

(22) Data de pedido: **2004.02.10**

(30) Prioridade(s): **2003.02.11 US 447747 P**

(43) Data de publicação do pedido: **2005.11.09**

(45) Data e BPI da concessão: **2011.11.23**
014/2012

(73) Titular(es):

SHIRE HUMAN GENETIC THERAPIES, INC.

700 MAIN STREET CAMBRIDGE, MA 02139 US

(72) Inventor(es):

MICHAEL W. HEARTLEIN

KURT VON FIGURA

BERNHARD SCHMIDT

THOMAS DIERKS

ANDREA BALLABIO

US

DE

DE

DE

IT

(74) Mandatário:

MARIA SILVINA VIEIRA PEREIRA FERREIRA

RUA CASTILHO, N.º 50, 5º - ANDAR 1269-163 LISBOA

PT

(54) Epígrafe: **DIAGNÓSTICO E TRATAMENTO DA DEFICIÊNCIA DE MÚLTIPLAS SULFATASES E OUTRAS UTILIZANDO UMA ENZIMA GERADORA DE FORMILGLICINA (FGE)**

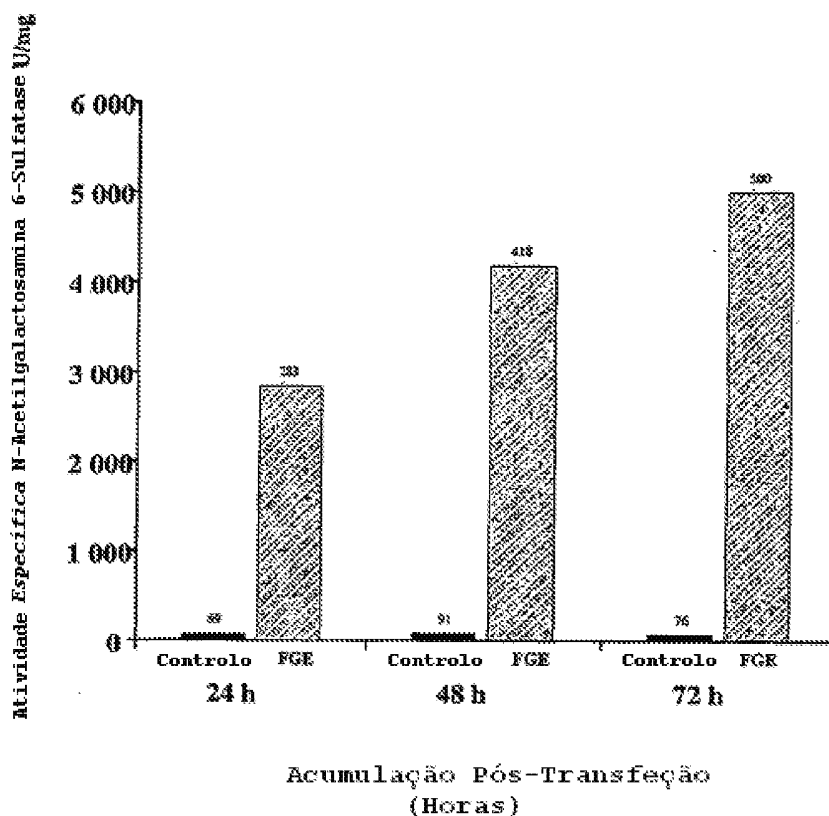
(57) Resumo:

ESTA INVENÇÃO REFERE-SE A MÉTODOS E COMPOSIÇÕES PARA O DIAGNÓSTICO E TRATAMENTO DA DEFICIÊNCIA DE MÚLTIPLAS SULFATASES (MSD), BEM COMO DE OUTRAS DEFICIÊNCIAS DE SULFATASES. MAIS ESPECIFICAMENTE, A INVENÇÃO REFERE-SE A MOLÉCULAS ISOLADAS QUE MODULAM MODIFICAÇÕES PÓS-TRADUÇÃO EM SULFATASES. ESSAS MODIFICAÇÕES SÃO ESSENCIAIS PARA A FUNÇÃO APROPRIADA DE SULFATASE.

RESUMO

"DIAGNÓSTICO E TRATAMENTO DA DEFICIÊNCIA DE MÚLTIPLAS SULFATASES E OUTRAS UTILIZANDO UMA ENZIMA GERADORA DE FORMILGLICINA (FGE) "

Esta invenção refere-se a métodos e composições para o diagnóstico e tratamento da Deficiência de Múltiplas Sulfatases (MSD), bem como de outras deficiências de sulfatases. Mais especificamente, a invenção refere-se a moléculas isoladas que modulam modificações pós-tradução em sulfatases. Essas modificações são essenciais para a função apropriada de sulfatase.



DESCRIÇÃO

"DIAGNÓSTICO E TRATAMENTO DA DEFICIÊNCIA DE MÚLTIPLAS SULFATASES E OUTRAS UTILIZANDO UMA ENZIMA GERADORA DE FORMILGLICINA (FGE) "

Esta invenção refere-se, entre outras coisas, a métodos e composições para o diagnóstico e tratamento da Deficiência de Múltiplas Sulfatases (MSD), bem como de outras deficiências de sulfatases. Mais especificamente, a invenção utiliza moléculas que modulam modificações pós-tradução em sulfatases. Essas modificações são essenciais para a função apropriada de sulfatase.

As sulfatases são membros de uma família genética altamente conservada, que partilham extensa homologia de sequências (Franco, B., et al., *Cell*, 1995, 81: 15-25; Parenti, G., et al., *Curr. Opin. Gen. Dev.*, 1997, 7: 386-391), um grau elevado de semelhança estrutural (Bond, C.S., et al., *Structure*, 1997, 5: 277-289; Lukatela, G., et al., *Biochemistry*, 1998, 37: 3654-64) e uma modificação pós-tradução única que é essencial para clivagem de éster de sulfato (Schmidt, B., et al., *Cell*, 1995, 82: 271-278; Selmer, T., et al., *Eur. J. Biochem.*, 1996, 238: 341-345). A modificação pós-tradução envolve a oxidação de um resíduo cisteína (em eucariotas) ou serina (em certos procariotas) conservado, em Cp, dando origem a L-C_α-formilglicina (também conhecida como FGly; ácido 2-amino-3-oxopropanoico) na qual um grupo aldeído substitui o grupo tiometilo da cadeia lateral. O aldeído é uma parte essencial do sítio catalítico da sulfatase e, provavelmente, atua como um aldeído hidratado. Um dos grupos hidroxilo geminais aceita o sulfato durante a clivagem do éster de sulfato,

conduzindo à formação de um intermediário enzimático sulfatado de modo covalente. O outro hidroxilo é requerido para a eliminação subsequente do sulfato e regeneração do grupo aldeído. Esta modificação ocorre no retículo endoplasmático durante, ou pouco após, a importação do polipéptido de sulfatase nascente e é dirigida por uma sequência linear curta que envolve o resíduo cisteína (ou serina) a ser modificado. Esta sequência altamente conservada é o hexapéptido L/V-C(S)-X-P-S-R (SEQ ID NO: 32), presente na região N-terminal de todas as sulfatases eucarióticas, e muito frequentemente contém um grupo hidroxilo ou tiol no resíduo X (Dierks, T., *et al.*, *Proc. Natl Acad Sci. U. S. A.*, 1997, 94: 11963-11968).

Até à data foram identificados treze genes de sulfatases em humanos. Codificam enzimas com diferentes especificidades para substratos e localização subcelular, como lisossomas, Golgi e ER. Quatro destes genes, *ARSC*, *ARSD*, *ARSE* e *ARSF*, que codificam as arilsulfatases C, D, E e F, respetivamente, estão localizados na mesma região cromossômica (Xp22.3). Partilham semelhança significativa de sequências e uma organização genómica quase idêntica, indicando que surgiram de eventos de duplicação que ocorreram recentemente durante a evolução (Franco B, *et al.*, *Cell*, 1995, 81: 15-25; Meroni G, *et al.*, *Hum Mol Genet*, 1996, 5: 423-31).

A importância das sulfatases no metabolismo humano é sublinhada pela identificação de pelo menos oito doenças monogénicas humanas causadas pela deficiência de atividades de sulfatases individuais. A maior parte destes estados consiste em perturbações de armazenamento lisossómico nas quais as consequências fenotípicas derivam do tipo e

distribuição em tecidos do material armazenado. Entre essas contam-se cinco tipos diferentes de mucopolissacaridoses (MPS tipos II, IIIA, IIID, IVA e VI), devido a deficiências de sulfatases que atuam no catabolismo dos glicosaminoglicanos (Neufeld e Muenzer, 2001, "The mucopolysaccharidoses", Em "The Metabolic and Molecular Bases of Inherited Disease", C.R. Scriver, A.L. Beaudet, W.S. Sly, D. Valle, B. Childs, K.W. Kinzler e B. Vogelstein, editores. Nova Iorque: Mc Graw-Hill, páginas 3421-3452), e leucodistrofia metacromática (MLD), que é caracterizada pelo armazenamento de sulfolípidos nos sistemas nervosos central e periférico conducente a deterioração neurológica grave e progressiva. Duas doenças humanas adicionais são causadas por deficiências de sulfatases não lisossômicas. Estas incluem ictiose ligada ao cromossoma X, uma perturbação da pele devido a deficiência de esteroide sulfatase (STS/ARSC), e condrodisplasia *punctata*, uma perturbação que afeta ossos e cartilagens devido a deficiência de arilsulfatase E (ARSE). As sulfatases também estão implicadas em síndromas de malformações humanas induzidas por fármacos, como embriopatia Varfarínica, causada por inibição da atividade da ARSE devido a exposição *in utero* a varfarina durante a gravidez.

Numa perturbação monogénica humana intrigante, deficiência de múltiplas sulfatases (MSD), todas as atividades de sulfatases estão simultaneamente deficientes. Em consequência, o fenótipo desta doença multissistémica grave combina as características observadas em deficiências de sulfatases individuais. Células de pacientes com MSD são deficientes em atividades de sulfatases mesmo após transfeção com cDNAs que codificam sulfatases humanas,

sugerindo a presença de um mecanismo comum requerido para a atividade de todas as sulfatases (Rommerskirch e von Figura, *Proc. Natl. Acad. Sci., USA*, 1992, 89: 2561-2565). Verificou-se que a modificação pós-tradução de sulfatases é defeituosa num paciente com MSD, sugerindo que esta perturbação é causada por uma mutação num gene, ou genes, implicados na maquinaria de conversão cisteína-em-formilglicina (Schmidt, B., et al., *Cell*, 1995, 82: 271-278). Apesar do interesse biológico e médico intenso, esforços destinados à identificação deste(s) gene(s) têm sido dificultados pela raridade de pacientes com MSD e consequente falta de casos familiares adequados para efetuar o mapeamento genético.

Esta invenção proporciona métodos e composições para o diagnóstico e tratamento de Deficiência de Múltiplas Sulfatases (MIM 272200) e o tratamento de outras deficiências de sulfatases. Mais especificamente, identificámos um gene que codifica a Enzima Geradora de Formilglicina (FGE), uma enzima responsável pela modificação pós-tradução única que ocorre em sulfatases que é essencial para a função de sulfatase (formação de L-C α -formilglicina; também conhecida como *FGly* e/ou *ácido 2-amino-3-oxopropanoico*). Foi descoberto, inesperadamente, que mutações no gene FGE conduzem ao desenvolvimento de Deficiência de Múltiplas Sulfatases (MSD) em sujeitos. Também foi descoberto, inesperadamente, que a FGE intensifica a atividade de sulfatases, incluindo mas não se limitando a Iduronato 2-Sulfatase, Sulfamidase, N-Acetilgalactosamina 6-Sulfatase, N-Acetilglucosamina 6-Sulfatase, Arilsulfatase A, Arilsulfatase B, Arilsulfatase C, Arilsulfatase D, Arilsulfatase E, Arilsulfatase F, Arilsulfatase G, HSulf-1, HSulf-2, HSulf-3, HSulf-4, HSulf-

5 e HSulf-6. Considerando estas descobertas, as moléculas da presente invenção podem ser utilizadas no diagnóstico e tratamento de Deficiência de Múltiplas Sulfatases, bem como de outras deficiências de sulfatases.

A invenção proporciona um método para determinar o nível de expressão da enzima geradora de C_{α} -formilglicina (FGE) num sujeito, cujo método compreende medir *in vitro*, numa amostra obtida de um sujeito, a atividade geradora de C_{α} -formilglicina de um péptido ou polipéptido possuindo uma sequência selecionada do grupo que consiste nos aminoácidos 34-374 da SEQ ID NO: 2, SEQ ID NO: 2,5, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76 e 78.

Também proporciona um método para identificar um agente útil na modulação da atividade geradora de C_{α} -formilglicina, que compreende:

- (a) contactar uma molécula possuindo atividade geradora de C_{α} -formilglicina com um agente candidato,
- (b) medir a atividade geradora de C_{α} -formilglicina da molécula, e
- (c) comparar a atividade geradora de C_{α} -formilglicina medida da molécula com um controlo para determinar se o agente candidato modula a atividade geradora de C_{α} -formilglicina da molécula;

em que a molécula é uma molécula de ácido nucleico possuindo a sequência de nucleótidos apresentada como SEQ ID NO: 1, ou é um respetivo produto de expressão.

Também proporciona um método para determinar uma deficiência de sulfatases num sujeito caracterizada por

expressão aberrante de uma molécula de ácido nucleico ou de um respetivo produto de expressão, que compreende:

monitorizar uma amostra de um paciente quanto a um parâmetro selecionado do grupo que consiste em:

- (i) uma molécula de ácido nucleico possuindo a sequência de nucleótidos apresentada como SEQ ID NO: 1,
- (ii) um polipéptido codificado pela molécula de ácido nucleico,
- (iii) um péptido derivado do polipéptido, e
- (iv) um anticorpo que se liga seletivamente ao polipéptido ou péptido,

como determinação de uma deficiência de sulfatases no sujeito.

A amostra pode ser, por exemplo, um fluido biológico ou um tecido.

O passo de monitorização pode incluir contactar a amostra com um agente detetável selecionado do grupo que consiste no seguinte:

- (a) uma molécula de ácido nucleico que hibridiza seletivamente, em condições restritas, com a molécula de ácido nucleico de (i),
- (b) um anticorpo que se liga seletivamente ao polipéptido de (ii) ou ao péptido de (iii), e
- (c) um polipéptido ou péptido que se liga ao anticorpo de (iv).

O anticorpo, o polipéptido, o péptido ou o ácido nucleico podem ser etiquetados com uma etiqueta radioativa ou uma enzima.

O método pode incluir avaliar a amostra quanto ao péptido.

A invenção também proporciona um estojo compreendendo uma embalagem que contém um agente e um controlo; em que o agente é um anticorpo, ou respetivo fragmento, que se liga seletivamente a um péptido ou polipéptido como descrito anteriormente, ou é um ácido nucleico que se liga seletivamente a um ácido nucleico que codifica o referido péptido ou polipéptido; e em que o controlo se destina a comparação com um valor medido de ligação do referido agente ao referido péptido, polipéptido ou ácido nucleico.

O controlo pode consistir num valor predeterminado para comparação com o valor medido.

Pode compreender um epítopo do péptido ou polipéptido.

A invenção também inclui a utilização de uma molécula de ácido nucleico, péptido ou polipéptido que aumenta a atividade geradora de C_{α} -formilglicina na preparação de um medicamento para o tratamento de uma deficiência de sulfatases num sujeito; em que:

(i) a molécula de ácido nucleico é selecionada do grupo que consiste em:

(a) o complemento de uma molécula de ácido nucleico que hibridiza, em condições restritas, com uma molécula que consiste

numa sequência de nucleótidos apresentada como SEQ ID NO: 1;

- (b) uma molécula de ácido nucleico que difere de uma molécula de ácido nucleico de (a) devido à degenerescência do código genético; e
 - (c) um ácido nucleico possuindo uma sequência selecionada do grupo que consiste em: SEQ ID NO: 1, 3, 4, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77 e 80-87; e
- (ii) o péptido ou polipéptido é selecionado do grupo que consiste em:
- (a) um péptido ou um polipéptido codificado por uma molécula de ácido nucleico como descrito acima; e
 - (b) um péptido ou um polipéptido que tem uma sequência selecionada do grupo que consiste em: aminoácidos 34-374 da SEQ ID NO: 2, SEQ ID NO: 2, 5, 46, 48, 50, 52, 54, 56; 58, 60, 62, 64, 66, 68, 70, 72, 74, 76 e 78.

O medicamento descrito relativamente à utilização acima mencionada também pode compreender um agente selecionado do grupo que consiste nos seguintes: uma molécula de ácido nucleico que codifica a Iduronato 2-Sulfatase, Sulfamidase, N-Acetilgalactosamina 6-Sulfatase, N-Acetilglucosamina 6-Sulfatase, Arilsulfatase A, Arilsulfatase B, Arilsulfatase C, Arilsulfatase D, Arilsulfatase E, Arilsulfatase F, Arilsulfatase G, HSulf-1, HSulf-2, HSulf-3, HSulf-4, HSulf-5 ou HSulf-6; um produto de expressão da molécula de ácido

nucleico; e um fragmento do produto de expressão da molécula de ácido nucleico.

O agente que aumenta a atividade geradora de C_{α} -formilglicina pode ser produzido por uma célula que expressa a referida molécula de ácido nucleico. A célula pode expressar uma molécula de ácido nucleico de FGE exógena ou endógena.

A invenção também inclui uma molécula de ácido nucleico, péptido ou polipéptido que modula a atividade geradora de C_{α} -formilglicina; para utilização no tratamento de uma deficiência de sulfatases num sujeito; em que a molécula de ácido nucleico, péptido ou polipéptido é como descrito acima.

Uma composição farmacêutica compreendendo a molécula de ácido nucleico, péptido ou polipéptido em conjunto com um transportador farmacêuticamente aceitável também pertence ao âmbito da presente invenção.

A invenção também inclui um método para identificar um agente candidato útil no tratamento de uma deficiência de sulfatases, que compreende:

determinar a expressão de um conjunto de moléculas de ácidos nucleicos *in vitro* numa célula ou tecido em condições tais que, na ausência de um agente candidato, permitem uma primeira quantidade de expressão do conjunto de moléculas de ácidos nucleicos; em que o conjunto de moléculas de ácidos nucleicos compreende pelo menos uma molécula de ácido nucleico que é uma molécula de ácido nucleico como descrito anteriormente;

contactar a célula ou tecido com o agente candidato *in vitro*, e
 detetar uma quantidade teste de expressão do conjunto de
 moléculas de ácidos nucleicos;
 em que um aumento da quantidade teste de expressão na
 presença do agente candidato, relativamente à primeira
 quantidade de expressão, indica que o agente candidato é
 útil no tratamento da deficiência de sulfatases.

Também inclui uma série de moléculas de ácidos nucleicos em
 fase sólida consistindo num conjunto de moléculas de ácidos
 nucleicos, respetivos produtos de expressão, ou respetivos
 fragmentos, quando fixados num substrato sólido; em que
 cada molécula de ácido nucleico codifica um polipéptido
 selecionado do grupo que consiste nos aminoácidos 34-374 da
 SEQ ID NO: 2, SEQ ID NO: 2, 5, 46, 48, 50, 52, 54, 56, 58,
 60, 62, 64, 66, 68, 70, 72, 74, 76 e 78, Iduronato 2-
 Sulfatase, Sulfamidase, N-Acetilgalactosamina 6-Sulfatase,
 N-Acetilglucosamina 6-Sulfatase, Arilsulfatase A,
 Arilsulfatase B, Arilsulfatase C, Arilsulfatase D,
 Arilsulfatase E, Arilsulfatase F, Arilsulfatase G, HSulf-1,
 HSulf-2, HSulf-3, HSulf-4, HSulf-5 e HSulf-6.

A série de moléculas de ácidos nucleicos em fase sólida
 pode compreender pelo menos uma molécula de ácido nucleico
 de controlo.

O conjunto de moléculas de ácidos nucleicos pode
 compreender pelo menos uma ou pelo menos duas moléculas de
 ácidos nucleicos que codificam um polipéptido selecionado
 do grupo que consiste nos seguintes: aminoácidos 34-374 da
 SEQ ID NO: 2, SEQ ID NO: 2, 5, 46, 48, 50, 52, 54, 56, 58,
 60, 62, 64, 66, 68, 70, 72, 74, 76 e 78, Iduronato 2-

Sulfatase, Sulfamidase, N-Acetilgalactosamina 6-Sulfatase, N-Acetilglucosamina 6-Sulfatase, Arilsulfatase A, Arilsulfatase B, Arilsulfatase C, Arilsulfatase D, Arilsulfatase E, Arilsulfatase F, Arilsulfatase G, HSulf-1, HSulf-2, HSulf-3, HSulf-4, HSulf-5 e HSulf-6.

No âmbito da presente invenção também está incluído um alelo variante isolado de um gene FGE humano que codifica um polipéptido FGE variante, em que o polipéptido compreende:

uma sequência de aminoácidos compreendendo pelo menos uma variação na SEQ ID NO: 2, em que a pelo menos uma variação compreende: Met1Arg; Met1Val; Ser155Pro; Cys218Tyr; Ala279Val; Arg327Stop; Cys336Arg; Arg345Cys; Arg349Trp; Arg349Gln; Ser359Stop; ou uma combinação destas.

A invenção inclui um polipéptido FGE humano variante isolado compreendendo uma sequência de aminoácidos com pelo menos uma variação na SEQ ID NO: 2, em que a pelo menos uma variação compreende: Met1Arg; Met1Val; Ser155Pro; Cys218Tyr; Ala279Val; Arg327Stop; Cys336Arg; Arg345Cys; Arg349Trp; Arg349Gln; Ser359Stop; ou uma combinação destas.

Um anticorpo possuindo um tal polipéptido FGE humano variante como imunogénio também pertence ao âmbito da presente invenção.

O anticorpo pode ser, por exemplo, um anticorpo policlonal, um anticorpo monoclonal, um anticorpo quimérico ou um anticorpo etiquetado de forma detetável.

Se estiver etiquetado de forma detetável, a etiqueta detetável pode compreender, por exemplo, um elemento radioativo, um químico fluorescente ou uma enzima.

A invenção também inclui um método para produzir uma sulfatase ativada, cujo método compreende:

- (a) contactar a sulfatase com uma Enzima Geradora de C_{α} -Formilglicina (FGE) *in vitro*; em que a Enzima Geradora de C_{α} -Formilglicina é um polipéptido ou um péptido como descrito na reivindicação 1; ou
- (b) proporcionar uma célula que coexpressa uma sulfatase e uma Enzima Geradora de C_{α} -Formilglicina (FGE), de modo que é produzida na célula sulfatase ativada; em que a célula compreende DNA ou RNA heterólogo que conduz a expressão aumentada da sulfatase ativada, relativamente ao que ocorreria na ausência do DNA ou RNA heterólogo; e em que a Enzima Geradora de C_{α} -Formilglicina é um polipéptido ou um péptido como descrito anteriormente.

A sulfatase pode ser seleccionada, por exemplo, do grupo que consiste em Iduronato 2-Sulfatase, Sulfamidase, N-Acetilgalactosamina 6-Sulfatase, N-Acetilglucosamina 6-Sulfatase, Arilsulfatase A, Arilsulfatase B, Arilsulfatase C, Arilsulfatase D, Arilsulfatase E, Arilsulfatase F, Arilsulfatase G, HSulf-1, HSulf-2, HSulf-3, HSulf-4, HSulf-5 e HSulf-6.

A deficiência de sulfatases referida relativamente aos métodos ou utilizações descritos anteriormente pode ser seleccionada do grupo que consiste em: Deficiência de

Múltiplas Sulfatases, Zucopolissacaridose II, Mucopolissacaridose IIIA, Mucopolissacaridose IVA, Mucopolissacaridose VI, Mucopolissacaridose VIII, Leucodistrofia Metacromática, Condrodisplasia *Punctata* 1 Recessiva ligada ao cromossoma X e Ictiose ligada ao cromossoma X.

Descrição Breve das Sequências

SEQ ID NO: 1 é a sequência de nucleótidos do cDNA da FGE humana.

SEQ ID NO: 2 é a sequência de aminoácidos prevista do produto de tradução do cDNA da FGE humana (SEQ ID NO: 1).

SEQ ID NO: 3 é a sequência de nucleótidos do cDNA da FGE humana que codifica o polipéptido da SEQ ID NO: 2 (isto é, nucleótidos 20-1141 da SEQ ID NO: 1).

SEQ ID NO: 4 é a sequência de nucleótidos do Número de Acesso da GenBank AK075459.

SEQ ID NO: 5 é a sequência de aminoácidos prevista do produto de tradução da SEQ ID NO: 4, um produto proteico sem nome com o Número de Acesso da GenBank BAC11634.

SEQ ID NO: 6 é a sequência de nucleótidos do cDNA da Iduronato 2-Sulfatase humana (Número de Acesso da GenBank M58342).

SEQ ID NO: 7 é a sequência de aminoácidos prevista do produto de tradução do cDNA da Iduronato 2-Sulfatase humana (SEQ ID NO: 6).

SEQ ID NO: 8 é a sequência de nucleótidos do cDNA da Sulfamidase humana (Número de Acesso da GenBank U30894).

SEQ ID NO: 9 é a sequência de aminoácidos prevista do produto de tradução do cDNA da Sulfamidase humana (SEQ ID NO: 8).

SEQ ID NO: 10 é a sequência de nucleótidos do cDNA da N-Acetilgalactosamina 6-Sulfatase humana (Número de Acesso da GenBank U06088).

SEQ ID NO: 11 é a sequência de aminoácidos prevista do produto de tradução do cDNA da N-Acetilgalactosamina 6-Sulfatase humana (SEQ ID NO: 10).

SEQ ID NO: 12 é a sequência de nucleótidos do cDNA da N-Acetilglucosamina 6-Sulfatase humana (Número de Acesso da GenBank Z12173).

SEQ ID NO: 13 é a sequência de aminoácidos prevista do produto de tradução do cDNA da N-Acetilglucosamina 6-Sulfatase humana (SEQ ID NO: 12).

SEQ ID NO: 14 é a sequência de nucleótidos do cDNA da Arilsulfatase A humana (Número de Acesso da GenBank X52151).

SEQ ID NO: 15 é a sequência de aminoácidos prevista do produto de tradução do cDNA da Arilsulfatase A humana (SEQ ID NO: 14).

SEQ ID NO: 16 é a sequência de nucleótidos do cDNA da Arilsulfatase B humana (Número de Acesso da GenBank J05225).

SEQ ID NO: 17 é a sequência de aminoácidos prevista do produto de tradução do cDNA da Arilsulfatase B humana (SEQ ID NO: 16).

SEQ ID NO: 18 é a sequência de nucleótidos do cDNA da Arilsulfatase C humana (Número de Acesso da GenBank J04964).

SEQ ID NO: 19 é a sequência de aminoácidos prevista do produto de tradução do cDNA da Arilsulfatase C humana (SEQ ID NO: 18).

SEQ ID NO: 20 é a sequência de nucleótidos do cDNA da Arilsulfatase D humana (Número de Acesso da GenBank X83572).

SEQ ID NO: 21 é a sequência de aminoácidos prevista do produto de tradução do cDNA da Arilsulfatase D humana (SEQ ID NO: 20).

SEQ ID NO: 22 é a sequência de nucleótidos do cDNA da Arilsulfatase E humana (Número de Acesso da GenBank X83573).

SEQ ID NO: 23 é a sequência de aminoácidos prevista do produto de tradução do cDNA da Arilsulfatase E humana (SEQ ID NO: 22).

SEQ ID NO: 24 é a sequência de nucleótidos do cDNA da Arilsulfatase F humana (Número de Acesso da GenBank X97868).

SEQ ID NO: 25 é a sequência de aminoácidos prevista do produto de tradução do cDNA da Arilsulfatase F humana (SEQ ID NO: 24).

SEQ ID NO: 26 é a sequência de nucleótidos do cDNA da Arilsulfatase G humana (Número de Acesso da GenBank BC012375).

SEQ ID NO: 27 é a sequência de aminoácidos prevista do produto de tradução da Arilsulfatase G humana (SEQ ID NO: 26).

SEQ ID NO: 28 é a sequência de nucleótidos do cDNA da HSulf-1 (Número de Acesso da GenBank AI101175).

SEQ ID NO: 29 é a sequência de aminoácidos prevista do produto de tradução do cDNA da HSulf-1 (SEQ ID NO: 28).

SEQ ID NO: 30 é a sequência de nucleótidos do cDNA da HSulf-2 (Número de Acesso da GenBank AY101176).

SEQ ID NO: 31 é a sequência de aminoácidos prevista do produto de tradução do cDNA da HSulf-2 (SEQ ID NO: 30).

SEQ ID NO: 32 é o hexapéptido altamente conservado L/V-FGly-X-P-S-R presente em sulfatases.

SEQ ID NO: 33 é um substrato de formação de *FGly* sintético; a sua sequência primária é derivada da Arilsulfatase A humana.

SEQ ID NO: 34 é o oligopéptido misturado **PVSLPTRSCAALLTGR**.

SEQ ID NO: 35 é o oligopéptido de Ser69 **PVSLSTPSRAALLTGR**.

SEQ ID NO: 36 é o "primer" específico para FGE humana 1199nc.

SEQ ID NO: 37 é o "primer" direto específico para FGE humana 1c.

SEQ ID NO: 38 é o "primer" reverso específico para FGE humana 1182c.

SEQ ID NO: 39 é o "primer" 5' específico para FGE humana contendo um sítio EcoRI.

SEQ ID NO: 40 é um "primer" específico para HA.

SEQ ID NO: 41 é um "primer" específico para c-myc.

SEQ ID NO: 42 é um "primer" específico para RGS-His₆.

SEQ ID NO: 43 é o oligopéptido tríptico **SONTPDSSASNLGFR** de uma preparação de FGE humana.

SEQ ID NO: 44 é o oligopéptido tríptico **MVPIPAGVFTMGTDPPQIK** de uma preparação de FGE humana.

SEQ ID NO: 45 é a sequência de nucleótidos do parálogo da FGE2 humana (GenBank GI: 24308053).

SEQ ID NO: 46 é a sequência de aminoácidos prevista do produto de tradução do parálogo da FGE2 humana (SEQ ID NO: 45).

SEQ ID NO: 47 é a sequência de nucleótidos do parálogo da FGE de ratinho (GenBank GI: 26344956).

SEQ ID NO: 48 é a sequência de aminoácidos prevista do produto de tradução do parálogo da FGE de ratinho (SEQ ID NO: 47).

SEQ ID NO: 49 é a sequência de nucleótidos do ortólogo da FGE de ratinho (GenBank GI: 22122361).

SEQ ID NO: 50 é a sequência de aminoácidos prevista do produto de tradução do ortólogo da FGE de ratinho (SEQ ID NO: 49).

SEQ ID NO: 51 é a sequência de nucleótidos do ortólogo da FGE da mosca da fruta (GenBank GI: 20130397).

SEQ ID NO: 52 é a sequência de aminoácidos prevista do produto de tradução do ortólogo da FGE da mosca da fruta (SEQ ID NO: 51).

SEQ ID NO: 53 é a sequência de nucleótidos do ortólogo da FGE de mosquito (GenBank GI: 21289310).

SEQ ID NO: 54 é a sequência de aminoácidos prevista do produto de tradução do ortólogo da FGE de mosquito (SEQ ID NO: 53).

SEQ ID NO: 55 é a sequência de nucleótidos do ortólogo intimamente relacionado da FGE de *S. coelicolor* (GenBank GI: 21225812).

SEQ ID NO: 56 é a sequência de aminoácidos prevista do produto de tradução do ortólogo da FGE de *S. coelicolor* (SEQ ID NO: 55).

SEQ ID NO: 57 é a sequência de nucleótidos do ortólogo intimamente relacionado da FGE de *C. efficiens* (GenBank GI: 25028125).

SEQ ID NO: 58 é a sequência de aminoácidos prevista do produto de tradução do ortólogo da FGE de *C. efficiens* (SEQ ID NO: 57).

SEQ ID NO: 59 é a sequência de nucleótidos do ortólogo da FGE de *N. aromaticivorans* (GenBank GI: 23108562).

SEQ ID NO: 60 é a sequência de aminoácidos prevista do produto de tradução do ortólogo da FGE de *N. aromaticivorans* (SEQ ID NO: 59).

SEQ ID NO: 61 é a sequência de nucleótidos do ortólogo da FGE de *M. loti* (GenBank GI: 13474559).

SEQ ID NO: 62 é a sequência de aminoácidos prevista do produto de tradução do ortólogo da FGE de *M. loti* (SEQ ID NO: 61).

SEQ ID NO: 63 é a sequência de nucleótidos do ortólogo da FGE de *B. fungorum* (GenBank GI: 22988809).

SEQ ID NO: 64 é a sequência de aminoácidos prevista do produto de tradução do ortólogo da FGE de *B. fungorum* (SEQ ID NO: 63).

SEQ ID NO: 65 é a sequência de nucleótidos do ortólogo da FGE de *S. meliloti* (GenBank GI: 16264068).

SEQ ID NO: 66 é a sequência de aminoácidos prevista do produto de tradução do ortólogo da FGE de *S. meliloti* (SEQ ID NO: 65).

SEQ ID NO: 67 é a sequência de nucleótidos do ortólogo da FGE de *Microscilla* sp. (GenBank GI: 14518334).

SEQ ID NO: 68 é a sequência de aminoácidos prevista do produto de tradução do ortólogo da FGE de *Microscilla* sp. (SEQ ID NO: 67).

SEQ ID NO: 69 é a sequência de nucleótidos do ortólogo da FGE de *P. putida* KT2440 (GenBank GI: 26990068).

SEQ ID NO: 70 é a sequência de aminoácidos prevista do produto de tradução do ortólogo da FGE de *P. putida* KT2440 (SEQ ID NO: 69).

SEQ ID NO: 71 é a sequência de nucleótidos do ortólogo da FGE de *R. metallidurans* (GenBank GI: 22975289).

SEQ ID NO: 72 é a sequência de aminoácidos prevista do produto de tradução do ortólogo da FGE de *R. metallidurans* (SEQ ID NO: 71).

SEQ ID NO: 73 é a sequência de nucleótidos do ortólogo da FGE de *P. marinus* (GenBank GI: 23132010).

SEQ ID NO: 74 é a sequência de aminoácidos prevista do produto de tradução do ortólogo da FGE de *P. marinus* (SEQ ID NO: 73).

SEQ ID NO: 75 é a sequência de nucleótidos do ortólogo da FGE de *C. crescentus* CB15 (GenBank GI: 16125425).

SEQ ID NO: 76 é a sequência de aminoácidos prevista do produto de tradução do ortólogo da FGE de *C. crescentus* CB 15 (SEQ ID NO: 75).

SEQ ID NO: 77 é a sequência de nucleótidos do ortólogo da FGE de *M. tuberculosis* Ht37Rv (GenBank GI: 15607852).

SEQ ID NO: 78 é a sequência de aminoácidos prevista do produto de tradução do ortólogo da FGE de *M. tuberculosis* Ht37Rv (SEQ ID NO: 77).

SEQ ID NO: 79 é o heptapéptido altamente conservado presente no subdomínio 3 de ortólogos e parálogos da FGE.

SEQ ID NO: 80 é a sequência de nucleótidos do fragmento EST do ortólogo da FGE com o Número de Acesso da GenBank: CA379852.

SEQ ID NO: 81 é a sequência de nucleótidos do fragmento EST do ortólogo da FGE com o Número de Acesso da GenBank: AI721440.

SEQ ID NO: 82 é a sequência de nucleótidos do fragmento EST do ortólogo da FGE com o Número de Acesso da GenBank: BJ505402.

SEQ ID NO: 83 é a sequência de nucleótidos do fragmento EST do ortólogo da FGE com o Número de Acesso da GenBank: BJ054666.

SEQ ID NO: 84 é a sequência de nucleótidos do fragmento EST do ortólogo da FGE com o Número de Acesso da GenBank: AL892419.

SEQ ID NO: 85 é a sequência de nucleótidos do fragmento EST do ortólogo da FGE com o Número de Acesso da GenBank: CA064079.

SEQ ID NO: 86 é a sequência de nucleótidos do fragmento EST do ortólogo da FGE com o Número de Acesso da GenBank: BF189614.

SEQ ID NO: 87 é a sequência de nucleótidos do fragmento EST do ortólogo da FGE com o Número de Acesso da GenBank: AV609121.

SEQ ID NO: 88 é a sequência de nucleótidos do cDNA da HSulf-3.

SEQ ID NO: 89 é a sequência de aminoácidos prevista do produto de tradução do cDNA da HSulf-3 (SEQ ID NO: 88).

SEQ ID NO: 90 é a sequência de nucleótidos do cDNA da HSulf-4.

SEQ ID NO: 91 é a sequência de aminoácidos prevista do produto de tradução do cDNA da HSulf-4 (SEQ ID NO: 90).

SEQ ID NO: 92 é a sequência de nucleótidos do cDNA da HSulf-5.

SEQ ID NO: 93 é a sequência de aminoácidos prevista do produto de tradução do cDNA da HSulf-5 (SEQ ID NO: 92).

SEQ ID NO: 94 é a sequência de nucleótidos do cDNA da HSulf-6.

SEQ ID NO: 95 é a sequência de aminoácidos prevista do produto de tradução do cDNA da HSulf-6 (SEQ ID NO: 94).

Descrição Breve das Figuras

Fig. 1: Espetros de massa MALDI-TOF esquemáticos de P23 após incubação na ausência (A) ou presença (B) de um extrato solúvel de microssomas de testículo bovino.

Fig. 2: Árvore filogenética derivada de um alinhamento de FGE humana e 21 proteínas da semente PFAM-DUF323.

Fig. 3: Organização do *locus* do gene FGE humano e murino. Os exões estão mostrados à escala na forma de retângulos e retângulos claros (*locus* murino). Os números acima das linhas de intrões indicam a dimensão dos intrões em quilobases.

Fig. 4: Diagrama que mostra um mapa do Plasmídeo de Expressão de FGE pXMG.1.3.

Fig. 5: Gráfico de barras que ilustra a Atividade da N-Acetilgalactosamina 6-Sulfatase em Células 36F Transfetadas de Modo Transitório com o Plasmídeo de Expressão de FGE.

Fig. 6: Gráfico de barras que ilustra a Atividade *Específica* da N-Acetilgalactosamina 6-Sulfatase em Células 36F Transfetadas de Modo Transitório com o Plasmídeo de Expressão de FGE.

Fig. 7: Gráfico de barras que ilustra a Produção de N-Acetilgalactosamina 6-Sulfatase em Células 36F Transfetadas de Modo Transitório com o Plasmídeo de Expressão de FGE.

Fig. 8: Gráfico que ilustra a Atividade da Iduronato 2-Sulfatase em Células 30C6 Transfetadas de Modo Transitório com o Plasmídeo de Expressão de FGE.

Fig. 9: Ilustra um estojo que implementa características da presente invenção.

Os inventores descobriram o gene, aquele gene, que codifica a Enzima Geradora de Formilglicina (FGE), uma enzima responsável pela modificação pós-tradução única que ocorre em sulfatases que é essencial para a função de sulfatase: A formação de L-C_α-formilglicina (também conhecida como *FGly* e/ou *ácido 2-amino-3-oxopropanoico*). Foi descoberto, inesperadamente, que mutações no gene FGE conduzem ao desenvolvimento de Deficiência de Múltiplas Sulfatases (MSD) em sujeitos. Também foi descoberto, inesperadamente,

que a FGE aumenta a atividade de sulfatases, incluindo mas não se limitando a Iduronato 2-Sulfatase, Sulfamidase, N-Acetilgalactosamina 6-Sulfatase, N-Acetilglucosamina 6-Sulfatase, Arilsulfatase A, Arilsulfatase B, Arilsulfatase C, Arilsulfatase D, Arilsulfatase E, Arilsulfatase F, Arilsulfatase G, HSulf-1, HSulf-2, HSulf-3, HSulf-4, HSulf-5 e HSulf-6, e sulfatases descritas nos Requerimentos provisórios U.S. com os números de publicação 20030073118, 20030147875, 20030148920, 20030162279 e 20030166283. Considerando estas descobertas, moléculas descritas aqui podem ser utilizadas no diagnóstico e/ou tratamento de Deficiência de Múltiplas Sulfatases, bem como no tratamento de outras deficiências de sulfatases.

São proporcionados métodos para utilização de moléculas no diagnóstico de Deficiência de Múltiplas Sulfatases.

Adicionalmente, também são proporcionados métodos para utilização destas moléculas *in vivo* ou *in vitro* com a finalidade de modular a formação de *FGly* em sulfatases, métodos de tratamento de estados associados a essa modificação e composições úteis na produção de preparações terapêuticas para o tratamento de Deficiência de Múltiplas Sulfatases, bem como de outras deficiências de sulfatases.

Também são proporcionados polipéptidos que modulam a formação de *FGly* em sulfatases, ácidos nucleicos isolados que codificam esses polipéptidos, modificações e variantes funcionais dos anteriores, fragmentos úteis dos anteriores, bem como terapêuticas e diagnósticos, métodos de investigação, composições e ferramentas relacionados com aqueles.

"Atividade geradora de C α -formilglicina" refere-se à capacidade de uma molécula para formar, ou intensificar a formação de, *FGly* num substrato. O substrato pode ser uma sulfatase, como descrito aqui noutro local, ou um oligopéptido sintético (ver, por exemplo, SEQ ID NO: 33, e os Exemplos). O substrato contém preferivelmente o hexapéptido conservado da SEQ ID NO: 32 [L/V-C(S)-X-P-S-R]. Métodos para avaliar a formação de *FGly* são os descritos na área (ver, por exemplo, Dierks, T., et al., *Proc. Natl. Acad. Sci. U. S. A.*, 1997, 94: 11963-11968), e aqui noutro local (ver, por exemplo, os Exemplos). Uma "molécula," como usado aqui, abrange "ácidos nucleicos" e "polipéptidos". Moléculas de FGE são capazes de formar, ou de intensificar/aumentar a formação de, *FGly* *in vivo* e *in vitro*.

"Intensificar (ou "aumentar")" a atividade geradora de C α -formilglicina, como usado aqui, refere-se tipicamente a expressão aumentada de FGE e/ou seu polipéptido codificado. Expressão aumentada refere-se a aumentar (isto é, numa extensão detetável) a replicação, transcrição e/ou tradução de quaisquer dos ácidos nucleicos (ácidos nucleicos de FGE, como descrito aqui noutro local), uma vez que a regulação positiva de quaisquer destes processos origina um aumento da concentração/quantidade do polipéptido codificado pelo gene (ácido nucleico). Intensificar (ou aumentar) a atividade geradora de C α -formilglicina também se refere a prevenir ou inibir a degradação (por exemplo, via ubiquitinação aumentada), regulação negativa, etc., da FGE, resultando, por exemplo, num t $\frac{1}{2}$ (semivida) da molécula de FGE aumentado ou estável em comparação com um controlo. Regulação negativa ou expressão decrescida refere-se a expressão decrescida de um gene e/ou do seu polipéptido codificado. A regulação positiva ou regulação negativa da

expressão genética pode ser diretamente determinada detetando um aumento ou decréscimo, respetivamente, do nível de mRNA para o gene (por exemplo, FGE), ou do nível de expressão proteica do polipéptido codificado pelo gene, utilizando quaisquer meios adequados conhecidos na área, como métodos de hibridização de ácidos nucleicos ou deteção de anticorpos, respetivamente, e em comparação com controlos. A regulação positiva ou regulação negativa da expressão do gene FGE também pode ser determinada indiretamente por deteção de uma alteração da atividade geradora de C_{α} -formilglicina.

"Expressão", como usado aqui, refere-se à expressão de ácidos nucleicos e/ou polipéptidos, bem como a atividade da molécula polipeptídica (por exemplo, atividade geradora de C_{α} -formilglicina da molécula).

Um método envolve a clonagem de um cDNA que codifica FGE. FGE é uma molécula de ácido nucleico isolado que compreende uma molécula de ácido nucleico da SEQ ID NO: 1, e codifica um polipéptido com atividade geradora de C_{α} -formilglicina. A sequência do cDNA da FGE humana está apresentada como SEQ ID NO: 1, e a sequência de aminoácidos prevista deste produto proteico codificado pelo cDNA está apresentada como SEQ ID NO: 2.

