotina e avidina.

Em algumas formas de realização, um dos grupos reativos é um grupo eletrofílico, e o segundo grupo reativo é um grupo nucleofílico. O grupo nucleofílico ou o grupo eletrofílico pode ser anexado à cadeia lateral do aminoácido antinatural; o grupo correspondente é então anexado para ao grupo sacárido.

Grupos eletrofílicos adequados que reagem com grupos nucleofílicos para formar uma ligação covalente são conhecidos pelos peritos na arte. Em certas formas de realização, tais grupos eletrofílicos incluem, mas não estão limitados a, p.ex., grupo carbonilo, um grupo sulfonil, um grupo aldeído, um grupo cetona, um grupo éster bloqueado, um grupo tioéster, um grupo imina estável, um grupo epóxido, um grupo aziridina, etc.

Grupos nucleofílicos adequados que podem reagir com grupos eletrofílicos são conhecidos por peritos na arte. Em certas formas de realização, tais nucleófilos incluem, por exemplo, aminas alifáticas ou aromáticas, tais como etilenodiamina. Em certas formas de realização, os grupos nucleofílicos incluem, mas não estão limitados a, p.ex., -NR₁-NH₂ (hidrazida), -NR₁(C=O)NR₂NH₂ (semicarbazida), -NR₁(C=S)NR₂NH₂ (tiosemicarbazida), -(C=O)NR₁NH₂ (carbonil-hidrazida), -(C=S) NR₁NH₂ (tiocarbonil-hidrazida), -(SO₂)NR₁NH₂ (sulfonil-hidrazida), -NR₁ NR₂(C=O)NR₃NH₂ (carbazida), NR₁NR₂(C=S)NR₃NH₂ (tiocarbazida), -O-NH₂ (hidroxilamina), e semelhantes, onde cada R₁, R₂ e R₃ são

independentemente H, ou alquilo possuindo 1-6 carbonos, preferencialmente H. Em certas formas de realização, o grupo reativo é uma hidrazida, hidroxilamina, semicarbazida, carbo-hidrazida, uma sulfonil-hidrazida, ou semelhantes.

O produto da reação entre o grupo nucleófilo e o eletrofílico tipicamente inclui os átomos originalmente presentes no grupo nucleofílico. Ligações típicas obtidas reagindo os aldeídos ou cetonas com os grupos nucleofílicos incluem produtos de reação tais como uma oxima, uma amida, uma hidrazona, uma hidrazona reduzida, um carbo-hidrazona, uma tiocarbo-hidrazona, uma sulfonil-hidrazona, uma semicarbazona, uma tiosemicarbazona, ou funcionalidade semelhante, dependendo do grupo nucleofílico usado e o grupo eletrofílico (p.ex., aldeído, cetona, e/ou semelhantes) que é reagido com o grupo nucleofílico. Ligações com os ácidos carboxílicos são geralmente referidas como carbohidrazidas ou ácidos hidroxâmicos. Ligações com ácidos sulfônicos são geralmente referidas como sulfonil-hidrazidas e N-sulfonil-hidroxilaminas. A ligação resultante pode ser subsequentemente estabilizada por redução química.

Estes métodos podem envolver adicionalmente contactar o grupo sacárido com uma glicosil transferase, um grupo dador de açúcar, e outros reagentes necessários para a atividade de glicosil transferase por um tempo suficiente e sob condições adequadas para transferir um açúcar do grupo dador de açúcar para o grupo sacárido. Em certas formas de realização, o método compreende adicionalmente contactar o produto da reação da glicosil transferase com pelo menos uma segunda glicosil transferase e um segundo grupo dador

de açúcar. Por outras palavras, certas formas de realização divulgadas neste documento fornecem métodos no qual um grupo sacárido ligado ao aminoácido ou um aminoácido antinatural que inclui um grupo sacárido é adicionalmente glicosilado. Estas etapas de glicosilação são preferencialmente (embora não necessariamente) realizadas enzimaticamente usando, por exemplo, uma glicosiltransferase, glicosidase, ou outra enzima conhecida dos peritos na arte. Em algumas formas de realização, uma pluralidade de etapas enzimáticas são realizadas numa mistura de reação única que contém duas ou mais glicosil transferases diferentes. Por exemplo, um pode realizar uma etapa de galactosilação e uma de sialilação simultaneamente, incluindo tanto sialil transferase e galactosil transferase na mistura de reação.

Para sínteses enzimáticas de sacáridos que envolvem reações de glicosil transferase, as células recombinantes opcionalmente contém pelo menos um gene heterólogo que codifica uma glicosil transferase. Muitas glicosil transferases são conhecidas, assim como as suas sequências polinucleotídicas. Ver, p.ex., "The WWW Guide To Cloned Glycosyl transferases," (disponível na World Wide Web). Sequências aminoacídicas de glicosil transferase e sequências nucleotídicas codificando glicosil transferases a partir da qual as sequências aminoacídicas podem ser deduzidas também estão presentes em várias bases de dados publicamente disponíveis, incluindo GenBank, Swiss-Prot, EMBL, e outros.

Em certas formas de realização, uma glicosil transferase inclui, mas não está limitada a, p.ex., uma galactosil transferase, uma fucosil transferase, uma glucosil

transferase, uma N-acetilgalactosaminil transferase, uma N-acetilglucosaminil transferase, uma glucuronil transferase, uma sialil transferase, uma manosil transferase, uma ácido glucurónico transferase, uma ácido galacturónico transferase, uma oligossacaril transferase, e semelhantes. Glicosil transferases adequadas incluem aquelas obtidas de eucariotas ou procariotas.

Um aceitador para as glicosil transferases estará presente na glicoproteína para ser modificado por métodos divulgados neste documento. Aceitadores apropriados, incluem, por exemplo, aceitadores de galactosil tais como Gal β 1,4GalNAc-; Gal β 1,3GalNAc-; lacto-N-tetraose-; Gal β 1,3GlcNAc-; Gal β 1,4GlcNAc-; Gal β 1,3Ara-; Gal β 1,6GlcNAc-; e Gal β 1,4Glc- (lactose). Outros aceitadores conhecidos dos peritos na arte (ver, p.ex., Paulson et al., J. Biol. Chem. 253: 5617-5624, 1978). Tipicamante, os aceitadores formam parte de uma cadeia de grupo sacárido que está acoplada à glicoproteína.

Numa forma de realização, o grupo sacárido compreende um GlcNAc terminal, o grupo dador de açúcar é UDP-GlcNAc e a glicosil transferase é uma β 1-4N-acetilglucosaminil transferase. N outra forma de realização, o grupo sacárido compreende um GlcNAc terminal, o grupo dador de açúcar é UDP-Gal e a glicosil transferase é uma β 1-4-galactosil transferase. Açúcares adicionais podem também ser adicionados.

As reações de glicosilação incluem, em adição à glicosil transferase e aceitador apropriados, um açúcar nucleotídico ativado que atua como um dador de açúcar para a glicosil transferase. As reações também podem incluir outros

ingredientes que facilitam a atividade da glicosil transferase. Estes ingredientes podem incluir um catião divalente (por exemplo, Mg^{2+} ou Mn^{2+}), materiais necessários para a regeneração de ATP, iões de fosfato, e solventes orgânicos. As concentrações ou quantidades dos vários reagentes usados nos processos dependem de vários fatores incluindo condições de reação tais como temperatura e valor de pH, e a escolha e quantidade de sacáridos aceitadores a ser glicosilados. O meio de reação também pode compreender detergentes solubilizantes (p.ex., Triton ou SDS) e solventes orgânicos tais como metanol ou etanol, se necessário.

Também fornecido por certas formas de realização para modificar uma glicoproteína são composições que incluem um sistema de tradução que pode ou não incluir uma célula hospedeira, um tRNA mutante externo, uma RS mutante externa, ou qualquer um ou todos estes.

Como usado neste documento, o termo "grupo sacárido" refere-se a grupos de açúcar natural e antinatural (ou seja, um grupo de açúcar que ocorre antinaturalmente, p.ex., um grupo de açúcar que é modificado, p.ex., em uma ou mais posições hidroxilo ou amina, p.ex., desidroxilado, desaminado, esterificado, etc, p.ex., 2-desoxiGal é um exemplo de um grupo de açúcar antinatural).

O termo "carboidrato" tem a fórmula geral $(CH_2O)_n$ e inclui, mas não está limitado a, p.ex., monossacáridos, dissacáridos, oligossacáridos e polissacáridos. Oligossacáridos são cadeias compostas por unidades de sacáridos, que são alternativamente conhecidas como açúcares. Unidades de sacáridos podem ser dispostas em

qualquer ordem e a ligação entre duas unidades de sacáridos pode ocorrer de qualquer uma das cerca de dez maneiras diferentes. As seguintes abreviações são usadas neste documento: Ara=arabinosil; Fru=fructosil; Fuc=fucosil; Gal=galactosil; GalNAc=N-acetilgalactosaminil; Glc=glicosil; GlcNAc=N-acetilglucosaminil; Man=manosil; e NeuAc=sialil (tipicamente N-acetilneuraminil).

Oligossacáridos são considerados possuir uma extremidade redutora e uma extremidade não-redutora, quer o sacárido na extremidade redutora seja ou não na verdade um açúcar redutor. Em conformidade com a nomenclatura aceite, oligossacáridos são descritos neste documento com a extremidade não-redutora à esquerda e a extremidade redutora à direita. Todos os oligossacáridos descritos neste documento são descritos com o nome ou a abreviatura para o sacárido não-redutor (p.ex., Gal), seguido pela configuração da ligação glicosídica (α ou β), a ligação do anel, a posição no anel do sacárido redutor envolvido na ligação, e depois o nome ou abreviatura do sacárido redutor (p.ex., GlcNAc). A ligação entre dois açúcares pode ser expressa, por exemplo, como 2,3; 2 $\rightarrow$ 3; 2-3; ou (2,3). Ligações naturais e antinaturais (p.ex., 1-2; 1-3; 1-4; 1-6; 2-3; 2-4; 2-6; etc.) entre dois açúcares são incluídas em certas formas de realização. Cada sacárido é uma piranose.

O termo "ácido siálico" (abreviado "Sia") refere-se a qualquer membro de uma família de açúcares de nove carbonos carboxilados. O membro mais comum da família de ácido siálico é ácido N-acetil-neuramínico (ácido 2-ceto-5-acetamido-3,5-didesoxi-D-glicero-D-galactononulopiranos-1-ónico) (frequentemente abreviado como Neu5Ac, NeuAc ou

NANA). Um segundo membro da família é ácido N-glicolil-neuramínico (Neu5Gc ou NeuGc), no qual o grupo de N-acetil de NeuAc é hidroxilado. Um terceiro membro da família de ácido siálico é ácido 2-ceto-3-desoxi-nonulosónico (KDN) (Nadano et al., J. Biol. Chem. 261: 11550-11557, 1986; Kanamori et al., J. Biol. Chem. 265: 21811-21819, 1990). Também incluídos estão ácidos siálicos 9-substituídos tais como um 9-O-C1-C6 acil-Neu5Ac como O-Neu5Ac-lactil-9 ou 9-O-acetil-Neu5Ac, 9-desoxi-9-fluoro-Neu5Ac e 9-azido-9-desoxi-Neu5Ac. Para revisão da família de ácido siálico, ver, p.ex., Varki, Glycobiology 2: 25-40, 1992; Sialic Acids: Chemistry, Metabolism and Function, R. Schauer, Ed. (Springer-Verlag, New York (1992)). A síntese e uso de compostos de ácido siálico num procedimento de sialilação é descrito em, por exemplo, o pedido internacional WO 92/16640 (todo o conteúdo incorporado neste documento por referência).

Substratos dadores para glicosil transferases são açúcares nucleotídicos ativados. Tais açúcares ativados geralmente consistem de uridina e guanosina difosfato, e citidina monofosfato, derivados dos açúcares nos quais o nucleósido difosfato ou monofosfato atua como um grupo solto. Sistemas bacterianos, de plantas e de fungos podem às vezes usar outros açúcares nucleotídicos ativados.

A incorporação de um aminoácido natural, p.ex., um aminoácido natural compreendendo um grupo onde um grupo sacárido pode ser acoplado, ou um aminoácido antinatural que inclui um grupo sacárido, pode ser feita para, p.ex., ajustar mudanças na estrutura e/ou função da proteína, p.ex., para alterar o tamanho, acidez, nucleofilicidade, ligação por pontes de hidrogénio, hidrofobicidade,

acessibilidade de sítios alvo de proteases, acesso alvo a um grupo proteico, etc. Proteínas que incluem um aminoácido antinatural, p.ex., um aminoácido antinatural compreendendo um grupo onde um grupo sacárido pode ser acoplado, ou um aminoácido antinatural que inclui um grupo sacárido, pode ter propriedades catalíticas ou físicas melhoradas ou mesmo inteiramente novas.

Por exemplo, as propriedades seguintes são opcionalmente modificadas pela inclusão de um aminoácido antinatural, p.ex., um aminoácido antinatural compreendendo um grupo onde um grupo sacárido pode ser acoplado, ou um aminoácido antinatural que inclui um grupo sacárido numa proteína: toxicidade, biodistribuição, propriedades estruturais, propriedades espectroscópicas, propriedades químicas e/ou fotoquímicas, capacidade catalítica, meia-vida (p.ex., meia-vida de soro), capacidade de reagir com outras moléculas, p.ex., covalentemente ou não covalentemente, e: semelhantes. As composições incluindo proteínas que incluem pelo menos um aminoácido antinatural, p.ex., um aminoácido natural compreendendo um grupo onde um grupo sacárido pode ser acoplado, ou um aminoácido antinatural que inclui um grupo sacárido são úteis para, p.ex., novos terapêuticos, diagnósticos, enzimas catalíticas, enzimas industriais, proteínas de ligação (por exemplo, anticorpos) e, p.ex., o estudo da estrutura e função da proteína. Ver, p.ex., Dougherty, Curr. Opin. in Chem. Biol., 4:645-652 (2000).

Num aspeto, uma composição inclui pelo menos uma proteína com pelo menos um, p.ex., pelo menos cerca de dois, três, quatro, cinco, seis, sete, oito, nove, ou pelo menos cerca de dez ou mais aminoácidos antinaturais, p.ex., um aminoácido antinatural compreendendo um grupo onde um grupo

sacárido pode ser acoplado, ou um aminoácido antinatural que inclui um grupo sacárido, e/ou que inclui outro aminoácido antinatural. Os aminoácidos antinaturais podem ser iguais ou diferentes, p.ex., podem existir 1, 2, 3, 4, 5, 6, 7, 8, 9 ou 10 ou mais sítios diferentes na proteína que compreende 1, 2, 3, 4, 5, 6, 7, 8, 9 ou 10 ou mais aminoácidos antinaturais diferentes. Noutro aspeto, uma composição inclui uma proteína com pelo menos um, mas menos do que todos, de um aminoácido particular presente na proteína substituída com o aminoácido antinatural, p.ex., um aminoácido antinatural compreendendo um grupo onde um grupo sacárido pode ser acoplado ou um aminoácido antinatural que inclui um grupo sacárido. Para uma determinada proteína com mais de um aminoácido natural, os aminoácidos antinaturais podem ser idênticos ou diferentes (p.ex., a proteína pode incluir dois ou mais tipos diferentes de aminoácidos antinaturais, ou pode incluir dois do mesmo aminoácido antinatural). Para uma determinada proteína com mais de dois aminoácidos antinaturais, os aminoácidos antinaturais podem ser o mesmo, diferentes, ou uma combinação de múltiplos aminoácidos antinaturais do mesmo tipo com pelo menos um aminoácido antinatural diferente.

Essencialmente qualquer proteína (ou parte desta) que inclui um aminoácido antinatural, p.ex., um aminoácido antinatural compreendendo um grupo onde um grupo sacárido é acoplado, tal como um aminoácido derivatizado de aldeído ou ceto, ou um aminoácido antinatural que inclui um grupo sacárido (e qualquer ácido nucleico de codificação correspondente, p.ex., que inclui um ou mais codões seletores) pode ser produzida usando os métodos e composições neste documento. Nenhuma tentativa é feita para

identificar as centenas de milhares de proteínas conhecidas, que podem ser modificadas para incluir um ou mais aminoácidos antinaturais, p.ex., ajustando quaisquer métodos de mutação disponíveis para incluir um ou mais códons degenerados adequados num sistema de tradução relevante. Repositórios comuns de sequência para proteínas conhecidas incluem GenBank EMBL, DDBJ e o NCBI. Outros repositórios podem ser facilmente identificados pesquisando a internet.

Normalmente, as proteínas são, p.ex., pelo menos cerca de 60%, 70%, 75%, 80%, 90%, 95%, ou pelo menos cerca de 99% ou mais idênticas a qualquer proteína disponível (p.ex., uma proteína terapêutica, uma proteína de diagnóstico, uma enzima industrial, ou parte da mesma, e semelhantes), e incluem um ou mais aminoácidos antinaturais.

Além de modificar um ou mais resíduos de aminoácidos da proteína, a composição de carboidratos da proteína pode ser modificada, ou seja, através de glicosilação. A modificação pós-traducional de proteínas por glicosilação pode afetar o enrolamento de proteínas e estabilidade, modificar a atividade intrínseca das proteínas, e modular as suas interações com outras biomoléculas. Ver, p.ex., Varki, *Glycobiology* 3: 97-130, 1993, incorporado na sua totalidade neste documento por referência. Glicoproteínas naturais estão frequentemente presentes como uma população de muitas glicoformas diferentes, o que torna difícil a análise da estrutura de glicano e o estudo dos efeitos de glicosilação na estrutura e função da proteína. Portanto, métodos para a síntese de proteínas homogeneamente glicosiladas naturais e antinaturais são necessários para a compreensão sistemática

da função dos glicanos, e para o desenvolvimento da terapêutica de glicoproteína melhorada.

Uma classe de proteínas que podem ser produzidas usando determinadas composições e métodos divulgados neste documento inclui moduladores transcricionais, enzimas, ou uma parte dos mesmos. Moduladores transcricionais exemplares incluem genes e proteínas moduladoras transcricionais que modulam o crescimento celular, diferenciação, regulação, ou semelhantes. Moduladores transcricionais estão presentes em procariotas, vírus, e eucariotas, incluindo plantas, fungos, leveduras, insetos, e animais, incluindo mamíferos, fornecendo uma ampla gama de alvos terapêuticos. Será apreciado que ativadores de expressão e transcricionais regulam a transcrição por vários mecanismos, p.ex., por ligação a recetores, estimulando uma cascata de transdução de sinal, regulando a expressão de fatores de transcrição, ligação a promotores e *enhancers*, ligação a proteínas que se ligam a promotores e *enhancers*, desenrolamento do DNA, *splicing* do pre-mRNA, poliadenilação do RNA, e degradando RNA. Alguns exemplos de enzimas incluem, mas não estão limitados a, por exemplo, amidases, aminoácido racemases, acilases, dehalogenases, dioxigenases, diarilpropano peroxidases, epimerases, epóxido hidrolases, esterases, isomerases, cinases, glicose isomerases, glicosidases, glicosil transferases, haloperoxidases, monooxygenases (p.ex., p450s), lipases, lenhina peroxidases, nitrilo hidratases, nitrilases, proteases, fosfatases, subtilisinas, transaminase, e nucleases.

Alguns dos polipeptídeos que podem ser modificados de acordo com determinadas formas de realização divulgadas

neste documento estão comercialmente disponíveis (ver, p.ex., a lista de catálogo e preço da Sigma Biosciences), e as correspondentes sequências de proteínas e genes e, normalmente, muitas variantes das mesmas, são conhecidas (ver, p.ex., Genbank).

Exemplos de propriedades terapeuticamente relevantes que podem ser manipuladas ou modificadas por qualquer um das formas de realização divulgadas neste documento (incluindo glicosilação e/ou adição de PEG, e/ou incorporação de aminoácidos não naturais) incluem meia-vida de soro, meia-vida de armazenamento, estabilidade, imunogenicidade, atividade terapêutica, detetabilidade (p.ex., pela inclusão de grupos repórter (p.ex., marcadores ou locais de ligação de marcadores) nos aminoácidos antinaturais, especificidade, redução de DL50 ou outros efeitos secundários, capacidade de entrar no corpo através do trato gástrico (p.ex., disponibilidade oral), ou semelhantes. Exemplos de propriedades de diagnóstico relevantes incluem meia-vida de armazenamento, estabilidade, atividade de diagnóstico, detetabilidade, especificidade, ou semelhantes. Exemplos de propriedades enzimáticas relevantes incluem meia-vida de armazenamento, estabilidade, especificidade, atividade enzimática, capacidade de produção, ou semelhantes.

Uma variedade de outras proteínas também podem ser modificadas para incluírem um ou mais aminoácidos antinaturais de acordo com certas formas de realização divulgadas neste documento. Por exemplo, as proteínas de fungos infecciosos, p.ex., espécies *Aspergillus*, *Candida*; bactérias, particularmente *E. coli*, que serve de modelo para bactérias patogénicas, bem como bactérias clinicamente

importantes tais como *Staphylococci* (p.ex., *aureus*), ou *Streptococci* (p.ex., *pneumoniae*); protozoários tais como esporozoários (p.ex., plasmódios), rizópodes (p.ex., *Entamoeba*) e flagelados (*Trypanosoma*, *Leishmania*, *Trichomonas*, *Giardia*, etc.); vírus tais como vírus de RNA(+) (exemplos incluem Poxvírus, p.ex., vaccinia; Picornavírus, p.ex., poliomielite; Togavírus, p.ex., rubéola; Flavivírus, p.ex., VHC; e Coronavírus), vírus de RNA(-) (p.ex., Rhabdovírus, p.ex., VSV; Paramixovírus, p.ex., RSV; Ortomixovírus, p.ex., gripe; Buniavírus; e Arenavírus), vírus dsDNA (Reovírus, por exemplo), vírus de RNA para DNA, ou seja, Retrovírus, p.ex., VIH e HTLV e certos vírus de DNA para RNA, como a Hepatite B.

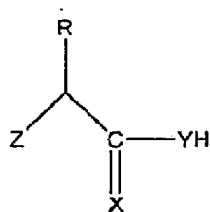
Proteínas agronomicamente relacionadas tais como proteínas de resistência a insetos (p.ex., as proteínas Cry), enzimas de produção de amido e lípidos, toxinas de plantas e insetos, proteínas de resistência a toxinas, proteínas de desintoxicação de micotoxina, enzimas de crescimento da planta (p.ex., Ribulose 1,5-Bisfosfato Carboxilase/Oxigenase, "RUBISCO"), lipoxigenase (LOX), e Fosfoenolpiruvato (PEP) carboxilase são também alvos adequados para modificação divulgadas por certas formas de realização neste documento.

Em certas formas de realização, a proteína ou polipéptido de interesse (ou parte dos mesmos) nos métodos e/ou composições divulgados neste documento é codificada por um ácido nucleico. Normalmente, o ácido nucleico compreende pelo menos um codão degenerado, pelo menos cerca de dois, três, quatro, cinco, seis, sete, oito, nove, ou pelo menos cerca de dez ou mais codões degenerados.

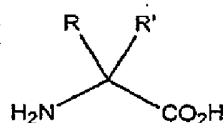
Assim os acima descritos polipéptidos e polinucleótidos artificiais (p.ex., criados pelo homem, e que não ocorrem naturalmente) são também característicos de certas formas de realização divulgadas neste documento. Um polinucleótido artificial pode incluir, p.ex., (a) um polinucleótido compreendendo uma sequência nucleotídica codificando um polipéptido artificial; (b) um polinucleótido que é complementar a ou que codifica uma sequência polinucleotídica de (a); (c) um ácido nucleico que hibridiza com um polinucleótido de (a) ou (b) sob condições de estringência em substancialmente todo o comprimento do ácido nucleico; (d) um polinucleótido que é pelo menos cerca de 95%, preferencialmente pelo menos cerca de 98% idêntico a um polinucleótido de (a), (b) ou (c); e, (e) um polinucleótido compreendendo uma variação conservativa de (a), (b), (c), ou (d).

Aminoácidos antinaturais são geralmente descritos acima. De particular interesse para fazer glicoproteínas como descrito neste documento são aminoácidos antinaturais nos quais R na fórmula I inclui um grupo que pode reagir com um grupo reativo que é acoplado a um grupo sacárido, para ligar o grupo sacárido a uma proteína que inclui o aminoácido antinatural. Grupos R apropriados incluem, por exemplo, ceto-, azido-, hidroxil-, hidrazina, ciano-, halo-, aminoóxi-, alceno, alcino, carbonilo, éter, tiol, seleno-, sulfonil-, borato, boronato, fosfo, fosfono, fosfina, heterocíclicos, enona, imina, aldeído, éster, tioácido, tioéster, éster bloqueado, hidroxilamina, amina, e semelhantes, ou qualquer combinação dos mesmos. Em algumas formas de realização, os aminoácidos antinaturais possuem um elemento de reticulação foto-ativável.

Além de aminoácidos antinaturais que contêm novas cadeias laterais, aminoácidos antinaturais também opcionalmente compreendem estruturas de cadeia principal modificadas, p.ex., conforme ilustrado pelas estruturas de Fórmula II e III:



Fórmula II



Fórmula III

em que Z tipicamente compreende OH, NH₂, SH, NH-R', ou S-R'; X e Y, que podem ser o mesmo ou diferentes, normalmente incluem S ou O e R e R', que opcionalmente são o mesmo ou diferentes, são tipicamente selecionados da mesma lista de componentes para o grupo R descrita acima para os aminoácidos antinaturais tendo Fórmula I assim como hidrogênio. Por exemplo, aminoácidos antinaturais divulgados neste documento são opcionalmente compreendem substituições no grupo amina ou carboxilo como ilustrado pelas Fórmulas II e III. Aminoácidos antinaturais deste tipo incluem, mas não estão limitados a, α-hidroxi ácidos, α α-tioácidos e aminotiocarboxilatos, p.ex., com cadeias laterais correspondentes ao vinte aminoácidos naturais comuns ou cadeias laterais antinaturais. Além disso, substituições no α-carbono opcionalmente incluem aminoácidos L, D ou α-α-disubstituídos como D-glutamato, D-alanina, D-O-metil-tirosina, ácido aminobutírico, e semelhantes. Outras alternativas estruturais incluem aminoácidos cíclicos, como análogos de prolina assim como análogos de anéis de prolina de 3, 4, 6, 7, 8 e 9 membros,

aminoácidos β e γ tais como β -alanina substituída e γ -amino ácido butírico.

Por exemplo, muitos aminoácidos antinaturais são baseados em aminoácidos naturais, tais como a tirosina, fenilalanina, glutamina, e semelhantes. Análogos de tirosina incluem tirosinas para-substituídas, tirosinas orto-substituídas, e tirosinas meta-substituídas, em que a tirosina substituída compreende um grupo acetilo, um grupo benzoílo, um grupo amina, uma hidrazina, um hidroxiamina, um grupo tiol, um grupo carboxilo, um grupo isopropilo, um grupo metilo, uma cadeia linear C6-C20 ou hidrocarboneto ramificado, um hidrocarboneto saturado ou insaturado, um grupo O-metilo, um grupo poliéter, um grupo nitro, ou semelhantes. Além disso, múltiplos anéis de arilo substituído também são contemplados. Análogos de glutamina incluem, mas não estão limitados a, derivados de α -hidroxi, derivados γ -substituídos, derivados cíclicos e derivados de glutamina de amida substituída. Exemplos de análogos de fenilalanina incluem, mas não estão limitados a, fenilalaninas meta-substituídas, orto-substituídas, e/ou para-substituídas, em que o substituinte compreende um grupo hidroxil, um grupo metóxi, um grupo metilo, um grupo alilo, um grupo aldeído ou ceto, ou semelhantes.

Exemplos específicos de aminoácidos antinaturais incluem, mas não estão limitados a, p-acetil-L-fenilalanina, O-metil-L-tirosina, uma L-3-(2-naftil)alanina, uma 3-metil-fenilalanina, uma O-4-alil-L-tirosina, uma 4-propil-L-tirosina, uma tri-O-acetil-GlcNAc β -serina, β -O-GlcNAc-L-serina, uma tri-O-acetil-GalNAc- α -treonina, uma α -GalNAc-L-treonina, uma L-dopa, uma fenilalanina fluorada, uma isopropil-L-fenilalanina, uma p-azido-L-fenilalanina, uma

p-acil-L-fenilalanina, uma p-benzoil-L-fenilalanina, uma L-fosfoserina, um fosfonoserina, uma fosfonotirosina, uma p-iodo-fenilalanina, uma p-bromofenilalanina, uma p-amino-L-fenilalanina, uma isopropil-L-fenilalanina, aqueles listados abaixo, ou noutro lugar neste documento, e semelhantes.

Aminoácidos antinaturais adequados para uso em alguns métodos divulgados neste documento também incluem aqueles que têm um grupo sacárido acoplado à cadeia lateral do aminoácido. Numa forma de realização, um aminoácido antinatural com um grupo sacárido inclui um aminoácido serina ou treonina com um grupo Man, GalNAc, Glc, Fuc ou Gal. Exemplos de aminoácidos antinaturais que incluem um grupo sacárido incluem, mas não estão limitados a, p.ex., uma tri-O-acetil-GlcNAc β -serina, uma β -O-GlcNAc-L-serina, uma tri-O-acetil-GalNAc- α -treonina, uma α -GalNAc-L-treonina, uma O-Man-L-serina, uma tetra-acetil-O-Man-L-serina, uma O-GalNAc-L-serina, uma tri-acetil-O-GalNAc-L-serina, uma Glc-L-serina, uma tetra-acetil-Glc-L-serina, uma fuc-L-serina, uma tri-acetil-fuc-L-serina, uma O-Gal-L-serina, uma tetra-acetil-O-Gal-L-serina, uma beta-O-GlcNAc-L-treonina, uma tri-acetil-beta-GlcNAc-L-treonina, um O-homem-L-treonina, um tetra-acetil-O-Man-L-treonina, uma O-GalNAc-L-treonina, uma tri-acetil-O-GalNAc-L-treonina, uma Glc-L-treonina, uma tetra-acetil-Glc-L-treonina, uma fuc-L-treonina, uma tri-acetil-fuc-L-treonina, uma O-Gal-L-treonina, uma tetra-acetil-O-Gal-L-serina, e semelhantes. Certas formas de realização também incluem formas desprotegidas e acetiladas das acima.

Em algumas formas de realização, o *design* de aminoácido antinaturais é enviesado por informação conhecida acerca

dos sítios ativos das sintetases, p.ex., tRNA sintetases mutantes externas usadas para aminoacilar um tRNA mutante externo. Por exemplo, três classes de análogos de glutamina são fornecidas, incluindo derivados substituídos no nitrogénio da amida (1), um grupo metil na γ -posição (2), e um derivado N-C γ -cíclico (3). Baseado na estrutura cristalina de raios-X da GlnRS de *E. coli*, na qual os resíduos do sítio de ligação chave são homólogos da GlnRS de levedura, os análogos foram concebidos para complementar uma gama de mutações de cadeia lateral de resíduos dentro de uma zona de 10 Å da cadeia lateral de glutamina, p.ex., uma mutação do sítio ativo Phe233 para um pequeno aminoácido hidrofóbico pode ser complementada por aumento de volume estérico na posição C γ de Gin.

Por exemplo, N-ftaloil-L-glutâmico 1,5-anidrido (composto número 4 na Fig. 23 de WO 02/085923) é opcionalmente usado para sintetizar análogos de glutamina com substituintes no nitrogénio da amida. Ver, p.ex., King and Kidd, J. Chem. Soc., 3315-3319, 1949; Friedman and Chatterrji, J. Am. Chem. Soc. 81, 3750-3752, 1959; Craig et al., J. Org. Chem. 53, 1167-1170, 1988; e Azoulay et al., Eur. J. Med. Chem. 26, 201-5, 1991. O anidrido é tipicamente preparado a partir de ácido glutâmico inicialmente pela proteção de amina como a ftalimida seguida por refluxo em ácido acético. O anidrido é então aberto com um número de amins, resultando numa variedade de substituintes na amida. Desproteção do grupo ftaloil com hidrazina proporciona um aminoácido livre, como apresentado na Fig. 23 de WO 2002/085923.

Substituição na γ -posição é normalmente realizada através da alquilação de ácido glutâmico. Ver, p.ex., Koskinen and

Rapoport, J. Org. Chem. 54, 1859-1866, 1989. Um aminoácido protegido, p.ex., conforme ilustrado pelo Composto número 5 na Fig. 24 de WO 02/085923, é opcionalmente preparado primeiro pela alquilação do grupo amina com 9-bromo-9-fenilfluoreno (PhflBr) (ver, p.ex., Christie and Rapoport, J. Org. Chem. 1989, 1859-1866, 1985) e depois esterificação do grupo ácido usando O-terc-butit-N,N'-diisopropilisourea. Adição de $\text{KN}(\text{Si}(\text{CH}_3)_3)_2$ desprotona seletivamente a região na β -posição do éster metílico para formar o enolato, que é depois opcionalmente alquilado com uma gama de iodetos de alquilo. Hidrólise do éster t-butílico e grupo Phfl produziu análogo de glutamina γ -metil desejado (Composto número 2 na Fig. 24 de WO 02/085923).

Certas outras formas de realização incluem um imunoconjugado compreendendo um anticorpo (ou o seu fragmento funcional) específico para um alvo (p.ex., uma célula alvo), o anticorpo (ou fragmento ou equivalente funcional do mesmo) conjugado, em posições específicas, pré-determinadas, com duas ou mais moléculas terapêuticas, em que cada uma das posições compreende um aminoácido antinatural. Em certas formas de realização, os fragmentos de anticorpos são $\text{F}(\text{ab}')_2$, Fab' , Fab , ou fragmentos de Fv .

Em certas formas de realização, as duas ou mais moléculas terapêuticas são as mesmas. Em certas formas de realização, as duas ou mais moléculas terapêuticas são diferentes. Em certas formas de realização, as moléculas terapêuticas são conjugadas com os mesmos aminoácidos antinaturais. Em certas formas de realização, as moléculas terapêuticas são conjugadas com aminoácidos antinaturais diferentes.

Em certas formas de realização, a natureza ou química da ligação aminoácido antinatural / molécula terapêutica permite a clivagem da ligação sob certas condições, tais como condições acídicas suaves ou fracas (p.ex., cerca de pH 4-6, preferencialmente cerca de pH 5), ambiente redutor (p.ex., a presença de um agente redutor), ou catiões divalentes, e opcionalmente é acelerada por calor.

Em certas formas de realização, a molécula terapêutica é conjugada com um anticorpo através um *linker* / espaçador (p.ex., uma ou mais repetições de metileno ($-\text{CH}_2-$), metilenoóxi ($-\text{CH}_2-\text{O}-$), metilenocarbonil ($-\text{CH}_2-\text{CO}-$), aminoácidos, ou combinações dos mesmos).

Complexos multiproteicos

Aminoácidos antinaturais também podem ser usados para juntar duas ou mais proteínas ou subunidades de proteína com funcionalidades únicas. Por exemplo, anticorpos biespecíficos podem ser gerados ligando dois anticorpos (ou partes funcionais dos mesmos ou derivados dos mesmos, tais como os fragmentos Fab, Fab', Fd, Fv, scFv, etc.) através de aminoácidos antinaturais incorporados naqueles.

Portanto certas formas de realização neste documento fornecem métodos para a síntese de conjugados multiproteicos. Estes métodos envolvem, em algumas formas de realização, incorporar numa primeira proteína (p.ex., um primeiro anticorpo) um primeiro aminoácido antinatural que compreende um primeiro grupo reativo; e contactar a primeira proteína com uma segunda proteína (p.ex., um segundo anticorpo) que inclui um segundo aminoácido antinatural que compreende um segundo grupo reativo, em que

o primeiro grupo reativo reage com o segundo grupo reativo, deste modo formando uma ligação covalente que acopla a segunda proteína à primeira proteína.

O primeiro grupo reativo é, em algumas formas de realização, um grupo eletrofílico (p.ex., um grupo ceto, um grupo aldeído, e/ou semelhantes), e o segundo grupo reativo é um grupo nucleofílico. Em algumas formas de realização, o primeiro grupo reativo é um grupo nucleofílico e o segundo grupo reativo é um grupo eletrofílico (por exemplo, um grupo ceto, um grupo aldeído, e/ou semelhantes). Por exemplo, um grupo eletrofílico é acoplado ao aminoácido antinatural do primeiro Ab, e o grupo nucleofílico é acoplado ao aminoácido antinatural do segundo Ab.

Domínios funcionais diferentes de proteínas diferentes podem ser ligados entre si através de forma similar para criar novas proteínas com novas funções (p.ex., novos fatores de transcrição com uma combinação única de domínios de ligação a DNA e ativação de transcrição, novas enzimas com novos domínios regulatórios, etc.).

Ligação sensível ao pH

Muitas interações proteicas são sensíveis ao pH, no sentido de que a afinidade de ligação de uma proteína ao seu parceiro habitual de ligação pode mudar conforme o pH ambiente muda. Por exemplo, muitos ligandos (tais como insulina, interferões, hormonas de crescimento, etc.) ligam-se aos seus respetivos recetores de superfície para desencadear transdução de sinal. O complexo ligando-recetor irá então ser internalizado por endocitose mediada por recetores, e passam por uma série sucessiva de endossomas

cada vez mais ácidos. Eventualmente, a interação ligando-recetor fica enfraquecida a um determinado pH ácido (p.ex., cerca de pH 5,0) e o ligando dissocia-se do recetor. Alguns recetores (e talvez alguns ligandos) podem ser reciclados de volta para a superfície da célula. Ali, poderão ligar-se aos seus respetivos parceiros de ligação normal.

Se a ligação sensível ao pH pode ser modulada de tal modo que o complexo ligando-recetor pode ser dissociado a um pH relativamente mais elevado, então certos ligandos podem ser dissociados dos seus recetores mais cedo, e são preferencialmente reciclados para a superfície da célula, em vez de serem degradados. Isto resultará numa meia-vida *in vivo* aumentada de tais ligandos, o que pode ser desejável já que menos insulina pode ser necessária para a mesma (ou melhor) eficácia em pacientes diabéticos. Noutras situações, pode ser desejável para modular a ligação sensível ao pH favorecendo a ligação a um pH mais baixo.

Por exemplo, anticorpos monoclonais são geralmente muito específicos para os seus alvos. No entanto, em muitas aplicações, tais como na terapia do cancro, eles tendem a desencadear efeitos secundários por, por exemplo, ligando-se a tecidos não-tumorais. Uma razão pode ser que os alvos tumorais contra os quais os anticorpos monoclonais são gerados não são especificamente expressos em células tumorais, mas também são expressos (embora possam ser em menor número) em algumas células saudáveis. Estes efeitos secundários são geralmente indesejáveis, e há uma necessidade de anticorpos com uma especificidade melhorada.

O pH do sangue humano é altamente regulado e mantido na faixa de cerca de 7,6-7,8. Por outro lado, as células tumorais têm um pH extracelular de 6,3-6,5, devido à acumulação de ácidos metabólicos que são ineficientemente removidos devido à fraca vascularização do tumor. Se a interação entre um antígeno tumoral e o seu anticorpo terapêutico pode ser modulada de tal modo a que a pH baixo, a ligação é favorecida, o tumor-anticorpo pode ter uma especificidade / afinidade / seletividade adicionada para aqueles antígenos tumorais, embora os mesmos antígenos tumorais também sejam ocasionalmente encontrados em tecidos normais.

Na verdade, tais anticorpos modificados podem ser desejáveis não só para a terapia do cancro, mas também desejáveis para qualquer ligação antígeno-anticorpo que pode ocorrer a um nível de pH mais baixo do que o normal.

Técnicas gerais

A prática das formas de realização divulgadas neste documento vai utilizar, salvo indicação em contrário, técnicas convencionais de biologia molecular, biologia celular, cultura celular, microbiologia e DNA recombinante, que estão no âmbito da arte. Tais técnicas são explicadas totalmente na literatura. Ver, por exemplo, Molecular Cloning: A Laboratory Manual, 2nd Ed., ed. By Sambrook, Fritsch and Maniatis (Cold Spring Harbor Laboratory Press: 1989); DNA Cloning, Volumes I and II (D. N. Glover ed., 1985); Oligonucleotide Synthesis (M. J. Gait ed., 1984); Mullis et al.; U.S. Patent No: 4,683,195; Nucleic Acid Hybridization (B. D. Hames & S. J. Higgins eds. 1984); Transcription And Translation (B. D. Hames & S. J. Higgins

eds. 1984); B. Perbal, A Practical Guide To Molecular Cloning (1984); the treatise, Methods In Enzymology (Academic Press, Inc., N.Y.); Methods In Enzymology, Vols. 154 and 155 (Wu et al, eds.), Immunochemical Methods In Cell And Molecular Biology (Mayer and Walker, eds., Academic Press, London, 1987).

Além disso, textos gerais divulgando clonagem em geral, mutação, cultura de células e semelhantes, incluem Berger and Kimmel, Guide to Molecular Cloning Techniques, Methods in Enzymology vol. 152 Academic Press, Inc., San Diego, Calif. (Berger); Sambrook et al., Molecular Cloning-A Laboratory Manual (3rd Ed.), Vol. 1-3, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., 2000 ("Sambrook") and Current Protocols in Molecular Biology, F. M. Ausubel et al., eds., Current Protocols, a joint venture between Greene Publishing Associates, Inc. and John Wiley & Sons, Inc., (supplemented through 2002) ("Ausubel")) são todos incorporados na sua totalidade neste documento por referência. Estes textos descrevem mutagénese, o uso de vetores, promotores e muitos outros tópicos relevantes relacionados com, p.ex., a geração de tRNA mutante externo, sintetases mutantes externas, e pares dos mesmos.

Vários tipos de mutagénese são usados em certas formas e realização, p.ex., para produzir novas sintetases ou tRNAs. Eles incluem mas não estão limitados a sítio-dirigida (tais como através da utilização de âmbar, ocre, opala ou outros codões stop), através de mutagénese de codão *wobble*, mutagénese pontual aleatória, recombinação homóloga (DNA *shuffling*), mutagénese usando moldes contendo uracilo, mutagénese dirigida por oligonucleótido, mutagénese de DNA modificado por fosforotioato, mutagénese usando DNA de

cadeia dupla intervalada ou semelhantes. Métodos adicionais adequados incluem reparação pontual *mismatch*, mutagénese usando estirpes hospedeiras deficientes na reparação, seleção por restrição e purificação por restrição, mutagénese por deleção, mutagénese por síntese total do gene, reparação da quebra de cadeia dupla, e semelhantes. Mutagénese, p.ex., envolvendo construções quiméricas, também está incluída em certas formas de realização. Numa forma de realização, mutagénese pode ser conduzida por informações conhecidas da molécula natural ou da molécula natural alterada ou mutada, p.ex., sequência, comparações de sequência, propriedades físicas, estrutura cristalina ou semelhantes.

Os textos e exemplos acima presentes neste documento descrevem estes procedimentos assim como as seguintes publicações e referências citadas em: Sieber, et al., *Nature Biotechnology*, 19:456-460 (2001); Ling et al., *Approaches to DNA mutagenesis: an overview*, *Anal. Biochem.* 254(2): 157-178 (1997); Date et al., *Methods Mol. Biol.* 57:369-374 (1996); I. A. Lorimer, I. Pastan, *Nucleic Acids Res.* 23, 3067-8 (1995); W. P. C. Stemmer, *Nature* 370, 389-91 (1994); Arnold, *Curr. Opin. in Biotech.* 4:450-455 (1993); Bass et al., *Science* 242:240-245 (1988); Fritz et al., *Nucl. Acids Res.* 16: 6987-6999 (1988); Kramer et al., *Nucl. Acids Res.* 16: 7207 (1988); Sakamar and Khorana, *Nucl. Acids Res.* 14: 6361-6372 (1988); Sayers et al., *Nucl. Acids Res.* 16:791-802 (1988); Sayers et al., *Nucl. Acids Res.* 16: 803-814 (1988); Carter, *Methods in Enzymol.* 154: 382-403 (1987); Kramer & Fritz *Methods in Enzymol.* 154:350-367 (1987); Kunkel, *Nucleic Acids & Mol. Biol.* (Eckstein, F. and Lilley, D. M. J. eds., Springer Verlag, Berlin) (1987); Kunkel et al., *Methods in Enzymol.* 154, 367-382

(1987); Zoller & Smith, *Methods in Enzymol.* 154:329-350 (1987); Carter, *Biochem. J.* 237:1-7 (1986); Eghtedarzadeh & Henikoff, *Nucl. Acids Res.* 14: 5115 (1986); Mandecki, *PNAS, USA*, 83:7177-7181 (1986); Nakamaye & Eckstein, *Nucl. Acids Res.* 14: 9679-9698 (1986); Wells et al., *Phil. Trans. R. Soc. Lond. A* 317: 415-423 (1986); Botstein & Shortle, *Science* 229:1193-1201(1985); Carter et al., *Nucl. Acids Res.* 13: 4431-4443 (1985); Grundström et al., *Nucl. Acids Res.* 13: 3305-3316 (1985); Kunkel, *PNAS, USA* 82:488-492 (1985); Smith, *Ann. Rev. Genet.* 19:423-462(1985); Taylor et al., *Nucl. Acids Res.* 13: 8749-8764 (1985); Taylor et al., *Nucl. Acids Res.* 13: 8765-8787 (1985); Wells et al., *Gene* 34:315-323 (1985); Kramer et al., *Nucl. Acids Res.* 12: 9441-9456 (1984); Kramer et al., *Cell* 38:879-887 (1984); Nambiar et al., *Science* 223: 1299-1301 (1984); Zoller & Smith, *Methods in Enzymol.* 100:468-500 (1983); e Zoller & Smith, *Nucl Acids Res.* 10:6487-6500 (1982). Detalhes adicionais sobre muitos dos métodos acima podem ser encontrados em *Methods in Enzymology Volume 154*, que também descreve controlos úteis para resolução de problemas com vários métodos de mutagénese.

Oligonucleótidos, p.ex., para uso em mutagénese em certas formas de realização, p.ex., mutando bibliotecas de sintetases, ou alterando tRNAs, são tipicamente sintetizados quimicamente segundo o método de triéster fosforamidita de fase sólida descrito por Beaucage and Caruthers, *Tetrahedron Letts.* 22(20):1859-1862, (1981), p.ex., usando um sintetizador automatizado, como descrito em Needham-VanDevanter et al., *Nucl Acids Res.*, 12:6159-6168 (1984).

Além disso, essencialmente qualquer ácido nucleico personalizado ou padrão pode ser encomendado a partir de uma variedade de fontes comerciais, tais como The Midland Certified Reagent Company, The Great American Gene Company, ExpressGen Inc., Operon Technologies Inc. (Alameda, Calif.) e muitas outras.

Todas as formas de realização descritas neste documento destinam-se a serem capazes de ser combinadas com uma ou várias outras formas de realização, mesmo para aquelas descritas em diferentes secções da divulgação.

Todas as patentes dos EUA, publicações de pedidos de patente dos EUA, pedidos de patentes dos EUA, patentes estrangeiras, pedidos de patentes estrangeiros e publicações não-patente acima referidos nesta descrição e/ou listados na Application Data Sheet, são incorporados na sua totalidade neste documento por referência.

EXEMPLOS

Estes exemplos ilustram a incorporação de um análogo de aminoácido em proteínas em posições codificadas por códons que normalmente codificam especificamente fenilalanina (Phe) ou codificam especificamente triptofano (Trp). Um diagrama esquemático é mostrado na Figura 1. Abordagens similares podem ser usadas para quaisquer outros análogos.

Phe é codificada por dois códons, UUC e UUU. Ambos os códons são lidos por um único tRNA, que está equipado com a sequência anticodão GAA. O codão UUC é portanto reconhecido através de emparelhamento de bases de Watson-Crick padrão entre codão e anticodão; UUU é lido através de um par de

bases *wobble* G-U na primeira posição do anticodão (Crick, J. Mol. Biol. 19: 548, 1966; Soll and RajBhandary, J. Mol. Biol. 29: 113, 1967). Desnaturação térmica de duplas hélices de RNA produziu estimativas de energias livres de Gibbs de derretimento dos pares de base G-U, G-C, A-U e A-C de 4,1, 6,3, 6,5 e 2,6 kcal/mol, respetivamente, a 37°C. Portanto, o par de bases *wobble*, G-U, é menos estável do que o par de bases de Watson-Crick, A-U. Um tRNA^{Phe} modificado, equipado com o anticodão AAA (tRNA^{Phe}_{AAA}) foi projetado para ler o codão UUU, e foi previsto ler tais codões mais rápido do que o tRNA^{Phe}_{GAA} de tipo selvagem.

Dihidrofolato redutase murina (mDHFR), que contém nove resíduos Phe, foi escolhida como a proteína de teste. O plasmídeo de expressão pQE16 codifica mDHFR sob o controlo de um promotor de bacteriófago T5; a proteína está equipada com um marcador C-terminal de hexa-histidina (HIS6) para facilitar a purificação através de cromatografia de afinidade de metal imobilizado.

A PheRS modificada de levedura (mu-yPheRS) foi preparada pela introdução de uma mutação Thr415Gly ou Thr415Ala na subunidade α -a da sintetase (Datta et al., J. Am. Chem. Soc. 124: 5652, 2002). As cinéticas de ativação do Nal e Phe por mu-yPheRS foram analisadas *in vitro* através do ensaio de troca de adenosina trifosfato-pirofosfato. A constante de especificidade (k_{cat}/K_M) para a ativação de Nal por mu-yPheRS foi descoberta ser $1,55 \times 10^{-3} (\text{s}^{-1} (\text{M}^{-1}))$, 8 ordens de grandeza maior do que a de Phe. Portanto, quando o rácio de Nal para Phe no meio de cultura é elevado, ytrNA^{Phe}_{AAA} deverá ser carregado predominantemente com Nal. Além disso, o mutante T415G foi gerado por mutagénese de quatro iniciadores.

Ambas as sintetases de *E. coli* e de levedura são $\alpha_2\beta_2$ hetero-tetrâmeros e os pesos moleculares de cada subunidade são bastante diferentes $\alpha(\text{ePheRS}) = 37 \text{ kDa}$; $\alpha(\text{yPheRS}) = 57 \text{ kDa}$; $\beta(\text{ePheRS}) = 87 \text{ kDa}$; e $\beta(\text{yPheRS}) = 67,5$, todos aproximadamente.

Assim, os exemplos seguintes são fornecidos como forma de ilustração e não por meio de limitação.

EXEMPLO 1

A fim de alterar a capacidade de uma aminoacil-tRNA sintetase de levedura, o gene yPheRS foi amplificado a partir do molde do plasmídeo pUC-ASab2 codificando as alfa e beta subunidades do gene PheRS. A amplificação foi realizada com uma sequência intergênica de 14 pares de base contendo um local de reiniciação de tradução a montante do codão de iniciação ATG do gene da subunidade beta.

Os seguintes iniciadores oligo foram usados para o PCR: 5'-CGA TTT TCA CAC AGG ATC CAG ACC ATG ATT CTA G-3' (SEQ ID NO:7) (iniciador 1 com local de restrição *Bam*HI) e 5'-GAC GGC CAG TGA ATT CGA GCT CGG TAC-3' (SEQ ID NO: 8) (iniciador 2 com local de restrição *Kpn*I). O produto de DNA resultante foi introduzido nos locais *Bam*HI e *Kpn*I de pQE32 para produzir pQE32-yFRS. A polinucleótido de yPheRS mutante foi gerado usando mutagénese por iniciador por técnicas padrão.