Como usado aqui, um sujeito é um mamífero ou um mamífero não humano. Em todas as formas de realização é preferida FGE humana e sujeitos humanos.

São divulgados um polipéptido FGE isolado, o cDNA que codifica este polipéptido, modificações e variantes funcionais dos anteriores, fragmentos úteis dos anteriores,

bem como diagnósticos e terapêuticas relacionadas com aqueles.

Como usado aqui relativamente a ácidos nucleicos, o termo "isolado" significa: (i) amplificado *in vitro*, por exemplo, por reação em cadeia de polimerase (PCR); (ii) produzido de modo recombinante por clonagem; (iii) purificado, tal como por clivagem e separação em gel; ou (iv) sintetizado, por exemplo, por síntese química. Um ácido nucleico isolado é tal que é facilmente manipulado por técnicas de DNA recombinante bem conhecidas na área. Assim, uma sequência de nucleótidos contida num vetor no qual são conhecidos sítios de restrição 5' e 3' ou para a qual foram divulgadas sequências de "primers" para a reação em cadeia de polimerase (PCR) é considerada isolada, mas uma sequência de ácido nucleico existente no seu estado nativo no seu hospedeiro natural não está. Um ácido nucleico isolado pode estar substancialmente purificado, mas tal não é necessário. Por exemplo, um ácido nucleico que está isolado num vetor de clonagem ou expressão não é puro, pois pode compreender apenas uma percentagem muito pequena do material da célula onde reside. No entanto, esse ácido nucleico está isolado, tal como o termo é usado aqui, pois é facilmente manipulado por técnicas comuns conhecidas dos profissionais.

Como usado aqui relativamente a polipéptidos, o termo "isolado" significa separado do seu ambiente nativo em forma suficientemente pura para poder ser manipulado ou usado para qualquer uma das finalidades descritas aqui. Assim, isolado significa suficientemente puro para ser utilizado (i) para dirigir e/ou isolar anticorpos, (ii)

como reagente num ensaio, (iii) para sequenciação, (iv) como agente terapêutico, etc.

Moléculas de ácidos nucleicos isolados que codificam um polipéptido FGE com atividade geradora de C_{α} -formilglicina incluem: (a) moléculas de ácidos nucleicos que hibridizam, em condições restritas, com uma molécula que consiste num ácido nucleico da SEQ ID NO: 1 e que codificam um polipéptido FGE com atividade geradora de C_{α} -formilglicina, (b) deleções, adições e substituições de (a) que codificam um polipéptido FGE respetivo com atividade geradora de C_{α} -formilglicina, (c) moléculas de ácidos nucleicos que diferem das moléculas de ácidos nucleicos de (a) ou (b) quanto à sequência de codões devido à degenerescência do código genético, e (d) complementos de (a), (b) ou (c). "Complementos", como usado aqui, inclui "filamentos complementares completos ou filamentos 100% complementares de (a), (b) ou (c).

São divulgados homólogos e alelos dos ácidos nucleicos FGE também possuindo atividade geradora de C_{α} -formilglicina. Homólogos, como descrito aqui, incluem as moléculas identificadas aqui noutra local (ver, por exemplo, SEQ ID NOS: 4, 5, 45-78 e 80-87), isto é, ortólogos e parálogos. Podem ser identificados outros homólogos seguindo os ensinamentos apresentados aqui, bem como por técnicas convencionais. Uma vez que todos os homólogos de FGE descritos aqui partilham atividade geradora de C_{α} -formilglicina, podem ser utilizados de forma intermutável com a molécula FGE humana em todos os aspetos descritos aqui.

Sequências de ácidos nucleicos que codificam polipéptidos FGE podem hibridizar com o complemento de uma molécula de

ácido nucleico que consiste na região codificadora da SEQ ID NO: 1, em condições restritas. Numa forma de realização importante, o termo "condições restritas", como usado aqui, refere-se a parâmetros familiares na área. Com ácidos nucleicos, é afirmado que condições de hibridização são restritas tipicamente em condições de baixa força iônica e uma temperatura logo abaixo da temperatura de fusão (T_f) do complexo híbrido de DNA (tipicamente, cerca de 3°C inferior à T_f do híbrido). Uma restrição mais elevada permite uma correlação mais específica entre a sequência da sonda e o alvo. Condições restritas utilizadas na hibridização de ácidos nucleicos são bem conhecidas na área e podem ser encontradas em referências que compilam esses métodos, por exemplo, "Molecular Cloning: A Laboratory Manual", J. Sambrook, et al., editores, Segunda Edição, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, Nova Iorque, 1989, ou "Current Protocols in Molecular Biology", F.M. Ausubel, et al., editores, John Wiley & Sons, Inc., Nova Iorque. Um exemplo de "condições restritas" consiste em hibridização a 65°C em 6 x SSC. Outro exemplo de condições restritas consiste em hibridização a 65°C em tampão de hibridização que consiste em 3,5 x SSC, 0,02% Ficoll, 0,02% polivinilpirrolidona, 0,02% Albumina de Soro Bovino, 2,5 mM NaH_2PO_4 [pH7], 0,5% SDS, 2 mM EDTA. (SSC é 0,15 M cloreto de sódio/0,15 M citrato de sódio, pH7; SDS é dodecilsulfato de sódio; e EDTA é ácido etilenodiaminotetracético). Após a hibridização, a membrana para onde é transferido o DNA é lavada a 2 x SSC à temperatura ambiente e depois a 0,1 x SSC/0,1 x SDS a temperaturas até 68°C. Noutro exemplo, uma alternativa à utilização de uma solução de hibridização aquosa é a utilização de uma solução de hibridização de formamida. Assim, condições de hibridização restritas podem ser obtidas utilizando, por exemplo, uma solução 50%

formamida e 42°C. Há outras condições, reagentes, e assim por diante, que podem ser utilizados e que originariam um grau semelhante de restrição. O profissional estará familiarizado com essas condições; como tal, elas não são apresentadas aqui. No entanto, será entendido que o profissional será capaz de manipular as condições de modo a permitir a identificação clara de homólogos e alelos de ácidos nucleicos FGE descritos aqui. O profissional também está familiarizado com a metodologia de rastreamento de células e bibliotecas quanto a expressão dessas moléculas que, então, são isoladas de modo rotineiro, seguido de isolamento da molécula de ácido nucleico pertinente e sequenciação.

Em geral, homólogos e alelos partilharão tipicamente pelo menos 40% de identidade de nucleótidos e/ou pelo menos 50% de identidade de aminoácidos com a SEQ ID NO: 1 e SEQ ID NO: 2, respetivamente; nalguns casos partilharão pelo menos 50% identidade de nucleótidos e/ou pelo menos 65% de identidade de aminoácidos e ainda noutros casos partilharão pelo menos 60% de identidade de nucleótidos e/ou pelo menos 75% de identidade de aminoácidos. Noutros casos, homólogos e alelos partilharão tipicamente pelo menos 90%, 95% ou mesmo 99% de identidade de nucleótidos e/ou pelo menos 95%, 98% ou mesmo 99% de identidade de aminoácidos com a SEQ ID NO: 1 e SEQ ID NO: 2, respetivamente. A homologia pode ser calculada utilizando várias ferramentas de "software" publicamente disponíveis desenvolvidas pelo NCBI (Bethesda, Maryland). Ferramentas exemplificativas incluem o algoritmo heurístico de Altschul SF, *et al.*, (*J Mol Biol*, 1990, 215: 403-410), também conhecido como BLAST. Podem ser obtidos alinhamentos de pares e ClustalW (empregando a matriz BLOSUM30) bem como análise hidropática de Kyte-Doolittle

utilizando ferramentas públicas (EMBL, Heidelberg, Alemanha) e comerciais (por exemplo, o "software" de análise de sequências MacVector do Oxford Molecular Group/Genetics Computer Group, Madison, WI). Também estão abrangidos complementos de Watson-Crick dos ácidos nucleicos anteriores.

No rastreio de genes relacionados com FGE, como homólogos e alelos de FGE, pode realizar-se uma coloração "Southern" utilizando as condições anteriores, juntamente com uma sonda radioativa. Após lavagem da membrana para onde é finalmente transferido o DNA, a membrana pode ser colocada contra um filme de raios X ou uma placa PhosphoImager para detetar o sinal radioativo.

Dados os ensinamentos apresentados aqui de um clone de cDNA de FGE humana completo, outras sequências de mamífero, como o clone de cDNA de ratinho correspondente ao gene FGE humano, podem ser isoladas de uma biblioteca de cDNA, utilizando técnicas comuns de hibridização de colónias.

Ácidos nucleicos degenerados incluem codões alternativos aos presentes nos materiais nativos. Por exemplo, resíduos serina são codificados pelos codões TCA, AGT, TCC, TCG, TCT e AGC. Assim, será claro para o profissional que quaisquer dos tripletos de nucleótidos codificadores de serina podem ser empregues para dirigir o aparato da síntese proteica, *in vitro* ou *in vivo*, para incorporar um resíduo serina num polipéptido FGE em alongamento. De modo semelhante, tripletos de sequências de nucleótidos que codificam outros resíduos de aminoácidos incluem mas não estão limitados a: CCA, CCC, CCG e CCT (codões de prolina); CGA, CGC, CGG, CGT, AGA e AGG (codões de arginina); ACA, ACC, ACG e ACT

(codões de treonina); AAC e AAT (codões de asparagina); e ATA, ATC e ATT (codões de isoleucina). Outros resíduos de aminoácidos podem ser codificados de modo semelhante por múltiplas sequências de nucleótidos. Assim, podem ser utilizados ácidos nucleicos degenerados que diferem dos ácidos nucleicos isolados de modo biológico quanto à sequência de codões, devido à degenerescência do código genético.

São proporcionados fragmentos únicos isolados da SEQ ID NO: 1 ou SEQ ID NO: 3, ou respectivos complementos. Um fragmento único é tal que é uma "assinatura" do ácido nucleico maior. Por exemplo, o fragmento único é suficientemente longo para assegurar que a sua sequência precisa não se encontra em moléculas do genoma humano fora dos ácidos nucleicos FGE definidos acima (e alelos humanos). Os profissionais podem aplicar não mais do que procedimentos de rotina para determinar se um fragmento é único no genoma humano. No entanto, fragmentos únicos excluem fragmentos completamente compostos pelas sequências de nucleótidos selecionadas do grupo que consiste na SEQ ID NO: 4, e/ou outras sequências previamente publicadas à data de registo deste requerimento.

Um fragmento que é completamente composto pela sequência descrita nos depósitos da GenBank anteriores é tal que não inclui nenhum dos nucleótidos únicos das sequências descritas aqui. Assim, um fragmento único deve conter uma sequência de nucleótidos diferente da sequência exata daquelas dos depósitos da GenBank, ou respectivos fragmentos. A diferença pode ser uma adição, deleção ou substituição relativamente à sequência da GenBank, ou pode

ser uma sequência totalmente separada da sequência da GenBank.

Fragmentos únicos podem ser utilizados como sondas em ensaios de coloração "Southern" e "Northern" para identificar esses ácidos nucleicos, ou podem ser utilizados em ensaios de amplificação, como os que empregam PCR. Como é conhecido dos profissionais, sondas grandes, tais como com 200, 250, 300 ou mais nucleótidos, são preferidas para certas utilizações, como colorações "Southern" e "Northern", ao passo que fragmentos mais pequenos serão preferidos para utilizações tais como PCR. Fragmentos únicos também podem ser utilizados para produzir proteínas de fusão para a geração de anticorpos ou determinação da ligação dos fragmentos polipeptídicos, como demonstrado nos Exemplos, ou para a geração de componentes de imunoenaios. De igual modo, fragmentos únicos podem ser empregues para produzir fragmentos não fundidos dos polipéptidos FGE, úteis, por exemplo, na preparação de anticorpos, imunoenaios ou aplicações terapêuticas. Fragmentos únicos também podem ser utilizados como moléculas antissentido para inibir a expressão de ácidos nucleicos e polipéptidos FGE, respetivamente.

Como será reconhecido pelos profissionais, a dimensão do fragmento único dependerá da sua conservação no código genético. Assim, algumas regiões da SEQ ID NO: 1 ou SEQ ID NO: 3 e complementos irão requerer que fragmentos maiores sejam únicos, ao passo que outras irão requerer apenas segmentos pequenos, tipicamente entre 12 e 32 nucleótidos de comprimento (por exemplo, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31 e 32 bases) ou mais, até ao comprimento total da sequência

divulgada. Como mencionado acima, pretende-se que esta divulgação abranja cada um e todos os fragmentos de cada sequência, começando no primeiro nucleótido, o segundo nucleótido e assim por diante até 8 nucleótidos de distância da extremidade, e terminando em qualquer parte desde o nucleótido número 8, 9, 10 e assim por diante para cada sequência, até ao último nucleótido (desde que a sequência seja única, como descrito acima). Será único virtualmente qualquer segmento da região da SEQ ID NO: 1 começando no nucleótido i e terminando no nucleótido 1180, ou SEQ ID NO: 3 começando no nucleótido 1 e terminando no nucleótido 1122, ou respectivos complementos, que tenha 20 ou mais nucleótidos de comprimento. Os profissionais estão bem versados em métodos de seleção dessas sequências, tipicamente com base na capacidade do fragmento único para distinguir seletivamente a sequência de interesse de outras sequências do genoma humano, do fragmento com aquelas em bases de dados conhecidas é tipicamente tudo o que é necessário, apesar de poder ser realizada análise de hibridização e sequenciação de confirmação *in vitro*.

Como mencionado acima, oligonucleótidos antissentido que se ligam seletivamente a uma molécula de ácido nucleico que codifica um polipéptido FGE podem ser utilizados para diminuir a atividade da FGE.

Como usado aqui, o termo "oligonucleótido antissentido" ou "antissentido" descreve um oligonucleótido que é um oligorribonucleótido, oligodesoxirribonucleótido, oligorribonucleótido modificado ou oligodesoxirribonucleótido modificado que hibridiza, em condições fisiológicas, com DNA compreendendo um gene particular ou com um transcrito de mRNA desse gene e, desse

modo, inibe a transcrição desse gene e/ou a tradução desse mRNA. As moléculas antissentido são concebidas de modo a interferirem na transcrição ou tradução de um gene alvo por hibridização com o gene alvo ou transcrito. Os profissionais reconhecerão que o comprimento exato do oligonucleótido antissentido e o seu grau de complementaridade com o seu alvo dependerão do alvo específico selecionado, incluindo a sequência do alvo e as bases particulares que compreendem essa sequência. É preferido que o oligonucleótido antissentido seja construído e disposto de modo a ligar-se seletivamente ao alvo em condições fisiológicas, isto é, de modo a hibridizar substancialmente mais com a sequência alvo do que com qualquer outra sequência na célula alvo, em condições fisiológicas. Com base na SEQ ID NO: 1 ou em sequências genômicas e/ou de cDNA alélicas ou homólogas, o profissional pode facilmente escolher e sintetizar quaisquer de algumas moléculas antissentido apropriadas. De modo a serem suficientemente seletivos e potentes para inibição, esses oligonucleótidos antissentido devem compreender pelo menos 10 e, mais preferivelmente, pelo menos 15 bases consecutivas que são complementares ao alvo, apesar de, em certos casos, oligonucleótidos modificados tão pequenos quanto com 7 bases de comprimento tenham sido usados com êxito como oligonucleótidos antissentido (Wagner *et al.*, *Nat. Med.*, 1995, 1(11): 1116-1118; *Nat. Biotech.*, 1996, 14: 840-844). Muito preferivelmente, os oligonucleótidos antissentido compreendem uma sequência complementar de 20-30 bases. Apesar de poderem ser escolhidos oligonucleótidos que são antissentido para qualquer região do gene ou transcritos de mRNA, em formas de realização preferidas os oligonucleótidos antissentido correspondem a sítios N-terminais ou a montante 5', como

sítios de iniciação da tradução, iniciação da transcrição ou promotores. Além disso, regiões 3' não traduzidas podem ser abordadas seletivamente por oligonucleótidos antissentido. O direcionamento para sítios de processamento de mRNA também tem sido utilizado na área, mas pode ser menos preferido se ocorrer processamento alternativo de mRNA. Adicionalmente, o antissentido é dirigido, preferivelmente, para sítios nos quais não é esperada uma estrutura secundária de mRNA (ver, por exemplo, Sainio *et al.*, *Cell Mol. Neurobiol.* 14(5): 439-457, 1994) e aos quais não é esperada ligação de proteínas. Por fim, não obstante a SEQ ID NO: 1 divulgar uma sequência de cDNA, o profissional pode facilmente derivar o DNA genómico correspondente a esta sequência. Assim, podem ser proporcionados oligonucleótidos antissentido que são complementares ao DNA genómico correspondente à SEQ ID NO: 1. De modo semelhante, antissentido para cDNAs de FGE alélicos ou homólogos e DNAs genómicos são permitidos sem experimentação desnecessária.

Num conjunto de formas de realização, os oligonucleótidos antissentido podem ser compostos por desoxirribonucleótidos, ribonucleótidos "naturais" ou qualquer combinação destes. Isto é, a extremidade 5' de um nucleótido nativo e a extremidade 3' de outro nucleótido nativo podem ser ligadas de modo covalente, tal como em sistemas naturais, via uma ligação internucleosídica fosfodiéster. Estes oligonucleótidos podem ser preparados por métodos reconhecidos na área, que podem ser implementados manualmente ou por intermédio de um sintetizador automático. Também podem ser produzidos de modo recombinante por vetores.

No entanto, em formas de realização preferidas, os oligonucleótidos antissentido também podem incluir oligonucleótidos "modificados". Isto é, os oligonucleótidos podem ser modificados de alguns modos que não evitam a sua hibridização com o seu alvo, mas que aumentam a sua estabilidade ou direcionamento ou que então aumentam a sua eficácia terapêutica.

O termo "oligonucleótido modificado", como usado aqui, descreve um oligonucleótido no qual (1) pelo menos dois dos seus nucleótidos estão ligados de modo covalente via uma ligação internucleosídica sintética (isto é, uma ligação diferente de uma ligação fosfodiéster entre a extremidade 5' de um nucleótido e a extremidade 3' de outro nucleótido) e/ou (2) um grupo químico normalmente não associado a ácidos nucleicos foi ligado ao oligonucleótido de modo covalente. Ligações internucleosídicas sintéticas preferidas são fosforotioatos, alquilfosfonatos, fosforoditioatos, ésteres de fosfato, alquilfosfonotioatos, fosforamidatos, carbamatos, carbonatos, triésteres de fosfato, acetamidatos, ésteres de carboximetilo e péptidos.

O termo "oligonucleótido modificado" também abrange oligonucleótidos com uma base e/ou açúcar modificado de modo covalente. Por exemplo, oligonucleótidos modificados incluem oligonucleótidos possuindo açúcares na cadeia principal que estão ligados de modo covalente a grupos orgânicos de baixo peso molecular, diferentes de um grupo hidroxilo na posição 3' e diferentes de um grupo fosfato na posição 5'. Assim, oligonucleótidos modificados podem incluir um grupo ribose 2'-O-alquilado. Adicionalmente, oligonucleótidos modificados podem incluir açúcares como arabinose em vez de ribose. Assim, são contempladas

preparações farmacêuticas contendo moléculas antissentido modificadas que são complementares e estão aptas a hibridizar, em condições fisiológicas, com ácidos nucleicos que codificam polipéptidos FGE, juntamente com transportadores farmacêuticamente aceitáveis. Oligonucleótidos antissentido podem ser administrados como parte de uma composição farmacêutica. Essa composição farmacêutica pode incluir os oligonucleótidos antissentido em combinação com quaisquer transportadores comuns fisiológica e/ou farmacêuticamente aceitáveis que são conhecidos na área. As composições devem ser esterilizadas e conter uma quantidade terapeuticamente eficaz dos oligonucleótidos antissentido numa unidade de peso ou volume adequada para administração a um paciente. O termo "farmacêuticamente aceitável" significa um material não tóxico que não interfere na eficácia da atividade biológica dos ingredientes ativos. O termo "fisiologicamente aceitável" refere-se a um material não tóxico que é compatível com um sistema biológico, como uma célula, cultura de células, tecido ou organismo. As características do transportador dependerão da via de administração. Transportadores fisiológica e farmacêuticamente aceitáveis incluem diluentes, agentes de enchimento, sais, tampões, estabilizadores, solubilizadores e outros materiais que são bem conhecidos na área.

Também são proporcionados métodos para aumentar a atividade geradora de C_{α} -formilglicina numa célula. Em formas de realização importantes, isto é conseguido através da utilização de vetores ("vetores de expressão" e/ou "vetores de direcionamento").

"Vetores", como usado aqui, podem ser quaisquer de alguns ácidos nucleicos nos quais uma sequência desejada pode ser inserida por restrição e ligação, para transporte entre diferentes ambientes genéticos ou para expressão numa célula hospedeiro. Vetores são tipicamente compostos por DNA, apesar de também estarem disponíveis vetores de RNA. Vetores incluem mas não estão limitados a plasmídeos, fagemídeos e genomas de vírus. Um vetor de clonagem é tal que é capaz de se replicar numa célula hospedeiro e que é adicionalmente caracterizado por um ou mais sítios de endonucleases de restrição, nos quais o vetor pode ser cortado de um modo determinável e aos quais pode ser ligada uma sequência de DNA desejada de modo que o novo vetor recombinante retenha a sua capacidade de se replicar na célula hospedeiro. No caso de plasmídeos, a replicação da sequência desejada pode ocorrer muitas vezes, à medida que o número de cópias do plasmídeo aumenta na bactéria hospedeira, ou apenas uma única vez por hospedeiro antes de o hospedeiro se reproduzir por mitose. No caso de fagos, a replicação pode ocorrer ativamente durante uma fase de lise, ou passivamente durante uma fase lisogénica. Um "vetor de expressão" é um vetor onde pode ser inserida uma sequência de DNA desejada (por exemplo, o cDNA de FGE da SEQ ID NO: 3) por restrição e ligação, de modo que fica operativamente ligada a sequências reguladoras e pode ser expressa como um transcrito de RNA. Vetores também podem conter uma ou mais sequências marcadoras adequadas para utilização na identificação de células que foram ou não foram transformadas ou transfetadas com o vetor. Marcadores incluem, por exemplo, genes que codificam proteínas que aumentam ou diminuem a resistência ou sensibilidade a antibióticos ou outros compostos, genes que codificam enzimas cujas atividades são detetáveis por ensaios comuns

conhecidos na área (por exemplo, β -galactosidase ou fosfatase alcalina) e genes que afetam visivelmente o fenótipo de células transformadas ou transfetadas, hospedeiros, colônias ou placas (por exemplo, proteína verde fluorescente).

Um "vetor de direcionamento" é tal que tipicamente contém construções/sequências de direcionamento que são usadas, por exemplo, para inserir uma sequência reguladora num gene endógeno (por exemplo, nas sequências de um exão e/ou intrão), nas sequências promotoras do gene endógeno ou a montante das sequências promotoras do gene endógeno. Noutro exemplo, um vetor de direcionamento pode conter o gene de interesse (por exemplo, codificado pelo cDNA da SEQ ID NO: 1) e outras sequências necessárias para dirigir o gene para uma localização preferida no genoma (por exemplo, uma localização ativa quanto à transcrição, por exemplo, a jusante de um promotor endógeno de um gene não relacionado). A produção de construções e vetores de direcionamento é descrita em pormenor nas Patentes U.S. 5,641,670 e 6,270,989.

Podem ser utilizadas virtualmente quaisquer células, procarióticas ou eucarióticas, que possam ser transformadas com DNA ou RNA heterólogo e que possam crescer ou ser mantidas em cultura. Exemplos incluem células bacterianas, como *Escherichia coli*, células de inseto e células de mamífero, como humanas, de ratinho, hamster, porco, cabra, primata, etc. Podem ser estirpes celulares primárias ou secundárias (que exibem um número finito de duplicações da população média em cultura e não são imortalizadas) e linhas de células imortalizadas (que exibem um tempo de vida aparentemente ilimitado em cultura). Células primárias

e secundárias incluem, por exemplo, fibroblastos, queratinócitos, células epiteliais (por exemplo, células epiteliais mamárias, células epiteliais intestinais), células endoteliais, células da glia, células neurais, elementos formados do sangue (por exemplo, linfócitos, células da medula óssea), células musculares e precursores destes tipos de células somáticas, incluindo células estaminais embrionárias. Quando as células forem destinadas a serem utilizadas em terapia genética, células primárias são preferivelmente obtidas do indivíduo ao qual são administradas as células manipuladas. No entanto, células primárias podem ser obtidas de um dador (diferente do recetor) da mesma espécie. Exemplos de linhas de células humanas imortalizadas que podem ser utilizadas com as construções de DNA e métodos descritos aqui incluem, mas não estão limitadas a células HT-1080 (ATCC CCL 121), células HeLa e derivados de células HeLa (ATCC CCL 2, 2.1 e 2.2), células de cancro da mama MCF-7 (ATCC BTH 22), células de leucemia K-562 (ATCC CCL 243), células de carcinoma KB (ATCC CCL 17), células de carcinoma do ovário 2780AD (Van der Blick, A. M. *et al.*, *Cancer Res*, 48: 5927-5932 (1988), células Raji (ATCC CCL 86), células de adenocarcinoma do cólon WiDr (ATCC CCL 218), células de adenocarcinoma do cólon SW620 (ATCC CCL 227), células Jurkat (ATCC TIB 152), células Namalwa (ATCC CRL 1432), células HL-60 (ATCC CCL 240), células Daudi (ATCC CCL 213), células RPMI 8226 (ATCC CCL 155), células U-937 (ATCC CRL 1593), células de Melanoma de Bowes (ATCC CRL 9607), células WI-38VA13 sublinha 2R4 (ATCC CLL 75.1), e células MOLT-4 (ATCC CRL 1582), células CHO, e células COS, bem como células de hetero-hibridoma produzidas por fusão de células humanas e células de outra espécie. Também podem ser utilizadas estirpes de fibroblastos secundários

humanos, como WI-38 (ATCC CCL 75) e MRC-5 (ATCC CCL 171). Discussões adicionais dos tipos de células que podem ser utilizados são descritas nas Patentes U.S. 5,641,670 e 6,270,989. Em vez de células também podem ser utilizados sistemas de transcrição sem células.

As células são mantidas em condições, como é conhecido na área, que originam expressão da proteína FGE, ou respectivos fragmentos funcionais. Proteínas expressas utilizando os métodos descritos podem ser purificadas de lisados celulares ou sobrenadantes celulares. As proteínas produzidas de acordo com este método podem ser preparadas na forma de uma formulação farmacêuticamente útil e distribuídas a um humano ou animal não humano por vias farmacêuticas convencionais, como é conhecido na área (por exemplo, oral, intravenosa, intramuscular, intranasal, intratraqueal ou subcutânea). Como descrito aqui noutro local, as células recombinantes podem ser células imortalizadas, primárias ou secundárias, preferivelmente humanas. A utilização de células de outra espécie pode ser desejável em casos em que as células não humanas são vantajosas para a finalidade de produção de proteínas onde a FGE não humana produzida é terapeuticamente útil.

Como usado aqui, é afirmado que uma sequência codificadora e sequências reguladoras estão "operativamente" ligadas quando estão ligadas de forma covalente de tal modo que colocam a expressão ou transcrição da sequência codificadora sob a influência ou controlo das sequências reguladoras. Se for desejado que as sequências codificadoras sejam traduzidas numa proteína funcional, é afirmado que duas sequências de DNA estão operativamente ligadas se a indução de um promotor nas sequências

reguladoras 5' originar a transcrição da sequência codificadora, e se a natureza da ligação entre as duas sequências de DNA não (1) originar a introdução de uma mutação por deslocamento do quadro de leitura, (2) interferir na capacidade da região promotora para dirigir a transcrição das sequências codificadoras, ou (3) interferir na capacidade do transcrito de RNA correspondente para ser traduzido numa proteína. Assim, uma região promotora estará operativamente ligada a uma sequência codificadora se a região promotora for capaz de efetuar a transcrição dessa sequência de DNA de modo que o transcrito resultante possa ser traduzido na proteína ou polipéptido desejado.

A natureza precisa das sequências reguladoras necessárias para a expressão genética pode variar entre espécies ou tipos de células, mas em geral incluirá, consoante o necessário, sequências 5' não transcritas e 5' não traduzidas envolvidas na iniciação da transcrição e tradução, respetivamente, como uma caixa TATA, sequência de capeamento, sequência CAAT, e afins. Especialmente, essas sequências reguladoras 5' não transcritas incluirão uma região promotora que inclui uma sequência promotora para controlo da transcrição do gene operativamente ligado. Sequências reguladoras também podem incluir sequências intensificadoras ou sequências ativadoras a montante, consoante o desejado. Os vetores podem opcionalmente incluir sequências 5' líder ou de sinal. A escolha e conceção de um vetor apropriado pertencem ao âmbito e critério do profissional.

Vetores de expressão contendo todos os elementos necessários para expressão estão disponíveis no mercado e são conhecidos dos profissionais. Ver, por exemplo,

Sambrook et al., "Molecular Cloning: A Laboratory Manual", Segunda Edição, Cold Spring Harbor Laboratory Press, 1989. Células são geneticamente manipuladas pela introdução nas células de DNA (RNA) heterólogo que codifica o polipéptido FGE, ou respetivo fragmento ou variante. Esse DNA (RNA) heterólogo é colocado sob o controlo operável de elementos de transcrição, para permitir a expressão do DNA heterólogo na célula hospedeiro.

Sistemas preferidos para expressão de mRNA em células de mamífero são aqueles como pRc/CMV (disponibilizado pela Invitrogen, Carlsbad, CA) que contém um marcador selecionável, como um gene que confere resistência a G418 (que facilita a seleção de linhas de células transfetadas de modo estável) e as sequências intensificadoras-promotoras do citomegalovírus (CMV) humano. Adicionalmente, para expressão em linhas de células de primata ou canino é adequado o vetor pCEP4 (Invitrogen, Carlsbad, CA), que contém uma origem da replicação do vírus Epstein Barr (EBV), que facilita a manutenção do plasmídeo como um elemento extracromossómico de múltiplas cópias. Outro vetor de expressão é o plasmídeo pEF-BOS, que contém o promotor do Fator de Alongamento de polipéptido 1 α , que estimula uma transcrição eficiente *in vitro*. O plasmídeo é descrito por Mishizuma e Nagata (*Nuc. Acids Res.* 18: 5322, 1990), e a sua utilização em experiências de transfeção é divulgada, por exemplo, por Demoulin (*Mol. Cell. Biol.* 16: 4710-4716, 1996). Ainda outro vetor de expressão preferido é um adenovírus, descrito por Stratford-Perricaudet, que é deficiente para as proteínas E1 E3 (*J. Clin. Invest.* 90: 626-630, 1992). A utilização do adenovírus como um Adeno.P1A recombinante é divulgada por Warnier et al., na

injeção intradérmica em ratinhos para imunização contra PlA (*Int. J. Cancer*, 67: 303-310, 1996).

Podem ser proporcionados estojos de expressão que permitem ao técnico preparar um vetor ou vetores de expressão desejados. Esses estojos de expressão incluem pelo menos porções separadas de cada uma das sequências codificadoras previamente descritas. Podem ser adicionados outros componentes, consoante o desejado, desde que estejam incluídas as sequências previamente mencionadas, que são requeridas.

Também será reconhecido que os vetores de expressão acima descritos contendo a sequência de cDNA de FGE podem ser utilizados para transfetar células hospedeiro e linhas de células, sejam procarióticas (por exemplo, *Escherichia coli*) ou eucarióticas (por exemplo, células CHO, células COS, sistemas de expressão de levedura e expressão de baculovírus recombinante em células de inseto). São especialmente úteis células de mamífero, como humanas, de ratinho, hamster, porco, cabra, primata, etc. Podem provir de uma grande variedade de tipos de tecidos, e incluem células primárias e linhas de células imortalizadas, como descrito aqui noutro local. Exemplos específicos incluem células HT-1080, células CHO, células dendríticas, células U293, leucócitos do sangue periférico, células estaminais da medula óssea, células estaminais embrionárias e células de inseto. A construção de "knock-outs" do gene FGE em células e em animais é útil para proporcionar materiais para estudar certos aspetos da atividade da FGE.

Podem ser proporcionados polipéptidos isolados (incluindo proteínas completas e proteínas parciais) codificados pelos

ácidos nucleicos FGE anteriores, e incluem o polipéptido da SEQ ID NO: 2 e respectivos fragmentos únicos. Esses polipéptidos são úteis, por exemplo, isolados ou como parte de proteínas de fusão para gerar anticorpos, como componentes de um imunoensaio, etc. Polipéptidos podem ser isolados de amostras biológicas, incluindo homogenatos teciduais ou celulares, e também podem ser expressos de modo recombinante numa variedade de sistemas de expressão procarióticos e eucarióticos por construção de um vetor de expressão apropriado para o sistema de expressão, introdução do vetor de expressão no sistema de expressão e isolamento da proteína expressa de modo recombinante. Polipéptidos curtos, incluindo péptidos antigénicos (como os apresentados por moléculas MHC na superfície de uma célula para reconhecimento imunológico) também podem ser sintetizados quimicamente utilizando métodos bem estabelecidos de síntese de péptidos.

Um fragmento único de um polipéptido FGE, em geral, tem os aspetos e características de fragmentos únicos, como discutido acima relacionado com ácidos nucleicos. Como será reconhecido pelos profissionais, a dimensão do fragmento único dependerá de fatores tais como se o fragmento constitui uma porção de um domínio proteico conservado. Assim, algumas regiões da SEQ ID NO: 2 irão requerer que segmentos maiores sejam únicos, ao passo que outras irão requerer apenas segmentos pequenos, tipicamente possuindo entre 5 e 12 aminoácidos (por exemplo, 5, 6, 7, 8, 9, 10, 11 e 12 aminoácidos de comprimento ou mais, incluindo cada inteiro até ao comprimento completo, 287 aminoácidos de comprimento).

Fragmentos únicos de um polipéptido são preferivelmente aqueles fragmentos que retêm uma capacidade funcional distinta do polipéptido. Capacidades funcionais que podem ser retidas num fragmento único de um polipéptido incluem interação com anticorpos, interação com outros polipéptidos ou respetivos fragmentos, interação com outras moléculas, etc. Uma atividade importante é a capacidade para atuar como assinatura para identificar o polipéptido. Os profissionais são bem versados em métodos para selecionar sequências de aminoácidos únicas, tipicamente com base na capacidade do fragmento único para distinguir seletivamente a sequência de interesse de membros não pertencentes à família. Tipicamente, tudo o que é necessário é uma comparação da sequência do fragmento com as de bases de dados conhecidas.

São úteis variantes dos polipéptidos FGE descritos acima. Como usado aqui, uma "variante" de um polipéptido FGE é um polipéptido que contém uma ou mais modificações na sequência primária de aminoácidos de um polipéptido FGE. Modificações que criam uma variante de um polipéptido FGE são tipicamente feitas no ácido nucleico que codifica o polipéptido FGE, e podem incluir deleções, mutações pontuais, truncamentos, substituições de aminoácidos e adição de aminoácidos ou de frações diferentes de aminoácidos, com a finalidade de: 1) reduzir ou eliminar uma atividade de um polipéptido FGE; 2) intensificar uma propriedade de um polipéptido FGE, como estabilidade proteica num sistema de expressão ou a estabilidade da ligação proteína-ligando; 3) dotar um polipéptido FGE de uma atividade ou propriedade nova, como adição de um epítopo antigénico ou adição de uma fração detetável; ou 4) proporcionar ligação equivalente ou melhor a um recetor de

um polipéptido FGE ou outra molécula. Alternativamente, podem ser feitas modificações diretamente no polipéptido, tais como por clivagem, adição de uma molécula de ligação, adição de uma fração detetável, como biotina, adição de um ácido gordo, e afins. Modificações também abrangem proteínas de fusão compreendendo a totalidade ou parte da sequência de aminoácidos de FGE. O profissional estará familiarizado com métodos para prever o efeito, na conformação proteica, de uma alteração da sequência proteica e, assim, pode "conceber" um polipéptido FGE variante de acordo com métodos conhecidos. Um exemplo de um desses métodos é descrito por Dahiyal e Mayo em *Science* 278: 82-87, 1997, de modo que proteínas podem ser concebidas de novo. O método pode ser aplicado a uma proteína conhecida para variar apenas uma porção da sequência polipeptídica. Por aplicação dos métodos computacionais de Dahiyal e Mayo, variantes específicas do polipéptido FGE podem ser propostas e testadas para determinar se a variante retém uma conformação desejada.

Variantes podem incluir polipéptidos FGE que são modificados especificamente para alterar uma característica do polipéptido que não está relacionada com a sua atividade fisiológica. Por exemplo, resíduos cisteína podem ser substituídos ou deletados, para prevenir ligações dissulfureto indesejadas. De modo semelhante, certos aminoácidos podem ser alterados para intensificar a expressão de um polipéptido FGE por eliminação de proteólise efetuada por proteases num sistema de expressão (por exemplo, resíduos de aminoácidos dibásicos em sistemas de expressão de levedura onde está presente atividade de protease KEX2).

Mutações de um ácido nucleico que codifica um polipéptido FGE preferivelmente preservam a fase de leitura de aminoácidos da sequência codificadora, e preferivelmente não criam regiões no ácido nucleico com probabilidade de hibridizarem para formar estruturas secundárias, como grampos ou alças, que podem ser prejudiciais para a expressão do polipéptido variante.

Podem ser efetuadas mutações por seleção de uma substituição de aminoácidos, ou por mutagénese aleatória de um sítio selecionado num ácido nucleico que codifica o polipéptido. Polipéptidos variantes são então expressos e testados quanto a uma ou mais atividades para determinar qual a mutação que dota um polipéptido variante das propriedades desejadas. Podem ser realizadas mutações adicionais em variantes (ou em polipéptidos FGE não variantes) que são silenciosas quanto à sequência de aminoácidos do polipéptido, mas que proporcionam códons preferidos para tradução num hospedeiro particular, ou alteram a estrutura do mRNA, por exemplo, para aumentar a estabilidade e/ou expressão. Os códons preferidos para tradução de um ácido nucleico, por exemplo, em *Escherichia coli*, células de mamífero, etc., são bem conhecidos dos profissionais. Ainda outras mutações podem ser feitas nas sequências não codificadoras de um gene FGE ou clone de cDNA para aumentar a expressão do polipéptido.

O profissional estará ciente de que podem ser feitas substituições de aminoácidos conservativas em polipéptidos FGE para proporcionar variantes funcionalmente equivalentes dos polipéptidos anteriores, isto é, as variantes retêm as capacidades funcionais dos polipéptidos FGE. Como usado aqui, uma "substituição de aminoácidos conservativa"

refere-se a uma substituição de aminoácidos que não altera significativamente a estrutura terciária e/ou atividade do polipéptido. Variantes podem ser preparadas de acordo com métodos de alteração de sequências polipeptídicas conhecidos dos profissionais, e incluem aqueles que se encontram em referências que compilam esses métodos, por exemplo, "Molecular Cloning: A Laboratory Manual", J. Sambrook, et al., editores, Segunda Edição, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, Nova Iorque, 1989, ou "Current Protocols in Molecular Biology", F.M. Ausubel, et al., editores, John Wiley & Sons, Inc., Nova Iorque. Variantes funcionalmente equivalentes exemplificativas dos polipéptidos FGE incluem substituições de aminoácidos conservativas da SEQ ID NO: 2. Substituições conservativas de aminoácidos incluem substituições feitas entre aminoácidos pertencentes aos grupos seguintes: (a) M, I, L, V; (b) F, Y, W; (c) K, R, H; (d) A, G; (e) S, T; (f) Q, N; e (g) E, D.

Assim, são contempladas variantes funcionalmente equivalentes de polipéptidos FGE, isto é, variantes de polipéptidos FGE que retêm a função dos polipéptidos FGE naturais. Substituições de aminoácidos conservativas na sequência de aminoácidos de polipéptidos FGE para produzir variantes funcionalmente equivalentes de polipéptidos FGE são tipicamente feitas por alteração de um ácido nucleico que codifica polipéptidos FGE (SEQ ID NOs: 1, 3). Essas substituições podem ser feitas por uma variedade de métodos conhecidos do profissional. Por exemplo, substituições de aminoácidos podem ser feitas por mutação dirigida por PCR, mutagênese dirigida a sítios de acordo com o método de Kunkel (Kunkel, *Proc. Nat. Acad. Sci. U.S.A.* 82: 488-492, 1985), ou por síntese química de um gene que codifica um

polipéptido FGE. A atividade de fragmentos funcionalmente equivalentes de polipéptidos FGE pode ser testada por clonagem do gene que codifica o polipéptido FGE alterado num vetor de expressão bacteriano ou de mamífero, introdução do vetor numa célula hospedeiro apropriada, expressão do polipéptido FGE alterado e teste quanto a uma capacidade funcional dos polipéptidos FGE, como divulgado aqui (por exemplo, atividade geradora de C_α -formilglicina, etc.).

A informação proporcionada aqui tem algumas utilizações, algumas das quais são descritas aqui noutro local. Em primeiro lugar, permite o isolamento de polipéptidos FGE. Pode ser utilizada uma variedade de metodologias bem conhecidas do profissional para se obterem moléculas FGE isoladas. O polipéptido pode ser purificado a partir de células que produzem naturalmente o polipéptido por meios cromatográficos ou reconhecimento imunológico. Alternativamente, um vetor de expressão pode ser introduzido em células para originar a produção do polipéptido. Noutro método, transcritos de mRNA podem ser microinjetados ou introduzidos em células de qualquer outra forma para causar a produção do polipéptido codificado. Também pode ser utilizada tradução de mRNA de FGE em extratos sem células, como o sistema de lisados de reticulócitos, para produzir polipéptidos FGE. Os profissionais também podem facilmente seguir métodos conhecidos para o isolamento de polipéptidos FGE. Estes incluem mas não estão limitados a imunocromatografia, HPLC, cromatografia de exclusão por tamanhos, cromatografia de permuta iónica e cromatografia de imunoafinidade.

Em certas formas de realização podem ser proporcionados polipéptidos "dominantes negativos", derivados de polipéptidos FGE. Um polipéptido dominante negativo é uma variante inativa de uma proteína que, ao interagir com a maquinaria celular, desloca uma proteína ativa da sua interação com a maquinaria celular ou compete com a proteína ativa, desse modo reduzindo o efeito da proteína ativa. Por exemplo, um recetor dominante negativo que se liga a um ligando mas que não transmite um sinal em resposta à ligação do ligando pode reduzir o efeito biológico da expressão do ligando. De igual modo, uma quinase dominante negativa cataliticamente inativa que interage normalmente com proteínas alvo mas que não fosforila as proteínas alvo pode reduzir a fosforilação das proteínas alvo em resposta a um sinal celular. De modo semelhante, um fator de transcrição dominante negativo que se liga a um sítio promotor na região de controlo de um gene mas que não aumenta a transcrição do gene pode reduzir o efeito de um fator de transcrição normal por ocupação de sítios de ligação promotores sem aumentar a transcrição.

O resultado final da expressão de um polipéptido dominante negativo numa célula é uma redução da função de proteínas ativas. O profissional pode avaliar o potencial de uma variante dominante negativa de uma proteína e utilizar técnicas comuns de mutagénese para criar um ou mais polipéptidos variantes dominantes negativos. Ver, por exemplo, Patente U.S. Nº 5,580,723 e Sambrook *et al.*, "Molecular Cloning: A Laboratory Manual", Segunda Edição, Cold Spring Harbor Laboratory Press, 1989. O profissional pode então testar a população de polipéptidos que sofreram mutagénese quanto a diminuição de uma atividade selecionada e/ou quanto a retenção dessa atividade. Outros métodos

semelhantes para criar e testar variantes dominantes negativas de uma proteína serão claros para o profissional.

O isolamento de cDNA de FGE também possibilita ao técnico diagnosticar uma perturbação caracterizada por uma expressão aberrante de FGE. Estes métodos envolvem determinar a expressão do gene FGE e/ou de polipéptidos FGE daí derivados. Na primeira situação, essas determinações podem ser efetuadas via qualquer ensaio padrão de determinação de ácidos nucleicos, incluindo a reação em cadeia de polimerase, ou avaliação com sondas de hibridização etiquetadas, como exemplificado em baixo. Na última situação, essa determinação pode ser efetuada via qualquer ensaio imunológico padrão utilizando, por exemplo, anticorpos que se ligam à proteína FGE segregada. Uma perturbação preferida que pode ser diagnosticada é Deficiência de Múltiplas Sulfatases.