Resumidamente, dois oligonucleótidos complementares: 5'-CTA CCT ACA ATC CTT ACG GCG AGC CAT CAA TGG AAA TC-3' (SEQ ID NO:9) (iniciador 3) e 5'-GAT TTC CAT TGA TGG CTC GCC GTA

AGG ATT GTA GGT AG-3' (SEQ ID NO: 10) (iniciador 4) foram sintetizados para conter a mutação específica na posição 415 de subunidade alfa do polinucleótido de yPheRS.

EXEMPLO 2

O plasmídeo pQE32-yFRS, e pQE32-T415G, pQE32-T415A foram cada um transformados em células hospedeiras de *E. coli* da estirpe BLR (da NOVAGEN®) para formar estirpes de expressão BLR (pQE32-yFRS_e BLR(pQE32-T415G)). As células foram cultivadas em meio LB, a uma concentração de 0,6 a DO 600. A expressão então foi induzida com 1 mM IPTG durante 4 horas. As células foram coletadas e polipéptidos foram purificados por meio de uma coluna de afinidade de ácido nitrilotriacético-níquel sob condições nativas de acordo com o protocolo do fabricante (QIAGEN®). O imidazole no tampão de eluição foi removido pela coluna de dessalinização e polipéptidos foram eluídos num tampão contendo 50 mM Tris-HCl (pH = 7,5), 1 mM DTT. Aliquotas de polipéptidos foram armazenadas a -80°C com 50% de glicerol.

EXEMPLO 3

A reação de troca ATP-PP_i dependente do aminoácido foi usada para avaliar a ativação de aminoácidos não-naturais por yPheRS. O ensaio foi realizado em 200 microlitros de tampão de reação contendo 50 mM HEPES ((pH=7,6), 20 mM MgCl₂, 1 mM DTT, 2 mM ATP e 2 mM [³²P]-PP_i com atividade específica de 0,2-0,5 TBq/mol. Dependendo da atividade dos vários aminoácidos não-naturais com a sintetase, a concentração de aminoácido variou de 10 microM a 5 mM e a concentração de enzima variou de 10 nM a 100 nM. Aliquotas de 20 microlitros foram removidas da solução de reação em

vários pontos temporais e reprimidos em 500 microlitros de solução tampão contendo 200 mM NaPP_i, 7% p/v HClO₄ e 3% p/v carvão ativado. O carvão foi centrifugado e lavado duas vezes com 500 microlitros de 10 mM NaPP_i e solução 0,5% HClO₄. O ATP radiomarcado absorvido no carvão foi quantificado através de métodos de cintilação líquida. As constantes de especificidade foram calculadas pelo ajuste de regressão não-linear dos dados a um modelo de Michaelis Menten. Os parâmetros cinéticos para a troca ATP-PP_i de aminoácidos pela yPheRS (T415G), yPheRS de tipo selvagem, e variante yPheRS_naph são apresentados na tabela abaixo.

Aminoácido	Enzima	Km (μM)	Kcat (s ⁻¹)	Kcat/Km (M ⁻¹ s ⁻¹)	Kcat/Km (relativo)
Phe	T415G	55+/-14	0,202+/-0,11	3512+/-1134	1 ^a
Trp	T415G	2,83+/-1,6	0,153+/-0,003	63190+/-34590	18 ^a
2NaI	T415G	7,03+/-0,14	0,208+/-0,04	29535+/-5848	8,4 ^a
Phe	Tipo selvagem	3,85+/-0,99	0,181+/-0,011	50994+/-22655	15 ^a
Phe	naph	11010+/-2688	0,0095+/-0,0021	0,855+/-0,007	1 ^b
Trp	naph	1424+/-597	0,0035+/-0,0009	2,52+/-0,44	2,9 ^b
2NaI	naph	2030+/-691	0,030+/-0,018	14,54+/-4,22	17 ^b

EXEMPLO 4

O plasmídeo de expressão, pQE16 (QIAGEN®) foi usado com marcador de polipéptido dihidrofolato redutase murina (mDHFR) com um gene marcador C-terminal de hexa-histidina sob o controlo de um promotor e terminador t₀ de bacteriófago T5.

Um codão âmbar (TAG) foi colocado na posição 38 do mDHFR usando um *kit* de mutagénese QUICK-CHANGE®. Dois iniciadores oligo complementares (5'-CCG CTC AGG AAC GAG TAG AAG TAC TTC CAA AGA ATG-3' (SEQ ID NO: 11) e 5'-CAT TCT TTG GAA GTA CTT CTA CTC GTT CCT GAG CGG-3' (SEQ ID NO: 12)) foram usados para produzir pQE16am. O gene yPheRS T415G mutante foi amplificado a partir de pQE32-T415G e um promotor constitutivo *tac* com um sítio de ligação do repressor *lac* abolido foi adicionado a montante do codão de iniciação do gene.

A cassette de expressão inteira de T415G foi inserida no local de PvuII de pQE16 para produzir pQE16am-T415G. O tRNA supressor mutante de levedura (*mutRNA*^{Phe}(CUA)) foi constitutivamente expresso sob o controlo do promotor de *Ipp*. A cassette de expressão de *mutRNA*^{Phe}(CUA) foi inserida no plasmídeo repressor pREP4 para formar pREP4-tRNA usando métodos conhecidos.

Uma estirpe bacteriana auxotrófica de fenilalanina (Phe); AF (K10, Hfr(Cavalli) *pheS93rel-1 tonA22 thi T2^R pheA18*) foi usada como estirpe hospedeira. Uma estirpe dupla auxotrófica dupla de Phe/Trp, AFW (K10, Hfr(Cavalli) *pheS13rel-1 tonA22 thi T2^R pheA18, trpB114*) e uma estirpe auxotrófica tripla de Phe/Trp/Lys AFWK (K10, Hfr(Cavalli) *pheS13rel-1 tonA22 thi T2^R pheA18, trpB114, lysA*) foram preparadas por transdução mediada pelo fago P1 com transposões *trpB::Tn10* e *lysA::Tn10*.

EXEMPLO 5

As estirpes de células hospedeiras auxotróficas AF, AFW, e AFWK foram cada uma transformada com o plasmídeo pQE16am-T415G e a pREP4-tRNA para produzir estirpes de expressão AF[pQE16am-T415G/pREP4-tRNA] e AFW[pQE16am-T415G/pREP4-tRNA], respetivamente. As células foram cultivadas em meio mínimo M9, suplementado com glicose, tiamina, MgSO_4 , CaCl_2 , 20 aminoácidos (20 mg/L), antibióticos (kanamicina e ampicilina). Quando as células alcançaram uma leitura de DO600 de 1,0, foram sedimentadas por centrifugação, lavadas duas vezes com 0,9% NaCl frio e mudadas para meio M9 suplementado contendo 17 aminoácidos (20 mg/L), 3 mM aminoácido não-natural de interesse, e as concentrações indicadas de Phe, Trp e Lys. A expressão da proteína foi induzida pela adição de IPTG (1 mM). Após 4 horas, as células foram peletizadas e a proteína foi purificada por meio de um marcador C-terminal de hexa-histidina e uma coluna de centrifugação de níquel-NTA de acordo com as instruções do fabricante. (QIAGEN®).

EXEMPLO 6

mDHFR mutante foi purificada sob condições de desnaturação e eluída com tampão padrão (ureia 8 M, 100 mM NaH_2PO_4 , 10mM Tris, pH 4,5). Os polipéptidos foram digeridos com tripsina com 10 microlitros da solução diluída em 90 microlitros de 75 mM $((\text{NH}_4)_2\text{CO}_3$ e o pH foi ajustado para 8. Dois microlitros de tripsina modificada (0,2 microgramas/microlitro) foram adicionados, e a amostra foi incubada à temperatura ambiente *overnight*. Os polipéptidos foram digeridos com endoproteinase Lys-C, 10 microlitros de solução diluída em 90 microlitros de 25 mM Tris-HCl, pH 8 e

1 mM EDTA. Depois, 2 microlitros de Lys-C (0,2 microgramas/microlitro; CALBIOCHEM®) foram adicionados e a reação foi incubada a 37°C durante 10 horas. A reação de digestão foi interrompida pela adição de 2 microlitros de ácido trifluoroacético (TFA). A solução foi purificada por meio de ZIPTIP_{C18}® (MILLIPORE®) e os péptidos foram digeridos eluídos com 3 microlitros de 50% CH₃CN, 0,1% TFA. Um microlitro foi usado para análise de espectrometria de massa de ionização e dessorção a laser assistida por matriz (MALDI-MS) com ácido alfa-ciano-4-hidroxicinâmico e ácido 2,5-benzóico como a matriz. A análise foi realizada usando um espectrômetro de massa PERSEPTIVE BIOSYSTEMS® Voyager DE PRO MALDI-TOF em modos iônicos linear e positivos.

Análise de LC-MS/MS, de péptidos digeridos por proteases foi conduzida em espectrometria de massa ion trap FINNIGAN® LCQ com bomba HPLC e sonda ESI. Sequenciação de massa em *tandem* foi realizada pela fragmentação do ião precursor com m/z correspondendo ao fragmento digerido por proteases incluindo o resíduo na posição 38 do mutante mDHFR.

EXEMPLO 7

Plasmídeos foram construídos para yPheRS de tipo selvagem e yPheRS (T415G) conforme descrito no Exemplo 1 neste documento. Além disso, o gene *lysS* de *E. coli* foi amplificado por PCR a partir do molde do plasmídeo pXLLysKS1, usando os seguintes iniciadores: 5'-GCA CTG ACC ATG GCT GAA CAA CAC GCA CAG-3' (SEQ ID NO: 13) (com local de restrição *NcoI*) e 5'-GGA CTT CGG ATC CTT TCT GTG GGC GCA TCG C-3' (SEQ ID NO: 14) (com local de restrição *BamHI*). O DNA resultante foi introduzido nos locais *NcoI* e *BamHI* de pQE60 para produzir pQE60-eLysS. As enzimas clonadas contêm

marcadores N-terminais ou C-terminais de hexa-histidina para facilitar a purificação de proteínas.

Nas primeiras duas reações, (iniciador 1 e iniciador 4) e (iniciador 2 e iniciador 3) foram adicionados a tubos individuais e dois fragmentos de DNA foram gerados a partir destas duas reações de PCR. Com a mistura de dois produtos de reação e iniciadores adicionais exteriores, um fragmento de DNA de 3400bp foi obtido. O fragmento foi purificado por métodos padrão e digerido com *Bam*HI e *Kpn*I e inserido em pQE32 para produzir pQE32-T415G. As enzimas PheRS clonadas continham uma sequência marcadora N-terminal de hexa-histidina conhecida para purificação. O DNA do gene yPheRS completo foi sequenciado para verificação.

EXEMPLO 8

O plasmídeo PQE32-T415A, e pQE60-eLysS foram individualmente co-transformados com um plasmídeo repressor pREP4 numa estirpe BLR de *E. coli* para formar estirpes de expressão BLR (pQE32-yFRS), BLR (pQE32-T415G), BLR (pQE32-T415A) e BLR (pQE60-eLysS). Sobre-expressão foi realizada em meios 2x YT com 100 microgramas/mL de ampicilina e 35 microgramas/mL de kanamicina. A DO 600 = 0,6, expressão de variantes de yPheRS e lisil-tRNA sintetase de *E. coli* codificada pelo gene *lysS* (eLysS) foram induzidas com 1 mM IPTG. Após expressão de 4 horas, as células foram coletadas e proteínas foram purificadas numa coluna de afinidade de ácido nitrilotriacético-níquel sob condições nativas de acordo com o protocolo do fabricante (QIAGEN®). O imidazole no tampão de eluição foi removido por coluna de dessalinização e polipéptidos foram eluídos num tampão contendo 50 mM Tris-HCl (pH 7,5), 1 mM DTT. Aliquotas de

polipéptidos foram armazenadas a -80°C com 50% de glicerol. Concentrações de variantes de yPheRS e eLysS foram determinadas por absorção UV a 280 nm.

EXEMPLO 9

O péptido 38 (resíduos 26-39; NGDLPWPPLRNEamber codonK) (SEQ ID NO: 15) contém o codão âmbar na posição número 38. Péptidos (K38S, K38L), Péptido W38 e Péptido pBrF (Z)38 foram separados e detetados por MS. Polipéptidos foram sintetizados em células hospedeiras auxotróficas triplas com (a) tRNA^{Phe}_{CUA} e yPheRS (T415G); (b) tRNA^{Phe}_{CUA_UG} e yPheRS (T415G); (c) tRNA^{Phe}_{CUA} e yPheRS (T415A); (d) tRNA^{Phe}_{CUA_UG} e yPheRS (T415A) ou (e) numa estirpe auxotrófica simples tRNA^{Phe}_{CUA_UG} e yPheRS (T415A). Os meios mínimos de expressão foram suplementados com 6,0 mM pBrF, 0,01 mM Trp, 1,0 mM Lys, 0,03 mM Phe (a e b) ou 0,01 mM Phe (c, d e e) e 25 mg/L de 17 aminoácidos, resultados são apresentados na Figura 9.

EXEMPLO 10

A reação de troca ATP-PP_i dependente do aminoácido foi usada para avaliar a ativação dos análogos de aminoácidos por yPheRS conforme descrito nos exemplos acima. Brevemente, uma alíquota de 200 microlitros de tampão de reação contém 50 mM HEPES (pH 7,6), 20 nM MgCl₂, 1 mM DTT, 2 mM ATP, e 2mM ³²P-pirofosfato (PP_i) com atividade específica de 10-50 Ci/mol. Dependendo da atividade de análogos pela sintetase, a concentração de aminoácido variou de 10 microM a 2,5 mM e a concentração de enzima variou de 10 nM a 100 nM. Alíquotas de 20 microlitros foram removidas da solução de reação em vários pontos temporais e

reprimidas em 500 microlitros de solução de tampão contendo 200 mM NaPP_i, 7% p/v HClO₄, e 3% carvão ativado. O carvão foi centrifugado e lavado duas vezes com 500 microlitros de 10mM NaPP_i e 0,5% solução de HClO₄. O ATP radiomarcado absorvido no carvão foi quantificado através de métodos de cintilação líquida. As constantes de especificidade foram calculadas por regressão não linear para os dados de um modelo de Michaelis Menten.

Os resultados dos parâmetros cinéticos são apresentados na Tabela I. Substituição no anel indol (especialmente na posição 6) foi altamente favorável para alguns análogos (8-10).

Tabela I: Parâmetros Cinéticos para troca ATP-PP_i de aminoácidos exemplares (1-11) pela PheRS mutante externa de levedura.

Aminoácido	Enzima	Km (μM)	kcat (s ⁻¹)	kcat/Km (M ⁻¹ s ⁻¹)	kcat/Km (relativo)*
1	T415G	264+/-42	0,05+/-0,002	184+/-30	1
2	T415G	22+/-3	0,03+/-0,001	1538+/-228	8
3	T415G	12+/-2	0,05+/-0,001	4365+/-797	24
4	T415G	11+/-3	0,05+/-0,002	4558+/-1186	25
5	T415G	757+/-149	0,4+/-0,003	48+/-10	1\4
6	T415G	20+/-5	0,30+/-0,006	15000+/-4063	82
7	T415G	27+/-2	0,04+/-0,001	1550+/-125	8
8	T415G	20+/-8	0,20+/-0,018	10256+/-4562	56
9	T415G	8+/-4	0,55+/-0,097	70876+/-34843	385
10	T415G	31+/-18	0,06+/-0,005	1939+/-1149	10
11	T415G	94+/-50	0,05+/-	533+/-293	3

Aminoácido	Enzima	Km (μM)	kcat (s^{-1})	kcat/Km ($\text{M}^{-1}\text{s}^{-1}$)	kcat/Km (relativo)*
			0,006		
1	T415G	68+/-20	0,52+/-0,093	7,627+/-2,664	41
Onde os aminoácidos estão indicados como na Figura 2.					

EXEMPLO 11

O tRNA âmbar supressor mutante de levedura ($\text{ytrNA}^{\text{Phe}}_{\text{CUA}}$) foi constitutivamente expresso sob o controlo do promotor de Ipp. A cassette de expressão de $\text{ytrNA}^{\text{Phe}}_{\text{CUA}}$ foi inserida no plasmídeo repressor pREP4 para formar pREP4-ytrNA, conforme descrito nos Exemplos neste documento. O $\text{ytrNA}^{\text{Phe}}_{\text{CUA\_30U40G}}$ ($\text{ytrNA}^{\text{Phe}}_{\text{CUA\_UG}}$) supressor mutante de levedura foi construído a partir de $\text{ytrNA}^{\text{Phe}}_{\text{CUA}}$ pelo uso de um *kit* de mutagénese QUICK-CHANGE®. Dois oligonucleótidos complementares, designados como iniciador UG-f (5'-GAA CAC AGG ACC TCC ACA TTT AGA GTA TGG CGC TCT CCC-3') (SEQ ID NO: 16) para o iniciador dianteiro e iniciador UG-r (5'-GGG AGA GCG CCA TAC TCT AAA TGT GGA GGT CCT GTG TTC-3') (SEQ ID NO: 17) para o iniciador reverso foram sintetizados para conter a mutação específica na posição 30 ou posição 40 do tRNA supressor mutante de levedura. O plasmídeo resultante contendo o gene que codifica o $\text{ytrNA}^{\text{Phe}}_{\text{CUA\_UG}}$ é designado como pREP4-ytrNA_UG. A fim de construir o plasmídeo para transcrição *in vitro* de $\text{ytrNA}^{\text{Phe}}$, os genes $\text{ytrNA}^{\text{Phe}}_{\text{CUA}}$ e $\text{ytrNA}^{\text{Phe}}_{\text{CUA\_UG}}$ foram amplificados a partir do plasmídeo molde pREP4-ytrNA e pREP4-ytrNA_UG, respetivamente. No final da sequência de tRNA, um local *BstNI* foi inserido para produzir a transcrição exata de $\text{ytrNA}^{\text{Phe}}$. Uma sequência de promotor T7 foi adicionada para transcrição *in vitro* de $\text{ytrNA}^{\text{Phe}}$ por uma RNA T7 polimerase. Os seguintes iniciadores foram usados para o PCR: 5'-CTG GGT AAG CTT CGC TAA GGA TCT

GCC CTG GTG CGA ACT CTG-3' (SEQ ID NO: 18) (com locais de restrição *HindIII* e *BstNI*) e 5'-GAT TAC GGA TTC CTA ATA CGA CTC ACT ATA GCG GAC TTA GCT C-3' (SEQ ID NO: 19) (com local de restrição *EcoRI* e uma sequência de promotor T7). O DNA resultante foi introduzido nos locais *HindIII* e *EcoRI* de pUC18 para produzir pUC18-ytRNA^{Phe}_{CUA} e pUC18- ytRNA^{Phe}_{CUA_UG}.

A fim de facilitar a manipulação de DNA, um local de clivagem *BstNI* próximo da sequência do promotor T7 de pUC18- ytRNA^{Phe}_{CUA} foi removido para aumentar o tamanho do fragmento de DNA contendo o gene ytRNA^{Phe}_{CUA} de 180 bp para 500 bp após digestão de *BstNI*. Dois oligonucleótidos complementares, 5'-CGG AAG CAG AAA GTG TAA AGA GCG GGG TGC CTA ATG AGT G-3' (SEQ ID NO: 20) para o iniciador dianteiro e 5'-CAC TCA TTA GGC ACC CCG CTC TTT ACA CTT TAT GCT TCC G-3' (SEQ ID NO: 21) para o iniciador reverso foram sintetizados para conter a mutação específica.

EXEMPLO 12

DNA linearizado foi preparado por digestão de *BstNI* de pUC18-ytRNA^{Phe}_{CUA} e pUC18- ytRNA^{Phe}_{CUA_UG} conforme descrito anteriormente (Ver Sampson, Uhlenbeck, PNAS USA 85: 1033-1037 (1988)). Transcrição *in vitro* de moldes de DNA linearizados e purificação de transcritos foram realizadas conforme descrito anteriormente (Ver Nowak, et al., Ion Channels pt. B 293: 504-529 (1998)). A transcrição *in vitro* do DNA linearizado para produzir transcritos de tRNA de 76 bases foi realizada com o kit AMBION® T7-MEGASHORTSCRIPT®. Transcritos foram isolados com uma extração (PCI) de fenol: CHCl₃:álcool isoamílico 25:24:1. A camada orgânica foi re-extraída com água e uma extração (CI) CHCl₃:álcool isoamílico 24:1 foi realizada nas camadas aquosas. A camada

de água foi depois misturada com igual volume de isopropanol, precipitada *overnight* a -20°C , peletizada, seca e re-dissolvida em água. Nucleótidos não reagidos na solução de tRNA foram eliminados usando CHROMA SPIN-30® DEPC- H_2O (BD Bioscience®). Concentrações dos transcritos foram determinadas por absorvância UV a 260 nm.

A Aminoacilação de $\text{ytRNA}^{\text{Phe}}_{\text{GAA}}$ de tipo selvagem com Phe e Trp por variantes de yPheRS foi realizada conforme descrito anteriormente (Ver Sampson, Uhlenbeck, PNAS USA 85: 1033-1037 (1988)). Reações de aminoacilação foram realizadas no tampão contendo 30 mM HEPES (pH 7,45), 15 mM MgCl_2 , 4 mM DTT, 25 mM KCl, e 2 mM ATP a 30°C , em volumes de reação de 100 microlitros. tRNA total de levedura purificado foi usado no ensaio a uma concentração final de 4 mg/mL (concentração de $\text{ytRNA}^{\text{Phe}}_{\text{GAA}}$ aproximadamente 2,24 microM). Para aminoacilação de Phe, 13,3 microM [^3H]-Phe (5,3 Ci/mmol) e 80 nM de variantes de yPheRS foram usados; para aminoacilação de Trp, 3,3 microM [^3H]-Trp (30,0 Ci/mmol) e 160 nM de variantes de yPheRS foram usados. Aminoacilação de transcritos de $\text{ytRNA}^{\text{Phe}}$ foi realizada em volumes de reação de 100 microlitros em solução tampão contendo 100 mM potássio-HEPES (pH 7,4), 10 mM MgCl_2 , 1 mM DTT, 0,2 mM EDTA, 2 mM ATP, e 4 unidades/mL de pirofosfatase inorgânica de levedura a 37°C para eLysS. Para aminoacilação de Lys, 4 microM de transcrito de $\text{ytRNA}^{\text{Phe}}$, 1,1 microM [^3H]-Lys (91 Ci/mmol) e 80 nM eLysS foram usados. Os tRNAs foram hibridizados antes do uso por aquecimento até 85°C durante 4 minutos no tampão de hibridação (60 mM Tris, pH 7,8, 2 mM MgCl_2) seguidos de arrefecimento lento a temperatura ambiente. As reações foram iniciadas pela adição de enzima e alíquotas de 10 microlitros foram reprimidas colocando-os em discos de filtro Whatman embebidos com 5% TCA. Os

filtros foram lavados durante três períodos de 10 minutos em 5% TCA gelado, lavados com etanol 95% gelado, e contados através de métodos de cintilação líquida.

EXEMPLO 13

Construção do plasmídeo para incorporação *in vivo* de um aminoácido não-natural foi realizada numa estirpe auxotrófica dupla de Phe/Trp, AFW, e numa estirpe auxotrófica tripla de Phe/Trp/Lys, AFWK. As estirpes auxotróficas foram construídas a partir de uma estirpe auxotrófica de Phe, AF (K10, Hfr(Cavalli) *pheS13rel-1 tonA22 thi T2^R pheA18, trpB114*) (Ver Furter, Prot. Sci, 7:419-426 (1998)) por transdução por transposição mediada pelo fago P1. Um vetor pQE16 (QIAGEN®) foi escolhido como o plasmídeo de expressão, que codifica uma proteína marcadora dihidrofolato redutase murina (mDHFR) com gene marcador C-terminal hexa-histidina sob o controlo de um promotor e terminador t_0 do bacteriófago T5. Kit de mutagénese quick-change foi usado para colocar um codão âmbar (TAG) na 38^a posição do mDHFR com dois oligonucleótidos complementares (5'-CCG CTC AGG AAC GAG TAG AAG TAC TTC CAA AGA ATG-3' (SEQ ID NO: 11); 5'-CAT TCT TTG GAA GTA CTT CTA CTC GTT CCT GAG CGG-3') (SEQ ID NO: 12) para produzir pQE16am. Os genes mutantes de yPheRS T415G e T415A foram amplificados a partir de pQE32-T415G e pQE32-T415A e um promotor constitutivo *tac* com um sítio de ligação do repressor *lac* abolido foi adicionado a montante do codão de iniciação do gene. A cassette de expressão completa de T415G e T415A foram inseridas no local *PvuII* de pQE16am-T415G e pQE16am-T415A.

EXEMPLO 14

As estirpes bacterianas auxotróficas AF, AFW, e AFWK foram transformadas com plasmídeo pQE16am contendo variantes de yPheRS e vetores pREP4-ytRNA contendo variantes de ytRNA para investigar a incorporação de pBrF. As estirpes de expressão de *E. coli* foram cultivadas em meio mínimo M9 suplementado com glicose, tiamina, MgSO_4 , CaCl_2 , 20 aminoácidos (a 25 mg/L), antibióticos (35 microgramas/mL de kanamicina) e 100 microgramas/mL de ampicilina. Quando as células alcançaram uma OD_{600} de 0,8-1,0, foram sedimentadas por centrifugação, lavadas duas vezes com 0,9% NaCl frio, e transferidas para o meio de expressão suplementado com 17 aminoácidos (a 20 mg/L), 6 mM de pBrF (p-bromo-fenilalanina) ou plodoF(p-iodo-fenilalanina), e as concentrações indicadas de fenilalanina, triptofano e lisina. Expressão da proteína foi induzida pela adição de 1 mM IPTG. Depois da expressão de quatro horas, as células foram peletizadas por centrifugação, e a proteína foi purificada em virtude do marcador C-terminal de hexa-histidina através de uma coluna de centrifugação de níquel-NTA de acordo com as instruções do fabricante (QIAGEN®). Após a purificação, níveis de expressão de mDHFR foram determinados por absorvância UV a 280 nm.

EXEMPLO 15

Análise de LC-MS/MS de digestões tripticas de mDHFR foi realizada num espectrômetro de massa *ion trap* LCQ Finnigan com bomba HPLC e sonda ESI. mDHFR mutante purificada sob condições de desnaturação estava em tampão de eluição (8 M urea, 100 mM NaH_2PO_4 , 10 mM Tris, pH 4,5).). Para a digestão por tripsina, 10 microL da solução foram diluídos em 90 microL de 75 mM $(\text{NH}_4)_2\text{CO}_3$. Um microlitro de tripsina

modificada (0,2 microgramas/microlitro) foi adicionado. A amostra foi incubada a 37°C durante 2 a 6 horas. A reação de digestão foi interrompida pela adição de 12 microL de solução 5% TFA. Solução de péptido digerido foi submetida a dessalinização com coluna C18 Vydac Microspin (grupo Nest) e eluída com 50 microL de 80% de acetonitrilo e 20% de ácido fórmico 0,1% p/v. Solução de péptido digerido eluída da coluna Microspin foi seca, evaporada em acetonitrilo 10% e 90% da solução 0,1% TFA, e injetada no bomba HPLC. Péptidos foram separados por coluna Magic C18 (Michrom, 300 Å, 0,3 x 150 mm) e eluídos a uma taxa de fluxo de 30 microL/min usando um gradiente de 10 a 95% de solvente A (90% de acetonitrilo e 10% de solução de ácido acético 0,1 M) e solvente B (2% de acetonitrilo e 98% de solução de ácido acético 0,1 M) durante 30 minutos. O fluxo de eluente da coluna para a fonte de *electrospray* e cada sinal da digestão tríptica foi detetado. Sequenciação de massa em *tandem* foi realizada simultaneamente pela fragmentação do ião precursor com m/z correspondendo ao fragmento digerido por proteases, incluindo o resíduo na posição 38 de mDHFR mutante. Assim, os polipéptidos DHFR foram sintetizados numa célula hospedeira auxotrófica tripla com (a), (b) tRNA^{Phe}_{CUA} de levedura e PheRS (T415G) de levedura; (c) tRNA^{Phe}_{CUA_UG} de levedura e PheRS (T415G) de levedura; (d) tRNA^{Phe}_{CUA} de levedura e PheRS (T415A) de levedura; (e) tRNA^{Phe}_{CUA-UG} de levedura e PheRS (T415A) de levedura ou (f) numa estirpe auxotrófica simples com tRNA^{Phe}_{CUA_UG} de levedura e PheRS (T415A) de levedura, os resultados dos quais são apresentados na Figura 12.

EXEMPLO 17

Uma biblioteca de PheRS de levedura (usando a proteína verde fluorescente ou GFP) foi analisada para identificar mutações de PheRS que permitem a incorporação de Nal. Posições de sequência de aminoácidos específicos que foram mutadas incluem resíduos número 412, 415, 418, e 437, que estão localizados no sítio de ligação e contactam o aminoácido. Conforme indicado na Figura 13, p-etinil-fenilalanina foi incorporada numa proteína de teste usando a PheRS modificada de levedura.

Brevemente, GFP foi ligado a um vetor contendo o gene de PheRS mutante T415G ou de levedura de tipo selvagem de acordo com os procedimentos padrão. O tRNA âmbar supressor mutante de levedura ($\text{ytrNA}^{\text{Phe}}_{\text{AAA}}$) foi constitutivamente expresso sob o controlo do promotor *lpp* e transformado numa estirpe auxotrófica dupla de Phe/Trp AFW de *E. coli*. As células foram crescidas em meio mínimo M9, suplementado com glicose, tiamina, MgSO_4 , CaCl_2 , 20 aminoácidos (a 25 mg/L), antibióticos (35 $\mu\text{g/mL}$ de kanamicina e 100 $\mu\text{g/mL}$ de ampicilina). Quando as células alcançaram uma DO_{600} de 0,8-1,0, as células foram peletizadas e lavadas duas vezes por 0,9% NaCl gelado e mudado para meio de expressão suplementado com 18 aminoácidos (a 20 mg/L) e várias concentrações de fenilalanina, triptofano e 2Nal. Expressão da proteína foi induzida por IPTG.

EXEMPLO 18

Mutagéneses dos quatro resíduos de aminoácidos selecionados (N412 T145, S418, e S437) foram conduzidas por mutação de PCR a duas etapas. Brevemente, uma série de mutagéneses de

PCR foram realizadas no gene GFP_{UV} num plasmídeo pQE9_GFP_{UV} (STRATAGENE®), usando quatro pares de iniciadores complementares (F64LS65T_f: 5'-CTT GTC ACT ACT CTG ACC TAT GGT GTT CAA TGC TTC TCC CGT-3' (SEQ ID NO: 22); F64LS65T_r: 5'-ACG GGA GAA GCA TTG AAC ACC ATA GGT CAG AGT AGT GAC AAG-3' (SEQ ID NO: 23); S99F_f: 5'-GTA CAG GAA CGC ACT ATA TTC TTC AAA GAT GAC GGG AAC-3' (SEQ ID NO: 24); S99F_r : 5'-GTT CCC GTC ATC TTT GAA GAA TAT AGT GCG TTC CTG TAC-3' (SEQ ID NO: 25); T153M_f: 5'-CAC AAT GTA TAC ATC ATG GCA GAC AAA CAA AAG AAT GGA-3' (SEQ ID NO: 26); T153M_r: 5'-TCC ATT CTT TTG TTT GTC TGC CAT GAT GTA TAC ATT GTG -3' (SEQ ID NO: 27)).

Os mutantes de GFP foram gerados conforme descrito neste documento. Brevemente, um GFP3 tem 12 resíduos de Phe dos quais cinco são codificados por codões *wobble* de Phe (UUU). Uma variante de GFP5 e GFP6 foram preparados substituindo codões UUC com codões UUU usando reações de PCR a duas etapas seguidas de ligação. Um GFP5 foi preparado substituindo quatro codões UUC e um codão Leu nos resíduos F8, L64, F84, F99 e F165 com codões UUU usando doze iniciadores (1: 5'-GTG CCA CCT GAC GTC TAA GAA ACC ATT ATT ATC ATG ACA TTA ACC-3' (SEQ ID NO: 28) 2: 5'-GAG TAA AGG AGA AGA ACT TTT TAC TGG AGT TGT CCC AAT TC-3' (SEQ ID NO: 29) 3: 5'-GAA TTG GGA CAA CTC CAG TAA AAA GTT CTT CTC CTT TAC TC-3' (SEQ ID NO: 30) 4: 5'-GGC CAA CAC TTG TCA CTA CTT TTA CCT ATG GTG TTC AAT GCT T-3' (SEQ ID NO: 31) 5: 5'-AAG CAT TGA ACA CCA TAG GTA AAA GTA GTG ACA AGT GTT GGC C-3' (SEQ ID NO: 32) 6: 5'-CAT ATG AAA CGG CAT GAC TTT TTT AAG AGT GCC ATG CCC GAA G-3' (SEQ ID NO: 33) 7: 5'-CTT CGG GCA TGG CAC TCT TAA AAA AGT CAT GCC GTT TCA TAT G (SEQ ID NO: 34) 8: 5'-GTT ATG TAC AGG AAC GCA CTA TAT TTT TCA AAG ATG ACG GGA ACT ACA A-3' (SEQ ID NO: 35) 9: 5'-TTG TAG TTC

CCG TCA TCT TTG AAA AAT ATA GTG CGT TCC TGT ACA TAA C-3'
 (SEQ ID NO: 36) 10: 5'- ACA AAA GAA TGG AAT CAA AGC TAA CTT
 TAA AAT TCG CCA CAA CAT TGA AGA TG-3' (SEQ ID NO: 37) 11:
 5'- CAT CTT CAA TGT TGT GGC GAA TTT TAA AGT TAG CTT TGA TTC
 CAT TCT TTT GT-3' (SEQ ID NO: 38); 12: 5'- CGC CAA GCT AGC
 TTG GAT TCT CAC CAA TAA AAA ACG CCC-3' (SEQ ID NO: 39).
 Cinco fragmentos parcialmente coincidentes de cassetes de
 expressão de GFP3 foram obtidos por cinco reações de PCR
 com cinco conjuntos de primers (1 e 3; 2 e 5; 4 e 7; 6 e 9;
 8 e 10).

Estes produtos de PCR foram purificados por eletroforese em
 gel de agarose seguido de extração de gel. Um GFP6 do qual
 todos os resíduos Phe são codificados por UUU foi preparado
 substituindo dois codões Phe (F71 e F99) de GFP5 com codões
 UUU usando seis iniciadores. (1 e 12 são os mesmos de
 acima; 13: 5'- TAC CTA TGG TGT TCA ATG CTT TTC CCG TTA TCC
 GGA TCA TAT G-3' (SEQ ID NO: 40); 14: 5'-CAT ATG ATC CGG
 ATA ACG GGA AAA GCA TTG AAC ACC ATA GGT A-3' (SEQ ID NO:
 41); 15: 5'- GTT ATG TAC AGG AAC GCA CTA TAT TTT TTA AAG
 ATG ACG GGA ACT ACA AG-3' (SEQ ID NO: 42); 16: 5'- CTT GTA
 GTT CCC GTC ATC TTT AAA AAA TAT AGT GCG TTC CTG TAC ATA AC-
 3' (SEQ ID NO: 43)).

EXEMPLO 19

Construção da biblioteca foi também realizada em duas
 etapas. Brevemente, mutagênese de saturação em quatro
 resíduos (N412, T415, S418 e S437) foi alcançada com
 mutagênese de PCR a duas etapas. Em primeiro lugar, codões
 degenerados foram introduzidos em S437 por mutagênese de
 PCR com dois iniciadores complementares (437_f: 5'-GTC GAA
 ATC GGT AAC NNK GGT ATG TTC AGA CCA GAA ATG CTC G-3' (SEQ

ID NO: 44); 437_r: 5'- C GAG CAT TTC TGG TCT GAA CAT ACC MNN GTT AC C GAT TTC GAC-3' (SEQ ID NO: 45)). Após 1 hr de digestão do produto de PCR com *DpnI*, o produto de PCR foi transformado em hospedeiro de clonagem XL-1 azul. Os plasmídeos da 437^a posição foram saturados e isolados e usados como um molde para a 2^a mutagénese de PCR para introduzir mutação nos resíduos N412, T415 e S418. A 2^a mutagénese de PCR foi realizada com outro par de iniciadores complementares (412_418_f: 5'-C AAG CCT ACC TAC NNK CCT TAC NNK GAG CCA NNK ATG GAA ATC TTT T-3' (SEQ ID NO: 46); 412_418_r: 5'-A AAA GAT TTC CAT MNN TGG CTC MNN GTA AGG MN N GTA GG T AGG CTT G-3' (SEQ ID NO: 47)). Após o PCR, os produtos foram digeridos com *DpnI* durante 1 hora, foram limpos e concentrados por coluna de centrifugação. Eluato foi eletroporado em células eletrocompetentes ElectroTen-Blue (Stratagene) de acordo com o protocolo do fabricante. 8 milhões de transformantes foram obtidos. O plasmídeo de biblioteca foi expandido em cultura e digerido com *NsiI* e *BglII*. Após a purificação destes insertos, estes foram ligados com fragmentos longos de pQE9_GFP6_yPheRS (T415G) e pQE9_GFP9_yPheRS (T415G) obtidos por digestão com *NsiI* e *BglII*.

A biblioteca foi transformada em células AFW e DHF químico-competentes de *E. coli*. Estas células foram depois inoculadas em meio 2x YT com kanamicina e crescidas *overnight*. Quando as células alcançaram uma DO₆₀₀ de 0,8, as células foram peletizadas e ressuspensas em água destilada. Stocks de glicerol da biblioteca foram expressos como é padrão na arte.

Após a expressão de GFP durante 3 horas, 1 mL de células (baseadas em DO₆₀₀ de 1,0) foram lavadas com PBS e diluídas

em água destilada, de seguida submetidas a análise de citometria de fluxo (MoFlo cell sorter®, DakoCytomation, Fort Collins, CO), usando um comprimento de onda de excitação de 488 nm, emissão de 525 nm e um filtro limite de 495 nm. Pelo menos 20000 eventos foram recolhidos em cada medição. Os dados foram analisados com o *software* Summit (DakoCytomation). Triagem de biblioteca foi feita tanto positiva como negativamente, isto é, as variantes de yPheRS que permitem a elevada incorporação de 2Nal ou quaisquer outros aminoácidos naturais exceto Phe em codões UUU vai desenrolar GFP e são menos brilhantes, e portanto células de fluorescência baixa são coletadas. As variantes de yPheRS que não permitem a incorporação de quaisquer outros aminoácidos naturais exceto Phe em codões UUU não afetará o enrolamento de GFP e são brilhantes. Portanto, células brilhantes são coletadas. A Figura 14 ilustra histogramas de triagem de biblioteca de yPheRS GFP.

LISTAGEM DE SEQUÊNCIAS

<110> California Institute of Technology

Wang, Pin

Kwon, Inchan

Son, Soojin

Tang, Yi

Tirrell, David A.

<120> INCORPORAÇÃO SÍTIO-ESPECÍFICA DE AMINOÁCIDOS EM MOLÉCULAS

<130> 110197 412PC

<140> PCT

<141> 2007-03-05

<150> US 60/779,375

<151> 2006-03-03

<150> US 60/779,376

<151> 2006-03-03

<160> 104

<170> FastSEQ para Windows Versão 4.0

<210> 1

<211> 7790

<212> DNA

<213> Sequência Artificial

<220>

<223> Rato modificado gene de dihidrofolato redutase
modificado de ratinho e gene MetRS de E. coli

<400> 1

```
ctcgagaaat cataaaaaat ttatttgctt tgtgagcgga taacaattat aatagattca 60
attgtgagcg gataacaatt tcacacagaa ttcattaaag aggagaaatt aactatgaga 120
ggatcgcatc accatcacca tcacggatcc ggcatcatgg ttcgaccatt gaactcgatc 180
gtcgccgtgt cccaaaaat ggggattggc aagaacggag acctaccctg gcctccgctc 240
aggaacgagt tcaagtactt ccaaagaatg accacaacct cttcagtggg aggtaaacag 300
aatctggtga ttatgggtag gaaaacctgg ttctccatcc ctgagaagaa tcgaccttta 360
aaggacagaa ttaatatagt tctcagtaga gaactcaaag aaccaccacg aggagctcat 420
tttcttgcca aaagtttgga tgatgcctta agacttattg aacaaccgga attggcaagt 480
aaagtagaca tggtttggtt agtcggaggg agttctgttt accaggaagc catgaatcaa 540
ccaggccacc ttagactctt tgtgacaagg atcatgcagg aatttgaaag tgacacgttt 600
ttcccagaaa ttgatttggg gaaatataaa cttctcccag aatacccagg cgtcctctct 660
gaggtccagg aggaaaaagg catcaagtat aagtttgaag tctacgagaa gaaaggttgg 720
aagatcttaa gcttaattag ctgagcttgg actcctgttg atagatccag taatgacctc 780
agaactccat ctggatttgt tcagaacgct cggttgccgc cgggcgtttt ttattggtga 840
gaatccaagc tagctctaga gacgtccggc cggagctcca ccgcggtggc ggccgctcta 900
gagtcactta cttaacattt tcccatttgg tactatctaa ccccttttca ctattaagaa 960
gtaatgccta ctatgactca agtcgcgaag aaaattctgg tgacgtgcgc actgccgtac 1020
gctaacggct caatccacct cggccatatg ctggagcaca tccaggctga tgtctgggtc 1080
cgttaccagc gaatgcggcg ccacgaggtc aacttcattc gcgcgacga tgcceacggt 1140
acaccgatca tgctgaaagc tcagcagctt ggtatcacc cggagcagat gattggcgaa 1200
```

```

atgagtcagg agcatcagac tgatttcgca ggctttaaca tcagctatga caactatcac 1260
tcgacgcaca gcgaagagaa ccgccagtlg tcagaactta tctactctcg cctgaaagaa 1320
aacgggttta ttaaaaaccg caccatctct cagctgtacg atccggaaaa aggcattgtc 1380
ctgccggacc gttttgtgaa aggcacctgc ccgaaatgta aatccccgga tcaatacggc 1440
gataactgcy aagtctgcgg ccgcacctac agcccgactg aactgatega gccgaaatcg 1500
gtgggtttcty gcgctacgcc ggtaatgcgt gattctgaac acttcttctt tgatctgccc 1560
tctttcagcy aatgttgca ggcattggacc ccgagcggtg cgttcagga gcagggtggca 1620
aataaaatgc aggagtggtt tgaatctggc ctgcaacagt gggatatctc ccgcgacgcc 1680
ccttacttcy gttttgaaat tccgaacgcg ccgggcaaat atttctacgt ctggctggac 1740
gcaccgattg gctacatggg ttctttcaag aatctgtgcg acaagcgcyg cgacagcgta 1800
agcttcgatg aatactggaa gaaagactcc accgccgagc tgtaccactt catcggtaaa 1860
gatattgttt acttccacag cctgttctgg cctgccatgc tggaaaggag caacttccgc 1920
aagccggtcca acctgtttgt tcatggctat gtgacggtga accgcgcaaa gatgtccaa 1980
tctccgggca cctttattaa agccagcacc tggctgaatc attttgacgc agacagccty 2040
cgtaactact acactgcgaa actctcttcg cgcattgatg atatcgatct caacctggaa 2100
gatttcgctc agcgtgtgaa tgcgcatatc gtttaacaaag tggttaacct ggctcccg 2160
aatccgggct ttatcaacaa cgtttttgac ggcgtgctgy caagcgaaat ggctgacctg 2220
cagttgtaca aaaccttcac tgatgccgct gaagtgattg gtgaagcgtg ggaagccgt 2280
gaatttggtg aagccgtgcg cgaatcatg gcgctggctg atctggctaa ccgctatgtc 2340
gatgaacagc ctccgtgggt ggtggcgaaa caggaaaggcc gcgatccga cctgcaggca 2400
atltgctcaa tgggcatcaa cctgttccgc gtgctgatga cttacctgaa gccggtaactg 2460
ccgaaactga ccgagcgtgc agaagcattc ctcaatacgg aactgacctg ggatggtatc 2520
cagcaaccgc tgcggggcca caaagtgaat ccgttcaagg cgtgtataaa ccgcatcgat 2580
atgaggcagg ttgaagcact ggtggaagcc tctaaatgag aagtataaagc cgctgccgcy 2640
ccgtaactg gcccgctggc agatgatccg attcaggaaa ccatcacctt tgacgacttc 2700
gctaaggtt acctgcgcgt ggcgtgatg gaaaacgcag agtttgttga aggttctgac 2760
aaactgtgc gccgacgct ggatctcgcc ggtgaaaaac gcaatgtctt ctccggtatt 2820
cgtttctgct acccgatcc gcaggcactg attggtcgtc acaccattat ggtgggctaac 2880
ctggcaccac gtaaaatcgc cttcggtatc tctgaaggca tgggtgatggc tgcgcgtcct 2940
ggcggggaaag atattttctt gctaagcccg gatgccggtg ctaaacccgg tcatacaggtg 3000
aaataactcc cttcaaggc gctgcacga cagccttttg ctttataaat tcttaaaagt 3060
gttttcttgc gatttgtct ctcttaacc cgcataaata ctggtagcat ctgcattcaa 3120
ctggataaaa ttacagggat gcagaatgag acactttatc tatcaggagc aaaaatcaca 3180
taaattcagg gcagttgagc aacagggaaa cgagttgcac atcagttggg gaaaagttgg 3240
caccaaagcc aaagccagat aaaaagtttt tcagatgctg ccgagcggc aaaagcggag 3300
cccgacctcg agggggggcc cggtaaccgg ccggacgtct ctagagctag cttggcgaga 3360
ttttcaggag ctaagggaagc taaaaaggag aaaaaaatca ctggatatac caccgttgat 3420
atatcccaat ggcacgttaa agaactttt gaggcatttc agtcagttgc tcaatgtacc 3480
tataaccaga ccgttcagct ggatattacg gcctttttaa agaccgtaaa gaaaaataag 3540
cacaagtttt atccggcctt tatccacatt cttgcccgcc tgatgaatgc tcatccggaa 3600
tttcgtatcg caatgaaaga cggtgagctg gtgatatggg atagtgttca cctttgttac 3660
accgttttcc atgagcaaac tgaaacgtt tcatcgctct ggagtgaata ccacyacgat 3720
ttccggcagt ttctacacat atattcgcaa gatgtggcgt gttacggtga aaacctggcc 3780
tattcccta aagggtttat tgagaatatg ttttctgct cagccaatcc ctgggtgagt 3840
ttcaccagt ttgatttaaa cgtggccaat atggacaact tcttcgcccc cgttttccac 3900
atgggcaaat attatacgca aggcgacaag gtgctgatgc cgtggcgat tcagggttcat 3960
catgccgttt gtgatggctt ccattgtggc agaattgctta atgaattaca acagtaactg 4020
gatgagtgcc agggcggggc gtaatttttt taaggcagtt attggtgccc ttaaacgcct 4080
ggggtaatga ctctctagct tgaggcataa aataaaacga aaggctcagt cgaaagactg 4140
ggcctttcgt tttatctggt gtttgcggt gaacgctctc ctgagtagga caaatccgcc 4200
ctctagatta cgtgcagtcg atgataagct gtcaaactg agaattgtgc ctaatgagt 4260
agctaactta cattaattgc gttgcgctca ctgcccgct tccagtcggg aaacctgtcg 4320
tgccagctgc attaatgaat ccggccaacg gcggggagag gcggtttgcy tattgggcgc 4380
cagggtggtt tttcttttca ccagtgagac gggcaacagc tgattgcctt tcaccgctg 4440
gccctgagag agttgcagca agcggteccac gctgggtttg ccagcaggg gaaaatcctg 4500
tttgatggtg gttaacggcg ggatataaca tgagctgctc tcgggtatcgt cgtatccac 4560
taccgagata tccgcaccaa ccgcgagccc ggactcggtg atggcgcgca ttgcgcccag 4620
cgccatctga tcyttggcaa ccagcatcgc agtgggaacg atgcccctcat tcagcatttg 4680

```



```

catggtttgt tgaaaaccgg acatggcact ccagtcgcct tcccgttccg ctatcggctg 4740
aattttgattg cgagtggagat atttatgcca gccagccaga cgcagacgcg ccgagacaga 4800
acttaattggg cccgctaaca gcgcgatttg ctggtgaccc aatgcgacca gatgctccac 4860
gcccagtcgc gtaccgtcctt catgggagaa aataaactg ttgatgggtg tctggtcaga 4920
gacatcaaga aataacgcgg gaacattagt gcaggcagct tccacagcaa tggcatcctg 4980
gtcatccagc ggatagttaa tgatcagccc actgacgcgt tgcgcgagaa gatgtgtcac 5040
cgccgcttta caggcttcga cgcgcgttcg ttctaccatc gacaccacca cgtcggcacc 5100
cagttgatcg gcgcgagatt taatcgccgc gacaatttgc gacggcgcgt gcagggccag 5160
actggagggtg gcaacgcaa tcagcaacga ctgtttgccc gccagttgtt gtgccacgcg 5220
gttgggaatg taattcagct ccgccatcgc cgtttccact ttttccgcg ttttcgcaga 5280
aacgtggctg gcctggttca ccacgcggga aacggtctga taagagacac cgcatactc 5340
tgccagatcg tataacgcta ctggtttcac attcaccacc ctgaattgac tctcttccg 5400
gcgctatcat gccataccgc gaaagggttt gcaccattcg atggtgtcgg aatttcgggc 5460
agcgttgggt cctggccacg ggtgcgcgtg atctagagct gcctcgcgcg tttcgggtgat 5520
gacggtgaaa acctctgaca catgcagctc ccggagacgg tcacagcttg tctgtaagcg 5580
gatgccggga gcagacaagc ccgtcagggc gcgtcagcgg gtgttgscgg gtgtcggggc 5640
gcagccatga cccagtcacg tagcgatagc ggagtgtata ctggcttaac tatcggcat 5700
cagagcagat tgtactgaga gtgcaccata tgcggtgtga aataccgcac agatgcgtaa 5760
ggagaaaata ccgcatcagg cgtctcttcg ctctctcgct cactgactcg ctgcgctcgg 5820
tcgttccggt gcggcgagcg gttacagctc actcaaaggc ggtaatatcg ttatccacag 5880
aatcaggggg taacgcaggg aagaacatgt gagcaaaagg ccagcaaaag gccaggaacc 5940
gtaaaaaggc cgcgttgctg gcgtttttcc ataggctccg cccctctgac gagcatcaca 6000
aaaatcgacg ctcaagtcag aggtggcgaa acccgacagg actataaaga taccaggcgt 6060
ttccccctgg aagctccctc gtgcgctctc ctgttccgac cctgcccgtt accggatacc 6120
tgtccgctt tctcccttcg ggaagcgtgg cgctttctca tagctcacgc tgtaggatc 6180
tcagttcgggt gtaggtcggt cgctccaagc tgggctgtgt gcacgaacc cccgttcagc 6240
ccgaccgctg cgccttatcc ggttaactatc gtcttgagtc caaccggta agacacgact 6300
tatcgccact ggagcagacc actggttaaca ggattagcag agcgagggtat gtaggcgtg 6360
ctacagagtt cttgaagtgg tggcctaact acggctacac tagaaggaca gtatttgta 6420
tctgcgctct gctgaagcca gttaccttcg gaaaaagagt tggtagctct tgatccggca 6480
aacaaccac ccgtcggtagc cgtgtgtttt ttgtttgcaa gcagcagatt acgcgcagaa 6540
aaaaaggatc tcaagaagat cctttgatct ttctacggg gtctgacgct cagtggaaag 6600
aaaactcacg ttaagggatt ttggtcatga gattatcaaa aaggatcttc acctagatcc 6660
ttttaaatca aaaatgaagt tttaaatcaa tctaaagtat atatgagtaa acttggctcg 6720
acagttacca atgcttaatc agtgaggcac ctatctcagc gatctgtcta tttcgttcat 6780
ccatagttgc ctgactcccc gtctgttaga taactacgat acgggagggg ttaccatctg 6840
gccccagtcg tgcaatgata ccgcgagacc cacgctcacc ggctccagat ctatcagcaa 6900
taaacagacc agccgggaagg gccgagcgca gaagtggtec tgcaacttta tccgcctcca 6960
tccagttctat taattgttgc cgggaagcta gagtaagtag ttccgccagt aatagtttgc 7020
gcaacgtttgt tgccattgct acaggcatcg tgggtgcacg ctctgctgtt ggtatggctt 7080
cattcagctc cgtttcccaa cgatcaaggc gagttacatg atcccccatg ttgtgcaaaa 7140
aagcgggttag ctcttccggt cctccgatcg ttgtcagaag taagttggcc gcagtgttat 7200
cactcatggt tatggcagca ctgcataatt ctcttactgt catgccatcc gtaagatgct 7260
tttctgtgac tgggtgagtac tcaaccaagt cattctgaga atagtgtatg cggcgaccga 7320
gttctcttgg ccggcgctca ataccgcgcc acatagcaga actttaaaag 7380
tgctcatcat tggaaaacgt tcttcggggc gaaaactctc aaggatctta ccgctgttga 7440
gatccagttc gatgtaaccc actcgtgcac ccaactgatc ttcagcatct ttacttttca 7500
ccagcgtttc tgggtgagca aaaaacaggaa ggcaaaatgc cgcaaaaaag ggaataaggg 7560
cgacacggaa atgttgaata ctcatactct tcctttttca atattattga agcattttac 7620
agggttattg tctcatgagc ggatacatat ttgaatgtat ttgaaaaaat aaacaaatag 7680
gggttccgcg cacatttccc cgaaaagtgc cacctgacgt ctaagaaacc attattatca 7740
tgacattaac ctataaaaat aggcgtatca cgaggccctt tctcttccac 7790