São úteis agentes de ligação de péptidos isolados que, por exemplo, podem ser anticorpos ou fragmentos de anticorpos ("polipéptidos de ligação") com capacidade para se ligarem seletivamente a polipéptidos FGE. Anticorpos incluem anticorpos policlonais e monoclonais, preparados de acordo com metodologia convencional. Em certas formas de realização estão excluídos agentes de ligação (por exemplo, anticorpos) que se ligam aos polipéptidos codificados pelos ácidos nucleicos da SEQ ID NO: 4.

De modo significativo, como é bem conhecido na área, apenas uma pequena porção de uma molécula de anticorpo, o paratopo, está envolvida na ligação do anticorpo ao seu epítopo (ver, em geral, Clark, W.R. (1986) "The Experimental Foundations of Modern Immunology" Wiley &

Sons, Inc., Nova Iorque; Roitt, L (1991) "Essential Immunology", 7ª Edição, Blackwell Scientific Publications, Oxford). As regiões pFc' e Fc, por exemplo, são efetores da cascata do complemento mas não estão envolvidas na ligação de antígenos. Um anticorpo de onde a região pFc' foi clivada de modo enzimático, ou que foi produzido sem a região pFc', designado fragmento F(ab')₂, retém os sítios de ligação de antígenos de um anticorpo intacto. De modo semelhante, um anticorpo de onde a região Fc foi clivada de modo enzimático, ou que foi produzido sem a região Fc, designado fragmento Fab, retém um dos sítios de ligação de antígenos de uma molécula de anticorpo intacto. Prosseguindo, fragmentos Fab consistem numa cadeia leve de anticorpo ligada de modo covalente e uma porção da cadeia pesada de anticorpo designada Fd. Os fragmentos Fd são os determinantes principais da especificidade de anticorpos (um único fragmento Fd pode estar associado a até dez cadeias leves diferentes sem alterar a especificidade do anticorpo) e fragmentos Fd retêm a capacidade de ligação de epítopos por isolamento.

Na porção de ligação de antígenos de um anticorpo, como é bem conhecido na área, há regiões determinantes de complementaridade (CDRs) que interagem diretamente com o epítipo do antígeno, e regiões de arcabouço (FRs) que mantêm a estrutura terciária do paratopo (ver, em geral, Clark, 1986; Roitt, 1991). No fragmento Fd da cadeia pesada e na cadeia leve de imunoglobulinas IgG há quatro regiões de arcabouço (FR1 até FR4) separadas, respetivamente, por três regiões determinantes de complementaridade (CDR1 até CDR3). As CDRs, e em particular as regiões CDR3, e mais particularmente a CDR3 da cadeia pesada, são

maioritariamente responsáveis pela especificidade dos anticorpos.

Está agora bem estabelecido na área que as regiões não CDR de um anticorpo de mamífero podem ser substituídas por regiões semelhantes de anticorpos conspecíficos ou heteroespecíficos mas retendo a especificidade epitópica do anticorpo original. Isto é muito claramente manifestado no desenvolvimento e utilização de anticorpos "humanizados", nos quais CDRs não humanas são ligadas de modo covalente a regiões FR e/ou Fc/pFc' humanas para produzir um anticorpo funcional. Ver, por exemplo, patentes U.S. 4,816,567, 5,225,539, 5,585,089, 5,693,762 e 5,859,205. Assim, por exemplo, a Publicação Internacional PCT Número WO 92/04381 ensina a produção e utilização de anticorpos RSV murinos humanizados nos quais pelo menos uma porção das regiões FR murinas foi substituída por regiões FR de origem humana. Esses anticorpos, incluindo fragmentos de anticorpos intactos com capacidade de ligação de antígenos, são muitas vezes chamados de anticorpos "quiméricos".

Assim, como será claro para o profissional, podem ser proporcionados os seguintes: fragmentos $F(ab')_2$, Fab, Fv e Fd; anticorpos quiméricos nos quais as regiões Fc e/ou FR e/ou CDR1 e/ou CDR2 e/ou CDR3 de cadeia leve foram substituídas por sequências humanas ou não humanas homólogas; anticorpos de fragmentos $F(ab')_2$ quiméricos nos quais as regiões FR e/ou CDR1 e/ou CDR2 e/ou CDR3 de cadeia leve foram substituídas por sequências humanas ou não humanas homólogas; anticorpos de fragmentos Fab quiméricos nos quais as regiões FR e/ou CDR1 e/ou CDR2 e/ou CDR3 de cadeia leve foram substituídas por sequências humanas ou não humanas homólogas; e anticorpos de fragmentos Fd

quiméricos nos quais as regiões FR e/ou CDR1 e/ou CDR2 foram substituídas por sequências humanas ou não humanas homólogas. Também podem ser proporcionados os denominados anticorpos de cadeia simples.

Assim, podem ser proporcionados polipéptidos de numerosas dimensões e tipos que se ligam especificamente a polipéptidos FGE, e complexos de polipéptidos FGE e seus parceiros de ligação. Estes polipéptidos também podem ser derivados de fontes diferentes de tecnologia de anticorpos. Por exemplo, esses agentes de ligação de polipéptidos podem ser proporcionados por bibliotecas de péptidos degenerados que podem ser facilmente preparadas em solução, em forma imobilizada, na forma de bibliotecas de expressão de péptidos derivados de flagelos bacterianos ou na forma de bibliotecas de expressão em fagos. Também podem ser sintetizadas bibliotecas combinatoriais de péptidos contendo um ou mais aminoácidos. Também podem ser sintetizadas bibliotecas de péptidos e frações sintéticas não peptídicas.

A expressão em fagos pode ser particularmente eficaz na identificação de péptidos de ligação úteis. Em resumo, prepara-se uma biblioteca de fagos (utilizando, por exemplo, m13, fd ou o fago lambda) expressando inserções com 4 até cerca de 80 resíduos de aminoácidos utilizando procedimentos convencionais. As inserções podem representar, por exemplo, uma série completamente degenerada ou distorcida. Em seguida, é possível selecionar fagos contendo inserções que se ligam ao polipéptido FGE ou a um complexo de FGE e um parceiro de ligação. Este processo pode ser repetido ao longo de vários ciclos de nova seleção de fagos que ligam ao polipéptido ou complexo

FGE. Rondas repetidas conduzem ao enriquecimento de fagos contendo sequências particulares. Pode ser conduzida análise de sequências de DNA para identificar as sequências dos polipéptidos expressos. Pode ser determinada a porção linear mínima da sequência que se liga ao polipéptido ou complexo FGE. É possível repetir o procedimento utilizando uma biblioteca distorcida contendo inserções que contêm parte ou a totalidade da porção linear mínima mais um ou mais resíduos degenerados adicionais a montante ou a jusante. Também podem ser utilizados métodos de rastreamento de dois híbridos de levedura para identificar polipéptidos que se ligam aos polipéptidos FGE. Assim, os polipéptidos FGE, ou respectivo fragmento, ou complexos de FGE e um parceiro de ligação, podem ser utilizados para rastrear bibliotecas de peptídeos, incluindo bibliotecas de expressão em fagos, para identificar e selecionar parceiros de ligação de peptídicos dos polipéptidos FGE. Essas moléculas podem ser utilizadas, como descrito, para ensaios de rastreamento, para protocolos de purificação, para interferir diretamente no funcionamento de FGE e para outras finalidades que serão claras para os profissionais.

Um polipéptido FGE, ou um respectivo fragmento, também pode ser utilizado para isolar os seus parceiros de ligação nativos. O isolamento de parceiros de ligação pode ser realizado de acordo com métodos bem conhecidos. Por exemplo, polipéptidos FGE isolados podem ser ligados a um substrato e depois pode ser aplicada no substrato uma solução suspeita de conter um parceiro de ligação FGE. Se o parceiro de ligação de polipéptidos FGE estiver presente na solução, então irá ligar-se ao polipéptido FGE ligado ao substrato. O parceiro de ligação pode depois ser isolado. Outras proteínas que são parceiros de ligação de FGE podem

ser isoladas por métodos semelhantes sem experimentação desnecessária. Um parceiro de ligação preferido é uma sulfatase.

Podem ser proporcionados métodos de medição do nível de expressão de FGE num sujeito. Isto pode ser efetuado obtendo primeiramente uma amostra de teste do sujeito. A amostra de teste pode ser tecido ou fluído biológico. Tecidos incluem cérebro, coração, soro, mama, cólon, bexiga, útero, próstata, estômago, testículo, ovário, pâncreas, glândula pituitária, glândula suprarrenal, glândula tiroide, glândula salivar, glândula mamária, rim, fígado, intestino, baço, timo, vasos sanguíneos, medula óssea, traqueia e pulmão. Em certas formas de realização, amostras de teste são originárias de tecidos do coração e vasos sanguíneos, e fluidos biológicos incluem sangue, saliva e urina. Podem ser utilizadas técnicas invasivas e não invasivas para se obterem essas amostras, e estão bem documentadas na área. Ao nível molecular podem ser utilizadas PCR e coloração "Northern" para determinar o nível de mRNA de FGE utilizando produtos descritos aqui, e protocolos bem conhecidos na área que se encontram em referências que compilam tais métodos. Ao nível proteico, a expressão de FGE pode ser determinada utilizando soros anti-FGE policlonais ou monoclonais em combinação com ensaios imunológicos comuns. Os métodos preferidos irão comparar o nível medido de expressão de FGE da amostra de teste com um controlo. Um controlo pode incluir uma quantidade conhecida de uma sonda de ácido nucleico, um epítipo de FGE (como um produto de expressão de FGE) ou uma amostra de teste semelhante de um sujeito com um nível de controlo ou "normal" de expressão de FGE.

Polipéptidos FGE são preferivelmente produzidos de modo recombinante, apesar de esses polipéptidos poderem ser isolados de extratos biológicos. Polipéptidos FGE produzidos de modo recombinante incluem proteínas quiméricas compreendendo uma fusão de uma proteína FGE com outro polipéptido, por exemplo, um polipéptido capaz de proporcionar ou intensificar a ligação proteína-proteína, ligação de ácidos nucleicos específicos para sequências (como GAL4), intensificar a estabilidade do polipéptido FGE em condições de ensaio ou proporcionar uma fração detetável, como proteína verde fluorescente. Um polipéptido fundido a um polipéptido FGE ou fragmento também pode proporcionar meios para detetar facilmente a proteína de fusão, por exemplo, por reconhecimento imunológico ou por etiquetagem fluorescente.

A informação proporcionada aqui também é útil na geração de animais transgênicos não humanos. Como usado aqui, "animais transgênicos não humanos" inclui animais não humanos com uma ou mais moléculas de ácidos nucleicos exógenos incorporadas em células da linha germinativa e/ou células somáticas. Assim, os animais transgênicos incluem animais "knockout" possuindo uma fragmentação genética homozigótica ou heterozigótica por recombinação homóloga, animais possuindo vetores de expressão episomais ou cromossomicamente incorporados, etc. Animais "knockout" podem ser preparados por recombinação homóloga utilizando células estaminais embrionárias, como é bem conhecido na área. A recombinação pode ser facilitada utilizando, por exemplo, o sistema cre/lox ou outros sistemas de recombinases conhecidos do profissional. Em certas formas de realização, o próprio sistema de recombinases é expresso condicionalmente, por exemplo, em certos tecidos ou tipos

de células, em certas fases do desenvolvimento embrionárias ou pós-embrionárias, é induzido pela adição de um composto que aumenta ou diminui a expressão, e afins. Em geral, os vetores de expressão condicional utilizados nesses sistemas empregam uma variedade de promotores que conferem o padrão de expressão genética desejado (por exemplo, temporal ou espacial). Promotores condicionais também podem ser operativamente ligados a moléculas de ácidos nucleicos FGE para aumentar a expressão de FGE de um modo regulado ou condicional. Reguladores negativos de atuação trans da atividade ou expressão de FGE também podem ser operativamente ligados a um promotor condicional, como descrito acima. Esses reguladores de atuação trans incluem moléculas de ácidos nucleicos FGE antissentido, moléculas de ácidos nucleicos que codificam moléculas FGE dominantes negativas, moléculas de ribozimas específicas para ácidos nucleicos FGE, e afins. Os animais transgênicos não humanos são úteis em experiências dirigidas ao teste de efeitos bioquímicos ou fisiológicos de diagnósticos ou terapêuticas para estados caracterizados por expressão de FGE aumentada ou decrescida. Outras utilizações serão claras para o profissional.

Também é contemplada terapia genética. O procedimento de realização de terapia genética *ex vivo* é esboçado na Patente U.S. 5,399,346 e em apresentações submetidas no histórico de arquivo dessa patente, em que todas são documentos publicamente disponíveis. Em geral, envolve a introdução *in vitro* de uma cópia funcional de um gene em célula(s) de um sujeito que contém(contêm) uma cópia defeituosa do gene, e repor no sujeito a(s) célula(s) geneticamente manipulada(s). A cópia funcional do gene está sob o controle operável de elementos reguladores que

permitem a expressão do gene na(s) célula(s) geneticamente manipulada(s). Numerosas técnicas de transfeção e transdução, bem como vetores de expressão apropriados, são bem conhecidas dos profissionais, algumas das quais são descritas no requerimento PCT W095/00654. Também é contemplada terapia genética *in vivo* utilizando vetores tais como adenovírus, retrovírus, vírus herpes e lipossomas dirigidos.

Podem ser proporcionados métodos eficientes de identificação de agentes ou compostos guia para agentes ativos ao nível de uma função celular dependente de FGE ou fragmento de FGE. Em particular, essas funções incluem interação com outros polipéptidos ou fragmentos. Em geral, os métodos de rastreio envolvem ensaiar quanto a compostos que interferem na atividade de FGE (como atividade geradora de C_{α} -formilglicina), apesar de compostos que aumentam a atividade geradora de C_{α} -formilglicina também poderem ser avaliados utilizando os métodos de rastreio. Esses métodos são adaptáveis ao rastreio automático e de alto rendimento de compostos. Indicações alvo incluem processos celulares modulados pela FGE, como atividade geradora de C_{α} -formilglicina.

É proporcionada uma grande variedade de ensaios para agentes candidatos (farmacológicos), incluindo ensaios de ligação proteína-ligando *in vitro* etiquetados, ensaios de desvio da mobilidade eletroforética, imunoensaios, ensaios à base de células, como rastreios de dois ou três híbridos, ensaios de expressão, etc. Os ácidos nucleicos transfetados podem codificar, por exemplo, bibliotecas de péptidos combinatoriais ou bibliotecas de cDNA. Reagentes convenientes para esses ensaios, por exemplo, proteínas de

fusão com GAL4, são conhecidos na área. Um ensaio exemplificativo à base de células envolve a transfeção de uma célula com um ácido nucleico que codifica um polipéptido FGE fundido a um domínio de ligação de DNA de GAL4 e um ácido nucleico que codifica um gene repórter operativamente ligado a uma região reguladora da expressão genética, como um ou mais sítios de ligação de GAL4. Ocorre ativação da transcrição do gene repórter quando o FGE e polipéptido de fusão de repórter se ligam de modo a permitir a transcrição do gene repórter. Agentes que modulam uma função mediada por um polipéptido FGE são então detetados por uma alteração da expressão do gene repórter. Métodos de determinação de alterações da expressão de um gene repórter são conhecidos na área.

Fragmentos FGE utilizados nos métodos, quando não produzidos por um ácido nucleico transfetado, são adicionados a uma mistura de ensaio na forma de um polipéptido isolado. Preferivelmente, polipéptidos FGE são produzidos de modo recombinante, não obstante esses polipéptidos poderem ser isolados de extratos biológicos. Polipéptidos FGE produzidos de modo recombinante incluem proteínas quiméricas compreendendo uma fusão de uma proteína FGE com outro polipéptido, por exemplo, um polipéptido capaz de proporcionar ou intensificar a ligação proteína-proteína, ligação de ácidos nucleicos específicos para sequências (como GAL4), intensificar a estabilidade do polipéptido FGE em condições de ensaio ou proporcionar uma fração detetável, como proteína verde fluorescente ou epítopo Flag.

A mistura de ensaio é compreendida por um alvo de ligação de FGE intracelular natural capaz de interagir com a FGE.

Apesar de poderem ser utilizados alvos de ligação de FGE naturais, é frequentemente preferido utilizar porções (por exemplo, péptidos - ver, por exemplo, o péptido da SEQ ID NO: 33 - ou fragmentos de ácidos nucleicos) ou análogos (isto é, agentes que imitam as propriedades de ligação de FGE do alvo de ligação natural para finalidades do ensaio) do alvo de ligação de FGE, desde que a porção ou análogo proporcione afinidade de ligação e avidéz para o fragmento FGE mensuráveis no ensaio.

A mistura de ensaio também compreende um agente candidato. Tipicamente, uma pluralidade de misturas de ensaio é processada em paralelo com diferentes concentrações do agente, para se obter uma resposta diferente às várias concentrações. Tipicamente, uma destas concentrações serve de controlo negativo, isto é, a uma concentração nula do agente ou a uma concentração do agente inferior aos limites de deteção do ensaio. Agentes candidatos abrangem numerosas classes químicas, apesar de tipicamente serem compostos orgânicos. Preferivelmente, os agentes candidatos são pequenos compostos orgânicos, isto é, aqueles que têm um peso molecular superior a 50 mas inferior a cerca de 2500, preferivelmente inferior a cerca de 1000 e mais preferivelmente inferior a cerca de 500. Agentes candidatos compreendem grupos químicos funcionais necessários para interações estruturais com polipéptidos e/ou ácidos nucleicos, e tipicamente incluem pelo menos um grupo amina, carbonilo, hidroxilo ou carboxilo, preferivelmente pelo menos dois dos grupos químicos funcionais e mais preferivelmente pelo menos três dos grupos químicos funcionais. Os agentes candidatos podem compreender uma estrutura carbonada cíclica ou heterocíclica e/ou estruturas aromáticas ou poliaromáticas substituídas com um

ou mais dos grupos funcionais acima identificados. Agentes candidatos também podem ser biomoléculas, como péptidos, sacáridos, ácidos gordos, esteróis, isoprenoides, purinas, pirimidinas, derivados ou análogos estruturais dos acima, ou respectivas combinações e afins. Quando o agente for um ácido nucleico, o agente é tipicamente uma molécula de DNA ou RNA, apesar de também serem contemplados ácidos nucleicos modificados como definidos aqui.

Agentes candidatos são obtidos de uma grande variedade de fontes, incluindo bibliotecas de compostos sintéticos ou naturais. Por exemplo, estão disponíveis numerosos meios para síntese aleatória e dirigida de uma grande variedade de compostos orgânicos e biomoléculas, incluindo expressão de oligonucleótidos aleatórios, bibliotecas combinatoriais orgânicas sintéticas, bibliotecas de expressão em fagos de péptidos aleatórios, e afins. Alternativamente, bibliotecas de compostos naturais na forma de extratos bacterianos, fúngicos, vegetais e animais estão disponíveis ou são facilmente produzidas. Adicionalmente, bibliotecas e compostos naturais e produzidos de modo sintético podem ser modificados por meios químicos, físicos e bioquímicos convencionais. Além disso, agentes (farmacológicos) conhecidos podem ser sujeitos a modificações químicas dirigidas ou aleatórias, como acilação, alquilação, esterificação, amidação, etc., para produzir análogos estruturais dos agentes.

Uma variedade de outros reagentes também pode ser incluída na mistura. Estes incluem reagentes tais como sais, tampões, proteínas neutras (por exemplo, albumina), detergentes, etc., que podem ser utilizados para facilitar uma ligação ótima proteína-proteína e/ou proteína-ácido

nucleico. Esse reagente também pode reduzir interações inespecíficas ou de fundo dos componentes da reação. Também podem ser utilizados outros reagentes que melhoram a eficiência do ensaio, como protease, inibidores, inibidores de nucleases, agentes antimicrobianos e afins.

A mistura dos materiais de ensaio anteriores é incubada em condições de modo que, excetuando quanto à presença do agente candidato, o polipéptido FGE se liga especificamente a um alvo de ligação celular, uma respetiva porção ou respetivo análogo. A ordem da adição dos componentes, temperatura de incubação, tempo de incubação e outros parâmetros do ensaio podem ser facilmente determinados. Essa experimentação envolve meramente otimização dos parâmetros do ensaio, não a composição fundamental do ensaio. Temperaturas de incubação situam-se tipicamente entre 4°C e 40°C. Tempos de incubação são preferivelmente minimizados para facilitar um rastreio rápido de alto rendimento, e tipicamente situam-se entre 0,1 e 10 horas.

Após a incubação, a presença ou ausência de ligação específica entre o polipéptido FGE e um ou mais alvos de ligação é detetada por qualquer método conveniente disponibilizado ao utilizador. Para ensaios do tipo ligação sem células é muitas vezes utilizado um passo de separação para separar componentes ligados dos não ligados. O passo de separação pode ser realizado de uma variedade de modos. Convenientemente, pelo menos um dos componentes é imobilizado num substrato sólido, de onde podem ser facilmente separados os componentes não ligados. O substrato sólido pode ser feito de uma grande variedade de materiais e ter uma grande variedade de configurações, por exemplo, placa de microtítulo, microesférula, tira

reagente, partícula de resina, etc. O substrato é preferivelmente escolhido de modo a maximizar razões sinal-ruído, principalmente para minimizar a ligação de fundo, bem como por facilidade de separação e custo.

A separação pode ser efetuada, por exemplo, por remoção de uma esférula ou tira reagente de um reservatório, esvaziamento ou diluição de um reservatório, como uma cavidade de uma placa de microtítulo, enxaguamento de uma esférula, partícula, coluna cromatográfica ou filtro com uma solução de lavagem ou solvente. Preferivelmente, o passo de separação inclui múltiplos enxaguamentos ou lavagens. Por exemplo, quando o substrato sólido é uma placa de microtítulo, as cavidades podem ser lavadas várias vezes com uma solução de lavagem, que tipicamente inclui aqueles componentes da mistura de incubação que não participam em ligações específicas, como sais, tampão, detergente, proteína inespecífica, etc. Quando o substrato sólido for uma esférula magnética, as esférulas podem ser lavadas uma ou mais vezes com uma solução de lavagem e isoladas utilizando um ímã.

A detecção pode ser efetuada de qualquer modo conveniente para ensaios à base de células, tais como rastreios de dois ou três híbridos. O transcrito resultante de um ensaio de transcrição de gene repórter de polipéptido FGE interagindo com uma molécula alvo codifica tipicamente um produto direta ou indiretamente detetável, por exemplo, atividade de β -galactosidase, atividade de luciferase, e afins. Para ensaios de ligação sem células, um dos componentes habitualmente compreende, ou está acoplado a, uma etiqueta detetável. Pode ser utilizada uma grande variedade de etiquetas, como as que proporcionam detecção direta (por

exemplo, radioatividade, luminescência, densidade ótica ou eletrônica, etc.) ou detecção indireta (por exemplo, epítipo marcador, como o epítipo FLAG, enzima marcadora, como peroxidase de rábano bravo, etc.). A etiqueta pode ser ligada a um parceiro de ligação de FGE, ou pode ser incorporada na estrutura do parceiro de ligação.

Pode ser utilizada uma variedade de métodos para detetar a etiqueta, dependendo da natureza da etiqueta e de outros componentes do ensaio. Por exemplo, a etiqueta pode ser detetada enquanto está ligada ao substrato sólido ou subsequentemente à separação do substrato sólido. Etiquetas podem ser diretamente detetadas através de densidade ótica ou eletrônica, emissões radioativas, transferências de energia não radiativa, etc., ou indiretamente detetadas com conjugados de anticorpos, conjugados de estreptavidina-biotina, etc. Métodos de detecção das etiquetas são bem conhecidos na área.

Podem ser proporcionados agentes de ligação específicos para FGE, métodos de identificação e preparação desses agentes, e sua utilização em diagnóstico, terapia e desenvolvimento farmacêutico. Por exemplo, agentes farmacológicos específicos para FGE são úteis numa variedade de aplicações de diagnóstico e terapêuticas, especialmente quando a doença ou prognóstico da doença estiver associado a características alteradas de ligação de FGE, como em Deficiência de Múltiplas Sulfatases. Novos agentes de ligação específicos para FGE incluem anticorpos específicos para FGE, recetores da superfície celular, e outros agentes de ligação naturais intracelulares ou extracelulares identificados com ensaios tais como rastreios de dois híbridos, e agentes de ligação não

naturais intracelulares e extracelulares identificados em rastreios de bibliotecas químicas e afins.

Em geral, a especificidade da ligação de FGE a uma molécula específica é determinada por constantes de equilíbrio de ligação. Alvos que são capazes de se ligarem seletivamente a um polipéptido FGE têm, preferivelmente, constantes de equilíbrio de ligação de pelo menos cerca de 10^7 M^{-1} , mais preferivelmente pelo menos cerca de 10^8 M^{-1} e muito preferivelmente pelo menos cerca de 10^9 M^{-1} . Pode ser utilizada uma grande variedade de ensaios à base de células e sem células para demonstrar a ligação específica de FGE. Ensaaios à base de células incluem rastreios de um, dois e três híbridos, ensaios nos quais a transcrição mediada pela FGE é inibida ou aumentada, etc. Ensaaios sem células incluem ensaios de ligação FGE-proteína, imunensaaios, etc. Outros ensaios úteis para o rastreio de agentes que se ligam a polipéptidos FGE incluem transferência de energia de ressonância de fluorescência (FRET) e análise de desvio da mobilidade eletroforética (EMSA).

Pode ser proporcionado um método de identificação de um agente útil na modulação da atividade geradora de C_α -formilglicina. O método envolve (a) contactar uma molécula com atividade geradora de C_α -formilglicina com um agente candidato, (b) medir a atividade geradora de C_α -formilglicina da molécula, e (c) comparar a atividade geradora de C_α -formilglicina medida da molécula com um controlo, para determinar se o agente candidato modula a atividade geradora de C_α -formilglicina da molécula, em que a molécula é uma molécula de ácido nucleico FGE como descrito aqui, ou um respetivo produto de expressão. "Contactar" refere-se ao contacto direto e indireto de uma

molécula com atividade geradora de C_{α} -formilglicina com o agente candidato. Contacto "indireto" significa que o agente candidato exerce os seus efeitos na atividade geradora de C_{α} -formilglicina da molécula via um terceiro agente (por exemplo, uma molécula mensageira, um recetor, etc.). Em certas formas de realização, o controlo é a atividade geradora de C_{α} -formilglicina da molécula medida na ausência do agente candidato. Métodos de ensaio e agentes candidatos são como descritos acima nas formas de realização anteriores relativamente à FGE.

Pode ser proporcionado um método de diagnóstico de uma perturbação caracterizada por expressão aberrante de uma molécula de ácido nucleico, um respetivo produto de expressão, ou um fragmento de um respetivo produto de expressão. O método envolve contactar uma amostra biológica isolada de um sujeito com um agente que se liga especificamente à molécula de ácido nucleico, um respetivo produto de expressão, ou um fragmento de um respetivo produto de expressão, e determinar a interação entre o agente e a molécula de ácido nucleico ou o produto de expressão como determinação da perturbação, em que a molécula de ácido nucleico é uma molécula FGE como descrito aqui. A perturbação é Deficiência de Múltiplas Sulfatases. Mutações no gene FGE que causam a expressão aberrante de moléculas FGE originam as seguintes alterações de aminoácidos na SEQ ID NO: 2: Met1Arg; Met1Val; Leu20Phe; Ser155Pro; Ala177Pro; Cys218Tyr; Arg224Trp; Asn259De; Pro266Leu; Ala279Val; Arg327Stop; Cys336Arg; Arg345Cys; Ala348Pro; Arg349Gln; Arg349Trp; Arg349Trp; Ser359Stop; ou uma combinação destas.

No caso de a molécula ser uma molécula de ácido nucleico, essas determinações podem ser efetuadas via qualquer ensaio comum de determinação de ácidos nucleicos, incluindo a reação em cadeia de polimerase, ou ensaio com sondas de hibridização etiquetadas, como exemplificado aqui. No caso de a molécula ser um produto de expressão da molécula de ácido nucleico, ou um fragmento de um produto de expressão da molécula de ácido nucleico, essa determinação pode ser efetuada via qualquer ensaio imunológico comum utilizando, por exemplo, anticorpos que se ligam a quaisquer dos produtos de expressão polipeptídicos.

"Expressão aberrante" refere-se a expressão decrescida (subexpressão) ou expressão aumentada (superexpressão) de moléculas FGE (ácidos nucleicos e/ou polipéptidos) em comparação com um controlo (isto é, expressão da mesma molécula num sujeito saudável ou "normal"). Um "sujeito saudável", como usado aqui, refere-se a um sujeito que, de acordo com padrões médicos comuns, não tem nem está em risco de desenvolver Deficiência de Múltiplas Sulfatases. Além disso, sujeitos saudáveis também não exibem sintomas de doença. Por outras palavras, esses sujeitos, se examinados por um profissional clínico, seriam caracterizados como sendo saudáveis e sem sintomas de uma Deficiência de Múltiplas Sulfatases. Estes incluem características de leucodistrofia metacromática e de uma mucopolissacaridose, como quantidades aumentadas de mucopolissacáridos ácidos em vários tecidos, "gargoilismo" ligeiro, rápida deterioração neurológica, presença excessiva de mucopolissacáridos e sulfatídeos na urina, proteína do fluido cerebrospinal aumentada e degeneração metacromática de mielina em nervos periféricos.

Novos estojos podem ser utilizados para medir os níveis dos ácidos nucleicos descritos aqui, ou respetivos produtos de expressão.

Numa forma de realização, um estajo compreende uma embalagem contendo um agente que se liga seletivamente a quaisquer dos ácidos nucleicos FGE isolados anteriores, ou respetivos produtos de expressão, e um controlo para comparação com um valor medido de ligação do referido agente a quaisquer dos ácidos nucleicos FGE isolados anteriores, ou respetivos produtos de expressão. Nalgumas formas de realização, o controlo consiste num valor predeterminado para comparação com o valor medido. Em certas formas de realização, o controlo compreende um epítipo do produto de expressão de quaisquer dos ácidos nucleicos FGE isolados anteriores. Numa forma de realização, o estajo também compreende um segundo agente que se liga seletivamente a um polipéptido selecionado do grupo que consiste em Iduronato 2-Sulfatase, Sulfamidase, N-Acetilgalactosamina 6-Sulfatase, N-Acetilglucosamina 6-Sulfatase, Arilsulfatase A, Arilsulfatase B, Arilsulfatase C, Arilsulfatase D, Arilsulfatase E, Arilsulfatase F, Arilsulfatase G, HSulf-1, HSulf-2, HSulf-3, HSulf-4, HSulf-5 e HSulf-6, ou um respetivo péptido, e um controlo para comparação com um valor medido de ligação do referido segundo agente ao referido polipéptido ou respetivo péptido.

No caso de deteção de ácidos nucleicos podem ser incluídos pares de "primers" para amplificação de uma molécula de ácido nucleico. Os estojos preferidos incluem controlos tais como quantidades conhecidas de sondas de ácidos nucleicos, epítipos (como Iduronato 2-Sulfatase,

Sulfamidase, N-Acetilgalactosamina 6-Sulfatase, N-Acetilglucosamina 6-Sulfatase, Arilsulfatase A, Arilsulfatase B, Arilsulfatase C, Arilsulfatase D, Arilsulfatase E, Arilsulfatase F, Arilsulfatase G, HSulf-1, HSulf-2, HSulf-3, HSulf-4, HSulf-5 e HSulf-6, produtos de expressão) ou anticorpos anti-epítomos, bem como instruções ou outro material impresso. Em certas formas de realização, o material impresso pode caracterizar o risco de desenvolvimento de um estado de deficiência de sulfatases com base no resultado do ensaio. Os reagentes podem ser embalados em recipientes e/ou podem ser revestidos em cavidades em quantidades predeterminadas, e os estojos podem incluir materiais comuns, tais como reagentes imunológicos etiquetados (como anticorpos anti-IgG etiquetados) e afins. Um estojo consiste numa placa de microtítulo de poliestireno embalada revestida com proteína FGE e um recipiente contendo anticorpos anti-IgG humana etiquetados. Uma cavidade da placa é contactada, por exemplo, com um fluido biológico, é lavada e então é contactada com o anticorpo anti-IgG. Em seguida, a etiqueta é detetada. Um estojo genericamente denominado número 11 está ilustrado na Figura 25. O estojo 11 é compreendido pelos seguintes elementos principais: embalagem 15, um agente como descrito aqui 17, um agente de controlo 19 e instruções 21. A embalagem 15 é uma estrutura do tipo caixa para conter um frasco (ou alguns frascos) contendo o agente 17, um frasco (ou alguns frascos) contendo um agente de controlo 19, e instruções 21. Os peritos podem facilmente modificar a embalagem 15 para satisfazer necessidades individuais.

Podem ser proporcionados métodos de tratamento de Deficiência de Múltiplas Sulfatases num sujeito. Um método

envolve administrar, a um sujeito necessitado desse tratamento, um agente que modula a atividade geradora de C α -formilglicina, numa quantidade eficaz para aumentar a atividade geradora de C α -formilglicina no sujeito. Nalgumas formas de realização, o método também compreende coadministrar um agente selecionado do grupo que consiste numa molécula de ácido nucleico que codifica Iduronato 2-Sulfatase, Sulfamidase, N-Acetilgalactosamina 6-Sulfatase, N-Acetilglucosamina 6-Sulfatase, Arilsulfatase A, Arilsulfatase B, Arilsulfatase C, Arilsulfatase D, Arilsulfatase E, Arilsulfatase F, Arilsulfatase G, HSulf-1, HSulf-2, HSulf-3, HSulf-4, HSulf-5 e HSulf-6, um produto de expressão da molécula de ácido nucleico e/ou um fragmento do produto de expressão da molécula de ácido nucleico.

"Agentes que modulam a expressão" de um ácido nucleico ou de um polipéptido, como usado aqui, são conhecidos na área, e referem-se a ácidos nucleicos sentido e antissentido, ácidos nucleicos dominantes negativos, anticorpos para os polipéptidos, e afins. São úteis quaisquer agentes que modulem a expressão de uma molécula (e, como descrito aqui, modulem a sua atividade). Em certas formas de realização, o agente que modula a atividade geradora de C α -formilglicina é uma molécula de ácido nucleico isolado (por exemplo, um ácido nucleico com a SEQ ID NO: 3). Em formas de realização importantes, o agente que modula a atividade geradora de C α -formilglicina é um péptido (por exemplo, um péptido com a SEQ ID NO: 2). Nalgumas formas de realização, o agente que modula a atividade geradora de C α -formilglicina é um ácido nucleico sentido.

Pode ser proporcionado um método para aumentar a atividade geradora de C α -formilglicina num sujeito. O método envolve

administrar a um sujeito uma molécula de ácido nucleico FGE isolada descrita aqui, e/ou um respetivo produto de expressão, numa quantidade eficaz para aumentar a atividade geradora de C α -formilglicina no sujeito.

Pode ser proporcionado um método para aumentar a atividade geradora de C α -formilglicina numa célula. O método envolve contactar a célula com uma molécula de ácido nucleico isolado (por exemplo, um ácido nucleico com a SEQ ID NO: 1), ou um respetivo produto de expressão (por exemplo, um péptido com a SEQ ID NO: 2), numa quantidade eficaz para aumentar a atividade geradora de C α -formilglicina na célula. Em formas de realização importantes, o método envolve ativar o gene FGE endógeno para aumentar a atividade geradora de C α -formilglicina na célula.

Em quaisquer das formas de realização anteriores, o ácido nucleico pode ser operativamente acoplado a uma sequência de expressão genética que dirige a expressão da molécula de ácido nucleico numa célula eucariótica, como uma célula HT-1080. A "sequência de expressão genética" é qualquer sequência de nucleótidos reguladora, como uma sequência promotora ou combinação promotor-intensificador, que facilita a transcrição e tradução eficientes do ácido nucleico ao qual está operativamente ligada. A sequência de expressão genética pode ser, por exemplo, um promotor de mamífero ou viral, como um promotor constitutivo ou indutível. Promotores constitutivos de mamífero incluem mas não estão limitados aos promotores para os genes seguintes: hipoxantina fosforribosil transferase (HFTR), adenosina desaminase, piruvato quinase, promotor da α -actina e outros promotores constitutivos. Promotores virais exemplificativos que funcionam de modo constitutivo em

células eucarióticas incluem, por exemplo, promotores do vírus símio, vírus papiloma, adenovírus, vírus da imunodeficiência humana (HIV), vírus do sarcoma de Rous, citomegalovírus, as repetições terminais longas (LTRs) do vírus da leucemia de Moloney e outros retrovírus, e o promotor de timidina quinase do vírus herpes simplex. Outros promotores constitutivos são conhecidos dos profissionais. Os promotores úteis como sequências de expressão genética também incluem promotores indutíveis. Promotores indutíveis são ativados na presença de um agente indutor. Por exemplo, o promotor da metalotioneína é ativado para aumentar a transcrição e tradução na presença de certos íons metálicos. Outros promotores indutíveis são conhecidos dos profissionais.

Em geral, a sequência de expressão genética incluirá, consoante o necessário, sequências 5' não transcritas e 5' não traduzidas envolvidas na iniciação da transcrição e tradução, respetivamente, como uma caixa TATA, sequência de capeamento, sequência CAAT, e afins. Especialmente, essas sequências 5' não transcritas incluirão uma região promotora que inclui uma sequência promotora para controlo da transcrição do ácido nucleico operativamente ligado. As sequências de expressão genética incluem opcionalmente sequências intensificadoras ou sequências ativadoras a montante, consoante o desejado.

Preferivelmente, qualquer uma das moléculas de ácidos nucleicos FGE descritas aqui está ligada a uma sequência de expressão genética que permite a expressão da molécula de ácido nucleico numa célula de uma linhagem específica de células, por exemplo, um neurónio. Uma sequência que permite a expressão da molécula de ácido nucleico numa

célula como um neurónio é uma sequência que é seletivamente ativa nesse tipo de célula, desse modo originando expressão da molécula de ácido nucleico nestas células. O promotor da sinapsina-1, por exemplo, pode ser utilizado para expressar num neurónio quaisquer das moléculas de ácidos nucleicos anteriores da invenção; e o promotor do gene do fator de von Willebrand, por exemplo, pode ser utilizado para expressar uma molécula de ácido nucleico numa célula vascular endotelial. Os profissionais conseguirão facilmente identificar promotores alternativos que são capazes de expressar uma molécula de ácido nucleico em quaisquer das células preferidas.

É afirmado que a sequência de ácido nucleico e a sequência de expressão genética estão "operativamente ligadas" quando estão ligadas de modo covalente por forma a colocar a transcrição e/ou tradução da sequência codificadora do ácido nucleico (por exemplo, no caso de FGE, SEQ ID NO: 3) sob a influência ou controlo da sequência de expressão genética. Se for desejado que a sequência de ácido nucleico seja traduzida numa proteína funcional, é afirmado que duas sequências de DNA estão operativamente ligadas se a indução de um promotor na sequência de expressão genética 5' der origem à transcrição da sequência de ácido nucleico e se a natureza da ligação entre as duas sequências de DNA não (1) originar a introdução de uma mutação por deslocamento do quadro de leitura, (2) interferir na capacidade da região promotora para dirigir a transcrição da sequência de ácido nucleico, e/ou (3) interferir na capacidade do transcrito de RNA correspondente para ser traduzido numa proteína. Assim, uma sequência de expressão genética estará operativamente ligada a uma sequência de ácido nucleico se a sequência de expressão genética for capaz de efetuar a

transcrição dessa sequência de ácido nucleico de modo que o transcrito resultante possa ser traduzido na proteína ou polipéptido desejado.

As moléculas descritas aqui podem ser distribuídas nos tipos de células preferidos isoladamente ou associadas a um vetor (ver também a discussão anterior sobre vetores). No seu sentido mais lato (e consistente com a descrição de vetores de expressão e direcionamento aqui noutro local), um "vetor" é qualquer veículo capaz de facilitar: (1) a distribuição de uma molécula numa célula alvo e/ou (2) a captação da molécula por uma célula alvo. Preferivelmente, os vetores de distribuição transportam a molécula para o interior da célula alvo com degradação reduzida relativamente à extensão de degradação que ocorreria na ausência do vetor. Opcionalmente, um "ligando de direcionamento" pode ser ligado ao vetor para distribuir seletivamente o vetor numa célula que expressa na sua superfície o recetor cognato para o ligando de direcionamento. Deste modo, o vetor (contendo um ácido nucleico ou uma proteína) pode ser seletivamente distribuído num neurónio. Metodologias de direcionamento incluem conjugados, como os descritos na Patente U.S. 5,391,723, atribuída a Priest. Outro exemplo de um veículo de direcionamento bem conhecido é um lipossoma. Lipossomas são disponibilizados no mercado pela Gibco BRL. Estão publicados numerosos métodos para preparar lipossomas dirigidos.

Em geral, vetores úteis incluem mas não estão limitados a plasmídeos, fagemídeos, vírus, outros veículos derivados de fontes virais ou bacterianas que foram manipulados pela inserção ou incorporação das sequências de ácidos nucleicos

descritas aqui, e fragmentos de ácidos nucleicos adicionais (por exemplo, intensificadores, promotores) que podem ser ligados às sequências de ácidos nucleicos. Vetores virais são um tipo preferido de vetor e incluem mas não estão limitados a sequências de ácidos nucleicos dos vírus seguintes: adenovírus; vírus adeno-associados; retrovírus, tais como o vírus da leucemia murina de Moloney; vírus do sarcoma murino de Harvey; vírus de tumores mamários murinos; vírus do sarcoma de Rous; vírus do tipo SV40; vírus polioma; vírus Epstein-Barr; vírus papiloma; vírus herpes; vírus vacínia; vírus polio; e vírus de RNA, como um retrovírus. É possível empregar facilmente outros vetores não designados mas conhecidos na área.

Um vírus particularmente preferido para certas aplicações é o vírus adeno-associado, um vírus de DNA de filamentação dupla. O vírus adeno-associado é capaz de infectar uma gama larga de tipos de células e espécies e pode ser manipulado para ser deficiente quanto à replicação. Tem outras vantagens, como estabilidade térmica e em solventes lipídicos, elevadas frequências de transdução em células de diversas linhagens, incluindo células hematopoiéticas, e ausência de inibição de superinfecção, desse modo permitindo múltiplas séries de transduções. Como relatado, o vírus adeno-associado pode ser integrado em DNA celular humano de um modo específico para sítios, desse modo minimizando a possibilidade de mutagênese por inserção e variabilidade da expressão do gene inserido. Adicionalmente, infecções com vírus adeno-associado de tipo selvagem têm sido seguidas em cultura de tecidos durante mais de 100 passagens na ausência de pressão seletiva, implicando que a integração genômica do vírus adeno-associado é um evento relativamente

estável. O vírus adeno-associado também pode funcionar de um modo extracromossômico.

Em geral, outros vetores virais preferidos baseiam-se em vírus eucarióticos não citopáticos nos quais genes não essenciais foram substituídos pelo gene de interesse. Vírus não citopáticos incluem retrovírus, cujo ciclo de vida envolve transcrição reversa de RNA viral genômico em DNA, com subsequente integração pró-viral em DNA celular hospedeiro. Adenovírus e retrovírus têm sido aprovados para ensaios de terapia genética humana. Em geral, os retrovírus são deficientes quanto à replicação (isto é, capazes de dirigir a síntese das proteínas desejadas, mas incapazes de produzir uma partícula infecciosa). Esses vetores de expressão retrovirais geneticamente alterados têm utilidade geral para a transdução com elevada eficiência de genes *in vivo*. Protocolos comuns para a produção de retrovírus deficientes quanto à replicação (incluindo os passos de incorporação de material genético exógeno num plasmídeo, transfeção de uma célula de empacotamento revestida com plasmídeo, produção de retrovírus recombinantes pela linha de células de empacotamento, recolha de partículas virais de meios de cultura de tecidos, e infecção das células alvo com partículas virais) são proporcionados em Kriegler, M., "Gene Transfer and Expression, A Laboratory Manual," W.H. Freeman C.O., Nova Iorque (1990), e Murry, E.J. Editor "Methods in Molecular Biology," volume 7, Humana Press, Inc., Cliffton, Nova Jérσία (1991).