```

<210> 2

<211> 8294

<212> DNA

<213> Gene de fenilalanina aminoacil sintetase de levedura

<400> 2

```

ctcgagaaat cataaaaaat ttatttgcct tgtgagcgga taacaattat aatagattca 60
attgtgagcg gataacaatt tcacacagaa ttcatataag aggagaaatt aactatgaga 120
ggatcgcatc accatcacca tcacggatcc gtcgacctgc agccccgggt accggtagaa 180
aaaatgagta aaggagaaga actttttact ggagtgtgcc caattcttgt tgaattagat 240
ggtgatgtta atgggcacaa attttctgtc agtggagagg gtgaagggtga tgcacatac 300
ggaaaaactta cctttaaatt tatttgcact actggaaaac tacctgttcc atggccaaca 360
cttgtcacta cttttaccta tgggtttcaa tgettttccc gttatccgga tcatatgaaa 420
cggcatgact tttttaagag tgccatgccc gaaggttatg tacaggaaacg cactatattt 480
tttaaagatg acgggaacta caagacgcgt gctgaagtea agtttgaagg tgataccctt 540
gttaatcgta tcgagttaaa aggtattgat tttaaagaag atggaaacat tctcggcac 600
aaactcgagt acaactataa ctacacacat gtatacatca tggcagacaa acaaaagaat 660
ggaatcaaag ctaactttaa aattcgccac aacattgaag atggatccgt tcaactagca 720
gaccattatc aacaaaatac tccaattggc gatggccctg tccctttacc agacaacat 780
tacctgtcga cacaatctgc cctttcgaaa gatcccaacg aaaagcgtga ccacatggtc 840
cttcttgagt ttgttaactgc tgctgggatt acacatggca tggatgagct ctacaaactg 900
cagccaagct taattagctg aagcttggac tctgttgtat agatccagta atgacctcag 960
aactecatct ggatttgttc agaacgcctg gttgccgcg ggcgtttttt attggtgaga 1020
atccaagcta gcttggcgag attttcagga gctaaggaa gtaaaatgga gaaaaaatc 1080
actggatata ccacggttga tatatcccaa tggcatcgta aagaacattt tgaggcattt 1140
cagtcaagtg ctcaatgtac ctataaccag accgttcagc tggcacgaca ggtttcccg 1200
ctggaaaagc ggacgtgagc gcaacgcaat taatgtgagt tagctcactc attaggcacc 1260
ccaggcttta cactttatgc ttcgggtcgt tatgttgtgt ggaattgtga gcggaataca 1320
atttcacaca ggaaacagct atgaccatga ttacgccaag cttgcattgc tgcagttgac 1380
aattaatcat cggctcgat atgtgaccca attgtgagcg gaatcgattt tcacacagga 1440
aacagaccat gaatctagag atgtctgact tccaattaga aattctaaag aaactagatg 1500
aattggatga gatcaagtc acactggcaa ctttccctca gcaaggctct caagatgttc 1560
tttccgcttt gaactctttg aaagcccaca acaagttaga gttttccaa gtcgacacg 1620
ttacgtatga cttgaccaa gaaggtgtc aaattttgaa tgaaggttcg tacgaaatta 1680
aactcttcaa gctcatccaa gagttgggtc aacttcaaat caaagatgtg atgtccaaac 1740
taggccctca agttggttaag gtcggtcagg cttagagctt caagaacgac tggatcgcca 1800
aaaacgcctc aaacgagctt gaactctcgg caaaattgca aataccgat ttaaatgagc 1860
ttactgatga aacgcaatct attctagcgc aaatcaagaa caactcgcat ctggatagca 1920
ttgacgccaa gattttgaac gacttgaaga aaagaaagtt aattgtctca ggtaaaatca 1980
cagattttag tctaccaa tggccagagt tctcgaccga cctcaccaaa ttggaaaccg 2040
atcttacctc cgacatggtc tccaccaatg catacaagga cttgaagttc aagccttaca 2100
atttcaattc tcaaggtgtg caaatatctt caggtgtctc tcacccctta aacaaagtca 2160
gagagggaatt tagacaaatt ttcttttcca tgggattcac agagatgcc tcgaaccaat 2220
acgtcgagac aggtttcttg aacttcgatg cctttactgt cccacaacag catcctgtc 2280
gtgacctgca agacactttc tacatcaagg accactaac cgtcgagttg cccgatgaca 2340
agacatacat ggacaatatc aaagccgttc acgaacaggg gagattcggg tccatcggtt 2400
atcgttacaa ctggaagcca gaagaatgtc aaaaattggt cttgagaact cactccacag 2460
ccatctctgc cagaatgtg cactgatttg ccaaagatcc aaagcccacc agattgtttt 2520
ctatcgaccg tgttttccgt aacgaagcag ttgacgccac ccatttggcc gaattccacc 2580
agggtggaag tgtctctgcc gactacaaca ttactctggg tgacctgatc aagttcatgg 2640
aagagttttt cgaaagaatg ggtgtcaccc gtttgagatt caagcctacc tacaatectt 2700
acaccgagcc atcaatggaa atcttttctt ggcacgaagg tttgcaaaaa tgggtcgaaa 2760
tcggtaactc tgggtatgtt agaccagaaa tgctcgagtc catgggtcta ccaaaggatc 2820
taagagtcct tgggtggggg ttatccttgg aaagacctac catgatcaaa tataaggttc 2880
aaaacatcag agaactgtta ggtcataaag tctctttgga ctttatcgaa accaatcctg 2940
ctgctagatt ggacgaagac ttgtacgaat aaggcaggaa tagattatgc ctaccgtctc 3000
cgtgaacaag cagcaattat ttgatcttct aggcgaagac tacacttccc aagagttcga 3060
cgaattatgt tttgaattcg gtatggaaat ggacgaagac accacagaag aggccttgaa 3120
aaccggggag gagccggaat tgaagcttga tatcagtgcc aatcgttacg atttgctttg 3180
tatcgaaggt atttcacagt cgttgaacga atacttgga cgtaaagaaa gacctgacta 3240
taaattaagc aagccaacca ctaagttgat catcgacaaa tcaacggagc aaattagacc 3300
ttttgctacc gctgctgtat tgagaaatat caagcttaac gaaaaatctt acgcttcttt 3360
tattgccttg caagataaat tacatgccaa tctatgtaga aacagaagct tggttgccat 3420

```

```

gggtactcac gatttagatt caattgaagg tccattccat tacagagctc taccacaaaa 3480
ggacatcaag ttcgtaccat tgaatcaaac ccaagagttt actggtgaca aattgatacga 3540
gttttataaa totccagaac agaaaaacaa catagggaga tacgttcaca ttattgagga 3600
ttctccagtc ttcccagtta ttatggacag caaagatcgt gtttgcctcc tgccaccatt 3660
aatcaatagt gaaccattca agatctctgt gaacacccgt aacattttga ttgatataac 3720
cgccaccgat aagaccaaag ccgagatcgt ttgaaacata ttaactacaa tgttctcacg 3780
ttattgtgac gaaccattca cggttgagcc tgtagaaatt gtctctgaac acaatggcca 3840
atccccgttg gcgccaaact tcaacgatag aattatggat gtctccatta agtacatcaa 3900
ctcctgtctt ggcctagatc aatccgctga tgaattgct cattgtctaa aaaagatgto 3960
gttgcattgc gttcaatcaa aggaagacaa ggacatcttg cacgttgaca ttccggtaac 4020
tagacctgat attttgcacg cttgtgatat aatggaagat gccgctgtcg gttatggttt 4080
caataatttg ccaaaggggt agaaattatc caatgccaac ttcatggcca aaccattacc 4140
aatcaacaag gtttctgata ttctcagagt tgcattctct caagccacgt gggttgaggt 4200
tttaccattg accttatgtt cgcacgatga aaacttttaa tttctaagac aatccgacaa 4260
tgatgattta gctgtcaaat tggccaaccc aaagactttg gaataccaag ttgttagaac 4320
cactttattg cctggatatc taaagacagt caaggaaaac agaaaacatt ccttaccata 4380
caaatgtctt gaaacgggtg acgttctatt taaagacgac aaactagaaa ggaaggcgta 4440
caatgaacgt cactgggctg ccatctacgt gggtaaagat tctgggtttg aaatcattca 4500
agggttattg ggtaaaatca tgcacacttt tagaacagag tggattgacg actacgggtg 4560
tgctgtcttc ggcatagggt actggattga agaagacgat tctgtgaaaa cctatttccc 4620
aggtagaggt gccaaaggtc tgttcagatc caaagaaggc gctgagccaa agcaaatcgg 4680
ccacttgggt gtcttgcatc ctgaagtcac gatgaatttc gacgttccat tccgtgcac 4740
ctttgtagag gttaatgccg aagtcttctc ataattgaat gttctaacia aaatttttac 4800
tgatttataa aacttatata gatagataga catatatata tatctatata tagaaacaca 4860
actaaagttt acctgtttt atatagggta ccgagctcga attcactggc cgtcgtttta 4920
caacgtcgtg actgggaaaa ccgcggttac gtgcagtcga tgataagctg tcaaacatga 4980
gaattgtgcc taatgagtga gctaacttac attaatggc ttgcgctcac tgcccgctt 5040
ccagtcggga aacctgtcgt gccagctgca ttaatgaatc ggccaacgag cggggagagg 5100
cgggttgcgt attggggcgc aggggtggtt ttcttttcac cagtgcagac ggcaacagct 5160
gatttgcctt caccgcctgg cctgagaga gtgcagcaa gcggtccacg ctggtttgc 5220
ccagcaggcg aaaatcctgt ttgatgttgg ttaacggcgg gatataacat gactgtctt 5280
cgggtatcgt gtatccact accgagatat ccgcaccaac gcgcagccc gactcggtaa 5340
tggcgcgcat tgcgcccagc gccatctgat cgttggaac cagcatcgca gtgggaacga 5400
tgccctcatt cagcatttgc atggtttgtt gaaaaccgga catggcactc cagtcgctt 5460
cccgttccgc tatcggtga atttgattgc gagtgcagata ttatgccag ccagccagac 5520
gcagacgcgc cgagacagaa cttaattggc ccgctaacag cgcgatttgc tggtgacca 5580
atgcgaccag atgctccacg cccagtcgag taccgtcttc atgggagaaa ataactgt 5640
tgatgggtgt ctggtcagag acatcaagaa ataacgcccg aacattagtg caggcagctt 5700
ccacagcaat ggcatcctgg tcatccagcg gatagttaat gatcagccca ctgacgcgtt 5760
gcgcgagaag attgtgcacc gccgctttac aggtctcgac gccgcttctg tctaccatcg 5820
acaccaccac gctggcacc agttgatcgg cgcgagattt aatcgccgag acaatttgcg 5880
acggcgcggt cagggccaga ctggaggtgg caacgccaat cagcaacgac tgtttgccg 5940
ccagctgttg tgcacgcgg ttgggaatgt aattcagctc cgccatcgcc gcttccactt 6000
tttcccgctt tttcgcagaa acgtggctgg cctggttcac cagcggggaa acggtctgat 6060
aagagacacc ggcatactct gcgacatcgt ataacyttac tggtttcaca ttcaccaccc 6120
tgaattgact ctcttccggg cgtatcatg ccataccgag aaaggttttg caccattcga 6180
tggtgtcgga atttcgggca gcgttgggtc ctggccacgg gtgcgcatga cttaaagcgt 6240
gtgaaatacc gcacagatgc gtaaggagaa aataccgcat caggcgctct tccgcttctt 6300
cgctcactga ctgcctgcgc tgggtctgtc ggctgcggcg agcgggtatc gctcactcaa 6360
aggcggtaat acggttatcc acagaatcag gggataacgc aggaagaac atgtgagcaa 6420
aaggccagca aaaggccagg aaccgtaaaa aggcgcggtt gctggcgctt ttccataggc 6480
tccgcccccc tgacgagcat cacaaaaatc gacgctcaag tcagaggtgg cgaaacccga 6540
caggactata aagataccag gcgtttcccc ctggaagctc cctcgtgcgc tctcctgttc 6600
cgaccctgcc gcttaccgga tactgtccg cctttctccc ttcgggaagc gtggcgcttt 6660
ctcaatgctc acgctgtagg tatctcagtt cgggtgaggt cgttcgctcc aagctgggct 6720
gtgtgcacga accccccgtt cagcccgacc gctgcgctt atccggtaac tatcgtctt 6780
agtccaaccc ggtaagacac gacttatcgc cactggcagc agccactggg aacaggatta 6840
gcagagcgag gtatgtaggc ggtgctacag agttcttgaa gtggtggcct aactacggct 6900

```

```

acactagaag gacagtattt ggtatctgcg ctctgctgaa gccagttacc ttcggaaaaa 6960
gagttggtag ctcttgatcc ggcaaaacaaa ccaccgctgg tagcgggtgg ttttttgttt 7020
gcaagcagca gattacgcgc agaaaaaaag gatctcaaga agatcctttg atcttttcta 7080
cggggctctga cgctcagtg aacgaaaact cacgttaagg gatttttggtc atgagattat 7140
caaaaaggat cttcacctag atccttttaa attaaaaatg aagttttaaa tcaatctaaa 7200
gtatatatga gtaaaccttg tctgacagtt accaatgctt aatcagtgag gcacctatct 7260
cagcgatctg tctatttcgt tcatccatag ctgectgact ccccgctcgt tagataacta 7320
cgatacggga gggcttacca tctggcccca gtgctgcaat gataccgca gaccacgct 7380
caccggctcc agatttatca gcaataaacc agccagccgg aagggccgag cgcagaagtg 7440
gtcctgcaac tttatccgcc tccatccagt ctattaattg ttgccgggaa gctagagtaa 7500
gtagttcgcc agttaatagt ttgcgcaacg ttgttgccat tgctacaggc atcgtggtgt 7560
cacgctcgtc gtttggtatg gcttcattca gctccggttc ccaacgatca agcgagtta 7620
catgatcccc catgttggtg aaaaaagcgg ttagctcett cggtcctccg atcgttgta 7680
gaagtaagtt ggccgcagtg ttatcactca tggttatggc agcactgcat aattctctta 7740
ctgtcatgcc atccgtaaga tgcttttctg tgactggtga gtaactcaacc aagtcattct 7800
gagaatagtg tatgcggcga ccgagttgct cttgcccggc gtcaatacgg gataataccg 7860
cgccacatag cagaacttta aaagtgtcca tcattgaaa acgttcttcg gggcgaaaac 7920
tctcaaggat cttaccgctg ttgagatcca gttcgatgta acccactcgt gcacccaact 7980
gatcttcagc atcttttact ttcaccagcg tttctgggtg agcaaaaaca ggaaggcaaa 8040
atgccgcaaa aaagggaata agggcgacac ggaaatgttg aatactcata ctcttccttt 8100
ttcaatatta ttgaagcatt tatcagggtt attgtctcat gaggcgatac atatttgaat 8160
gtatttagaa aaataaacia ataggggttc cgcgcacatt tccccgaaa gtgccacctg 8220
acgtctaaga aaccattatt atcatgacat taacctataa aaataggcgt atcacgaggc 8280
cctttcgtct tcac                                     8294

```

<210> 3

<211> 1512

<212> DNA

<213> Sequência Artificial

<220>

<223> Sub-unidade alfa do gene de fenilalanina aminoacil
sintetase de levedura

<220>

<221> misc_feature

<222> 1234, 1235, 1236, 1243, 1244, 1245, 1252, 1253, 1254,
1309,

1310, 1311

<223> n = A,T,C or G

<400> 3

```

atgtctgact tccaattaga aattctaaag aaactagatg aattggatga gatcaagtcc 60
acactggcaa ctttccctca gcacggctct caagatgttc tttccgcttt gaactccttg 120
aaagcccaca acaagttaga gttttccaag gtgcacacgg ttacgtatga cttgaccaa 180
gaagggtgctc aaattttgaa tgaagggttcg tacgaaatta aactagtcaa gctcatccaa 240
gagttgggtc aacttcaaat caaagatgtg atgtccaaac taggcccctca agttggtaag 300
gtcggtcagg ctagagcttt caagaacggc tggatcgcca aaaacgcctc aaacgagctt 360
gaactctccg caaaattgca aaataccgat ttaaatgagc ttactgatga aacgcaatct 420
attctagcgc aaatcaagaa caactcgcat ctggatagca ttgacgcca gattttgaac 480
gacttgaaga aaagaaagt ttattgctcaa ggtaaaatca cagatttcag tgtcaccaaa 540
gggccagagt tctcgaccga cctcaccaaa ttggaaaccg atcttacctc cgacatgggtc 600
tccaccaatg catacaagga cttgaagttc aagccttaca atttcaattc tcaaggtgtg 660
caaatactct cagggtgctct tcacccctta aacaaagtca gagaggaatt tagacaaatt 720
ttcttttcca tgggattcac agagatgccc tcgaaccaat acgtcgagac aggtttctgg 780
aacttcgatg ccctttacgt ccacacaacag catcctgtctc gtgacctgca agacacttcc 840
tacatcaagg acccactaac cgctgagttg cccgatgaca agacatacat ggacaatatc 900
aaagccgttc acgaacaggg gagattcggg tccatcggtt atcgttacaa ctggaagcca 960
gaagaatgtc aaaaattggt cttgagaact cactccacag ccactctctc cagaatgctg 1020

```

```

cacgatttgg ccaaagatcc aaagcccacc agattgtttt ctatcgaccg tgttttccgt 1080
aacgaagcag ttgacgccac ccatttggcc gaattccacc aggtggaagg tgttcttgcc 1140
gactacaaca ttactctggg tgacctgatc aagttcatgg aagagttttt cgaagaagt 1200
gggtgtcaccg gtttgagatt caagcctacc tacnnnctt acnnngagcc annnatggaa 1260
atcttttctt ggcacgaagg ttgcaaaaa tgggtcgaaa tcggtaacnn nggtatgttc 1320
agaccagaaa tgctcgagtc catgggtcta ccaaaggatc taagagtcct tgggtggggg 1380
ttatccttgg aaagacctac catgatcaaa tataagggtc aaaacatcag agaactgtta 1440
ggtcataaag tctctttgga ctttatcgaa accaatcctg ctgctagatt ggacgaagac 1500
ttgtacgaat aa 1512

```

<210> 4

<211> 13

<212> PRT

<213> Sequência Artificial

<220>

<223> Derivado artificial da dihidrofolato redutase

<400> 4

```

Asn Gly Asp Leu Pro Trp Pro Pro Leu Arg Asn Glu Lys
  1             5             10

```

<210> 5

<211> 14

<212> PRT

<213> Sequência Artificial

<220>

<223> Derivado artificial da dihidrofolato redutase

<220>

<221> VARIANTE

<222> 13

<223> Xaa = aminoácido suppressor

<400> 5

Asn	Gly	Asp	Leu	Pro	Trp	Pro	Pro	Leu	Arg	Asn	Glu	Xaa	Lys
1				5					10				

<210> 6

<211> 8

<212> PRT

<213> Sequência Artificial

<220>

<223> Derivado artificial da dihidrofolato redutase

<400> 6

Pro	Trp	Pro	Pro	Leu	Arg	Asn	Glu
1				5			

<210> 7

<211> 34

<212> DNA

<213> Sequência Artificial

<220>

<223> Oligonucleótido iniciador

<400> 7

cgattttcac acaggatcca gaccatgatt ctag 34

<210> 8

<211> 27

<212> DNA

<213> Sequência Artificial

<220>

<223> Oligonucleótido iniciador

<400> 8

gacggccagt gaattcgagc tcggtac 27

<210> 9

<211> 38

<212> DNA

<213> Sequência Artificial

<220>

<223> Oligonucleótido iniciador

<400> 9

ctacctacaa tccttacggc gagccatcaa tggaaatc 38

<210> 10

<211> 38

<212> DNA

<213> Sequência Artificial

<220>

<223> Oligonucleótido iniciador

<400> 10

gatttccatt gatggctcgc cgtaaggatt gtaggtag 38

<210> 11

<211> 36

<212> DNA

<213> Sequência Artificial

<220>

<223> Oligonucleótido iniciador

<400> 11

ccgctcagga acgagtagaa gtacttccaa agaatg 36

<210> 12
<211> 36
<212> DNA
<213> Sequência Artificial
<220>
<223> Oligonucleótido iniciador
<400> 12
cattcttttg aagtacttct actcggttcct gagcgg 36

<210> 13
<211> 30
<212> DNA
<213> Sequência Artificial
<220>
<223> Oligonucleótido iniciador
<400> 13
gcactgacca tggctgaaca acacgcacag 30

<210> 14
<211> 31
<212> DNA
<213> Sequência Artificial
<220>
<223> Oligonucleótido iniciador
<400> 14
ggacttcgga tcctttctgt gggcgcatcg c 31

<210> 15
<211> 14

<212> PRT

<213> Sequência Artificial

<220>

<223> Sub-unidade alfa do gene de fenilalanina aminoacil
sintetase de levedura

<220>

<221> VARIANTE

<222> 13

<223> Xaa = aminoácido suppressor

<400> 15

Asn Gly Asp Leu Pro Trp Pro Pro Leu Arg Asn Glu Xaa Lys
1 5 10

<210> 16

<211> 39

<212> DNA

<213> Sequência Artificial

<220>

<223> Oligonucleótido iniciador

<400> 16

gaacacagga cctccacatt tagagtatgg cgctctccc 39

<210> 17

<211> 39

<212> DNA

<213> Sequência Artificial

<220>

<223> Oligonucleótido iniciador

<400> 17

gggagagcgc catactctaa atgtggaggt cctgtgttc 39

<210> 18

<211> 42
<212> DNA
<213> Sequência Artificial
<220>
<223> Oligonucleótido iniciador
<400> 18
ctgggtaagc ttcgctaagg atctgccctg gtgcgaactc tg 42

<210> 19
<211> 43
<212> DNA
<213> Sequência Artificial
<220>
<223> Oligonucleótido iniciador
<400> 19
gattacggat tcctaatacg actcactata gcggacttag ctc 43

<210> 20
<211> 40
<212> DNA
<213> Sequência Artificial
<220>
<223> Oligonucleótido iniciador
<400> 20
cggaagcaga aagtgtaaag agcgggggtgc ctaatgagtg 40

<210> 21
<211> 40
<212> DNA
<213> Sequência Artificial

<220>

<223> Oligonucleótido iniciador

<400> 21

cactcattag gcaccccgct ctttacactt tatgcttccg 40

<210> 22

<211> 42

<212> DNA

<213> Sequência Artificial

<220>

<223> Oligonucleótido iniciador

<400> 22

cttgtcacta ctctgaccta tgggtgttcaa tgcttctccc gt 42

<210> 23

<211> 42

<212> DNA

<213> Sequência Artificial

<220>

<223> Oligonucleótido iniciador

<400> 23

acgggagaag cattgaacac cataggtcag agtagtgaca ag 42

<210> 24

<211> 39

<212> DNA

<213> Sequência Artificial

<220>

<223> Oligonucleótido iniciador

<400> 24

gtacaggaac gcactatatt cttcaaagat gacgggaac 39

<210> 25

<211> 39

<212> DNA

<213> Sequência Artificial

<220>

<223> Oligonucleótido iniciador

<400> 25

gttcccgta tctttgaaga atatagtgcg ttctgtac 39

<210> 26

<211> 39

<212> DNA

<213> Sequência Artificial

<220>

<223> Oligonucleótido iniciador

<400> 26

cacaatgtat acatcatggc agacaaacaa aagaatgga 39

<210> 27

<211> 39

<212> DNA

<213> Sequência Artificial

<220>

<223> Oligonucleótido iniciador

<400> 27

tccattcttt tgtttgtctg ccatgatgta tacattgtg 39

<210> 28
<211> 45
<212> DNA
<213> Sequência Artificial
<220>
<223> Oligonucleótido iniciador
<400> 28
gtgccacctg acgtctaaga aaccattatt atcatgacat taacc 45

<210> 29
<211> 41
<212> DNA
<213> Sequência Artificial
<220>
<223> Oligonucleótido iniciador
<400> 29
gagtaaagga gaagaacttt ttactggagt tgtcccaatt c 41

<210> 30
<211> 41
<212> DNA
<213> Sequência Artificial
<220>
<223> Oligonucleótido iniciador
<400> 30
gaattgggac aactccagta aaaagttctt ctcctttact c 41

<210> 31
<211> 43
<212> DNA

<213> Sequência Artificial

<220>

<223> Oligonucleótido iniciador

<400> 31

ggccaacact tgtcactact ttacctatg gtgttcaatg ctt 43

<210> 32

<211> 43

<212> DNA

<213> Sequência Artificial

<220>

<223> Oligonucleótido iniciador

<400> 32

aagcattgaa caccataggt aaaagtagtg acaagtgttg gcc 43

<210> 33

<211> 43

<212> DNA

<213> Sequência Artificial

<220>

<223> Oligonucleótido iniciador

<400> 33

catatgaaac ggcattgactt ttttaagagt gccatgcccg aag 43

<210> 34

<211> 43

<212> DNA

<213> Sequência Artificial

<220>

<223> Oligonucleótido iniciador

<400> 34

cttcgggcat ggcactctta aaaaagtcac gccgtttcat atg 43

<210> 35

<211> 49

<212> DNA

<213> Sequência Artificial

<220>

<223> Oligonucleótido iniciador

<400> 35

gttatgtaca ggaacgcact atatttttca aagatgacgg gaactacaa 49

<210> 36

<211> 49

<212> DNA

<213> Sequência Artificial

<220>

<223> Oligonucleótido iniciador

<400> 36

ttgtagttcc cgtcatcttt gaaaaatata gtgcgttcct gtacataac 49

<210> 37

<211> 53

<212> DNA

<213> Sequência Artificial

<220>

<223> Oligonucleótido iniciador

<400> 37

acaaaagaat ggaatcaaag ctaactttaa aattcgccac aacattgaag atg

53

<210> 38

<211> 53

<212> DNA

<213> Sequência Artificial

<220>

<223> Oligonucleótido iniciador

<400> 38

catcttcaat gttgtggcga attttaagat tagctttgat tccattcttt tgt
53

<210> 39

<211> 39

<212> DNA

<213> Sequência Artificial

<220>

<223> Oligonucleótido iniciador

<400> 39

cgccaagcta gcttggattc tcaccaataa aaaacgccc 39

<210> 40

<211> 43

<212> DNA

<213> Sequência Artificial

<220>

<223> Oligonucleótido iniciador

<400> 40

tacctatggt gttcaatgct tttcccgta tccggatcat atg 43

<210> 41
<211> 43
<212> DNA
<213> Sequência Artificial
<220>
<223> Oligonucleótido iniciador
<400> 41
catatgatcc ggataacggg aaaagcattg aacaccatag gta 43

<210> 42
<211> 50
<212> DNA
<213> Sequência Artificial
<220>
<223> Oligonucleótido iniciador
<400> 42
gttatgtaca ggaacgcact atatttttta aagatgacgg gaactacaag 50

<210> 43
<211> 50
<212> DNA
<213> Sequência Artificial
<220>
<223> Oligonucleótido iniciador
<400> 43
cttgtagttc ccgtcatctt taaaaaatat agtgcgttcc tgtacataac 50

<210> 44
<211> 43
<212> DNA

<213> Sequência Artificial

<220>

<223> Oligonucleótido iniciador

<220>

<221> misc_feature

<222> 16, 17

<223> n = A,T,C ou G

<400> 44

gtcgaaatcg gtaacnnkgg tatgttcaga ccagaaatgc tcg 43

<210> 45

<211> 43

<212> DNA

<213> Sequência Artificial

<220>

<223> Oligonucleótido iniciador

<220>

<221> misc_feature

<222> 27, 28

<223> n = A,T,C ou G

<400> 45

cgagcatttc tggctctgaac ataccmnngt taccgatttc gac 43

<210> 46

<211> 47

<212> DNA

<213> Sequência Artificial

<220>

<223> Oligonucleótido iniciador

<220>

<221> misc_feature

<222> 14, 15, 23, 24, 32, 33

<223> n = A,T,C ou G

<400> 46

caagcctacc tacnnkcctt acnnkgagcc annkatggaa atctttt 47

<210> 47

<211> 47

<212> DNA

<213> Sequência Artificial

<220>

<223> Oligonucleótido iniciador

<220>

<221> misc_feature

<222> 15, 16, 24, 25, 33, 34

<223> n = A,T,C ou G

<400> 47

aaaagatttc catmnntggc tcmnngtaag gmnnntaggt aggcttg 47

<210> 48

<211> 432

<212> PRT

<213> Thermus thermophilus

<400> 48

```

Met Ala Gly Thr Gly His Thr Pro Glu Glu Ala Leu Ala Leu Leu Lys
 1          5          10          15
Arg Gly Ala Glu Glu Ile Val Pro Glu Glu Glu Leu Leu Ala Lys Leu
 20          25          30
Lys Glu Gly Arg Pro Leu Thr Val Lys Leu Gly Ala Asp Pro Thr Arg
 35          40          45
Pro Asp Leu His Leu Gly His Ala Val Val Leu Arg Lys Met Arg Gln
 50          55          60
Phe Gln Glu Leu Gly His Lys Val Val Leu Ile Ile Gly Asp Phe Thr
 65          70          75          80
Gly Met Ile Gly Asp Pro Ser Gly Arg Ser Lys Thr Arg Pro Pro Leu
 85          90          95
Thr Leu Glu Glu Thr Arg Glu Asn Ala Lys Thr Tyr Val Ala Gln Ala
100          105          110
Gly Lys Ile Leu Arg Gln Glu Pro His Leu Phe Glu Leu Arg Tyr Asn
115          120          125
Ser Glu Trp Leu Glu Gly Leu Thr Phe Lys Glu Val Val Arg Leu Thr
130          135          140
Ser Leu Met Thr Val Ala Gln Met Leu Glu Arg Glu Asp Phe Lys Lys
145          150          155          160
Arg Tyr Glu Ala Gly Ile Pro Ile Ser Leu His Glu Leu Leu Tyr Pro
165          170          175
Phe Ala Gln Ala Tyr Asp Ser Val Ala Ile Arg Ala Asp Val Glu Met
180          185          190
Gly Gly Thr Asp Gln Arg Phe Asn Leu Leu Val Gly Arg Glu Val Gln
195          200          205
Arg Ala Tyr Gly Gln Ser Pro Gln Val Cys Phe Leu Met Pro Leu Leu
210          215          220
Val Gly Leu Asp Gly Arg Glu Lys Met Ser Lys Ser Leu Asp Asn Tyr
225          230          235          240
Ile Gly Leu Thr Glu Pro Pro Glu Ala Met Phe Lys Lys Leu Met Arg
245          250          255
Val Pro Asp Pro Leu Leu Pro Ser Tyr Phe Arg Leu Leu Thr Asp Leu
260          265          270
Glu Glu Glu Glu Ile Glu Ala Leu Leu Lys Ala Gly Pro Val Pro Ala
275          280          285          290
His Arg Val Leu Ala Arg Leu Leu Thr Ala Ala Tyr Ala Leu Pro Gln

```

```

290          295          300
Ile Pro Pro Arg Ile Asp Arg Ala Phe Tyr Glu Ser Leu Gly Tyr Ala
305          310          315          320
Trp Glu Ala Phe Gly Arg Asp Lys Glu Ala Gly Pro Glu Glu Val Arg
325          330          335
Arg Ala Glu Ala Arg Tyr Asp Glu Val Ala Lys Gly Gly Ile Pro Glu
340          345          350
Glu Ile Pro Glu Val Thr Ile Pro Ala Ser Glu Leu Lys Glu Gly Arg
355          360          365
Ile Trp Val Ala Arg Leu Phe Thr Leu Ala Gly Leu Thr Pro Ser Asn
370          375          380
Ala Glu Ala Arg Arg Leu Ile Gln Asn Arg Gly Leu Arg Leu Asp Gly
385          390          395          400
Glu Val Leu Thr Asp Pro Met Leu Gln Val Asp Leu Ser Arg Pro Arg
405          410          415
Ile Leu Gln Arg Gly Lys Asp Arg Phe Val Arg Val Arg Leu Ser Asp
420          425          430

```

<210> 49

<211> 372

<212> PRT

<213> Homo sapiens

<400> 49

```

Met Gly Asp Ala Pro Ser Pro Glu Glu Lys Leu His Leu Ile Thr Arg
 1          5          10          15
Asn Leu Gln Glu Val Leu Gly Glu Glu Lys Leu Lys Glu Ile Leu Lys
 20          25          30
Glu Arg Glu Leu Lys Ile Tyr Trp Gly Thr Ala Thr Thr Gly Lys Pro
 35          40          45
His Val Ala Tyr Phe Val Pro Met Ser Lys Ile Ala Asp Phe Leu Lys
 50          55          60
Ala Gly Cys Glu Val Thr Ile Leu Phe Ala Asp Leu His Ala Tyr Leu
 65          70          75          80
Asp Asn Met Lys Ala Pro Trp Glu Leu Leu Glu Leu Arg Val Ser Tyr
 85          90          95
Tyr Glu Asn Val Ile Lys Ala Met Leu Glu Ser Ile Gly Val Pro Leu
100          105          110
Glu Lys Leu Lys Phe Ile Lys Gly Thr Asp Tyr Gln Leu Ser Lys Glu
115          120          125
Tyr Thr Leu Asp Val Tyr Arg Leu Ser Ser Val Val Thr Gln His Asp
130          135          140
Ser Lys Lys Ala Gly Ala Glu Val Val Lys Gln Val Glu His Pro Leu
145          150          155          160
Leu Ser Gly Leu Leu Tyr Pro Gly Leu Gln Ala Leu Asp Glu Glu Tyr
165          170          175
Leu Lys Val Asp Ala Gln Phe Gly Gly Ile Asp Gln Arg Lys Ile Phe
180          185          190
Thr Phe Ala Glu Lys Tyr Leu Pro Ala Leu Gly Tyr Ser Lys Arg Val
195          200          205          210
His Leu Met Asn Pro Met Val Pro Gly Leu Thr Gly Ser Lys Met Ser
215          220          225
Ser Ser Glu Glu Glu Ser Lys Ile Asp Leu Leu Asp Arg Lys Glu Asp
230          235          240
Val Lys Lys Lys Leu Lys Lys Ala Phe Cys Glu Pro Gly Asn Val Glu
245          250          255
Asn Asn Gly Val Leu Ser Phe Ile Lys His Val Leu Phe Pro Leu Lys
260          265          270
Ser Glu Phe Val Ile Leu Arg Asp Glu Lys Trp Gly Gly Asn Lys Thr
275          280          285
Tyr Thr Ala Tyr Val Asp Leu Glu Lys Asp Phe Ala Ala Glu Val Val
290          295          300
His Pro Gly Asp Leu Lys Asn Ser Val Glu Val Ala Leu Asn Lys Leu
305          310          315          320
Leu Asp Pro Ile Arg Glu Lys Phe Asn Thr Pro Ala Leu Lys Lys Leu
325          330          335
Ala Ser Ala Ala Tyr Pro Asp Pro Ser Lys Gln Lys Pro Met Ala Lys
340          345          350
Gly Pro Ala Lys Asn Ser Glu Pro Glu Glu Val Ile Leu Glu His His
355          360          365
His His His His
370

```

<210> 50

<211> 424

<212> PRT

<213> Escherichia coli

<400> 50

```

Met Ala Ser Ser Asn Leu Ile Lys Gln Leu Gln Glu Arg Gly Leu Val
 1      5      10      15
Ala Gln Val Thr Asp Glu Glu Ala Leu Ala Glu Arg Leu Ala Gln Gly
 20      25      30
Pro Ile Ala Leu Tyr Cys Gly Phe Asp Pro Thr Ala Asp Ser Leu His
 35      40      45
Leu Gly His Leu Val Pro Leu Leu Cys Leu Lys Arg Phe Gln Gln Ala
 50      55      60
Gly His Lys Pro Val Ala Leu Val Gly Gly Ala Thr Gly Leu Ile Gly
 65      70      75      80
Asp Pro Ser Phe Lys Ala Ala Glu Arg Lys Leu Asn Thr Glu Glu Thr
 85      90      95
Val Gln Glu Trp Val Asp Lys Ile Arg Lys Gln Val Ala Pro Phe Leu
100      105      110
Asp Phe Asp Cys Gly Glu Asn Ser Ala Ile Ala Ala Asn Asn Tyr Asp
115      120      125
Trp Phe Gly Asn Met Asn Val Leu Thr Phe Leu Arg Asp Ile Gly Lys
130      135      140
His Phe Ser Val Asn Gln Met Ile Asn Lys Glu Ala Val Lys Gln Arg
145      150      155      160
Leu Asn Arg Glu Asp Gln Gly Ile Ser Phe Thr Glu Phe Ser Tyr Asn
165      170      175
Leu Leu Gln Gly Tyr Asp Phe Ala Cys Leu Asn Lys Gln Tyr Gly Val
180      185      190
Val Leu Gln Ile Gly Gly Ser Asp Gln Trp Gly Asn Ile Thr Ser Gly
195      200      205
Ile Asp Leu Thr Arg Arg Leu His Gln Asn Gln Val Phe Gly Leu Thr
210      215      220
Val Pro Leu Ile Thr Lys Ala Asp Gly Thr Lys Phe Gly Lys Thr Glu
225      230      235      240
Gly Gly Ala Val Trp Leu Asp Pro Lys Lys Thr Ser Pro Tyr Lys Phe
245      250      255
Tyr Gln Phe Trp Ile Asn Thr Ala Asp Ala Asp Val Tyr Arg Phe Leu
260      265      270
Lys Phe Phe Thr Phe Met Ser Ile Glu Glu Ile Asn Ala Leu Glu Glu
275      280      285
Glu Asp Lys Asn Ser Gly Lys Ala Pro Arg Ala Gln Tyr Val Leu Ala
290      295      300
Glu Gln Val Thr Arg Leu Val His Gly Glu Glu Gly Leu Gln Ala Ala
305      310      315      320
Lys Arg Ile Thr Glu Cys Leu Phe Ser Gly Ser Leu Ser Ala Leu Ser
325      330      335
Glu Ala Asp Phe Glu Gln Leu Ala Gln Asp Gly Val Pro Met Val Glu
340      345      350
Met Glu Lys Gly Ala Asp Leu Met Gln Ala Leu Val Asp Ser Glu Leu
355      360      365
Gln Pro Ser Arg Gly Gln Ala Arg Lys Thr Ile Ala Ser Asn Ala Ile
370      375      380
Thr Ile Asn Gly Glu Lys Gln Ser Asp Pro Glu Tyr Phe Phe Lys Glu
385      390      395      400
Glu Asp Arg Leu Phe Gly Arg Phe Thr Leu Leu Arg Arg Gly Lys Lys
405      410      415
Asn Tyr Cys Leu Ile Cys Trp Lys
420

```

<210> 51

<211> 442

<212> PRT

<213> Drosophila melanogaster

<400> 51

```

Met Ser Gln Lys Asn Leu Leu Glu Leu Thr Asp Arg Gly Phe Phe His
 1      5      10      15
Gly Ile Phe Pro Asp Thr Ala Ala Pro Arg Met Lys Gln Leu Phe Thr
 20      25      30
Arg Gly Gln Gln Ser Ile Tyr Ala Gly Phe Asp Pro Thr Ala Asp Ser
 35      40      45
Leu His Val Gly Asn Leu Leu Val Ile Met Gly Leu Ile His Cys Gln
 50      55      60
Arg Ala Gly His Arg Pro Ile Ala Leu Val Gly Gly Ala Thr Gly Leu
 65      70      75      80
Ile Gly Asp Pro Ser Gly Arg Lys Thr Glu Arg Asn Gln Leu Gly Glu
 85      90      95
Thr Val Ile Glu Thr Asn Leu Lys Ala Ile Glu Gln Gln Leu Arg Arg
100      105      110
Val Phe Glu Asn His Glu Asn Cys Leu Trp Asp Ser Lys Lys Gln Lys
115      120      125
Leu Pro Leu Ala Pro Leu Ile Ile Val Asn Asn Ala Asp Trp Tyr Ala
130      135      140
Asp Leu Gln Leu Ile Asp Phe Val Ala Asn Met Gly Arg His Phe Arg
145      150      155      160
Met Gly Ser Met Leu Ser Arg Ser Ser Val Gln Ser Arg Leu Glu Ser
165      170      175
Glu Asp Gly Met Ser Phe Thr Glu Phe Thr Tyr Gln Ile Phe Gln Ala
180      185      190
Tyr Asp Trp Leu His Leu Leu Arg Arg His Asn Cys Cys Phe Gln Met
195      200      205
Gly Gly Ser Asp Gln Thr Gly Asn Leu Met Thr Gly His Glu Leu Ile
210      215      220
Ser Arg Val Glu Arg Lys Arg Glu Val Phe Gly Leu Thr Leu Pro Leu
225      230      235      240
Val Thr Thr Glu Glu Gly Asp Lys Phe Gly Lys Ser Ala Gly Asn Ala
245      250      255
Val Trp Leu Asp Gly Asn Lys Thr Ser Pro Phe Ala Leu Tyr Gln Phe
260      265      270
Phe Leu Arg Met Pro Asp Ser Glu Val Glu Lys Leu Leu Lys Leu Phe
275      280      285
Thr Phe Ile Pro Leu Pro Gln Val Glu Gln Leu Met Arg Glu His Thr
290      295      300
Lys Glu Pro Glu Lys Arg Lys Ala Gln Thr Leu Leu Ala Glu Asp Val
305      310      315      320
Thr Leu Leu Val His Gly Glu Ser Gly Leu Lys Gln Ala Glu Arg Val
325      330      335
Thr Asn Ala Leu Tyr Lys Gly Asn Val Glu Gly Leu Ala Glu Leu Asn
340      345      350
Leu Ser Glu Ile Gln Gln Thr Phe Gln Gly Ala Thr Met Val Asn Leu
355      360      365
Leu Thr Glu Pro Gly Met Ser Ile Leu Glu Leu Ala Met Lys Ala Lys
370      375      380
Cys Phe Pro Thr Glu Thr Asp Ala Val Arg Ile Ile Asn Ala Gly Gly
385      390      395      400
Phe Tyr Val Asn Gln Lys Arg Val Gln Asn Ile Ala Glu Val Leu Thr
405      410      415
Thr Gly Val His Ile Leu Arg Asn Gly Ile Ser Leu Leu Arg Val Gly
420      425      430
Lys Arg Asn Phe Tyr Ile Val Arg Trp Gln
435      440

```

<210> 52

<211> 492

<212> PRT

<213> *Saccharomyces cerevisiae*

<400> 52

```

Met Leu Glu Leu Arg Ser Cys Ser Asn Leu Val Asn Ser Ser Arg Arg
 1          5          10          15
Leu Val Pro Leu Val Thr Tyr Ser Gly Leu Ser Ala Ile Thr Leu Pro
 20          25          30
Lys Ser Arg Phe Tyr Ser Gln Pro Ser Ala Leu Glu Val Gln Gly Thr
 35          40          45
Ser Asp Ser Arg Ser Asp Asn Ile Leu Asp Glu Leu Lys Gln Arg Gly
 50          55          60
Leu Val Ser Gln Val Ser Gln Pro Glu Ser Phe Leu Arg Thr Lys Leu
 65          70          75          80
Asn Gly Asn Asp Lys Ile Lys Leu Tyr Cys Gly Val Asp Pro Thr Ala
 85          90          95
Gln Ser Leu His Leu Gly Asn Leu Val Pro Leu Met Val Leu Leu His
 100          105          110
Phe Tyr Val Lys Gly His Asp Ile Val Thr Val Ile Gly Gly Ala Thr
 115          120          125
Gly Lys Val Gly Asp Pro Ser Gly Arg Lys Thr Glu Arg Asp Val Met
 130          135          140
Glu Asn Asp Ile Arg Gln Ser Asn Val Ala Ser Ile Ser Gln Gln Leu
 145          150          155          160
Gln Arg Phe Phe Lys Asn Gly Leu Glu Tyr Tyr Arg Asn Arg Cys Ala
 165          170          175
Leu Thr Glu Asp Val Pro Ser Gly Lys Tyr Thr Pro Arg Asn Asn Phe
 180          185          190
Asn Trp Trp Lys Asp Ile Lys Met Leu Asp Phe Leu Ala Asp Phe Gly
 195          200          205
Arg His Ile Arg Val Gln Ser Met Leu Ala Arg Asp Ser Ile Ser Ser
 210          215          220
Arg Leu Gln Thr Lys Asn Gly Leu Gly Phe Asn Glu Phe Thr Tyr Gln
 225          230          235          240
Val Leu Gln Ala Tyr Asp Phe Tyr His Leu Tyr Lys Glu Glu Asn Val
 245          250          255
Thr Ile Gln Val Gly Gly Asn Asp Gln Trp Gly Asn Ile Thr Ala Gly
 260          265          270
Ile Asp Leu Ile Asn Arg Ile Gln Pro Ile Lys Asn Lys Gly Leu Pro
 275          280          285
Phe Gly Ile Thr Val Pro Leu Leu Thr Thr Ala Thr Gly Glu Lys Phe
 290          295          300
Gly Lys Ser Ala Gly Asn Ala Val Phe Ile Asp Pro Ser Ile Asn Thr
 305          310          315          320
Ala Tyr Asp Val Tyr Gln Phe Phe Tyr Asn Thr Leu Asp Ala Asp Val
 325          330          335
Pro Lys Phe Leu Lys Ile Phe Thr Phe Leu Asn Ser Ser Glu Ile Lys
 340          345          350
Lys Ile Val Glu Thr His Ile Lys Ser Pro Ser Leu Arg Tyr Gly Gln
 355          360          365
Thr Leu Leu Ala Lys Glu Val Thr Asp Met Leu Tyr Gly Val Gly Ser
 370          375          380
Gly Ser Asp Ser Glu Ala Leu Ser Asn Ile Ile Phe Gly Arg Tyr Asp
 385          390          395          400
Gly Thr Leu Ser Ala Ala Lys Leu Val Asp Leu Cys Lys Lys Ala Arg
 405          410          415
Ile Leu Gln Tyr Ala Asp Arg Glu Ile Asp Leu Ile Lys Leu Ile Cys
 420          425          430
Lys Leu Val Asn Cys Ser Val Ser Glu Ala Arg Arg Lys Leu Ser Gln
 435          440          445
Gly Ser Val Tyr Leu His His Ser Lys Ser Lys Val Asn Glu Asn Ile
 450          455          460
Ser Asn Leu Ala Pro Phe Leu Ile Asp Asp Arg Val Leu Ile Leu Arg
 465          470          475          480
Ile Gly Lys Gln Lys Cys Phe Ile Ile Glu Met Arg
 485          490

```

<210> 53

<211> 445

<212> PRT

<213> Schizosaccharomyces pombe

<400> 53

```

Met Ser Arg Leu Leu Ala Cys Leu Lys Gln Leu Gln Ala Arg Ser Leu
 1      5      10      15
Ile His Asn Thr Thr Leu Leu Gln Pro Ser Cys Asn Val Asn Ser Val
      20      25      30
Tyr Leu Gly Ala Asp Pro Thr Ala Ala Ser Leu His Val Gly Asn Leu
      35      40      45
Val Ala Leu Met Pro Leu Val His Phe Phe Leu Asn Gly Phe Pro Val
      50      55      60
Phe Thr Val Ile Gly Asp Ala Thr Ala Gln Leu Gly Asp Pro Ser Gly
65      70      75      80
Arg Ser Thr Ser Arg Lys Gln Met Ala Glu Thr Thr Arg Thr Ala Asn
      85      90      95
Ser Asn Ser Ile His Asn Gln Leu Lys Asp Leu Ser Ser Ser Ile Leu

```

```

      100      105      110
Ser Tyr Ala Gln Asp Cys Asn Tyr Pro Phe Ser Gln Met Pro Ser Ser
      115      120      125
Ser Gln Trp Ser Ile Val Arg Asn Ser Ser Trp Tyr Glu Asn Leu Lys
      130      135      140
Leu Leu Lys Phe Leu Ser Ser Val Gly Pro His Val Arg Val Ser Gln
145      150      155      160
Met Leu Ala Arg Asp Ser Val Thr Thr Arg Leu Gln Ser Pro Ser Gly
      165      170      175
Leu Ser Phe Ala Glu Leu Thr Tyr Gln Leu Leu Gln Ala Tyr Asp Tyr
      180      185      190
Ser Tyr Leu Tyr Glu Asn His Ser Val Asn Leu Gln Ile Gly Gly Ser
      195      200      205
Asp Gln Trp Gly Asn Ile Thr Ala Gly Thr Asp Leu Val Arg Arg Thr
210      215      220
His Pro Asn Ala Asn Val Tyr Ala Leu Thr Thr Pro Leu Leu Thr Ser
225      230      235      240
Ser Ser Gly Gln Lys Leu Gly Lys Ser Ala Gly Asn Ala Ile Trp Leu
      245      250      255
Asp Pro Lys Leu Thr Asp Ser Tyr Ser Leu Tyr Gln Tyr Phe Ile Ser
260      265      270
Ala Pro Asp Asp Leu Ala Cys Lys Cys Leu Asp Met Leu Thr Leu Leu
275      280      285
Pro Leu Glu Gln Leu Glu Gln Ile Lys Ala Glu His Glu Lys Asp Pro
290      295      300
Ser Gln Arg Ile Val His Lys Tyr Leu Ala Ser Asn Val Val Arg Met
305      310      315      320
Val His Gly Lys Lys Ala Leu Glu Leu Ala Gln Ile Gln Thr Lys Leu
      325      330      335
Leu His Gly Ala His Gln Ala Pro Phe Gly Phe Tyr Ser Glu Ala Pro
340      345      350
Gln Gln Gly Asp Ser Phe Pro Ser Leu Pro Glu Ile Arg Ala Leu Phe
355      360      365
Lys Asp Cys Lys Phe Tyr Arg Thr Ile Asp Ser Ser Ile Lys Asp Gln
370      375      380
Pro Phe Ser Arg Leu Leu Arg Thr Leu Gln Ile Tyr Thr Ser Arg Lys
385      390      395      400
Glu Ala Thr Glu His Ile Leu Ser Gly Ala Val Ser Leu Gly His Lys
      405      410      415
Pro Ile Leu Asp Ser Asn Tyr Lys Phe Pro Asp Asn Ser Leu Phe Val
420      425      430
Leu Arg Ala Gly Lys Arg Thr Phe Val Leu Asp Ser Leu
435      440      445