Outro vetor retroviral preferido é o vetor derivado do vírus da leucemia murina de Moloney, como descrito em Nabel, E.G., *et al.*, *Science*, 1990, 249: 1285-1288. Foi relatado que estes vetores foram eficazes para a

distribuição de genes nas três camadas da parede arterial, incluindo a média. Outros vetores preferidos são divulgados em Flugelman, *et al.*, *Circulation*, 1992, 85: 1110-1117. Vetores adicionais que são úteis para a distribuição de moléculas são descritos na Patente U.S. Nº 5,674,722 de Mulligan, *et al.*

Para além dos vetores anteriores, podem ser utilizados outros métodos de distribuição para distribuir uma molécula descrita aqui numa célula, como um neurónio, célula do fígado, fibroblasto e/ou uma célula vascular endotelial, e desse modo facilitar a captação daquela.

Um tal método de distribuição preferido é um sistema de dispersão coloidal. Sistemas de dispersão coloidal incluem sistemas à base de lípidos, incluindo emulsões óleo-em-água, micélios, micélios misturados e lipossomas. Um sistema coloidal preferido é um lipossoma. Lipossomas são vasos membranares artificiais que são úteis como vetor de distribuição *in vivo* ou *in vitro*. Foi mostrado que vasos unilamelares grandes (LUV), cuja dimensão varia entre 0,2 - 4,0 μm , podem encapsular grandes macromoléculas. RNA, DNA e viriões intactos podem ser encapsulados no interior aquoso e podem ser distribuídos em células numa forma biologicamente ativa (Fraley, *et al.*, *Trends Biochem. Sci.*, 1981, 6: 77). Para que um lipossoma seja um vetor de transferência genética eficiente devem estar presentes uma ou mais das seguintes características: (1) encapsulação do gene de interesse com elevada eficiência, com retenção da atividade biológica; (2) ligação preferencial e substancial a uma célula alvo em comparação com células não alvo; (3) distribuição do conteúdo aquoso da vesícula no citoplasma

da célula alvo com elevada eficiência; e (4) expressão precisa e eficaz da informação genética.

Os lipossomas podem ser dirigidos para um tecido particular, como o miocárdio ou a parede de células vasculares, por acoplamento do lipossoma a um ligando específico, como um anticorpo monoclonal, açúcar, glicolípido ou proteína. Ligandos que podem ser úteis para dirigir um lipossoma para a parede vascular incluem mas não estão limitados à proteína do envelope viral do vírus de Hemaglutinação do Japão. Adicionalmente, o vetor pode ser acoplado a um péptido de direcionamento nuclear, que irá dirigir o ácido nucleico para o núcleo da célula hospedeiro.

Lipossomas são disponibilizados comercialmente pela Gibco BRL, por exemplo, como LIPOFECTIN™ e LIPOFECTACE™, constituídos por lípidos catiónicos tais como cloreto de N-[1-(2,3 dioleíloxi)-propil]-N,N,N-trimetilamónio (DOTMA) e brometo de dimetildiocetadecilamónio (DDAB). Métodos de preparação de lipossomas são bem conhecidos na área e têm sido descritos em muitas publicações. Lipossomas também foram revistos por Gregoriadis, G., em "Trends in Biotechnology". Volume 3, página 235-241 (1985). Novos lipossomas para a distribuição intracelular de macromoléculas, incluindo ácidos nucleicos, também são descritos no requerimento Internacional PCT nº PCT/US96/07572 (Publicação Nº WO 96/40060, intitulado "Intracellular Delivery of Macromolecules").

Numa forma de realização particular, o veículo preferido é uma micropartícula ou implante biocompatível que é adequado para implantação no recetor mamífero. Implantes

bioerodíveis exemplificativos que são úteis de acordo com este método são descritos no requerimento Internacional PCT nº PCT/LUS/03307 (Publicação Nº WO 95/24929, intitulada "Polymeric Gene Delivery System", que reivindica prioridade para o requerimento de patente U.S. nº de série 213,668, registado em 15 de março de 1994). O documento PCT/US/0307 descreve uma matriz polimérica biocompatível, preferivelmente biodegradável, para conter um gene exógeno sob o controlo de um promotor apropriado. A matriz polimérica é utilizada para se obter um aumento sustentado do gene exógeno no paciente. Os ácidos nucleicos descritos aqui podem ser encapsulados ou dispersos na matriz polimérica biocompatível, preferivelmente biodegradável, divulgada em PCT/US/03307. A matriz polimérica está preferivelmente na forma de uma micropartícula, como uma microesfera (em que um ácido nucleico é disperso numa matriz polimérica sólida) ou uma microcápsula (em que um ácido nucleico é armazenado no núcleo de um invólucro polimérico). Outras formas da matriz polimérica para conter os ácidos nucleicos incluem filmes, revestimentos, géis, implantes e endopróteses. A dimensão e composição do dispositivo de matriz polimérica são seleccionadas de modo a originar uma cinética de libertação favorável no tecido onde está implantado o dispositivo matricial. A dimensão do dispositivo de matriz polimérica é adicionalmente seleccionada de acordo com o método de distribuição que se pretende utilizar, tipicamente injeção num tecido ou administração de uma suspensão por aerossol nas áreas nasais e/ou pulmonares. A composição da matriz polimérica pode ser seleccionada de modo a exibir taxas de degradação favoráveis e também de modo a ser formada por um material que é bioadesivo, para aumentar adicionalmente a eficácia da transferência quando o dispositivo é administrado numa

superfície vascular. A composição da matriz também pode ser selecionada de modo a não ocorrer degradação mas sim libertação por difusão durante um período de tempo prolongado.

Matrizes poliméricas não biodegradáveis e biodegradáveis podem ser utilizadas para distribuir os ácidos nucleicos no sujeito. São preferidas matrizes biodegradáveis. Esses polímeros podem ser polímeros naturais ou sintéticos. São preferidos polímeros sintéticos. O polímero é selecionado com base no período de tempo durante o qual é desejada a libertação, geralmente da ordem de algumas horas até um ano ou mais. Tipicamente, é muito desejável libertação durante um período de tempo que varia entre algumas horas e três até doze meses. O polímero está opcionalmente na forma de um hidrogel que pode absorver até 90% do seu peso em água e mais, está opcionalmente reticulado com iões multivalentes ou outros polímeros.

Em geral, os ácidos nucleicos descritos aqui podem ser distribuídos utilizando o implante bioerodível através de difusão ou, mais preferivelmente, por degradação da matriz polimérica. Polímeros sintéticos exemplificativos que podem ser utilizados para formar o sistema de distribuição biodegradável incluem: poliamidas, policarbonatos, polialquilenos, polialquilenoglicóis, óxidos de polialquileno, tereftalatos de polialquileno, álcoois polivinílicos, éteres de polivinilo, ésteres de polivinilo, haletos de polivinilo, polivinilpirrolidona, poliglicólidos, polissiloxanos, poliuretanos e respetivos copolímeros, alquilcelulose, hidroxialquilceluloses, éteres de celulose, ésteres de celulose, nitroceluloses, polímeros de ésteres acrílicos e metacrílicos, metilcelulose,

etilcelulose, hidroxipropilcelulose, hidroxipropilmetilcelulose, hidroxibutilmetilcelulose, acetato de celulose, propionato de celulose, butirato de acetato de celulose, ftalato de acetato de celulose, carboxietilcelulose, triacetato de celulose, sal de sódio de sulfato de celulose, poli(metacrilato de metilo), poli(metacrilato de etilo), poli(metacrilato de butilo), poli(metacrilato de isobutilo), poli(metacrilato de hexilo), poli(metacrilato de isodecilo), poli(metacrilato de laurilo), poli(metacrilato de fenilo), poli(acrilato de metilo), poli(acrilato de isopropilo), poli(acrilato de isobutilo), poli(acrilato de octadecilo), polietileno, polipropileno, poli(etilenoglicol), poli(óxido de etileno), poli(tereftalato de etileno), poli(álcoois vinílicos), acetato de polivinilo, cloreto de polivinilo, poliestireno e polivinilpirrolidona.

Exemplos de polímeros não biodegradáveis incluem etileno acetato de vinilo, ácido poli(met)acrílico, poliamidas, copolímeros e misturas destes.

Exemplos de polímeros biodegradáveis incluem polímeros sintéticos, como polímeros de ácido láctico e ácido glicólico, polianidridos, poli(orto)ésteres, poliuretanos, poli(ácido bórico), poli(ácido valérico), e poli(láctido-co-caprolactona), e polímeros naturais, como alginato e outros polissacáridos incluindo dextrano e celulose, colagénio, respetivos derivados químicos (substituições, adições de grupos químicos, por exemplo, alquilo, alquilenos, hidroxilações, oxidações, e outras modificações rotineiramente feitas pelos profissionais), albumina e outras proteínas hidrófilas, zeína e outras prolamínas e proteínas hidrófobas, respetivos copolímeros e misturas. Em

geral, estes materiais degradam-se por hidrólise enzimática ou exposição a água *in vivo*, por erosão superficial ou em volume.

Polímeros bioadesivos de interesse particular incluem hidrogéis bioerodíveis descritos por H.S. Sawhney, C.P. Patak e J.A. Hubel em *Macromolecules*, 1993, 26, 581-587, cujos ensinamentos são incorporados aqui, ácidos polihialurónicos, caseína, gelatina, glutina, polianidridos, ácido poliacrílico, alginato, quitosano, poli(metacrilatos de metilo), poli(metacrilatos de etilo), poli(metacrilato de butilo), poli(metacrilato de isobutilo), poli(metacrilato de hexilo), poli(metacrilato de isodecilo), poli(metacrilato de laurilo), poli(metacrilato de fenilo), poli(acrilato de metilo) poli(acrilato de isopropocrilo), poli(acrilato de isobutilo), e poli(acrilato de octadecilo). Assim, podem ser proporcionados os seguintes: uma composição das moléculas descritas acima para utilização como medicamento, métodos de preparação do medicamento e métodos para a libertação prolongada do medicamento *in vivo*.

Agentes de compactação também podem ser utilizados em combinação com um vetor. Um "agente de compactação", como usado aqui, refere-se a um agente, como uma histona, que neutraliza as cargas negativas do ácido nucleico e, desse modo, permite a compactação do ácido nucleico num grânulo fino. A compactação do ácido nucleico facilita a captação do ácido nucleico pela célula alvo. Os agentes de compactação podem ser utilizados isoladamente, isto é, para distribuir um ácido nucleico isolado como descrito aqui numa forma que é mais eficientemente absorvida pela célula

ou, mais preferivelmente, em combinação com um ou mais dos vetores acima descritos.

Outras composições exemplificativas que podem ser utilizadas para facilitar a captação, pela célula alvo, dos ácidos nucleicos incluem fosfato de cálcio e outros mediadores químicos do transporte intracelular, composições para microinjeção e eletroporação.

Podem ser proporcionados métodos para aumentar a atividade de sulfatase numa célula. Esses métodos envolvem contactar uma célula que expressa uma sulfatase com uma molécula de ácido nucleico isolado como descrito aqui (por exemplo, uma molécula de ácido nucleico isolado como reivindicado em qualquer uma das Reivindicações 1-8, uma molécula de ácido nucleico FGE possuindo uma sequência selecionada do grupo que consiste na SEQ ID NO: 1, 3, 4, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77 e 80-87), ou um respetivo produto de expressão (por exemplo, um polipéptido como reivindicado nas Reivindicações 11-15, 19, 20, ou um péptido possuindo uma sequência selecionada do grupo que consiste na SEQ ID NO: 2, 5, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76 e 78), numa quantidade eficaz para aumentar a atividade de sulfatase na célula. "Aumentar" a atividade de sulfatase, como usado aqui, refere-se a afinidade aumentada para, e/ou conversão de, o substrato específico para a sulfatase, tipicamente o resultado de um aumento da formação de *FGly* na molécula de sulfatase. Numa forma de realização, a célula expressa uma sulfatase em níveis mais elevados do que os de células de tipo selvagem. "Aumentar a atividade de sulfatase numa célula" também se refere a aumentar a atividade de uma sulfatase que é segregada pela célula. A célula pode

expressar uma sulfatase endógena e/ou exógena. O referido contacto da molécula FGE também se refere à ativação do gene FGE endógeno da célula. Em formas de realização importantes, a sulfatase endógena é ativada. Em certas formas de realização, a sulfatase é Iduronato 2-Sulfatase, Sulfamidase, N-Acetilgalactosamina 6-Sulfatase, N-Acetilglucosamina 6-Sulfatase, Arilsulfatase A, Arilsulfatase B, Arilsulfatase C, Arilsulfatase D, Arilsulfatase E, Arilsulfatase F, Arilsulfatase G, HSulf-1, HSulf-2, HSulf-3, HSulf-4, HSulf-5 e/ou HSulf-6. Em certas formas de realização, a célula é de um mamífero.

Pode ser proporcionada uma composição farmacêutica. A composição pode compreender uma sulfatase que é produzida por uma célula, numa quantidade farmacologicamente eficaz para tratar uma deficiência de sulfatases, e um transportador farmacologicamente aceitável, em que a referida célula foi contactada com um agente que compreende uma molécula de ácido nucleico isolado como descrito aqui (por exemplo, como reivindicado nas Reivindicações 1-8, ou um ácido nucleico como reivindicado nas Reivindicações 1-8, ou uma molécula de ácido nucleico possuindo uma sequência selecionada do grupo que consiste na SEQ ID NO: 1, 3, 4, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77 e 80-87), ou um respetivo produto de expressão (por exemplo, um péptido selecionado do grupo que consiste na SEQ ID NO: 2, 5, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76 e 78). Em formas de realização importantes, a sulfatase é expressa em níveis mais elevados do que em células normais/de controlo.

Também pode ser proporcionada uma célula produtora de sulfatase; em que a razão entre sulfatase ativa e sulfatase

total produzida pela célula é aumentada. A célula compreende: (i) uma sulfatase com uma atividade aumentada em comparação com um controlo, e (ii) uma Enzima geradora de Formilglicina com uma atividade aumentada em comparação com um controlo, em que a razão entre sulfatase ativa e sulfatase total produzida pela célula é aumentada em pelo menos 5% relativamente à razão entre sulfatase ativa e sulfatase total produzida pela célula na ausência da Enzima Geradora de Formilglicina. É conhecido na área que a superexpressão de sulfatases pode diminuir a atividade de sulfatases endógenas (Anson *et al.*, *Biochem. J.*, 1993, 294: 657-662). Além disso, apenas uma fração das sulfatases recombinantes é ativa. Descobrimos, inesperadamente, que uma expressão/atividade aumentada de FGE numa célula com expressão/atividade aumentada de uma sulfatase leva à produção de uma sulfatase que é mais ativa. Uma vez que a presença de *FGly* numa molécula de sulfatase está associada à atividade de sulfatase, a "sulfatase ativa" pode ser quantificada determinando a presença de *FGly* no produto celular de sulfatase utilizando espetrometria de massa MALDI-TOF, como descrito aqui noutra local. A razão com sulfatase total pode então ser facilmente determinada.

Podem ser proporcionados métodos para o diagnóstico e terapia de deficiências de sulfatases. Essas perturbações incluem mas não estão limitadas a Deficiência de Múltiplas Sulfatases, Mucopolissacaridose II (MPS II; Síndrome de Hunter), Mucopolissacaridose IIIA (MPS IIIA; Síndrome de Sanfilippo A), Mucopolissacaridose VIII (MPS VIII), Mucopolissacaridose IVA (MPS IVA; Síndrome de Morquio A), Mucopolissacaridose VI (MPS VI; Síndrome de Maroteaux-Lamy), Leucodistrofia Metacromática (MLD), Condrodisplasia *Punctata* Recessiva ligada ao Cromossoma X 1, e Ictiose

ligada ao Cromossoma X (Deficiência de Esteróide Sulfatase).

Os métodos são úteis no tratamento agudo e profilático de quaisquer dos estados anteriores. Como usado aqui, um tratamento agudo refere-se ao tratamento de sujeitos com um estado particular. Tratamento profilático refere-se ao tratamento de sujeitos em risco de contrair o estado mas que presentemente não têm nem sentem os sintomas do estado.

No seu sentido mais lato, os termos "tratamento" ou "tratar" referem-se a tratamentos agudos e profiláticos. Se o sujeito necessitado de tratamento sentir um estado (ou sofre ou está a sofrer de um estado particular), então o tratamento do estado refere-se a melhorar, reduzir ou eliminar o estado ou um ou mais sintomas provenientes do estado. Nalgumas formas de realização preferidas, tratamento do estado refere-se a melhorar, reduzir ou eliminar um sintoma específico ou um subconjunto específico de sintomas associados ao estado. Se o sujeito necessitado de tratamento estiver em risco de sofrer de um estado, então tratar o sujeito refere-se a reduzir o risco do sujeito contrair o estado.

O modo de administração e dosagem de um agente terapêutico irão variar com a fase particular do estado a ser tratado, a idade e estado físico do sujeito a ser tratado, a duração do tratamento, a natureza da terapia concorrente (se existir), a via de administração específica e fatores afins no âmbito do conhecimento e perícia do profissional de saúde.

Como descrito aqui, agentes terapêuticos podem ser administrados em quantidades eficazes para tratar quaisquer das deficiências de sulfatases anteriores. Em geral, uma quantidade eficaz é qualquer quantidade que possa causar uma alteração benéfica num tecido desejado de um sujeito. Preferivelmente, uma quantidade eficaz é aquela quantidade suficiente para causar uma alteração fenotípica favorável num estado particular, como atenuação, alívio ou eliminação de um sintoma ou de todo um estado.

Em geral, uma quantidade eficaz é aquela quantidade de uma preparação farmacêutica que isoladamente, ou em conjunto com outras doses, produz a resposta desejada. Isto pode envolver somente o abrandamento temporário da progressão do estado, apesar de, mais preferivelmente, envolver a interrupção permanente da progressão do estado ou retardar o surgimento ou evitar a ocorrência do estado. Isto pode ser monitorizado por métodos de rotina. Em geral, doses de compostos ativos serão desde cerca de 0,01 mg/kg por dia até 1000 mg/kg por dia. É esperado que sejam adequadas doses variando entre 50 µg - 500 mg/kg, preferivelmente pela via oral e numa ou várias administrações por dia.

Essas quantidades dependerão, obviamente, do estado particular a ser tratado, da gravidade do estado, dos parâmetros do paciente individual, incluindo a idade, condição física, estatura e peso, a duração do tratamento, a natureza de terapia concorrente (se existir), a via de administração específica e fatores afins no âmbito do conhecimento e perícia do profissional de saúde. Certas formas de administração originarão doses mais baixas, como administração intravenosa. No caso de uma resposta num sujeito ser insuficiente às doses iniciais aplicadas, podem

empregar-se doses mais elevadas (ou doses efetivamente mais elevadas por uma via de distribuição diferente, mais localizada) tanto quanto o permitir a tolerância do paciente. São contempladas múltiplas doses por dia para se obterem níveis sistêmicos apropriados de compostos. É preferido geralmente utilizar uma dose máxima, isto é, a dose segura mais elevada de acordo com uma avaliação médica fundamentada. No entanto, os profissionais entenderão que um paciente pode insistir numa dose mais baixa ou dose tolerável por motivos médicos, motivos psicológicos ou virtualmente quaisquer outros motivos.

Os agentes descritos aqui podem ser combinados, opcionalmente, com um transportador farmacêuticamente aceitável para formar uma preparação farmacêutica. O termo "transportador farmacêuticamente aceitável", como usado aqui, significa um ou mais agentes de enchimento, diluentes ou substâncias de encapsulamento sólidos ou líquidos compatíveis que são adequados para administração a um humano. O termo "transportador" designa um ingrediente orgânico ou inorgânico, natural ou sintético, com o qual é combinado o ingrediente ativo para facilitar a aplicação. Os componentes das composições farmacêuticas também podem ser comisturados com as moléculas descritas aqui, e entre si, de modo tal a não haver nenhuma interação que enfraqueça substancialmente a eficácia farmacêutica desejada. Nalguns aspetos, as preparações farmacêuticas compreendem um agente como descrito aqui numa quantidade eficaz para tratar uma perturbação.

As preparações farmacêuticas podem conter agentes de tamponamento adequados, incluindo: ácido acético num sal; ácido cítrico num sal; ácido bórico num sal; ou ácido

fosfórico num sal. As composições farmacêuticas também podem conter, opcionalmente, conservantes adequados, tais como: cloreto de benzalcónio; clorobutanol; parabenos ou timerosal.

Está disponível uma variedade de vias de administração. O modo particular selecionado dependerá, obviamente, do fármaco particular selecionado, da gravidade do estado a ser tratado e a dosagem requerida para se obter eficácia terapêutica. Os métodos descritos aqui, de maneira geral, podem ser implementados utilizando qualquer modo de administração que seja aceitável do ponto de vista médico, significando qualquer modo que produza níveis eficazes dos compostos ativos sem causar efeitos adversos clinicamente inaceitáveis. Esses modos de administração incluem as vias oral, retal, tópica, nasal, intradérmica, transdérmica ou parentérica. O termo "parentérico" inclui subcutâneo, intravenoso, intraomental, intramuscular ou infusão. A via intravenosa ou intramuscular não é particularmente adequada para terapia e profilaxia de longa duração. Como exemplo, composições farmacêuticas para o tratamento agudo de sujeitos com uma dor de cabeça devido a enxaqueca podem ser formuladas numa variedade de diferentes modos e para uma variedade de modos de administração, incluindo comprimidos, cápsulas, pós, supositórios, injeções e pulverizações nasais.

As preparações farmacêuticas podem ser convenientemente apresentadas em forma galénica unitária e podem ser preparadas por quaisquer dos métodos bem conhecidos na área da farmácia. Todos os métodos incluem o passo de associar o agente ativo a um transportador, que constitui um ou mais ingredientes acessórios. Em geral, as composições são

preparadas por associação uniforme e íntima do composto ativo a um transportador líquido, a um transportador sólido finamente dividido, ou ambos, e depois, se necessário, configuração do produto.

Composições adequadas para administração oral podem ser apresentadas como unidades discretas, como cápsulas, comprimidos, rebuçados, em que cada uma contém uma quantidade predeterminada do composto ativo. Outras composições incluem suspensões em líquidos aquosos ou líquidos não aquosos, como um xarope, elixir ou uma emulsão.

Composições adequadas para administração parentérica compreendem convenientemente uma preparação aquosa esterilizada de um agente descrito aqui, que é preferivelmente isotónica com o sangue do recetor. Esta preparação aquosa pode ser formulada de acordo com métodos conhecidos utilizando agentes dispersantes ou humedecedores e agentes de suspensão adequados. A preparação injetável esterilizada também pode ser uma solução ou suspensão injetável esterilizada num diluente ou solvente não tóxico aceitável do ponto de vista parentérico, por exemplo, na forma de uma solução em 1,3-butanodiol. Entre os veículos e solventes aceitáveis que podem ser empregues contam-se água, solução de Ringer e solução isotónica de cloreto de sódio. Adicionalmente, óleos fixos esterilizados são convencionalmente empregues como solvente ou meio de suspensão. Para esta finalidade pode ser empregue qualquer óleo fixo mole, incluindo mono- ou diglicéridos sintéticos. Adicionalmente, ácidos gordos, como ácido oleico podem ser utilizados na preparação de injetáveis. Formulações adequadas para administração oral, subcutânea, intravenosa,

intramuscular, etc., podem ser encontradas em "Remington's Pharmaceutical Sciences", Mack Publishing Co., Easton, PA.

Pode ser proporcionado um método para aumentar a atividade geradora de C_{α} -formilglicina numa célula. O método envolve contactar a célula com uma molécula de ácido nucleico isolado (por exemplo, um ácido nucleico com a SEQ ID NO: 1), ou um respetivo produto de expressão (por exemplo, um péptido com a SEQ ID NO: 2), numa quantidade eficaz para aumentar a atividade geradora de C_{α} -formilglicina na célula. Em formas de realização importantes, o método envolve ativar o gene FGE endógeno para aumentar a atividade geradora de C_{α} -formilglicina na célula. Nalgumas formas de realização, o contacto é efetuado em condições que permitem a entrada de uma molécula na célula.

O termo "permitir a entrada" de uma molécula numa célula tem os significados seguintes, dependendo da natureza da molécula. Para um ácido nucleico isolado, pretende-se que descreva a entrada do ácido nucleico através da membrana celular e para o núcleo celular, de modo que o "transgene de ácido nucleico" pode utilizar a maquinaria celular para produzir polipéptidos funcionais codificados pelo ácido nucleico. Por "transgene de ácido nucleico" pretende-se descrever todos os ácidos nucleicos descritos aqui, com ou sem os vetores associados. Para um polipéptido, pretende-se que descreva a entrada do polipéptido através da membrana celular e para o citoplasma celular e, se necessário, utilização da maquinaria citoplasmática da célula para modificar funcionalmente o polipéptido (por exemplo, numa forma ativa).

Podem ser empregues várias técnicas para introduzir ácidos nucleicos em células, dependendo de os ácidos nucleicos serem introduzidos *in vitro* ou *in vivo* num hospedeiro. Essas técnicas incluem transfeção de precipitados ácido nucleico- CaPO_4 , transfeção de ácidos nucleicos associados a DEAE, transfeção com um retrovírus incluindo o ácido nucleico de interesse, transfeção mediada por lipossomas, e afins. Para certas utilizações, é preferido dirigir o ácido nucleico para células particulares. Nesses casos, um veículo utilizado para distribuir um ácido nucleico numa célula (por exemplo, um retrovírus, ou outro vírus; um lipossoma) pode estar ligado a uma molécula de direcionamento. Por exemplo, uma molécula como um anticorpo específico para uma proteína membranar de superfície na célula alvo, ou um ligando para um recetor na célula alvo, pode ser ligada ou incorporada no veículo de distribuição do ácido nucleico. Por exemplo, quando forem empregues lipossomas para distribuir os ácidos nucleicos, proteínas que se ligam a uma proteína membranar de superfície associada a endocitose podem ser incorporadas na formulação de lipossoma, para dirigir e/ou facilitar a captação. Essas proteínas incluem proteínas capsídicas ou respetivos fragmentos trópicos para um tipo particular de células, anticorpos para proteínas que sofrem interiorização por ciclicização, proteínas que abordam seletivamente uma localização intracelular e que aumentam a semivida intracelular, e afins. Sistemas de distribuição poliméricos também têm sido utilizados com êxito para distribuir ácidos nucleicos em células, como é conhecido dos profissionais. Esses sistemas permitem mesmo a distribuição oral de ácidos nucleicos.

Outros sistemas de distribuição podem incluir sistemas de distribuição com libertação calendarizada, libertação retardada ou libertação prolongada. Esses sistemas podem evitar administrações repetidas de um agente como descrito aqui, aumentando a conveniência para o sujeito e para o médico. Muitos tipos de sistemas de distribuição com libertação estão disponíveis e são conhecidos dos profissionais. Incluem sistemas de base polimérica, como poli(láctido-glicólido), copolioxalatos, policaprolactonas, poliésteramidas, poliortoésteres, ácido poli-hidroxibutírico e polianidridos. Microcápsulas dos polímeros anteriores contendo fármacos são descritos, por exemplo, na patente U.S. 5,075,109. Sistemas de distribuição também incluem sistemas não poliméricos que são: lípidos incluindo esteróis, como colesterol, ésteres de colesterol e ácidos gordos ou gorduras neutras, como mono-, di- e triglicéridos; sistemas de libertação de hidrogel; sistemas silásticos; sistemas à base de péptidos; revestimentos cerosos; comprimidos preparados utilizando aglutinantes e excipientes convencionais; implantes parcialmente fundidos; e afins. Exemplos específicos incluem mas não estão limitados a: (a) sistemas por erosão, nos quais um agente como descrito aqui está contido numa forma dentro de uma matriz, como as descritas nas Patentes U.S. N°s 4,452,775, 4,675,189 e 5,736,152, e (b) sistemas por difusão, nos quais um componente ativo é permeado, a uma velocidade controlada, a partir de um polímero, como descrito nas Patentes U.S. N°s 3,854,480, 5,133,974 e 5,407,686. Adicionalmente, podem ser utilizados sistemas de distribuição por "hardware" à base de bombas, alguns dos quais estão adaptados para implantação.

Pode ser desejável a utilização de um implante de libertação prolongada de longa duração. Libertação de longa duração, como usado aqui, significa que o implante é construído e disposto para distribuir níveis terapêuticos do ingrediente ativo durante pelo menos 30 dias, e preferivelmente 60 dias. Implantes de libertação prolongada de longa duração são bem conhecidos dos profissionais e incluem alguns dos sistemas de libertação descritos acima. Exemplos específicos incluem mas não estão limitados a implantes de libertação prolongada de longa duração descritos na Patente U.S. N° 4,748,024 e Patente Canadiana N° 1330939.

A administração e, nalgumas formas de realização, a coadministração de agentes podem ser realizadas com agentes diferentes das moléculas FGE descritas aqui de tal modo que, quando administrados em quantidades eficazes, podem atuar de forma cooperativa, aditiva ou sinérgica com uma molécula como descrito aqui para: (i) modular a atividade geradora de C_{α} -formilglicina, e (ii) tratar qualquer um dos estados onde está envolvida a atividade geradora de C_{α} -formilglicina de uma molécula da invenção (por exemplo, uma deficiência de sulfatases, incluindo MSD). Agentes diferentes das moléculas FGE incluem Iduronato 2-Sulfatase, Sulfamidase, N-Acetilgalactosamina 6-Sulfatase, N-Acetilglucosamina 6-Sulfatase, Arilsulfatase A, Arilsulfatase B, Arilsulfatase C, Arilsulfatase D, Arilsulfatase E, Arilsulfatase F, Arilsulfatase G, HSulf-1, HSulf-2, HSulf-3, HSulf-4, HSulf-5 ou HSulf-6 (ácidos nucleicos e polipéptidos, e/ou respectivos fragmentos), e/ou combinações destes.

"Coadministração", como usado aqui, refere-se à administração simultânea de dois ou mais compostos como descrito aqui (por exemplo, um ácido nucleico e/ou polipéptido FGE, e um agente que é conhecido ser benéfico no tratamento, por exemplo, de uma deficiência de sulfatases - por exemplo, Iduronato 2-Sulfatase no tratamento de MPSII), na forma de uma mistura numa única composição, ou sequencial, em instantes suficientemente próximos para que os compostos possam exercer um efeito aditivo ou mesmo sinérgico.

Podem ser proporcionadas séries de moléculas de ácidos nucleicos em fase sólida. A série pode consistir essencialmente num conjunto de moléculas de ácidos nucleicos, respetivos produtos de expressão, ou respetivos fragmentos (da molécula de ácido nucleico ou de polipéptido), em que cada molécula de ácido nucleico é selecionada do grupo que consiste em FGE, Iduronato 2-Sulfatase, Sulfamidase, N-Acetilgalactosamina 6-Sulfatase, N-Acetilglucosamina 6-Sulfatase, Arilsulfatase A, Arilsulfatase B, Arilsulfatase C, Arilsulfatase D, Arilsulfatase E, Arilsulfatase F, Arilsulfatase G, HSulf-1, HSulf-2, HSulf-3, HSulf-4, HSulf-5 e HSulf-6, fixa num substrato sólido. Nalgumas formas de realização, a série em fase sólida também compreende pelo menos uma molécula de ácido nucleico de controlo. Em certas formas de realização, o conjunto de moléculas de ácidos nucleicos compreende pelo menos uma, pelo menos duas, pelo menos três, pelo menos quatro ou mesmo pelo menos cinco moléculas de ácidos nucleicos, cada uma selecionada do grupo que consiste em FGE, Iduronato 2-Sulfatase, Sulfamidase, N-Acetilgalactosamina 6-Sulfatase, N-Acetilglucosamina 6-Sulfatase, Arilsulfatase A, Arilsulfatase B, Arilsulfatase

C, Arilsulfatase D, Arilsulfatase E, Arilsulfatase F, Arilsulfatase G, HSulf-1, HSulf-2, HSulf-3, HSulf-4, HSulf-5 e HSulf-6. Em formas de realização preferidas, o conjunto de moléculas de ácidos nucleicos compreende um número máximo de 100 moléculas de ácidos nucleicos diferentes. Em formas de realização importantes, o conjunto de moléculas de ácidos nucleicos compreende um número máximo de 10 moléculas de ácidos nucleicos diferentes.

A técnica de hibridização padrão da tecnologia de microsséries pode ser utilizada para avaliar padrões de expressão de ácidos nucleicos e identificar a expressão de ácidos nucleicos. A tecnologia de microsséries - que também é conhecida por outros nomes, incluindo: tecnologia de "chips" de DNA, tecnologia de "chips" genéticos e tecnologia de séries de ácidos nucleicos em fase sólida - é bem conhecida dos profissionais e baseia-se, sem assim ficar limitada, na obtenção de uma série de sondas de ácidos nucleicos identificados (por exemplo, moléculas descritas aqui noutra local, tais como FGE, Iduronato 2-Sulfatase, Sulfamidase, N-Acetilgalactosamina 6-Sulfatase, N-Acetilglucosamina 6-Sulfatase, Arilsulfatase A, Arilsulfatase B, Arilsulfatase C, Arilsulfatase D, Arilsulfatase E, Arilsulfatase P, Arilsulfatase G, HSulf-1, HSulf-2, HSulf-3, HSulf-4, HSulf-5 e/ou HSulf-6) num substrato fixo, etiquetagem de moléculas alvo com moléculas repórter (por exemplo, marcadores radioativos, quimioluminescentes ou fluorescentes, como fluoresceína, Cy3-dLTP ou Cy5-dUTP), hibridização de ácidos nucleicos alvo com as sondas e avaliação da hibridização alvo-sonda. Uma sonda com uma sequência de ácido nucleico que corresponde perfeitamente à sequência alvo originará, em geral, detecção de um sinal mais forte da molécula repórter

do que sondas com correspondências menos perfeitas. Muitos componentes e técnicas utilizados na tecnologia de microsséries de ácidos nucleicos são apresentados em "The Chipping Forecast", *Nature Genetics*, Volume 21, janeiro de 1999.

Substratos para microsséries podem incluir mas não estão limitados a vidro, sílica, aluminossilicatos, borossilicatos, óxidos metálicos, como alumina e óxido de níquel, várias argilas, nitrocelulose ou *nylon*. Em todas as formas de realização é preferido um substrato de vidro. Sondas podem ser selecionadas do grupo de ácidos nucleicos incluindo mas não se limitando aos seguintes: DNA, DNA genómico, cDNA e oligonucleótidos; e podem ser naturais ou sintéticas. Sondas oligonucleotídicas consistem preferivelmente em oligonucleótidos 20 até 25-meros, e sondas de DNA/cDNA têm preferivelmente 500 até 5000 bases de comprimento, apesar de poderem ser utilizados outros comprimentos. O comprimento apropriado de sondas pode ser determinado pelo profissional seguindo procedimentos conhecidos na área. Numa forma de realização, sondas preferidas consistem em conjuntos de duas ou mais das moléculas de ácidos nucleicos apresentadas como SEQ ID NOs: 1, 3, 4, 6, 8, 10 e/ou 12. As sondas podem ser purificadas, para remover contaminantes, utilizando métodos comuns conhecidos dos profissionais, como filtração em gel ou precipitação.

Numa forma de realização, o substrato da microssérie pode ser revestido com um composto para aumentar a síntese da sonda no substrato. Esses compostos incluem mas não estão limitados a oligoetilenoglicóis. Noutra forma de realização, agentes ou grupos de acoplamento presentes no

substrato podem ser utilizados para ligar de modo covalente o primeiro nucleótido ou oligonucleótido ao substrato. Estes agentes ou grupos podem incluir mas não estão limitados aos seguintes: grupos amino, hidroxí, bromo e carboxi. Estes grupos reativos são preferivelmente ligados ao substrato através de um radical hidrocarbilo, como um radical divalente alquilenos ou fenileno, com uma posição da valência ocupada pela ligação da cadeia e o restante ligado aos grupos reativos. Estes grupos hidrocarbilo podem conter até cerca de dez átomos de carbono, preferivelmente até cerca de seis átomos de carbono. São habitualmente preferidos radicais alquilenos contendo dois até quatro átomos de carbono na cadeia principal. Estes e pormenores adicionais do processo são divulgados, por exemplo, na Patente U.S. 4,458,066.

Numa forma de realização, sondas são sintetizadas diretamente no substrato, num padrão de grelha predeterminado, utilizando métodos tais como síntese química dirigida por luz, desproteção fotoquímica ou distribuição de precursores nucleotídicos no substrato e subsequente produção de sondas.

Noutra forma de realização, o substrato pode ser revestido com um composto para aumentar a ligação da sonda ao substrato. Esses compostos incluem mas não estão limitados aos seguintes: polilisina, aminossilanos, silanos reativos com amino ("Chipping Forecast", 1999) ou crómio (Gwyne e Page, 2000). Nesta forma de realização, sondas pré-sintetizadas são aplicadas no substrato num volume e padrão de grelha precisos e predeterminados, utilizando um autómato controlado por computador para aplicar a sonda no substrato de um modo de impressão por contacto ou de um

modo sem contacto, como distribuição por jato de tinta ou piezoelétrica. Sondas podem ser ligadas ao substrato de modo covalente com métodos que incluem mas não estão limitados a irradiação de UV. Noutra forma de realização, sondas são ligadas ao substrato por ação de calor.

Alvos são ácidos nucleicos selecionados do grupo que inclui mas não está limitado aos seguintes: DNA, DNA genómico, cDNA, RNA, mRNA, e podem ser naturais ou sintéticos. Em todas as formas de realização são preferidas moléculas de ácidos nucleicos de sujeitos suspeitos de desenvolverem ou de sofrerem de uma deficiência de sulfatases. Em certas formas de realização, uma ou mais moléculas de ácidos nucleicos de controlo podem ser ligadas ao substrato. Preferivelmente, moléculas de ácidos nucleicos de controlo permitem a determinação de fatores incluindo mas não se limitando aos seguintes: qualidade e características de ligação de ácidos nucleicos; qualidade e eficácia de reagentes; sucesso da hibridização, e limites e sucesso da análise. Ácidos nucleicos de controlo podem incluir mas não estão limitados a produtos de expressão de genes, como genes domésticos, ou respetivos fragmentos.

Para seleccionar um conjunto de marcadores de doença de deficiência de sulfatases, os dados de expressão gerados, por exemplo, por análise de microsséries da expressão genética são preferivelmente analisados para determinar quais os genes em diferentes categorias de pacientes (em que cada categoria de pacientes consiste numa perturbação diferente de deficiência de sulfatases) que são significativamente expressos de modo diferencial. A significância da expressão genética pode ser determinada utilizando "software" computacional Permax, apesar de poder

ser utilizado qualquer pacote estatístico comum que consiga distinguir diferenças significativas da expressão. Permax realiza testes t de 2 amostras de permutação em grandes séries de dados. Para vetores de observações de dimensionalidade mais elevada, o "software" Permax calcula estatística t para cada atributo e avalia a significância utilizando a distribuição de permutação dos atributos globais máximo e mínimo. A utilização principal consiste em determinar os atributos (genes) mais diferentes entre dois grupos (por exemplo, sujeito saudável de controlo e um sujeito com uma deficiência de sulfatase particular), medindo os "mais diferentes" utilizando o valor da estatística t, e seus níveis de significância.

A expressão de moléculas de ácidos nucleicos relacionadas com doença de deficiência de sulfatases também pode ser determinada utilizando métodos de medições de proteínas para determinar a expressão da SEQ ID NO: 2, por exemplo, determinando a expressão de polipéptidos codificados pelas SEQ ID NOS: 1 e/ou 3. Métodos preferidos para medir proteínas específica e quantitativamente incluem mas não estão limitados aos seguintes: métodos à base de espectroscopia de massa, como desorção ionização por laser realçado de superfície (SELDI; por exemplo, Ciphergen ProteinChip System), métodos não baseados em espectroscopia de massa e métodos à base de imuno-histoquímica, como eletroforese em gel 2-dimensional.

A metodologia SELDI pode ser utilizada, através de procedimentos conhecidos dos profissionais, para vaporizar quantidades microscópicas de proteína e para criar uma "impressão digital" de proteínas individuais, desse modo permitindo a medição simultânea da abundância de muitas

proteínas numa única amostra. Preferivelmente, ensaios à base de SELDI podem ser utilizados para caracterizar deficiência de múltiplas sulfatases, bem como fases desses estados. Preferivelmente, esses ensaios incluem mas não estão limitados aos seguintes exemplos. Produtos genéticos descobertos por microsséries de RNA podem ser seletivamente medidos por captura específica (mediada por anticorpos) para o disco de proteínas SELDI (por exemplo, SELDI seletiva). Produtos genéticos descobertos por rastreamento de proteínas (por exemplo, com géis 2-D) podem ser resolvidos por "SELDI proteica total" otimizada para visualizar aqueles marcadores de interesse particulares de entre as SEQ ID NOs: 1, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26 e/ou 28. Modelos de previsão de uma deficiência de sulfatases específica a partir da medição SELDI de múltiplos marcadores de entre as SEQ ID NOs: 1, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26 e/ou 28 podem ser utilizados para as estratégias de SELDI.

A utilização de quaisquer dos métodos de microsséries anteriores para determinar a expressão de ácidos nucleicos relacionados com uma doença de deficiência de sulfatases pode ser feita com métodos de rotina conhecidos dos profissionais, e a expressão pode ser determinada por métodos de medição de proteínas, pode ser correlacionada com níveis predeterminados de um marcador utilizados como método de prognóstico para selecionar estratégias de tratamento para pacientes com doença de deficiência de sulfatases.

Pode ser proporcionada uma célula produtora de sulfatase, em que a razão entre sulfatase ativa e sulfatase total produzida (isto é, a atividade específica) pela célula é

aumentada. A célula compreende: (i) uma sulfatase com uma expressão aumentada, e (ii) uma Enzima Geradora de Formilglicina com uma expressão aumentada, em que a razão entre sulfatase ativa e sulfatase total produzida pela célula é aumentada em pelo menos 5% relativamente à razão entre sulfatase ativa e sulfatase total produzida pela célula na ausência da Enzima Geradora de Formilglicina.

Uma "sulfatase com uma expressão aumentada", como usado aqui, refere-se tipicamente a expressão aumentada de uma sulfatase e/ou do seu polipéptido codificado em comparação com um controlo. Expressão aumentada refere-se ao aumento (isto é, numa extensão detetável) da replicação, transcrição e/ou tradução de quaisquer dos ácidos nucleicos de sulfatases (ácidos nucleicos de sulfatases como descrito aqui noutro local), uma vez que a regulação positiva de quaisquer destes processos origina um aumento da concentração/quantidade do polipéptido codificado pelo gene (ácido nucleico). Isto pode ser conseguido utilizando alguns métodos conhecidos na área, também descritos aqui noutro local, como transfeção de uma célula com o cDNA da sulfatase e/ou DNA genómico abrangendo o *locus* da sulfatase, ativação do gene de sulfatase endógeno por colocação, por exemplo, de um elemento promotor forte a montante do *locus* genómico do gene de sulfatase endógeno utilizando recombinação homóloga (ver, por exemplo, a tecnologia de ativação genética descrita em pormenor nas Patentes U.S. N°s 5,733,761, 6,270,989 e 6,565,844), etc. Um controlo típico consistirá numa célula idêntica transfetada com plasmídeo(s) vetorial(vetoriais). Intensificar (ou aumentar) a atividade de sulfatase também se refere a prevenir ou inibir a degradação (por exemplo, via ubiquitinação), regulação negativa, etc. de sulfatases,

originando, por exemplo, um $t\frac{1}{2}$ (semivida) da molécula de sulfatase aumentado ou estável em comparação com um controlo. Regulação negativa ou expressão decrescida refere-se a expressão decrescida de um gene e/ou do seu polipéptido codificado. A regulação positiva ou regulação negativa da expressão genética pode ser determinada diretamente detetando um aumento ou decréscimo, respetivamente, do nível de mRNA para o gene (por exemplo, uma sulfatase), ou o nível de expressão proteica do polipéptido codificado pelo gene, utilizando quaisquer meios adequados conhecidos na área, como métodos de hibridização de ácidos nucleicos ou de deteção com anticorpos, respetivamente, e em comparação com controlos. A regulação positiva ou regulação negativa da expressão de genes de sulfatases também pode ser determinada indiretamente por deteção de uma alteração da atividade de sulfatase.

De modo semelhante, uma "Enzima Geradora de Formilglicina com expressão aumentada", como usado aqui, refere-se tipicamente a expressão aumentada de um ácido nucleico FGE e/ou do seu polipéptido codificado em comparação com um controlo. Expressão aumentada refere-se ao aumento (isto é, numa extensão detetável) da replicação, transcrição e/ou tradução de quaisquer dos ácidos nucleicos FGE (descrito aqui noutro local), uma vez que a regulação positiva de quaisquer destes processos origina um aumento da concentração/quantidade do polipéptido codificado pelo gene (ácido nucleico). Isto pode ser conseguido utilizando os métodos descritos acima (para as sulfatases), e aqui noutro local.

Em certas formas de realização, a razão entre sulfatase ativa e sulfatase total produzida pela célula é aumentada em pelo menos 10%, 15%, 20%, 50%, 100%, 200%, 500%, 1000%, relativamente à razão entre sulfatase ativa e sulfatase total produzida pela célula na ausência da Enzima geradora de Formilglicina.

Pode ser proporcionado um método aperfeiçoado para tratar uma deficiência de sulfatases num sujeito. O método envolve administrar, a um sujeito necessitado desse tratamento, uma sulfatase numa quantidade eficaz para tratar a deficiência de sulfatases no sujeito, em que a sulfatase é contactada com uma Enzima Geradora de Formilglicina numa quantidade eficaz para aumentar a atividade específica da sulfatase. Como descrito aqui noutro local, "atividade específica" refere-se à razão produzida entre sulfatase ativa e sulfatase total. "Contactado", como usado aqui, refere-se à modificação pós-tradução com FGE da sulfatase como descrito aqui noutro local. Será claro para o profissional que uma FGE pode contactar uma sulfatase e modificá-la se ácidos nucleicos que codificam FGE e uma sulfatase forem coexpressos numa célula, ou mesmo se um polipéptido FGE isolado contactar um polipéptido sulfatase isolado *in vivo* ou *in vitro*. Mesmo apesar de um polipéptido FGE isolado poder ser coadministrado com um polipéptido sulfatase isolado a um sujeito para tratar uma deficiência de sulfatases no sujeito, é preferido que o contacto entre FGE e a sulfatase ocorra *in vitro* antes da administração da sulfatase ao sujeito. Este método de tratamento aperfeiçoado é benéfico para um sujeito, pois é necessário administrar quantidades menores da sulfatase e/ou com menos frequência, uma vez que a sulfatase tem maior atividade específica.

A invenção será mais completamente compreendida por referência aos exemplos seguintes. No entanto, pretende-se meramente que estes exemplos ilustrem as formas de realização da invenção e que não sejam considerados limitadores do âmbito da invenção.

Exemplos

Exemplo 1:

A Deficiência de Múltiplas Sulfatases é causada por mutações no gene que codifica a enzima geradora de C α -formilglicina (FGE) humana

Procedimentos Experimentais

Materiais e Métodos

Ensaio *in vitro* para FGE

Para monitorizar a atividade da FGE, o péptido de 23-meros N-acetilado e C-amidado P23 (MTDFYVPVSLCTPSRAALLTGRS) (SEQ ID NO: 33) foi utilizado como substrato. A conversão do resíduo Cisteína na posição 11 em FGly foi monitorizada por espetrometria de massa MALDI-TOP. Preparou-se uma solução-mãe 6 μ M de P23 em 30% acetonitrilo e 0,1% ácido trifluoroacético (TFA). Em condições padrão, incubaram-se 6 pmol de P23 a 37°C com até 10 μ L de enzima num volume final de 30 μ L 50 mM Tris/HCl, pH 9,0, contendo 67 mM NaCl, 15 μ M CaCl₂, 2 mM DTT e 0,33 mg/mL de albumina do soro bovino. Para parar a reação enzimática adicionaram-se 1,5 μ L de 10% TFA. Em seguida, P23 foi ligado a ZipTip C18 (Millipore), foi lavado com 0,1% TFA e eluiu em 3 μ L de 50%

acetonitrilo, 0,1% TFA. Misturaram-se 0,5 µL do eluato com 0,5 µL de solução matricial (5 mg/mL de ácido α -ciano-4-hidroxicinâmico (Bruker Daltonics, Billerica, MA) em 50% acetonitrilo, 0,1% TFA) num alvo de aço inoxidável. Realizou-se espectrometria de massa MALDI-TOF com um Reflex III (Bruker Daltonics) utilizando o modo de refletor e energia laser imediatamente acima do limite de desorção/ionização. Todos os espectros consistiram em médias de 200-300 disparos de várias manchas no alvo. O eixo da massa foi calibrado utilizando péptidos de massas moleculares variando desde 1000 até 3000 Da como padrões externos. O MH⁺ monoisotópico do P23 é 2526,28 e o do produto contendo FGly é 2508,29. A atividade (pmol de produto / h) foi calculada com base na altura do pico do produto dividida pela soma das alturas dos picos de P23 e do produto.

Purificação de FGE de testículo bovino

Obtiveram-se testículos de bovino do matadouro local, tendo sido armazenados em gelo durante um período até 20 horas. O parênquima foi libertado do tecido conjuntivo e homogeneizado num misturador Waring e por três rondas no recipiente de motor. A preparação de microssomas grosseiros (RM) por fracionamento celular do homogenato obtido foi realizada como descrito (Meyer *et al.*, *J. Biol. Chem.*, 2000, 275: 14550-14557) com as seguintes modificações. Realizaram-se três passos de centrifugação diferencial, 20 minutos cada a 4°C, a 500 g (rotor JA10), 3000 g (JA10) e 10000 g (JA20). Do último sobrenadante, as membranas de RM foram sedimentadas (125000 g, rotor Ti45, 45 minutos, 4°C), homogeneizadas no recipiente de motor e dispostas em camadas numa almofada de sucrose (50 mM Hepes, pH 7,6, 50 mM KAc, 6 mM MgAc₂, 1 mM EDTA, 1,3 M sucrose, 5 mM β -

mercaptoetanol). Os RMs foram recuperados do grânulo após rotação durante 210 minutos a 45000 rpm num rotor Ti45 a 4°C. Habitualmente obtiveram-se 100000-150000 equivalentes de RM, como definido por Walter e Blobel (*Methods Enzymol.*, 1983, 96: 84-93), a partir de 1 kg de tecido de testículo. O reticuloplasma, isto é, o conteúdo luminal do RM, foi obtido por extração diferencial a baixas concentrações de desoxi Big Chap, como descrito (Fey et al., *J. Biol. Chem.*, 2001, 276: 47021-47028). Para a purificação de FGE, 95 mL de reticuloplasma foram dialisados durante 20 horas a 4 °C contra 20 mM Tris/HCl, pH 8,0, 2,5 mM DTT, e foram clarificados por centrifugação a 125000 g durante 1 hora. Alíquotas de 32 mL do reticuloplasma clarificado foram carregadas numa coluna MonoQ HR10/10 (Amersham Biosciences, Piscataway, NJ) à temperatura ambiente, foram lavadas e eluíram a 2 mL/minuto com um gradiente linear de 0 até 0,75 M NaCl em 80 mL do tampão Tris. As frações contendo atividade de FGE, eluindo a 50-165 mM NaCl, de três rondas foram reunidas (42 mL) e misturadas com 2 mL de Concanavalina A-Sefarose (Amersham Biosciences) que tinham sido lavados com 50 mM tampão Hepes, pH 7,4, contendo 0,5 M KCl, 1 mM MgCl₂, 1 mM MnCl₂, 1 mM CaCl₂ e 2,5 mM DTT. Após incubação durante 16 horas a 4 °C, a Concanavalina A-Sefarose foi recolhida numa coluna e lavada com 6 mL do mesmo tampão Hepes. O material ligado eluiu por incubação da coluna durante 1 hora à temperatura ambiente com 6 mL de 0,5 M a-metilmanósido em 50 mM Hepes, pH 7,4, 2,5 mM DTT. A eluição foi repetida com 4 mL do mesmo eluente. Os eluatos combinados (10 mL) de Concanavalina A-Sefarose foram ajustados para pH 8,0 com 0,5 M Tris/HCl, pH 9,0, e foram misturados com 2 mL de Affigel 10 (Bio-Rad Laboratories, Hercules, CA) que tinham sido derivatizados com 10 mg do péptido misturado (PVSLPTRSCAALLTGR) (SEQ ID NO: 34) e

lavados com tampão A (50 mM Hepes, pH 8,0, contendo 0,15 M acetato de potássio, 0,125 M sucrose, 1 mM MgCl₂ e 2,5 mM DTT). Após incubação durante 3 horas a 4 °C recolheu-se a matriz de afinidade numa coluna. O fluxo e uma fração de lavagem com 4 mL de tampão A foram recolhidos, combinados e misturados com 2 mL de Affigel 10 que tinha sido substituído com 10 mg do péptido Ser69 (PVSLSTPSRAALLTGR) (SEQ ID NO: 35) e lavado com tampão A. Após incubação durante a noite a 4°C, a matriz de afinidade foi recolhida numa coluna, foi lavada 3 vezes com 6 mL de tampão B (tampão A contendo 2 M NaCl e uma mistura dos 20 aminoácidos proteinogénicos, cada um a 50 mg/mL). O material ligado eluiu da matriz de afinidade por incubação do Affigel duas vezes, durante 90 minutos cada, com 6 mL de tampão B contendo 25 mM péptido Ser69. Uma alíquota do eluato foi substituída com 1 mg/mL de albumina do soro bovino, foi dialisada contra tampão A e analisada quanto à afinidade. A parte remanescente da atividade (11,8 mL) foi concentrada num concentrador Vivaspin 500 (Vivascience AG, Hanôver, Alemanha), e solubilizada a 95 °C em tampão de amostra Laemmli SDS. A composição polipeptídica do material de partida e preparações obtidas após os passos cromatográficos foi monitorizada por SDS-PAGE (15% acrilamida, 0,16% bisacrilamida) e coloração com SYPRO Ruby (Bio-Rad Laboratories).

Identificação de FGE por espectrometria de massa

Para análise das impressões digitais de massa peptídicas, os polipéptidos purificados foram digeridos *in gel* com tripsina (Shevchenko *et al.*, *Anal. Chem.*, 1996, 68: 850-855), dessalinizados em C18 ZipTip e analisados por espectrometria de massa MALDI-TOF, utilizando como matriz

ácido di-hidrobenczoico e como padrões internos dois péptidos autolíticos de tripsina (m/z 842,51 e 2211,10). Para a análise de espetrometria de massa em série, os péptidos selecionados foram analisados por espetrometria de massa com decaimento pós-fonte MALDI-TOF. Os seus iões correspondentes duplamente carregados foram isolados e fragmentados por espetrometria de massa com armadilha de iões nano-ESI em diferido (EsquireLC, Bruker Daltonics). Os dados espetrométricos de massa foram utilizados pelo algoritmo de pesquisa Mascot para a identificação de proteínas na base de dados de proteínas NCBI nr e na base de dados de nucleótidos NCBI EST.

Bioinformática

Péptidos de sinal e sítios de clivagem foram descritos com o método de von Heijne (von Heijne, *Nucleic Acids Res.*, 1986, 14: 4683-90) implementado em EMBOSS (Rice et al., *Trends in Genetics*, 2000, 16: 276-277). Sítios de N-glicosilação foram previstos utilizando o algoritmo de Brunak (Gupta e Brunak, *Pac. Symp. Biocomput.*, 2002, 310-22). Detetaram-se domínios funcionais por busca em PFAM-Hidden-Markov-Models (versão 7.8) (Sonnhammer et al., *Nucleic Acids Res.*, 1998, 26: 320-322). Para pesquisar homólogos de FGE, fizeram-se buscas nas bases de dados do National Center for Biotechnology Information (Wheeler et al., *Nucleic Acids Res.*, 2002, 20: 13-16) com BLAST (Altschul et al., *Nucleic Acids Res.*, 1997, 25: 3389-3402). Calcularam-se semelhanças de sequências utilizando ferramentas comuns de EMBOSS. A organização e sintenia de *loci* genómicos foram determinadas utilizando os recursos dos genomas humano e de ratinho do NCBI e o Mapa de Homologia Humano-Ratinho também do NCBI, Bethesda, MD).

Clonagem de cDNA de FGE humana

RNA total, preparado a partir de fibroblastos humanos utilizando o estojo RNEASY™ Mini (Qiagen, Inc., Valencia, CA) foi transcrito de forma reversa utilizando o estojo OMNISCRIPT RT™ (Qiagen, Inc., Valencia, CA) e um "primer" oligo(dT) ou o "primer" específico para FGE 1199nc (CCAATGTAGGTCAGACAAG) (SEQ ID NO: 36). O primeiro filamento de cDNA foi amplificado por PCR utilizando o "primer" direto 1c (ACATGGCCCGCGGGAC) (SEQ ID NO: 37) e, como "primer" reverso, 1199nc ou 1182nc (CGACTGCTCCTTGGACTGG) (SEQ ID NO: 38). Os produtos de PCR foram clonados diretamente no vetor pCR4-TOPO™ (Invitrogen Corporation, Carlsbad, CA). Por sequenciação de múltiplos dos produtos de PCR clonados, que haviam sido obtidos de vários indivíduos e de reações de RT e PCR independentes, determinou-se a sequência codificadora do cDNA da FGE (SEQ ID NOs: 1 e 3).

Deteção de mutações, sequenciação genómica, mutagénese dirigida a sítios e análise de coloração "Northern"

Os protocolos padrão utilizados neste estudo foram essencialmente como descritos em Lübke *et al.* (Nat. Gen., 2001, 28: 73-76) e Hansske *et al.* (J. Clin. Invest., 2002, 109: 725-733). As colorações "Northern" foram hibridizadas com uma sonda de cDNA abrangendo a região codificadora completa e uma sonda de cDNA de β -actina como controlo da carga de RNA.

Linhas de células e cultura de células

Os fibroblastos dos pacientes com MSD 1-6 foram obtidos de E. Christenson (Rigshospitalet Copenhaga), M. Beck

(Universitätskinderklinik Mainz), A. Kohlschütter (Universitätskrankenhaus Eppendorf, Hamburgo), E. Zammarchi (Hospital Meyer, Universidade de Florença), K. Harzer (Institut für Hirnforschung, Universität Tübingen), e A. Fensom (Guy's Hospital, Londres), respectivamente. Fibroblastos de pele humana, células HT-1080, BHK21 e CHO foram mantidos a 37°C sob 5% CO₂ em meio de Eagle modificado de Dulbecco contendo 10% soro fetal de vitelo.

Transfeção, imunofluorescência indireta, análise de coloração "Western" e detecção da atividade de FGE

O cDNA da FGE foi equipado com um sítio 5' EcoRI e uma sequência marcadora 3' HA, c-Myc ou RGS-His₆, seguida de um codão de terminação e um sítio HindIII, por PCR "add-on" utilizando Pfu polimerase (Stratagene, La Jolla, CA) e os "primers" seguintes: GGAATTCGGGACAACATGGCTGCG (EcoRI) (SEQ ID NO: 39), CCCAAGCTTATGC GTAGTCAGGCACATCATACGGATAGTCCATGGTGGGCAGGC(HA) (SEQ ID NO: 40), CCCAAGCTTACAGGTCTT-CTTCAGAAATCAGCTTTTGTTCGTCCATGGTGGGCAG GC (c-Myc) (c-Myc) (SEQ ID NO: 41), CCCAAGCTTAGTGTATGGT-GATGGTGATGCGATC CTCTGTCCATGGTGGGCAGGC (RGS-His₆) (SEQ ID NO:42). Os produtos de PCR resultantes foram clonados, na forma de fragmentos EcoRI/HindIII, em pMPSVEH (Artelt *et al.*, *Gene*, 1988, 68: 213-219). Os plasmídeos obtidos foram transfetados de modo transitório em células HT-1080, BHK21 e CHO, cresceram em lamelas de cobertura utilizando EFFECTENE™ (Qiagen) como reagente de transfeção. Quarenta e oito horas após a transfeção, as células foram analisadas por imunofluorescência indireta como descrito previamente (Lübke *et al.*, *Nat. Gen.*, 2001, 28: 73-76; Hansske *et al.*, *J. Clin. Invest.*, 2002, 109: 725-733), utilizando anticorpos monoclonais IgG1 contra HA (Berkeley Antibody

Company, Richmond, CA), c-Myc (Santa Cruz Biotechnology, Inc., Santa Cruz, CA) ou RGS-His (Qiagen) como anticorpos primários. A proteína marcadora do retículo endoplasmático proteína dissulfureto isomerase (PDI) foi detetada com um anticorpo monoclonal de subtipo diferente (IgG2A, Stressgen Biotech., Victoria BC, Canadá). Os anticorpos primários foram detetados com anticorpos secundários de cabra específicos para o isotipo acoplados a CY2 ou CY3, respetivamente (Molecular Probes, Inc., Eugene, OR). Obtiveram-se imagens de imunofluorescência num microscópio de varrimento laser Leica TCS Sp2 AOBS. Para a análise de coloração "Western" utilizaram-se os mesmos anticorpos monoclonais e uma IgG anti-ratinho conjugada a HRP como anticorpo secundário. Para determinação da atividade de FGE, as células submetidas a tripsinação foram lavadas com solução salina tamponada com fosfato contendo uma mistura de inibidores de proteinases (208 μ M cloridrato de fluoreto de 4-(2-aminoetil)benzenossulfonilo, 0,16 μ M aprotinina, 4,2 μ M leupeptina, 7,2 μ M bestatina, 3 μ M pepstatina A, 2,8 μ M E-64), solubilizados em 10 mM Tris, pH 8,0, contendo 2,5 mM DTT, os inibidores de proteinases e 1% Triton X-100, e foram clarificadas por centrifugação a 125 000 g durante 1 hora. O sobrenadante foi sujeito a cromatografia numa coluna MonoQ PC 1.6/5 usando as condições descritas acima. As frações que eluíram a 50-200 mM NaCl foram reunidas, liofilizadas e reconstituídas num décimo do volume original da reunião antes da determinação da atividade da FGE com o péptido P23.

Transdução retroviral

Os cDNAs de interesse foram clonados no vetor à base do vírus da leucemia murina de Moloney pLPCX e pLNCX2 (BD

Biosciences Clontech, Palo Alto, CA). A transfeção de células FNX-Eco ecotrópicas (ATCC, Manassas, VA) e a transdução de células RETROPACK™ PT67 anfotrópicas (BD Biosciences Clontech) e de fibroblastos humanos foram realizadas como descrito (Lübke et al., *Nat. Gen.*, 2001, 28: 73-76; Thiel et al., *Biochem. J.*, 2002, 376, 195-201). Para algumas experiências, células PT67 transduzidas com pLPCX foram selecionadas com puromicina antes da determinação das atividades de sulfatases.

Ensaio de sulfatases

A atividade de ASA, STS e GalNAc6S foi determinada como descrito em Rommerskirch e von Figura, *Proc. Natl. Acad. Sci., USA*, 1992, 89: 2561-2565; Glössl e Kresse, *Clin. Chim. Acta*, 1978, 88: 111-119.

Resultados

Um ensaio rápido à base de péptidos para atividade de FGE

Desenvolvemos um ensaio para determinar a atividade da FGE em extratos de microssomas utilizando como substrato fragmentos [³⁵S] ASA sintetizados *in vitro*. Os fragmentos foram adicionados à mistura de ensaio na forma de complexos de cadeia nascente associados a ribossomas. A quantificação do produto incluiu digestão triptica, separação dos péptidos por RP-HPLC e identificação e quantificação do péptido triptico contendo FGly etiquetado com [³⁵S] por uma combinação de derivatização química em hidrazonas, separação por RP-HPLC e contagem de cintilações líquidas (Fey et al., *J. Biol. Chem.*, 2001, 276: 47021-47028). Para monitorizar a atividade enzimática durante a purificação

foi necessário modificar este procedimento pesado. Um péptido sintético de 16-meros correspondente aos resíduos 65-80 de ASA e contendo o motivo de sequência requerido para a formação de *FGly* inibiu a atividade da FGE no ensaio *in vitro*. Isto sugeriu que péptidos tais como ASA65-80 podem servir de substratos para a FGE. Sintetizámos o péptido de 23-meros P23 (SEQ ID NO: 33), que corresponde aos resíduos 60-80 de ASA com uma metionina N-acetilada adicional e um resíduo serina C-amidado para proteger o terminal N e C, respetivamente. A cisteína e as formas contendo *FGly* de P23 puderam ser identificadas e quantificadas por espetrometria de massa tempo de voo de desorção/ionização com laser assistido por matriz (MALDI-TOF). A presença do resíduo *FGly* na posição 11 do P23 foi verificada por espetrometria de massa com decaimento pós-fonte MALDI-TOF (ver Peng *et al.*, *J. Mass Spec.*, 2003, 38: 80-86). A incubação de P23 com extratos de microssomas de pâncreas bovino ou testículo bovino converteu até 95% do péptido num derivado contendo *FGly* (Fig. 1). Em condições padrão, a reação foi proporcional à quantidade de enzima e tempo de incubação, desde que menos de 50% do substrato tivesse sido consumido e o período de incubação não excedesse 24 horas. A k_m para o P23 foi 13 nM. Os efeitos de glutatona reduzida e oxidada, Ca^{2+} e pH foram comparáveis aos observados no ensaio utilizando como substrato complexos de cadeia nascente associados a ribossomas (Fey *et al.*, *J. Biol. Chem.*, 2001, 276: 47021-47028).

Purificação de FGE

Para a purificação da FGE, a fração solúvel (reticuloplasma) de microssomas de testículo bovino serviu

de material de partida. A atividade específica da FGE foi 10-20 vezes mais elevada do que no reticuloplasma de microssomas de pâncreas bovino (Fey *et al.*, *J. Biol. Chem.*, 2001, 276: 47021-47028). A FGE foi purificada por uma combinação de quatro passos cromatográficos. Os primeiros dois passos consistiram em cromatografia num permutador aniônico MonoQ e em Concanavalina A-Sefarose. A pH 8, a atividade de FGE ligou-se a MonoQ e foi eluída a 50-165 mM NaCl com recuperação de 60-90%. Quando esta fração foi misturada com Concanavalina A-Sefarose, a FGE foi ligada. Foi possível eluir 30-40% da atividade de partida com 0,5 M α -metilmanósido. Os dois passos de purificação finais consistiram em cromatografia em matrizes de afinidade derivatizadas com péptidos de 16-meros. A primeira matriz de afinidade consistiu em Affigel 10 substituído com uma variante do péptido ASA65-80, na qual os resíduos Cys69, Pro71 e Arg73, críticos para a formação de *FGly*, foram misturados (péptido misturado **PVSLPTRSCAALL-TGR** - SEQ ID NO: 34). Este péptido não inibiu a atividade da FGE quando adicionado ao ensaio *in vitro* a uma concentração de 10 mM e, quando imobilizado em Affigel 10, não reteve a atividade de FGE. Cromatografia realizada na matriz de afinidade do péptido misturado removeu proteínas de ligação ao péptido, incluindo "chaperones" do retículo endoplasmático. A segunda matriz de afinidade consistiu em Affigel 10 substituído com uma variante do péptido ASA65-80, na qual o Cys69 foi substituído por uma serina (péptido Ser69 **PVSLSTPSRAALLTGR** - SEQ ID NO: 35). A matriz de afinidade do péptido Ser69 ligou FGE de modo eficiente. Foi possível eluir a atividade de FGE com 2 M KSCN ou 25 mM péptido Ser69 com recuperação de 20-40%. Antes da determinação da atividade foi necessário remover por diálise o KSCN ou péptido Ser69. A substituição de Cys69 por serina foi

crucial para a eluição de FGE ativa. Affigel 10 substituído com o péptido ASA65-80 de tipo selvagem ligou FGE de modo eficiente. No entanto, não foi possível recuperar atividade quase nenhuma em eluatos com sais caotrópicos (KSCN, MgCl₂), péptidos (péptido ASA65-80 ou Ser69) ou tampões com pH baixo ou alto. Na Fig. 2 apresenta-se o padrão polipeptídico do material de partida e das frações ativas obtidas após os quatro passos cromatográficos de uma purificação típica. Na fração final recuperaram-se 5% da atividade de FGE de partida e 0,0006% da proteína de partida (purificação de 8333 vezes).

Os polipéptidos purificados de 39,5 e 41,5 kDa são codificados por um único gene

Os polipéptidos de 39,5 e 41,5 kDa presentes na preparação de FGE purificada foram sujeitos a análise de impressões digitais de massa peptídicas. Os espectros de massa dos péptidos trípticos dos dois polipéptidos obtidos por espectrometria de massa MALDI-TOF apresentaram grande sobreposição, sugerindo que as duas proteínas são originárias do mesmo gene. Entre os péptidos trípticos dos dois polipéptidos encontraram-se dois péptidos abundantes MH⁺ 1580,73, ~~SQNTPDSSASNLGFR~~ (SEQ ID NO: 43), e MH⁺ 2049,91, ~~MVPFAGVFTMGTDPPQIK~~ - SEQ ID NO: 44 mais duas oxidações de metionina), que corresponderam à proteína codificada por um cDNA com o Número de Acesso da GenBank AK075459 (SEQ ID NO: 4). A sequência de aminoácidos dos dois péptidos foi confirmada por espectros com decaimento pós-fonte MALDI-TOF e por análise MS/MS utilizando espectrometria de massa com armadilha de iões com ionização por nanoeletropulverização (ESI) em diferido. Uma sequência EST do ortólogo bovino do cDNA humano abrangendo a parte C-terminal da FGE e

correspondendo às sequências de ambos os péptidos proporcionou informação de sequência adicional para a FGE bovina.

Conservação evolutiva e estrutura de domínios da FGE

O gene da FGE humana é codificado pelo cDNA da (SEQ ID NOs: 1 e/ou 3) e está localizado no cromossoma 3p26. Abrange ~105 kb e a sequência codificadora está distribuída por 9 exões. Encontram-se três ortólogos do gene FGE humano em ratinho (87% de identidade), *Drosophila melanogaster* (48% de identidade) e *Anopheles gambiae* (47% de identidade). Encontram-se sequências EST ortólogas para mais 8 espécies, incluindo vaca, porco, *Xenopus laevis*, *Silurana tropicalis*, peixe zebra, salmão e outras espécies de peixe (quanto a pormenores ver o Exemplo 2). A estrutura exão-intrão entre o gene humano e o de ratinho está conservada e o gene de ratinho no cromossoma 6E2 está localizado numa região sinténica com o cromossoma humano 3p26. Os genomas de *S. cerevisiae* e *C. elegans* não têm homólogos de FGE. Em procariotas encontraram-se 12 homólogos da FGE humana. Prevê-se que o cDNA da FGE humana codifique uma proteína com 374 resíduos (Fig. 3 e SEQ ID NO: 2). A proteína contém uma sequência de sinal apta a ser clivada com 33 resíduos, o que indica translocação da FGE para o retículo endoplasmático, e contém um único sítio de N-glicosilação em Asn141. A ligação da FGE à concanavalina A sugere que este sítio de N-glicosilação é utilizado. Os resíduos 87-367 da FGE estão listados na base de dados de motivos proteicos PFAM na forma de um domínio de função desconhecida (PFAM: DUF323). A análise de comparação de sequências da FGE humana e seus ortólogos eucarióticos

identificados em bases de dados indica que este domínio é composto por três subdomínios distintos.

O subdomínio N-terminal (resíduos 91-154 da FGE humana) tem uma identidade de sequências de 46% e uma semelhança de 79% nos quatro ortólogos de FGE eucarióticos conhecidos. Na FGE humana, este domínio contém o sítio de N-glicosilação em Asn 141, que está conservado nos outros ortólogos. A parte do meio da FGE (resíduos 179-308 da FGE humana) é representada por um subdomínio rico em triptofano (12 triptofanos por 129 resíduos). A identidade dos ortólogos eucarióticos neste subdomínio é 57%, a semelhança é 82%. O subdomínio C-terminal (resíduos 327-366 da FGE humana) é a sequência mais altamente conservada na família FGE. A identidade de sequências do subdomínio C-terminal humano com os ortólogos eucarióticos (3 sequências completas e 8 ESTs) é 85%, a semelhança é 97%. Nos 40 resíduos do subdomínio 3, quatro resíduos cisteína estão completamente conservados. Três das cisteínas também estão conservadas nos ortólogos de FGE procarióticos. Os 12 membros procarióticos da família FGE (quanto a pormenores ver o Exemplo 2) partilham a estrutura de subdomínios com FGEs eucarióticas. As fronteiras entre os três subdomínios são mais evidentes na família de FGE procariótica devido a sequências não conservadas de comprimento variável que separam os subdomínios entre si. O genoma humano e o de ratinho codificam dois homólogos intimamente relacionados da FGE (SEQ ID NOs: 43 e 44, Número de Acesso da GenBank NM_015411, no homem, e SEQ ID NOs: 45 e 46, Número de Acesso da GenBank AK076022, no ratinho). Os dois parálogos são 86% idênticos. Os seus genes estão localizados em regiões cromossômicas sinténicas (7q11 em humano, 5G1 em ratinho). Os dois parálogos partilham a estrutura de

subdomínios com os ortólogos da FGE e são 35% idênticos e 47% semelhantes à FGE humana. No terceiro subdomínio, que é 100% idêntico em ambos os homólogos, falta a sequência de undecâmero contendo cisteína do subdomínio 3.

Expressão, localização subcelular e formas moleculares

Um único transcrito de 2,1 kb é detetável por análise de coloração "Northern" de RNA total de fibroblastos da pele e RNA poli A⁺ de coração, cérebro, placenta, pulmão, fígado, musculoesquelético, rim e pâncreas. Relativamente a RNA de β -actina, a abundância varia em uma ordem de magnitude e é mais elevada em pâncreas e rim e menor em cérebro. Avaliaram-se várias linhas de células eucarióticas que expressam, de forma estável ou transitória, o cDNA da FGE humana ou derivados da FGE estendidos de forma C-terminal por um marcador HA, Myc ou His₆, quanto à atividade de FGE e localização subcelular de FGE. A expressão transitória de FGE marcada e não marcada aumentou a atividade de FGE em 1,6 - 3,9 vezes. A expressão estável de FGE em células PT67 aumentou a atividade de FGE em cerca de 100 vezes. A detecção da forma marcada de FGE por imunofluorescência indireta em células BHK 21, CHO e HT1080 revelou uma colocalização das formas de FGE marcadas de modo variado com proteína dissulfureto isomerase, uma proteína luminal do retículo endoplasmático. A análise por coloração "Western" de extratos de células BHK21 transfetadas de modo transitório com cDNA que codifica formas marcadas da FGE revelou uma única banda imunorreativa com uma dimensão aparente entre 42 e 44 kDa.

O gene FGE contém mutações em MSD

A MSD é causada por uma deficiência na geração de resíduos FGly em sulfatases (Schmidt, B., et al., *Cell*, 1995, 82: 271-278). Em consequência, o gene FGE é um gene candidato para MSD. Amplificámos e sequenciámos o cDNA que codifica a FGE de sete pacientes com MSD e descobrimos dez mutações diferentes que foram confirmadas por sequenciação do DNA genómico (Tabela 1).

Tabela 1: Mutações em pacientes com MSD

Mutação	Efeito na Proteína	Observações	Paciente
1076C>A	S359X	Truncamento dos 16 resíduos C-terminais	1*
IVS3+5-8 del	Deleção dos resíduos 149-173	Deleção na estrutura do exão 3	1, 2
979C>T	R327X	Perda do subdomínio 3	2
1045C>T	R349W	Substituição de um resíduo conservado no subdomínio 3	3, 7
1046G>A	R349Q	Substituição de um resíduo conservado no subdomínio 3	4
1006T>C	C336R	Substituição de um resíduo conservado no subdomínio 3	4
836C>T	A279V	Substituição de um resíduo conservado no subdomínio 2	5
243delC	deslocamento do quadro de leitura e truncamento	Perda dos três subdomínios	5
661delG	deslocamento do quadro de leitura e truncamento	Perda da terça parte C-terminal da FGE incluindo o subdomínio 3	6**
IVS6-1G>A	Deleção dos resíduos 281-318	Deleção na estrutura do exão 7	5
*O paciente 1 é o paciente com MSD Mo. em Schmidt, B., et al., <i>Cell</i> , 1995, <u>82</u> : 271-278 e Rommerskirch e von Figura, <i>Proc. Natl. Acad. Sci., USA</i> , 1992, <u>89</u> : 2561-2565. **O paciente 6 é o paciente com MSD relatado por Burk et al., <i>J. Pediatr.</i> , 1984, <u>104</u> : 574-578.			

Os outros pacientes representam casos não publicados.

O primeiro paciente era heterozigótico para uma substituição 1076C>A que converte o códon para serina 359 num códon de terminação (S359X) e uma mutação que causa a deleção dos 25 resíduos 149-173 que são codificados pelo exão 3 e procedem ao espaçamento do primeiro e segundo

domínios da proteína. A sequenciação genómica revelou uma deleção dos nucleótidos +5-8 do terceiro intrão (IVS3+5-8 del), desse modo destruindo o sítio dador de processamento do intrão 3. O segundo paciente era heterozigótico para a mutação que causa a perda do exão 3 (IVS3+5-8 del) e uma substituição 979C>T que converte o codão para arginina 327 num codão de terminação (R327X). A FGE truncada codificada pelo alelo 979C>T está desprovida da maior parte do subdomínio 3. O terceiro paciente era homozigótico para uma substituição 1045C>T que substitui a arginina 349 conservada no subdomínio 3 por triptofano (R349W). O quarto paciente era heterozigótico para duas mutações de sequência trocada que substituem resíduos conservados no domínio da FGE: uma substituição 1046>T que substitui arginina 349 por glutamina (R349Q) e uma substituição 1006T>C que substitui cisteína 336 por arginina (C336R). O quinto paciente era heterozigótico para uma substituição 836 C>T que substitui a alanina 279 conservada por valina (A279V). A segunda mutação é uma deleção de um único nucleótido (243delC) que altera a sequência após prolina 81 e causa uma paragem da tradução após o resíduo 139. O sexto paciente era heterozigótico para a deleção de um único nucleótido (661delG) que altera a sequência de aminoácidos após o resíduo 220 e introduz um codão de terminação após o resíduo 266. A segunda mutação é uma mutação do sítio aceitador de processamento do intrão 6 (IVS6-1G>A) que causa uma deleção na estrutura dos resíduos 281-318 codificadores do exão 7. No sétimo paciente encontrou-se a mesma substituição 1045C>T do terceiro paciente. Adicionalmente, detetámos dois polimorfismos na região codificadora de 18 alelos da FGE de controlos e pacientes com MSD. Vinte e dois por cento continham uma substituição

188G>A, que substitui serina 63 por asparagina (S63N), e 28% tinham uma substituição 1116C>T silenciosa.

Transdução de fibroblastos MSD com cDNA de FGE de tipo selvagem e mutante

Para confirmar a deficiência de FGE como causa da inatividade de sulfatases sintetizadas em MSD, expressámos o cDNA de FGE em fibroblastos MSD utilizando transferência genética retroviral. Como controlo procedemos à transdução do vetor retroviral sem inserção de cDNA. Para monitorizar a complementação do defeito metabólico na atividade de ASA medimos esteroide sulfatase (STS) e N-acetilgalactosamina 6-sulfatase (GalNAc6S) nos fibroblastos transduzidos antes ou após a seleção. A transdução da FGE de tipo selvagem restabeleceu parcialmente a atividade catalítica das três sulfatases em duas linhas de células MSD (Tabela 2) e para a STS numa terceira linha de células MSD. Deve notar-se que, para ASA e GalNAc6S, o restabelecimento foi apenas parcial após a seleção dos fibroblastos, atingindo 20 até 50% da atividade normal. Para STS verificou-se que a atividade foi restabelecida para a de fibroblastos de controlo após a seleção. A seleção aumentou a atividade da ASA e STS em 50 até 80%, o que é compatível com a observação anterior de que 15 até 50% dos fibroblastos são transduzidos (Lübke *et al.*, *Nat. Gen.*, 2001, 28: 73-76). As atividades de sulfatase nos fibroblastos MSD transduzidos com o vetor retroviral isoladamente (Tabela 2) foram comparáveis às de fibroblastos MSD não transduzidos (não mostrado). A transdução de cDNA de FGE contendo a mutação IVS3+5-8del não conseguiu restabelecer as atividades de sulfatase (Tabela 2).

Tabela 2: Complementação de fibroblastos MSD por transdução de cDNA de FGE de tipo selvagem ou mutante

Fibroblastos	Inserção de FGE	Sulfatase		
		ASA ¹	STS ¹	GalNAc6S ¹
MSD 3°	-	1,9 ± 0,2	< 3	56,7 ± 32
	FGE ⁺	7,9	13,5	n. d.
	FGE ⁺⁺	12,2 ± 0,2	75,2	283 ± 42
	FGE-IVS3+5-8del ⁺	1,8	< 3	n. d.
	FGE-IVS3+5-8del ⁺⁺	2,1	< 3	98,5
MSD 4°	-	1,1 ± 0,3	<3	n. d.
	FGE ⁺	4,7	17,0	n. d.
Fibroblastos de controlo		58 ± 11	66 ± 31	828 ± 426

¹ Os valores apresentam a razão entre ASA (mU/mg de proteína celular), STS (μU/mg de proteína celular), GalNAc6S (μU/mg de proteína celular) e de β-hexosaminidase (U/mg de proteína celular). Para fibroblastos de controlo apresenta-se a média e a variação de 6-11 linhas de células. Quando indicado, a gama de duas culturas transduzidas em paralelo é apresentada para fibroblastos MSD.

° O número de fibroblastos MSD refere-se ao do paciente na Tabela 1.

+ Determinação da atividade antes da seleção.

++ Determinação da atividade após a seleção.

n.d.: não determinado

Discussão

FGE é uma glicoproteína altamente conservada do retículo endoplasmático

A purificação de FGE de testículo bovino deu origem a dois polipéptidos de 39,5 e 41,5 kDa originários do mesmo gene. A expressão de três versões de FGE marcadas de modo diferente em três linhas de células eucarióticas diferentes numa única forma sugere que uma das duas formas observadas na preparação de FGE purificada de testículo bovino pode ter sido gerada por proteólise limitada durante a purificação. A substituição de Cys69 no péptido ASA65-80 por serina foi crítica para a purificação de FGE por cromatografia de afinidade. A FGE tem uma sequência de sinal apta a ser clivada que medeia a translocação através

da membrana do retículo endoplasmático. A maior parte da proteína madura (275 resíduos de 340) define um domínio único que provavelmente é composto por três subdomínios (ver Exemplo 2), pois não existe nenhum dos homólogos dos três subdomínios em proteínas com função conhecida. O reconhecimento do motivo de modificação *FGly* linear em polipéptidos sulfatase sintetizados de novo (Dierks *et al.*, *EMBO J.*, 1999, 18: 2084-2091) pode ser a função de um subdomínio da FGE. O domínio catalítico pode catalisar a formação de *FGly* de vários modos. Foi proposto que a FGE abstrai eletrões do grupo tiol da cisteína e transfere-os para um aceitador. O tioaldeído resultante sofrerá hidrólise espontânea em *FGly* e H_2S (Schmidt, B., *et al.*, *Cell*, 1995, 82: 271-278). Alternativamente, a FGE pode atuar como uma oxigenase (mono-oxigenase) de função mista, introduzindo um átomo de O_2 na cisteína e o outro em H_2O com o auxílio de um dador de eletrões, como $FADH_2$. O derivado tioaldeído hidratado resultante de cisteína reagirá espontaneamente formando *FGly* e H_2S . Experiências preliminares com uma preparação de FGE parcialmente purificada revelaram uma dependência crítica da formação de *FGly* do oxigénio molecular. Isto sugere que a FGE atua como uma oxigenase de função mista. A elevada conservação particular do subdomínio 3 e a presença aí de três resíduos cisteína completamente conservados tornam este subdomínio num candidato provável para o sítio catalítico. Será interessante ver se os elementos estruturais que medeiam o reconhecimento do motivo *FGly* e a ligação de um aceitador de eletrões ou dador de eletrões estão correlacionados com a estrutura de domínios da FGE.

A FGE recombinante está localizada no retículo endoplasmático, o que é compatível com o sítio proposto da

sua ação. Resíduos *FGly* são gerados em sulfatases sintetizadas de novo durante ou pouco tempo após a sua translocação para o retículo endoplasmático (Dierks *et al.*, *Proc. Natl. Acad. Sci. U.S.A.*, 1997, 94: 11963-11968; Dierks *et al.*, *FEBS Lett.*, 1998, 423: 61-65). A própria FGE não contém um sinal de retenção no ER do tipo KDEL. Em consequência, a sua retenção no retículo endoplasmático pode ser mediada pela interação com outras proteínas ER. Componentes da maquinaria de translocação/N-glicosilação são candidatos atrativos para esses parceiros de interação.

Mutações na FGE causam MSD

Mostrámos que mutações no gene codificador de FGE causam MSD. A FGE também pode interagir com outros componentes, e defeitos em genes que codificam os últimos podem igualmente bem causar MSD. Em sete pacientes com MSD descobrimos, de facto, dez mutações diferentes no gene FGE. Todas as mutações têm efeitos graves na proteína FGE, substituindo resíduos altamente conservados no subdomínio 3 (três mutações) ou subdomínio 2 (uma mutação) ou truncamentos C-terminais de vários comprimentos (quatro mutações) ou grandes deleções na estrutura (duas mutações). Para duas linhas de células MSD e uma das mutações MSD verificou-se que a transdução do cDNA de FGE de tipo selvagem, mas não do mutante, restabelece parcialmente as atividades de sulfatase. Isto identifica claramente o gene FGE como o sítio de mutação e a natureza causadora de doença da mutação. A MSD é clínica e bioquimicamente heterogénea. Foram diferenciadas uma forma neonatal rara, que se apresenta na altura do nascimento e desenvolve um hidrocéfalo, uma forma comum, que se assemelha inicialmente a uma leucodistrofia metacromática infantil e

subsequentemente desenvolve características do tipo ictiose e mucopolissacaridose, e uma forma ligeira menos frequente, na qual prevalecem as características clínicas de uma mucopolissacaridose. Bioquimicamente, é característico que possa ser detetada uma atividade residual de sulfatases, a qual, para a maior parte dos casos em fibroblastos da pele cultivados, é inferior a 10% dos controlos (Burch *et al.*, *Clin. Genet.*, 1986, 30: 409-15; Basner *et al.*, *Pediatr. Res.*, 1979, 13: 1316-1318). No entanto, nalgumas linhas de células MSD, a atividade de sulfatases selecionadas pode atingir a gama normal (Yutaka *et al.*, *Clin. Genet.*, 1981, 20: 296-303). Além disso, foi relatado que a atividade residual está sujeita a variações, dependendo das condições de cultura de células e de fatores desconhecidos. Bioquimicamente, a MSD foi classificada em dois grupos. No grupo I, a atividade residual de sulfatases é inferior a 15%, incluindo a de ASB. No grupo II, a atividade residual de sulfatases é mais elevada, e particularmente a da ASB pode atingir valores até 50-100% do controlo. Todos os pacientes relatados aqui caem no grupo I excetuando o paciente 5, que cai no grupo II (atividade de ASB na gama do controlo) do fenótipo bioquímico. Com base em critérios clínicos, os pacientes 1 e 6 são casos neonatais, ao passo que os pacientes 2-4 e 7 têm a forma comum e o paciente 5 tem a forma do tipo mucopolissacaridose de MSD.

A heterogeneidade fenotípica sugere que as diferentes mutações em pacientes com MSD estão associadas a diferentes atividades residuais da FGE. Dados preliminares em células PT67 que expressam de modo estável FGE IV3+5-8del indicam que a deleção na estrutura do exão 3 abole completamente a atividade de FGE. A caracterização das mutações em MSD, das propriedades bioquímicas da FGE mutante e do teor residual

de *FGly* em sulfatases, utilizando um método espectrométrico de massa altamente sensível recentemente desenvolvido (Peng *et al.*, *J. Mass Spec.*, 2003, 38: 80-86), proporcionará uma melhor compreensão da correlação genótipo-fenótipo em MSD.

Exemplo 2:

O gene FGE humano define uma nova família de genes que modificam sulfatases que está conservada desde procariotas até eucariotas

Bioinformática

Péptidos de sinal e sítios de clivagem foram descritos com o método de von Heijne (*Nucleic Acids Res.*, 1986, 14: 4683) implementado em EMBOSS (Rice *et al.*, *Trends in Genetics*, 2000, 16: 276-277), e o método de Nielsen *et al.* (*Protein Engineering*, 1997, 10: 1-6). Sítios de N-glicosilação foram previstos utilizando o algoritmo de Brunak (Gupta e Brunak, *Pac. Symp. Biocomput.*, 2002, 310-22).