```

<210> 54

<211> 528

<212> PRT

<213> *Xenopus laevis*

<400> 54

```

Met Gly Asp Ser Leu Thr Leu Glu Gly Lys Ala Gln Leu Ile Thr Arg
 1      5      10      15
Asn Leu Gln Glu Leu Leu Gly Glu Asp Lys Met Lys Glu Ile Leu Lys
 20      25      30
Glu Arg Pro Leu Arg Ile Tyr Trp Gly Thr Ala Thr Thr Gly Lys Pro
 35      40      45
His Val Ala Tyr Phe Val Pro Met Ser Lys Ile Ala Asp Phe Leu Lys

50      55      60
Ala Gly Cys Glu Val Thr Ile Leu Phe Ala Asp Leu His Ala Tyr Leu
65      70      75      80
Asp Asn Met Lys Ala Pro Trp Asp Leu Leu Glu Leu Arg Thr Arg Tyr
 85      90      95
Tyr Glu Gln Val Ile Gln Ala Met Leu Gln Ser Ile Gly Val Pro Leu
100      105      110
Glu Arg Leu Arg Phe Ile Arg Gly Thr Glu Phe Gln Leu Ser Lys Glu
115      120      125
Tyr Thr Leu Asp Val Tyr Arg Leu Ser Ser Val Val Thr Gln His Asp
130      135      140
Ala Lys Lys Ala Gly Ala Glu Val Val Lys Gln Val Glu His Pro Leu
145      150      155      160
Leu Ser Gly Leu Leu Tyr Pro Gly Leu Gln Ala Leu Asp Glu Glu Tyr
165      170      175
Leu Lys Val Asp Ala Gln Phe Gly Gly Val Asp Gln Arg Lys Ile Phe
180      185      190
Thr Phe Ala Glu Lys Tyr Leu Pro Ala Leu Gly Tyr Ala Lys Arg Ile
195      200      205
His Leu Met Asn Pro Met Val Pro Gly Leu Thr Gly Ala Lys Met Ser
210      215      220
Ser Ser Glu Glu Glu Ser Lys Ile Asp Leu Leu Asp Ser Pro Ala Asp
225      230      235      240
Val Lys Lys Lys Leu Lys Lys Ala Phe Cys Glu Pro Gly Asn Val Glu
245      250      255
Asn Asn Gly Val Leu Ser Phe Val Arg His Val Leu Phe Pro Leu Lys
260      265      270
Ser Glu Phe Val Val Leu Arg Asp Glu Lys Phe Gly Gly Asn Lys Thr
275      280      285
Tyr Thr Asp Phe Glu Thr Leu Glu Lys Asp Phe Ala Glu Glu Leu Val
290      295      300
His Pro Gly Asp Leu Lys Ala Ser Val Glu Lys Ala Leu Asn Lys Leu
305      310      315      320
Leu His Pro Ile Arg Glu Lys Phe Asn Ser Pro Glu Met Lys Lys Leu
325      330      335
Ser Asn Asp Ala Tyr Pro Asp Ala Ser Lys Gln Lys Ser Val Pro Lys
340      345      350
Gly Ser Thr Lys Asn Ser Gly Thr Glu Glu Ile Asp Pro Ser Leu Leu
355      360      365
Asp Leu Arg Val Gly Lys Ile Leu Ser Val Ser Gln His Pro Asp Ala
370      375      380
Asp Ser Leu Tyr Val Glu Ser Val Asp Val Gly Glu Ala Asn Pro Arg
385      390      395      400
Cys Val Val Ser Gly Leu Val Gln Tyr Val Pro Ser Asp Gln Leu Leu
405      410      415
Gly Arg Ser Val Val Leu Leu Cys Asn Leu Lys Pro Gln Lys Met Arg
420      425      430
Gly Ile Glu Ser Gln Gly Met Leu Leu Cys Ala Ser Thr Glu Gly Glu
435      440      445
Gln Lys Gln Val Glu Pro Leu Asp Pro Pro Ser Gly Ser Ala Pro Gly
450      455      460
Glu Arg Ile Tyr Ile Glu Gly Tyr Glu Asn Gly Glu Pro Glu Gly Glu
465      470      475      480
Leu Lys Pro Lys Lys Lys Val Phe Glu Lys Leu Gln Val Asp Phe Arg
485      490      495
Ile Ser Asp Asp Leu Cys Ala Gln Trp Lys Gly Lys Asn Phe Leu Thr
500      505      510
Lys Leu Gly Ser Val Thr Cys Lys Thr Leu Arg Gly Gly Ser Ile Gly

```

515

520

525

<210> 55

<211> 402

<212> PRT

<213> *Helicobacter pylori*

<400> 55

```

Met Glu Gln Lys Ile Ala Ile Ala Leu Lys Glu Ile Ala Arg Gly Thr
 1      5      10      15
Asn Glu Ile Ile Gly Leu Glu Tyr Ile Glu Lys Leu Val Arg Lys Tyr
      20      25      30
Tyr Glu Thr Asn Glu Arg Phe Ile Val Lys Ala Gly Phe Asp Pro Thr
      35      40      45
Ala Pro Asp Leu His Leu Gly His Thr Val Leu Ile Gln Lys Leu Ala
      50      55      60
Leu Leu Gln Gln Tyr Gly Ala Arg Val Lys Phe Leu Ile Gly Asp Phe
65      70      75      80
Thr Ala Met Ile Gly Asp Pro Thr Gly Lys Asn Glu Thr Arg Lys Pro
      85      90      95
Leu Asn Arg Glu Gln Val Leu Glu Asn Ala Lys Thr Tyr Glu Glu Gln
      100      105      110
Ile Tyr Lys Ile Leu Asp Glu Lys His Thr Glu Val Cys Phe Asn Ser
      115      120      125
Thr Trp Leu Asp Ala Leu Gly Ala Lys Gly Met Ile Glu Leu Cys Ala
      130      135      140
Lys Phe Ser Val Ala Arg Met Leu Glu Arg Asp Asp Phe Thr Lys Arg
145      150      155      160
Tyr Lys Glu Asn Arg Pro Ile Ser Ile Val Glu Phe Leu Tyr Pro Leu
      165      170      175
Leu Gln Gly Tyr Asp Ser Val Ala Met Asp Ala Asp Ile Glu Leu Gly
      180      185      190
Gly Asn Asp Gln Lys Phe Asn Leu Leu Val Gly Arg Phe Leu Gln Arg
      195      200      205
Ala Tyr Gly Leu Asn Lys Glu Gln Ser Val Ile Thr Met Pro Leu Leu
      210      215      220
Glu Gly Leu Asp Gly Val Gln Lys Met Ser Lys Ser Leu Gly Asn Tyr
225      230      235      240
Val Gly Ile Thr Glu Glu Pro Asn Ala Met Phe Gly Lys Ile Met Ser
      245      250      255
Val Ser Asp Asp Leu Met Trp Arg Tyr Tyr Thr Leu Leu Ser Thr Lys
      260      265      270
Thr Leu Glu Glu Ile Glu Asp Leu Lys His Gly Ile Leu Asn Gln Thr
      275      280      285
Leu His Pro Lys Ala Val Lys Glu Asp Leu Ala Ser Glu Ile Val Ala
      290      295      300
Arg Tyr Tyr Asp Asn Asp Gln Ala Ile Lys Ala Lys Glu Gln Phe Ser
305      310      315      320
Lys Val Phe Ser Ala Asn Leu Leu Pro Glu Ile Leu Ser Glu Ser Asp
      325      330      335
Phe Asp Glu Gly Val Gly Ile Leu Asp Val Leu Lys Gln Ile Gly Phe
      340      345      350
Cys Pro Ser Thr Ser Gln Ala Arg Arg Asp Ile Gln Gly Gly Gly Val
      355      360      365
Lys Ile Asn Gln Glu Val Ile Lys Asn Glu Ser Tyr Arg Phe Val Lys
      370      375      380
Gly Asn Tyr Val Ile Gln Leu Gly Lys Lys Arg Phe Met Lys Leu Asn
      385      390      395      400
Ile Asn

```

<210> 56
 <211> 447
 <212> PRT
 <213> Caenorhabditis elegans
 <400> 56

```

Met Phe Asn Ser Lys Arg Ile Gly Asn Val Ala Leu Lys Thr Val Arg
 1      5      10      15
Ala Pro Arg Glu Ser Ser Phe Val Asp Tyr Ile Thr Asp Leu Asn Ala
      20      25      30
Arg Lys Gln Leu Gln His Ser Tyr Pro Thr Asp Leu Leu Ser Lys Cys
      35      40      45
Ser Glu Asp Leu Arg Gln Leu Pro Pro Tyr Val Tyr Ala Gly Phe Asp
      50      55      60
Pro Thr Ala Glu Ser Leu His Ile Gly Asn Leu Leu Ile Leu Val Asn
      65      70      75      80
Leu Ile Arg Ala Gln Gln Phe Gly Ile Arg Pro Ile Ala Leu Ile Gly
      85      90      95
Glu Phe Thr Ala Ser Ile Gly Asp Pro Ser Gly Lys Lys Ser Glu Arg
      100      105      110
Gly Leu Leu Ala Gly Asp Val Ile Met His Asn Ser Arg Lys Val Thr
      115      120      125
Asp Gln Ile Cys Lys Ile Phe Glu Asn Ala Pro Gly Ser Ser Glu Lys
      130      135      140
Pro Ile Ile Val Asn Asn Asn Asp Trp Leu Gly Lys Ile Ser Leu Arg
      145      150      155      160
Asp Phe Leu Arg Glu Cys Lys Asn Met Gln Val Gly Lys Met Leu Arg
      165      170      175
Met Asn Thr Ile Lys Asn Arg Leu Glu Val Gly Leu Ser Tyr Thr Glu
      180      185      190
Phe Ser Tyr Gln Thr Met Gln Ala Phe Asp Trp Tyr Thr Leu Ser Glu
      195      200      205
Lys Tyr Gly Cys Arg Phe Gln Leu Gly Gly Tyr Asp Gln Leu Gly His
      210      215      220
Leu Asp Phe Gly Ala His Tyr Ile Lys Lys Met Met Asn Gln Ala Phe
      225      230      235      240
Ala Ala Gly Val Cys Phe Pro Ile Leu Thr Asp Ser Thr Gly Ala Lys
      245      250      255
Leu Gly Lys Ser Glu Gly Gly Gly Ala Leu Trp Leu Asp Ala Thr Lys
      260      265      270
Thr Ser Pro Phe His Phe Tyr Gln Phe Phe Ala Gln Leu His Asp Asp
      275      280      285
Lys Ala Glu Glu Leu Leu Leu Leu Phe Ser Leu Gln Asp Ile Glu His
      290      295      300
Ile Arg Asp Val Leu Lys Asn His Arg Ser Asn Leu Gly Gln Trp Ile
      305      310      315      320
Ala Gln Arg Glu Leu Ala Ala Glu Ile Thr Arg Ile Val His Gly Lys
      325      330      335
Glu Gly Leu Glu Val Ala Met Arg Cys Thr Lys Ala Met Phe Gly Ala
      340      345      350
Lys Lys Ala Asp Leu Ser Gly Leu Ser Arg Ser Glu Val Leu Gln Leu
      355      360      365
Phe Arg Thr Thr Ile Asp Leu Lys Lys Glu Asn Val Ala Thr Met Gly

      370      375      380
Gln Leu Ala Asp Ala Ser Arg Leu Gly Ser Gly Lys Gly His Leu Leu
      385      390      395      400
Met Gln Gln Gly Ala Phe Ser Val Asn Gly Glu Lys Lys Arg Ser Pro
      405      410      415
Ser Glu Ser Ile Ala Asp Val Phe Leu Glu Asn Ala Ser Asp Leu Thr
      420      425      430
Leu Val Cys Trp Gly Lys Arg Gly Tyr Gln Leu Val Arg Trp Val
      435      440      445

```

<210> 57

<211> 412

<212> PRT

<213> *Pseudomonas aeruginosa*

<400> 57

```

Met Ser Val Pro Thr His Gln Gln Asp Leu Ile Ala Leu Leu Glu Glu
 1      5      10      15
Arg Gly Phe Val His Gln Cys Thr Asp Arg Asp Gly Leu Ala Ala His
      20      25      30
Leu Ala Ala Gly Pro Ala Thr Ala Tyr Leu Gly Phe Asp Ala Thr Ala
      35      40      45
Asp Ser Leu His Val Gly His Leu Gln Gly Leu Met Leu Met Arg Trp
 50      55      60
Leu Gln Lys Ala Gly His Arg Pro Leu Leu Leu Ile Gly Gly Ala Thr
 65      70      75      80
Thr Arg Ile Gly Asp Pro Ser Phe Arg Asp Ser Ser Arg Pro Ile Leu
      85      90      95
Thr Glu Ala Gln Ile Gln Ala Asn Ile Asp Gly Ile Ala Arg Val Phe
      100      105      110
Ser Arg Tyr Val Glu Leu His Asp Asp Ser Leu Val Asn Asn Ala Glu
      115      120      125
Trp Leu Asp Gly Val Gly Tyr Leu Glu Phe Leu Asp Arg Val Gly Arg
      130      135      140
His Phe Ser Ile Asn Arg Leu Thr Phe Asp Ala Ile Arg Gln Arg
      145      150      155      160
Leu Asp Arg Glu His Ser Leu Ser Phe Leu Glu Phe Gly Tyr Thr Leu
      165      170      175
Leu Gln Ala Tyr Asp Phe Val Glu Leu Ser Arg Arg Arg Gly Cys Thr
      180      185      190
Leu Gln Leu Gly Gly Ala Asp Gln Trp Ala Asn Ile Ile Asn Gly Val
      195      200      205
Glu Leu Ser Arg Arg Gln Gly Gly Ala Gln Leu Phe Gly Leu Thr Met
      210      215      220
Pro Leu Leu Ala Thr Ser Asp Gly Arg Lys Met Gly Lys Ser Ala Gln
      225      230      235      240
Gly Ala Val Trp Leu Asn Ala Glu Arg Leu Ala Pro Phe Asp Phe Trp
      245      250      255
Gln Phe Trp Arg Asn Cys Asp Asp Arg Asp Val Gly Arg Phe Leu Ala
      260      265      270
Leu Phe Ser Glu Leu Pro Met Asp Glu Val Arg Arg Leu Gly Ala Leu
      275      280      285
Gln Gly Ala Glu Leu Asn Glu Ala Lys Val Val Leu Ala Asn Ala Ala
      290      295      300
Thr Ala Leu Ala His Gly Glu His Ala Ala Arg Ser Ala Ala Asp Ala
      305      310      315      320
Ala Arg Gly Val Phe Ala Asp Gly Thr Arg Asp Ser Gly Leu Pro Val
      325      330      335
Met Lys Leu Ser Arg Ala Arg Leu Ala Gln Gly Leu Ser Leu Thr Asp
      340      345      350
Leu Leu Leu Glu His Ala Ile Gln Pro Ser Arg Ser Ala Val Arg Arg
      355      360      365
Leu Ala Ala Gly Gly Gly Leu Arg Leu Asp Gly Thr Pro Val Ser Asp
      370      375      380
Pro Asp Thr Pro Leu Ala Gly Glu Val Asp Gly Leu Arg Leu Ser Leu
      385      390      395      400
Gly Lys Lys Gln His Leu His Leu Arg Leu Glu Asp
      405      410

```

<210> 58
 <211> 399
 <212> PRT
 <213> *Pseudomonas aeruginosa*
 <400> 58

```

Met Lys Ser Val Glu Glu Gln Leu Ala Leu Ile Gln Arg Gly Ala Asp
 1      5      10      15
Glu Ile Leu Val Glu Ala Glu Leu Val Ala Lys Leu Lys Arg Gly Gln
 20      25      30
Pro Leu Arg Ile Lys Ala Gly Phe Asp Pro Thr Ala Pro Asp Leu His
 35      40      45
Leu Gly His Thr Val Leu Ile Asn Lys Leu Arg Gln Phe Gln Asp Leu
 50      55      60
Gly His Gln Val Ile Phe Leu Ile Gly Asp Phe Thr Gly Met Ile Gly
 65      70      75
Asp Pro Ser Gly Lys Ser Val Thr Arg Pro Pro Leu Thr Arg Glu Gln
 85      90      95
Val Leu Glu Asn Ala Glu Thr Tyr Lys Ser Gln Val Phe Lys Ile Leu
100      105      110
Asp Pro Ala Lys Thr Glu Val Ala Phe Asn Ser Thr Trp Met Asp Gln
115      120      125
Leu Thr Pro Ala Asp Phe Ile Arg Leu Ala Ser Gln Tyr Thr Val Ala
130      135      140
Arg Met Leu Glu Arg Asp Asp Phe Ser Lys Arg Tyr Ala Ser Asn Gln
145      150      155
Pro Ile Ala Ile His Glu Phe Leu Tyr Pro Leu Val Gln Gly Tyr Asp
165      170      175
Ser Val Ala Leu Lys Ala Asp Val Glu Leu Gly Gly Thr Asp Gln Lys
180      185      190
Phe Asn Leu Leu Met Gly Arg Glu Leu Gln Arg Ala Tyr Gly Gln Glu
195      200      205
Ala Gln Val Ile Leu Thr Met Pro Leu Leu Glu Gly Leu Asp Gly Val
210      215      220
Lys Lys Met Ser Lys Ser Leu Gly Asn Tyr Ile Gly Ile Gln Glu Ala
225      230      235
Pro Gly Val Met Tyr Ser Lys Leu Val Ser Ile Pro Asp Thr Leu Met
245      250      255
Trp Arg Tyr Phe Glu Leu Leu Ser Phe Arg Ser Leu Asp Glu Ile Asp
260      265      270
Ser Phe Arg Lys Asp Val Glu Ala Gly Ala Aon Pro Arg Asp Ile Lys
275      280      285
Ile Lys Leu Ala Glu Glu Ile Val Ala Arg Phe His Gly Glu Glu Ala
290      295      300
Ala Ala Ser Ala His Lys Ser Ala Gly Asn Arg Leu Lys Glu Gly Glu

305      310      315      320
Leu Pro Glu Asp Leu Pro Glu Ile Glu Leu Ser Ser Pro Glu Asp Met
325      330      335
Pro Val Ala Ser Val Leu Asn Lys Ala Gly Leu Val Lys Asn Ala Ala
340      345      350
Ala Ala Arg Asp Leu Leu Gly Ala Gly Ser Val Lys Val Asp Gly Gln
355      360      365
Val Val Asp Arg Thr Phe Met Leu Ala Leu Gly Glu Thr Arg Val Phe
370      375      380
Gln Ala Gly Lys Lys Ala Phe Ala Arg Ile Thr Leu Lys Ala Glu
385      390      395

```

<210> 59
 <211> 334

<212> PRT

<213> Escherichia coli

<400> 59

```

Met Thr Lys Pro Ile Val Phe Ser Gly Ala Gln Pro Ser Gly Glu Leu
 1      5      10      15
Thr Ile Gly Asn Tyr Met Gly Ala Leu Arg Gln Trp Val Asn Met Gln
 20      25      30
Asp Asp Tyr His Cys Ile Tyr Cys Ile Val Asp Gln His Ala Ile Thr
 35      40      45
Val Arg Gln Asp Ala Gln Lys Leu Arg Lys Ala Thr Leu Asp Thr Leu
 50      55      60
Ala Leu Tyr Leu Ala Cys Gly Ile Asp Pro Glu Lys Ser Thr Ile Phe
 65      70      75      80
Val Gln Ser His Val Pro Glu His Ala Gln Leu Gly Trp Ala Leu Asn
 85      90      95
Cys Tyr Thr Tyr Phe Gly Glu Leu Ser Arg Met Thr Gln Phe Lys Asp
100      105      110
Lys Ser Ala Arg Tyr Ala Glu Asn Ile Asn Ala Gly Leu Phe Asp Tyr
115      120      125
Pro Val Leu Met Ala Ala Asp Ile Leu Leu Tyr Gln Thr Asn Leu Val
130      135      140
Pro Val Gly Glu Asp Gln Lys Gln His Leu Glu Leu Ser Arg Asp Ile
145      150      155      160
Ala Gln Arg Phe Asn Ala Leu Tyr Gly Glu Ile Phe Lys Val Pro Glu
165      170      175
Pro Phe Ile Pro Lys Ser Gly Ala Arg Val Met Ser Leu Leu Glu Pro
180      185      190
Thr Lys Lys Met Ser Lys Ser Asp Asp Asn Arg Asn Asn Val Ile Gly
195      200      205
Leu Leu Glu Asp Pro Lys Ser Val Val Lys Lys Ile Lys Arg Ala Val
210      215      220
Thr Asp Ser Asp Glu Pro Pro Val Val Arg Tyr Asp Val Gln Asn Lys
225      230      235      240
Ala Gly Val Ser Asn Leu Leu Asp Ile Leu Ser Ala Val Thr Gly Gln
245      250      255
Ser Ile Pro Glu Leu Glu Lys Gln Phe Glu Gly Lys Met Tyr Gly His
260      265      270
Leu Lys Gly Glu Val Ala Asp Ala Val Ser Gly Met Leu Thr Glu Leu
275      280      285
Gln Glu Arg Tyr His Arg Phe Arg Asn Asp Glu Ala Phe Leu Gln Gln
290      295      300
Val Met Lys Asp Gly Ala Glu Lys Ala Ser Ala His Ala Ser Arg Thr
305      310      315      320
Leu Lys Ala Val Tyr Glu Ala Ile Gly Phe Val Ala Lys Pro
325      330

```

<210> 60

<211> 326

<212> PRT

<213> Helicobacter pylori

<400> 60

```

Met His Lys Lys Arg Val Phe Ser Gly Ile Gln Pro Thr Gly Gln Ile
 1      5      10      15
His Leu Gly Asn Tyr Leu Gly Ala Ile Lys His Trp Val Glu Met Gln
 20      25      30
Asp Glu Tyr Glu Asn Leu Phe Cys Val Val Asn Ser His Ala Ile Thr
 35      40      45
Leu Pro Ile Asp Pro Ala Phe Leu Lys Ser Gln Ser Tyr Glu Leu Val
 50      55      60
Lys Leu Leu Leu Ala Cys Gly Ile Asp Pro Lys Gln Ser Gly Leu Phe
 65      70      75      80
Ile Gln Ser Glu Val Asp Glu His Pro Ala Leu Ala Trp Leu Leu Asn
 85      90      95
Cys Gln Val Ser Met Gly Glu Met Gln Arg Met Thr Gln Phe Lys Asp
100      105      110
Lys Ser Leu Lys Asn Pro Lys Ser Val Asn Val Gly Leu Phe Asn Tyr
115      120      125
Pro Ile Leu Met Ala Ser Asp Ile Leu Leu Tyr Gln Ser Asp Leu Val
130      135      140
Pro Val Gly Glu Asp Gln Lys Gln His Leu Glu Leu Thr Arg Asn Ile
145      150      155      160
Ala Glu Lys Phe Asn Arg Asp Phe Gly Asn Cys Phe Lys Val Pro Glu
165      170      175
Pro Leu Ile Ala Lys Val Gly Ala Arg Val Met Gly Leu Asp Asp Pro
180      185      190
Lys Val Lys Met Ser Lys Ser His Gln Gly Ala Asn His Ala Ile Phe
195      200      205
Leu Leu Asp Glu Pro Asp Ile Ile Val Lys Lys Ile Lys Lys Ala Ala
210      215      220
Thr Asp Ser Met Gly Val Ile Ala Phe Asp Glu Lys Arg Glu Gly Ile
225      230      235      240
Phe Asn Leu Leu Asn Ile Tyr Met Leu Leu Ser Asn Glu Ser Pro Glu
245      250      255
Asn Ile Glu Glu Arg Phe Lys Asn Lys Gly Tyr Gly Asp Phe Lys Lys
260      265      270
Glu Leu Ala Glu Val Met Ile Gln Ala Leu Lys Pro Ile Gln Glu Arg
275      280      285
Tyr Lys Glu Ile Ser Asp Asp Glu Val Lys Ala Ile Leu Asn Gly Gly
290      295      300
Ala Glu Lys Ala Arg Pro Leu Ala Arg Ala Thr Tyr Gln Lys Ala Lys
305      310      315      320
Glu Leu Met Gly Leu Ile
325

```

<210> 61

<211> 361

<212> PRT

<213> Schizosaccharomyces pombe

<400> 61


```

Met Ala Lys Leu Pro Lys Ile Thr Ser Leu Leu Pro His Ser Arg Val
 1          5          10          15
Val Ser Gly Ile Gln Pro Thr Gly Ile Pro His Ile Gly Asn Tyr Leu
 20          25          30
Gly Ser Leu Lys Gln Trp Val Gln Leu Gln Glu Glu Ala Ala Arg Thr
 35          40          45
Pro Phe Ser Lys Cys Phe Phe Phe Val Ala Asp Leu His Ala Leu Thr
 50          55          60
Val Pro Gln Asp Pro Leu Lys Phe Arg Gln Ala Arg Leu Asp Met Leu
 65          70          75          80
Ala Ala Leu Leu Ala Ile Gly Ile Asn Pro Gln Lys Ser Thr Leu Phe
 85          90          95
Phe Gln Ser Asp Val Ala Gln His Ser Glu Leu Ala Trp Leu Leu Ala
100          105          110
Cys Ser Thr Ser Met Gly Gln Leu Asn Arg Met Thr Gln Trp Lys Ser
115          120          125
Lys Leu His Leu His Asp His Asp Asp Leu Ser Phe Leu Asp Ala Ser
130          135          140
Ala Thr Ser Ser Thr Arg Phe Asn Leu Gly Leu Phe Ser Tyr Pro Val
145          150          155          160
Leu Gln Ala Ala Asp Ile Leu Leu Tyr Gly Ala Thr His Ile Pro Val
165          170          175
Gly Lys Asp Gln Ser Gln His Val Glu Leu Thr Arg Ser Ile Ala Arg
180          185          190
Ser Phe Asn Ser Ser Tyr Lys Glu Lys Ile Leu Thr Val Pro Asp Ile
195          200          205
Ile Leu Asn Ser Ser Ser Ser Ile Met Ala Leu Cys Gln Pro Glu Lys
210          215          220
Lys Met Ser Lys Ser Asp Ile Asn Ser Lys Asn Tyr Ile Leu Leu Ser
225          230          235          240
Asp Ser Thr Gly Glu Ile Arg Lys Lys Ile Ser Arg Ala Gln Thr Asp
245          250          255
Asn Ile Lys Gly Ile Thr Tyr Gly Asp Ser Asn Arg Pro Gly Ile Asn
260          265          270
Asn Leu Ile Asn Ile Phe Ala Ala Ile Ser Asp Ser Thr Pro Ser Asp
275          280          285
Ile Ala Gln Ala Asn Ala Ser Cys Ser Asn Ala Glu Phe Lys Glu Lys
290          295          300
Val Ser Ser Ala Ile Ile Arg Cys Leu Gln Pro Ile Ser Thr Ser Phe
305          310          315          320
Asn Glu Trp Arg Gln Asn Arg Glu Leu Leu Arg Asp Ile Ala Lys Lys
325          330          335
Gly Ala Glu Glu Ala Val Ala Glu Ala Ser Ser Cys Met His Lys Leu
340          345          350
Lys Thr Leu Thr Gly Leu Ser Val Tyr
355          360

```

<210> 62

<211> 665

<212> PRT

<213> Drosophila melanogaster

<400> 62

Met Phe Arg Phe Gly Lys Leu His Arg Ser Leu Ser Lys Val Ala Lys

1	5	10	15
Lys Gly Gln Thr Ala Val Tyr Ser ¹ Gln Arg Arg Cys Ile Gln Ala Arg			
	20	25	30
Ser Phe Leu Pro Gly Leu Ala Ser Arg Thr Ala Gly Thr Pro Ser Ala			
	35	40	45
Thr Ala Gln Val Thr Gly Gly Asn Lys Gln His Val His Pro His Gly			
	50	55	60
Asn Ser Val Asn Ser Gly His Glu Glu His Asn Thr Arg Trp Pro Arg			
65	70	75	80
Lys Val Phe Ser Gly Ile Gln Pro Thr Gly Ser Leu His Leu Gly Asn			
	85	90	95
Tyr Leu Gly Ala Val Arg Lys Trp Val Gln Leu Gln Asn Ala Arg Asp			
	100	105	110
Asp Val Thr Val Cys Ile Val Asp Leu His Ser Ile Thr Met Pro His			
	115	120	125
Asn Pro Pro Leu Leu Arg Glu Asn Ile Phe Thr Met Ala Thr Leu			
	130	135	140
Leu Ala Cys Gly Ile Asp Pro Thr Lys Ser Thr Leu Phe Val Gln Ser			
145	150	155	160
Ala Val Ala Glu His Ala Glu Phe Asn Trp Ile Leu Ser Ser Leu Thr			
	165	170	175
Thr Met Pro Arg Leu Ala Gln Leu Pro Gln Phe Arg Glu Lys Ser Arg			
	180	185	190
Leu Leu Lys Asp Val Pro Leu Gly Leu Tyr Val Tyr Pro Val Leu Gln			
	195	200	205
Ala Ala Asp Ile Met Leu Tyr Lys Ser Thr His Val Pro Val Gly Ala			
	210	215	220
Asp Gln Ile Gln His Ile Gln Leu Ala Gln His Leu Ala Arg Ile Tyr			
225	230	235	240
Asn Gly Arg Tyr Gly Glu Thr Phe Pro Val Cys Thr Ala Ile Ile Glu			
	245	250	255
Asp Gly Asp Ala Ser Arg Val Leu Ser Leu Arg Asp Pro Ser Lys Lys			
	260	265	270
Met Ser Lys Ser Glu Ala Asn Pro Lys Ala Thr Ile Asn Leu Cys Asp			
	275	280	285
Ser Pro Asp Leu Ile Thr Gln Lys Ile Lys Lys Ala Val Thr Asp Phe			
	290	295	300
Thr Ser Asp Ile Thr Tyr Asn Pro Gly Lys Arg Ala Gly Val Ser Asn			
	305	310	315
Leu Val Asn Ile His Ala Gln Val Thr Gly Gln Ser Ile Lys Thr Val			
	325	330	335
Val Asn Glu Ala Ala Thr Leu Asp Thr Ala Lys Tyr Lys Asp Arg Val			
	340	345	350
Ala Glu Ala Val Val Glu His Leu Arg Pro Ile Arg Glu Gln Ile His			
	355	360	365
His His Met Thr Lys Arg Asn Glu Met Ile Tyr Leu Leu Glu Val Gly			
	370	375	380
Ala Glu Lys Ala Arg Gln Gln Ala Arg Gln Thr Leu Asn Asp Val Lys			
385	390	395	400
Gln Arg Leu Gly Leu Gly Thr Ser Ala Asn Ile Pro Ala Ala Val His			
	405	410	415
Val Ala Pro Leu Leu Pro Ala Pro Asp Ile Ser Lys Thr Ser Ala Ser			
	420	425	430
Arg Arg Met Ser Lys Asp Phe Asp Gly Asn Val His Glu Arg His Glu			
	435	440	445
Arg Met Tyr Gly Phe Gly Asp Gln Pro Ser Gly Gly Ala Gly Gln Ser			
	450	455	460
Ala Thr Gly Glu Val Ala Asp Met Pro Gln Ala Val Val Pro Arg Val			

```

465          470          475          480
Ile Arg Arg Ala Pro Ile Val Pro Thr Gln Ser Met Arg Val Glu Lys
          485          490          495
Gln Leu Gly Ser Thr Arg Thr Arg His Val Phe Gln Met Asp Thr Cys
          500          505          510
Asn Ala Pro His Ile Gly Phe Lys Pro Pro Pro Gly Phe Ser Ala Ser
          515          520          525
Met Val Thr Gly Ala Leu Ala Lys Asn Lys Arg Ala Thr Val Lys Pro
          530          535          540
Val Thr Gly Asn Leu Gly Phe Gly Gln Gln Ser Gln Gly Gly Ser Tyr
545          550          555          560
Lys His Ser Gln Asn Val Asn Pro Ser Ala Val Ser Thr Ile Ala Arg
          565          570          575
Glu Val Asn Ala Ala Lys Lys Gln Glu Phe Thr Cys Asn Pro Ala Leu
          580          585          590
Tyr Gly Ser Ala Leu Glu Ser Leu Ser Asn Ser Ser Arg Ser Ser Glu
          595          600          605
Lys Thr Asn Ser Thr Ala Gly Asp Ser Glu Phe Ile Ile Val Ala Ser
          610          615          620
Glu Glu Glu Glu Gln Ala Ala Ser Ser Glu Glu Gln Arg Glu Glu Ala
625          630          635          640
Glu Glu Glu Glu Ser Gly Glu Arg Glu Arg Glu Lys Gly Glu Arg Asn
          645          650          655
Ala Thr Glu Ala Thr Lys Val Glu Ile
          660          665

```

<210> 63

<211> 360

<212> PRT

<213> Homo sapiens

<400> 63

```

Met Ala Leu His Ser Met Arg Lys Ala Arg Glu Arg Trp Ser Phe Ile
 1          5          10          15
Arg Ala Leu His Lys Gly Ser Ala Ala Pro Ala Leu Gln Lys Asp
      20          25          30
Ser Lys Lys Arg Val Phe Ser Gly Ile Gln Pro Thr Gly Ile Leu His
      35          40          45
Leu Gly Asn Tyr Leu Gly Ala Ile Glu Ser Trp Val Arg Leu Gln Asp
      50          55          60
Glu Tyr Asp Ser Val Leu Tyr Ser Ile Val Asp Leu His Ser Ile Thr
      65          70          75          80
Val Pro Gln Asp Pro Ala Val Leu Arg Gln Ser Ile Leu Asp Met Thr
      85          90          95
Ala Val Leu Leu Ala Cys Gly Ile Asn Pro Glu Lys Ser Ile Leu Phe
      100          105          110
Gln Gln Ser Gln Val Ser Glu His Thr Gln Leu Ser Trp Ile Leu Ser
      115          120          125
Cys Met Val Arg Leu Pro Arg Leu Gln His Leu His Gln Trp Lys Ala
      130          135          140
Lys Thr Thr Lys Gln Lys His Asp Gly Thr Val Gly Leu Leu Thr Tyr
      145          150          155          160
Pro Val Leu Gln Ala Ala Asp Ile Leu Leu Tyr Lys Ser Thr His Val
      165          170          175
Pro Val Gly Glu Asp Gln Val Gln His Met Glu Leu Val Gln Asp Leu
      180          185          190
Ala Gln Gly Phe Asn Lys Lys Tyr Gly Glu Phe Phe Pro Val Pro Glu
      195          200          205
Ser Ile Leu Thr Ser Met Lys Lys Val Lys Ser Leu Arg Asp Pro Ser
      210          215          220
Ala Lys Met Ser Lys Ser Asp Pro Asp Lys Leu Ala Thr Val Arg Ile
      225          230          235          240
Thr Asp Ser Pro Glu Glu Ile Val Gln Lys Phe Arg Lys Ala Val Thr
      245          250          255
Asp Phe Thr Ser Glu Val Thr Tyr Asp Pro Ala Gly Arg Ala Gly Val
      260          265          270
Ser Asn Ile Val Ala Val His Ala Ala Val Thr Gly Leu Ser Val Glu
      275          280          285
Glu Val Val Arg Arg Ser Ala Gly Met Asn Thr Ala Arg Tyr Lys Leu
      290          295          300
Ala Val Ala Asp Ala Val Ile Glu Lys Phe Ala Pro Ile Lys Arg Glu
      305          310          315          320
Ile Glu Lys Leu Lys Leu Asp Lys Asp His Leu Glu Lys Val Leu Gln
      325          330          335
Ile Gly Ser Ala Lys Ala Lys Glu Leu Ala Tyr Thr Val Cys Gln Glu
      340          345          350
Val Lys Lys Leu Val Gly Phe Leu
      355          360

```

<210> 64

<211> 379

<212> PRT

<213> *Saccharomyces cerevisiae*

<400> 64

```

Met Ser Asn Lys Gln Ala Val Leu Lys Leu Ile Ser Lys Arg Trp Ile
 1      5      10      15
Ser Thr Val Gln Arg Ala Asp Phe Lys Leu Asn Ser Glu Ala Leu His
 20      25      30
Ser Asn Ala Thr Val Phe Ser Met Ile Gln Pro Thr Gly Cys Phe His
 35      40      45
Leu Gly Asn Tyr Leu Gly Ala Thr Arg Val Trp Thr Asp Leu Cys Glu
 50      55      60
Leu Lys Gln Pro Gly Gln Glu Leu Ile Phe Gly Val Ala Asp Leu His
 65      70      75      80
Ala Ile Thr Val Pro Lys Pro Asp Gly Glu Met Phe Arg Lys Phe Arg
 85      90      95
His Glu Ala Val Ala Ser Ile Leu Ala Val Gly Val Asp Pro Glu Lys
100      105      110
Ala Ser Val Ile Tyr Gln Ser Ala Ile Pro Gln His Ser Glu Leu His
115      120      125
Trp Leu Leu Ser Thr Leu Ala Ser Met Gly Leu Leu Asn Arg Met Thr
130      135      140
Gln Trp Lys Ser Lys Ser Asn Ile Lys Gln Ser Thr Asn Gly Asp Tyr
145      150      155      160
Leu Val Asn Asp Ser Asp Val Gly Lys Val Arg Leu Gly Leu Phe Ser
165      170      175
Tyr Pro Val Leu Gln Ala Ala Asp Ile Leu Leu Tyr Lys Ser Thr His
180      185      190
Val Pro Val Gly Asp Asp Gln Ser Gln His Leu Glu Leu Thr Arg His
195      200      205
Leu Ala Glu Lys Phe Asn Lys Met Tyr Lys Lys Asn Phe Phe Pro Lys
210      215      220
Pro Val Thr Met Leu Ala Gln Thr Lys Lys Val Leu Ser Leu Ser Thr
225      230      235      240
Pro Glu Lys Lys Met Ser Lys Ser Asp Pro Asn His Asp Ser Val Ile
245      250      255
Phe Leu Asn Asp Glu Pro Lys Ala Ile Gln Lys Lys Ile Arg Lys Ala
260      265      270
Leu Thr Asp Ser Ile Ser Asp Arg Phe Tyr Tyr Asp Pro Val Glu Arg
275      280      285
Pro Gly Val Ser Asn Leu Ile Asn Ile Val Ser Gly Ile Gln Arg Lys
290      295      300
Ser Ile Glu Asp Val Val Glu Asp Val Ser Arg Phe Asn Asn Tyr Arg
305      310      315      320
Asp Phe Lys Asp Tyr Val Ser Glu Val Ile Ile Glu Glu Leu Lys Gly
325      330      335
Pro Arg Thr Glu Phe Glu Lys Tyr Ile Asn Glu Pro Thr Tyr Leu His
340      345      350
Ser Val Val Glu Ser Gly Met Arg Lys Ala Arg Glu Lys Ala Ala Lys
355      360      365
Asn Leu Ala Asp Ile His Lys Ile Met Gly Phe
370      375

```

<210> 65

<211> 475

<212> PRT

<213> *Xenopus laevis*

<400> 65

```

Met Ser Asp Ser Gly Thr Ser Thr Ser Gly Ser Leu Leu Pro Met Asp
1      5      10      15
Leu Tyr Asn Gln Val Thr Ala Gln Gly Asp Lys Ile Arg Val Leu Lys
20      25      30
Ser Glu Lys Ser Pro Lys Glu Glu Ile Asp Ala Ala Val Lys Leu Leu
35      40      45
Leu Ala Leu Lys Val Asp Tyr Lys Asn Val Thr Gly Gln Asp Tyr Lys
50      55      60
Pro Gly Val Pro Pro Ala Asp Asp Met Pro Thr Asn Arg Asn Gly Pro
65      70      75      80
Ser Thr Pro Asn Asp Gly Asp Asp Phe Val Asp Pro Trp Thr Val Gln
85      90      95
Thr Gly Ser Ala Lys Gly Val Asp Tyr Asp Lys Leu Ile Val Arg Phe
100     105     110
Gly Ser Ser Lys Ile Asp Gln Ala Leu Ile Asp Arg Ile Glu Arg Val
115     120     125
Thr Gly Gln Lys Ala His His Phe Leu Arg Arg Gly Ile Phe Phe Ser
130     135     140
His Arg Asp Met His Gln Val Leu Asp Ala Tyr Glu Lys Lys Lys Pro
145     150     155     160
Phe Tyr Leu Tyr Thr Gly Arg Gly Pro Ser Ser Glu Ala Met His Val
165     170     175
Gly His Leu Ile Pro Phe Ile Phe Thr Lys Trp Leu Gln Glu Val Phe
180     185     190
Asn Val Pro Leu Val Val Gln Leu Thr Asp Asp Glu Lys Tyr Leu Trp
195     200     205
Lys Asp Leu Thr Leu Glu Lys Ala Tyr Gln Tyr Ala Thr Glu Asn Ala
210     215     220
Lys Asp Ile Ile Ala Cys Gly Phe Asp Val Asn Lys Thr Phe Ile Phe
225     230     235     240
Ser Asp Leu Glu Tyr Met Gly Lys Ser Ser Gly Phe Tyr Gln Asn Val

245      250      255
Val Lys Val Gln Lys His Val Thr Phe Asn Gln Val Lys Gly Ile Phe
260      265      270
Gly Phe Thr Asp Ser Asp Cys Ile Gly Lys Ile Ser Phe Pro Ala Ile
275      280      285
Gln Ala Ala Pro Ser Phe Ser Ser Ser Phe Pro Glu Ile Phe Lys Gly
290      295      300
Arg Lys Asp Ile Gln Cys Leu Ile Pro Cys Ala Ile Asp Gln Asp Pro
305      310      315     320
Tyr Phe Arg Met Thr Arg Asp Val Ala Pro Arg Ile Asn Tyr Pro Lys
325      330      335
Pro Ala Leu Met His Ser Thr Phe Phe Pro Ala Leu Gln Gly Ala Gln
340      345      350
Thr Lys Met Ser Ala Ser Asp Pro Asn Ser Ser Ile Phe Leu Thr Asp
355      360      365
Thr Ala Lys Gln Ile Lys Ser Lys Ile Asn Lys His Ala Phe Ser Gly
370      375      380
Gly Lys Asp Thr Ile Glu Glu His Arg Gln Phe Gly Gly Asn Trp Glu
385      390      395     400
Val Asp Val Ser Tyr Met Tyr Leu Thr Phe Phe Leu Glu Asp Asp Glu
405      410      415
Arg Leu Glu Gln Ile Lys Gln Asp Tyr Ser Ser Gly Ala Leu Leu Thr
420      425      430
Gly Asp Leu Lys Lys Ile Leu Thr Glu Thr Leu Gln Pro Met Ile Ser
435      440      445
Ala His Gln Glu Arg Arg Lys His Ile Thr Glu Glu Thr Val Lys Gln
450      455      460
Phe Met Met Pro Arg Lys Leu Ala Phe Asp Phe
465      470      475

```

<210> 66

<211> 417

<212> PRT

<213> *Caenorhabditis elegans*

<400> 66

```

Met Ala Ala Pro Ala Glu Gln Val Ala Ala Glu Ile Glu Asn Leu Lys
 1          5          10          15
Val Asn Gly Gly Ala Ala Gly Gly Gly Val Gln Glu Asp Glu Asp
 20          25          30
Arg Val Thr Pro Trp Glu Val Thr Thr Thr Lys Ala Thr Gly Ile Asp
 35          40          45
Tyr Asp Lys Leu Ile Val Lys Phe Gly Cys Arg Lys Leu Asp Glu Glu
 50          55          60
Ile Ile Ala Arg Phe Glu Arg Val Thr Gly His Lys Ala Ser Pro Met
 65          70          75
Leu Arg Arg Gly Met Phe Phe Ala His Arg Asp Leu Thr Ala Ile Leu
 85          90          95
Asp Arg Lys Glu Gln Gly Lys Pro Phe Tyr Leu Tyr Thr Gly Arg Gly
100          105          110
Ala Ser Ser Gly Ser Leu His Leu Gly His Leu Val Pro Phe Ile Phe
115          120          125
Thr Lys Trp Leu Gln Glu Val Phe Asp Val Pro Leu Val Ile Gln Met
130          135          140
Thr Asp Asp Glu Lys Phe Leu Trp Lys Asp Met Lys Val Asp Glu Ala
145          150          155          160
Lys Lys Met Ala Arg Glu Asn Met Lys Asp Ile Ile Ser Val Gly Phe

```

```

          165          170          175
Asp Pro Thr Lys Thr Phe Ile Phe Asn Asn Phe Asp Tyr Met Cys Pro
          180          185          190
Pro Phe Tyr Glu Asn Ile Val Lys Ile Trp Lys Val Val Asn Thr Asn
          195          200          205
Gln Ala Arg Ala Ile Phe Gly Phe Thr Pro Glu Asp Cys Leu Gly Lys
210          215          220
Ala Ala Phe Pro Ala Val Glu Ala Ala Pro Cys Phe Ala Ser Ser Phe
225          230          235          240
Pro Gln Ile Phe Gly Lys Arg Asn Asp Ile Pro Cys Leu Ile Pro Cys
          245          250          255
Ala Ile Asp Gln Asp Pro Phe Phe Arg Met Thr Arg Asp Val Ala Pro
260          265          270
Arg Leu Lys Ala Ser Lys Pro Ser Leu Ile Phe Ser Thr Phe Leu Pro
275          280          285
Ala Leu Thr Gly Ala Gln Thr Lys Met Ser Ala Ser Glu Pro Asn Thr
290          295          300
Cys Ile Phe Leu Ser Asp Thr Ala Lys Gln Ile Lys Asn Lys Ile Asn
305          310          315          320
Lys Tyr Ala Phe Ser Gly Gly Gln Gln Thr Val Gln Glu His Arg Glu
          325          330          335
Lys Gly Gly Asn Cys Asp Val Asp Ile Ser Tyr Gln Phe Leu Arg Phe
          340          345          350
Phe Leu Asp Asp Asp Glu Lys Leu Ala Glu Ile Arg Glu Asn Tyr Thr
355          360          365
Lys Gly Glu Met Leu Ser Gly Glu Leu Lys Ala Leu Ala Thr Gln Lys
370          375          380
Val Gln Glu Ile Val Leu Glu Met Gln Glu Arg Arg Lys Leu Val Thr
385          390          395          400
Asp Glu Thr Val Glu Glu Phe Val Lys Val Arg Pro Leu Ala Tyr Lys
          405          410          415
Tyr

```

<210> 67

<211> 337

<212> PRT

<213> Salmonella enterica subsp. enterica serovar Typhi8

<400> 67

```

Met Asn Thr Leu Pro Ile Ile Leu Thr Gly Asp Arg Pro Thr Gly Ser
 1      5      10      15
Leu His Leu Gly His Tyr Val Gly Ser Leu Arg Gln Arg Val Ala Leu
 20      25      30
Gln Gln Asp His Gln Gln Tyr Val Leu Ile Ala Asp Leu Gln Gly Leu
 35      40      45
Thr Asp Asn Gly Ser Asn Pro Gln Lys Ile Arg Asp Asn Ile Pro Gln
 50      55      60
Val Leu Ala Asp Tyr Leu Ala Val Gly Ile Asp Pro Ala Leu Thr Thr
 65      70      75      80
Ile Cys Leu Gln Ser Ala Leu Pro Ala Leu Ala Glu Leu Thr Val Leu
 85      90      95
Tyr Met Asn Ile Val Thr Val Ala Arg Val Glu Arg Asn Pro Thr Val
100      105      110
Lys Asn Glu Ile Ala Gln Lys Gly Phe Thr Arg Ser Leu Pro Val Gly
115      120      125
Phe Met Ala Tyr Pro Ile Ser Gln Ala Ala Asp Ile Thr Ala Phe Lys

130      135      140
Ala Glu Met Val Pro Val Gly Asp Asp Gln Leu Pro Met Ile Glu Gln
145      150      155      160
Thr Asn Glu Ile Val His Lys Met Asn Ser Leu Phe Ser Ser Pro Val
165      170      175
Leu Arg Pro Cys Gln Ala Leu Leu Ser Asp Thr Gly Arg Leu Pro Gly
180      185      190
Ile Asp Gly Ser Ala Lys Met Ser Lys Ser Leu Val Asn Thr Leu Leu
195      200      205
Leu Ser Ala Ser Glu Glu Thr Ile His Arg Ala Val Ser Ala Met Tyr
210      215      220
Thr Asp Pro Asn His Leu Lys Ile Ser Asp Pro Gly Lys Ile Glu Gly
225      230      235      240
Asn Val Val Phe Thr Trp Leu Asp Ala Phe His Pro Asp Lys Ala Lys
245      250      255
Val Ala Ala Met Lys Ala His Tyr Gln Gln Gly Gly Leu Gly Asp Arg
260      265      270
Val Cys Lys Asn Glu Leu Glu Thr Cys Leu Gln Glu Leu Ile Ala Pro
275      280      285
Ile Arg Glu Arg Arg Ala Thr Phe Ile Ala Asp Lys Gly Met Leu Met
290      295      300
Glu Leu Leu Lys Lys Gly Ser Glu Arg Ala His Glu Val Thr Gln Lys
305      310      315      320
Thr Leu Gln Glu Val Lys Arg Gly Leu Gly Leu Pro Thr Leu Phe Gln
325      330      335
Val

```

<210> 68

<211> 448

<212> PRT

<213> Pseudomonas aeruginosa

<400> 68


```

Met Thr Thr Arg Ile Leu Thr Gly Ile Thr Pro Thr Gly Thr Pro His
1      5      10      15
Leu Gly Asn Tyr Ala Gly Ala Ile Arg Pro Ala Ile Leu Ala Ser Arg
20      25      30
Arg Ser Asp Val Asp Ser Phe Tyr Phe Leu Ala Asp Tyr His Ala Leu
35      40      45
Ile Lys Cys Asp Asp Pro Ala Arg Ile Gln Arg Ser Arg Leu Glu Ile
50      55      60
Ala Ala Thr Trp Leu Ala Gly Gly Leu Asp Val Glu Arg Ala Thr Phe
65      70      75      80
Tyr Arg Gln Ser Asp Ile Pro Glu Ile Pro Glu Leu Thr Trp Leu Leu
85      90      95
Thr Cys Val Ser Ala Lys Gly Leu Leu Asn Arg Ala His Ala Tyr Lys
100     105     110
Ala Ala Val Asp Arg Asn Val Glu Ala Gly Glu Asp Pro Asp Ala Gly
115     120     125
Val Thr Met Gly Leu Tyr Ser Tyr Pro Val Leu Met Ala Ala Asp Ile
130     135     140
Leu Met Phe Asn Ala His Lys Ile Pro Val Gly Arg Asp Gln Val Gln
145     150     155     160
His Val Glu Met Ala Arg Asp Ile Gly Gln Arg Phe Asn His Leu Phe
165     170     175     180
Gly Asn Gly Arg Glu Phe Phe Val Leu Pro Glu Ala Val Ile Glu Glu

180     185     190
Asn Val Ala Thr Leu Pro Gly Leu Asp Gly Arg Lys Met Ser Lys Ser
195     200     205
Tyr Asp Asn Thr Ile Pro Leu Phe Ser Pro Ser Arg Gln Leu Lys Asp
210     215     220
Ala Ile Ala Arg Ile Val Thr Asp Ser Arg Ala Pro Gly Glu Pro Lys
225     230     235     240
Asp Pro Asp Ser Ser His Leu Phe Leu Leu Tyr Ser Ala Phe Ala Ser
245     250     255
Ala Glu Gln Val Ala Ala Phe Arg Gln Glu Leu Leu Glu Gly Leu Ala
260     265     270
Trp Gly Glu Ala Lys Gln Arg Leu Phe Gln Leu Leu Asp Asn Glu Leu
275     280     285
Gly Glu Ala Arg Glu Arg Tyr Gln Ala Leu Ile Ala Lys Pro Asp Asp
290     295     300
Ile Glu Asp Ile Leu Leu Ala Gly Ala Ala Lys Ala Arg Arg Ile Ala
305     310     315     320
Thr Pro Phe Ile Ala Glu Leu Arg Glu Ala Val Gly Leu Arg Ser Leu
325     330     335
Arg Glu Pro Leu Lys Ser Ala Glu Ser Gly Lys Lys Lys Ala Ala Lys
340     345     350
Ala Ala Arg Leu Val Ser Phe Arg Asp Asp Asp Gly Ser Phe Arg Phe
355     360     365
Arg Leu Leu Asp Ala Ala Gly Glu Gln Leu Leu Ser Arg Ala Phe
370     375     380
Ala Asp Gly Lys Ala Ala Gly Ala Val Ser Lys Arg Leu Leu Ala Gly
385     390     395     400
Glu Thr Ala Asp Leu Arg Ala Glu Gly Asn Ala Phe Gly Leu Trp Leu
405     410     415
Asp Gly Glu Ala Val Ala Gln Ser Pro Ala Phe Ala Asp Ala Ala Ala
420     425     430
Arg Asp Ala Ala Ile Glu Arg Thr Arg Glu Ala Leu Ala Pro Gln Glu
435     440     445

```

<210> 69

<211> 350

<212> PRT

<213> Thermus thermophilus

<400> 69

```

Met Leu Glu Glu Ala Leu Ala Ala Ile Gln Asn Ala Arg Asp Leu Glu
 1      5      10      15
Glu Leu Lys Ala Leu Lys Ala Arg Tyr Leu Gly Lys Lys Gly Leu Leu
 20      25      30
Thr Gln Glu Met Lys Gly Leu Ser Ala Leu Pro Leu Glu Arg Arg
 35      40      45
Lys Arg Gly Gln Glu Leu Asn Ala Ile Lys Ala Ala Leu Glu Ala Ala
 50      55      60
Leu Glu Ala Arg Glu Lys Ala Leu Glu Glu Ala Ala Leu Lys Glu Ala
 65      70      75      80
Leu Glu Arg Glu Arg Val Asp Val Ser Leu Pro Gly Ala Ser Leu Phe
 85      90      95
Ser Gly Gly Leu His Pro Ile Thr Leu Met Glu Arg Glu Leu Val Glu
100      105      110
Ile Phe Arg Ala Leu Gly Tyr Gln Ala Val Glu Gly Pro Glu Val Glu
115      120      125
Ser Glu Phe Phe Asn Phe Asp Ala Leu Asn Ile Pro Glu His His Pro

      130      135      140
Ala Arg Asp Met Trp Asp Thr Phe Trp Leu Thr Gly Glu Gly Phe Arg
145      150      155      160
Leu Glu Gly Pro Leu Gly Glu Glu Val Glu Gly Arg Leu Leu Leu Arg
      165      170      175
Thr His Thr Ser Pro Met Gln Val Arg Tyr Met Val Ala His Thr Pro
      180      185      190
Pro Phe Arg Ile Val Val Pro Gly Arg Val Phe Arg Phe Glu Gln Thr
      195      200      205
Asp Ala Thr His Glu Ala Val Phe His Gln Leu Glu Gly Leu Val Val
210      215      220
Gly Glu Gly Ile Ala Met Ala His Leu Lys Gly Ala Ile Tyr Glu Leu
225      230      235      240
Ala Gln Ala Leu Phe Gly Pro Asp Ser Lys Val Arg Phe Gln Pro Val
      245      250      255
Tyr Phe Pro Phe Val Glu Pro Gly Ala Gln Phe Ala Val Trp Trp Pro
260      265      270
Glu Gly Gly Lys Trp Leu Glu Leu Gly Gly Ala Gly Met Val His Pro
275      280      285
Lys Val Phe Gln Ala Val Asp Ala Tyr Arg Glu Arg Leu Gly Leu Pro
290      295      300
Pro Ala Tyr Arg Gly Val Thr Gly Phe Ala Phe Gly Leu Gly Val Glu
305      310      315      320
Arg Leu Ala Met Leu Arg Tyr Gly Ile Pro Asp Ile Arg Tyr Phe Phe
      325      330      335
Gly Gly Arg Leu Lys Phe Leu Glu Gln Phe Lys Gly Val Leu
      340      345      350

```

<210> 70

<211> 350

<212> PRT

<213> Thermus thermophilus

<400> 70

```

Met Leu Glu Glu Ala Leu Ala Ala Ile Gln Asn Ala Arg Asp Leu Glu
 1      5      10      15
Glu Leu Lys Ala Leu Lys Ala Arg Tyr Leu Gly Lys Lys Gly Leu Leu
 20      25      30
Thr Gln Glu Met Lys Gly Leu Ser Ala Leu Pro Leu Glu Arg Arg
 35      40      45
Lys Arg Gly Gln Glu Leu Asn Ala Ile Lys Ala Ala Leu Glu Ala Ala
 50      55      60
Leu Glu Ala Arg Glu Lys Ala Leu Glu Glu Ala Ala Leu Lys Glu Ala
 65      70      75      80
Leu Glu Arg Glu Arg Val Asp Val Ser Leu Pro Gly Ala Ser Leu Phe
 85      90      95
Ser Gly Gly Leu His Pro Ile Thr Leu Met Glu Arg Glu Leu Val Glu
100      105      110
Ile Phe Arg Ala Leu Gly Tyr Gln Ala Val Glu Gly Pro Glu Val Glu
115      120      125
Ser Glu Phe Phe Asn Phe Asp Ala Leu Asn Ile Pro Glu His His Pro
130      135      140
Ala Arg Asp Met Trp Asp Thr Phe Trp Leu Thr Gly Glu Gly Phe Arg
145      150      155      160
Leu Glu Gly Pro Leu Gly Glu Glu Val Glu Gly Arg Leu Leu Leu Arg
165      170      175      180
Thr His Thr Ser Pro Met Gln Val Arg Tyr Met Val Ala His Thr Pro

180      185      190
Pro Phe Arg Ile Val Val Pro Gly Arg Val Phe Arg Phe Glu Gln Thr
195      200      205
Asp Ala Thr His Glu Ala Val Phe His Gln Leu Glu Gly Leu Val Val
210      215      220
Gly Glu Gly Ile Ala Met Ala His Leu Lys Gly Ala Ile Tyr Glu Leu
225      230      235      240
Ala Gln Ala Leu Phe Gly Pro Asp Ser Lys Val Arg Phe Gln Pro Val
245      250      255
Tyr Phe Pro Phe Val Glu Pro Gly Ala Gln Phe Ala Val Trp Trp Pro
260      265      270
Glu Gly Gly Lys Trp Leu Glu Leu Gly Gly Ala Gly Met Val His Pro
275      280      285
Lys Val Phe Gln Ala Val Asp Ala Tyr Arg Glu Arg Leu Gly Leu Pro
290      295      300
Pro Ala Tyr Arg Gly Val Thr Gly Phe Ala Phe Gly Leu Gly Val Glu
305      310      315      320
Arg Leu Ala Met Leu Arg Tyr Gly Ile Pro Asp Ile Arg Tyr Phe Phe
325      330      335
Gly Gly Arg Leu Lys Phe Leu Glu Gln Phe Lys Gly Val Leu
340      345      350

```

<210> 71

<211> 344

<212> PRT

<213> *Bacillus subtilis*

<400> 71

```

Met Glu Glu Lys Leu Lys Gln Leu Glu Gln Glu Ala Leu Glu Gln Val
 1      5      10      15
Glu Ala Ala Ser Ser Leu Lys Val Val Asn Asp Ile Arg Val Gln Tyr
 20      25      30
Leu Gly Lys Lys Gly Pro Ile Thr Glu Val Leu Arg Gly Met Gly Lys
 35      40      45
Leu Ser Ala Glu Glu Arg Pro Lys Met Gly Ala Leu Ala Asn Glu Val
 50      55      60
Arg Glu Arg Ile Ala Asn Ala Ile Ala Asp Lys Asn Glu Lys Leu Glu
 65      70      75      80
Glu Glu Glu Met Lys Gln Lys Leu Ala Gly Gln Thr Ile Asp Val Thr
 85      90      95
Leu Pro Gly Asn Pro Val Ala Val Gly Gly Arg His Pro Leu Thr Val
100      105      110
Val Ile Glu Glu Ile Glu Asp Leu Phe Ile Gly Met Gly Tyr Thr Val
115      120      125
Glu Glu Gly Pro Glu Val Glu Thr Asp Tyr Tyr Asn Phe Glu Ser Leu
130      135      140
Asn Leu Pro Lys Glu His Pro Ala Arg Asp Met Gln Asp Ser Phe Tyr
145      150      155      160
Ile Thr Glu Glu Thr Leu Met Arg Thr Gln Thr Ser Pro Val Gln Thr
165      170      175
Arg Thr Met Glu Lys His Glu Gly Lys Gly Pro Val Lys Ile Ile Cys
180      185      190
Pro Gly Lys Val Tyr Arg Arg Asp Asn Asp Asp Ala Thr His Ser His
195      200      205
Gln Phe Met Gln Ile Glu Gly Leu Val Val Asp Lys Asn Ile Ser Met
210      215      220
Ser Asp Leu Lys Gly Thr Leu Glu Leu Val Ala Lys Lys Met Phe Gly
225      230      235      240
Gln Asp Arg Glu Ile Arg Leu Ala Pro Ser Phe Phe Pro Phe Thr Glu
245      250      255
Pro Ser Val Glu Val Asp Val Thr Cys Phe Lys Cys Gly Gly Asn Gly
260      265      270
Cys Ser Val Cys Lys Gly Thr Gly Trp Ile Glu Ile Leu Gly Ala Gly
275      280      285
Met Val His Pro Asn Val Leu Lys Met Ala Gly Phe Asp Pro Lys Glu
290      295      300
Tyr Gln Gly Phe Ala Phe Gly Met Gly Val Glu Arg Ile Ala Met Leu
305      310      315      320
Lys Tyr Gly Ile Asp Asp Ile Arg His Phe Tyr Thr Asn Asp Val Arg
325      330      335
Phe Ile Ser Gln Phe Lys Gln Ala
340

```

<210> 72

<211> 328

<212> PRT

<213> Helicobacter pylori

<400> 72

```

Met His Thr Leu Ile Glu Arg Leu Glu Lys Val Thr Asn Ser Lys Glu
 1      5      10      15
Leu Glu Glu Ala Arg Leu Asn Ala Leu Gly Lys Lys Gly Val Phe Ala
 20      25      30
Asp Lys Phe Asn Gln Leu Lys His Leu Asn Gly Glu Glu Lys Asn Ala
 35      40      45
Phe Ala Lys Glu Ile His His Tyr Lys Gln Ala Phe Glu Lys Ala Phe
 50      55      60
Glu Trp Lys Lys Lys Ala Ile Ile Glu Leu Glu Leu Glu Glu Arg Leu
 65      70      75      80
Lys Lys Glu Lys Ile Asp Val Ser Leu Phe Asn Ala Ile Lys Thr Ser
 85      90      95
Ser Ser His Pro Leu Asn Tyr Thr Lys Asn Lys Ile Ile Glu Phe Phe
100      105      110
Thr Pro Leu Gly Tyr Lys Leu Glu Ile Gly Ser Leu Val Glu Asp Asp
115      120      125
Phe His Asn Phe Ser Ala Leu Asn Leu Pro Pro Tyr His Pro Ala Arg
130      135      140
Asp Met Gln Asp Thr Phe Tyr Phe Lys Asp His Lys Leu Leu Arg Thr
145      150      155      160
His Thr Ser Pro Val Gln Ile His Thr Met Gln Glu Gln Thr Pro Pro
165      170      175
Ile Lys Met Ile Cys Leu Gly Glu Thr Phe Arg Arg Asp Tyr Asp Leu
180      185      190
Thr His Thr Pro Met Phe His Gln Ile Glu Gly Leu Val Val Asp Gln
195      200      205
Lys Gly Asn Ile Arg Phe Thr His Leu Lys Gly Val Ile Glu Asp Phe
210      215      220
Leu His Tyr Phe Phe Gly Gly Val Lys Leu Arg Trp Arg Ser Ser Phe
225      230      235      240
Phe Pro Phe Thr Glu Pro Ser Ala Glu Val Asp Ile Ser Cys Val Phe
245      250      255
Cys Lys Gln Glu Gly Cys Arg Val Cys Ser His Thr Gly Trp Leu Glu
260      265      270
Val Leu Gly Cys Gly Met Val Asn Asn Ala Val Phe Glu Ala Ile Gly
275      280      285
Tyr Glu Asn Val Ser Gly Phe Ala Phe Gly Met Gly Ile Glu Arg Leu
290      295      300
Ala Met Leu Thr Cys Gln Ile Asn Asp Leu Arg Ser Phe Phe Glu Thr
305      310      315      320
Asp Leu Arg Val Leu Glu Ser Phe
325

```

<210> 73

<211> 327

<212> PRT

<213> Escherichia coli

<400> 73

```

Met Ser His Leu Ala Glu Leu Val Ala Ser Ala Lys Ala Ala Ile Ser
 1          5          10          15
Gln Ala Ser Asp Val Ala Ala Leu Asp Asn Val Arg Val Glu Tyr Leu
 20          25          30
Gly Lys Lys Gly His Leu Thr Leu Gln Met Thr Thr Leu Arg Glu Leu
 35          40          45
Pro Pro Glu Glu Arg Pro Ala Ala Gly Ala Val Ile Asn Glu Ala Lys
 50          55          60
Glu Gln Val Gln Gln Ala Leu Asn Ala Arg Lys Ala Glu Leu Glu Ser
 65          70          75          80
Ala Ala Leu Asn Ala Arg Leu Ala Ala Glu Thr Ile Asp Val Ser Leu
 85          90          95
Pro Gly Arg Arg Ile Glu Asn Gly Gly Leu His Pro Val Thr Arg Thr
100          105          110
Ile Asp Arg Ile Glu Ser Phe Phe Gly Glu Leu Gly Phe Thr Val Ala
115          120          125
Thr Gly Pro Glu Ile Glu Asp Asp Tyr His Asn Phe Asp Ala Leu Asn
130          135          140
Ile Pro Gly His His Pro Ala Arg Ala Asp His Asp Thr Phe Trp Phe
145          150          155          160
Asp Thr Thr Arg Leu Leu Arg Thr Gln Thr Ser Gly Val Gln Ile Arg
165          170          175
Thr Met Lys Ala Gln Gln Pro Pro Ile Arg Ile Ile Ala Pro Gly Arg
180          185          190
Val Tyr Arg Asn Asp Tyr Asp Gln Thr His Thr Pro Met Phe His Gln
195          200          205
Met Glu Gly Leu Ile Val Asp Thr Asn Ile Ser Phe Thr Asn Leu Lys
210          215          220
Gly Thr Leu His Asp Phe Leu Arg Asn Phe Phe Glu Glu Asp Leu Gln
225          230          235          240
Ile Arg Phe Arg Pro Ser Tyr Phe Pro Phe Thr Glu Pro Ser Ala Glu
245          250          255
Val Asp Val Met Gly Lys Asn Gly Lys Trp Leu Glu Val Leu Gly Cys
260          265          270
Gly Met Val His Pro Asn Val Leu Arg Asn Val Gly Ile Asp Pro Glu
275          280          285
Val Tyr Ser Gly Phe Ala Phe Gly Met Gly Met Glu Arg Leu Thr Met
290          295          300
Leu Arg Tyr Gly Val Thr Asp Leu Arg Ser Phe Phe Glu Asn Asp Leu
305          310          315          320
Arg Phe Leu Lys Gln Phe Lys
325

```

<210> 74

<211> 453

<212> PRT

<213> Drosophila melanogaster

<400> 74

```

Met Leu Leu Thr Leu Arg Val Gln Gly Ala Arg His Trp Leu Lys Ser
 1      5      10      15
Thr Arg Cys Leu Ala Ser Ser Ala Ala Pro Ala Lys Ser Pro Ser Ser
 20      25      30
Pro Pro Gln Leu Glu Val Ser Gly Ser Thr Tyr Ala Thr Asp Gly Trp
 35      40      45
Thr Asn Val Thr Pro Lys Ile Leu Ser Tyr Val Gly Ala Asn Lys His
 50      55      60
Leu Gln Thr Asp His Pro Leu Ser Ile Ile Arg Gln Arg Ile Val Asn
 65      70      75      80
Tyr Phe Tyr Gly Ala Tyr Arg Asn Gln Arg Gly Asn Pro Leu Phe Ser
 85      90      95
Val Tyr Asp Gln Met Asn Pro Val Val Thr Val Gln Gln Asn Phe Asp
100      105      110
Asn Leu Leu Ile Pro Ala Asp His Val Ser Arg Gln Lys Ser Asp Cys
115      120      125
Tyr Tyr Ile Asn Gln Gln His Leu Leu Arg Ala His Thr Thr Ala His
130      135      140
Gln Val Glu Leu Ile Ser Gly Gly Leu Asp Asn Phe Leu Val Val Gly
145      150      155      160
Glu Val Tyr Arg Arg Asp Glu Ile Asp Ser Thr His Tyr Pro Val Phe
165      170      175
His Gln Ala Asp Ala Val Arg Leu Val Thr Lys Asp Lys Leu Phe Glu
180      185      190
Arg Asn Pro Gly Leu Glu Leu Phe Glu Glu Thr Trp Ser Gly Thr Leu
195      200      205
Ala Asp Pro Lys Leu Ile Leu Pro Ser Ser Lys Phe Met Asp Gln Thr
210      215      220
Lys Gln Pro Cys His Thr Leu Glu Ala Val Lys Leu Met Glu His Glu
225      230      235      240
Met Lys His Val Leu Val Gly Leu Thr Lys Asp Leu Phe Gly Pro Arg
245      250      255
Ile Lys Tyr Arg Trp Val Asp Thr Tyr Phe Pro Phe Thr Gln Pro Ser
260      265      270
Trp Glu Leu Glu Ile Tyr Phe Lys Asp Asn Trp Leu Glu Val Leu Gly
275      280      285
Cys Gly Ile Met Arg His Glu Ile Leu Gln Arg Ser Gly Val His Gln
290      295      300
Ser Ile Gly Tyr Ala Phe Gly Val Gly Leu Glu Arg Leu Ala Met Val
305      310      315      320
Leu Phe Asp Ile Pro Asp Ile Arg Leu Phe Trp Ser Asn Asp Ser Gly
325      330      335
Phe Leu Ser Gln Phe Ser Glu Lys Asp Leu His Asn Leu Pro Lys Tyr
340      345      350
Lys Pro Ile Ser His Tyr Pro Gln Cys Thr Asn Asp Leu Ser Phe Trp
355      360      365
Leu Pro Gln Asp Ile Glu Val Asp Ala Gly Phe Ser Pro Asn Asp Phe
370      375      380
Tyr Asp Leu Val Arg Ser Val Ala Gly Asp Met Val Glu Gln Ile Ser
385      390      395      400
Leu Val Asp Lys Phe Lys His Pro Lys Thr Gly Lys Ser Ser Val Cys
405      410      415
Phe Arg Ile Val Tyr Arg His Met Glu Arg Thr Leu Thr Gln Ala Glu
420      425      430
Val Asn Glu Ile His Lys Gln Ile Ala Ser Ala Ser Val Asp Ser Phe
435      440      445
Asn Val Gln Ile Arg
450

```

<210> 75

<211> 339

<212> PRT

<213> Chlamydomphila pneumoniae

<400> 75

```

Met Glu Met Lys Glu Glu Ile Glu Ala Val Lys Gln Gln Phe His Ser
 1      5      10      15
Glu Leu Asp Gln Val Asn Ser Ser Gln Ala Leu Ala Asp Leu Lys Val
 20      25      30
Arg Tyr Leu Gly Lys Lys Gly Ile Phe Arg Ser Phe Ser Glu Lys Leu
 35      40      45
Lys Gln Cys Thr Asp Lys Ala Lys Leu Gly Ser Leu Ile Asn Asp Phe
 50      55      60
Lys Thr Tyr Val Glu Asp Leu Leu Gln Glu Lys Ser Leu Val Leu Leu
 65      70      75      80
Ala Ser Glu Gln Ala Glu Ala Phe Ser Lys Glu Lys Ile Asp Ser Ser
 85      90      95
Leu Pro Gly Asp Ser Gln Pro Ser Gly Gly Arg His Ile Leu Lys Ser
100      105      110
Ile Leu Asp Asp Val Val Asp Ile Phe Val His Leu Gly Phe Cys Val
115      120      125
Arg Glu Ala Pro Asn Ile Glu Ser Glu Ala Asn Asn Phe Thr Leu Leu
130      135      140
Asn Phe Thr Glu Asp His Pro Ala Arg Gln Met His Asp Thr Phe Tyr
145      150      155      160
Leu Asn Ala Thr Thr Val Leu Arg Thr His Thr Ser Asn Val Gln Ala
165      170      175
Arg Glu Leu Lys Lys Gln Gln Pro Pro Ile Lys Val Val Ala Pro Gly
180      185      190
Leu Cys Phe Arg Asn Glu Asp Ile Ser Ala Arg Ser His Val Leu Phe
195      200      205
His Gln Val Glu Ala Phe Tyr Val Asp His Asn Val Thr Phe Ser Asp
210      215      220
Leu Thr Ala Ile Leu Ser Ala Phe Tyr His Ser Phe Phe Gln Arg Lys
225      230      235      240
Thr Glu Leu Arg Phe Arg His Ser Tyr Phe Pro Phe Val Glu Pro Gly
245      250      255
Ile Glu Val Asp Val Ser Cys Glu Cys Cys Gly Lys Gly Cys Ala Leu
260      265      270
Cys Lys His Thr Gly Trp Leu Glu Val Ala Gly Ala Gly Met Ile His
275      280      285
Pro Gln Val Leu Arg Asn Gly Asn Val Asp Pro Glu Ile Tyr Ser Gly
290      295      300
Tyr Ala Val Gly Met Gly Ile Glu Arg Leu Ala Met Leu Lys Tyr Gly
305      310      315      320
Val Ser Asp Ile Arg Leu Phe Ser Glu Asn Asp Leu Arg Phe Leu Gln
325      330      335
Gln Phe Ser

```

<210> 76

<211> 451

<212> PRT

<213> Homo sapiens

<400> 76


```

Met Val Gly Ser Ala Leu Arg Arg Gly Ala His Ala Tyr Val Tyr Leu
 1      5      10      15
Val Ser Lys Ala Ser His Ile Ser Arg Gly His Gln His Gln Ala Trp
    20      25      30
Gly Ser Arg Pro Pro Ala Ala Glu Cys Ala Thr Gln Arg Ala Pro Gly
 35      40      45
Ser Val Val Glu Leu Leu Gly Lys Ser Tyr Pro Gln Asp Asp His Ser
 50      55      60
Asn Leu Thr Arg Lys Val Leu Thr Arg Val Gly Arg Asn Leu His Asn
 65      70      75
Gln Gln His His Pro Leu Trp Leu Ile Lys Glu Arg Val Lys Glu His
    85      90      95
Phe Tyr Lys Gln Tyr Val Gly Arg Phe Gly Thr Pro Leu Phe Ser Val
   100      105      110
Tyr Asp Asn Leu Ser Pro Val Val Thr Thr Trp Gln Asn Phe Asp Ser
   115      120      125
Leu Leu Ile Pro Ala Asp His Pro Ser Arg Lys Lys Gly Asp Asn Tyr
   130      135      140
Tyr Leu Asn Arg Thr His Met Leu Arg Ala His Thr Ser Ala His Gln
   145      150      155
Trp Asp Leu Leu His Ala Gly Leu Asp Ala Phe Leu Val Val Gly Asp
   165      170      175
Val Tyr Arg Arg Asp Gln Ile Asp Ser Gln His Tyr Pro Ile Phe His
   180      185      190
Gln Leu Glu Ala Val Arg Leu Phe Ser Lys His Glu Leu Phe Ala Gly
   195      200      205
Ile Lys Asp Gly Glu Ser Leu Gln Leu Phe Glu Gln Ser Ser Arg Ser
   210      215      220
Ala His Lys Gln Glu Thr His Thr Met Glu Ala Val Lys Leu Val Glu
   225      230      235
Phe Asp Leu Lys Gln Thr Leu Thr Arg Leu Met Ala His Leu Phe Gly
   245      250      255
Asp Glu Leu Glu Ile Arg Trp Val Asp Cys Tyr Phe Pro Phe Thr His
   260      265      270
Pro Ser Phe Glu Met Glu Ile Asn Phe His Gly Glu Trp Leu Glu Val
   275      280      285
Leu Gly Cys Gly Val Met Glu Gln Gln Leu Val Asn Ser Ala Gly Ala
   290      295      300
Gln Asp Arg Ile Gly Trp Ala Phe Gly Leu Gly Leu Glu Arg Leu Ala
   305      310      315
Met Ile Leu Tyr Asp Ile Pro Asp Ile Arg Leu Phe Trp Cys Glu Asp
   325      330      335
Glu Arg Phe Leu Lys Gln Phe Cys Val Ser Asn Ile Asn Gln Lys Val
   340      345      350
Lys Phe Gln Pro Leu Ser Lys Tyr Pro Ala Val Ile Asn Asp Ile Ser
   355      360      365
Phe Trp Leu Pro Ser Glu Asn Tyr Ala Glu Asn Asp Phe Tyr Asp Leu
   370      375      380
Val Arg Thr Ile Gly Gly Asp Leu Val Glu Lys Val Asp Leu Ile Asp
   385      390      395
Lys Phe Val His Pro Lys Thr His Lys Thr Ser His Cys Tyr Arg Ile
   405      410      415
Thr Tyr Arg His Met Glu Arg Thr Leu Ser Gln Arg Glu Val Arg His
   420      425      430
Ile His Gln Ala Leu Gln Glu Ala Ala Val Gln Leu Leu Gly Val Glu
   435      440      445
Gly Arg Phe
   450

```

<210> 77

<211> 341

<212> PRT

<213> Mycobacterium tuberculosis

<400> 77

```

Met Leu Ser Pro Glu Ala Leu Thr Thr Ala Val Asp Ala Ala Gln Gln
 1          5          10          15
Ala Ile Ala Leu Ala Asp Thr Leu Asp Val Leu Ala Arg Val Lys Thr
 20          25          30
Glu His Leu Gly Asp Arg Ser Pro Leu Ala Leu Ala Arg Gln Ala Leu
 35          40          45
Ala Val Leu Pro Lys Glu Gln Arg Ala Glu Ala Gly Lys Arg Val Asn
 50          55          60
Ala Ala Arg Asn Ala Ala Gln Arg Ser Tyr Asp Glu Arg Leu Ala Thr
 65          70          75          80
Leu Arg Ala Glu Arg Asp Ala Ala Val Leu Val Ala Glu Gly Ile Asp
 85          90          95
Val Thr Leu Pro Ser Thr Arg Val Pro Ala Gly Ala Arg His Pro Ile
100          105          110
Ile Met Leu Ala Glu His Val Ala Asp Thr Phe Ile Ala Met Gly Trp
115          120          125
Glu Leu Ala Glu Gly Pro Glu Val Glu Thr Glu Gln Phe Asn Phe Asp
130          135          140
Ala Leu Asn Phe Pro Ala Asp His Pro Ala Arg Gly Glu Gln Asp Thr
145          150          155          160
Phe Tyr Ile Ala Pro Glu Asp Ser Arg Gln Leu Leu Arg Thr His Thr
165          170          175
Ser Pro Val Gln Ile Arg Thr Leu Leu Ala Arg Glu Leu Pro Val Tyr
180          185          190
Ile Ile Ser Ile Gly Arg Thr Phe Arg Thr Asp Glu Leu Asp Ala Thr
195          200          205
His Thr Pro Ile Phe His Gln Val Glu Gly Leu Ala Val Asp Arg Gly
210          215          220
Leu Ser Met Ala His Leu Arg Gly Thr Leu Asp Ala Phe Ala Arg Ala
225          230          235          240
Glu Phe Gly Pro Ser Ala Arg Thr Arg Ile Arg Pro His Phe Phe Pro
245          250          255
Phe Thr Glu Pro Ser Ala Glu Val Asp Val Trp Phe Ala Asn Lys Ile
260          265          270
Gly Gly Ala Ala Trp Val Glu Trp Gly Gly Cys Gly Met Val His Pro
275          280          285
Asn Val Leu Arg Ala Thr Gly Ile Asp Pro Asp Leu Tyr Ser Gly Phe
290          295          300
Ala Phe Gly Met Gly Leu Glu Arg Thr Leu Gln Phe Arg Asn Gly Ile
305          310          315          320
Pro Asp Met Arg Asp Met Val Glu Gly Asp Val Arg Phe Ser Leu Pro
325          330          335
Phe Gly Val Gly Ala
340

```

<210> 78

<211> 458

<212> PRT

<213> *Caenorhabditis elegans*

<400> 78

```

Met Arg Phe Leu Ser Thr Tyr Gln Thr Gly Ala Arg Arg Cys Cys Thr
1      5      10      15
Met Leu Phe Gln Lys Asn Ala Lys Ser Ile Arg Ala Thr Leu Met Ser
20     25     30
Thr Thr Ala Ala Ala Glu Glu Ser Arg Lys Ile Val Lys Glu Glu Val
35     40     45
Phe Glu Leu Asp Gly Arg Lys Tyr Thr Pro Asp Ala Leu Tyr Asn Leu
50     55     60
Ser Pro Gly Val Arg Arg Leu Leu Asp Arg Arg Ile Leu Gln Glu Ser
65     70     75     80
Ser Asn Pro Leu Asn Leu Leu Lys Arg Arg Ile Val Asp Tyr Val His
85     90     95
Gln Thr Tyr Arg Lys Pro Gly Asn Arg Ser Pro Leu Phe Thr Ile Cys
100    105    110
Glu Ser Glu Pro Arg Val Val Thr Thr Tyr Gln Asn Phe Asp Ser Leu
115    120    125
Leu Thr Pro Glu Asp His Val Ser Arg Arg Pro Ser Asp Thr Tyr Tyr
130    135    140
Val Asn His Glu His Cys Leu Arg Ala His Thr Ser Ala His Gln His
145    150    155    160
Asn Leu Met Gln Ser Gly Leu Asp Ala Phe Leu Val Ile Gly Asp Val
165    170    175
Tyr Arg Arg Asp Glu Val Asp Arg Thr His Tyr Pro Cys Phe His Gln
180    185    190
Ile Glu Gly Val Arg Leu Tyr Ser Lys Asp Glu Leu Leu Gly Lys Lys
195    200    205
Pro Asp Gly Lys Asn Val Ala Glu Leu Phe Ser Thr Ala Ala Ser Ala
210    215    220
Ala Thr Glu Arg Ser Pro Glu Lys Gln Glu Lys His Thr Leu Asp Ala
225    230    235    240
Thr Lys Ala Ala Glu Ile Gln Leu Lys Gln Phe Leu Glu Asn Leu Cys
245    250    255
Asp Glu Leu Phe Gly Lys Asp Ala Glu Lys Arg Trp Val Asp Ala Tyr
260    265    270
Phe Pro Phe Thr His Pro Ser Trp Glu Leu Glu Val Phe Tyr Asn Gly
275    280    285
Gln Trp Leu Glu Val Leu Gly Cys Gly Ile Met Glu Gln Lys Leu Leu
290    295    300
Glu Ser Ala Gly Val Thr Asp Lys Ile Gly Trp Ala Phe Gly Ile Gly
305    310    315    320
Leu Glu Arg Ile Ala Met Val Leu Tyr Gly Ile Pro Asp Ile Arg Leu
325    330    335
Phe Trp Ser Lys Asp Thr Gly Phe Leu Ser Gln Phe Ala Gly Lys Met
340    345    350
Pro Gly Glu Asp Val Lys Tyr Lys Gln Ile Ser Ala His Pro Gln Val
355    360    365
Ile Phe Asp Ile Ser Phe Phe Leu Pro Ser Thr Val Gln Phe Asn Asp
370    375    380
Met Thr Ser Asp Val Tyr Asp Thr Ile Arg Thr Val Gly Gly Glu Leu
385    390    395    400
Val Glu Gln Val Lys Leu Thr Asp Glu Phe Glu Asn Lys Lys Lys Glu
405    410    415
Lys Lys Ser Gln Thr Tyr Arg Ile Val Tyr Arg Ser His Glu Arg Ala
420    425    430
Leu Thr Lys Glu Glu Val Asn Val Ile His Lys Gln Ile Glu Gln Ser
435    440    445
Leu Ala Ser Ser Phe Gly Val Thr Leu Arg
450    455

```

<210> 79

<211> 469

<212> PRT

<213> *Saccharomyces cerevisiae*

<400> 79

```

Met Phe Leu Asn Arg Met Met Lys Thr Arg Thr Gly Leu Tyr Arg Leu
 1      5      10      15
Tyr Ser Thr Leu Lys Val Pro His Val Glu Ile Asn Gly Ile Lys Tyr
 20      25      30
Lys Thr Asp Pro Gln Thr Thr Asn Val Thr Asp Ser Ile Ile Lys Leu
 35      40      45
Thr Asp Arg Ser Leu His Leu Lys Glu Ser His Pro Val Gly Ile Leu
 50      55      60
Arg Asp Leu Ile Glu Lys Lys Leu Asn Ser Val Asp Asn Thr Phe Lys
 65      70      75      80
Ile Phe Asn Asn Phe Lys Pro Val Val Thr Thr Met Glu Asn Phe Asp
 85      90      95
Ser Leu Gly Phe Pro Lys Asp His Pro Gly Arg Ser Lys Ser Asp Thr
100      105      110
Tyr Tyr Ile Asn Glu Thr His Leu Leu Arg Thr His Thr Ser Ala His
115      120      125
Glu Leu Glu Cys Phe Gln Lys Ile Arg Asn Asp Ser Asp Asn Ile Lys
130      135      140
Ser Gly Phe Leu Ile Ser Ala Asp Val Tyr Arg Arg Asp Glu Ile Asp
145      150      155      160
Lys Thr His Tyr Pro Val Phe His Gln Met Glu Gly Ala Thr Ile Trp
165      170      175
Lys Arg Thr Lys Ala Asp Val Gly Val Lys Glu Pro Met Tyr Ile Glu
180      185      190
Lys Ile Arg Glu Asp Ile Arg Gln Val Glu Asn Leu Leu Asn Lys Glu
195      200      205
Asn Val Lys Ile Thr Val Asp Asp Asp Thr Ile Pro Leu Lys Glu Asn
210      215      220
Asn Pro Lys Gln Glu Tyr Met Ser Asp Leu Glu Val Asp Leu Cys Ser
225      230      235      240
Gln His Leu Lys Arg Ser Ile Glu Leu Ile Val Ser Glu Val Phe Asn
245      250      255
Lys Lys Ile Ser Ser Met Ile Lys Asn Lys Ala Asn Asn Thr Pro Lys
260      265      270
Glu Leu Lys Val Arg Trp Ile Asn Ala Tyr Phe Pro Trp Thr Ala Pro
275      280      285
Ser Trp Glu Ile Glu Val Trp Trp Gln Gly Glu Trp Leu Glu Leu Cys
290      295      300
Gly Cys Gly Leu Ile Arg Gln Asp Val Leu Leu Arg Ala Gly Tyr Lys
305      310      315      320
Pro Ser Glu Thr Ile Gly Trp Ala Phe Gly Leu Gly Leu Asp Arg Ile
325      330      335
Ala Met Leu Leu Phe Glu Ile Pro Asp Ile Arg Leu Leu Trp Ser Arg
340      345      350
Asp Glu Arg Phe Ser Arg Gln Phe Ser Lys Gly Leu Ile Thr Ser Phe
355      360      365
Lys Pro Tyr Ser Lys His Pro Gly Ser Phe Arg Asp Val Ala Phe Trp
370      375      380
Leu Pro Glu Asp Lys Pro Asp Ile His Gln Val His Glu Asn Asp Leu
385      390      395      400
Met Glu Ile Ile Arg Asn Ile Ala Gly Asp Leu Val Glu Ser Val Lys
405      410      415
Leu Val Asp Ser Phe Thr His Pro Lys Thr Gly Arg Lys Ser Met Cys
420      425      430
Tyr Arg Ile Asn Tyr Gln Ser Met Asp Arg Asn Leu Thr Asn Ala Glu
435      440      445
Val Asn Thr Leu Gln Asp Met Val Cys Ser Lys Leu Val Lys Glu Tyr
450      455      460
Ser Val Glu Leu Arg
465

```

<210> 80

<211> 499

<212> PRT

<213> Schizosaccharomyces pombe

<400> 80

```

Met Ser Lys Leu Glu Ala Leu Gln Val Phe Leu Leu Glu Lys Leu Asn
 1      5      10      15
Glu Lys Asn Glu Ile Pro Asn Thr Ser His Leu Glu Phe Asp Gly Lys
      20      25      30
Lys Leu Gly Pro Gln Glu Ala Gln Ser Ala Ile Leu Ser Leu Ala Ala
      35      40      45
Lys Asn Met Ile Glu Phe Ser Arg His Glu Ile Glu Ile Tyr Asn Leu
      50      55      60
Thr Ala Glu Gly Glu Asn Ile Cys Ala Asn Gly Ser His Glu Ala Lys
      65      70      75      80
Val Tyr Asn Glu Ile Cys Ala Ser Met Ser Gly Leu Asn Ile Gly Glu
      85      90      95
Leu Lys Lys Lys Leu Gly Asn Ser Ala Gly Ile Gly Gln Gly Arg Ala
      100      105      110
Phe Lys Leu Gly Trp Ile Lys Lys Asp Gly Asp Lys Leu Val Lys Asn
      115      120      125
Thr Asp Ser Ile Thr Asp Glu Thr Pro Lys Val Leu Ser Glu Ile Lys
      130      135      140
Glu His Gly Thr Ile Ser Asp Ser Lys Thr Leu Thr Asp Leu Lys Lys
      145      150      155      160
Arg Lys Leu Val Glu Arg Asn Lys Ile Met Tyr Phe Ser Leu Arg Lys
      165      170      175
Gly Pro Asn Phe Ser Leu Gln Ile Glu Lys Leu Asn Thr Asp Leu Thr
      180      185      190
Ala Glu Met Ile Thr Ser Arg Ser Trp Glu Ser Ala Lys Phe Lys Ser
      195      200      205
Tyr Asn Phe Ala Ala Glu Gly Ile Pro Pro Ala Gly Gly Cys Leu His
      210      215      220
Pro Leu Met Lys Val Arg Glu Glu Phe Arg Lys Phe Phe Phe Glu Leu

```

```

225      230      235      240
Gly Phe Glu Glu Met Pro Thr Asn Asn Phe Val Glu Ser Gly Phe Trp
      245      250      255
Asn Phe Asp Ala Leu Phe Val Pro Gln Gln His Ser Ala Arg Asp Ala
      260      265      270
Gln Asp Thr Phe Phe Leu Lys Val Pro Ala Ser Thr Asp Lys Leu Pro
      275      280      285
Asp Pro Glu Tyr Val Ala Arg Val Lys Ala Thr His Glu Asn Gly Gly
      290      295      300
Glu Thr Lys Gly Ile Gly Tyr Arg Ala Pro Phe Ser Leu Glu Glu Thr
      305      310      315      320
Arg Lys Leu Val Leu Arg Thr His Thr Thr Ala Val Ser Ala Asn Met
      325      330      335
Leu Tyr Lys Leu Ala Gln Asn Gly Phe His Pro Ala Lys Tyr Phe Ser
      340      345      350
Ile Asp Arg Val Phe Arg Asn Glu Thr Val Asp Ala Thr His Leu Ala
      355      360      365
Glu Phe His Gln Val Glu Gly Val Ile Cys Asp Arg Asn Ile Thr Leu
      370      375      380
Gly Asp Leu Ile Gly Phe Leu Glu Val Phe Phe Gly Lys Met Asn Val
      385      390      395      400
Lys Asn Leu Arg Phe Lys Pro Ala Tyr Asn Pro Tyr Thr Glu Pro Ser
      405      410      415
Leu Glu Val Phe Ser Tyr His Glu Lys Leu Gly Lys Trp Val Glu Val
      420      425      430
Gly Asn Ser Gly Met Phe Arg Pro Glu Met Leu Glu Pro Met Gly Leu
      435      440      445
Pro Lys Asp Val Arg Cys Leu Gly Phe Gly Leu Ser Leu Glu Arg Pro
      450      455      460
Thr Met Ile Lys Tyr Gly Val Ala Asp Ile Arg Gln Leu Ile Gly Pro
      465      470      475      480
Lys Val Asn Leu Asp Leu Ile Glu Ala Ser Pro Ala Val Arg Leu Asp
      485      490      495
Lys Glu Glu

```

<210> 81
 <211> 338
 <212> PRT
 <213> *Pseudomonas aeruginosa*
 <400> 81

```

Met Glu Asn Leu Asp Ala Leu Val Ser Gln Ala Leu Glu Ala Val Arg
 1      5      10      15
His Thr Glu Asp Val Asn Ala Leu Glu Gln Ile Arg Val His Tyr Leu
      20      25      30
Gly Lys Lys Gly Glu Leu Thr Gln Val Met Lys Thr Leu Gly Asp Leu
      35      40      45
Pro Ala Glu Glu Arg Pro Lys Val Gly Ala Leu Ile Asn Val Ala Lys
      50      55      60
Glu Lys Val Gln Asp Val Leu Asn Ala Arg Lys Thr Glu Leu Glu Gly
65      70      75      80
Ala Ala Leu Ala Ala Arg Leu Ala Ala Glu Arg Ile Asp Val Thr Leu
      85      90      95
Pro Gly Arg Gly Gln Leu Ser Gly Gly Leu His Pro Val Thr Arg Thr
      100      105      110
Leu Glu Arg Ile Glu Gln Cys Phe Ser Arg Ile Gly Tyr Glu Val Ala

      115      120      125
Glu Gly Pro Glu Val Glu Asp Asp Tyr His Asn Phe Glu Ala Leu Asn
      130      135      140
Ile Pro Gly His His Pro Ala Arg Ala Met His Asp Thr Phe Tyr Phe
145      150      155      160
Asn Ala Asn Met Leu Leu Arg Thr His Thr Ser Pro Val Gln Val Arg
      165      170      175
Thr Met Glu Ser Gln Gln Pro Pro Ile Arg Ile Val Cys Pro Gly Arg
      180      185      190
Val Tyr Arg Cys Asp Ser Asp Leu Thr His Ser Pro Met Phe His Gln
      195      200      205
Val Glu Gly Leu Leu Val Asp Glu Gly Val Ser Phe Ala Asp Leu Lys
      210      215      220
Gly Thr Ile Glu Glu Phe Leu Arg Ala Phe Phe Glu Lys Gln Leu Glu
225      230      235      240
Val Arg Phe Arg Pro Ser Phe Phe Pro Phe Thr Glu Pro Ser Ala Glu
      245      250      255
Val Asp Ile Gln Cys Val Ile Cys Ser Gly Asn Gly Cys Arg Val Cys
      260      265      270
Lys Gln Thr Gly Trp Leu Glu Val Met Gly Cys Gly Met Val His Pro
      275      280      285
Asn Val Leu Arg Met Ser Asn Ile Asp Pro Glu Lys Phe Gln Gly Phe
      290      295      300
Ala Phe Gly Met Gly Ala Glu Arg Leu Ala Met Leu Arg Tyr Gly Val
305      310      315      320
Asn Asp Leu Arg Leu Phe Phe Asp Asn Asp Leu Arg Phe Leu Gly Gln
      325      330      335
Phe Arg

```

<210> 82
 <211> 785
 <212> PRT

<213> Thermus thermophilus

<400> 82

```

Met Arg Val Pro Phe Ser Trp Leu Lys Ala Tyr Val Pro Glu Leu Glu
 1           5           10           15
Ser Pro Glu Val Leu Glu Glu Arg Leu Ala Gly Leu Gly Phe Glu Thr
      20           25           30
Asp Arg Ile Glu Arg Val Phe Pro Ile Pro Arg Gly Val Val Phe Ala
      35           40           45
Arg Val Leu Glu Ala His Pro Ile Pro Gly Thr Arg Leu Lys Arg Leu
      50           55           60
Val Leu Asp Ala Gly Arg Thr Val Glu Val Val Ser Gly Ala Glu Asn
      65           70           75           80
Ala Arg Lys Gly Ile Gly Val Ala Leu Ala Leu Pro Gly Thr Glu Leu
      85           90           95
Pro Gly Leu Gly Gln Lys Val Gly Glu Arg Val Ile Gln Gly Val Arg
      100          105          110
Ser Phe Gly Met Ala Leu Ser Pro Arg Glu Leu Gly Val Gly Glu Tyr
      115          120          125
Gly Gly Gly Leu Leu Glu Phe Pro Glu Asp Ala Leu Pro Pro Gly Thr
      130          135          140
Pro Leu Ser Glu Ala Trp Pro Glu Glu Val Val Leu Asp Leu Glu Val
      145          150          155          160
Thr Pro Asn Arg Pro Asp Ala Leu Gly Leu Leu Gly Leu Ala Arg Asp

```

```

165      170      175
Leu His Ala Leu Gly Tyr Ala Leu Val Glu Pro Glu Ala Ala Leu Lys
180      185      190
Ala Glu Ala Leu Pro Leu Pro Phe Ala Leu Lys Val Glu Asp Pro Glu
195      200      205
Gly Ala Pro His Phe Thr Leu Gly Tyr Ala Phe Gly Leu Arg Val Ala
210      215      220
Pro Ser Pro Leu Trp Met Gln Arg Ala Leu Phe Ala Ala Gly Met Arg
225      230      235      240
Pro Ile Asn Asn Val Val Asp Val Thr Asn Tyr Val Met Leu Glu Arg
245      250      255
Ala Gln Pro Met His Ala Phe Asp Leu Arg Phe Val Gly Glu Gly Ile
260      265      270
Ala Val Arg Arg Ala Arg Glu Gly Glu Arg Leu Lys Thr Leu Asp Gly
275      280      285
Val Glu Arg Thr Leu His Pro Glu Asp Leu Val Ile Ala Gly Trp Arg
290      295      300
Gly Glu Glu Ser Phe Pro Leu Gly Leu Ala Gly Val Met Gly Gly Ala
305      310      315      320
Glu Ser Glu Val Arg Glu Asp Thr Glu Ala Ile Ala Leu Glu Val Ala
325      330      335
Cys Phe Asp Pro Val Ser Ile Arg Lys Thr Ala Arg Arg His Gly Leu
340      345      350
Arg Thr Glu Ala Ser His Arg Phe Glu Arg Gly Val Asp Pro Leu Gly
355      360      365
Gln Val Pro Ala Gln Arg Arg Ala Leu Ser Leu Leu Gln Ala Leu Ala
370      375      380
Gly Ala Arg Val Ala Glu Ala Leu Leu Glu Ala Gly Ser Pro Lys Pro
385      390      395      400
Pro Glu Ala Ile Pro Phe Arg Pro Glu Tyr Ala Asn Arg Leu Leu Gly
405      410      415
Thr Ser Tyr Pro Glu Ala Glu Gln Ile Ala Ile Leu Lys Arg Leu Gly
420      425      430
Cys Arg Val Glu Gly Glu Gly Pro Thr Tyr Arg Val Thr Pro Pro Ser
435      440      445
His Arg Leu Asp Leu Arg Leu Glu Glu Asp Leu Val Glu Glu Val Ala
450      455      460
Arg Ile Gln Gly Tyr Glu Thr Ile Pro Leu Ala Leu Pro Ala Phe Phe
465      470      475      480
Pro Ala Pro Asp Asn Arg Gly Val Glu Ala Pro Tyr Arg Lys Glu Gln
485      490      495
Arg Leu Arg Glu Val Leu Ser Gly Leu Gly Phe Gln Glu Val Tyr Thr
500      505      510
Tyr Ser Phe Met Asp Pro Glu Asp Ala Arg Arg Phe Arg Leu Asp Pro
515      520      525
Pro Arg Leu Leu Leu Leu Asn Pro Leu Ala Pro Glu Lys Ala Ala Leu
530      535      540
Arg Thr His Leu Phe Pro Gly Leu Val Arg Val Leu Lys Glu Asn Leu
545      550      555      560
Asp Leu Asp Arg Pro Glu Arg Ala Leu Leu Phe Glu Val Gly Arg Val
565      570      575
Phe Arg Glu Arg Glu Glu Thr His Leu Ala Gly Leu Leu Phe Gly Glu
580      585      590
Gly Val Gly Leu Pro Trp Ala Lys Glu Arg Leu Ser Gly Tyr Phe Leu
595      600      605
Leu Lys Gly Tyr Leu Glu Ala Leu Phe Ala Arg Leu Gly Leu Ala Phe
610      615      620
Arg Val Glu Ala Gln Ala Phe Pro Phe Leu His Pro Gly Val Ser Gly

```



```

625          630          635          640
Arg Val Leu Val Glu Gly Glu Glu Val Gly Phe Leu Gly Ala Leu His
          645          650          655
Pro Glu Ile Ala Gln Glu Leu Glu Leu Pro Pro Val His Leu Phe Glu
          660          665          670
Leu Arg Leu Pro Leu Pro Asp Lys Pro Leu Ala Phe Gln Asp Pro Ser
          675          680          685
Arg His Pro Ala Ala Phe Arg Asp Leu Ala Val Val Val Pro Ala Pro
          690          695          700
Thr Pro Tyr Gly Glu Val Glu Ala Leu Val Arg Glu Ala Ala Gly Pro
705          710          715
Tyr Leu Glu Ser Leu Ala Leu Phe Asp Leu Tyr Gln Gly Pro Pro Leu
          725          730          735
Pro Glu Gly His Lys Ser Leu Ala Phe His Leu Arg Phe Arg His Pro
          740          745          750
Lys Arg Thr Leu Arg Asp Glu Glu Val Glu Glu Ala Val Ser Arg Val
          755          760          765
Ala Glu Ala Leu Arg Ala Arg Gly Phe Gly Leu Arg Gly Leu Asp Thr
          770          775          780
Pro
785