Domínios funcionais foram detetados por pesquisa em PFAM-Hidden-Markov-Models (versão 7.8) (Sonnhammer *et al.*, *Nucleic Acids Res.*, 1998, 26: 320-322). Sequências da semente PFAM DUF323 foram obtidas da TrEMBL (Bairoch, A. e Apweiler, R., *Nucleic Acids Res.*, 2000, 28: 45-48). Alinhamentos múltiplos e construções de árvores filogenéticas foram efetuados com Clustal W (Thompson, J., *et al.*, *Nucleic Acids Res.*, 1994, 22: 4673-4680). Para o cálculo de árvores filogenéticas, posições de intervalos foram excluídas e múltiplas substituições foram corrigidas. Realizou-se "bootstrapping" de árvores para se obterem resultados significativos. As árvores foram visualizadas

utilizando Njplot (Perriere, G. e Gouy, M., *Biochimie*, 1996, 78: 364-369). Os alinhamentos foram representados graficamente utilizando o comando "pret-typlot" de EMBOSS.

Para procurar homólogos da FGE, as bases de dados NR, NT e EST do National Center for Biotechnology Information (NCBI) (Wheeler *et al.*, *Nucleic Acids Res.*, 2002, 20: 13-16) foram pesquisadas com BLAST (Altschul *et al.*, *Nucleic Acids Res.*, 1997, 25: 3389-3402). Para sequências proteicas, a pesquisa foi efetuada utilizando Psi-Blast de convergência iterativa contra a versão corrente da base de dados NR utilizando um limite do valor de expectativa de 10^{-40} , e parâmetros por defeito. Obteve-se convergência após 5 iterações. Para sequências de nucleótidos, a pesquisa foi efetuada com Psi-TBlastn: utilizando NR e a sequência proteica da FGE humana como entrada, construiu-se uma matriz de pontuação para a hFGE com Psi-Blast de convergência iterativa. Esta matriz foi usada como entrada para blastall, para pesquisar as bases de dados de nucleótidos NT e EST. Para ambos os passos utilizou-se um limite do valor de expectativa de 10^{-20} .

A previsão da estrutura proteica secundária foi efetuada utilizando Psipred (Jones, D., *J Mol Biol.*, 1999, 292: 1950-202; McGuffin, L., *et al.*, *Bioinformatics*, 2000, 16: 404-405).

Calcularam-se pontuações de semelhança dos subdomínios a partir de alinhamentos utilizando o algoritmo cons da EMBOSS com parâmetros por defeito. Os meta-alinhamentos foram gerados por alinhamento de sequências de consenso dos subgrupos da família FGE. A organização e sintenia de *loci* genómicos foram determinadas utilizando os recursos dos genomas humano e de ratinho do NCBI em NCBI (Bethesda, MD)

e Human-Mouse-Rat Synteny da Softberry (Mount Kisco, NY). Sequências genômicas bacterianas foram descarregadas do servidor NCBI-FTP. A anotação genômica microbiana do NCBI foi utilizada para se obter uma visão global dos *loci* genômicos de genes FGE bacterianos.

Resultados e Discussão

Características e motivos básicos da FGE humana e proteínas relacionadas

O gene FGE humano (SEQ ID NOs: 1, 3) codifica a proteína FGE (SEQ ID NO: 2) que se prevê ter 374 resíduos. Um sinal de clivagem entre os resíduos 22-33 (Pontuação de Heijne de 15,29) e uma pontuação de hidropatia (Kyte, J. e Doolittle, R., *J Mol Biol.*, 1982, 157: 105-132) dos resíduos 17-29 entre 1,7 e 3,3 indicam que os 33 resíduos N-terminais são removidos por clivagem após translocação para o ER. No entanto, com o algoritmo de Nielsen *et al.* (*Protein Engineering*, 1997, 10: 1-6), a clivagem da sequência de sinal é prevista após o resíduo 34. A proteína tem um único sítio de N-glicosilação potencial em Asn 141.

Uma pesquisa com a sequência proteica FGE contra a base de dados de motivos de proteínas PFAM (Sonnhammer *et al.*, *Nucleic Acids Res.*, 1998, 26: 320-322) revelou que os resíduos 87-367 da FGE humana podem ser classificados como domínio proteico DUF323 ("domínio de função desconhecida", PF03781) com um valor de expectativa altamente significativo de $7:9.10^{-114}$. A semente de PFAM que define DUF323 consiste em 25 sequências proteicas, das quais a maioria consiste em proteínas hipotéticas derivadas de dados de sequenciação. Para analisar a relação entre a FGE humana e DUF323

efetuou-se um múltiplo alinhamento de FGE com as sequências da semente DUF323. Com base nisto construiu-se uma árvore filogenética que foi sujeita a "bootstrap". Quatro das sequências hipotéticas (TrEMBL-IDs Q9CK12, Q9I761, 094632 e Q9Y405) apresentaram uma divergência tão forte dos outros membros da semente que impediram um "bootstrap" bem-sucedido e tiveram de ser removidas do conjunto. A Figura 2 mostra a árvore sujeita a "bootstrap" exibindo a relação entre a FGE humana e as restantes 21 proteínas da semente DUF323. A árvore pode ser utilizada para subdividir os membros da semente em duas categorias: homólogos intimamente relacionados com a FGE humana e os genes restantes menos relacionados.

As 7 proteínas do topo têm uma distância filogenética entre 0,41 e 0,73 para a FGE humana. Contêm apenas um único domínio, DUF323. A homologia neste grupo estende-se ao longo de toda a sequência de aminoácidos, cuja maior parte consiste no domínio DUF323. O domínio DUF323 está fortemente conservado neste grupo de homólogos, ao passo que as outras 15 proteínas da semente estão menos relacionadas com a FGE humana (distância filogenética entre 1,14 e 1,93). Os seus domínios DUF323 divergem consideravelmente do domínio DUF323 altamente conservado do primeiro grupo (ver secção "Subdomínios de FGE e mutações no gene FGE"). A maior parte destas 15 proteínas é hipotética, seis das quais foram adicionalmente investigadas. Uma delas, uma serina/treonina quinase (TrEMBL: 084147) de *C. trachomatis*, contém outros domínios para além do DUF323: um domínio de ligação de ATP e um domínio de quinase. As sequências de *R. sphaeroides* (TrEMBL: Q9ALV8) e *Pseudomonas* sp. (TrEMBL: 052577) codificam a proteína NirV, um gene cotranscrito com a

nitrito redutase contendo cobre nirK (Jain, R. e Shapleigh, J., *Microbiology*, 2001, 147: 2505-2515). CarC (TrEMBL: Q9XB56) é uma oxigenase envolvida na síntese de um antibiótico β -lactama de *E. carotovora* (McGowan, S., et al., *Mol Microbiol.*, 1996, 22: 415-426; Khaleeli N, T. C., e Busby RW, *Biochemistry*, 2000, 39: 8666-8673). XylR (TrEMBL: 031397) e BH0900 (TrEMBL: Q9KEF2) são proteínas de ligação intensificadoras envolvidas na regulação da utilização de pentose (Rodionov, D., et al., *FEMS Microbiol Lett.*, 2001, 205: 305-314) em *bacillaceae* e *clostridiaceae*. A comparação de FGE e DUF323 conduziu ao estabelecimento de um limite de homologia que diferencia a família FGE de homólogos distantes contendo DUF323 com diferentes funções. Os últimos incluem uma serina/treonina quinase e XylR, um intensificador da transcrição bem como FGE, uma enzima geradora de *FGly* e CarC, uma oxigenase. Como discutido aqui noutro local, a FGE também poderá exercer a sua função modificadora de cisteína na forma de uma oxigenase, sugerindo que a FGE e membros da semente DUF323 diferentes de FGE podem partilhar uma função de oxigenase.

Homólogos da FGE

A presença de homólogos intimamente relacionados da FGE humana na semente DUF323 dirigiu-nos para a pesquisa de homólogos da FGE humana na base de dados NR do NCBI (Wheeler et al., *Nucleic Acid Res.*, 2002, 20: 13-16). O limite da pesquisa foi escolhido de modo tal que fossem obtidos os 6 homólogos presentes na semente DUF323 e outros homólogos intimamente relacionados sem encontrar os outros membros da semente. Esta pesquisa conduziu à identificação de três ortólogos da FGE em eucariotas, 12 ortólogos em procariotas e dois parálogos no homem e ratinho (Tabela 3).

Tabela 3: A família de genes FGE em eucariotas e procariotas

SEQ ID NOs: NA, AA [GI]	ESPÉCIE	COMPRIMENTO [AA]	SUBGRUPO
1/3,2	<i>Homo sapiens</i>	374	E1
49,50 [22122361]	<i>Mus musculus</i>	372 ^f	E1
51,52 [20130397]	<i>Drosophila melanogaster</i>	336	E1
53,54 [21289310]	<i>Anopheles gambiae</i>	290	E1
47, 48 [26344956]	<i>Mus musculus</i>	308	E2
45, 46 [24308053]	<i>Homo sapiens</i>	301	E2
55,56 [21225812]	<i>Streptomyces coelicolor</i> A3(2)	314	P1
57,58 [25028125]	<i>Corynebacterium efficiens</i> YS-314	334	P1
59,60 [23108562]	<i>Novosphingobium aromaticivorans</i>	338	P2
61, 62 [13474559]	<i>Mesorhizobium loti</i>	372	P2
63,64 [22988809]	<i>Burkholderia fungorum</i>	416	P2
65, 66 [16264068]	<i>Sinorhizobium meliloti</i>	303	P2
67,68 [14518334]	<i>Microscilla</i> sp.	354	P2
69,70 [26990068]	<i>Pseudomonas putida</i> KT2440	291	P2
71, 72 [22975289]	<i>Ralstonia metallidurans</i>	259	P2
73,74 [23132010]	<i>Prochlorococcus marinus</i>	291	P2
75, 76 [16125425]	<i>Caulobacter crescentus</i> CB15	338	P2
77, 78 [15607852]	<i>Mycobacterium tuberculosis</i> Ht37Rv	299	P2
GI- Identificador de proteínas da GenBank NA- ácido nucleico AA - aminoácidos, E1 - ortólogos eucarióticos E2 - parálogos eucarióticos P1 - ortólogos procarióticos intimamente relacionados P2 - outros ortólogos procarióticos f- sequência proteica mal prevista na GenBank			

De notar que a Genbank prevê que a sequência de ratinho GI 22122361 codifique uma proteína de 284 aminoácidos, apesar de a sequência de cDNA NM 145937 codificar uma proteína de

372 resíduos. Esta previsão errada baseia-se na omissão do primeiro exão do gene FGE murino. Todas as sequências encontradas na base de dados NR provêm de eucariotas superiores ou procariotas. Não foram detetados homólogos da FGE em *archaeobacterae* ou plantas. Pesquisas efetuadas mesmo com limites menores nos genomas totalmente sequenciados de *C. elegans* e *S. cerevisiae* e nas bases de dados de ORF relacionadas não revelaram quaisquer homólogos. Uma pesquisa nas sequências eucarióticas das bases de dados de nucleótidos NT e EST conduziram à identificação de 8 ESTs de ortólogos de FGE adicionais com fragmentos da sequência de cDNA 3'-terminais exibindo um elevado grau de conservação ao nível proteico que não estão listados na base de dados NR. Estas sequências não abrangem a parte codificadora completa dos mRNAs e todas provêm de eucariotas superiores (Tabela 4).

Tabela 4: Fragmentos ESTs de ortólogos da FGE em eucariotas

SEQ ID NOs: NA [GB]	ESPÉCIE
80 [CA379852]	<i>Oncorhynchus mykiss</i>
81 [AI721440]	<i>Danio rerio</i>
82 [BJ505402]	<i>Oryzias latipes</i>
83 [BJ054666]	<i>Xenopus laevis</i>
84 [AL892419]	<i>Silurana tropicalis</i>
85 [CA064079]	<i>Salmo salar</i>
86 [BF189614]	<i>Sus scrofa</i>
87 [AV609121]	<i>Bos taurus</i>
GB- N° de Acesso da GenBank; NA- ácido nucleico	

Múltiplos alinhamentos e a construção de uma árvore filogenética (utilizando ClustalW) das sequências codificadoras da base de dados NR permitiram a definição de

quatro subgrupos de homólogos: ortólogos eucarióticos (FGE humana, de ratinho, mosquito e mosca da fruta), parálogos eucarióticos (parálogos da FGE humana e de ratinho), ortólogos procarióticos intimamente relacionados com a FGE (*Streptomyces* e *Corynebacterium*) e outros ortólogos procarióticos (*Caulobacter*, *Pseudomonas*, *Mycobacterium*, *Prochlorococcus*, *Mesorhizobium*, *Sinorhizobium*, *Novosphingobium*, *Ralstonia*, *Burkholderia* e *Microscilla*). Os ortólogos eucarióticos exibem uma identidade global com a FGE humana de 87% (ratinho), 48% (mosca da fruta) e 47% (*Anopheles*). Apesar de se encontrarem ortólogos da FGE em procariotas e eucariotas superiores, estão ausentes dos genomas completamente sequenciados de eucariotas inferiores filogeneticamente situados entre *S. cerevisiae* e *D. melanogaster*. Adicionalmente, homólogos da FGE estão ausentes dos genomas completamente sequenciados de *E. coli* e do peixe-balão.

Como discutido aqui noutro local, os parálogos da FGE encontrados em humanos e ratinho podem ter uma atividade geradora de *FGly* muito pequena e contribuir para as atividades residuais de sulfatases presentes em pacientes com MSD.

Subdomínios da FGE

Os membros da família de genes FGE têm três partes/domínios altamente conservados (como descrito aqui noutro local). Para além das duas sequências não conservadas que separam os primeiros, têm extensões não conservadas no terminal N e C. Considera-se que as três partes conservadas representam subdomínios do domínio DUF323 porque estão espaçadas por partes não conservadas de comprimento variável. O

comprimento da parte de espaçamento dos subdomínios 1 e 2 varia entre 22 e 29 resíduos, e o da parte de espaçamento dos subdomínios 2 e 3 varia entre 7 e 38 aminoácidos. As partes não conservadas N e C-terminais exibem uma variação de comprimento ainda mais forte (N-terminal: 0-90 aminoácidos, C-terminal: 0-28 aminoácidos). A sequência do gene FGE de *Ralstonia metallidurans* está provavelmente incompleta, pois não tem o primeiro subdomínio.

Para verificar a plausibilidade de definir subdomínios de DUF323, efetuámos uma previsão da estrutura secundária da proteína FGE humana utilizando Psipred. Prevê-se que o sinal de ER hidrofóbico (resíduos 1-33) contenha estruturas em hélice confirmando a previsão de sinal do algoritmo de von-Heijne. A região não conservada N-terminal (aminoácidos 34-89) e a região de espaçamento entre os subdomínios 2 e 3 (aminoácidos 308-327) contêm secções enroladas. A região de espaçamento dos subdomínios 1 e 2 contém um enrolamento. A hélice α nos aminoácidos 65/66 tem baixa confiança de previsão e é provavelmente um artefacto da previsão. As fronteiras dos subdomínios estão situadas em enrolamentos e não interrompem hélices α nem filamentos β . O primeiro subdomínio é constituído por vários filamentos β e uma hélice α , o segundo subdomínio contém dois filamentos β e quatro hélices α . O terceiro subdomínio tem uma região em hélice α flanqueada por uma folha no início e no final do subdomínio. Em resumo, a estrutura secundária está de acordo com a estrutura de subdomínios proposta, pois as fronteiras dos subdomínios estão situadas em enrolamentos e os subdomínios contêm elementos estruturais hélices α e filamentos β).

Deve ser notado que nenhum dos subdomínios existe como um módulo isolado em sequências listadas em bases de dados. Em cada um dos quatro subgrupos da família FGE, os subdomínios estão altamente conservados, com o terceiro subdomínio exibindo a homologia mais elevada (Tabela 5). Este subdomínio também exibe a homologia mais forte através dos subgrupos.

Tabela 5: Homologia (% semelhança) dos subdomínios da família FGE

Subfamília	Membros	Subdomínio		
		1	2	3
E1	4	79	82	100
E2	2	90	94	100
P1	2	70	79	95
P2	10	59	79	80
E1 - ortólogos eucarióticos; E2 - parálogos eucarióticos P1 - ortólogos procarióticos intimamente relacionados; P2 - outros ortólogos procarióticos				

O primeiro subdomínio da família FGE exibe a homologia mais fraca através dos subgrupos. Nos ortólogos eucarióticos contém o sítio de N-glicosilação: no resíduo Asn 141 em humanos, em Asn 139 no ratinho e Asn 120 na mosca da fruta. Em *anopheles* não se encontra asparagina no resíduo 130 homólogo a Asn 120 de *D. melanogaster*. No entanto, uma alteração de dois nucleótidos criaria um sítio de N-glicosilação Asn 130 em *anopheles*. Assim, é necessário sequenciar novamente a sequência que abrange o resíduo 130. O segundo subdomínio é rico em triptofanos, com 12 Trp em 129 resíduos da FGE humana. Dez destes triptofanos estão conservados na família FGE.

Conservação elevada do subdomínio 3: os subdomínios 3 entre ortólogos eucarióticos são 100% semelhantes e 90% idênticos. A importância do terceiro subdomínio para a função da proteína é sublinhada pela observação de que este

subdomínio é um ponto quente para mutações causadoras de doença em pacientes com MSD. Sete de nove mutações identificadas em seis pacientes com MSD descritos no Exemplo 1 estão localizadas em sequências que codificam os 40 resíduos do subdomínio 3. Os resíduos contêm quatro cisteínas, três das quais estão conservadas entre os ortólogos procarióticos e eucarióticos. Os dois parálogos eucarióticos exibem a homologia mais baixa com os outros membros da família FGE, por exemplo, não têm duas das três cisteínas conservadas do subdomínio 3. Características conservadas entre sequências do subdomínio 3 de ortólogos e parálogos são o motivo **RVXXGG(A)S** inicial (SEQ ID NO: 79), um heptâmero contendo três argininas (resíduos 19-25 da sequência de consenso do subdomínio) e o motivo GFR terminal. Uma comparação com o domínio DUF323 das 15 sequências de semente que não são homólogos próximos da FGE mostra diferenças de sequências acentuadas: as 15 sequências de semente têm um primeiro e segundo subdomínios menos conservados, apesar de a estrutura global de subdomínios também ser visível. O subdomínio 3, que está fortemente conservado na família FGE, é mais pequeno e tem uma homologia significativamente mais fraca com o subdomínio 3 eucariótico (semelhança de cerca de 20%) em comparação com os membros procarióticos da família FGE (semelhança de cerca de 60%). Assim, estão desprovidos de todos os resíduos cisteína conservados do subdomínio 3. As únicas características conservadas são o motivo **RVXXGG(A)S** inicial (SEQ ID NO: 79) e o motivo GFR terminal.

Organização genómica do gene FGE humano e murino

O gene FGE humano está localizado no cromossoma 3p26. Abrange 105 kb e 9 exões para a sequência traduzida. O gene

FGE murino tem um comprimento de 80 kb e está localizado no cromossoma 6E2. Os 9 exões do gene FGE murino têm quase o mesmo tamanho dos exões humanos (Figura 3). As grandes diferenças entre o gene humano e o de ratinho são a menor conservação da 3'-UTR no exão 9 e o comprimento do exão 9, que é mais longo em 461 bp no gene murino. O segmento 6E2 do cromossoma de ratinho 6 é altamente sinténico com o segmento do cromossoma humano 3p26. Na direção do telómero, os *loci* FGE humano e murino são flanqueados pelos genes que codificam LMCD1, KIAA0212, ITPR1, AXCAM e IL5RA. Na direção centromérica, ambos os *loci* FGE são flanqueados pelos *loci* de CAV3 e OXTR.

Organização genómica dos genes FGE procarióticos

Em procariotas, as sulfatases são classificadas como sulfatases do tipo cisteína ou serina, dependendo do resíduo que é convertido em *FGly* no seu centro ativo (Miech, C., et al., *J Biol Chem.*, 1998, 273: 4835-4837; Dierks, T., et al., *J Biol Chem.*, 1998, 273: 25560-25564). Em *Klebsiella pneumoniae*, *E. coli* e *Yersinia pestis*, as sulfatases do tipo serina fazem parte de um operão com *AtsB*, que codifica uma proteína citosólica contendo motivos agregados de ferro-enxofre e é crítica para a geração de *FGly* a partir de resíduos serina (Marquardt, C., et al., *J Biol Chem.*, 2003, 278: 2212-2218; Szameit, C., et al., *J Biol Chem.*, 1999, 274: 15375-15381).

Em consequência, foi interessante examinar se genes FGE procarióticos estão localizados na proximidade de sulfatases do tipo cisteína que são os substratos da FGE. Entre os genes FGE procarióticos apresentados na Tabela 3, sete têm genomas totalmente sequenciados, permitindo uma

análise da vizinhança dos *loci* de FGE. De facto, em quatro dos 7 genomas (*C. efficiens*: PID 25028125, *P. putida*: PID 26990068, *C. crescentus*: PID 16125425 and *M. tuberculosis*: PID 15607852) uma sulfatase do tipo cisteína encontra-se na vizinhança direta de FGE, compatível com uma cotranscrição de FGE e da sulfatase. Em dois deles (*C. efficiens* e *P. putida*), FGE e a sulfatase têm mesmo ORFs sobrepostas, apontando fortemente para a sua coexpressão. Além disso, a vizinhança genómica de genes FGE e sulfatase em quatro procariotas proporciona evidências adicionais para a presunção de que as FGEs bacterianas são ortólogos funcionais.

Os três organismos restantes contêm sulfatases do tipo cisteína (*S. coelicolor*: PID 24413927, *M. loti*: PID 13476324, *S. meliloti*: PIDs 16262963, 16263377, 15964702); no entanto, os genes vizinhos da FGE nestes organismos não contêm uma assinatura de sulfatase canónica (Dierks, T., et al., *J Biol Chem.*, 1998, 273: 25560-25564) nem um domínio que indique a sua função. Em consequência, é provável que, nestes organismos, a expressão de FGE e sulfatases do tipo cisteína seja *regulada in trans*.

Conclusões

A identificação de FGE humana, cuja deficiência causa a doença de armazenamento lisossómico transmitida de forma autossómica-regressiva Deficiência de Múltiplas Sulfatases, permite a definição de uma nova família de genes que compreende ortólogos da FGE de procariotas e eucariotas, bem como um parálogo da FGE no ratinho e no homem. A FGE não se encontra nos genomas completamente sequenciados de *E. coli*, *S. cerevisiae*, *C. elegans* e *Fugu rubripes*.

Adicionalmente, há um hiato filogenético entre procariotas e eucariotas superiores, com ausência de FGE em qualquer espécie situada filogeneticamente entre procariotas e *D. melanogaster*. No entanto, alguns destes eucariotas inferiores, por exemplo, *C. elegans*, têm genes sulfatase do tipo cisteína. Isto aponta para a existência de um segundo sistema gerador de *FGly* que atua em sulfatases do tipo cisteína. Esta presunção é suportada pela observação de que *E. coli*, que não tem FGE, pode gerar *FGly* em sulfatases do tipo cisteína (Dierks, T., et al., *J Biol Chem.*, 1998, 273: 25560-25564).

Exemplo 3:

A expressão de FGE origina aumentos significativos da atividade de sulfatase em linhas de células que superexpressam uma sulfatase

Quisemos examinar os efeitos da FGE em células que expressam/superexpressam uma sulfatase. Para esta finalidade, células HT-1080 que expressam as sulfatases humanas Iduronato 2-Sulfatase (I2S) ou N-Acetilgalactosamina 6-Sulfatase (GALNS) foram transfetadas em duplicado com uma construção de expressão de FGE, pXMG.1.3 (Tabela 7 e Fig. 4), ou um plasmídeo de controlo, pXMG.1.2 (FGE em orientação antissentido incapaz de produzir FGE funcional, Tabela 7). Recolheram-se amostras de meio às 24, 48 e 72 horas após uma mudança do meio pós-eletroporação às 24 horas. As amostras de meio foram testadas quanto à atividade de sulfatase respetiva por ensaio da atividade, e o nível proteico de sulfatase total foi estimado por ELISA específica para Iduronato 2-Sulfatase ou N-Acetilgalactosamina 6-Sulfatase.

Tabela 6. Linhas de Células Transfetadas que Expressam Sulfatases Utilizadas como Substrato para Transfeção

Estirpe de Células	Plasmídeo	Sulfatase Expressa
36F	XFM4A.1	N-Acetilgalactosamina 6-Sulfatase
30C6	pXI2S6	Iduronato 2-Sulfatase

Tabela 7. FGE e Plasmídeos de Controlo Utilizados para a Transfeção de Células HT-1080 que Expressam Iduronato 2-Sulfatase e N-Acetilgalactosamina 6-Sulfatase

Plasmídeo	Configuração de Elementos Principais da Sequência de DNA*
pXMG.1.3 (expressão de FGE)	> intensificador/promotor do CMV de 1,6 kb > cDNA de FGE de 1,1 kb > sequência não traduzida hGH3' < amp < cassete DHFR < cassete Cdneo (neomicina fosfotransferase)
pXMG.1.2 (controlo, orientação reversa de FGE)	> intensificador/promotor do CMV de 1,6 kb > cDNA de FGE de 1,1 kb < sequência não traduzida hGH3' < amp < cassete DHFR < cassete Cdneo (neomicina fosfotransferase)
* > designa orientação 5' para 3'	

Procedimentos Experimentais

Materiais e Métodos

Transfeção de células HT-1080 produtoras de Iduronato 2-Sulfatase e N-Acetilgalactosamina 6-Sulfatase

Recolheram-se células HT-1080 para se obterem $9-12 \times 10^6$ células para cada eletroporação. Dois plasmídeos foram transfetados em duplicado: um a ser testado (FGE) e um controlo; neste caso, o plasmídeo de controlo continha o cDNA de FGE clonado na orientação reversa relativamente ao promotor do CMV. As células foram centrifugadas aproximadamente a 1000 RPM durante 5 minutos. As células foram suspensas em 1X PBS a 16×10^6 células/mL. À base da cuvete de eletroporação adicionaram-se 100 µg de DNA plasmídico, à solução de DNA presente na cuvete adicionaram-se 750 µL de suspensão de células (12×10^6

células). As células e DNA foram misturados suavemente com uma pipeta de transferência de plástico, tendo o cuidado de não criar bolhas. As células foram submetidas a eletroporação a 450 V, 250 μ F (BioRad Gene Pulser). Registrou-se a constante temporal.

As células sujeitas a eletroporação assentaram sem perturbações durante 10-30 minutos. Depois adicionaram-se a cada cuvete 1,25 mL de DMEM/10% soro de vitelo, o sistema foi misturado e todas as células foram transferidas para um balão T75 fresco contendo 20 mL de DMEM/10. Passadas 24 horas, o balão foi novamente nutrido com 20 mL de DMEM/10, para remover células mortas. Quarenta e oito - setenta e duas horas após a transfeção recolheram-se amostras de meio e as células foram recolhidas de balões T75 em duplicado.

Preparação do Meio

Um litro de DMEM/10 (contém: 23 mL de 2 mM L Glutamina, 115 mL de soro de vitelo).

As células foram transfetadas em meio sem metotrexato (MTX). Vinte e quatro horas mais tarde, as células foram novamente nutridas com meio contendo as quantidades apropriadas de MTX (36F = 1,0 μ M MTX, 30C6 = 0,1 M MTX). O meio foi recolhido e as células foram recolhidas às 24, 48 e 72 horas após a nova nutrição.

Ensaio de Atividade

Iduronato 2-Sulfatase (I2S). Colunas de dessalinização NAP5 (Amersham Pharmacia Biotech AB, Uppsala, Suécia) foram equilibradas com Tampão de Diálise (5 mM acetato de sódio,

5 mM tris, pH 7,0). A amostra contendo I2S foi aplicada na coluna e deixada entrar no leito. A amostra eluiu em 1 mL de Tampão de Diálise. Amostras dessalinizadas foram adicionalmente diluídas para aproximadamente 100 ng/mL de I2S em Tampão Reacional (5 mM acetato de sódio, 0,5 mg/L de BSA, 0,1 % Triton X-100, pH 4,5). Adicionaram-se 10 µL de cada amostra de I2S à linha de cima de uma Placa Fluorométrica de 96 cavidades (Perkin Elmer, Norwalk, CT) e o sistema foi pré-incubado durante 15 minutos a 37°C. Preparou-se substrato por dissolução de sulfato de 4-metilumbeliferilo (Fluka, Buchs, Suíça) em Tampão de Substrato (5 mM acetato de sódio, 0,5 mg/mL de BSA, pH 4,5) a uma concentração final de 1,5 mg/mL. Adicionaram-se 100 µL de Substrato a cada cavidade contendo amostra de I2S e a placa foi incubada durante 1 hora a 37°C no escuro. Após a incubação adicionaram-se 190 µL de Tampão de Terminação (332,5 mM glicina, 207,5 mM carbonato de sódio, pH 10,7) a cada cavidade contendo amostra. Preparou-se um lote de 4-metilumbeliferona (4-MUF, Sigma, St. Louis, MO) como padrão de produto em água com qualidade reagente para uma concentração final de 1 µM. Adicionaram-se 150 µL de Lote 1 µM 4-MUF e 150 µL de Tampão de Terminação a uma cavidade da linha de cima da placa. Adicionaram-se 150 µL de Tampão de Terminação a cada cavidade restante da placa de 96 cavidades. Fizeram-se diluições em série de duas vezes desde a linha de cima de cada coluna até à última linha da placa. A placa foi lida num Analisador de Microplacas Fusion Universal (Packard, Meriden, CT) com um comprimento de onda de filtração de excitação de 330 nm e um comprimento de onda de filtração de emissão de 440 nm. Gerou-se uma curva padrão de µmoles do lote de 4-MUF versus fluorescência, e a partir desta curva extrapolou-se a fluorescência de amostras desconhecidas. Os resultados

estão relatados em Unidades/mL, em que uma Unidade de atividade iguala 1 μ mole de 4-MUF produzida por minuto a 37°C.

N-Acetilgalactosamina 6-Sulfatase (GALNS). O ensaio de atividade de GALNS emprega o substrato fluorescente 4-metilumbeliferil- β -D-galactopiranosido-6-sulfato (Toronto Research Chemicals Inc., Número do Catálogo M33448). O ensaio consistia em dois passos. No primeiro passo, 75 μ L do substrato 1,3 mM, preparado em tampão reacional (0,1 M acetato de sódio, 0,1 M cloreto de sódio, pH 4,3), foram incubados durante 4 horas a 37°C com 10 μ L de meio/amostra proteica ou suas diluições correspondentes. A reação foi terminada pela adição de 5 μ L de 2 M fosfato de sódio monobásico, para inibir a atividade de GALNS. Após a adição de aproximadamente 500 U de β -galactosidase de *Aspergillus oryzae* (Sigma, Número do Catálogo G5160), a mistura reacional foi incubada a 37°C durante uma hora adicional, para libertar a fração fluorescente do substrato. A segunda reação foi terminada pela adição de 910 μ L de solução de terminação (1% glicina, 1% carbonato de sódio, pH 10,7). Mediu-se a fluorescência da mistura resultante utilizando um comprimento de onda de medição de 359 nm e um comprimento de onda de referência de 445 nm, com 4-metilumbeliferona (sal de sódio da Sigma, Número do Catálogo M1508) servindo de padrão de referência. Uma unidade da atividade corresponde a nmoles de 4-metilumbeliferona libertada por hora.

Imunoensaios (ELISA)

Iduronato 2-Sulfatase (I2S). Uma placa de fundo plano de 96 cavidades foi revestida com um anticorpo monoclonal anti-

I2S de ratinho diluído para 10 µg/mL em 50 mM bicarbonato de sódio pH 9,6 durante 1 hora a 37°C. O anticorpo monoclonal anti-I2S de ratinho foi desenvolvido, sob contrato com a Maine Biotechnology Services, Inc. (Portland, ME), num polipéptido I2S humano purificado, produzido de modo recombinante e completo, utilizando tecnologia comum de produção de hibridomas. A placa foi lavada três vezes com 1X PBS contendo 0,1% Tween-20 e foi bloqueada durante 1 hora com 2% BSA em tampão de lavagem a 37°C. Utilizou-se tampão de lavagem com 2% BSA para diluir amostras e padrões. O padrão de I2S foi diluído e utilizado desde 100 ng/mL até 1,56 ng/mL. Após remoção do tampão de bloqueio, amostras e padrões foram aplicados na placa e incubados durante 1 hora a 37°C. O anticorpo de detecção, anticorpo anti-I2S de ratinho conjugado com peroxidase de rábano bravo, foi diluído para 0,15 µg/mL em tampão de lavagem com 2% BSA. A placa foi lavada três vezes, o anticorpo de detecção foi adicionado à placa e o sistema foi incubado durante 30 minutos a 37°C. Para desenvolver a placa preparou-se substrato TMB (Bio-Rad, Hercules, CA). A placa foi lavada 3 vezes, adicionaram-se a cada cavidade 100 µL de substrato e o sistema foi incubado durante 15 minutos a 37°C. A reação foi terminada com 2 N ácido sulfúrico (100 µL/cavidade) e a placa foi lida num leitor de placas de microtítulo a 450 nm, utilizando 655 nm como comprimento de onda de referência. N-Acetilgalactosamina 6-Sulfatase (GALNS). Dois anticorpos monoclonais anti-GALNS de ratinho proporcionaram a base para ELISA de GALNS. Os anticorpos monoclonais anti-GALNS de ratinho também foram desenvolvidos, sob contrato com a Maine Biotechnology Services, Inc. (Portland, ME), num polipéptido GALNS humano purificado, produzido de modo recombinante e completo, utilizando tecnologia comum de produção de hibridomas. O

primeiro anticorpo, para captura de GALNS, foi utilizado para revestir uma Placa F96 MaxiSorp Nunc-Immuno (Nalge Nunc, Número do Catálogo 442404) num tampão de revestimento (50 mM bicarbonato de sódio, pH 9,6). Após incubação durante uma hora a 37°C e lavagem com um tampão de lavagem, a placa foi bloqueada com tampão de bloqueio (PBS, 0,05% Tween-20, 2% BSA) durante uma hora a 37°C. Em seguida, amostras experimentais e de controlo juntamente com padrões de GALNS foram carregados na placa e o sistema foi adicionalmente incubado durante uma hora a 37°C. Após lavagem com um tampão de lavagem, o segundo anticorpo, de deteção, conjugado com HRP foi aplicado em tampão de bloqueio, seguido de 30 minutos de incubação a 37°C. Após nova lavagem da placa adicionou-se o reagente de substrato TMB da Bio-Rad e o sistema foi incubado durante 15 minutos. Depois adicionou-se 2 N ácido sulfúrico, para terminar a reação, e os resultados foram pontuados de forma espectralométrica utilizando um leitor de placas Molecular Device a um comprimento de onda de 450 nm.

Discussão

Efeito da FGE na Atividade de Sulfatase

GALNS. Observou-se um aumento de aproximadamente 50 vezes da atividade de GALNS total relativamente aos níveis de controlo (Figura 5). Este nível de atividade aumentada foi observado nos três instantes de amostragem do meio. Além disso, a atividade da GALNS foi-se acumulando linearmente ao longo do tempo, com um aumento de quatro vezes entre as 24 e as 48 horas e um aumento de duas vezes entre os instantes às 48 horas e 72 horas.

I2S. Apesar da menor magnitude absoluta, observou-se um efeito semelhante para a atividade de I2S total, tendo-se observado um aumento de aproximadamente 5 vezes da atividade de I2S total relativamente aos níveis de controlo. Este nível de atividade aumentada manteve-se durante toda a experiência. A atividade da I2S foi-se acumulando linearmente no meio ao longo do tempo, semelhante aos resultados observados com a GALNS (2,3 vezes entre as 24 e as 48 horas e 1,8 vezes entre as 48 e as 72 horas).

Efeito da FGE na Atividade *Específica* de Sulfatase

GALNS. A expressão de FGE em células 36F aumentou a atividade específica aparente da GALNS (razão entre atividade enzimática e enzima total estimada por ELISA) em 40-60 vezes relativamente aos níveis de controlo (Figura 6). O aumento da atividade específica manteve-se ao longo dos três instantes do estudo e aparentou aumentar ao longo dos três dias de acumulação pós-transfecção.

I2S. Observou-se um efeito semelhante com a I2S, tendo sido observado um aumento de 6-7 vezes da atividade específica (3-5 U/mg) relativamente aos valores de controlo (0,5 - 0,7 U/mg).

Os valores de ELISA para a GALNS (Figura 7) e I2S não foram significativamente afetados pela transfecção da FGE. Isto indica que a expressão de FGE não enfraquece as vias de tradução e secreção envolvidas na produção de sulfatases.

Em resumo, todos estes resultados para ambas as sulfatases indicam que a expressão de FGE aumenta dramaticamente a

atividade *específica* de sulfatase em linhas de células que superexpressam a GALNS e I2S.

Coexpressão de FGE (SUMF1) e outros genes de sulfatases

Para testar o efeito da FGE (SUMF1) em atividades de sulfatases adicionais em células normais, superexpressamos cDNAs de ARSA (SEQ ID NO: 14), ARSC (SEQ ID NO: 18) e ARSE (SEQ ID NO: 22) em várias linhas de células com e sem cotransfecção do cDNA de FGE (SUMF1), e medimos atividades de sulfatases. A superexpressão de cDNAs de sulfatases em células Cos-7 originou um aumento moderado da atividade de sulfatase, ao passo que se observou um acentuado aumento sinérgico (20 até 50 vezes) quando foram coexpressos um gene de sulfatase e o gene FGE (SUMF1). Observou-se um efeito semelhante, se bem que menor, em três linhas de células adicionais, HepG2, LE293 e U2OS. A superexpressão simultânea de cDNAs de múltiplas sulfatases originou um aumento menor de cada atividade de sulfatase específica em comparação com a superexpressão de uma única sulfatase, indicando a presença de competição das diferentes sulfatases para a maquinaria de modificação.

Para testar a conservação funcional do gene FGE (SUMF1) durante a evolução, superexpressamos cDNAs de ARSA, ARSC e ARSE em várias linhas de células com e sem cotransfecção do cDNA de MSD, e medimos atividades de sulfatases. Ambos os genes FGE (SUMF1) murino e de *Drosophila* foram ativos nas três sulfatases humanas, com FGE (SUMF1) de *Drosophila* sendo menos eficiente. Estes dados demonstram um elevado grau de conservação funcional de FGE (SUMF1) durante a evolução, implicando uma importância biológica significativa para a função e sobrevivência celulares.

Observou-se um efeito semelhante e consistente, se bem que muito mais fraco, utilizando o gene FGE2 (SUMF2), sugerindo que a proteína codificada por este gene também tem atividade modificadora de sulfatase. Estes dados demonstram que a quantidade da proteína codificada por FGE (SUMF1) é um fator limitador das atividades de sulfatases, uma descoberta com implicações importantes para a produção em larga escala de sulfatases ativas a serem utilizadas em terapia de substituição enzimática.

Exemplo 4:

Identificação do gene mutado em MSD através de complementação funcional utilizando transferência cromossômica mediada por microcélulas.

Numa experiência separada utilizando transferência cromossômica mediada por microcélulas através de complementação funcional, confirmámos que o gene mutado em MSD é FGE. As nossas descobertas proporcionam uma compreensão mais profunda de um novo mecanismo biológico que afeta toda uma família de proteínas em organismos remotamente relacionados. Para além de identificar a base molecular de uma doença genética rara, os nossos dados confirmam adicionalmente um efeito intensificador poderoso do produto do gene *FGE* na atividade de sulfatases. A última descoberta tem implicações clínicas diretas para a terapia de pelo menos oito doenças humanas causadas por deficiências de sulfatases.

O gene para MSD é mapeado no cromossoma 3p26

Para identificar a localização cromossômica do gene mutado em MSD tentámos resgatar as enzimas sulfatase deficientes por complementação funcional via transferência cromossômica mediada por microcélulas. Um painel de linhas de células híbridas humanas/ratinho, contendo cromossomas humanos normais individuais marcados com o marcador selecionável dominante HyTK, foi utilizado como fonte de cromossomas humanos dadores e foi fundido a uma linha de células imortalizadas de um paciente com MSD. Todos os 22 autossomas humanos foram transferidos, um por um, para a linha de células do paciente e selecionaram-se híbridos em higromicina. Foram escolhidas aproximadamente 25 colónias sobreviventes em cada uma das 22 experiências de transferência. Estas cresceram separadamente e foram recolhidas para testes enzimáticos subsequentes. As atividades da Arilsulfatase A (ARSA) (SEQ ID NO: 15), Arilsulfatase B (ARSB) (SEQ ID NO: 17) e Arilsulfatase C (ARSC) (SEQ ID NO: 19) foram testadas para cada um dos aproximadamente 440 clones (20 x 22). Esta análise indicou claramente que atividades de sulfatases de vários clones derivados da transferência do cromossoma 3 foram significativamente mais elevadas em comparação com as de todos os outros clones. Observou-se uma variabilidade acentuada quando se analisaram as atividades de cada clone individual da transferência do cromossoma 3. Para verificar se cada clone tinha um cromossoma 3 humano intacto da linha de células dadoras, utilizámos um painel de 23 marcadores genéticos polimórficos para o cromossoma 3, distribuídos uniformemente por todo o comprimento do cromossoma e previamente selecionados com base no facto de haver diferentes alelos entre as linhas de células dadoras e do paciente. Isto permitiu-nos examinar a presença do cromossoma dador e identificar possíveis perdas de regiões

específicas devido a quebra cromossômica acidental. Cada clone possuindo elevada atividade enzimática reteve a totalidade do cromossoma 3 da linha de células dadoras, ao passo que clones com baixas atividades aparentaram ter perdido todo o cromossoma com base na ausência de alelos do cromossoma 3 da linha de células dadoras. Os últimos clones provavelmente retiveram uma pequena região do cromossoma dador contendo o gene marcador selecionável que lhes permitiu sobreviver em meio contendo higromicina. Estes dados indicam que um cromossoma 3 humano normal conseguiu complementar o defeito observado na linha de células do paciente com MSD.

Para determinar a região cromossômica específica contendo o gene responsável pela atividade de complementação utilizámos híbridos do cromossoma 3 marcados com Neo que se verificou terem perdido várias porções do cromossoma. Adicionalmente, efetuámos transferência cromossômica mediada por microcélulas irradiadas de cromossomas 3 humanos marcados com HyTK. Cento e quinze híbridos irradiados do cromossoma 3 foram testados quanto a atividades de sulfatase e foram submetidos a genotipagem utilizando um painel de 31 marcadores microssatélite polimórficos abrangendo todo o cromossoma. Todos os clones exibindo elevadas atividades enzimáticas aparentaram ter retido o cromossoma 3p26. Uma análise de maior resolução utilizando marcadores adicionais desta região mapeou a localização putativa do gene de complementação entre os marcadores *D3S3630* e *D3S2397*.

Identificação do gene mutado em MSD

Investigámos genes da região genómica 3p26 quanto a mutações em pacientes com MSD. Cada exão, incluindo junções de processamento, foi amplificado por PCR e analisado por sequenciação direta. Efetuou-se análise de mutações em doze indivíduos afetados não relacionados; cinco pacientes com MSD previamente descritos e sete casos não publicados. Identificaram-se várias mutações a partir do nosso coorte de MSD no marcador de sequências expressas (EST) AK075459 (SEQ ID NOs: 4, 5), correspondendo a um gene de função desconhecida, sugerindo fortemente que este era o gene envolvido em MSD. Verificou-se que cada mutação estava ausente em 100 indivíduos de controlo, desse modo excluindo a presença de um polimorfismo de sequências. Efetuou-se análise de mutações para confirmação adicional em RNAs de pacientes transcritos de forma reversa, particularmente naqueles casos onde a análise de DNA genómico revelou a presença de uma mutação num sítio de processamento ou próximo deste, possivelmente afetando o processamento. Também foram identificadas mutações por deslocamento do quadro de leitura, sem sentido, de processamento e de sentido trocado, sugerindo que a doença é causada por uma perda do mecanismo funcional, como antecipado para uma perturbação recessiva. Isto também é consistente com a observação de que quase todas as mutações de sentido trocado afetam aminoácidos que estão altamente conservados ao longo da evolução (ver em baixo).