```

<210> 83

<211> 598

<212> PRT

<213> Arabidopsis thaliana

<400> 83

Met	Pro	Thr	Ile	Ser	Val	Gly	Arg	Asp	Arg	Leu	Phe	Ala	Ala	Leu	Gly
1				5					10					15	
Glu	Ser	Tyr	Thr	Gln	Glu	Lys	Phe	Glu	Glu	Leu	Cys	Phe	Ser	Phe	Gly
			20					25					30		
Ile	Glu	Leu	Asp	Asp	Val	Thr	Thr	Glu	Lys	Ala	Ile	Ile	Arg	Lys	Glu
		35					40					45			
Lys	His	Ile	Asp	Glu	Glu	Ala	Asp	Asp	Asp	Glu	Glu	Ile	Ile	Tyr	Lys
	50					55					60				
Ile	Glu	Ile	Pro	Ala	Asn	Arg	Pro	Asp	Leu	Leu	Cys	Leu	Glu	Gly	Leu
65					70					75					80
Ala	Gln	Ser	Leu	Arg	Val	Phe	Ile	Glu	Lys	Gln	Glu	Ile	Pro	Thr	Tyr
				85					90					95	
Thr	Leu	Ala	Asp	Ile	Ser	Lys	Asp	Lys	Ile	Leu	Gln	Met	Asn	Val	Lys
			100					105					110		
Pro	Glu	Thr	Ser	Lys	Ile	Arg	Pro	Phe	Val	Val	Cys	Ala	Val	Leu	Arg
		115					120					125			
Gly	Val	Thr	Phe	Asp	Glu	Ala	Arg	Tyr	Asn	Ser	Phe	Ile	Asp	Leu	Gln
	130					135					140				
Asp	Lys	Leu	His	Gln	Asn	Ile	Cys	Arg	Arg	Arg	Ser	Leu	Val	Ala	Ile
145					150					155					160
Gly	Thr	His	Asp	Leu	Asp	Thr	Leu	Gln	Gly	Pro	Phe	Thr	Tyr	Glu	Ala
			165						170					175	
Leu	Pro	Pro	Thr	Asp	Ile	Asn	Phe	Val	Pro	Leu	Lys	Gln	Thr	Lys	Ser
			180					185					190		
Phe	Arg	Ala	Asp	Glu	Leu	Ile	Glu	Phe	Tyr	Lys	Ser	Asp	Met	Lys	Leu
		195					200					205			
Lys	Lys	Phe	Leu	His	Ile	Ile	Glu	Asn	Ser	Pro	Val	Phe	Pro	Val	Leu
	210					215					220				
Tyr	Asp	Ser	Lys	Arg	Thr	Val	Leu	Ser	Leu	Pro	Pro	Ile	Ile	Asn	Gly

```

225      230      235      240
Ala His Ser Ala Ile Thr Leu Gln Thr Lys Asn Val Phe Ile Glu Cys
      245      250      255
Thr Ala Thr Asp Leu Thr Lys Ala Lys Ile Val Leu Asn Thr Met Val
      260      265      270
Thr Thr Phe Ser Glu Phe Cys Ala Arg Lys Phe Glu Ile Glu Pro Val
      275      280      285
Glu Val Thr Tyr Asp Asp Gly Lys Ser Tyr Ile Tyr Pro Asp Leu Ala
      290      295      300
Val Tyr Asp Met Glu Val Pro Leu Ser Phe Ile Thr Asp Ser Ile Gly
305      310      315      320
Val Ser Leu Lys Val Glu Gln Val Thr Ser Leu Leu Thr Arg Met Gln
      325      330      335
Leu Gln Ala Glu Gln Ala Lys Ser Ser Asp Asn Gln Cys Ala Ile Lys
      340      345      350
Val His Val Pro Pro Ser Arg Ser Asp Val Leu His Pro Cys Asp Val
      355      360      365
Met Glu Asp Val Ala Ile Ala Tyr Gly Phe Asn Asn Ile Pro Thr Arg
370      375      380
Lys Pro Ala Ser Ile Lys Pro Leu Thr Leu Asn Glu Leu Thr Asp Leu
385      390      395      400
Leu Arg Ile Glu Ile Ala Met Cys Val Tyr Thr Glu Val Val Thr Trp
      405      410      415
Leu Leu Cys Ser His Lys Glu Asn Phe Ala Met Leu Asn Arg Glu Asp
      420      425      430
Val Asn Ser Ala Val Ile Val Gly Asn Pro Arg Ser Ala Asp Phe Glu
      435      440      445
Ala Met Arg Arg Ala Leu Met Pro Gly Leu Leu Lys Thr Val Gly His
      450      455      460
Asn Asn Lys Tyr Pro Lys Pro Ile Lys Ile Phe Glu Ile Ser Asp Val
465      470      475      480
Val Met Leu Asp Glu Ser Lys Asp Val Gly Ala Ser Asn Arg Arg His
      485      490      495
Leu Ala Ala Leu Tyr Cys Gly Ala Thr Ser Gly Phe Glu Leu Ile His
      500      505      510
Gly Leu Val Asp Arg Ile Met Glu Val Met Ala Ile Pro Phe Leu Thr
      515      520      525
Ile His Glu Asn Asn Val Pro Ile Asn Glu Lys Asp Gly Tyr Tyr Val
      530      535      540
Lys Leu Ser Gln Glu Pro Glu Phe Leu Pro Gly Arg Gln Ala Ser Ile
545      550      555      560
Ile Val Arg Gly Lys His Ile Gly Asn Phe Gly Ile Val His Pro Glu
      565      570      575
Val Leu Asn Asn Phe Asp Ile Pro Asp Pro Cys Ser Tyr Leu Glu Leu
      580      585      590
Asp Ile Glu Ala Ile Leu
      595

```

<210> 84

<211> 589

<212> PRT

<213> *Drosophila melanogaster*

<400> 84

```

Met Pro Thr Ile Gly Val Lys Arg Asp Leu Leu Phe Glu Ala Leu Gly
1      5      10      15
Lys Thr Tyr Thr Asp Asp Glu Phe Gln Asp Leu Cys Phe Ala Phe Gly

```


<210> 85

<211> 591

<212> PRT

<213> *Caenorhabditis elegans*

<400> 85

Met	Pro	Thr	Val	Gly	Ile	Lys	Lys	Val	Ile	Leu	Asp	Lys	His	Phe	Lys
1				5					10					15	
Arg	Val	Tyr	Ser	Glu	Lys	Glu	Phe	Asp	Glu	Leu	Cys	Phe	Glu	Tyr	Gly
			20					25					30		
Leu	Glu	Leu	Asp	Glu	Ile	Thr	Ser	Glu	Lys	Ala	Ala	Val	Glu	Lys	Glu
		35					40					45			
Gln	Gly	Thr	Arg	Ala	Ala	Ser	Asp	Leu	Asn	Asp	Gln	Glu	Val	Tyr	Lys
	50					55					60				
Ile	Asp	Ile	Pro	Ala	Asn	Arg	Tyr	Asp	Leu	Leu	Ser	Val	Glu	Gly	Leu
65					70					75					80
Ala	Arg	Ala	Ile	Arg	Ile	Phe	Lys	Gln	Glu	Ile	Pro	Ser	Pro	Ala	Tyr
			85					90						95	
Lys	Tyr	Ala	Asp	Val	Pro	Lys	Thr	Gly	Leu	Gln	Lys	Ile	Ile	Val	Lys
			100					105					110		
Lys	Glu	Thr	Ala	Gln	Val	Arg	Pro	Phe	Val	Val	Gly	Ala	Val	Leu	Arg
		115					120					125			
Asp	Ile	Ser	Phe	Asp	Ala	Asp	Ser	Tyr	Ala	Ser	Phe	Ile	Asp	Leu	Gln
	130					135					140				
Asp	Lys	Leu	His	Gln	Asn	Ile	Cys	Arg	Lys	Arg	Thr	Leu	Val	Ala	Ile
145					150					155					160
Gly	Thr	His	Asp	Leu	Asp	Thr	Ile	Gln	Gly	Pro	Phe	Glu	Tyr	Arg	Ala
			165					170						175	
Glu	Ala	Pro	Lys	Asp	Ile	Lys	Phe	Lys	Pro	Leu	Asn	Gln	Thr	Lys	Glu
			180					185					190		
Tyr	Thr	Ala	Glu	Glu	Leu	Met	Thr	Leu	Tyr	Ser	Thr	Asp	Ser	His	Leu
		195					200					205			
Lys	Ala	Tyr	Leu	Pro	Ile	Ile	Gln	Asn	His	Pro	Val	Tyr	Pro	Val	Ile
	210					215					220				
Tyr	Asp	Lys	Asn	Gly	Val	Val	Cys	Ser	Met	Pro	Pro	Ile	Ile	Asn	Gly
225					230					235					240
Glu	His	Ser	Lys	Ile	Thr	Leu	Asn	Thr	Lys	Asn	Val	Phe	Ile	Glu	Ala
			245					250					255		
Thr	Ala	Thr	Asp	Lys	Gln	Lys	Ala	Phe	Val	Val	Leu	Asp	Thr	Ile	Val
			260					265					270		
Thr	Leu	Phe	Ser	Gln	Tyr	Cys	Ala	Lys	Pro	Phe	Thr	Ile	Glu	Gln	Val
		275					280					285			
Glu	Val	Val	Tyr	Glu	Glu	Thr	Gly	Val	Lys	Glu	Leu	Tyr	Pro	Leu	Leu

```

      290      295      300
Ser Tyr Arg Glu Met Thr Val Thr Thr Pro Glu Ile Asn Thr Lys Ile
305      310      315      320
Gly Ile Asn Leu Lys Asp Glu Glu Met Ala Thr Leu Leu Asn Lys Met
      325      330      335
Ser Leu Lys Ala Glu Val Ala Ala Lys Glu Thr Leu Lys Ile Val Val
      340      345      350
Pro Pro Thr Arg His Asp Ile Leu His Ala Cys Asp Ile Ala Glu Asp
      355      360      365
Val Gly Val Ala Phe Gly Tyr Asn Asn Leu Ile Thr Lys Leu Pro Glu
      370      375      380
Ser Asn Thr Val Ala Val Ala Phe Pro Ile Asn Lys Leu Cys Asp Asn
385      390      395      400
Leu Arg Ile Glu Ile Ala Ala Ala Gly Trp Thr Glu Ala Leu Asn Phe
      405      410      415
Ala Leu Cys Ser Arg Asp Asp Ile Ser Ser Lys Leu Arg Gln Pro Asp
      420      425      430
Ala Leu Ser His Ala Val His Ile Gly Asn Pro Lys Thr Leu Glu Phe
      435      440      445
Gln Val Ala Arg Thr Ser Leu Leu Pro Gly Leu Leu Lys Thr Leu Ser
450      455      460
Ser Asn Arg Asp Met Pro Leu Pro Leu Lys Leu Phe Glu Leu Gln Asp
465      470      475      480
Val Ile Val Lys Asp Ser Asn Thr Asp Val Gly Ala Arg Asn Glu Arg
      485      490      495
Arg Leu Ala Ala Val Tyr Tyr Asn Arg Ala Ala Gly Phe Glu Ile Ile
      500      505      510
Gln Gly Phe Leu Asp Arg Ile Met Arg Met Leu Asn Val Asn Pro Ala
      515      520      525
Arg Asp Gly Thr Gly Tyr Tyr Ile Glu Ala Asp Glu Asn Ser Thr Tyr
530      535      540
Phe Pro Gly Arg Cys Ala Lys Ile Ile Gly Pro Lys Gly Val Val Leu
545      550      555      560
Gly His Ile Gly Ala Leu His Pro Glu Val Ile Thr Ser Phe Gly Leu
      565      570      575
Thr Leu Pro Cys Gly Ala Val Glu Ile Asn Val Glu Pro Phe Leu
      580      585      590

```

<210> 86

<211> 804

<212> PRT

<213> *Bacillus subtilis*

<400> 86

```

Met Phe Val Ser Tyr Lys Trp Leu Glu Asp Tyr Val Asp Leu Lys Gly
1      5      10      15
Met Asp Pro Ala Val Leu Ala Glu Lys Ile Thr Arg Ala Gly Ile Glu
      20      25      30
Val Glu Gly Ile Glu Tyr Lys Gly Glu Gly Ile Lys Gly Val Val Ile
      35      40      45
Gly His Val Leu Glu Arg Glu Gln His Pro Asn Ala Asp Lys Leu Asn
50      55      60
Lys Cys Leu Val Asp Ile Gly Ala Glu Ala Pro Val Gln Ile Ile Cys
65      70      75      80
Gly Ala Pro Asn Val Asp Lys Gly Gln Lys Val Ala Val Ala Thr Val
      85      90      95
Gly Ala Val Leu Pro Gly Asn Phe Lys Ile Lys Lys Ala Lys Leu Arg

```



```

                    565                    570                    575
Ala Leu Tyr Glu Val Gly Ser Val Phe Leu Thr Lys Glu Glu Asp Thr
                    580                    585                    590
Lys Pro Val Glu Thr Glu Arg Val Ala Gly Ala Val Thr Gly Leu Trp
                    595                    600                    605
Arg Lys Gln Leu Trp Gln Gly Glu Lys Lys Pro Val Asp Phe Phe Val
                    610                    615                    620
Val Lys Gly Ile Val Glu Gly Leu Leu Asp Lys Leu Asn Val Leu Asp
                    625                    630                    635                    640
Ser Ile Glu Phe Val Gln Ser Glu Arg Lys Gln Leu His Pro Gly Arg
                    645                    650                    655
Thr Ala Asn Ile Leu Leu Asn Gly Ser Leu Ile Gly Phe Ile Gly Gln
                    660                    665                    670
Val His Pro Ser Leu Glu Lys Glu Leu Asp Ile Lys Glu Thr Tyr Val
                    675                    680                    685
Phe Glu Leu Asp Leu His Ala Leu Leu Ala Ala Glu Thr Ala Pro Leu
                    690                    695                    700
Val Tyr Thr Ala Ile Pro Lys Tyr Pro Ser Val Thr Arg Asp Ile Ala
                    705                    710                    715                    720
Leu Val Thr Asp Lys Thr Val Thr Ser Gly Gln Leu Glu Ser Val Ile
                    725                    730                    735
Lys Glu Ala Gly Gly Lys Leu Leu Lys Glu Val Thr Val Phe Asp Val
                    740                    745                    750
Tyr Glu Gly Glu His Met Glu Glu Gly Lys Lys Ser Val Ala Phe Ser
                    755                    760                    765
Leu Gln Tyr Val Asn Pro Glu Gln Thr Leu Thr Glu Glu Glu Val Thr
                    770                    775                    780
Lys Ala His Ser Lys Val Leu Lys Ala Leu Glu Asp Thr Tyr Gln Ala
                    785                    790                    795                    800
Val Leu Arg Gly

```

<210> 87

<211> 595

<212> PRT

<213> *Saccharomyces cerevisiae*

<400> 87

Met	Pro	Thr	Val	Ser	Val	Asn	Lys	Gln	Gln	Leu	Phe	Asp	Leu	Leu	Gly
1				5				10					15		
Lys	Asn	Tyr	Thr	Ser	Gln	Glu	Phe	Asp	Glu	Leu	Cys	Phe	Glu	Phe	Gly
			20					25					30		
Met	Glu	Met	Asp	Glu	Asp	Thr	Thr	Glu	Glu	Ala	Leu	Lys	Thr	Gly	Glu
		35					40					45			
Glu	Pro	Glu	Leu	Lys	Leu	Asp	Ile	Ser	Ala	Asn	Arg	Tyr	Asp	Leu	Leu
	50					55				60					
Cys	Ile	Glu	Gly	Ile	Ser	Gln	Ser	Leu	Asn	Glu	Tyr	Leu	Glu	Arg	Lys
65					70					75					80
Glu	Arg	Pro	Asp	Tyr	Lys	Leu	Ser	Lys	Pro	Thr	Thr	Lys	Leu	Ile	Ile
			85						90					95	
Asp	Lys	Ser	Thr	Glu	Gln	Ile	Arg	Pro	Phe	Ala	Thr	Ala	Ala	Val	Leu
			100					105					110		
Arg	Asn	Ile	Lys	Leu	Asn	Glu	Lys	Ser	Tyr	Ala	Ser	Phe	Ile	Ala	Leu
		115					120					125			
Gln	Asp	Lys	Leu	His	Ala	Asn	Leu	Cys	Arg	Asn	Arg	Ser	Leu	Val	Ala
	130					135					140				
Met	Gly	Thr	His	Asp	Leu	Asp	Ser	Ile	Glu	Gly	Pro	Phe	His	Tyr	Arg

```

145      150      155      160
Ala Leu Pro Pro Lys Asp Ile Lys Phe Val Pro Leu Asn Gln Thr Gln
165      170      175
Glu Phe Thr Gly Asp Lys Leu Ile Glu Phe Tyr Lys Ser Pro Glu Gln
180      185      190
Lys Asn Asn Ile Gly Arg Tyr Val His Ile Ile Glu Asp Ser Pro Val
195      200      205
Phe Pro Val Ile Met Asp Ser Lys Asp Arg Val Cys Ser Leu Pro Pro
210      215      220
Leu Ile Asn Ser Glu His Ser Lys Ile Ser Val Asn Thr Arg Asn Ile
225      230      235
Leu Ile Asp Ile Thr Ala Thr Asp Lys Thr Lys Ala Glu Ile Val Leu
245      250      255
Asn Ile Leu Thr Thr Met Phe Ser Arg Tyr Cys Asp Glu Pro Phe Thr
260      265      270
Val Glu Pro Val Glu Ile Val Ser Glu His Asn Gly Gln Ser Arg Leu
275      280      285
Ala Pro Asn Phe Asn Asp Arg Ile Met Asp Val Ser Ile Lys Tyr Ile
290      295      300
Asn Ser Cys Leu Gly Leu Asp Gln Ser Ala Asp Glu Ile Ala His Cys
305      310      315
Leu Lys Lys Met Ser Leu His Ala Val Gln Ser Lys Glu Asp Lys Asp
325      330      335
Ile Leu His Val Asp Ile Pro Val Thr Arg Pro Asp Ile Leu His Ala
340      345      350
Cys Asp Ile Met Glu Asp Ala Ala Val Gly Tyr Gly Phe Asn Asn Leu
355      360      365
Pro Lys Gly Glu Lys Leu Ser Asn Ala Asn Phe Ile Ala Lys Pro Leu
370      375      380
Pro Ile Asn Lys Val Ser Asp Ile Phe Arg Val Ala Ser Ser Gln Ala
385      390      395
Thr Trp Val Glu Val Leu Pro Leu Thr Leu Cys Ser His Asp Glu Asn
405      410      415
Phe Lys Phe Leu Arg Gln Ser Asp Asn Gly Asp Leu Ala Val Lys Leu
420      425      430
Ala Asn Pro Lys Thr Leu Glu Tyr Gln Val Val Arg Thr Thr Leu Leu
435      440      445
Pro Gly Ile Leu Lys Thr Val Lys Glu Asn Arg Lys His Ser Leu Pro
450      455      460
Ile Lys Val Phe Glu Thr Gly Asp Val Val Phe Lys Asp Asp Lys Leu
465      470      475
Glu Arg Lys Ala Tyr Asn Glu Arg His Trp Ala Ala Ile Tyr Val Gly
485      490      495
Lys Asn Ser Gly Phe Glu Ile Ile Gln Gly Leu Leu Gly Lys Ile Met
500      505      510
Gln Thr Phe Arg Thr Glu Trp Ile Ala Asp Tyr Gly Ala Ala Ser
515      520      525
Gly Arg Gly Tyr Trp Ile Glu Asp Asp Ser Val Lys Thr Tyr Phe
530      535      540
Pro Gly Arg Gly Ala Lys Val Met Phe Arg Ser Lys Glu Gly Ala Glu
545      550      555
Pro Lys Gln Ile Gly His Leu Gly Val Leu His Pro Glu Val Met Met
565      570      575
Asn Phe Asp Val Pro Phe Ala Ala Ser Phe Val Glu Val Asn Ala Glu
580      585      590
Val Phe Leu
595

```

<210> 88

<211> 792

<212> PRT

<213> Chlamydophila pneumoniae

<400> 88

```

Met Arg Ile Pro Ile Thr Leu Leu Gln Thr Tyr Phe Ser Glu Pro Leu
 1          5          10          15
Ser Thr Lys Glu Ile Leu Glu Ala Cys Asp His Ile Gly Ile Glu Ala
 20          25          30
Glu Ile Glu Asn Thr Thr Leu Tyr Ser Phe Ala Ser Val Ile Thr Ala
 35          40          45
Lys Ile Leu His Thr Ile Pro His Pro Asn Ala Asp Lys Leu Arg Val
 50          55          60
Ala Thr Leu Thr Asp Gly Glu Lys Glu His Gln Val Val Cys Gly Ala
 65          70          75          80
Pro Asn Cys Glu Ala Gly Leu Ile Val Ala Leu Ala Leu Pro Gly Ala
 85          90          95
Lys Leu Phe Asp Ser Glu Gly Gln Ala Tyr Thr Ile Lys Lys Ser Lys
100          105          110
Leu Arg Gly Val Glu Ser Gln Gly Met Cys Cys Gly Ala Asp Glu Leu
115          120          125
Gly Leu Asp Glu Leu Gln Ile Gln Glu Arg Ala Leu Leu Glu Leu Pro
130          135          140
Glu Ala Thr Pro Leu Gly Glu Asp Leu Ala Thr Val Leu Gly Asn Thr
145          150          155          160
Ser Leu Glu Ile Ser Leu Thr Pro Asn Leu Gly His Cys Ala Ser Phe
165          170          175
Leu Gly Leu Ala Arg Glu Ile Cys His Val Thr Gln Ala Asn Leu Val
180          185          190
Ile Pro Lys Glu Phe Ser Phe Glu Asn Leu Pro Thr Thr Ala Leu Asp
195          200          205
Met Gly Asn Asp Pro Asp Ile Cys Pro Phe Phe Ser Tyr Val Val Ile
210          215          220
Thr Gly Ile Ser Ala Gln Pro Ser Pro Ile Lys Leu Gln Glu Ser Leu
225          230          235          240
Gln Ala Leu Lys Gln Lys Pro Ile Asn Ala Ile Val Asp Ile Thr Asn
245          250          255
Tyr Ile Met Leu Ser Leu Gly Gln Pro Leu His Ala Tyr Asp Ala Ser
260          265          270
His Val Ala Leu Asp Ser Leu Arg Val Glu Lys Leu Ser Thr Pro Glu
275          280          285
Ser Leu Thr Leu Leu Asn Gly Glu Thr Val Leu Leu Pro Ser Gly Val
290          295          300
Pro Val Val Arg Asp Asp His Ser Leu Leu Gly Leu Gly Val Met
305          310          315          320
Gly Ala Lys Ala Pro Ser Phe Gln Glu Thr Thr Thr Thr Val Ile
325          330          335
Lys Ala Ala Tyr Phe Leu Pro Glu Ala Leu Arg Ala Ser Gln Lys Leu
340          345          350
Leu Pro Ile Pro Ser Glu Ser Ala Tyr Arg Phe Thr Arg Gly Ile Asp
355          360          365
Pro Gln Asn Val Val Pro Ala Leu Gln Ala Ala Ile His Tyr Ile Leu
370          375          380
Glu Ile Phe Pro Glu Ala Thr Ile Ser Pro Ile Tyr Ser Ser Gly Glu
385          390          395          400
Ile Cys Arg Glu Leu Lys Glu Val Ala Leu Arg Pro Lys Thr Leu Gln

```

```

      405      410      415
Arg Ile Leu Gly Lys Ser Phe Ser Ile Glu Ile Leu Ser Gln Lys Leu
      420      425      430
Gln Ser Leu Gly Phe Ser Thr Thr Pro Gln Glu Thr Ser Leu Leu Val
      435      440      445
Lys Val Pro Ser Tyr Arg His Asp Ile Asn Glu Glu Ile Asp Leu Val
      450      455      460
Glu Glu Ile Cys Arg Thr Glu Ser Trp Asn Ile Glu Thr Gln Asn Pro
465      470      475      480
Val Ser Cys Tyr Thr Pro Ile Tyr Lys Leu Lys Arg Glu Thr Ala Gly
      485      490      495
Phe Leu Ala Asn Ala Gly Leu Gln Glu Phe Phe Thr Pro Asp Leu Leu
500      505      510
Asp Pro Glu Thr Val Ala Leu Thr Arg Lys Glu Lys Glu Glu Ile Ser
515      520      525
Leu Gln Gly Ser Lys His Thr Thr Val Leu Arg Ser Ser Leu Leu Pro
530      535      540
Gly Leu Leu Lys Ser Ala Ala Thr Asn Leu Asn Arg Gln Ala Pro Ser
545      550      555      560
Val Gln Ala Phe Glu Ile Gly Thr Val Tyr Ala Lys His Gly Glu Gln
565      570      575
Cys Gln Glu Thr Gln Thr Leu Ala Ile Leu Leu Thr Glu Asp Gly Glu
580      585      590
Ser Arg Ser Trp Leu Pro Lys Pro Ser Leu Ser Phe Tyr Ser Leu Lys
595      600      605
Gly Trp Val Glu Arg Leu Leu Tyr His His His Leu Ser Ile Asp Ala
610      615      620
Leu Thr Leu Glu Ser Ser Ala Leu Cys Glu Phe His Pro Tyr Gln Gln
625      630      635      640
Gly Val Leu Arg Ile His Lys Gln Ser Phe Ala Thr Leu Gly Gln Val
645      650      655
His Pro Glu Leu Ala Lys Lys Ala Gln Ile Lys His Pro Val Phe Phe
660      665      670
Ala Glu Leu Asn Leu Asp Leu Leu Cys Lys Met Leu Lys Lys Thr Thr
675      680      685
Lys Leu Tyr Lys Pro Tyr Ala Ile Tyr Pro Ser Ser Phe Arg Asp Leu
690      695      700
Thr Leu Thr Val Pro Glu Asp Ile Pro Ala Asn Leu Leu Arg Gln Lys
705      710      715      720
Leu Leu His Glu Gly Ser Lys Trp Leu Glu Ser Val Thr Ile Ile Ser
725      730      735
Ile Tyr Gln Asp Lys Ser Leu Glu Thr Arg Asn Lys Asn Val Ser Leu
740      745      750
Arg Leu Val Phe Gln Asp Tyr Glu Arg Thr Leu Ser Asn Gln Asp Ile
755      760      765
Glu Glu Glu Tyr Cys Arg Leu Val Ala Leu Leu Asn Glu Leu Leu Thr
770      775      780
Asp Thr Lys Gly Thr Ile Asn Ser
785      790

```

<210> 89

<211> 831

<212> PRT

<213> Mycobacterium tuberculosis

<400> 89

Met Arg Leu Pro Tyr Ser Trp Leu Arg Glu Val Val Ala Val Gly Ala

1	5	10	15
Ser Gly Trp Asp Val Thr Pro Gly Glu Leu Glu Gln Thr Leu Leu Arg	20	25	30
Ile Gly His Glu Val Glu Glu Val Ile Pro Leu Gly Pro Val Asp Gly	35	40	45
Pro Val Thr Val Gly Arg Val Ala Asp Ile Glu Glu Leu Thr Gly Tyr	50	55	60
Lys Lys Pro Ile Arg Ala Cys Ala Val Asp Ile Gly Asp Arg Gln Tyr	65	70	75
Arg Glu Ile Ile Cys Gly Ala Thr Asn Phe Ala Val Gly Asp Leu Val	85	90	95
Val Val Ala Leu Pro Gly Ala Thr Leu Pro Gly Gly Phe Thr Ile Ser	100	105	110
Ala Arg Lys Ala Tyr Gly Arg Asn Ser Asp Gly Met Ile Cys Ser Ala	115	120	125
Ala Glu Leu Asn Leu Gly Ala Asp His Ser Gly Ile Leu Val Leu Pro	130	135	140
Pro Gly Ala Ala Glu Pro Gly Ala Asp Gly Ala Gly Val Leu Gly Leu	145	150	155
Asp Asp Val Val Phe His Leu Ala Ile Thr Pro Asp Arg Gly Tyr Cys	165	170	175
Met Ser Val Arg Gly Leu Ala Arg Glu Leu Ala Cys Ala Tyr Asp Leu	180	185	190
Asp Phe Val Asp Pro Ala Ser Asn Ser Arg Val Pro Pro Leu Pro Ile	195	200	205
Glu Gly Pro Ala Trp Pro Leu Thr Val Gln Pro Glu Thr Gly Val Arg	210	215	220
Arg Phe Ala Leu Arg Pro Val Ile Gly Ile Asp Pro Ala Ala Val Ser	225	230	235
Pro Trp Trp Leu Gln Arg Arg Leu Leu Leu Cys Gly Ile Arg Ala Thr	245	250	255
Cys Pro Ala Val Asp Val Thr Asn Tyr Val Met Leu Glu Leu Gly His	260	265	270
Pro Met His Ala His Asp Arg Asn Arg Ile Ser Gly Thr Leu Gly Val	275	280	285
Arg Phe Ala Arg Ser Gly Glu Thr Ala Val Thr Leu Asp Gly Ile Glu	290	295	300
Arg Lys Leu Asp Thr Ala Asp Val Leu Ile Val Asp Asp Ala Ala Thr	305	310	315
Ala Ala Ile Gly Gly Val Met Gly Ala Ala Ser Thr Glu Val Arg Ala	325	330	335
Asp Ser Thr Asp Val Leu Leu Glu Ala Ala Ile Trp Asp Pro Ala Ala	340	345	350
Val Ser Arg Thr Gln Arg Arg Leu His Leu Pro Ser Glu Ala Ala Arg	355	360	365
Arg Tyr Glu Arg Thr Val Asp Pro Ala Ile Ser Val Ala Ala Leu Asp	370	375	380
Arg Cys Ala Arg Leu Leu Ala Asp Ile Ala Gly Gly Glu Val Ser Pro	385	390	395
Thr Leu Thr Asp Trp Arg Gly Asp Pro Pro Cys Asp Asp Trp Ser Pro	405	410	415
Pro Pro Ile Arg Met Gly Val Asp Val Pro Asp Arg Ile Ala Gly Val	420	425	430
Ala Tyr Pro Gln Gly Thr Thr Ala Arg Arg Leu Ala Gln Ile Gly Ala	435	440	445
Val Val Thr His Asp Gly Asp Thr Leu Thr Val Thr Pro Pro Ser Trp	450	455	460
Arg Pro Asp Leu Arg Gln Pro Ala Asp Leu Val Glu Glu Val Leu Arg			

```

465      470      475      480
Leu Glu Gly Leu Glu Val Ile Pro Ser Val Leu Pro Pro Ala Pro Ala
      485      490      495
Gly Arg Gly Leu Thr Ala Gly Gln Gln Arg Arg Arg Thr Ile Gly Arg
      500      505      510
Ser Leu Ala Leu Ser Gly Tyr Val Glu Ile Leu Pro Thr Pro Phe Leu
      515      520      525
Pro Ala Gly Val Phe Asp Leu Trp Gly Leu Glu Ala Asp Asp Ser Arg
      530      535      540
Arg Met Thr Thr Arg Val Leu Asn Pro Leu Glu Ala Asp Arg Pro Gln
      545      550      555
Leu Ala Thr Thr Leu Leu Pro Ala Leu Leu Glu Ala Leu Val Arg Asn
      565      570      575
Val Ser Arg Gly Leu Val Asp Val Ala Leu Phe Ala Ile Ala Gln Val
      580      585      590
Val Gln Pro Thr Glu Gln Thr Arg Gly Val Gly Leu Ile Pro Val Asp
      595      600      605
Arg Arg Pro Thr Asp Asp Glu Ile Ala Met Leu Asp Ala Ser Leu Pro
      610      615      620
Arg Gln Pro Gln His Val Ala Ala Val Leu Ala Gly Leu Arg Glu Pro
      625      630      635
Arg Gly Pro Trp Gly Pro Gly Arg Pro Val Glu Ala Ala Asp Ala Phe
      645      650      655
Glu Ala Val Arg Ile Ile Ala Arg Ala Ser Arg Val Asp Val Thr Leu
      660      665      670
Arg Pro Ala Gln Tyr Leu Pro Trp His Pro Gly Arg Cys Ala Gln Val
      675      680      685
Phe Val Gly Glu Ser Ser Val Gly His Ala Gly Gln Leu His Pro Ala
      690      695      700
Val Ile Glu Arg Ser Gly Leu Pro Lys Gly Thr Cys Ala Val Glu Leu
      705      710      715
Asn Leu Asp Ala Ile Pro Cys Ser Ala Pro Leu Pro Ala Pro Arg Val
      725      730      735
Ser Pro Tyr Pro Ala Val Phe Gln Asp Val Ser Leu Val Val Ala Ala
      740      745      750
Asp Ile Pro Ala Gln Ala Val Ala Asp Ala Val Arg Ala Gly Ala Gly
      755      760      765
Asp Leu Leu Glu Asp Ile Ala Leu Phe Asp Val Phe Thr Gly Pro Gln
      770      775      780
Ile Gly Glu His Arg Lys Ser Leu Thr Phe Ala Leu Arg Phe Arg Ala
      785      790      795
Pro Asp Arg Thr Leu Thr Glu Asp Asp Ala Ser Ala Ala Arg Asp Ala
      805      810      815
Ala Val Gln Ser Ala Ala Glu Arg Val Gly Ala Val Leu Arg Gly
      820      825      830

```

<210> 90

<211> 764

<212> PRT

<213> Helicobacter pylori

<400> 90

```

Met Lys Leu Ser Ile Asn Asp Leu Asn Val Phe Val Asn Thr Pro Lys
 1          5          10          15
Asp Ile Ala Lys Leu Cys Glu Asp Leu Ser Arg Leu Gly Leu Glu Val
      20          25          30
Glu Ser Cys Ile Pro Cys Ile Ala Pro Lys Asn Val Val Val Gly Lys

```

		35				40					45				
Ile Leu Glu Lys Ala Pro His Lys Asn Ala Glu Lys Leu Ser Val Cys															
Gln Val Asp Val Gly Lys Glu Val Leu Gln Ile Val Cys Gly Ala Lys	50				55					60					
Asn Val Ala Pro Asn Gln Phe Val Pro Val Ala Leu Asn Gly Ala Leu	65				70				75					80	
Ile Gly Ser Thr Thr Ile Ala Lys Thr Glu Leu Arg Gly Val Glu Ser				85				90					95		
His Gly Met Ile Cys Ser Ser Ile Glu Leu Gly Phe Pro Lys Ile Asn			100				105						110		
Asp Gly Ile Leu Glu Leu Asp Glu Ser Val Gly Glu Leu Val Leu Gly			115			120				125					
Lys Glu Leu Asn Glu Tyr Ala Pro Phe Asn Thr His Val Leu Glu Ile			130		135				140						
Ser Leu Thr Pro Asn Arg Gly Asp Cys Leu Ser Val Leu Gly Ile Ala	145			150				155					160		
Arg Glu Ile Ser Ala Phe Tyr His Thr Pro Leu Lys Pro Ile Lys Ala			165				170						175		
Leu Asn Phe Thr Pro Lys Ser Gly Leu Ile Thr Leu Ser Ala Gly Glu			180			185						190			
Asn Ile Glu Ser His Leu Ala Tyr Tyr Leu Ile Cys Asn His Ser Leu		195				200				205					
Lys Thr Pro Leu Asn Ile Lys Leu Ser Leu Ala His Asn Asn Ala Leu	210			215					220						
Ser Glu Asn Asp Leu Asn Asn Phe Ile Glu Phe Ser Thr His Phe Ser				230				235					240		
Gly Val Ile Met Asn Ala Tyr Ser Leu Asn Thr Thr Pro Met Asp Leu			245				250						255		
Ser Val Lys Asn Asp Glu Asn Asn Glu Ser Val Tyr Ile Asn His			260			265						270			
Gln Lys Arg Ser Thr Ile Ala Ile Lys His Gln Val Gln Lys Asp Leu		275			280					285					
Ser Glu Cys Leu Leu Leu Glu Ala Ser Tyr Thr Asp Pro Ile Ser Leu	290			295					300						
Ser Leu Lys Leu His Ala Leu Lys Asp Lys Thr Leu Gln Lys Asp Asn			310					315					320		
Ala Leu Ile Tyr Arg Ser Ala Arg Gly Ser Asn Pro Asn Leu Ser Asp			325				330						335		
Gly Leu Asn Phe Leu Ser Ala His Leu Lys Ala Thr Ile Leu Glu Ser			340				345					350			
Lys Gln Thr Glu His Ser Leu Lys Asp Arg Thr Leu Thr Phe Gln Leu		355			360					365					
Glu Asp Ile Thr Glu Ile Leu Gly Leu Ala Val Glu Lys Glu Lys Ile					370		375			380					
Gln Gly Ile Leu Lys Asn Leu Gly Phe Lys Val Ser Val Lys Glu Pro	385			390				395					400		
Asn Ser Lys Pro Gln Ile Leu Glu Val Ile Ala Pro Asn Phe Arg His			405				410						415		
Asp Ile Lys Thr Ile Gln Asp Ile Ala Glu Glu Ile Leu Arg Phe Val			420				425					430			
Gly Ile Asp Asn Leu Val Ser Lys Pro Leu His Cys Val Ser Ser Lys		435				440				445					
Asn Ser Asn Pro Asn Tyr Asp Thr His Arg Phe Phe Glu Asn Leu Lys					455				460						
His Lys Ala Leu Ala Cys Gly Phe Lys Glu Val Ile His Tyr Val Phe	465			470				475					480		
Tyr Ser Lys Glu Lys Gln Gln Lys Leu Gly Phe Glu Val Leu Glu Asp			485				490						495		

130	135	140
Asp Val Arg Thr Tyr Leu	Glu Leu Ala Asp Tyr Thr	Ile Glu Val Gly
145	150	155
Leu Thr Pro Asn Arg Gly	Asp Cys Leu Ser Leu	Ala Gly Leu Ala Arg
165	170	175
Glu Val Ser Ala Ile Tyr	Asp Val Pro Leu Ala	Pro Val Ala Val Asp
180	185	190
Ala Val Ala Ala Gln His	Asp Glu Thr Arg Pro	Val Glu Leu Ala Ala
195	200	205
Pro Ala Ala Cys Pro Arg	Tyr Leu Gly Arg Val	Ile Arg Asn Val Asp
210	215	220
Leu Ser Arg Pro Thr Pro	Leu Trp Met Val Glu	Arg Leu Arg Arg Ser
225	230	235
Asp Ile Arg Ser Ile Asp	Pro Val Val Asp Val	Thr Asn Tyr Val Met
245	250	255
Ile Glu Leu Gly Gln Pro	Met His Ala Phe Asp	Leu Ala Glu Ile Asn
260	265	270
Gly Gly Val Arg Val Arg	Met Ala Glu Asp Gly	Glu Lys Leu Val Leu
275	280	285
Leu Asp Gly Gln Glu Ile	Thr Leu Arg Ala Asp	Thr Leu Val Ile Ala
290	295	300
Asp His Gln Arg Ala Leu	Ala Ile Ala Gly Val	Met Gly Gly Glu His
305	310	315
Ser Gly Val Ser Asp Ser	Thr Arg Asp Leu Phe	Leu Glu Ala Ala Phe
325	330	335
Phe Asp Thr Ile Ala Leu	Ala Gly Lys Ala Arg	Ser Tyr Gly Leu His
340	345	350
Thr Asp Ser Ser His Arg	Phe Glu Arg Gly Val	Asp Ser Gln Leu Ala
355	360	365
Arg Lys Ala Met Glu Arg	Ala Thr Arg Leu Ile	Leu Asp Ile Val Gly
370	375	380
Gly Glu Pro Gly Pro Ile	Val Glu Gln Val Ser	Glu Ala His Leu Pro
385	390	395
Lys Val Ala Pro Ile Thr	Leu Arg Ala Glu Arg	Val Thr Gln Met Leu
405	410	415
Gly Met Pro Leu Asp Ala	Ala Glu Ile Val Arg	Leu Leu Gln Ala Leu
420	425	430
Glu Leu Thr Val Val Ala	Asp Gly Glu Gly Gln	Trp Ser Val Gly Val
435	440	445
Pro Ser His Arg Phe Asp	Ile Ser Leu Glu Val	Asp Leu Ile Glu Glu
450	455	460
Leu Ala Arg Leu Tyr Gly	Tyr Asn Arg Leu Pro	Val Arg Tyr Pro Gln
465	470	475
Ala Arg Leu Ala Pro Asn	Asn Lys Pro Glu Ala	Arg Ala Ala Leu Pro
485	490	495
Leu Leu Arg Arg Leu Leu	Val Ala Arg Gly Tyr	Gln Glu Ala Ile Thr
500	505	510
Phe Ser Phe Ile Asp Pro	Ala Leu Phe Glu Leu	Phe Asp Pro Gly Thr
515	520	525
Gln Pro Leu Thr Leu Ala	Asn Pro Ile Ser Ala	Asp Met Ala Ala Met
530	535	540
Arg Ser Ser Leu Trp Pro	Gly Leu Val Lys Ala	Leu Gln His Asn Leu
545	550	555
Asn Arg Gln Gln Ser Arg	Val Arg Leu Phe Glu	Ser Gly Leu Arg Phe
565	570	575
Val Gly Gln Leu Glu Gly	Leu Lys Gln Glu Ala	Met Leu Ala Gly Ala
580	585	590
Ile Cys Gly Lys Arg Leu	Pro Glu Gly Trp Ala	Asn Gly Arg Asp Gly

```

      595      600      605
Val Asp Phe Phe Asp Ala Lys Ala Asp Val Glu Ala Val Leu Ala Ser
610      615      620
Ala Gly Ala Leu Gly Asp Phe Ser Phe Val Pro Gly Glu His Pro Ala
625      630      635      640
Leu His Pro Gly Gln Thr Ala Arg Ile Glu Arg Glu Gly Arg Leu Val
      645      650      655
Gly Tyr Leu Gly Ala Leu His Pro Glu Leu Ala Lys Lys Leu Asp Leu
660      665      670
Glu Gln Pro Val Phe Leu Phe Glu Leu Leu Leu Ala Glu Val Val Asp
675      680      685
Gly His Leu Pro Lys Phe Arg Glu Leu Ser Arg Phe Pro Glu Val Arg
690      695      700
Arg Asp Leu Ala Leu Leu Val Asp Gln Asp Val Pro Ala Gln Asp Ile
705      710      715      720
Leu Thr Gln Ile Arg Ala Ala Ala Gly Glu Trp Leu Thr Asp Leu Arg
      725      730      735
Leu Phe Asp Val Tyr His Gly Lys Gly Ile Asp Pro His Arg Lys Ser
740      745      750
Leu Ala Val Gly Leu Thr Trp Gln His Pro Ser Arg Thr Leu Asn Asp
755      760      765
Asp Glu Val Asn Ser Thr Thr Gln Asn Ile Val Thr Ser Leu Glu Glu
770      775      780
Arg Phe Asn Ala Thr Leu Arg Lys
785      790

```

<210> 92

<211> 551

<212> PRT

<213> Escherichia coli

<400> 92

Met	Thr	Gln	Val	Ala	Lys	Lys	Ile	Leu	Val	Thr	Cys	Ala	Leu	Pro	Tyr
1				5					10					15	
Ala	Asn	Gly	Ser	Ile	His	Leu	Gly	His	Met	Leu	Glu	His	Ile	Gln	Ala
			20					25					30		
Asp	Val	Trp	Val	Arg	Tyr	Gln	Arg	Met	Arg	Gly	His	Glu	Val	Asn	Phe
		35					40					45			
Ile	Cys	Ala	Asp	Asp	Ala	His	Gly	Thr	Pro	Ile	Met	Leu	Lys	Ala	Gln
	50					55					60				
Gln	Leu	Gly	Ile	Thr	Pro	Glu	Gln	Met	Ile	Gly	Glu	Met	Ser	Gln	Glu
65					70					75					80
His	Gln	Thr	Asp	Phe	Ala	Gly	Phe	Asn	Ile	Ser	Tyr	Asp	Asn	Tyr	His
			85					90						95	
Ser	Thr	His	Ser	Glu	Glu	Asn	Arg	Gln	Leu	Ser	Glu	Leu	Ile	Tyr	Ser
			100					105						110	
Arg	Leu	Lys	Glu	Asn	Gly	Phe	Ile	Lys	Asn	Arg	Thr	Ile	Ser	Gln	Leu
	115						120					125			
Tyr	Asp	Pro	Glu	Lys	Gly	Met	Phe	Leu	Pro	Asp	Arg	Phe	Val	Lys	Gly
	130					135					140				
Thr	Cys	Pro	Lys	Cys	Lys	Ser	Pro	Asp	Gln	Tyr	Gly	Asp	Asn	Cys	Glu
145					150					155					160
Val	Cys	Gly	Ala	Thr	Tyr	Ser	Pro	Thr	Glu	Leu	Ile	Glu	Pro	Lys	Ser
				165					170					175	
Val	Val	Ser	Gly	Ala	Thr	Pro	Val	Met	Arg	Asp	Ser	Glu	His	Phe	Phe
		180						185					190		
Phe	Asp	Leu	Pro	Ser	Phe	Ser	Glu	Met	Leu	Gln	Ala	Trp	Thr	Arg	Ser

```

      195      200      205
Gly Ala Leu Gln Glu Gln Val Ala Asn Lys Met Gln Glu Trp Phe Glu
  210      215      220
Ser Gly Leu Gln Gln Trp Asp Ile Ser Arg Asp Ala Pro Tyr Phe Gly
  225      230      235      240
Phe Glu Ile Pro Asn Ala Pro Gly Lys Tyr Phe Tyr Val Trp Leu Asp
      245      250      255
Ala Pro Ile Gly Tyr Met Gly Ser Phe Lys Asn Leu Cys Asp Lys Arg
      260      265      270
Gly Asp Ser Val Ser Phe Asp Glu Tyr Trp Lys Lys Asp Ser Thr Ala
  275      280      285
Glu Leu Tyr His Phe Ile Gly Lys Asp Ile Val Tyr Phe His Ser Leu
  290      295      300
Phe Trp Pro Ala Met Leu Glu Gly Ser Asn Phe Arg Lys Pro Ser Asn
  305      310      315      320
Leu Phe Val His Gly Tyr Val Thr Val Asn Gly Ala Lys Met Ser Lys
      325      330      335
Ser Arg Gly Thr Phe Ile Lys Ala Ser Thr Trp Leu Asn His Phe Asp
      340      345      350
Ala Asp Ser Leu Arg Tyr Tyr Tyr Thr Ala Lys Leu Ser Ser Arg Ile
  355      360      365
Asp Asp Ile Asp Leu Asn Leu Glu Asp Phe Val Gln Arg Val Asn Ala
  370      375      380
Asp Ile Val Asn Lys Val Val Asn Leu Ala Ser Arg Asn Ala Gly Phe
  385      390      395      400
Ile Asn Lys Arg Phe Asp Gly Val Leu Ala Ser Glu Leu Ala Asp Pro
      405      410      415
Gln Leu Tyr Lys Thr Phe Thr Asp Ala Ala Glu Val Ile Gly Glu Ala
  420      425      430
Trp Glu Ser Arg Glu Phe Gly Lys Ala Val Arg Glu Ile Met Ala Leu
  435      440      445
Ala Asp Leu Ala Asn Arg Tyr Val Asp Glu Gln Ala Pro Trp Val Val
  450      455      460
Ala Lys Gln Glu Gly Arg Asp Ala Asp Leu Gln Ala Ile Cys Ser Met
  465      470      475      480
Gly Ile Asn Leu Phe Arg Val Leu Met Thr Tyr Leu Lys Pro Val Leu
      485      490      495
Pro Lys Leu Thr Glu Arg Ala Glu Ala Phe Leu Asn Thr Glu Leu Thr
  500      505      510
Trp Asp Gly Ile Gln Gln Pro Leu Leu Gly His Lys Val Asn Pro Phe
  515      520      525
Lys Ala Leu Tyr Asn Arg Ile Asp Met Arg Gln Val Glu Ala Leu Val
  530      535      540
Glu Ala Ser Lys Glu Glu Val
  545      550

```

<210> 93

<211> 594

<212> PRT

<213> Homo sapiens

<400> 93

```

Thr Met Leu Arg Thr Ser Val Leu Arg Leu Leu Gly Arg Thr Gly Ala
  1      5      10      15
Ser Arg Leu Ser Leu Leu Glu Asp Phe Gly Pro Arg Tyr Tyr Ser Ser
      20      25      30
Gly Ser Leu Ser Ala Gly Asp Asp Ala Cys Asp Val Arg Ala Tyr Phe

```


				500					505					510					
Val	Ala	Leu	Glu	Cys	Leu	Arg	Val	Phe	Gly	Thr	Leu	Leu	Gln	Pro	Val				
		515						520					525						
Thr	Pro	Ser	Leu	Ala	Asp	Lys	Leu	Leu	Ser	Arg	Leu	Gly	Val	Ser	Ala				
		530					535					540							
Ser	Glu	Arg	Ser	Leu	Gly	Glu	Leu	Tyr	Phe	Leu	Pro	Arg	Phe	Tyr	Gly				
545					550				555						560				
His	Pro	Cys	Pro	Phe	Glu	Gly	Arg	Arg	Leu	Gly	Pro	Glu	Thr	Gly	Leu				
				565					570					575					
Leu	Phe	Pro	Arg	Leu	Asp	Gln	Ser	Arg	Thr	Trp	Leu	Val	Lys	Ala	His				
			580					585					590						
Arg	Thr																		

<210> 94

<211> 562

<212> PRT

<213> *Xenopus laevis*

<400> 94

Met	Leu	Arg	Ser	Leu	Ala	Leu	Arg	Thr	Phe	Ala	Asn	Ile	Leu	Gly	Leu
1				5					10					15	
Ser	Ser	Arg	Ser	Gly	Ser	Thr	Ala	Ala	Met	Ser	Arg	Asn	Ala	Val	Ala
			20					25					30		
Ser	Arg	Pro	His	Leu	Tyr	Thr	Thr	Pro	Ile	Phe	Tyr	Val	Asn	Ala	Ala
		35					40					45			
Pro	His	Leu	Gly	His	Val	Tyr	Ser	Ala	Leu	Leu	Ala	Asp	Val	Gln	His
	50					55					60				
Arg	Tyr	Ser	Ala	Met	Cys	Gly	Ile	Glu	Ser	Lys	Leu	Ser	Thr	Gly	Thr
65					70					75				80	
Asp	Glu	His	Gly	Met	Lys	Val	Gln	Gln	Ala	Ala	Ser	Ala	Leu	Gly	Leu
			85						90					95	
Asp	Pro	Gln	Thr	Phe	Cys	Ser	Thr	Val	Ser	Leu	Gln	Phe	Arg	Thr	Ile
			100					105					110		
Phe	Asp	Ala	Leu	Asp	Ile	Ser	Tyr	Thr	Asp	Phe	Val	Arg	Thr	Thr	Glu
	115						120					125			
Pro	Arg	His	Ile	Glu	Ala	Val	Ser	Arg	Phe	Trp	Met	Thr	Leu	Glu	Glu
	130					135					140				
Gln	Gly	Tyr	Ile	Tyr	Lys	Gly	Thr	Tyr	Glu	Gly	Trp	Tyr	Cys	Thr	Ser
145					150					155				160	
Asp	Glu	Ala	Phe	Leu	Ser	Glu	Gly	Gln	Thr	Ala	Glu	His	Thr	Asp	Phe
			165						170					175	
Glu	Gly	Asn	Lys	Ile	Arg	Val	Ser	Leu	Glu	Ser	Gly	His	Gln	Val	His
		180						185					190		
Trp	Val	Ser	Glu	Glu	Asn	Tyr	Met	Phe	Arg	Leu	Ser	Ser	Leu	Arg	Pro
		195				200						205			
Ala	Leu	Leu	Asn	Trp	Leu	Gln	Thr	Glu	Pro	Val	His	Pro	Ala	Pro	Phe
	210					215					220				
Leu	Lys	Leu	Val	His	His	Trp	Leu	Glu	Glu	Glu	Leu	Pro	Asp	Leu	Ser
225					230					235				240	
Val	Ser	Arg	Gln	Arg	Ser	Arg	Leu	Ser	Trp	Gly	Ile	Pro	Val	Pro	Ser
			245						250					255	
Asp	Ser	Ser	His	Val	Ile	Tyr	Val	Trp	Leu	Asp	Ala	Leu	Val	Asn	Tyr
		260						265					270		
Leu	Thr	Ala	Ala	Gly	Tyr	Pro	Asn	Pro	Gln	Leu	Ala	Pro	Trp	Gly	Pro
		275					280					285			
Ser	Thr	His	Leu	Leu	Gly	Lys	Asp	Ile	Leu	Arg	Phe	His	Ala	Ile	Tyr