Tabela 8: Mutações MSD adicionais identificadas

Caso	referência	fenótipo	exão	alteração de nucleótidos	alteração de amino-ácidos
1. BA426	Conary et al, 1988	moderado	3	463T>C	S 155P
2. BA428	Burch et al, 1986	neonatal grave	3 5	463T>C 661delG	S155P deslocamento do quadro de leitura
3. BA431	Zenger et al, 1989	moderado	1	2T>G	MIR
			2	276delC	deslocamento do quadro de leitura
4. BA799	Burk et al, 1981	ligeiro-moderado	3	463T>C	S155P
5. BA806	não publicado	neonatal grave	3 9	463T>C 1045T>C	S155P R349W
6. BA807	Schmidt et al, 1995	desconhecido	3	c519+4delGTAA	omissão ex 3
7. BA809	Couchot et al, 1974	ligeiro-moderado	9 1	1076C>A 1A>G	S359X M1V
8. BA810	não publicado	grave	9 8	1042G>C 1006T>C	A348P C336R
9. BA811	não publicado	neonatal grave	9 3	1046G>A c519+4delGTAA	R349Q omissão ex 3
10. BA815	não publicado	moderado	8 5	979C>T c.603-6delC	R327X omissão ex 6
11. BA919	não publicado	ligeiro-moderado	6 9	836C>T 1033C>T	A279V R345C
12. BA920	não publicado	moderado	9 5	1033C>T 653G>A	R345C C218Y
			9	1033C>T	R345C

Foram identificadas mutações em cada paciente com MSD testado, desse modo excluindo heterogeneidade de *locus*. Não foi observada nenhuma correlação óbvia entre os tipos de mutações identificados e a gravidade do fenótipo relatado nos pacientes, sugerindo que a variabilidade clínica não é causada por heterogeneidade alélica. Em três casos, verificou-se que diferentes pacientes (caso 1 e 4, caso 6 e 9, e caso 11 e 12 na Tabela 6) continham a mesma mutação. Dois destes pacientes (caso 11 e 12) eram originários da mesma cidade da Sicília, sugerindo a presença de um efeito fundador que foi realmente confirmado por análise de haplótipos. Surpreendentemente, verificou-se que a maior

parte dos pacientes era heterozigótico composto, contendo diferentes mutações alélicas, ao passo que apenas alguns eram homozigóticos. Apesar de ser consistente com a ausência de consanguinidade relatada pelos pais, esta descoberta foi um pouco inesperada para uma perturbação recessiva muito rara como a MSD.

O gene e proteína FGE

A sequência de cDNA de consenso do cDNA da FGE humana (também usado aqui de forma intermutável com SUMF1) (SEQ ID NO: 1) foi reunida a partir de vários clones de marcadores de sequências expressas (EST) e parcialmente da sequência genômica correspondente. O gene contém nove exões e abrange aproximadamente 105 kb (ver Exemplo 1). A comparação de sequências também identificou a presença de um parálogo do gene FGE localizado no cromossoma 7 humano, que designamos FGE2 (também usado aqui de forma intermutável com SUMF2) (SEQ ID NOs: 45, 46).

Complementação funcional de deficiências de sulfatases

Fibroblastos de dois pacientes (caso 1 e 12 na Tabela 8) com MSD onde identificamos mutações do gene FGE (SUMF1) (linhas de células BA426 e BA920) foram infectados com vírus HSV contendo a forma de tipo selvagem e duas formas mutadas do cDNA de FGE (SUMF1) (R327X e Δ ex3). Testaram-se as atividades de ARSA, ARSB e ARSC 72 horas após a infecção. A expressão do cDNA de FGE (SUMF1) de tipo selvagem originou complementação funcional das três atividades, ao passo que cDNAs de FGE (SUMF1) mutantes não o fizeram (Tabela 9). Estes dados proporcionam evidências conclusivas da identidade de FGE (SUMF1) como sendo o gene de MSD e provam

a relevância funcional das mutações encontradas em pacientes. As mutações associadas a doença originam deficiência de sulfatase, desse modo demonstrando que a FGE (SUMF1) é um fator essencial para a atividade de sulfatase.

Tabela 9: Complementação funcional de deficiências de sulfatases

Linha de células MSD de recetores	construção	ARSA⁽¹⁾	ARSB⁽¹⁾	ARSC⁽¹⁾
BA426	Produto de amplificação do HSV	24,0	22,5	0,15
	SUMF1-Δex3	42,0	23,8	0,29
	SUMF1-R327X	33,6	24,2	0,16
	SUMF1	119,5 (4,9 x)	37,8 (1,7 x)	0,62 (4,1 x)
BA920	Produto de amplificação do HSV	16,6	11,3	0,15
	SUMF1-Δex3	17,2	14,4	0,07
	SUMF1-R327X	36,0	13,5	0,13
	SUMF1	66,5 (4,0 x)	21,6 (1,9 x)	0,42 (2,8 x)
Gama de controlo		123,7- 394,6	50,6- 60,7	1,80- 1,58
⁽¹⁾ Todas as atividades enzimáticas estão expressas em nmoles de 4-metilumbeliferona libertadas mg proteína ⁻¹ . 3 horas. As linhas de células MSD BA426 e BA920 foram infetadas com o produto de amplificação do HSV isoladamente, e com construções contendo cDNAs de SUMF1 mutantes ou de tipo selvagem. O aumento de atividades de arilsulfatases isoladas em fibroblastos infetados com o gene SUMF1 de tipo selvagem, em comparação com as de células infetadas com o vetor isoladamente, está indicado entre parênteses. Estão indicadas atividades medidas em fibroblastos de controlo não infetados.				

Base molecular da MSD

Com base na hipótese de que o gene da doença deve ser capaz de complementar a deficiência enzimática numa linha de células de um paciente, realizámos transferência cromossómica mediada por microcélulas para uma linha de células imortalizada de um paciente com MSD. Esta técnica tem sido utilizada com êxito para a identificação de genes cuja função prevista pode ser avaliada em linhas de células

(por exemplo, medindo a atividade enzimática ou detetando características morfológicas). Para abordar o problema da variabilidade estocástica da atividade enzimática medimos as atividades de três sulfatases diferentes (ARSA, ARSB e ARSC) no ensaio de complementação. Os resultados da transferência cromossômica indicaram claramente o mapeamento do gene de complementação no cromossoma 3. Obteve-se mapeamento sub-regional gerando um painel de híbridos de radiação para o cromossoma 3. Clones híbridos individuais foram caracterizados ao nível genômico, por tipagem de 31 marcadores microssatélite exibindo alelos diferentes entre linhas de células dadoras e recetoras, e ao nível funcional, testando atividades de sulfatases. A análise de 130 desses híbridos levou ao mapeamento da região de complementação no cromossoma 3p26.

Depois de definida a região genômica crítica, o gene FGE (SUMF1) também foi identificado por análise de mutações em DNA de pacientes. Encontraram-se mutações em todos os pacientes testados, provando que um único gene está envolvido em MSD. As mutações encontradas foram de tipos diferentes, a maioria (por exemplo, sítio de processamento, sítio de início, sem sentido, deslocamento do quadro de leitura) resultando putativamente numa função perdida da proteína codificada, como esperado para uma doença recessiva. A maior parte das mutações de sentido trocado afeta codões correspondentes a aminoácidos que têm estado altamente conservados durante a evolução, sugerindo que também estas mutações causam uma perda de função. Não foi possível extrair correlações entre o tipo de mutação e a gravidade do fenótipo, indicando que o último se deve a fatores não relacionados. Inesperadamente para uma doença genética rara, verificou-se que muitos pacientes eram

heterozigóticos compostos, contendo duas mutações diferentes. No entanto, foi identificado um efeito fundador para uma mutação originária de uma pequena cidade da Sicília.

Função do gene FGE (SUMF1)

A identidade do gene FGE (SUMF1) como "fator de complementação" foi demonstrada de modo definitivo por resgate da deficiência enzimática de quatro sulfatases diferentes por expressão de cDNA de FGE (SUMF1) exógeno, inserido num vetor viral, em duas linhas de células de pacientes diferentes. Em cada caso observou-se um restabelecimento consistente, se bem que parcial, de todas as atividades de sulfatases testadas, em comparação com linhas de células de pacientes de controlo transfectadas com vetores vazios. Em média, o aumento das atividades enzimáticas variou entre 1,7 e 4,9 vezes e atingiu aproximadamente metade dos níveis observados em linhas de células normais. A atividade enzimática está correlacionada com o número de partículas de vírus utilizado em cada experiência e com a eficiência da infeção, testada por análise de proteína marcadora (GFP). Nas mesmas experiências utilizaram-se vetores contendo cDNAs de FGE (SUMF1) contendo duas das mutações encontradas nos pacientes, R327X e Δ ex3, não tendo sido observado nenhum aumento significativo da atividade enzimática, desse modo demonstrando a relevância funcional destas mutações.

Como mencionado aqui noutro local, Schmidt *et al.* descobriram primeiramente que sulfatases sofrem uma modificação pós-tradução de uma cisteína altamente conservada, que se encontra no sítio ativo da maior parte

das sulfatases, em C α -formilglicina. Também mostraram que esta modificação era defeituosa em MSD (Schmidt, B., et al., *Cell*, 1995, 82: 271-278). Os nossos dados mutacionais e funcionais proporcionam evidências fortes de que o FGE (SUMF1) é responsável por esta modificação.

O gene FGE (SUMF1) exibe um grau extremamente elevado de conservação de sequência ao longo de todas as espécies remotamente relacionadas analisadas, desde bactérias até ao homem. Proporcionamos evidências de que o homólogo de *Drosophila* do gene FGE (SUMF1) humano é capaz de ativar sulfatases humanas superexpressas, provando que o elevado nível observado de semelhança de sequências dos genes FGE (SUMF1) de espécies remotamente relacionadas está correlacionado com uma acentuada conservação funcional. Uma exceção notável é levedura, que aparenta estar desprovida do gene FGE (SUMF1) bem como de quaisquer genes codificadores de sulfatases, indicando que a função de sulfatase não é requerida por este organismo e sugerindo a presença de uma influência recíproca na evolução de genes FGE (SUMF1) e de sulfatases.

É interessante notar que há dois genes homólogos, FGE (SUMF1) e FGE2 (SUMF2), nos genomas de todos os vertebrados analisados, incluindo humanos. Como é evidente da árvore filogenética, o gene FGE2 (SUMF2) aparenta ter evoluído independentemente do gene FGE (SUMF1). Nos nossos ensaios, o gene FGE2 (SUMF2) também é capaz de ativar sulfatases, mas fá-lo de um modo muito menos eficiente em comparação com o gene FGE (SUMF1). Isto pode ser responsável pela atividade de sulfatase residual encontrada em pacientes com MSD e sugere que uma deficiência completa de sulfatases seria letal. Neste momento não podemos excluir a

possibilidade de o gene FGE2 (SUMF2) ter uma função adicional, se bem que ainda desconhecida.

Impacto na terapia de doenças devidas a deficiência de sulfatases

Observou-se um forte aumento, até 50 vezes, de atividades de sulfatases em células que superexpressam cDNA de FGE (SUMF1) em conjunto com cDNAs de ARSA, ARSC ou ARSE, em comparação com células que superexpressam sulfatases únicas isoladas. Em todas as linhas de células encontrou-se um efeito sinérgico significativo, indicando que o FGE (SUMF1) é um fator limitador da atividade de sulfatase. No entanto, observou-se variabilidade entre diferentes sulfatases, possivelmente devido a diferente afinidade da proteína codificada por FGE (SUMF1) para as várias sulfatases. Também foi observada variabilidade entre diferentes linhas de células que podem ter níveis diferentes de enzima geradora de formilglicina endógena. De modo consistente com estas observações, verificámos que a expressão do gene MSD varia entre diferentes tecidos, com níveis significativamente elevados no rim e fígado. Isto pode ter implicações importantes, pois tecidos com níveis baixos de expressão do gene FGE (SUMF1) podem ser menos capazes de modificar efetivamente proteínas sulfatases distribuídas de modo exógeno (ver em baixo). Em conjunto, estes dados sugerem que a função do gene FGE (SUMF1) evoluiu até atingir um sistema regulador dual, em que cada sulfatase é controlada por um mecanismo individual, responsável pelos níveis de mRNA de cada gene sulfatase estrutural, e um mecanismo comum partilhado por todas as sulfatases. Adicionalmente, o FGE2 (SUMF2) proporciona redundância parcial para a modificação de sulfatases.

Estes dados têm implicações profundas na produção em massa de sulfatases ativas a serem utilizadas em terapia de substituição enzimática. Têm sido relatados estudos de substituição enzimática em modelos animais de deficiências de sulfatases, como um modelo felino de mucopolissacaridose VI, e têm demonstrado ser eficazes na prevenção e cura de vários sintomas. Estão presentemente a ser conduzidos ensaios terapêuticos em humanos para duas perturbações congénitas devidas a deficiência de sulfatases, MPSII (síndrome de Hunter) e MPSVI (síndrome de Maroteaux-Lamy), e brevemente serão estendidos a um número maior de pacientes.

Exemplo 5:

Terapia de Substituição Enzimática com GALNS ativada por FGE para Doença de Morquio MPS IVA

A causa primária da patologia esquelética em pacientes com Morquio é a acumulação de sulfato de queratano (KS) em condrócitos do disco (placa de crescimento) epifisário devido a deficiência da sulfatase lisossômica, GALNS. O objetivo principal de estudos de investigação *in vivo* foi determinar se GALNS ativada por FGE administrada intravenosamente (IV) era capaz de penetrar em condrócitos da placa de crescimento, bem como noutros tipos de células apropriados, em ratinhos normais. Não obstante uma ausência geral de anormalidades esqueléticas, um modelo de ratinho deficiente em GALNS (Morquio Knock-In -MKI, S. Tomatsu, St. Louis University, MO) também foi utilizado para demonstrar atividade bioquímica *in vivo* de GALNS ativada por FGE administrada repetidamente. A ausência de patologia esquelética em modelos de ratinho reflete o facto de o KS

esquelético estar grandemente reduzido ou ausente em roedores (Venn G, & Mason RM., *Biochem J.*, 1985, 228: 443-450). No entanto, estes ratinhos demonstraram acumulação detetável de GAG e outras anormalidades celulares em vários órgãos e tecidos. Em consequência, o objetivo global dos estudos foi demonstrar que GALNS *ativada por FGE* penetra na placa de crescimento (estudo de biodistribuição) e exibe atividade enzimática de GALNS funcional dirigida para a remoção de GAG acumulado em tecidos afetados (estudo farmacodinâmico).

Os resultados destes estudos demonstraram que GALNS *ativada por FGE* injetada IV foi interiorizada por condrócitos da placa de crescimento, apesar de a níveis relativamente baixos em comparação com outros tecidos. Adicionalmente, a injeção de GALNS *ativada por FGE* ao longo de 16 semanas em ratinhos MKI eliminou efetivamente GAG acumulado e reduziu a coloração de biomarcadores lisossômicos em todos os tecidos moles examinados. Em resumo, as experiências demonstraram com sucesso a distribuição de GALNS em condrócitos da placa de crescimento e demonstraram atividade bioquímica em termos de eliminação de GAG em múltiplos tecidos.

Estudo de Biodistribuição

Ratinhos ICR (normais) de quatro semanas de idade receberam uma única injeção IV de 5 mg/kg de GALNS *ativada por FGE*. Recolheram-se o fígado, fémur (osso), coração, rim e baço duas horas após a injeção, tendo sido preparados para exame histológico. Utilizou-se um anticorpo monoclonal anti-GALNS humana para detetar a presença de GALNS injetada nos vários tecidos. A GALNS foi detetada em todos os tecidos

examinados em comparação com controlos de veículo. Além disso, a GALNS foi facilmente observada em todos os tecidos examinados utilizando um sistema repórter de peroxidase de rábano bravo, com a exceção de osso. A demonstração da captação de GALNS na placa de crescimento requereu a utilização de um sistema repórter de isotiocianato de fluoresceína (FTTC) mais sensível e indica que, apesar de a GALNS penetrar na placa de crescimento, está menos rapidamente disponível para condrócitos da placa de crescimento do que para células de tecidos moles. Não obstante o requisito de um método de deteção fluorescente mais sensível, a distribuição de GALNS em condrócitos da placa de crescimento do osso foi observada em todas as secções da placa de crescimento examinadas, em comparação com os controlos de veículo.

Estudo Farmacodinâmico em Ratinhos MKI

Ratinhos MKI ou de tipo selvagem de quatro semanas de idade receberam semanalmente injeções IV (n=8 por grupo) até às 20 semanas de idade. Cada injeção semanal consistiu em 2 mg/kg de GALNS *ativada por FGE* ou controlo de veículo (nenhuma injeção para ratinhos de tipo selvagem). Todos os ratinhos foram sacrificados, para exame histológico, às 20 semanas de idade e foram corados empregando os métodos seguintes: hematoxilina e eosina para morfologia celular, azul de alcian para deteção de GAGs.

A eliminação de GAGs acumulados foi demonstrada por coloração reduzida ou ausente com azul de alcian em todos os tecidos moles examinados (fígado, coração, rim e baço). Isto foi observado apenas nos ratinhos injetados com GALNS. Apesar de a placa de crescimento nos ratinhos MKI ter

funcionado normalmente, evidenciado por morfologia esquelética normal, observaram-se anormalidades celulares mais subtis (incluindo vacuolização de condrócitos sem efeito patológico aparente). Os condrócitos vacuolizados das zonas hipertróficas e de proliferação da placa de crescimento não foram afetados pela administração de GALNS. Isto contrastou com os condrócitos da zona de calcificação da placa de crescimento, onde se observou uma redução da vacuolização em ratinhos injetados com GALNS. A vacuolização de condrócitos e acumulação de GAGs presumivelmente não KS na placa de crescimento em ratinhos MKI foi, em geral, surpreendente e inesperada, devido à ausência conhecida de KS na placa de crescimento de ratinhos. Estas observações particulares refletem provavelmente o facto de que, nos ratinhos "knock-in", estão presentes níveis elevados de GALNS mutante (em oposição a ratinhos "knock-out", onde não há GALNS mutante residual, nenhuma vacuolização de condrócitos da placa de crescimento e nenhuma acumulação de GAG - Tomatsu S. *et al.*, *Human Molecular Genetics*, 2003, 12: 3349-3358). O fenómeno de vacuolização na placa de crescimento pode ser indicador de um efeito secundário num subconjunto de células que expressam GALNS mutante. Ainda assim, a injeção de enzima ao longo de 16 semanas demonstrou evidências fortes de distribuição de GALNS *ativada por FGE* em múltiplos tecidos e atividade enzimática *in vivo*.

Descrição Pormenorizada das Figuras

Fig. 1: Espetros de massa MALDI-TOF de P23 após incubação na ausência (A) ou presença (B) de um extrato solúvel de microssomas de testículo bovino. Incubaram-se 6 pmol de P23, em condições padrão, durante 10 minutos a 37°C na

ausência ou presença de 1 μ L de extrato microssômico. As amostras foram preparadas para espectrometria de massa MALDI-TOF como descrito em Procedimentos Experimentais. Estão indicadas as massas monoisotópicas MH^+ do P23 (2526,28) e seu derivado FGly (2508,29).

Fig. 2: Árvore filogenética derivada de um alinhamento de FGE humana e 21 proteínas da semente PFAM-DUF323. Os números nos ramos indicam a distância filogenética. As proteínas são designadas pelo seu número TrEMBL ID e o nome da espécie. hFGE - FGE humana. Em cima à direita: escala das distâncias filogenéticas. Um asterisco indica que o gene foi adicionalmente investigado. Os sete genes de cima fazem parte da família de genes FGE.

Fig. 3: Organização do locus do gene FGE humano e murino. Os exões estão mostrados à escala na forma de retângulos pretos (*locus* humano) e retângulos claros (*locus* murino). A barra no canto inferior direito mostra a escala. As linhas entre os exões mostram os intrões (não à escala). Os números acima das linhas de intrões indicam a dimensão dos intrões em quilobases.

Fig. 4: Diagrama que mostra um mapa do Plasmídeo de Expressão de FGE pXMG.1.3.

Fig. 5: Gráfico de barras que ilustra a Atividade da N-Acetilgalactosamina 6-Sulfatase em Células 36F Transfetadas de Modo Transitório com o Plasmídeo de Expressão de FGE. Células foram transfetadas com um plasmídeo de controle, pXMG.1.2, com o cDNA de FGE na orientação reversa, ou um plasmídeo de expressão de FGE, pXMG.1.3 em meio sem metotrexato (MTX). Vinte e quatro horas mais tarde, as

células foram novamente nutridas com meio contendo 1,0 μ M MTX. O meio foi recolhido e as células foram recolhidas às 24, 48 e 72 horas após a nova nutrição. A atividade de N-Acetilgalactosamina 6-Sulfatase foi determinada por ensaio de atividade. Cada valor mostrado é a média de duas transfeções separadas, com desvios padrão indicados por barras de erro.

Fig. 6: Gráfico de barras que ilustra a Atividade Específica da N-Acetilgalactosamina 6-Sulfatase em Células 36F Transfetadas de Modo Transitório com o Plasmídeo de Expressão de FGE. Células foram transfetadas com um plasmídeo de controlo, pXMG.1.2, com o cDNA de FGE na orientação reversa, ou um plasmídeo de expressão de FGE, pXMG.1.3 em meio sem metotrexato (MTX). Vinte e quatro horas mais tarde, as células foram novamente nutridas com meio contendo 1,0 μ M MTX. O meio foi recolhido e as células foram recolhidas às 24, 48 e 72 horas após a nova nutrição. A atividade específica de N-Acetilgalactosamina 6-Sulfatase foi determinada por ensaio de atividade e ELISA e está representada como uma razão de atividade de N-Acetilgalactosamina 6-Sulfatase por mg de N-Acetilgalactosamina 6-Sulfatase reativa em ELISA. Cada valor mostrado é a média de duas transfeções separadas.

Fig. 7: Gráfico de barras que ilustra a Produção da N-Acetilgalactosamina 6-Sulfatase em Células 36F Transfetadas de Modo Transitório com o Plasmídeo de Expressão de FGE. Células foram transfetadas com um plasmídeo de controlo, pXMG.1.2, com o cDNA de FGE na orientação reversa, ou um plasmídeo de expressão de FGE, pXMG.1.3 em meio sem metotrexato (MTX). Vinte e quatro horas mais tarde, as células foram novamente nutridas com meio contendo 1,0 μ M

MTX. O meio foi recolhido e as células foram recolhidas às 24, 48 e 72 horas após a nova nutrição. A proteína total N-Acetilgalactosamina 6-Sulfatase foi determinada por ELISA. Cada valor mostrado é a média de duas transfeções separadas, com desvios padrão indicados por barras de erro.

Fig. 8: Gráfico de barras que ilustra a Atividade da Iduronato 2-Sulfatase em Células 30C6 Transfetadas de Modo Transitório com o Plasmídeo de Expressão de FGE. Células foram transfetadas com um plasmídeo de controlo, pXMG.1.2, com o cDNA de FGE na orientação reversa, ou um plasmídeo de expressão de FGE, pXMG.1.3 em meio sem metotrexato (MTX). Vinte e quatro horas mais tarde, as células foram novamente nutridas com meio contendo 0,1 μ M MTX. O meio foi recolhido e as células foram recolhidas às 24, 48 e 72 horas após a nova nutrição. A atividade da Iduronato 2-Sulfatase foi determinada por ensaio de atividade. Cada valor mostrado é a média de duas transfeções separadas.

Fig. 9: Ilustra um estojo que implementa características da presente invenção.

Todas as referências divulgadas aqui são incorporadas por referência na sua totalidade. O que é reivindicado está apresentado em baixo e é seguido de uma Listagem de Sequências.

LISTAGEM DE SEQUÊNCIAS

<110> Transkaryotic Therapies, Inc.

von Figura, Kurt

Schmidt, Bernhard

Dierks, Thomas

Heartlein, Michael W.

Cosma, Maria P.

Ballabio, Andrea

<120> DIAGNÓSTICO E TRATAMENTO DE DEFICIÊNCIA DE MÚLTIPLAS
SULFATASES E OUTRAS DEFICIÊNCIAS DE SULFATASES

<130> 0403WO

<150> US 60/447,747

<151> 2003-02-11

<160> 95

<170> PatentIn versão 3.2

<210> 1

<211> 1180

<212> DNA

<213> *Homo sapiens*

<220>

<221> CDS

<222> (20)..(1141)

<223> CDNA de FGE

<400> 1

acatggcccg cgggacaac atg gct gcg ccc gca cta ggg ctg gtg tgt gga	52
Met Ala Ala Pro Ala Leu Gly Leu Val Cys Gly	
1 5 10	
cgt tgc cct gag ctg ggt ctc gtc ctc ttg ctg ctg ctg ctc tcg ctg	100
Arg Cys Pro Glu Leu Gly Leu Val Leu Leu Leu Leu Leu Ser Leu	
15 20 25	
ctg tgt gga gcg gca ggg agc cag gag gcc ggg acc ggt gcg ggc gcg	148
Leu Cys Gly Ala Ala Gly Ser Gln Glu Ala Gly Thr Gly Ala Gly Ala	
30 35 40	
ggg tcc ctt gcg ggt tct tgc ggc tgc ggc acg ccc cag cgg cct ggc	196
Gly Ser Leu Ala Gly Ser Cys Gly Cys Gly Thr Pro Gln Arg Pro Gly	
45 50 55	
gcc cat ggc agt tcc gca gcc gct cac cga tac tcc cgg gag gct aac	244
Ala His Gly Ser Ser Ala Ala Ala His Arg Tyr Ser Arg Glu Ala Asn	
60 65 70 75	
gct ccg ggc ccc gta ccc gga gag cgg caa ctc gcg cac tca aag atg	292
Ala Pro Gly Pro Val Pro Gly Glu Arg Gln Leu Ala His Ser Lys Met	
80 85 90	
gtc ccc atc cct gct gga gta ttt aca atg ggc aca gat gat cct cag	340
Val Pro Ile Pro Ala Gly Val Phe Thr Met Gly Thr Asp Asp Pro Gln	
95 100 105	
ata aag cag gat ggg gaa gca cct gcg agg aga gtt act att gat gcc	388
Ile Lys Gln Asp Gly Glu Ala Pro Ala Arg Arg Val Thr Ile Asp Ala	
110 115 120	
ttt tac atg gat gcc tat gaa gtc agt aat act gaa ttt gag aag ttt	436
Phe Tyr Met Asp Ala Tyr Glu Val Ser Asn Thr Glu Phe Glu Lys Phe	
125 130 135	
gtg aac tca act ggc tat ttg aca gag gct gag aag ttt ggc gac tcc	484
Val Asn Ser Thr Gly Tyr Leu Thr Glu Ala Glu Lys Phe Gly Asp Ser	
140 145 150 155	
ttt gtc ttt gaa ggc atg ttg agt gag caa gtg aag acc aat att caa	532
Phe Val Phe Glu Gly Met Leu Ser Glu Gln Val Lys Thr Asn Ile Gln	
160 165 170	
cag gca gtt gca gct gct ccc tgg tgg tta cct gtg aaa ggc gct aac	580

Gln	Ala	Val	Ala	Ala	Ala	Pro	Trp	Trp	Leu	Pro	Val	Lys	Gly	Ala	Asn		
			175					180					185				
tgg	aga	cac	cca	gaa	ggg	cct	gac	tct	act	att	ctg	cac	agg	ccg	gat	628	
Trp	Arg	His	Pro	Glu	Gly	Pro	Asp	Ser	Thr	Ile	Leu	His	Arg	Pro	Asp		
		190					195					200					
cat	cca	gtt	ctc	cat	gtg	tcc	tgg	aat	gat	gcg	gtt	gcc	tac	tgc	act	676	
His	Pro	Val	Leu	His	Val	Ser	Trp	Asn	Asp	Ala	Val	Ala	Tyr	Cys	Thr		
		205				210					215						
tgg	gca	ggg	aag	cgg	ctg	ccc	acg	gaa	gct	gag	tgg	gaa	tac	agc	tgt	724	
Trp	Ala	Gly	Lys	Arg	Leu	Pro	Thr	Glu	Ala	Glu	Trp	Glu	Tyr	Ser	Cys		
		220			225					230					235		
cga	gga	ggc	ctg	cat	aat	aga	ctt	ttc	ccc	tgg	ggc	aac	aaa	ctg	cag	772	
Arg	Gly	Gly	Leu	His	Asn	Arg	Leu	Phe	Pro	Trp	Gly	Asn	Lys	Leu	Gln		
				240				245						250			
ccc	aaa	ggc	cag	cat	tat	gcc	aac	att	tgg	cag	ggc	gag	ttt	ccg	gtg	820	
Pro	Lys	Gly	Gln	His	Tyr	Ala	Asn	Ile	Trp	Gln	Gly	Glu	Phe	Pro	Val		
			255					260					265				
acc	aac	act	ggt	gag	gat	ggc	ttc	caa	gga	act	gcg	cct	gtt	gat	gcc	868	
Thr	Asn	Thr	Gly	Glu	Asp	Gly	Phe	Gln	Gly	Thr	Ala	Pro	Val	Asp	Ala		
		270				275						280					
ttc	cct	ccc	aat	ggt	tat	ggc	tta	tac	aac	ata	gtg	ggg	aac	gca	tgg	916	
Phe	Pro	Pro	Asn	Gly	Tyr	Gly	Leu	Tyr	Asn	Ile	Val	Gly	Asn	Ala	Trp		
		285				290					295						
gaa	tgg	act	tca	gac	tgg	tgg	act	gtt	cat	cat	tct	gtt	gaa	gaa	acg	964	
Glu	Trp	Thr	Ser	Asp	Trp	Trp	Thr	Val	His	His	Ser	Val	Glu	Glu	Thr		
					305					310					315		
ctt	aac	cca	aaa	ggt	ccc	cct	tct	ggg	aaa	gac	cga	gtg	aag	aaa	ggt	1012	
Leu	Asn	Pro	Lys	Gly	Pro	Pro	Ser	Gly	Lys	Asp	Arg	Val	Lys	Lys	Gly		
				320					325					330			
gga	tcc	tac	atg	tgc	cat	agg	tct	tat	tgt	tac	agg	tat	cgc	tgt	gct	1060	
Gly	Ser	Tyr	Met	Cys	His	Arg	Ser	Tyr	Cys	Tyr	Arg	Tyr	Arg	Cys	Ala		
			335					340					345				
gct	cgg	agc	cag	aac	aca	cct	gat	agc	tct	gct	tcg	aat	ctg	gga	ttc	1108	
Ala	Arg	Ser	Gln	Asn	Thr	Pro	Asp	Ser	Ser	Ala	Ser	Asn	Leu	Gly	Phe		
			350				355					360					
cgc	tgt	gca	gcc	gac	cgc	ctg	ccc	acc	atg	gac	tgaca	accaa	ggg	tagt	cctt	1161	
Arg	Cys	Ala	Ala	Asp	Arg	Leu	Pro	Thr	Met	Asp							
			365			370											
ccccagtc	cca	aggagc	cagt													1180	

<210> 2

<211> 374

<212> PRT

<213> *Homo sapiens*

<400> 2

Met Ala Ala Pro Ala Leu Gly Leu Val Cys Gly Arg Cys Pro Glu Leu
 1 5 10 15
 Gly Leu Val Leu Leu Leu Leu Leu Ser Leu Leu Cys Gly Ala Ala
 20 25 30
 Gly Ser Gln Glu Ala Gly Thr Gly Ala Gly Ala Gly Ser Leu Ala Gly
 35 40 45
 Ser Cys Gly Cys Gly Thr Pro Gln Arg Pro Gly Ala His Gly Ser Ser
 50 55 60
 Ala Ala Ala His Arg Tyr Ser Arg Glu Ala Asn Ala Pro Gly Pro Val

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

<210> 3

<211> 1122

<212> DNA

<213> *Homo sapiens*

<400> 3

```

atggctgcgc ccgcactagg gctgggtgtgt ggacgttgcc ctgagctggg tctcgctcctc    60
ttgctgctgc tgctctcgct gctgtgtgga gcggcaggga gccaggaggc cgggaccgggt    120
gggggcgcgg ggtcccttgc gggttcttgc ggctgcggca cgcgccagcg gcctggcgcc    180
catggcagtt cggcagccgc tcaccgatac tcgcgggagg ctaacgctcc ggcccccgta    240
cccggagagc ggcaactcgc gcactcaaag atgggtcccca tccttgcctg agtattttaca    300
atgggacacag atgacccetca gataaagcag gatggggaag cacctgcgag gagagttact    360
attgatgcct ttacatgga tgcttatgaa gtcagtaata ctgaatttga gaagtttgtg    420
aactcaactg gctatttgac agaggctgag aagtttggcg actcctttgt ctttgaaggc    480
atgttgagtg agcaagtga gaccaatatt caacaggcag ttgcagctgc tccttggtgg    540
ttacctgtga aaggcgctaa ctggagacac ccagaagggc ctgactctac tattctgcac    600
aggccggatc atccagttct ccattgtgctc tggaatgatg cggttgccta ctgcacttgg    660
gcagggaagc ggctgcccac ggaagctgag tgggaatata gctgtcgagg aggcctgcac    720
aatagacttt tcccctgggg caacaaactg cagcccaaag gccagcatta tgccaacatt    780
tggcagggcg agtttccggg gaccaacact ggtgaggatg gcttccaagg aactgcgcct    840
gttgatgcct tccctcccaa tggttatggc ttatacaaca tagtggggaa cgcattggaa    900
tggacttcag actggtggac tgttcatcat tctgttgaag aaacgcttaa cccaaaaggt    960
cccccttctg ggaaagaccg agtgaagaaa ggtggatcct acatgtgcca taggtcttat   1020
tgttacaggt atcgctgtgc tgctcggagc cagaacacac ctgatagctc tgcttcgaat   1080
ctgggattcc gctgtgcagc cgaccgcctg cccaccatgg ac                               1122

```

<210> 4

<211> 2130

<212> DNA

<213> *Homo sapiens*

<400> 4

acatggcccg	cgggacaaca	tggctgcgcc	cgcactaggg	ctggtgtgtg	gacgttgccc	60
tgagctgggt	ctcgtcctct	tgctgctgct	gctctcgctg	ctgtgtggag	cggcagggag	120
ccaggaggcc	gggaccggtg	cgggcgcggg	gtcccttgcg	ggttcttgcg	gctgcggcac	180
gccccagcgg	cctggcgccc	atggcagttc	ggcagccgct	caccgatact	cgcgggaggc	240
taacgctccg	ggccccgtac	cgggagagcg	gcaactcgcg	cactcaaaga	tggccccat	300
ccctgctgga	gtatttacaa	tgggcacaga	tgatccctcag	ataaagcagg	atggggaagc	360
acctgcgagg	agagtacta	ttgatgccct	ttacatggat	gcctatgaag	tcagtaatac	420
tgaatttgag	aagtttgtga	actcaactgg	ctatttgaca	gaggctgaga	agtttggcga	480
ctcctttgtc	tttgaaggca	tgttgagtga	gcaagtgaag	accaatattc	aacaggcagt	540
tgcagctgct	ccctggtggt	tacctgtgaa	aggcgctaac	tggagacacc	cagaagggcc	600
tgactctact	attctgcaca	ggccggatca	tccagttctc	catgtgtcct	ggaatgatgc	660
ggttgccctac	tgcacttggg	caggggaagcg	gctgcccacg	gaagctgagt	gggaatacag	720
ctgtcgagga	ggcctgcata	atagactttt	cccctggggc	aacaaactgc	agcccaaaag	780
ccagcattat	gccaacattt	ggcagggcga	ttttccggtg	accaacactg	gtgaggatgg	840
cttccaagga	actgcgcctg	ttgatgcctt	ccctcccaat	ggttatggct	tatacaacat	900
agtggggaac	gcatgggaat	ggacttcaga	ctggtggact	gttcatcatt	ctgttgaaga	960
aacgcttaac	ccaaaaggtc	ccccttcttg	gaaagaccga	gtgaagaaag	gtggatccta	1020
catgtgccat	aggctctatt	gttacaggta	tcgctgtgct	gctcggagcc	agaacacacc	1080
tgatagctct	gcttcgaatc	tgggattccg	ctgtgcagcc	gaccgcctgc	ccaccatgga	1140
ctgacaacca	agggtagtct	tccccagtcc	aaggagcagt	cgtgtctgac	ctacattggg	1200
ctttcctcag	aactttgaac	gatcccatgc	aaagaattcc	caccctgagg	tgggttacat	1260
acctgcccac	tggccaaagg	aaccgccttg	tgagaccaaa	ttgctgacct	gggtcagtgc	1320
atgtgcttta	tgggtgtgtg	catcttttga	gatcatcacc	atattttact	tttgagagtc	1380
tttaagagg	aaggggagtg	gaggggaacc	tgagctaggg	ttcaggaggc	ccgcatocta	1440
cgcaggctct	gccacagggg	ttagacccca	ggtccgacgc	ttgaccttcc	tgggcctcaa	1500
gtgcctctcc	ctatcaaatg	aaggaaatgga	cagcatgacc	tctgggtgtc	tctccaaactc	1560
accagttcta	aaaagggtat	cagattctat	tgtgacttca	tagaatttat	gatagattat	1620
tttttagcta	ttttttccat	gtgtgaacct	tgagtgtatc	taatcatgta	aagtaagagt	1680
tctcttatgt	attatgttct	gaagaggggt	gtggtgactc	ctttatatte	gtactgcact	1740
ttgtttttcc	aaggaaatca	gtgtctttta	cgttggttat	atgaatccca	catggggccg	1800
gtgatgggat	gctgaagtcc	agcctgtgaa	cacataggaa	tgtctgtggg	gtgactctac	1860
tgtgctttat	cttttaacat	taagtgcctt	tggttcagag	gggcagtcac	aagctctggt	1920
tccccctctc	cccaaagcct	tcagcgaacg	tgaaatgtgc	gctaaacggg	gaaacctgtt	1980
taattctaga	tatagggaac	aaggaaacgag	gaccttgaat	gagctatatt	cagggtatcc	2040
ggtattttgt	aatagggaat	aggaaacctt	gttggtgtgt	gaatatccga	tgctttgaat	2100
catgcactgt	gttgaataaa	cgtatctgct				2130

<211> 374

<212> PRT

<213> *Homo sapiens*

<400> 5

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

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

<210> 6

<211> 2297

<212> DNA

<213> *Homo sapiens*

<400> 6

```

cggctgtggt ggcagttctt catgggttcc cgacgaggag gtctctgtgg ctgcggcggc 60
tgctaactgc gccacctgct gcagcctgtc cccgccgttc tgaagcggcc gcgtcgagac 120
cgaaatgccc ccaccccgga ccggccgagg ccttctctgg ctgggtctgg ttctgagctc 180
cgtctgcgtc gccctcggtt ccgaaacgca ggccaactcg accacagatg ctctgaaagt 240
tcttctcacc atcgtggatg acctgcgccc ctccctgggc tgttatgggg ataagctggt 300
gagggtcccca aatattgacc aactggcacc ccacagcctc ctcttccaga atgcctttgc 360
gcagcaagca gtgtgcgccc cgagccgctt tcttttcttc actggcagga gacctgacac 420
caccgcctg tacgaactca actcctactg gagggtgcac gctggaaact tctccaccat 480
ccccagtag ttcaaggaga atggtatgtt gacctgtcgt gtgggaaaag tctttcaccc 540
tggtgatctt tctaaccata ccgatgatcc tccgtatagc tggctctttc caccctatca 600
tcttctctct gagaagtatg aaaacactaa gacatgtcga gggccagatg gagaactcca 660
tgccaacctg ctttgccttg tggatgtgct ggatgttccc gagggcacct tgcctgacaa 720
acagagcact gagcaagcca tacagttggt ggaagagatg aaaacgtcag ccagtccttt 780
cttctcgccc gttgggtatc ataagccaca catcccttc agatacccca aggaatttca 840
gaagttgtat cccttgaga acatcacctt ggcctccgat cccgaggtcc ctgatggcct 900
acccctgtg gctacaacc cctggatgga catcaggcaa cgggaagacg tccaagcctt 960
aaacatcagt gtgccgtatg gtccaattcc tgtggacttt cagcggaaaa tccgccagag 1020
ctactttgcc tctgtgtcat atttggtatc acaggtcggc cgcctcttga gtgctttgga 1080
cgatcttcag ctggccaaca gcaccatcat tgcatttacc tcggatcatg ggtgggctct 1140
aggtgaacat ggagaatggg ccaatacag caattttgat gttgctaccc atgttccctt 1200
gatattctat gttcctggaa ggacggcttc acttccggag gcaggcgaga agcttttccc 1260
ttacctcgac ctttttgatt ccgcctcaca gttgatggag ccaggcaggc aatccatgga 1320
ccttggtgaa cttgtgtctc tttttccac gctggctgga cttgcaggac tgcaggttcc 1380
acctcgctgc ccgttccctt catttcacgt tgagctgtgc agagaaggca agaaccctct 1440
gaagcatttt cgattccgtg acttggaaga ggatccgtac ctccctggta atcccggtga 1500
actgattgcc tatagccagt atccccggc ttcagacatc ctcagtgga attctgacaa 1560
gccgagttta aaagatataa agatcatggg ctattccata cgcaccatag actataggta 1620
tactgtgtgg gttggcttca atcctgatga atttctagct aacttttctg acatccatgc 1680
aggggaactg tattttgtgg attctgaccc attgcaggat cacaatatgt ataattgttc 1740
ccaaggtgga gatcttttcc agttgttgat gccttgagtt ttgccaaaca tggatggcaa 1800
atgtgatgtg ctcccttcca gctggtgaga ggaggagtta gagctggtcg ttttgtgatt 1860
accataata ttggaagcag cctgagggtt agttaatcca aacatgcac aacaatttgg 1920
cctgagaata tgtaacagcc aaaccttttc gtttagtctt tattaaaatt tataattggt 1980
aattggacca gttttttttt taatttccct ctttttaaaa cagttacggc ttatttactg 2040
aataaataca aagcaaaaca actcaagtta tgtcatacct ttggatacga agaccataca 2100
taataaccaa acataacatt atacacaaag aatactttca ttatttgtgg aatttagtgc 2160
atttcaaaaa gtaatcatat atcaaaactg gcaccacact aagttcctga ttratttght 2220
tataatttaa taatatatct tatgagccct atalattcaa aatattatgt taacatgtaa 2280
tccatgtttc tttttcc 2297