290	295	300
Trp Pro Ala Phe Leu Ile	Ala Ala Gly Leu Pro	Pro Pro Gln Lys Leu
305	310	315
Leu Val His Ser His Trp	Thr Ser Glu Gly Thr	Lys Met Ser Lys Ser
325	330	335
Leu Lys Asn Val Val Asp	Pro Ser Asp Cys Ile Arg	Arg Tyr Thr Thr
340	345	350
Asp Gly Leu Arg Tyr Tyr	Leu Leu Arg His Gly	Ala Pro Glu Arg Asp
355	360	365
Cys Asp Phe Thr His Arg	Thr Ala Arg Met Leu	Leu Asn Ser Glu Leu
370	375	380
Ala Asp Ala Leu Gly Gly	Leu Leu Asn Arg Cys	Thr Ala Pro Ala Ile
385	390	395
Asn Pro Met Gln His Phe	Pro Lys Phe Gln Tyr	Glu Asn Phe Pro Val
405	410	415
Ala Ser Arg Asp Gln Val	His Asp Leu Leu Gly	Ala Leu Gln Glu Leu
420	425	430
Pro Val Glu Val Asp Gln	Trp Ile Lys Lys Phe	Gln Val His Lys Ala
435	440	445
Leu Glu Cys Ile Asp Ala	Cys Val Arg Arg Ser	Asn Ala Phe Phe Gln
450	455	460
Ser Gln Ala Pro Trp Lys	Leu Gln Arg Gly Val	Glu Lys Glu Ala Ala
465	470	475
Leu Arg Asp Ser Val Ile	Tyr Leu Thr Leu Glu	Ala Leu Arg Leu Tyr
485	490	495
Ala Thr Leu Leu His Pro	Ala Val Pro Gly Leu	Ala Thr Val Val Leu
500	505	510
Asp Arg Leu Gly Val Pro	His Lys Met Arg Thr	Leu Lys Lys Asn Thr
515	520	525
Phe Leu Ala Ala Thr Arg	Gly Glu Ile Cys Tyr	Phe Gln Ala Gln Thr
530	535	540
Leu Gly Pro Asp Lys Gly	Leu Leu Phe Pro Arg	Leu Glu Lys Ser Glu
545	550	555
Ala Phe		560

<210> 95

<211> 539

<212> PRT

<213> Schizosaccharomyces pombe

<400> 95

Met Leu Arg Lys Gly Ile Cys Arg Leu Ile His Gln Val Ser Glu Ser
 1 5 10 15
 Ser Lys Lys Pro Tyr Phe Leu Thr Thr Pro Ile Phe Tyr Val Asn Ala
 20 25 30
 Ala Pro His Leu Gly His Leu Tyr Ser Leu Val Leu Thr Asp Ala Ile
 35 40 45
 Ala Arg Phe Gln Asn Leu Lys Pro Asp Val Ser Val Ile Ser Ser Thr
 50 55 60
 Gly Thr Asp Glu His Gly Leu Lys Val Gln Thr Val Ala Gln Thr Glu
 65 70 75 80
 Gly Val Ser Pro Leu Gln Leu Cys Asp Arg Asn Ser Lys Arg Phe Ala
 85 90 95
 Asp Leu Ala Val Ala Ala Asn Thr Lys Phe Thr His Phe Ile Arg Thr
 100 105 110
 Thr Asn Pro Lys His Gln Ala Ser Val Gln Glu Phe Trp Lys Thr Ile

115 120 125
 Gln Lys Ala Gly Met Ile Ser Phe Glu Arg His Glu Gly Trp Tyr Cys
 130 135 140
 Val Ser Asp Glu Thr Phe Tyr Pro Glu Ser Ala Ile Gln Lys Val Val
 145 150 155 160
 Asp Pro Ala Thr Lys Gln Glu Lys Arg Val Ser Met Glu Thr Gly Lys
 165 170 175
 Glu Val Gln Trp Ser Ser Glu Met Asn Tyr His Phe Leu Leu Ser Lys
 180 185 190
 Phe Gln Ser Arg Leu Ile Glu His Tyr Asn Lys Asn Pro Asn Phe Val
 195 200 205
 Gln Pro Ser Ile Phe His Thr Gln Val Leu Glu Glu Leu Lys Thr Gly
 210 215 220
 Ile Ser Asp Leu Ser Ile Ser Arg Pro Lys Gln Arg Leu Ser Trp Gly
 225 230 235 240
 Ile Pro Val Pro Gly Asn Ser Gln Gln Thr Ile Tyr Val Trp Leu Asp
 245 250 255
 Ala Leu Ile Asn Tyr Ile Ser Val Ile Gly Tyr Pro Trp Leu Asn Glu
 260 265 270
 Lys Ser Ser Leu Ser Ala Gly Trp Pro Ala Asn Met His Val Ile Gly
 275 280 285
 Lys Asp Ile Ile Arg Phe His Cys Ile Tyr Trp Pro Ala Phe Leu Met
 290 295 300
 Ala Ala Gly Leu Pro Leu Pro Glu Lys Ile Leu Val His Ser His Trp
 305 310 315 320
 Thr Met Asn Lys Val Lys Met Ser Lys Ser Leu Gly Asn Val Val Asp
 325 330 335
 Pro Phe Trp Leu Ile Glu Lys Tyr Gly Val Asp Thr Ile Arg Tyr Tyr
 340 345 350
 Leu Leu Lys Arg Gly Arg Leu Thr Ser Asp Ser Asn Phe Asp Ile Glu
 355 360 365
 Glu Leu Glu Lys Asp Glu Glu His Asp Leu Arg Arg Ser Leu Gly Val
 370 375 380
 Leu Leu Ser Arg Leu Gln Ser Lys Lys Leu Phe Ile Ser Asn Glu Ile
 385 390 395 400
 Gln Lys Gln Trp His Lys Lys Asp Asp Phe Thr Glu Tyr Glu Asp Ile
 405 410 415
 Val His Glu Leu Ile Glu Leu Pro Val Val Cys Ala Gln Ser Ile Asp
 420 425 430
 Gly Gly Cys Val Tyr Glu Val Ile Asn Leu Val Gln Ser Val Leu Arg
 435 440 445
 Arg Val Thr Lys Leu Phe Gln Leu Lys Glu Pro Trp Lys Leu Ser Asp
 450 455 460
 Asp Ser Gln Glu Lys Ile Asp Thr Leu Met Leu Val Ala His Ser Leu
 465 470 475 480
 Arg Ile Ser Gly Ile Leu Leu Gln Pro Ile Met Pro Thr Lys Ser Thr
 485 490 495
 Glu Leu Leu Asp Gln Leu Gly Ile Pro Lys Asn Gln Arg Ser Leu Gln
 500 505 510
 Asn Ala Thr Asn Val Phe Glu Pro Thr Glu Phe Thr Phe His Ser Gly
 515 520 525
 Asn Asn Ser His Leu Phe Asp Lys Arg Thr Gln
 530 535

<210> 96

<211> 913

<212> PRT

<213> Danio rerio

<400> 96

```

Met Lys Leu Phe Ile Gly Glu Gly Asn Pro His Cys Leu Lys Val Leu
 1      5      10      15
Ala Ala Leu Glu Leu Ser Gly Val Arg Cys Glu Thr Gln Leu Val Lys
 20      25      30
His Glu Glu Lys Val Val Pro Tyr Leu Thr His Pro Val Leu Pro Ile
 35      40      45
Leu Gln Leu Pro Ser Gly Gln His Leu Phe Ser Pro Asn Ser Ile Cys
 50      55      60
Gln Tyr Leu Phe Asp Ile Ser Gly Gln Lys Ala Thr Asp Ala Thr Asn
 65      70      75      80
Gln Trp Leu Glu Trp Glu Ala Thr Asn Leu Gln Pro Ala Val Leu Gln
 85      90      95
Ser Leu Gln Leu Val Ala Leu Gln Gly Lys Arg Val Glu Ser Ala Ala
100      105      110
Val Met Lys Glu Pro Leu Ser Trp Leu Glu Gln Ser Leu Ser Lys Arg
115      120      125
Lys Thr Ser Phe Leu Thr Asp Glu Val Val Ser Val Ala Asp Val Val
130      135      140
Leu Trp Ala Ala Leu Tyr Pro Leu Leu Ser Asp Ser Ala Phe Glu Pro
145      150      155      160
Gly Asp Leu Gln Ala Val Arg Gly Trp Phe Glu Arg Val Cys Ser Val
165      170      175
Ser Ala Cys Gln Ser Ala Ala Leu Arg Val Leu Gln Gly Lys Gly Ala
180      185      190
Glu Ala Leu Lys Ser Phe Leu Gln Lys Gln Pro Val Thr His Thr Pro
195      200      205
Arg Arg Asp Ser Pro Ser Asn Ser Thr Glu Ala Glu Asp Ser Glu Arg
210      215      220
Glu Leu Ser Pro Glu Glu Ile Glu Glu Ala Ala Gln Val Tyr Ser Glu
225      230      235      240
Gly Leu Lys Asp Phe Thr Val Val Thr Glu Arg Lys His Pro Val Leu
245      250      255
Pro Gln Glu Gly Lys Arg Asn Val Leu Ile Thr Ser Ala Leu Pro Tyr
260      265      270
Val Asn Asn Val Pro His Leu Gly Asn Ile Ile Gly Cys Val Leu Ser
275      280      285
Ala Asp Val Phe Ala Arg Tyr Gly Arg Leu Arg Gly Trp Asn Leu Leu
290      295      300
Tyr Ile Cys Gly Thr Asp Glu Tyr Gly Thr Ala Thr Glu Asn Lys Ala
305      310      315      320
Arg Glu Glu Gly Leu Thr Pro Gln Gln Ile Cys Asp Lys Tyr His Cys
325      330      335
Ile His Ala Ser Ile Tyr Gln Trp Phe Gln Ile Asp Phe Asp Phe Phe
340      345      350
Gly Arg Thr Thr Thr Gln His Gln Thr Glu Ile Ala Gln Asp Ile Phe
355      360      365
Trp Arg Leu His Glu Arg Gly Phe Leu Leu Glu Asp Thr Val Glu Gln
370      375      380
Leu Arg Cys Glu Gly Cys Gln Arg Phe Leu Ala Asp Arg Phe Val Glu
385      390      395      400
Gly Glu Cys Pro His Cys Arg Tyr Pro Glu Ala Arg Gly Asp Gln Cys
405      410      415
Asp Lys Cys Gly Arg Leu Ile Asn Ala Val Glu Leu Lys Asn Pro Gln
420      425      430
Cys Lys Val Cys Lys Glu Thr Pro Val Ile Arg Ser Ser Lys His Leu

```

	435		440		445
Phe	Leu	Asn	Leu	Pro	Lys
450			455		460
Thr	Ser	Thr	Ala	Ala	Gly
465			470		475
Arg	Ser	Trp	Leu	Arg	Asp
			485		490
Leu	Lys	Trp	Gly	Thr	Pro
			500		505
Phe	Tyr	Val	Trp	Phe	Asp
			515		520
Asn	Tyr	Thr	Asp	Gln	Trp
530				535	
Glu	Leu	Tyr	Asn	Phe	Met
545				550	
Val	Phe	Pro	Cys	Ser	Leu
			565		570
Asn	Asn	Leu	Ile	Ala	Thr
			580		585
Ser	Lys	Ser	Arg	Gly	Val
			595		600
Gly	Ile	Pro	Ser	Asp	Val
610				615	
Glu	Gly	Gln	Asp	Ser	Ala
625				630	
Asn	Ser	Glu	Leu	Leu	Asn
			645		650
Met	Phe	Val	Ser	Lys	Phe
			660		665
Asn	Asp	Asp	Asp	Lys	Arg
			675		680
Gln	Tyr	Ile	Gln	Leu	Leu
690				695	
Cys	Ile	Leu	Asn	Met	Ser
705				710	
Glu	Pro	Trp	Lys	Lys	Ile
			725		730
Gly	Thr	Val	Thr	Gly	Val
			740		745
Val	Leu	Glu	Pro	Phe	Met
			755		760
Leu	Gln	Ala	Pro	Glu	Ser
770				775	
Ala	Phe	Ile	Cys	Thr	Leu
785				790	
Pro	Leu	Phe	Gln	Lys	Leu
			805		810
Arg	Phe	Gly	Gly	Leu	Gln
			820		825
Ser	Lys	Pro	Asn	Ala	Gly
			835		840
Ala	Asp	Pro	Glu	Arg	Ala
850				855	
Gly	Glu	Lys	Val	Arg	Ala
865				870	
Ile	Gly	Val	Glu	Val	Ala
			885		890
Leu	Ala	Glu	Gly	Lys	Thr

Lys

900

905

910

<210> 97

<211> 650

<212> PRT

<213> *Helicobacter pylori*

<400> 97

Met	Gln	Lys	Ser	Leu	Ile	Thr	Thr	Pro	Ile	Tyr	Tyr	Val	Asn	Asp	Ile
1				S					10					15	
Pro	His	Ile	Gly	His	Ala	Tyr	Thr	Thr	Leu	Ile	Ala	Asp	Thr	Leu	Lys
			20					25					30		
Lys	Tyr	Tyr	Thr	Leu	Gln	Gly	Glu	Glu	Val	Phe	Phe	Leu	Thr	Gly	Thr
		35					40					45			
Asp	Glu	His	Gly	Gln	Lys	Ile	Glu	Gln	Ser	Ala	Arg	Leu	Arg	Asn	Gln
	50					55					60				
Ser	Pro	Lys	Ala	Tyr	Ala	Asp	Ser	Ile	Ser	Thr	Ile	Phe	Lys	Asp	Gln
65					70					75				80	
Trp	Asp	Phe	Phe	Asn	Leu	Asp	Tyr	Asp	Gly	Phe	Ile	Arg	Thr	Thr	Asp
				85					90					95	
Ser	Glu	His	Gln	Lys	Cys	Val	Gln	Asn	Ala	Phe	Glu	Ile	Met	Phe	Glu
			100					105					110		
Lys	Gly	Asp	Ile	Tyr	Lys	Gly	Ala	Tyr	Ser	Gly	Tyr	Tyr	Cys	Val	Ser
		115					120					125			
Cys	Glu	Ser	Tyr	Cys	Ala	Ile	Ser	Lys	Ala	Asp	Asn	Thr	Asn	Asp	Lys
	130					135					140				
Val	Leu	Cys	Pro	Asp	Cys	Leu	Arg	Glu	Thr	Thr	Leu	Leu	Glu	Glu	Glu
145					150					155				160	
Ser	Tyr	Phe	Phe	Arg	Leu	Ser	Ala	Tyr	Glu	Lys	Pro	Leu	Leu	Asp	Phe
				165					170					175	
Tyr	Ala	Lys	Asn	Pro	Glu	Ala	Ile	Leu	Pro	Val	Tyr	Arg	Lys	Asn	Glu
			180					185					190		
Val	Thr	Ser	Phe	Ile	Glu	Gln	Gly	Leu	Leu	Asp	Leu	Ser	Ile	Thr	Arg
		195					200					205			
Thr	Ser	Phe	Glu	Trp	Gly	Ile	Pro	Leu	Pro	Lys	Lys	Met	Asn	Asp	Pro
	210				215						220				
Lys	His	Val	Val	Tyr	Val	Trp	Leu	Asp	Ala	Leu	Leu	Asn	Tyr	Ala	Ser
225					230						235			240	
Ala	Leu	Gly	Tyr	Leu	Asn	Asp	Leu	Asp	Asn	Lys	Met	Ala	His	Phe	Glu
				245					250					255	
Cys	Ala	Arg	His	Ile	Val	Gly	Lys	Asp	Ile	Leu	Arg	Phe	His	Ala	Ile
			260					265					270		
Tyr	Trp	Pro	Ala	Phe	Leu	Met	Ser	Leu	Asn	Leu	Pro	Leu	Phe	Lys	Gln
		275					280					285			
Leu	Cys	Val	His	Gly	Trp	Trp	Thr	Ile	Glu	Gly	Val	Lys	Met	Ser	Lys
	290					295					300				
Ser	Leu	Gly	Asn	Val	Leu	Asp	Ala	Gln	Lys	Ile	Ala	Met	Glu	Tyr	Gly
305					310					315				320	
Ile	Glu	Glu	Leu	Arg	Tyr	Phe	Leu	Leu	Arg	Glu	Val	Pro	Phe	Gly	Gln
				325					330					335	
Asp	Gly	Asp	Phe	Ser	Lys	Lys	Ala	Leu	Ile	Glu	Arg	Ile	Asn	Ala	Asn
			340					345					350		
Leu	Asn	Asn	Asp	Leu	Gly	Asn	Leu	Leu	Asn	Arg	Leu	Leu	Gly	Met	Ala
		355					360					365			
Lys	Lys	Tyr	Phe	Asn	His	Ser	Leu	Lys	Ser	Thr	Lys	Ile	Thr	Ala	Tyr

```

      370      375      380
Tyr Ser Lys Glu Leu Glu Lys Val His Gln Ile Leu Asp Asn Ala Asn
385      390      395      400
Ser Phe Val Pro Lys Met Gln Leu His Lys Ala Leu Glu Glu Leu Phe
      405      410      415
Asn Val Tyr Asp Phe Leu Asn Lys Leu Ile Ala Lys Glu Glu Pro Trp
      420      425      430
Val Leu His Lys Asn Asn Glu Ser Glu Lys Leu Glu Ala Leu Leu Ser
      435      440      445
Leu Ile Ala Asn Ala Leu Leu Gln Ser Ser Phe Leu Leu Tyr Ala Phe
      450      455      460
Met Pro Lys Ser Ala Val Lys Leu Ala Asn Ala Phe Asn Thr Glu Ile
      465      470      475      480
Thr Pro Asp Asn Tyr Glu Arg Phe Phe Lys Ala Lys Lys Leu Gln Asp
      485      490      495
Met Ile Leu Gln Asp Thr Glu Pro Leu Phe Ser Lys Met Glu Lys Ile
      500      505      510
Glu Lys Thr Glu Lys Ala Gly Glu Ala Ser Pro Glu Lys Asn Glu Lys
      515      520      525
Glu Lys Lys Asp Ala Lys Glu Lys Ala Pro Leu Lys Gln Glu Asn Tyr
      530      535      540
Ile Gly Ile Glu Asp Phe Lys Lys Val Glu Ile Lys Val Gly Leu Ile
      545      550      555      560
Lys Glu Ala Gln Arg Ile Glu Lys Ser Asn Lys Leu Leu Arg Leu Lys
      565      570      575
Val Asp Leu Gly Glu Gly Arg Leu Arg Gln Ile Ile Ser Gly Ile Ala
      580      585      590
Leu Asp Tyr Glu Pro Glu Ser Leu Val Gly Gln Met Val Cys Val Val
      595      600      605
Ala Asn Leu Lys Pro Ala Lys Leu Met Gly Glu Met Ser Glu Gly Met
      610      615      620
Ile Leu Ala Val Arg Asp Ser Asp Asn Leu Ala Leu Ile Ser Pro Thr
      625      630      635      640
Arg Glu Lys Ile Ala Gly Ser Leu Ile Ser
      645      650

```

<210> 98

<211> 582

<212> PRT

<213> *Drosophila melanogaster*

<400> 98

```

Met Leu Ile Arg Arg Ile Lys Cys Leu Arg Tyr Leu Gly Thr Arg Tyr
 1      5      10      15
Asn Ser Ser His Tyr Val Thr Thr Pro Ile Phe Tyr Val Asn Ala Ala
 20      25      30
Pro His Ile Gly His Leu Tyr Ser Ala Val Ile Ala Asp Ala His Cys
 35      40      45
Arg Tyr Gln Arg Leu Arg Tyr Pro Glu Gln Asp Val Arg Leu Cys Thr
 50      55      60
Gly Thr Asp Glu His Gly Thr Lys Ile Gln Gln Ala Ala Ser Leu His
 65      70      75      80
Gly Val Pro Val Ala Lys Tyr Cys Asp Asp Ile Ser Gln Arg Tyr Arg
 85      90      95
Glu Val Phe Arg Ser Ala Ser Ile Gln Gln Asp Asp Phe Ile Arg Thr
100      105      110
Thr Glu Asp Arg His Lys Arg Ala Val Ala Asn Phe Trp Arg Thr Leu

```

115	120	125
His Thr Arg Gly His Ile Tyr Ser Ala Ala Tyr Ser Gly Trp Tyr Cys		
130	135	140
Val Ser Asp Glu Thr Phe Leu Thr Asp Ser Gln Leu Arg Leu Asp Glu		
145	150	155
Ala Thr Gly Thr Arg Tyr Ser Leu Glu Ser Gly His Pro Val Glu Trp		
165	170	175
Thr Glu Glu Thr Asn Tyr Met Phe Arg Leu Ser Gln Phe Gln Asp Asp		
180	185	190
Val Arg His Trp Val Lys Thr Glu Ala Arg Val Arg Pro Ala Lys Phe		
195	200	205
Glu Lys Ile Leu Leu Asp Thr Leu Ser Glu Pro Leu Pro Asp Val Ser		
210	215	220
Val Ser Arg Pro Ser Asn Arg Val His Trp Ala Ile Pro Val Pro Asp		
225	230	235
Asp Asp Ser Gln Thr Val Tyr Val Trp Leu Asp Ala Leu Val Asn Tyr		
245	250	255
Leu Ser Ser Val Gly Tyr Pro Asp Glu Lys Phe Ser Ala His Trp Pro		
260	265	270
Pro Ala Gln Gln Val Ile Gly Lys Asp Ile Leu Lys Phe His Gly Ile		
275	280	285
Tyr Trp Thr Ala Phe Leu Leu Ala Ala Gly Leu Glu Pro Pro Gly Gln		
290	295	300
Leu Tyr Val His Ser His Trp Thr Val Asp Gly Gln Lys Met Ser Lys		
305	310	315
Ser Lys His Asn Val Val Asp Pro Leu Gln Ala Ala Gln Gln Tyr Thr		
325	330	335
Met Glu Gly Leu Arg Tyr Phe Leu Leu Arg Glu Gly Val Ala His Ser		
340	345	350
Asp Gly Asn Tyr Ser His Val Lys Ala Gln Arg Ile Leu Asn Ser Glu		
355	360	365
Leu Ala Asp Thr Leu Gly Asn Leu Leu Ser Arg Ala Ser Ala Lys Ser		
370	375	380
Leu Asn Pro Gly Gln Ile Tyr Pro Ser Pro Ser Ala Glu His Leu Ala		
385	390	395
Asp Leu Leu Arg Ser Leu Asp Val Ala Lys Arg Leu Gln Asp Ser Leu		
405	410	415
Leu Gln Leu Ser Glu Arg Cys Glu Ser His Tyr Glu Cys Asn His Phe		
420	425	430
His Leu Val Ala Asp Thr Thr Met Ala Ala Leu His Ala Ala Asn Asn		
435	440	445
Phe Phe Glu Ser Ser Lys Pro Trp Thr Leu Lys Ala Gly Ala Pro Asp		
450	455	460
Gly Asn Gln Ala Arg Leu Glu Thr Ile Ile Ala Met Thr Met Asp Ala		
465	470	475
Leu Arg Leu Ser Gly Ile Val Leu Gln Pro Ile Ile Pro Gln Leu Ala		
485	490	495
Asn Arg Leu Leu Asp Lys Leu Ser Val Pro Thr Ala Gln Arg Gly Trp		
500	505	510
Asn Tyr Leu Ala Glu Ser Phe Ala Thr Ser Pro Asn Ser Ser Asn Ser		
515	520	525
Pro Ala Gly Leu Gly Glu Ser Arg Gln Leu Asp Gly Gln Thr Ser Ala		
530	535	540
Leu Leu Phe Gln Arg Ile Leu Glu Glu Thr Ser Ala Lys Glu Val Lys		
545	550	555
Glu Gln Lys Pro Gln Pro Ala Lys Arg Ser Lys Ser Lys Lys Lys Glu		
565	570	575
Arg Arg Glu Thr Met Ser		

580

<210> 99

<211> 519

<212> PRT

<213> Mycobacterium tuberculosis

<400> 99

Met Lys Pro Tyr Tyr Val Thr Thr Ala Ile Ala Tyr Pro Asn Ala Ala
 1 5 10 15
 Pro His Val Gly His Ala Tyr Glu Tyr Ile Ala Thr Asp Ala Ile Ala
 20 25 30
 Arg Phe Lys Arg Leu Asp Arg Tyr Asp Val Arg Phe Leu Thr Gly Thr
 35 40 45
 Asp Glu His Gly Leu Lys Val Ala Gln Ala Ala Ala Ala Gly Val
 50 55 60
 Pro Thr Ala Ala Leu Ala Arg Arg Asn Ser Asp Val Phe Gln Arg Met
 65 70 75 80
 Gln Glu Ala Leu Asn Ile Ser Phe Asp Arg Phe Ile Arg Thr Thr Asp
 85 90 95
 Ala Asp His His Glu Ala Ser Lys Glu Leu Trp Arg Arg Met Ser Ala
 100 105 110
 Ala Gly Asp Ile Tyr Leu Asp Asn Tyr Ser Gly Trp Tyr Ser Val Arg
 115 120 125
 Asp Glu Arg Phe Phe Val Glu Ser Glu Thr Gln Leu Val Asp Gly Thr
 130 135 140
 Arg Leu Thr Val Glu Thr Gly Thr Pro Val Thr Trp Thr Glu Glu Gln
 145 150 155 160
 Thr Tyr Phe Phe Arg Leu Ser Ala Tyr Thr Asp Lys Leu Leu Ala His
 165 170 175
 Tyr His Ala Asn Pro Asp Phe Ile Ala Pro Glu Thr Arg Arg Asn Glu
 180 185 190
 Val Ile Ser Phe Val Ser Gly Gly Leu Asp Asp Leu Ser Ile Ser Arg
 195 200 205
 Thr Ser Phe Asp Trp Gly Val Gln Val Pro Glu His Pro Asp His Val
 210 215 220
 Met Tyr Val Trp Val Asp Ala Leu Thr Asn Tyr Leu Thr Gly Ala Gly
 225 230 235 240
 Phe Pro Asp Thr Asp Ser Glu Leu Phe Arg Arg Tyr Trp Pro Ala Asp
 245 250 255
 Leu His Met Ile Gly Lys Asp Ile Ile Arg Phe His Ala Val Tyr Trp
 260 265 270
 Pro Ala Phe Leu Met Ser Ala Gly Ile Glu Leu Pro Arg Arg Ile Phe
 275 280 285
 Ala His Gly Phe Leu His Asn Arg Gly Glu Lys Met Ser Lys Ser Val
 290 295 300
 Gly Asn Ile Val Asp Pro Val Ala Leu Ala Glu Ala Leu Gly Val Asp
 305 310 315 320
 Gln Val Arg Tyr Phe Leu Leu Arg Glu Val Pro Phe Gly Gln Asp Gly
 325 330 335
 Ser Tyr Ser Asp Glu Ala Ile Val Thr Arg Ile Asn Thr Asp Leu Ala
 340 345 350
 Asn Glu Leu Gly Asn Leu Ala Gln Arg Ser Leu Ser Met Val Ala Lys
 355 360 365
 Asn Leu Asp Gly Arg Val Pro Asn Pro Gly Glu Phe Ala Asp Ala Asp
 370 375 380
 Ala Ala Leu Leu Ala Thr Ala Asp Gly Leu Leu Glu Arg Val Arg Gly

385 390 395 400
 His Phe Asp Ala Gln Ala Met His Leu Ala Leu Glu Ala Ile Trp Leu
 405 410 415
 Met Leu Gly Asp Ala Asn Lys Tyr Phe Ser Val Gln Gln Pro Trp Val
 420 425 430
 Leu Arg Lys Ser Glu Ser Glu Ala Asp Gln Ala Arg Phe Arg Thr Thr
 435 440 445
 Leu Tyr Val Thr Cys Glu Val Val Arg Ile Ala Ala Leu Leu Ile Gln
 450 455 460
 Pro Val Met Pro Glu Ser Ala Gly Lys Ile Leu Asp Leu Leu Gly Gln
 465 470 475 480
 Ala Pro Asn Gln Arg Ser Phe Ala Ala Val Gly Val Arg Leu Thr Pro
 485 490 495
 Gly Thr Ala Leu Pro Pro Pro Thr Gly Val Phe Pro Arg Tyr Gln Pro
 500 505 510
 Pro Gln Pro Pro Glu Gly Lys
 515

<210> 100

<211> 575

<212> PRT

<213> *Saccharomyces cerevisiae*

<400> 100

```

Met Gln Cys Arg Ser Ile Val His Arg Leu Tyr Ser Lys Val Ser His
 1      5      10      15
Val Thr Thr Pro Ile Phe Tyr Pro Asn Ala Lys Pro His Leu Gly His
 20      25      30
Leu Tyr Ser Ser Leu Leu Ser Asp Val Tyr His Arg Trp Gln Leu Phe
 35      40      45
Lys Gly Asn Leu Ser Phe Phe Thr Thr Gly Thr Asp Glu His Gly Leu
 50      55      60
Lys Ile Gln Cys Ala Ser Glu Ser Asn Gly Phe Asp Gln Pro Lys Lys
 65      70      75      80
Phe Val Asp Lys Leu Tyr Pro Glu Phe Val Gln Leu Asp Lys Ile Tyr
 85      90      95
Gly Ile Asn Tyr Thr Arg Phe Ile Arg Thr Thr Asp Pro Asp His Ile
 100     105     110
Glu Asn Val Met Lys Leu Trp Glu Leu Cys Leu Lys Asn Gly Tyr Ile
 115     120     125
Tyr Met Gly Glu His Lys Gly Trp Tyr Ser Ile Ser Asp Glu Thr Phe
 130     135     140
Tyr Pro Glu Ser Lys Val Ile Lys Asp Pro Lys Asn Asp Gly Lys Tyr
 145     150     155     160
Leu Asn Thr Glu Ser Lys Asn Glu Val Val Tyr Gln Ser Glu Thr Asn
 165     170     175
Tyr Phe Phe Arg Leu Ser Leu Phe Asn Lys Lys Ile Val Asp His Ile
 180     185     190
Arg Lys Asn Pro Asp Phe Ile Phe Pro Ala Ser Lys Arg Asp Gln Ile
 195     200     205
Leu Lys Glu Leu Glu Thr Gly Gly Thr Leu Pro Asp Leu Ser Ile Ser
 210     215     220
Arg Pro Ser Ala Arg Leu Lys Trp Gly Ile Pro Thr Pro Asn Asp Pro
 225     230     235     240
Ser Gln Lys Val Tyr Val Trp Phe Asp Ala Leu Cys Asn Tyr Leu Ser
 245     250     255
Ser Ile Gly Gly Ile Pro Ser Ile Leu Ser Asn Ala Thr Glu Val Val

```

```

                260                265                270
Ser Arg His Tyr Ser Asp Lys Ser Asn Val Lys Gly Gln Leu Leu Ile
      275                280                285
Pro Tyr Pro Lys Glu Val Gln Arg Asn Thr Ile His Val Ile Gly Lys
      290                295                300
Asp Ile Ala Lys Phe His Thr Val Tyr Trp Pro Ser Phe Leu Leu Ala
305                310                315                320
Ala Gly Leu Pro Leu Pro Arg Gln Ile Val Val His Gly His Trp Leu
      325                330                335
Cys Asn Gly Met Lys Met Ser Lys Ser Leu Gly Asn Val Val Asp Pro
      340                345                350
Ile Asp Met Ala Arg Tyr Tyr Gly Ala Asp Ile Val Arg Trp Phe Leu
      355                360                365
Leu Glu Asn Ser Lys Leu Glu Asp Gly Asp Phe Gln Glu Ala Lys
      370                375                380
Leu Tyr Glu Thr Arg Glu Leu Leu Val Ser Lys Trp Gly Asn Leu Ile
385                390                395                400
Asn Arg Cys Cys Gly Ser Lys Phe Asn Ile Glu Arg Ala Val Met Lys
      405                410                415
Phe Ser Asp Lys Ala Asn Phe Gln Phe Gln Glu Ile Phe Gln Asn Glu
      420                425                430
Pro Ile Val Ser Glu Arg Ile Glu Asn Leu Ala Lys Leu Leu Asn Lys
      435                440                445
Ser Gln Glu Val Phe Asp Glu Lys Ile Ala Ile Phe Gln Tyr Pro Gln
450                455                460
Leu Leu Arg His Val Trp Ser Ile Ile Asn Asp Ala Asn Thr Leu Val
465                470                475                480
Gln Asn Ser Lys Pro Trp Glu Arg Glu Leu Asp Gln Gln Asp Asn Ile
      485                490                495
Ile Phe Leu Ala Met Glu Thr Ser Arg Ile Leu Ser Ile Leu Cys Gln
      500                505                510
Ser Ile Ile Pro Ser Leu Ser Gln Ser Phe Leu Asp Arg Ile Asp Val
      515                520                525
Ser Lys Glu Lys Arg Thr Ile Asn Tyr Ala Arg Leu Gly Ser Asp Lys
530                535                540
Thr Tyr Gly Lys Gln Ser Asn Lys Lys Gly Arg Glu Val Pro Leu Lys
545                550                555                560
Lys Ile Pro Phe Arg Leu Gln Glu Glu Gln Thr Asn Met Arg Ser
      565                570                575

```

<210> 101

<211> 917

<212> PRT

<213> *Caenorhabditis elegans*

<400> 101

```

Met Gly His Asp Leu Ala Asp Ile Lys Lys Ser Phe Glu Ala Ser Leu
 1                5                10                15
Pro Gly Tyr Val Glu Lys Lys Asp Pro Lys Ser Ile Leu Pro Gln Pro
      20                25                30
Gly Lys Arg Asn Ile Leu Ile Thr Ala Ala Leu Pro Tyr Val Asn Asn
      35                40                45
Val Pro His Leu Gly Asn Ile Ile Gly Cys Val Leu Ser Ala Asp Val
      50                55                60
Phe Ala Arg Tyr Cys Asn Leu Arg Gly His Gln Thr Phe Tyr Val Gly
65                70                75                80
Gly Thr Asp Glu Tyr Gly Thr Ala Thr Glu Thr Lys Ala Leu Gln Glu

```



```

545          550          555          560
Leu Lys Ala Gly His Lys Ile Gly Gln Pro Ser Pro Leu Phe Gln Lys
          565          570          575
Leu Asp Pro Ala Gln Ile Ala Glu Phe Lys Ala Lys Phe Gly Gly Ser
          580          585          590
Gln Asp Ala Gln Ser Ser Ala Pro Lys Thr Ala Glu Lys Pro Lys Gln
          595          600          605
Gln Lys Lys Gln Ala Pro Thr Lys Asp Lys Lys Gly Asp Lys Lys Met
          610          615          620
Ala Ser Thr Ala Ala Phe Val Glu Leu Glu Gln Gly Ala Lys Val Ile
625          630          635          640
Ser Gln Leu Ile Ala Gln Asn Leu Lys Lys Phe Asp Gln Ala Lys Ala
          645          650          655
Leu Phe Thr Arg Asn Gln Leu Gln Arg Leu Asp Gly Glu Asn Lys Gln
          660          665          670
Leu Thr Ile Asp Val Lys Thr Leu Gln His Gln Leu Ile Glu Leu Glu
          675          680          685
Thr Ala Ala Gly Ile Lys Gln Val Pro Lys Pro Val Val Ser Cys Thr
690          695          700
Pro Thr Pro Thr Ser Thr Pro Ala Ser Gly Ile Ile Thr Glu Ala Pro
705          710          715          720
Lys Lys Glu Ala Pro Ser Thr Pro Ala Pro Ser Glu Pro Lys Lys Ala
          725          730          735
Lys Glu Gln Lys Lys Gly Lys Gly Gly Ala Ala Ala Ala Pro Val Asp
          740          745          750
Asp Thr Ile Asp Val Gly Arg Leu Asp Met Arg Val Gly Arg Ile Ile
          755          760          765
Lys Cys Glu Lys His Pro Asp Ala Asp Ala Leu Tyr Val Glu Gln Ile
          770          775          780
Asp Val Gly Glu Ser Ala Pro Arg Thr Val Val Ser Gly Leu Val Arg
785          790          795          800
His Val Pro Leu Asp Gln Met Gln Asn Arg Leu Val Val Val Leu Cys
          805          810          815
Asn Leu Lys Pro Ala Lys Met Arg Gly Val Glu Ser Arg Ala Met Val
          820          825          830
Met Cys Ala Ser Ser Pro Asp Lys Val Glu Ile Met Glu Val Pro Ala
          835          840          845
Asp Ser Lys Pro Gly Thr Pro Val Val Cys Pro Pro Tyr Thr His Arg
          850          855          860
Pro Asp Glu Gln Leu Asn Pro Lys Lys Lys Ile Trp Glu Thr Val Ala
865          870          875          880
Glu Asp Leu Lys Val Ser Ala Glu Gly Phe Ala Glu Trp Lys Gly Gln
          885          890          895
Pro Leu Leu Ile Gly Ser Glu Ser Lys Met Thr Ala Pro Thr Leu Arg
          900          905          910
Gly Val His Val Lys
          915

```

<210> 102

<211> 545

<212> PRT

<213> Streptococcus pneumoniae

<400> 102

```

Met Ser Ile Phe Ile Gly Gly Ala Trp Pro Tyr Ala Asn Gly Ser Leu
1          5          10          15
His Ile Gly His Ala Ala Ala Leu Leu Pro Gly Asp Ile Leu Ala Arg

```

20 25 30
 Tyr Tyr Arg Gln Lys Gly Glu Glu Val Leu Tyr Val Ser Gly Ser Asp
 35 40 45
 Cys Asn Gly Thr Pro Ile Ser Ile Arg Ala Lys Lys Glu Asn Lys Ser
 50 55 60
 Val Lys Glu Ile Ala Asp Phe Tyr His Lys Glu Phe Lys Glu Thr Phe
 65 70 75 80
 Glu Lys Leu Gly Phe Thr Tyr Asp Leu Tyr Ser Arg Thr Asp Ser Pro
 85 90 95
 Leu His His Glu Ile Val Gln Glu Leu Phe Leu Gln Leu Tyr Glu Lys
 100 105 110
 Lys Phe Leu Tyr Thr Lys Lys Ile Lys Gln Leu Tyr Cys Thr Phe Asp
 115 120 125
 Asn Gln Phe Leu Pro Asp Arg Phe Val Glu Gly Lys Cys Pro Asn Cys
 130 135 140
 Gly Thr His Ser Arg Gly Asp Gln Cys Asp Asn Cys Ser Ala Ile Leu
 145 150 155 160
 Asp Pro Ile Asp Leu Val Asp Lys Arg Cys Ser Ile Cys Ser Asn Glu
 165 170 175
 Pro Glu Val Arg Glu Thr Glu His Phe Tyr Tyr Val Phe Ser Glu Phe
 180 185 190
 Gln Asn Leu Leu Glu Thr Tyr Leu Asn Asp Ala Glu Glu Thr Val Arg
 195 200 205
 Trp Arg Lys Asn Ala Ile Asn Leu Thr Lys Arg Tyr Leu Arg Glu Gly
 210 215 220
 Leu Pro Asp Arg Ala Val Thr Arg Asp Leu Pro Asn Gly Ile Pro Val
 225 230 235 240
 Pro Ile Asp Gly Phe Arg Asp Lys Lys Ile Tyr Val Trp Phe Glu Ala
 245 250 255
 Val Ala Gly Tyr Thr Ala Ser Val Asp Trp Ala Gln Lys Leu Gln
 260 265 270
 Asn Asn Ile Thr Asp Phe Trp Asn Asn Arg Thr Lys Ser Tyr Tyr Val
 275 280 285
 His Gly Lys Asp Asn Ile Pro Phe His Thr Ile Ile Trp Pro Ala Ile
 290 295 300
 Leu Ser Gly Leu Glu Ile Glu Pro Leu Pro Glu Tyr Ile Ile Ser Ser
 305 310 315 320
 Glu Tyr Leu Thr Leu Glu Asn Lys Lys Ile Ser Thr Ser Asn Asn Trp
 325 330 335
 Ala Ile Trp Leu Asn Asp Ile Ile Lys Lys Tyr Asp Ala Asp Ser Ile
 340 345 350
 Arg Tyr Phe Leu Thr Ile Asn Ala Pro Glu Met Lys Asp Ala Asn Phe
 355 360 365
 Ser Trp Arg Glu Phe Ile Tyr Ser His Asn Ser Glu Leu Leu Gly Ser
 370 375 380
 Tyr Gly Asn Phe Ile Asn Arg Thr Leu Lys Phe Ile Glu Lys Tyr Phe
 385 390 395 400
 Glu Ser Glu Ile Pro Thr Lys Tyr Leu Glu Gly Glu Ile Leu Tyr Asn
 405 410 415
 Leu Lys Glu Leu Tyr Thr Thr Val Gly Asn Leu Val Glu Ser Gly His
 420 425 430
 Met Lys Gln Ala Leu Glu Glu Ile Phe Glu Tyr Ile Arg Ser Ala Asn
 435 440 445
 Lys Phe Tyr Asp Asp Met Lys Pro Trp Ala Leu Arg Glu Ser Asp Ile
 450 455 460
 Glu Lys Cys Lys Glu Val Leu Ala Thr Cys Val Ile Ile Ile Leu Asn
 465 470 475 480
 Leu Gly Gln Met Leu Asn Pro Phe Ile Pro Phe Ser Gly Lys Lys Ile

485 490 495
 Glu Asp Met Phe Lys Thr Lys Leu Asn Thr Trp Asn Tyr Ile Ser Asn
 500 505 510
 Leu Pro Asn Lys Leu Ser Asp Val Ser Met Leu Phe Asp Arg Ile Asp
 515 520 525
 Leu Lys Lys Ile Asp Glu Glu Val Leu Glu Leu Gln Gln Thr Ser Ser
 530 535 540
 Arg
 545

<210> 103

<211> 677

<212> PRT

<213> *Pseudomonas aeruginosa*

<400> 103

Met	Ser	Glu	Pro	Arg	Lys	Ile	Leu	Val	Thr	Ser	Ala	Leu	Pro	Tyr	Ala	1	5	10	15
Asn	Gly	Ser	Ile	His	Leu	Gly	His	Met	Leu	Glu	Tyr	Ile	Gln	Thr	Asp	20	25	30	
Met	Trp	Val	Arg	Phe	Gln	Lys	Met	Arg	Gly	Asn	Gln	Ala	Val	Tyr	Val	35	40	45	
Cys	Ala	Asp	Asp	Ala	His	Gly	Ser	Ala	Ile	Met	Leu	Arg	Ala	Glu	Arg	50	55	60	
Glu	Gly	Ile	Thr	Ser	Glu	Gln	Leu	Ile	Asp	Ala	Val	Arg	Ala	Glu	His	65	70	75	80
Met	Gly	Asp	Phe	Ala	Asp	Phe	Leu	Val	Asp	Phe	Asp	Asn	Tyr	His	Ser	85	90	95	
Thr	His	Ser	Glu	Asn	Arg	Glu	Leu	Ser	Ser	Ala	Ile	Tyr	Leu	Lys		100	105	110	
Leu	Arg	Asp	Ala	Gly	His	Ile	Asp	Thr	Arg	Pro	Val	Thr	Gln	Tyr	Phe	115	120	125	
Asp	Pro	Glu	Lys	Gln	Met	Phe	Leu	Ala	Asp	Arg	Phe	Ile	Lys	Gly	Thr	130	135	140	
Cys	Pro	Lys	Cys	Gly	Thr	Ala	Asp	Gln	Tyr	Gly	Asp	Asn	Cys	Glu	Ala	145	150	155	160
Cys	Gly	Ala	Thr	Tyr	Ala	Pro	Thr	Glu	Leu	Lys	Asp	Pro	Lys	Ser	Ala	165	170	175	
Ile	Ser	Gly	Ala	Thr	Pro	Val	Leu	Lys	Glu	Ser	Leu	His	Tyr	Phe	Phe	180	185	190	
Lys	Leu	Pro	Asp	Phe	Glu	Ala	Met	Leu	Lys	Gln	Trp	Thr	Arg	Ser	Gly	195	200	205	
Ala	Leu	Gln	Glu	Ser	Val	Ala	Asn	Lys	Leu	Ala	Glu	Trp	Leu	Asp	Ser	210	215	220	
Gly	Leu	Gln	Gln	Trp	Asp	Ile	Ser	Arg	Asp	Ala	Pro	Tyr	Phe	Gly	Phe	225	230	235	240
Glu	Ile	Pro	Asp	Ala	Pro	Gly	Lys	Tyr	Phe	Tyr	Val	Trp	Leu	Asp	Ala	245	250	255	
Pro	Ile	Gly	Tyr	Met	Ala	Ser	Phe	Lys	Asn	Leu	Cys	Ala	Arg	Arg	Pro	260	265	270	
Glu	Leu	Asp	Phe	Asp	Ala	Phe	Trp	Gly	Lys	Asp	Ser	Gly	Ala	Glu	Leu	275	280	285	
Tyr	His	Phe	Ile	Gly	Lys	Asp	Ile	Val	Asn	Phe	His	Ala	Leu	Phe	Trp	290	295	300	
Pro	Ala	Met	Leu	Glu	Gly	Ala	Gly	Tyr	Arg	Lys	Pro	Thr	Ala	Leu	Asn	305	310	315	320
Val	His	Gly	Tyr	Leu	Thr	Val	Asn	Gly	Gln	Lys	Met	Ser	Lys	Ser	Arg				

```

          325          330          335
Gly Thr Phe Val Lys Ala Arg Thr Tyr Leu Asp His Leu Asp Pro Glu
          340          345          350
Tyr Leu Arg Tyr Tyr Tyr Ala Ser Lys Leu Gly Arg Gly Val Glu Asp
          355          360          365
Leu Asp Leu Asn Leu Glu Asp Phe Val Gln Lys Val Asn Ser Asp Leu
          370          375          380
Val Gly Lys Val Val Asn Ile Ala Ser Arg Cys Ala Gly Phe Ile His
          385          390          395          400
Lys Gly Asn Ala Gly Val Leu Val Gly Ala Asp Pro Ala Pro Glu Leu
          405          410          415
Leu Ala Ala Phe Arg Glu Ala Ala Pro Gly Ile Ala Glu Ala Tyr Glu
          420          425          430
Ala Arg Asp Phe Asn Arg Ala Met Arg Glu Ile Met Ala Leu Ala Asp
          435          440          445
Arg Ala Asn Ala Trp Ile Ala Glu Gln Ala Pro Trp Ala Leu Ala Lys
          450          455          460
Gln Glu Gly Gln Gln Asp Lys Val Gln Ala Val Cys Gly Leu Gly Ile
          465          470          475          480
Asn Leu Phe Arg Gln Leu Val Ile Phe Leu Lys Pro Val Leu Pro Lys
          485          490          495
Leu Ala Ala Ala Glu Ala Phe Leu Asn Val Ala Pro Leu Thr Trp
          500          505          510
Ala Asp His Gln Thr Leu Leu Ala Asn His Gln Leu Asn Pro Phe Gln
          515          520          525
Pro Leu Met Thr Arg Ile Glu Pro Ala Lys Val Glu Ala Met Ile Glu
          530          535          540
Ala Ser Lys Glu Asp Leu Ala Ala Ala Ser Gln Pro Ala Gly Asn Gly
          545          550          555          560
Glu Leu Val Lys Glu Pro Ile Ala Ala Glu Ile Asp Phe Asp Ala Phe
          565          570          575
Ala Ala Val Asp Leu Arg Ile Ala Leu Ile Glu Lys Cys Glu Phe Val
          580          585          590
Glu Gly Ala Asp Lys Leu Leu Arg Leu Ser Leu Asp Ile Gly Asp Ala
          595          600          605
Lys Arg Asn Val Phe Ser Gly Ile Lys Ser Ala Tyr Pro Asp Pro Ser
          610          615          620
Ala Leu Glu Gly Arg Leu Thr Leu Tyr Val Ala Asn Leu Ala Pro Arg
          625          630          635          640
Lys Met Lys Phe Gly Val Ser Glu Gly Met Val Leu Ala Ala Gly Pro
          645          650          655
Gly Gly Glu Glu Ile Tyr Leu Leu Ser Pro Asp Ser Gly Ala Lys Pro
          660          665          670
Gly Gln Arg Val Lys
          675

```

<210> 104

<211> 618

<212> PRT

<213> Thermus thermophilus HB8

<400> 104

```

Met Glu Lys Val Phe Tyr Val Thr Thr Pro Ile Tyr Tyr Val Asn Ala
 1           5           10           15
Glu Pro His Leu Gly His Ala Tyr Thr Thr Val Val Ala Asp Phe Leu
          20           25           30
Ala Arg Trp His Arg Leu Asp Gly Tyr Arg Thr Phe Phe Leu Thr Gly

```


REIVINDICAÇÕES

1. Uma composição compreendendo um primeiro vetor contendo uma polinucleótido codificando uma fenilalanil-tRNA sintetase (PheRS) modificada de levedura, em que O polinucleótido referido está mutado em códons codificando o aminoácido em posições de sequência selecionadas do grupo consistindo dos números de posição de sequência do aminoácido (i) 412 e 415; (ii) 415 e 418; (iii) 415 e 437; (iv) 412, 415 e 437; (v) 415, 418 e 437; (vi) 412, 415 e 418; e (vii) 412, 415, 418 e 437 do polipéptido codificado por SEQ ID No: 3 e em que a sintetase modificada referida é capaz de carregar uma molécula de tRNA com um aminoácido não-naturais.
2. A composição da reivindicação 1, compreendendo adicionalmente um segundo vetor contendo um polinucleótido codificando uma molécula de tRNA modificada.
3. A composição da reivindicação 2 em que o primeiro e segundo vetores referidos são o mesmo vetor.
4. A composição da reivindicação 2 em que o primeiro e segundo vetores referidos são vetores diferentes.
5. A composição de qualquer uma das reivindicações 2 a 4 em que o tRNA referido é modificado de tal modo que contém um anticodão mutado que emparelha com um codão degenerado *wobble* correspondente com uma afinidade maior do que a afinidade do tRNA natural.

6. Um polipéptido compreendendo uma fenilalanil-tRNA sintetase (PheRS) modificada de levedura, em que a sintetase referida é mutada nas posições da sequência de aminoácidos selecionadas do grupo consistindo dos números de posição da sequência de aminoácidos (i) 412 e 415; (ii) 415 e 418; (iii) 415 e 437; (iv) 412, 415 e 437; (v) 415, 418 e 437; (vi) 412, 415 e 418; e (vii) 412, 415, 418 e 437 do polipéptido codificado por SEQ ID No: 3 e em que a referida sintetase modificada é capaz de carregar uma molécula de tRNA com um aminoácido não-natural.
7. Um sistema de tradução compreendendo um polinucleótido codificando fenilalanil-tRNA sintetase (PheRS) modificada de levedura, em que a sintetase modificada referida é mutada em um ou mais códons codificando o aminoácido em posições de sequência selecionadas do grupo consistindo dos números de posição da sequência de aminoácidos (i) 412 e 415; (ii) 415 e 418; (iii) 415 e 437; (iv) 412, 415 e 437; (v) 415, 418 e 437; (vi) 412, 415 e 418; e (vii) 412, 415, 418 e 437 do polipéptido codificado por SEQ ID No: 3 e em que a sintetase modificada referida é capaz de carregar uma molécula de tRNA com um aminoácido não-natural.
8. O sistema de tradução da reivindicação 7 em que o referido sistema compreende uma célula hospedeira.
9. O sistema de tradução da reivindicação 7 ou reivindicação 8 compreendendo adicionalmente um polinucleótido codificando uma molécula de tRNA modificada.

10. O sistema de tradução de qualquer uma das reivindicações 7 a 9 em que o tRNA referido é modificado de tal modo que contém um anticodão mutado que emparelha com um codão degenerado *wobble* correspondente com uma afinidade maior do que a afinidade do tRNA natural.
11. O sistema de tradução da reivindicação 10 em que o tRNA modificado é tRNA^{Phe} munido com o anticodão AAA.
12. O sistema de tradução da reivindicação 8 em que a PheRS modificada referida e/ou a molécula de tRNA modificada referida é derivada de um organismo diferente da célula hospedeira.
13. O sistema de tradução da reivindicação 12 em que a molécula de tRNA modificada referida é derivada de uma célula eucariótica e a célula hospedeira é uma célula procariótica.
14. O sistema de tradução da reivindicação 8 em que a célula hospedeira é uma auxotrófica.
15. O sistema de tradução de qualquer uma das reivindicações 7 a 14 compreendendo adicionalmente um meio de cultura contendo um ou mais aminoácidos não-naturais.
16. Um método para incorporar pelo menos um aminoácido não-natural num polipéptido alvo em uma ou mais localização(ões) especificada(s), o método compreendendo fornecer um sistema de tradução contendo pelo menos um aminoácido não-natural; fornecer ao

sistema de tradução uma ou mais fenilalanil-tRNA sintetase(s) (PheRS) modificada(s) de levedura em que a sintetase modificada referida é mutada em posições de sequência de aminoácidos selecionadas do grupo consistindo dos números de posição da sequência de aminoácidos (i) 412 e 415; (ii) 415 e 418; (iii) 415 e 437; (iv) 412, 415 e 437; (v) 415, 418 e 437; (vi) 412, 415 e 418; e (vii) 412, 415, 418 e 437 do polipéptido codificado por SEQ ID No: 3 e em que a sintetase modificada referida é capaz de carregar uma molécula de tRNA com um aminoácido não-natural; fornecer ao sistema de tradução um polinucleótido codificando um polipéptido alvo de interesse; e permitir tradução do polipéptido alvo de interesse, assim incorporando pelo menos um aminoácido não-natural no polipéptido alvo.

17. Um método de acordo com a reivindicação 16 em que a molécula de tRNA referida é modificada de modo a conter um anticodão mutante que emparelha com um correspondente codão degenerado *wobble* com afinidade maior do que a afinidade do tRNA natural.
18. Um método de acordo com a reivindicação 16 ou 17 em que o sistema de tradução compreende uma célula hospedeira, a molécula de tRNA é modificada e em que a PheRS modificada referida e/ou molécula de tRNA modificada referida é derivada de um organismo diferente da célula hospedeira.
19. Uma composição, polipéptido, sistema de tradução ou método de acordo com qualquer uma das reivindicações 1 a 18 em que a sintetase modificada é mutada de tal

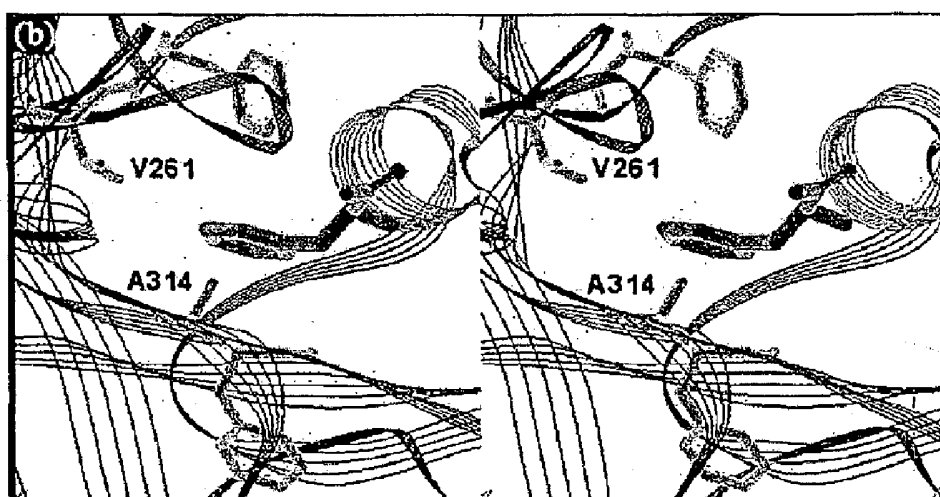
modo que os codões codificando os aminoácidos nas posições 412, 415 e 437 são mutados.

20. Uma composição, polipéptido, sistema de tradução ou método de acordo com qualquer uma das reivindicações 1 a 18 em que a sintetase modificada é mutada de tal modo que os codões codificando os aminoácidos nas posições 415, 418 e 437 são mutados.
21. Uma composição, polipéptido, sistema de tradução ou método de acordo com qualquer uma das reivindicações 1 a 18 em que a sintetase modificada é mutada de tal modo que os codões codificando os aminoácidos nas posições 412, 415, 418 e 437 são mutados.
22. Uma composição, polipéptido, sistema de tradução ou método de acordo com qualquer uma das reivindicações 1, 6, 7 ou 16 em que a sintetase modificada referida utiliza um codão stop âmbar para incorporação de um aminoácido não-natural numa localização particular no polipéptido.

Lisboa, 18 de Setembro de 2014

FIGURAS

	261	314
T. thermophilus:	RFQPVYFPFVEP...	GFAFGLGVERLAMLR
E. coli:	RFRPSTFPFTEP...	GFAFGMGMERLTMLR
S. cerevisiae:	RFKPTYNPYTEP...	VLGMGLSLERPTMIKY
	415	460

FIG. 1A*FIG. 1B*

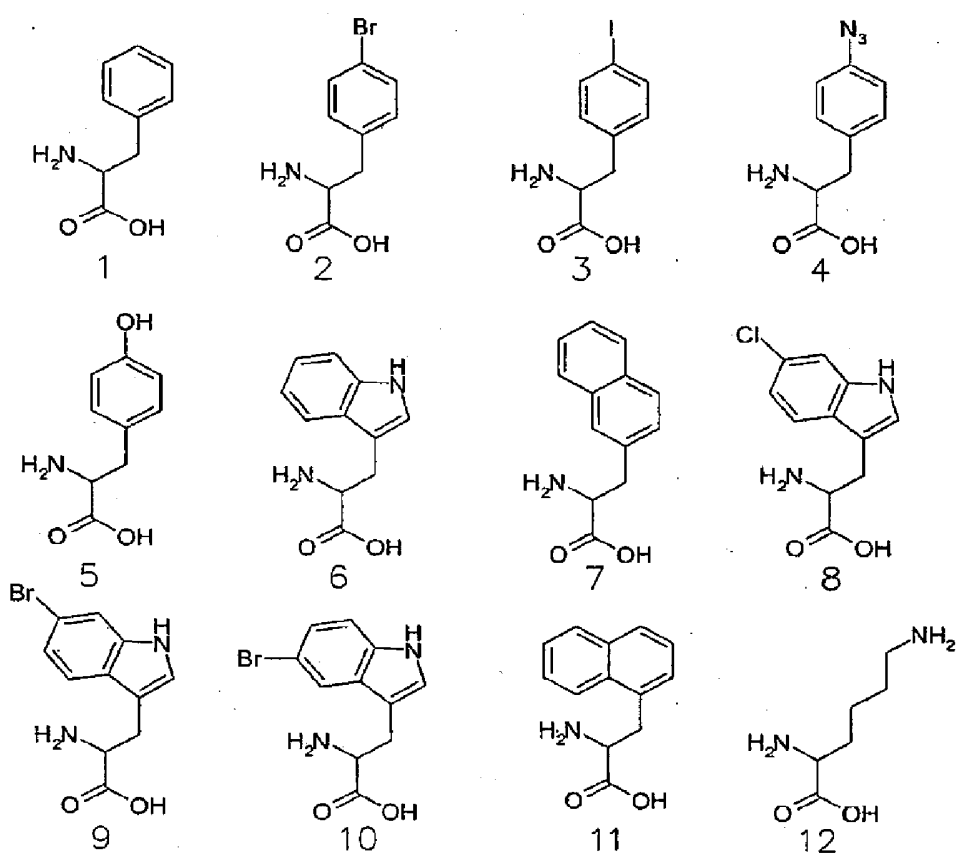


FIG. 2

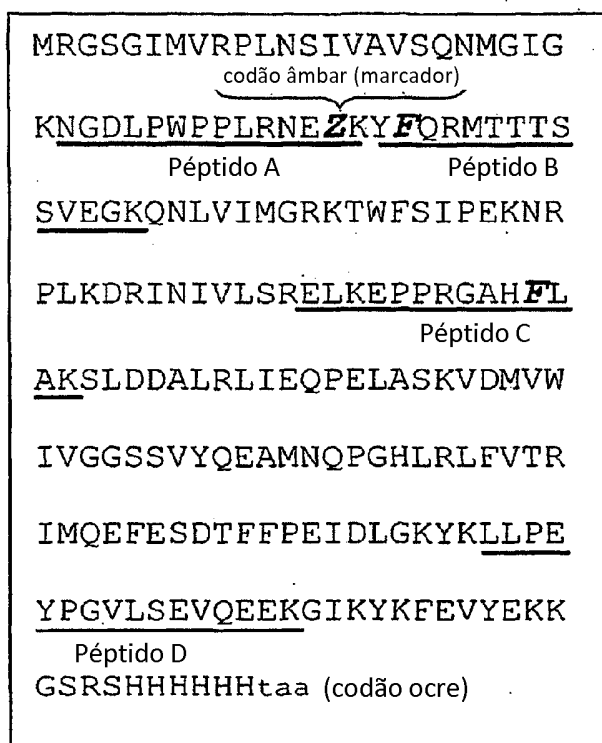
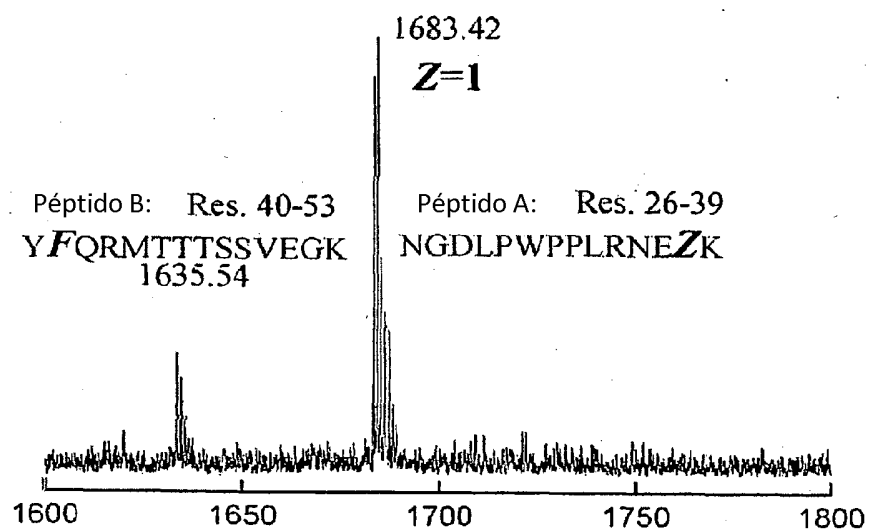
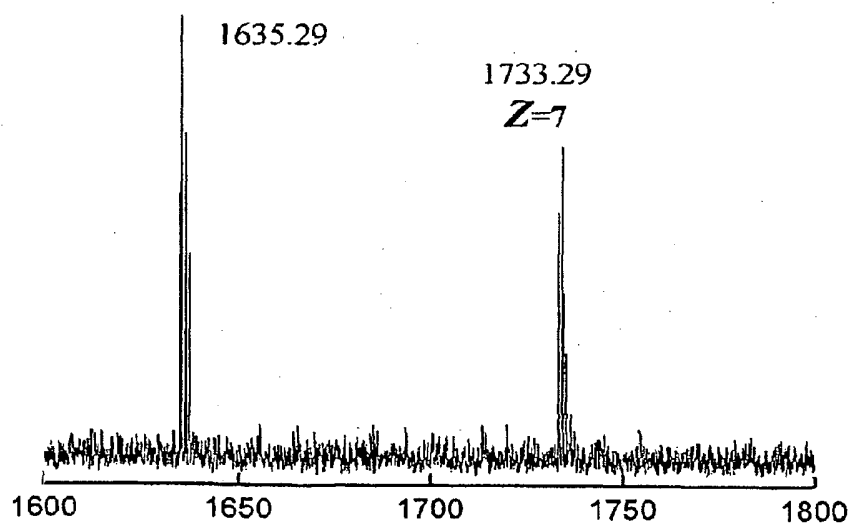
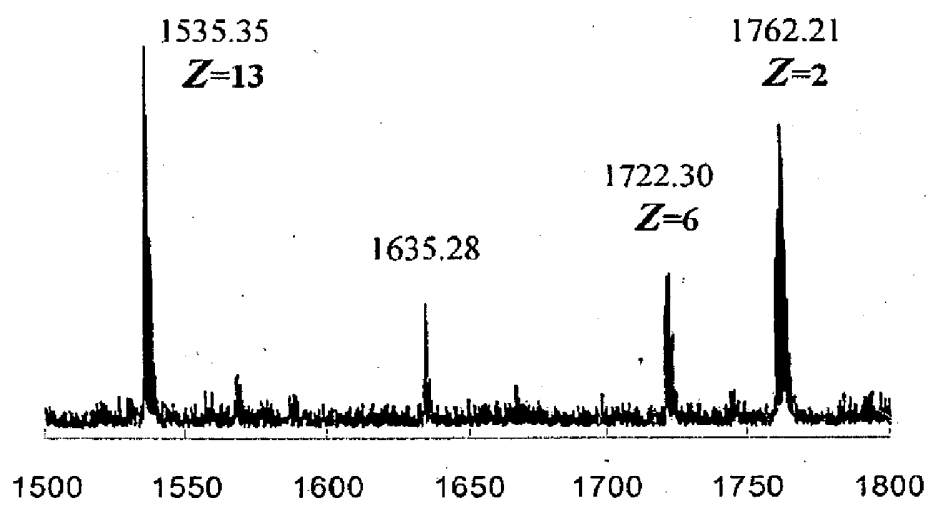
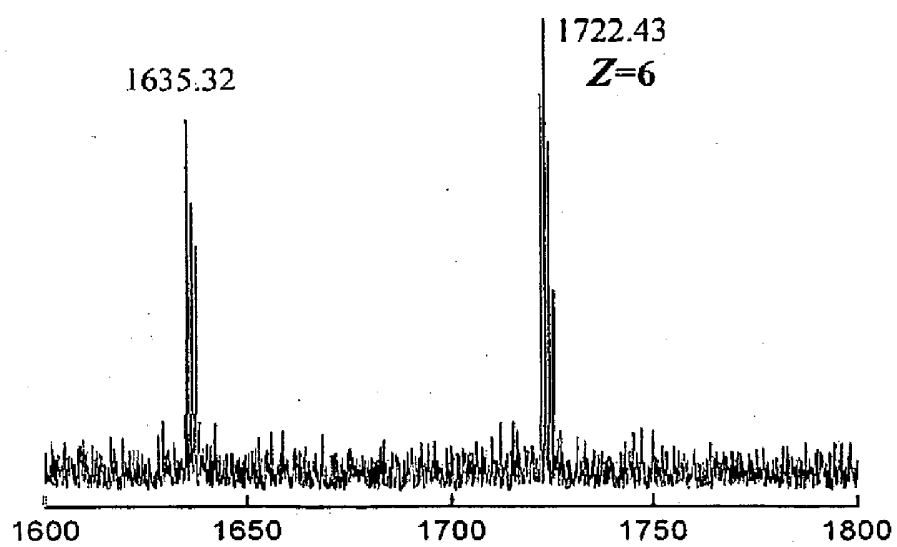


FIG. 3

*FIG. 4A**FIG. 4B*

*FIG. 4C**FIG. 4D*

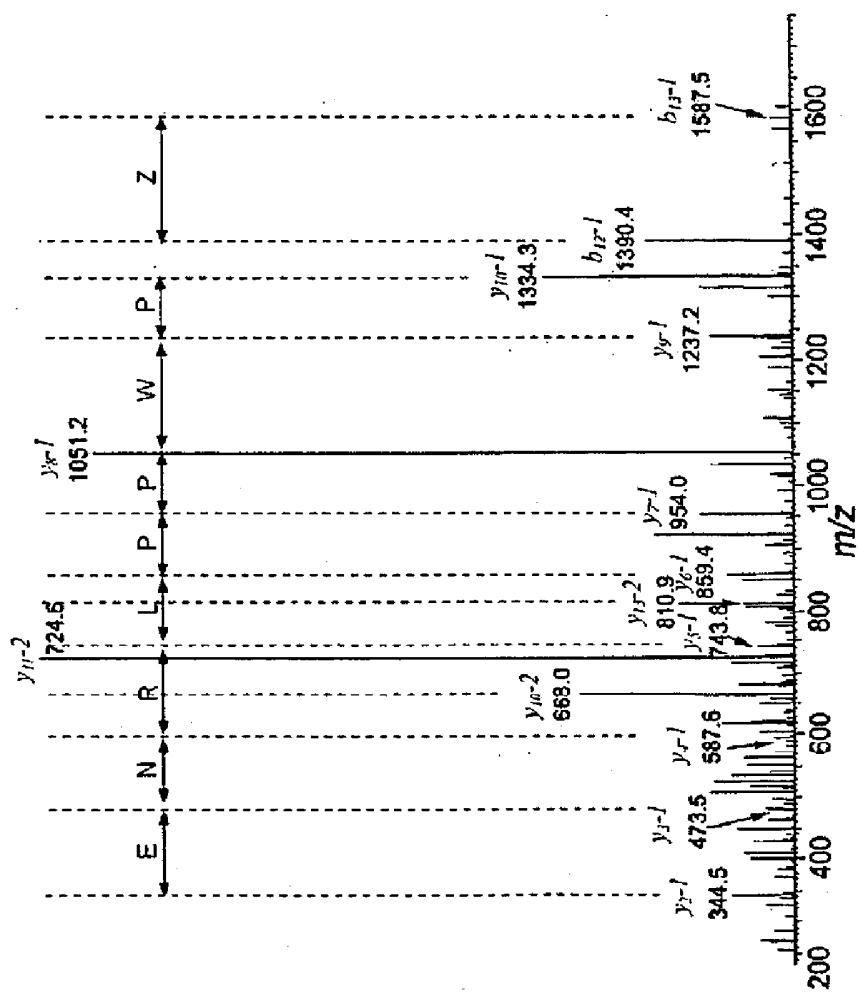
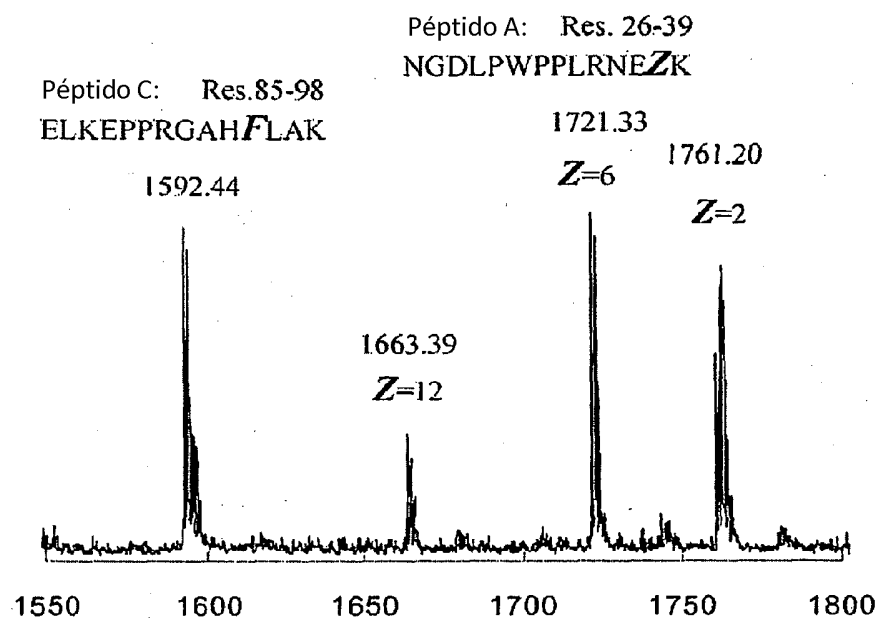
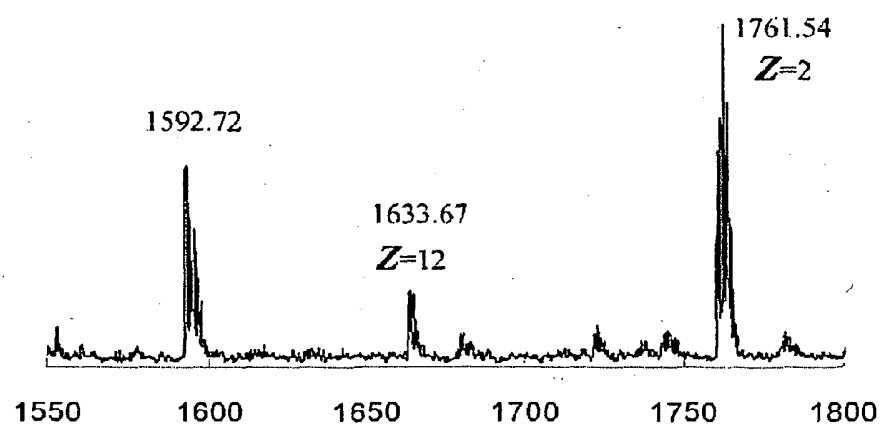
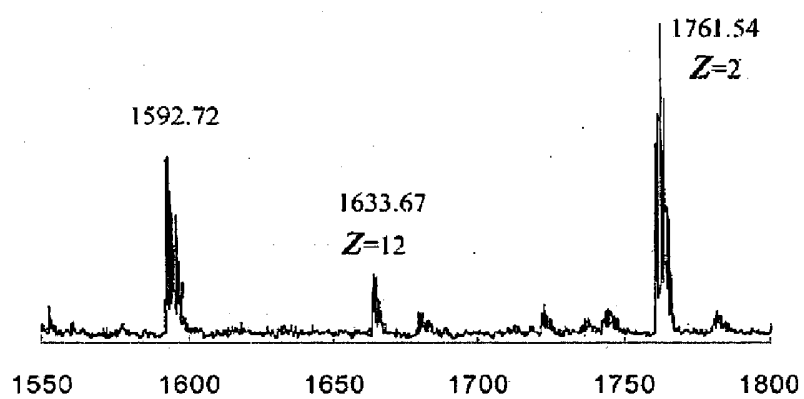
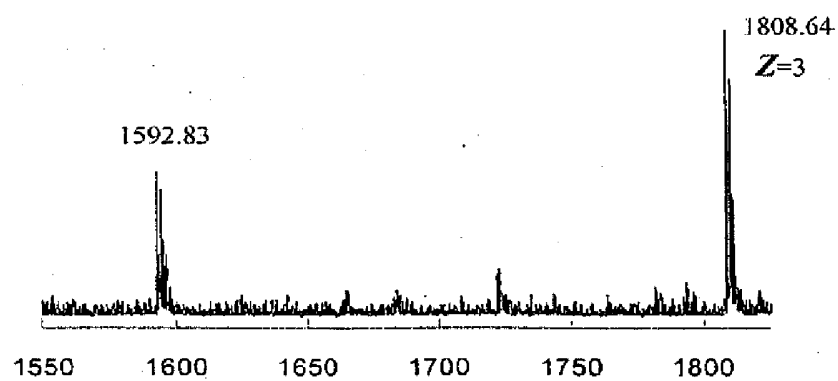


FIG. 5

*FIG. 6A**FIG. 6B*

*FIG. 6C**FIG. 6D*

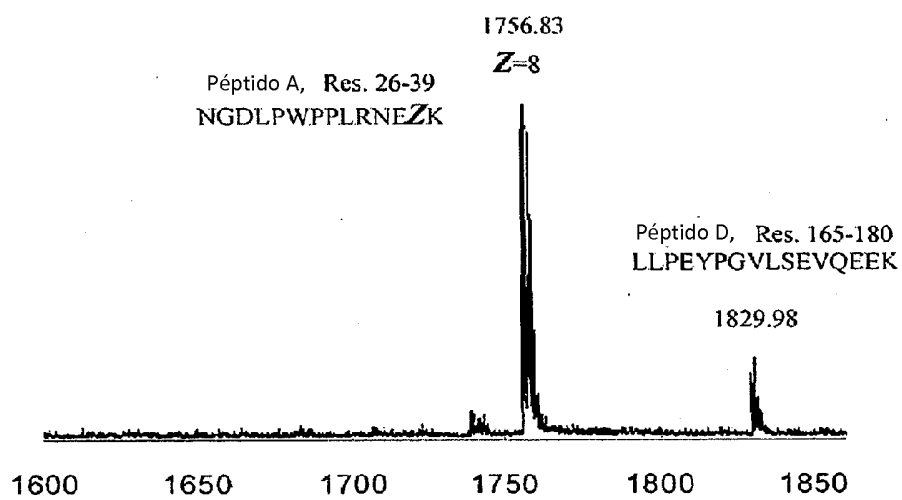


FIG. 7A

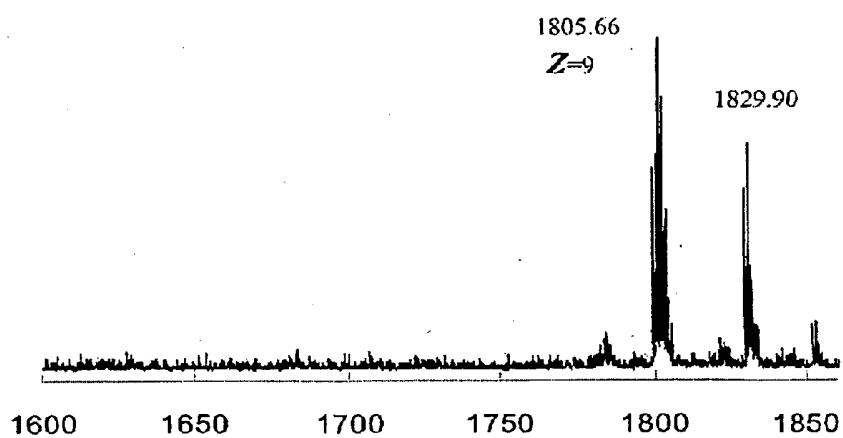


FIG. 7B

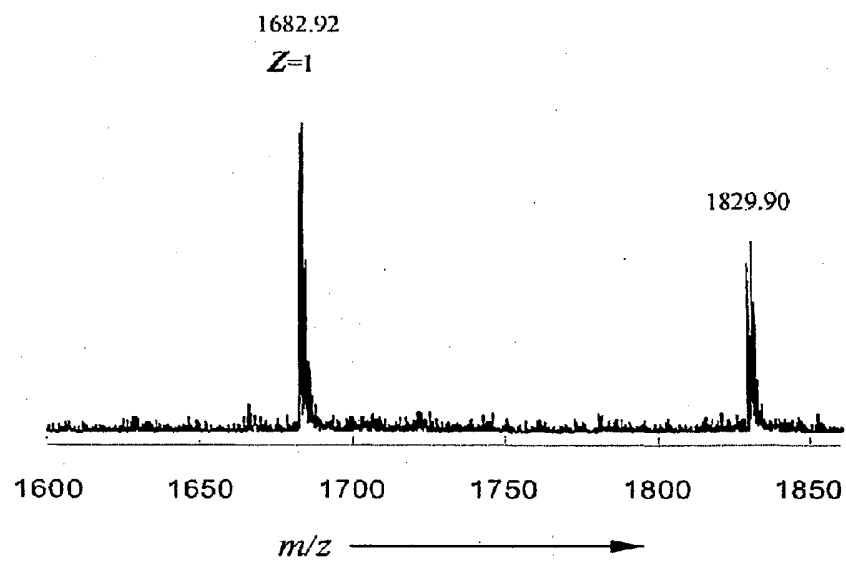


FIG. 7C

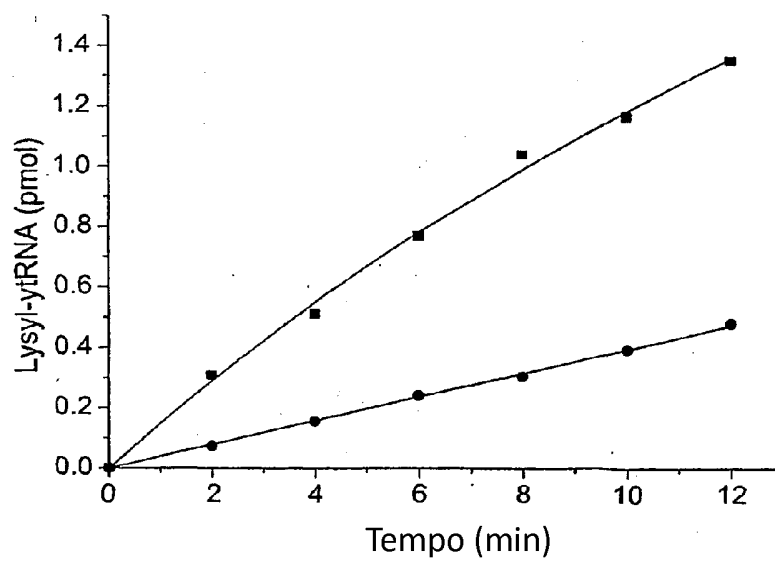
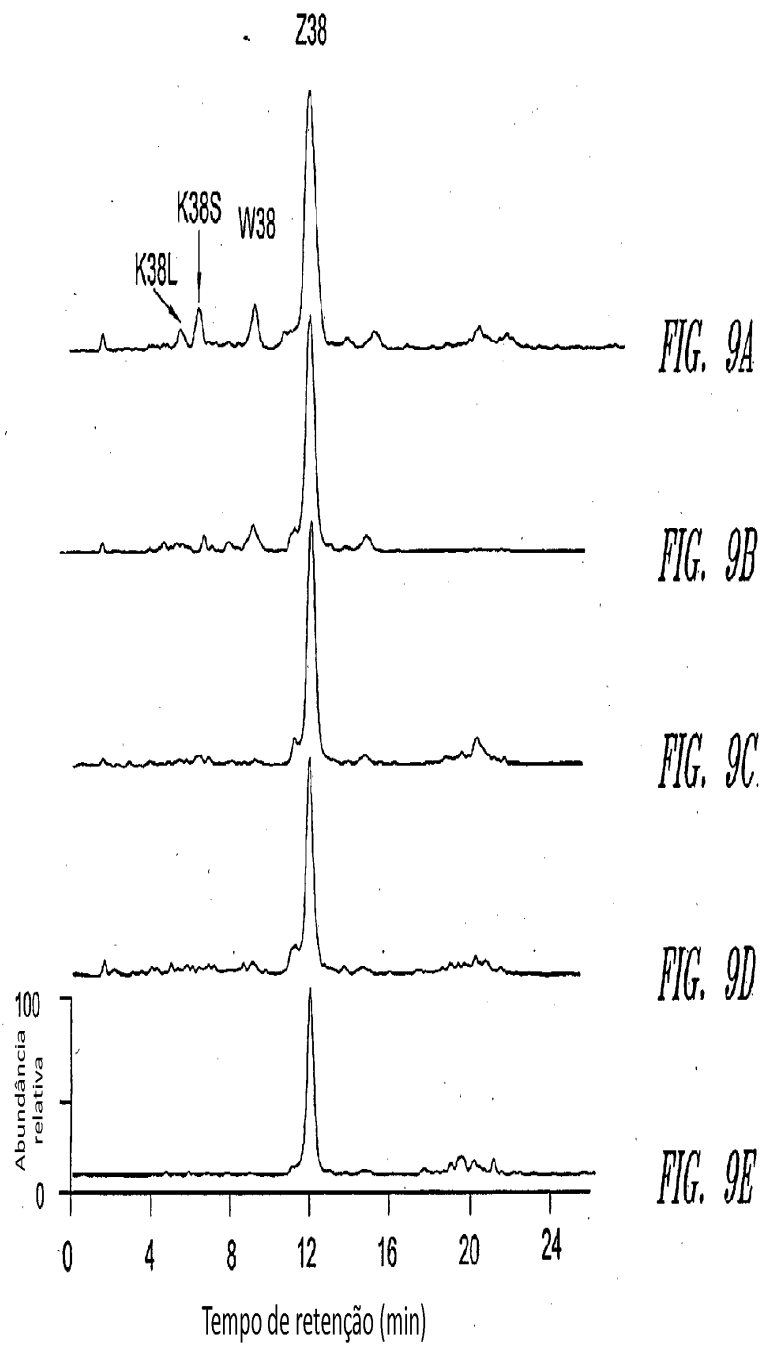
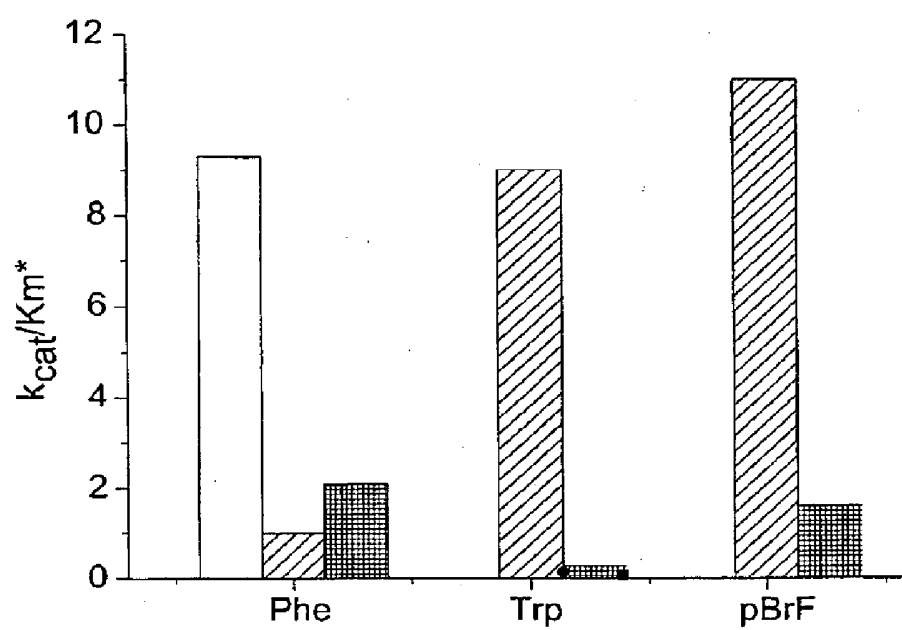
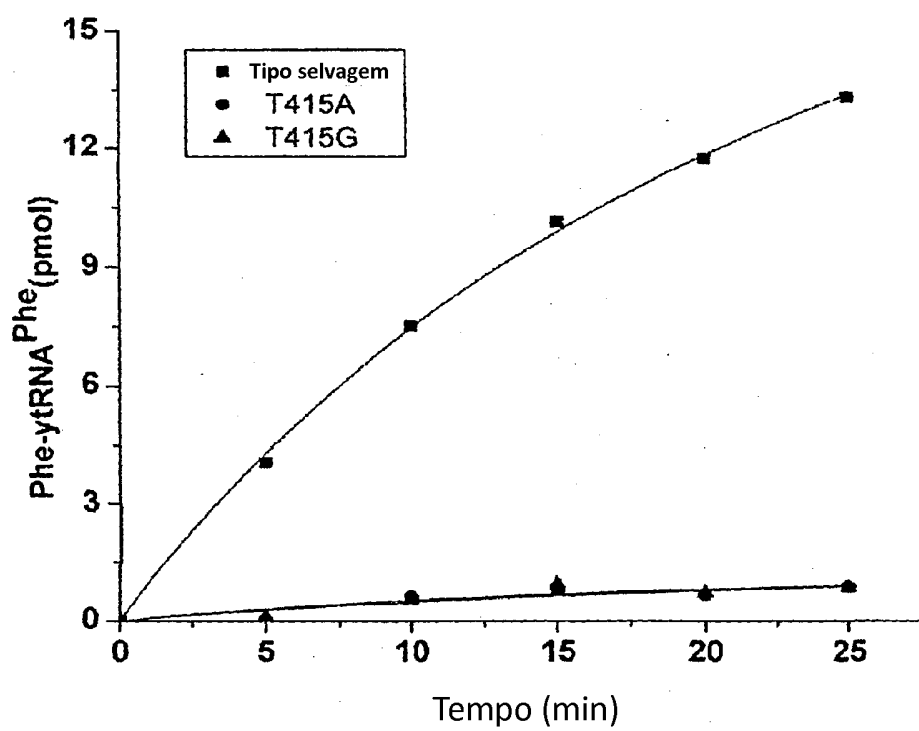
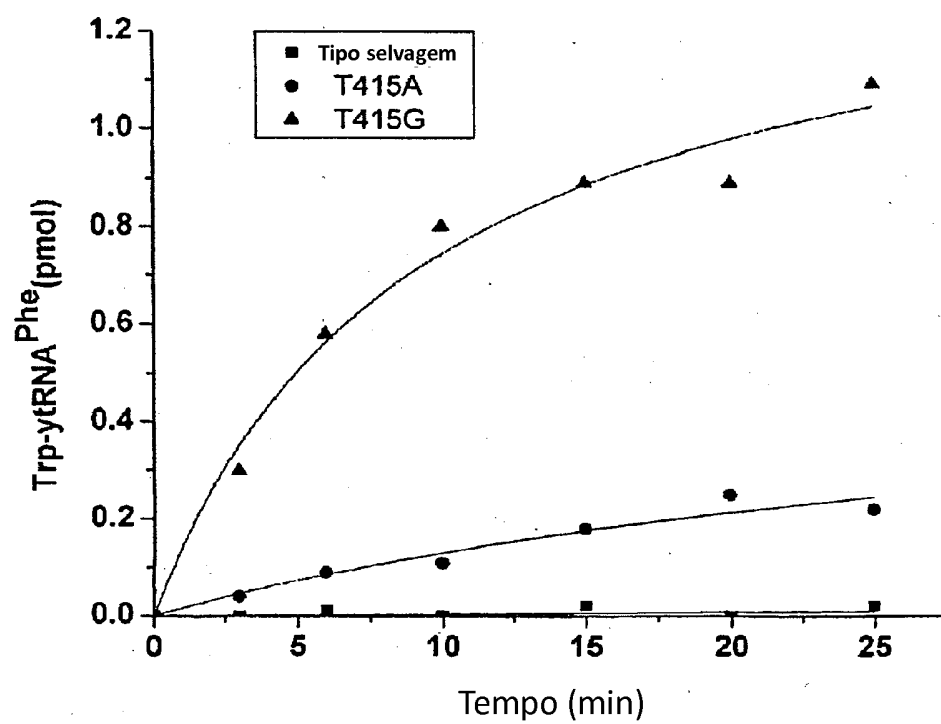


FIG. 8



*FIG. 10*

*FIG. 11A*

*FIG. 11B*

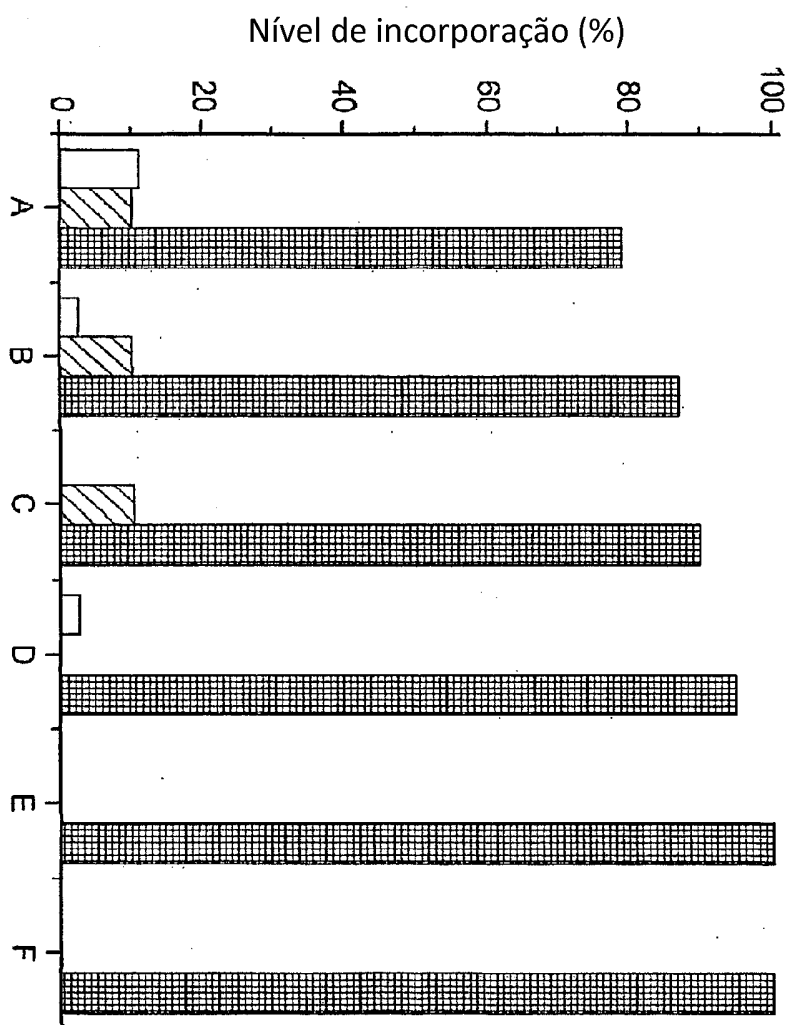


FIG. 12

Usando variante yPheRS (T415G) e ytRNA^{Phe} supressor âmbar, pEtF pode ser inserido em mDHFR em resposta a um codão âmbar no sistema de *E. coli*.

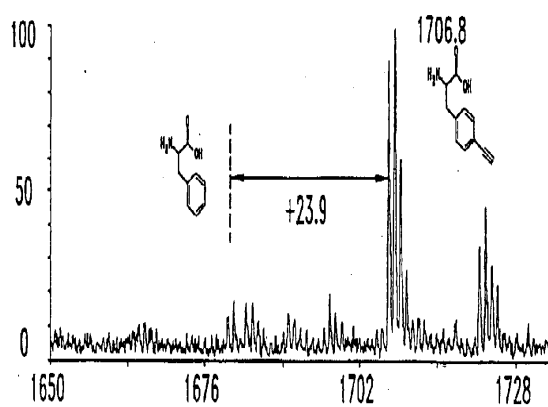
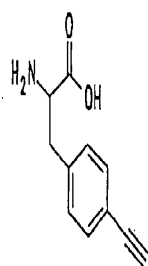
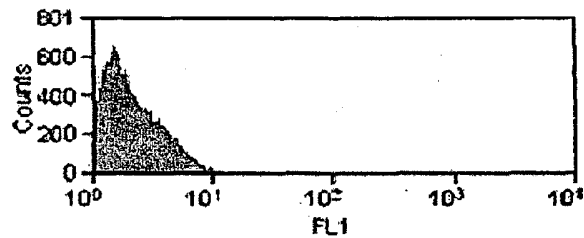


FIG. 13

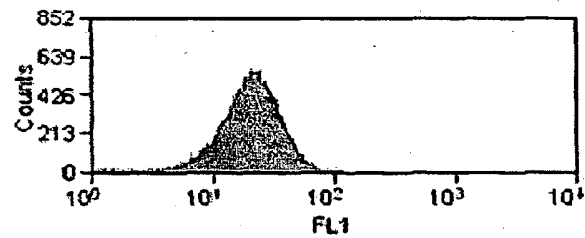
GFP supressão âmbar_081406

GFP_158 (L64, Am158)

Antes da
Indução



0.01 mM Phe
0.01 mM Trp



0.05 mM Phe
0.05 mM Trp
3.0 mM 2NaI

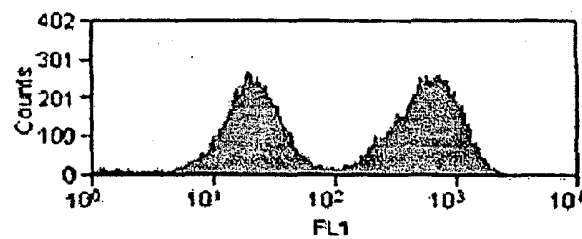
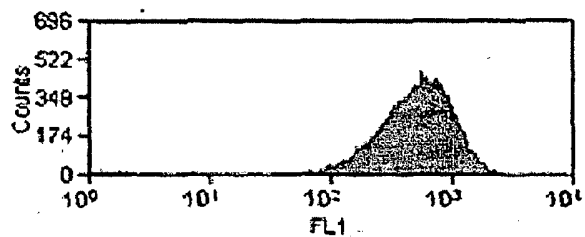
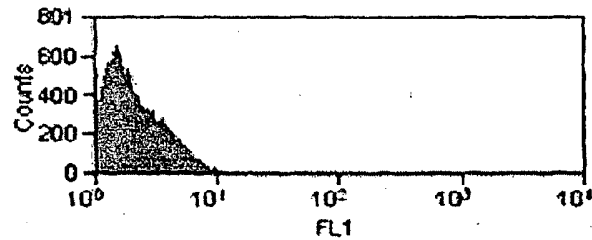


FIG. 14A

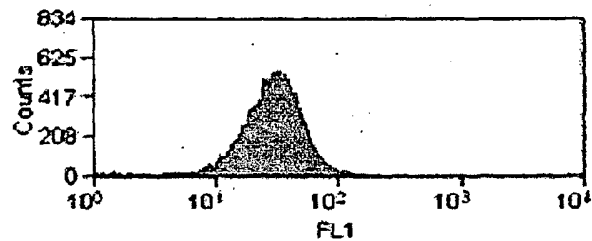
GFP supressão âmbar_081406

GFP_158 (L64, Am158)

Antes da
Indução



0.05 mM Phe
0.05 mM Trp



0.05 mM Phe
0.05 mM Trp
3.0 mM 2NaI

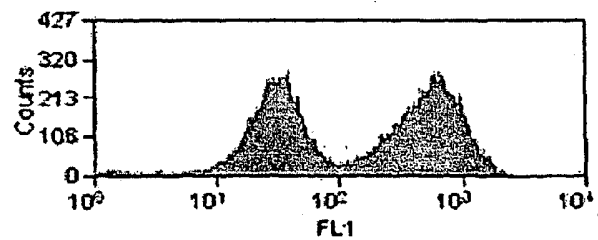
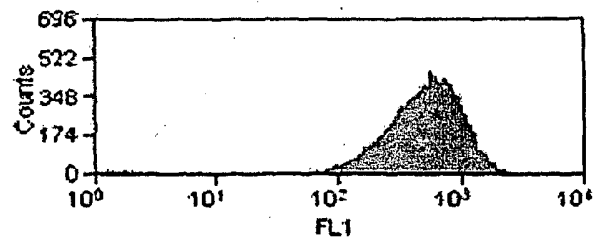


FIG. 14B

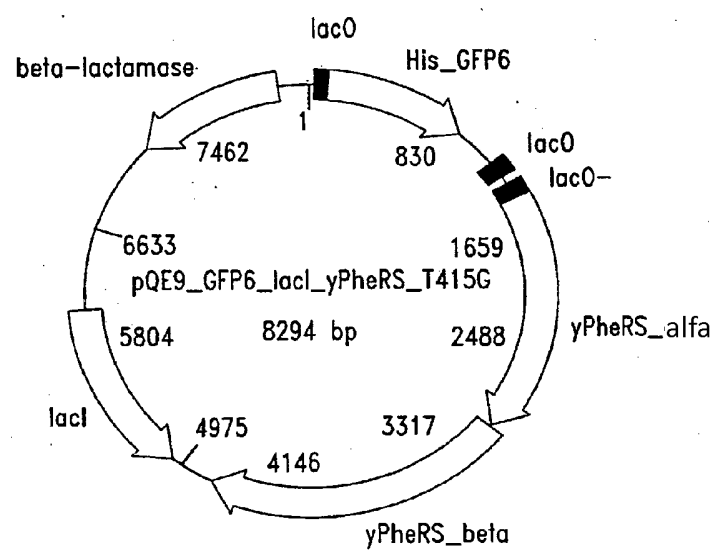


FIG. 15

Locais de mutação alfa de PheRS de levedura estão marcados com NNN

atgtctgact tccaattaga aattctaaag aaactagatg aattggatga gatcaagtc
 acactggcaa ctttcctca gcacggctct caagatgttc ttccgcttt gaactctttg
 aaagcccaca acaagttaga gttttccaag gtcgacacgg ttacgtatga cttgaccaa
 gaagtgctc aaattttgaa tgaagggtcg tacgaaatta aactagtcaa gtcacccaa
 gagggtggc aacttcaat caaagatgtg atgtccaaac taggcctca agttggaag
 gtcggtcagg ctgagcttt caagaacggc tggatcgcca aaaacgcctc aaacgagctt
 gaactctcg caaaattgca aaataccgat ttaaatgagc ttactgatga aacgcaatct
 attctagcg aaatcaagaa caactcgat ctggatagca ttgacgcaa gatittgaac
 gacttgaaga aaagaaagt aattgtcaa ggtaaaatca cagatttcag tgcaccaa
 gggccagagt tctcaccga ctcaccaa ttggaaccg atcttacctc cgaatggc
 tccaccaatg catacaagga ctgaagttc aagccttaca atttcaattc tcaaggtgtg
 caaatatctt caggtgctt tcaccctta aacaaagta gagaggaatt tagacaaatt
 ttcttttcca tgggattcac agagatgcc tcgaaccaat acgtcgagac aggtttctg
 aacttcgatg cctttacgt cccacaacag catctgtct gtacctgca agacatttc
 tacatcaagg accactaac cgtgagttg cccgatgaca agacatacat ggaatatac
 aaagccgtc acgaacagg gagattcgg tccatcggtt atcgttaca ctggaagcca
 gaagaatgtc aaaaattggt ctgagaact cactccacag ccactctgc cagaatgctg
 cagatttgg ccaagatcc aaagccacc agattgttt ctatcgaccg tgtttccgt
 aacgaagcag ttgacgcac ccatttggc gaattccacc aggtggaagg tgttctgcc
 gactacaaca ttactctgg tgacctgatc aagttcatgg aagagtttt cgaagaatg
 ggtgtcaccg gtttgagatt caagcctacc tacNNNcctt acNNNgagcc aNNNatggaa
 atcttttctt ggcacgaagg ttgcaaaaa tgggtcgaaa tggtaacNNNggtatgttc
 agaccagaaa tgctcgagtc catgggtcta ccaaaggatc taagagtcct tggttgggg
 ttatccttgg aaagacctac catgatcaaa tataaggttc aaaacatcag agaactgtta
 ggtcataaag tctctttgga cttatcgaa accaatctg ctgctagatt ggacgaagac
 ttgtacgaat aa

	Nome	412	415	418	437
	T415G	AAT	GGC	TCA	TCT
1	2Na1	GGG	GGG	TGT	TTT
2	412_415	GGG	GGC		
3	415_418		GGC	TGT	
4	415_437		GGC		TTT
5	412_415_437	GGG	GGC		TTT
6	415_418_437		GGC	TGT	TTT
7	412_415_418	GGG	GGC	TGT	

FIG. 16