```

<210> 7

<211> 550

<212> PRT

<213> *Homo sapiens*

<400> 7

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

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

<210> 8

<211> 2657

<212> DNA

<213> *Homo sapiens*

<400> 8

gaattccggg ccatgagctg ccccgtagcc gctgtgtgag cgctgtgtgt agtcctgggg 60

 ctctgccggg cgcgtccccc gaacgcactg ctgctcctcg cggatgacgg aggcctttgag 120
 agtgggcgct acaacaacag cgccatcgcc accccgcacc tggacgcctt ggcccgccgc 180
 agcctcctct ttcgcaatgc cttcacctcg gtcagcagct gctctcccag ccgcgccagc 240
 ctctcactcg gcttgcccca gcatcagaat gggatgtacg ggctgcacca ggacgtgcac 300
 cacttcaact ccttcgacaa ggtgaggagc ctgcccgtgc tgctcagcca agctgggtgtg 360
 cgcacaggca tcatcgggaa gaagcacgtg gggccggaga ccgtgtaccc gtttgacttt 420
 gcgtacacgg aggagaatgg ctccgtcctc caggtggggc ggaacatcac tagaattaag 480
 ctgctcgtcc ggaattcct gcagactcag gatgaccggc ctttcttctt ctacgtcgcc 540
 ttccacgacc cccaccgctg tgggcactcc cagccccagt accggaacctt ctgtgagaag 600
 tttggcaacg gagagagcgg catgggtcgt atcccagact ggacccccca ggctacgac 660
 ccactggacg tgetgggtgc ttacttcgtc cccaacaccc cggcagcccc agccgacctg 720
 gccgtcagc acaccacgt cgcccgcatg gaccaaggag ttggactggt gctccaggag 780
 ctgcgtgacg ccggtgtcct gaacgacaca ctggtgatct tcacgtccga caacgggac 840
 ccttcccca gcggcaggac caacctgtac tggccgggca ctgctgaacc cttactggtg 900
 tcatccccgg agcaccaaa acgctggggc caagtcagcg aggcctacgt gagcctccta 960
 gacctcacgc ccaccatctt ggattggttc tcgatcccg accccagcta cccatcttt 1020
 ggctcgaaga ccatccacct cactggccgg tcctcctgc cggcgctgga ggccgagccc 1080
 ctctgggcca cegtctttgg cagccagagc caccacgagg tcaccatgtc ctaccccatg 1140
 cgctccgtgc agcaccggca ctccgcctc gtgcacaacc tcaacttcaa gatgcccttt 1200
 cccatcgacc aggaattcta cgtctcacc acccttcagg acctcctgaa ccgcaccaca 1260
 gctggtcagc ccacgggctg gtacaaggac ctccgtcatt actactaccg ggccgctgg 1320
 gagctctacg accggagccg ggacccccac gagaccaga acctggccac cgaccgcgc 1380
 tttgtcagc ttctggagat gcttcgggac cagctggcca agtggcagtg ggagaccac 1440
 gacccctggg tgtgcgcccc cgacggcgtc ctggaggaga agctctctcc ccagtgccag 1500
 cccctccaca atgagctgtg accatcccag gaggcctgtg cacacatccc aggcagtctc 1560
 cagacacatc ccacacgtgt ccgtgtggcc ggccagcctg gggagtagtg gcaacagccc 1620
 ttccgtccac actcccatcc aaggagggtt ctctctctt gtggggtcac tcttgccatt 1680
 gcctggaggg ggaccagagc atgtgaccag agcatgtgcc cagccccctc accaccaggg 1740
 gcactgccgt catggcaggg gacacagttg tccttgtgtc tgaaccatgt ccagacgg 1800
 gaattctaga catacgtggt ctgcccagag ggcagcgc ccagcccatg acaagggagt 1860
 ettggtttct ggcttggttt ggggacctgc aaatgggagg cctgaggccc tcttcaggct 1920
 ttggcagcca cagatacttc tgaaccttc acagagagca ggcaggggct tcggtgccgc 1980
 gtgggcagta cgcaggtccc accgacactc acctgggagc acggcgccctg gctcttacca 2040
 gcgtctggcc tagaggaagc ctttgagcga cctttgggca ggtttctgt tcttctgttt 2100
 tgcccatggt caagtccctg ttccccaggc aggtttcagc tgattggcag caggctccct 2160
 gagtgatgag cttgaacctg tgggtttctt gggcagaagc ttatcttttt tgagagtgtc 2220
 cgaagatgaa ggcagtgcca tgcccgctct ctggcttggg ttaattcttc ggtgacactg 2280
 gcattgctgg gtggtgatgc ccgtcctctg gcttgggtta attcttcggg gacactggcg 2340
 ttgctgggtg gcaatgcccc tcctctgggt tgggttaatt cttcggtgac actggcggtg 2400

ctgggtggcg atgcccgtcc tctggcttgg gttaattctt ggatgacgtc ggcgttgctg	2460
ggagaatgtg ccgttcctgc cctgcctcca cccacctcgg gaggagaagc ccggcctgga	2520
cacccctcgg cctggacacc cctcgaagga gagggcgctt ccttgagtag gtgggctccc	2580
cttgcccttc cctccctatc actccatact ggggtgggct ggaggaggcc acaggccagc	2640
tattgtaaaa gcttttc	2657

- <210> 9
- <211> 502
- <212> PRT
- <213> *Homo sapiens*
- <400> 9

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

```

Thr Thr Val Gly Arg Met Asp Gln Gly Val Gly Leu Val Leu Gln Glu
245                250

Leu Arg Asp Ala Gly Val Leu Asn Asp Thr Leu Val Ile Phe Thr Ser
260                265                270

Asp Asn Gly Ile Pro Phe Pro Ser Gly Arg Thr Asn Leu Tyr Trp Pro
275                280                285

Gly Thr Ala Glu Pro Leu Leu Val Ser Ser Pro Glu His Pro Lys Arg
290                295                300

Trp Gly Gln Val Ser Glu Ala Tyr Val Ser Leu Leu Asp Leu Thr Pro
305                310                315

Thr Ile Leu Asp Trp Phe Ser Ile Pro Tyr Pro Ser Tyr Ala Ile Phe
325                330                335

Gly Ser Lys Thr Ile His Leu Thr Gly Arg Ser Leu Leu Pro Ala Leu
340                345                350

Glu Ala Glu Pro Leu Trp Ala Thr Val Phe Gly Ser Gln Ser His His
355                360                365

Glu Val Thr Met Ser Tyr Pro Met Arg Ser Val Gln His Arg His Phe
370                375                380

Arg Leu Val His Asn Leu Asn Phe Lys Met Pro Phe Pro Ile Asp Gln
385                390                395                400

Asp Phe Tyr Val Ser Pro Thr Phe Gln Asp Leu Leu Asn Arg Thr Thr
405                410                415

Ala Gly Gln Pro Thr Gly Trp Tyr Lys Asp Leu Arg His Tyr Tyr Tyr
420                425                430

Arg Ala Arg Trp Glu Leu Tyr Asp Arg Ser Arg Asp Pro His Glu Thr
435                440                445

Gln Asn Leu Ala Thr Asp Pro Arg Phe Ala Gln Leu Leu Glu Met Leu
450                455                460

Arg Asp Gln Leu Ala Lys Trp Gln Trp Glu Thr His Asp Pro Trp Val
465                470                475                480

Cys Ala Pro Asp Gly Val Leu Glu Glu Lys Leu Ser Pro Gln Cys Gln
485                490                495

Pro Leu His Asn Glu Leu
500

```

<210> 10

<211> 1014

<212> DNA

<213> *Homo sapiens*

<400> 10

cgtgcctgta atcccagcag ctactcactc aggaggctga ggcaggagaa tctcttgaac 60
 ccggaaggca gaggttgcag tgagccaaga tcgcgccact gaactccagc ctgggtgaca 120

gagtgagact gtctcagaac agcaacaaca aaatgcccgc tgctgctggg tccagaagag 180
 cttgaataac tgcattgtct ttttctcaat ttccatttcc cagaactggg cacctccggg 240
 ctgtgaaaag ttagggaagt gtctgacacc tccagaatcc attcccaaga agtgcctctg 300
 gtcccactag cacctgcgca gactcaggcc aggcctagaa tctccagttg gccctgcaag 360
 tgcctggagg aaggatggct ctggcctcgg tcctccccc accctgccc agccagacag 420
 acagcacctg cagacgcagg gggactgcac aattccacct gccaggacc tgaacctggc 480
 gtgtgcttgg ccctcctcct cgcccacggc gcctcagatt tcaggacct cctcctcgcc 540
 caccggcgct cagacctcag gacctgcg tcctcacgct ttgtgaacct caaatatctg 600
 agaccagtct cagtttatct tgccaaggct aaggatgcac ctgtgacagc ctcaggaggt 660
 cctgacaaca ggtgcccgag gtggctgggg atacagtctg cttttataca tcttagggag 720
 acacaagatc agtatgtgta tggcgtagat tgggttcagtc agccttcac tgaatacacg 780
 attgagtctg gcccagtga tccgcatttt tatgtaaaca gtaagggaac ggggcaatca 840
 tataagcgtt tgtctcaggg gagccccaga gggatgaact ccagttccgt ctgtcctttg 900
 tccacaagga atttccctgg gcgctaatta tgaggagggc gtgtagcttc ttatcattgt 960
 agctatgtta tttagaaata aaacgggagg caggtttgcc taattccag gttg 1014

<210> 11

<211> 522

<212> PRT

<213> *Homo sapiens*

<400> 11

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

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

Pro Ala Gln Pro Gln Leu Asn Val Cys Asn Trp Ala Val Met Asn Trp
 485 490 495

Ala Pro Pro Gly Cys Glu Lys Leu Gly Lys Cys Leu Thr Pro Pro Glu
 500 505 510

Ser Ile Pro Lys Lys Cys Leu Trp Ser His
 515 520

<210> 12

<211> 2379

<212> DNA

<213> *Homo sapiens*

<400> 12

```

ggaattccgg tcggcctctc gcccttcagc tacctgtgcg tccctccgtc ccgtcccgtc 60
ccgggggtcac cccggagcct gtccgctatg ccgtccctgc ctctagcccc aggtcggtctc 120
cggcggggca gcccccgcca cctgccctcc tgcagcccag cgtctctact gctgggtgctg 180
ggcggctgcc tgggggtctt cgggggtggc gcgggaaccc ggaggcccaa cgtgggtgctg 240
ctcctcacgg acgaccagga cgaagtgtgc ggccgcatga caccactaaa gaaaacccaa 300
gctctcatcg gagagatggg gatgaacttt tccagtgtct atgtgccaa gctctctctg 360
tgccccagca gagccagtat cctgacagga aagtaccac ataatcatca cgttgtgaac 420
aacactctgg aggggaactg cagtagtaag tcctggcaga agatccaaga accaaatact 480
ttcccagcaa ttctcagatc aatgtgtggt tatcagacct tttttgcagg gaaatattta 540
aatgagtacg gagccccaga tgcaggtgga ctagaacacg ttctctctgg ttggagtac 600
tggtatgcct tggaaaagaa ttctaagtat tataattaca cctgtctat caatgggaag 660
gcacggaagc atggtgaaaa ctatagtgtg gactacctga cagatgtttt ggctaattgtc 720
tccttggaact ttctggacta caagtccaac tttagaccct tcttcatgat gatcgccact 780
ccagcgcttc attcgcttg gacagctgca cctcagtacc agaaggcttt ccagaatgtc 840
tttgacccaa gaaacaagaa cttcaacatc catggaacga acaagcactg gtttaattagg 900
caagccaaga ctccaatgac taattcttca atacagtttt tagataatgc atttaggaaa 960
aggtggcaaa ctctctctc agttgatgac cttgtggaga aactgggtcaa gaggctggag 1020
ttcactgggg agctcaacaa cacttacatc ttctatacct cagacaatgg ctatcacaca 1080
ggacagtttt ccttgccaat agacaagaga cagctgtatg agtttgatat caaagttcca 1140
ctgttggttc gaggacctgg gatcaaacca aatcagacaa gcaagatgct ggttgccaac 1200
attgacttgg gtccctactat tttggacatt gctggctacg acctaaataa gacacagatg 1260
gatgggatgt ccttattgcc ctttttgaga ggtgccagta acttgacctg gcgatcagat 1320
gtcctggttg aataccaagg agaaggccgt aacgtcactg acccaacatg cccttccttg 1380
agtcctggcg tatctcaatg ctcccagac tgtgtatgtg aagatgttta taacaatacc 1440
tatgcctgtg tgaggacaat gtcagcattg tggaaattgc agtattgcga gtttgatgac 1500
caggaggtgt ttgtagaagt ctataatctg actgcagacc cagaccagat cactaacatt 1560
gctaaaacca tagaccaga gcttttagga aagatgaact atcggttaat gatgttacag 1620
tcctgttctg ggccaacctg tcgcactcca ggggtttttg accccggata cagggttgac 1680
ccccgtctca tgttcagcaa tcgcggcagt gtcaggactc gaagattttc caaacatctt 1740

```

```

ctgttagcgac ctcacacagc ctctgcagat ggatccctgc acgcctcttt ctgatgaagt 1800
gattgtagta ggtgtctgta gctagtcttc aagaccacac ctggaagagt ttctgggctg 1860
gctttaagtc ctgtttgaaa aagcaacca gtcagctgac ttcctcgtgc aatgtgttaa 1920
actgtgaact ctgcccattg gtcaggagtg gctgtctctg gtctcttctt ttagctgaca 1980
aggacactcc tgaggctctt gttctcactg tttttttttt atcctggggc cacagtctct 2040
gattattcct cttgtgggta aagactgaat ttgtaaacc attcagataa atggcagtac 2100
tttaggacac acacaaacac acagatacac cttttgatat gtaagcttga cctaaagtca 2160
aaggacctgt gtagcatttc agattgagca cttcactatc aaaaatacta acatcacatg 2220
gcttgaagag taaccatcag agctgaatca tccaagtaag aacaagtacc attgttgatt 2280
gataagtaga gatacatttt ttatgatggt catcacagtg tggttaagggt gcaaattcaa 2340
aacatgtcac ccaagctctg ttcattgttt tgtgaattc 2379

```

<210> 13

<211> 552

<212> PRT

<213> *Homo sapiens*

<400> 13

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

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

Cys Ser Gly Pro Thr Cys Arg Thr Pro Gly Val Phe Asp Pro Gly Tyr
 515 520 525

Arg Phe Asp Pro Arg Leu Met Phe Ser Asn Arg Gly Ser Val Arg Thr
 530 535 540

Arg Arg Phe Ser Lys His Leu Leu
 545 550

<210> 14

<211> 2022

<212> DNA

<213> *Homo sapiens*

<400> 14

```

ccggtaccgg ctccctctgg gctccctcta gcgccttccc cccggcccga ctgctgggtc 60
agcgccaagt gacttacgcc cccgacctg agcccgacc gctaggcgag gaggatcaga 120
tctccgctcg agaattctgaa ggtgccctgg tcctggagga gttccgtccc agccctcggg 180
tctcccggta ctgctcgccc cggccctctg gagcttcagg aggcggccgt cagggtcggg 240
gagtattttg gtccggggtc tcagggaagg gcggcgccgt ggtctgcggt atcggaaga 300
gcctgctgga gccaaagtag cctccctctc ttgggacaga cccctcggtc ccattgcat 360
gggggcaccg cggtcctcc tctggccct ggctgctggc ctggccgttg cccgtccgcc 420
caacatcggt ctgatctttg ccgacgacct cggctatggg gacctgggct gctatgggca 480
ccccagctct accactccca acctggacca gctggcgccg ggagggctgc ggttcacaga 540
cttctacgtg cctgtgtctc tgtgcacacc ctctagggcc gccctcctga ccggccggct 600
cccgttcgg atgggcatgt acctggcgt cctggtgcc agctcccggg ggggcctgcc 660
cctggaggag gtgaccgtgg ccgaagtcct ggctgccga ggctacctca cagggaatggc 720
cggcaagtgg cacctgggg tggggcctga gggggccttc ctgcccccc atcagggett 780
ccatcgatct ctaggcaccc cgtactccca cgaccagggc cctgcccaga acctgacctg 840
cttcccgccg gccactcctt gcgacgggtg ctgtgaccag ggctgggtcc ccattcccact 900
gttggccaac ctgtccgtgg aggcgcagcc cccctggctg cccggactag agggccgcta 960
catggetttt gcccatgacc tcattggcca cggccagcgc caggatcgcc ccttcttct 1020
gtactatgcc tctaccaca cccactacc tcagttcagt gggcagagct ttgcagagcg 1080
ttcaggccgc gggccatttg gggactccct gatggagctg gatgcagctg tggggacct 1140
gatgacagcc ataggggacc tggggctgct tgaagagacg ctggtcatct tcatgcaga 1200
caatggacct gagaccatgc gtatgtcccg aggcggctgc tccggtctct tgcggtgtgg 1260
aaagggaaac acctacgagg gcggtgtccg agagcctgcc ttggccttct ggccaggta 1320
tactgctccc ggcgtgacct acgagctggc cagctccctg gacctgctgc ctaccctggc 1380
agccctggct ggggccccac tgcccaatgt caccttggat ggctttgacc tcagccccct 1440
gctgctgggc acaggcaaga gccctcgga gtctctcttc ttctaccgt cctaccaga 1500
cgaggctcgt ggggtttttg ctgtgcggac tggaaagtac aaggctcact tcttcacca 1560
gggctctgcc cacagtgata ccactgcaga cctgctgc cagcctcca gctctctgac 1620
tgctcatgag ccccgcctgc tctatgacct gtccaaggac cctgggtgaga actacaacct 1680
gctggggggg gtggccgggg ccaccccaga ggtgtgcaa gccctgaaac agcttcagct 1740

gctcaaggcc cagttagacg cagctgtgac ctccggcccc agccagggtg cccggggcga 1800
ggaccccgcc ctgcagatct gctgtcatcc tggtgcacc ccccgcccag cttgctgcca 1860
ttgccagat ccccatgcct gagggccct cggctggcct ggcatgtga tggctcctca 1920
ctgggagcct gtgggggagg ctcagggtgc tggaggggt ttgtgcctga taacgtaata 1980
acaccagtgg agacttgac atctgaaaaa aaaaaaaaaa aa 2022

```

<210> 15

<211> 507

<212> PRT

<213> *Homo sapiens*

<400> 15

```

Met Gly Ala Pro Arg Ser Leu Leu Leu Ala Leu Ala Ala Gly Leu Ala
1      5      10      15

Val Ala Arg Pro Pro Asn Ile Val Leu Ile Phe Ala Asp Asp Leu Gly
20      25      30

Tyr Gly Asp Leu Gly Cys Tyr Gly His Pro Ser Ser Thr Thr Pro Asn
35      40      45

Leu Asp Gln Leu Ala Ala Gly Gly Leu Arg Phe Thr Asp Phe Tyr Val
50      55      60

Pro Val Ser Leu Cys Thr Pro Ser Arg Ala Ala Leu Leu Thr Gly Arg
65      70      75      80

Leu Pro Val Arg Met Gly Met Tyr Pro Gly Val Leu Val Pro Ser Ser
85      90      95

Arg Gly Gly Leu Pro Leu Glu Glu Val Thr Val Ala Glu Val Leu Ala
100     105     110

Ala Arg Gly Tyr Leu Thr Gly Met Ala Gly Lys Trp His Leu Gly Val
115     120     125

Gly Pro Glu Gly Ala Phe Leu Pro Pro His Gln Gly Phe His Arg Phe
130     135     140

Leu Gly Ile Pro Tyr Ser His Asp Gln Gly Pro Cys Gln Asn Leu Thr
145     150     155     160

Cys Phe Pro Pro Ala Thr Pro Cys Asp Gly Gly Cys Asp Gln Gly Leu
165     170     175

Val Pro Ile Pro Leu Leu Ala Asn Leu Ser Val Glu Ala Gln Pro Pro
180     185     190

Trp Leu Pro Gly Leu Glu Ala Arg Tyr Met Ala Phe Ala His Asp Leu
195     200     205

Met Ala Asp Ala Gln Arg Gln Asp Arg Pro Phe Phe Leu Tyr Tyr Ala
210     215     220

Ser His His Thr His Tyr Pro Gln Phe Ser Gly Gln Ser Phe Ala Glu
225     230     235     240

```

Arg Ser Gly Arg Gly Pro Phe Gly Asp Ser Leu Met Glu Leu Asp Ala
 245 250 255
 Ala Val Gly Thr Leu Met Thr Ala Ile Gly Asp Leu Gly Leu Leu Glu
 260 265 270
 Glu Thr Leu Val Ile Phe Thr Ala Asp Asn Gly Pro Glu Thr Met Arg
 275 280 285
 Met Ser Arg Gly Gly Cys Ser Gly Leu Leu Arg Cys Gly Lys Gly Thr
 290 295 300
 Thr Tyr Glu Gly Gly Val Arg Glu Pro Ala Leu Ala Phe Trp Pro Gly
 305 310 315 320
 His Ile Ala Pro Gly Val Thr His Glu Leu Ala Ser Ser Leu Asp Leu
 325 330 335
 Leu Pro Thr Leu Ala Ala Leu Ala Gly Ala Pro Leu Pro Asn Val Thr
 340 345 350
 Leu Asp Gly Phe Asp Leu Ser Pro Leu Leu Leu Gly Thr Gly Lys Ser
 355 360 365
 Pro Arg Gln Ser Leu Phe Phe Tyr Pro Ser Tyr Pro Asp Glu Val Arg
 370 375 380
 Gly Val Phe Ala Val Arg Thr Gly Lys Tyr Lys Ala His Phe Phe Thr
 385 390 395 400
 Gln Gly Ser Ala His Ser Asp Thr Thr Ala Asp Pro Ala Cys His Ala
 405 410 415
 Ser Ser Ser Leu Thr Ala His Glu Pro Pro Leu Leu Tyr Asp Leu Ser
 420 425 430
 Lys Asp Pro Gly Glu Asn Tyr Asn Leu Leu Gly Gly Val Ala Gly Ala
 435 440 445
 Thr Pro Glu Val Leu Gln Ala Leu Lys Gln Leu Gln Leu Leu Lys Ala
 450 455 460
 Gln Leu Asp Ala Ala Val Thr Phe Gly Pro Ser Gln Val Ala Arg Gly
 465 470 475 480
 Glu Asp Pro Ala Leu Gln Ile Cys Cys His Pro Gly Cys Thr Pro Arg
 485 490 495
 Pro Ala Cys Cys His Cys Pro Asp Pro His Ala
 500 505

<210> 16

<211> 2228

<212> DNA

<213> *Homo sapiens*

<400> 16

```
acaaggatgg gtccgcgcgg cgcggcgagc ttgccccgag gccccggacc tcggcggctg    60
ctctccccg tcgtctccc gctgctgctg ctgctgttgt tggcgcgcc gggtcgggc    120
```

```

--
gccggggcca gccggccgcc ccacctgggc ttcttgctgg cagacgacct aggctggaac 180
gacgtcggct tccacggctc ccgcatccgc acgcccgcacc tggacgcgct ggcggccggc 240
gggggtgctcc tggacaacta ctacacgcag ccgctgtgca cgcgctcgcg gagccagctg 300
ctcactggcc gctaccagat ccgtacaggt ttacagcacc aaataatctg gccctgtcag 360
cccagctgtg ttctcttgga tgaaaaactc ctgccccagc tcctaaaaga agcagggttat 420
actaccata tggtcggaaa atggcacctg ggaatgtacc ggaagaatg ccttccaacc 480
cgccgaggat ttgataccta ctttggatat ctctgggta gtgaagatta ttattcccat 540
gaacgcgtga cattaattga cgctctgaat gtcacacgat gtgctcttga ttttcgagat 600
ggcgaagaag ttgcaacagg atataaaaaat atgtattcaa caacatatt caccaaaagg 660
gctatagccc tcataactaa ccattccacca gagaagcctc tgtttctcta ccttgccttc 720
cagtctgtgc atgagccccc tcaggctccct gaggaatact tgaagccata tgactttatc 780
caagacaaga acaggcatca ctatgcagga atgggtgtccc ttatggatga agcagtagga 840
aatgtcactg cagctttaa aagcagtgagg ctctggaaca acacggtgtt catcttttct 900
acagataacg gagggcagac ttgggcaggg ggttaataact ggccccctcg aggaagaaaa 960
tggagcctgt ggaaggagg cgctcgaggg gtgggctttg tggcaagccc ctgtctgaag 1020
cagaaggcgc tgaagaaccg ggagctcatc cacatctctg actggctgcc aacactcgtg 1080
aagctggcca ggggacacac caatggcaca aagcctctgg atggcctcga cgtgtggaaa 1140
accatcagtg aaggaagccc atccccaga attgagctgc tgcataatat tgacccaaac 1200
ttcgtggact cttcacctg tcccaggaac agcatggctc cagcaaagga tgactcttct 1260
cttcacgaat attcagcctt taacacatct gtccatgctg caattagaca tggaaattgg 1320
aaactcctca cgggctaccc aggctgtggt tactggttcc ctccaccgtc tcaatacaat 1380
gtttctgaga taccctcatc agaccacca accaagacce tctggctctt tgatattgat 1440
cgggaccctg aagaaagaca tgacctgtcc agagaatata ctacatcgt cacaaagctc 1500
ctgtcccgcc tacagttcta ccataaacac tcagtccccc tgtacttccc tgcacaggac 1560
ccccgctgtg atcccaaggc cactgggggtg tggggccctt ggatgtagga ttccagggag 1620
gctagaaaac ctttcaattg gaagttggac ctccaggcctt ttctcagcac tcttgtctca 1680
tttgttatcc caacctgggt tcaactggcc cttctcttgc tcttaaacca caccgaggtg 1740
tctaatttca acccctaatt catttaagaa gctgataaaa tctgcaacac tctgtctgtt 1800
ggctggagca tgtgtctaga ggtgggggtg gctgggttta tcccccttcc ctaagccttg 1860
ggacagctgg gaacttaact tgaaatagga agttctcact gaatcctgga ggctggaaca 1920
gctggctctt ttagactcac aagtcagacg ttccattccc ctctgccaat agccagtttt 1980
attggagtga atcacatttc ttacgcaaat gaaggagca gacagtgatt aatggttctg 2040
ttggccaagg cttctccctg tgggtgaagg atcatgttca ggcactccaa gtgaaccacc 2100
cctcttggtt cacccttacc tcacttatct catcacagag cataaggccc attttgttgt 2160
tcaggteaac agcaaatgg cctgcacat gactgtggct tttaaaataa agaaatgtgt 2220
ttttatcg 2228

```

<210> 17

<211> 533

<212> PRT

<213> *Homo sapiens*

<400> 17

```

Met Gly Pro Arg Gly Ala Ala Ser Leu Pro Arg Gly Pro Gly Pro Arg
 1          5          10          15

Arg Leu Leu Leu Pro Val Val Leu Pro Leu Leu Leu Leu Leu Leu
 20          25          30

Ala Pro Pro Gly Ser Gly Ala Gly Ala Ser Arg Pro Pro His Leu Val
 35          40          45

Phe Leu Leu Ala Asp Asp Leu Gly Trp Asn Asp Val Gly Phe His Gly
 50          55          60

Ser Arg Ile Arg Thr Pro His Leu Asp Ala Leu Ala Ala Gly Gly Val
 65          70          75          80

Leu Leu Asp Asn Tyr Tyr Thr Gln Pro Leu Cys Thr Pro Ser Arg Ser
 85          90          95

Gln Leu Leu Thr Gly Arg Tyr Gln Ile Arg Thr Gly Leu Gln His Gln
100          105          110

Ile Ile Trp Pro Cys Gln Pro Ser Cys Val Pro Leu Asp Glu Lys Leu
115          120          125

Leu Pro Gln Leu Leu Lys Glu Ala Gly Tyr Thr Thr His Met Val Gly
130          135          140

Lys Trp His Leu Gly Met Tyr Arg Lys Glu Cys Leu Pro Thr Arg Arg
145          150          155          160

Gly Phe Asp Thr Tyr Phe Gly Tyr Leu Leu Gly Ser Glu Asp Tyr Tyr
165          170          175

Ser His Glu Arg Cys Thr Leu Ile Asp Ala Leu Asn Val Thr Arg Cys
180          185          190

Ala Leu Asp Phe Arg Asp Gly Glu Glu Val Ala Thr Gly Tyr Lys Asn
195          200          205

Met Tyr Ser Thr Asn Ile Phe Thr Lys Arg Ala Ile Ala Leu Ile Thr
210          215          220

Asn His Pro Pro Glu Lys Pro Leu Phe Leu Tyr Leu Ala Leu Gln Ser
225          230          235          240

Val His Glu Pro Leu Gln Val Pro Glu Glu Tyr Leu Lys Pro Tyr Asp
245          250          255

Phe Ile Gln Asp Lys Asn Arg His His Tyr Ala Gly Met Val Ser Leu
260          265          270

Met Asp Glu Ala Val Gly Asn Val Thr Ala Ala Leu Lys Ser Ser Gly
275          280          285

Leu Trp Asn Asn Thr Val Phe Ile Phe Ser Thr Asp Asn Gly Gly Gln
290          295          300

```

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

<210> 18

<211> 2401

<212> DNA

<213> *Homo sapiens*

<400> 18

gcctccagca gctgacggga cccagctgta gtgaggttgc agtgattgag taggattggc	60
ctgcttcaaa gcagaggttt ctcatgggaa tatgcttatt aaactcccac tgggtgcagaa	120
accatgaaca gaggatgaac aagtgaagtt gcaatctcct ccatcacagc tcagttcccc	180
aacaacagga tcacaagctg gagatgcctt taaggaagat gaagatccct ttctctctac	240
tgttctttct gtgggaagcc gagagccacg cagcatcaag gccgaacatc atcctggtga	300
tggctgacga cctcggcatt ggagatcctg ggtgctatgg gaacaaaact atcaggactc	360

```

ccaatatcga ccggttgccc agtgggggag tgaaaactcac tcagcacctg gcagcatcac 420
cgctgtgcac accaagcagg gcagccttca tgaactggccg gtacctgtgc cgatcaggaa 480
tggcatcttg gtcccgccact ggagttttcc tcttcacagc ctcttcggga ggacttccca 540
ccgatgagat taccttttgc aagcttctga aggatcaagg ttattcaaca gcactgatag 600
ggaaatggca ccttgggatg agctgtcaca gcaagactga cttctgtcac caccctttac 660
atcacggctt caattatttc tatgggatct ctttgaccaa tctgagagac tgcaagcccg 720
gagagggcag tgtcttcacc acgggcttca agaggctggc cttcctcccc ctgcagatcg 780
tcggggctac cctccttacc cttgtgtcac tcaattgtct ggggctactc cagtgctctc 840
taggcgtttt tttcagcctt ctcttcctag cagccctaatt cctgaccctt ttcttgggct 900
tccttcatta cttccggccc ctgaactgct tcatgatgag gaactacgag atcattcagc 960
agccccatgc ctatgacaat ctccccaga ggctaaccggg ggaggcgccc cagttcatac 1020
agcggaacac tgagactccg ttctgtcttg tcttgtctca cctccacgtg cacacagccc 1080
tgttctccag caaagacttt gctggcaaaa gtcaacacgg agtctacggg gatgctgttg 1140
aggaaatgga ctggagtgtg gggcagatct tgaaccttct ggatgagctg agattggcta 1200
atgataccct catctacttc acatcggacc agggagcaca tgtagaggag gtgtcttcca 1260
aaggagaaat tcatggcgga agtaatggga tctataaagg aggaaaagca aacaactggg 1320
aaggaggtat ccgggttcca ggcaccttc gttggcccag ggtgatacag gctggccaga 1380
agattgatga gccactatgc aacatggaca tatttcctac agtagccaag ctggctggag 1440
ctcccttgcc tgaggacagg atcattgatg gacgtgatct gatgcccctg cttgaaggaa 1500
aaagccaacg ctccgatcat gaggtttctt tccattactg caacgcctac ttaaatgctg 1560
tgcgctggca cctcagaac agcacatcca tctggaaggg ctttttcttc accccaact 1620
tcaaccccg tgggttccaac ggatgctttg ccacacacgt gtgcttctgt ttcgggagtt 1680
atgtcaccca tcacgacca cctttactct ttgatatttc caaagatccc agagagagaa 1740
acccactaac tccagcatcc gagccccggg ttatgaaat cctcaaagtc atgcaggaa 1800
ctgaggacag acacacccag accctgccag aggtgcccga tcagttttca tggaaacact 1860
ttctttggaa gccctggctt cagctgtgct gtccctccac cggcctgtct tgccagtgtg 1920
atagagaaaa acaggataag agactgagcc gctagcagcg cctggggacc agacagacgc 1980
atgtggcaaa gctcaccatc ttactacaa acacgcctga gagtggcact ggggaaacat 2040
aactccatct acacettgga ttggactga ttctccattt tatcacctga aggccttggc 2100
cagagctcaa cagctactca actggagggg tgagggggat aaggtctgta gtatacagac 2160
aggaagatgg taggtttatg cttctgttgg ccagagtctt ggactcatgg aaatagaatg 2220
aatagagggg cattcacaag gcacaccagt gcaagcagat gacaaaaagg tgcagaaggc 2280
aatcttaaaa cagaaagggt caggaggtag cttaactcac cctcagcaa atacctatgt 2340
caacagtata agttaccatt tactctataa tctgcagtga tgcaataacc agcataataa 2400
a 2401

```

<210> 19

<211> 583

<212> PRT

<213> *Homo sapiens*

<400> 19

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

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

<210> 20

<211> 1945

<212> DNA

<213> *Homo sapiens*

<400> 20

ggaagccttg gcactagcgg cggccgggcg cggagtgcg agggcaagg cctgcgctct 60

gggccagcgc tcggccatgc gatccgccgc gcggagggga cgcgccgcgc ccgccgccag 120
ggactctttg cgggtgctac tgtttttatg cttgcttctg aagacgtgtg aacctaaaac 180
tgcaaagtgc tttaaaccaa ataccctact gatcatggcg gatgatctag gcactgggga 240
tctcgggtgc tacgggaaca atacactgag aacgccgaat attgaccagc ttgcagagga 300
aggtgtgagg ctcaactcagc acctggcggc cggcccgctc tgcaccccaa gccgagctgc 360
attcctcaca gggagacatt ccttcagatc aggcattggac gccagcaatg gataccgggc 420
ccttcagtgg aacgcaggct caggtaggact ccttgagaac gaaaccactt ttgcaagaat 480
cttgccagcag catggctatg caaccggcct cataggaaaa tggcaccagg gtgtgaattg 540
tgcacccgcg ggggataact gccaccaccc cctgaaccac ggatttgact atttctacgg 600
catgcccttc acgctcacia acgactgtga cccaggcagg ccccccgaag tggacgccgc 660
cctgagggcg cagctctggg gttacaacca gttcctggcg ctggggatc tcacctggc 720
tgccggccag acctgcggtt tcttctctgt ctcccgaga gcagtcaccg gcattggcgg 780
cgtgggctgc ctgtttttca tctcttggtt ctctctcttc gggtttgtgc gacgtggaa 840
ctgtatcctg atgagaaacc atgacgtcac ggagcaacc atggttctgg agaaaacagc 900
gagtccttatg ctaaaggaag ctgtttccta tattgaaaga cacaagcatg ggccatttct 960
cctcttccct tctttgctgc atgtgcacat tccccctgtg accacgagtg cattctggg 1020
gaaaagtcag catggcttat atggtgataa tgtggaggag atggactggc tcataggtaa 1080
ggttcttaat gccatcgaag acaatgggtt aaagaactca acattcacgt atttcacctc 1140
tgaccatgga ggacatttag aggcaagaga tggacacagc cagttagggg gatggaacgg 1200
aatttacaaa ggtgggaagg gcatgggagg atgggaagg gggatccgag tgcccgggat 1260
cttccactgg cggggggtgc tcccggccgg ccgagtgatt ggagagccca cgagcctgat 1320
ggacgtgttc cctactgtgg tccagctggt gggtaggcgag gtgccccagg acagggtgat 1380
tgatggccac agcctggtag ccttgctgca gggagctgag gcacgctcg cactgagtt 1440
cctgtttcat tactgtgggc agcatcttca cgcagcacgc tggcaccaga aggacagtgg 1500
aagcgtctgg aaggttcatt acacgacccc gcagttccac cccgaggagc ggggcctgct 1560
aacggccgag gcgtctgccc atgtgaaatg gggaggcgtg acccatcaca gaccccttt 1620
gctctttgac ctctccaggg acccctccga ggcacggccc ctgaccccg actccgagcc 1680
cctgtaccac gccgtgatag caagggtagg tgccgcggtg tcggagcacc ggcagaccct 1740
gagtcctgtg cccagcagc tttccatgag caacatcctg tggaaagccg ggcctgcagc 1800
gtgctgcgga catttccgt tctgttcatg ccacgaggat ggggatggca cccctgaat 1860
gccaggactg tgagagagga tccaggagag cctgactgcg ttgcaacaa aattctccaa 1920
gcttggttct atcttcagtc cggaa 1945

<210> 21

<211> 593

<212> PRT

<213> *Homo sapiens*

<400> 21

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

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

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

<210> 22

<211> 1858

<212> DNA

<213> *Homo sapiens*

<400> 22

ccttcctctt cttgatcggg gattcaggaa ggagcccagg agcagaggaa gtagagagag 60
 agacaacatg ttacatctgc accattcttg tttgtgttgc aggagctggc tgccagcgat 120

gctcgtgta ctgctaagtt tggcaccatc agcttcacgc gacatttccg cctcccagacc 180
 gaacatcctt cttctgatgg cggacgacct tggcattggg gacattggct gctatggcaa 240
 caacaccatg aggactccga atattgaccg ccttgccagag gacggcgtga agctgaccca 300
 acacatctct gccgcctctt tgtgcacccc aagcagagcc gccttctcca cgggcagata 360
 ccttggtgca tcagggatgg tttccagcat tggttaccgt gttcttcagt ggaccggagc 420
 atctggaggt cttccaacaa atgagacaac ttttgcaaaa atactgaaag agaaaggcta 480
 tgccactgga ctcatggaa aatggcatct gggctctaac tgtgagtcag ccagtgatca 540
 ttgccaccac cctctccatc atggccttga gcatttctac ggaatgcctt tctccttgat 600
 gggtgattgc gcccgctggg aactctcaga gaagcgtgtc aacctggaac aaaaactcaa 660
 cttctcttgc caagtccctg ccttggttgc cctcacactg gtagcaggga agctcacaca 720
 cctgataccc gtctcgtgga tgccgggtcat ctggctcagc ctttcggcgg tctctctct 780
 cgcaagctcc tattttgtgg gtgctctgat tgtccatgcc gattgcttgc tgatgagaaa 840
 ccacaccatc acggagcagc ccatgtgctt ccaaagaacg acacccctta ttctgcagga 900
 ggttgcgctc tttctcaaaa ggaataagca tgggccttgc ctctctcttg tttcctttct 960
 acacgttcac atccctctta tcatctatga gaacttcctc gggaagagtc tccacgggct 1020
 gtatggggac aacgtagagg agatggactg gatggtagga cggatccttg acactttgga 1080
 cgtgaggggt ttgagcaaca gcacccctcat ttatcttacg tcggatcacg gcggttccct 1140
 agagaatcaa cttggaaaca cccagtatgg tggctggaat ggaatttata aagtggggaa 1200
 gggcatggga ggatgggaag gtgggatccg cgtgcccggg atcttccgct ggcccggggt 1260
 gctcccggcc ggccgagtga ttggcgagcc cagcagctct atggacgtgt tccccaccgt 1320
 ggtccggctg gcggggcgcg aggtgcccga ggacagagtg attgacggcc aagaccttct 1380
 gcccttgctc ctggggacag cccaacactc agaccacgag ttcttgatgc attattgtga 1440
 gaggtttctg cacgcagcca ggtggcatca acgggacaga ggaacaatgt ggaaagtcca 1500
 ctttgtgacg cctgtgttcc agccagaggg agccgggtgc tgcctatggaa gaaaggctctg 1560
 cccgtgcttt ggggaaaaag tagtccacca cgatccacct ttgctctttg acctctcaag 1620
 agaccttctt gagaaccaca tcttcacacc agcctcagag cccgtgttct atcagggtgat 1680
 ggaacgagtc cagcaggcgg tgtgggaaca ccagcggaca ctacgcccag ttctcttgca 1740
 gctggacagg ctgggcaaca tctggagacc gtggctgcag cctgctgtg gcccgttccc 1800
 cctctgctgg tgcccttaggg aagatgacct acaataaatg tctgcagtga aaagctgg 1858

<210> 23

<211> 589

<212> PRT

<213> *Homo sapiens*

<400> 23

Met	Leu	His	Leu	His	His	Ser	Cys	Leu	Cys	Phe	Arg	Ser	Trp	Leu	Pro
1				5					10					15	
Ala	Met	Leu	Ala	Val	Leu	Leu	Ser	Leu	Ala	Pro	Ser	Ala	Ser	Ser	Asp
			20					25					30		
Ile	Ser	Ala	Ser	Arg	Pro	Asn	Ile	Leu	Leu	Leu	Met	Ala	Asp	Asp	Leu
		35					40					45			

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

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

<210> 24

<211> 1996

<212> DNA

<213> *Homo sapiens*

<400> 24

```

gggttctgct cctagacatt agagagataa tacggctgat agacaacaag aaggtattcc      60
aagctgcaca atgaggccca ggagaccgtt ggtcttcatg tctttggtgt gtgcactctt      120
gaacacatgg ccagggcaca cagggtgcat gacgacaagg cctaattattg tcctaatacat      180
ggttgatgac ctgggtattg gagatctggg ctgctacggc aatgacacca tgaggacgcc      240
tcacatcgac cgccttgcca gggaaggcgt gcgactgact cagcacatct ctgccgcctc      300
cctctgcagc ccaagccggt ccgcgttctt gacgggaaga taccatcc gatcagggtat      360
ggtttctagt ggtaatagac gtgtcatcca aaatcttgca gtccccgcag gcctccctct      420

taatgagaca acacttgtag ccttgctaaa gaagcaagga tacagcacgg ggcttatagg      480
caaatggcac caaggcttga actgcgactc ccgaagtac cagtgccacc atccatataa      540
ttatgggttt gactactact atggcatgcc gttcactctc gttgacagct gctggccgga      600
ccccctctgt aacacggaat tagcctttga gagtcagctc tggctctgtg tgcagctagt      660
tgccattgcc atcctcacc taacctttgg gaagctgagc ggctgggtct ctgttccctg      720
gctcctgac ttctccatga ttctgtttat ttctctcttg ggctatgctt ggttctccag      780
ccacacgtcc cctttatact gggactgcct cctcatgcgg gggcacgaga tcacggagca      840
gcccataaag gctgaacgag ctggatccat tatggtgaag gaagcgattt cctttttaga      900
aaggcacagt aaggaaactt tccttctctt ttctctctt ctccacgtgc acacacctct      960
ccccaccacg gacgatttca ctggcaccag caagcatggc ttgtatgggg ataattgtgga      1020
agagatggac tccatggtgg gcaagattct tgatgctatc gatgattttg gcctaaggaa      1080
caacaccctt gtctacttta catcagatca cggagggcac ttggaagcta ggcgagggca      1140
tgcccaactt ggtggatgga atggaatata caaagggtgga aaaggcatgg ggggctggga      1200
aggtggaatc cgcgtccag gaattgtccg atggcctgga aaggtagcag ctggacgggt      1260
gattaaggaa cctacaagtt taatggatat ttaccact gtcgcatcag tgcaggagg      1320
aagtctccct caggacaggg tcattgacgg ccgagacctc atgcccttgc tgcagggcaa      1380
cgtcaggcac tcggagcatg aatttctttt ccactactgt ggctcctacc tgcacgccgt      1440
gcggtggatc cccaaggacg acagtgggtc agtttggag gctcactatg tgaccccggt      1500
attccagcca ccagcttctg gtggctgcta tgtcacctca ttatgcagat gtttcggaga      1560
acaggttacc taccacaacc cccctctgct ctctgatctc tccagggacc cctcagagtc      1620
cacaccctg acacctgcca cagagccctc ctatgatttt gtgattaaa aggtggccaa      1680
cgccctgaag gaacacacag aaaccatcgt gcctgtgacc taccactctc cagaactgaa      1740
tcagggcagg acgtggctga agccttgcgt tggggtgttc ccattttgtc tgtgtgacaa      1800
ggaagaggaa gtctctcagc ctccgggtcc taacgagaag agataattac aatcaggcta      1860
ccagaggaa cctttgtgct taacgagaag agataattac aatcaggcta ccaaaggaa      1920
cactaacttt ggtgctttca agttggcaag gagtgcatct aatagtcaat aaattcatct      1980
accattccag attatt                                     1996

```

<210> 25

<211> 591

<212> PRT

<213> *Homo sapiens*

<400> 25

Met	Arg	Pro	Arg	Arg	Pro	Leu	Val	Phe	Met	Ser	Leu	Val	Cys	Ala	Leu
1				5					10					15	
Leu	Asn	Thr	Trp	Pro	Gly	His	Thr	Gly	Cys	Met	Thr	Thr	Arg	Pro	Asn
			20					25					30		
Ile	Val	Leu	Ile	Met	Val	Asp	Asp	Leu	Gly	Ile	Gly	Asp	Leu	Gly	Cys
		35					40					45			
Tyr	Gly	Asn	Asp	Thr	Met	Arg	Thr	Pro	His	Ile	Asp	Arg	Leu	Ala	Arg
	50					55					60				

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

```

Ile Val Arg Trp Pro Gly Lys Val Pro Ala Gly Arg Leu Ile Lys Glu
385          390          395          400

Pro Thr Ser Leu Met Asp Ile Leu Pro Thr Val Ala Ser Val Ser Gly
          405          410          415

Gly Ser Leu Pro Gln Asp Arg Val Ile Asp Gly Arg Asp Leu Met Pro
          420          425          430

Leu Leu Gln Gly Asn Val Arg His Ser Glu His Glu Phe Leu Phe His
          435          440          445

Tyr Cys Gly Ser Tyr Leu His Ala Val Arg Trp Ile Pro Lys Asp Asp
          450          455          460

Ser Gly Ser Val Trp Lys Ala His Tyr Val Thr Pro Val Phe Gln Pro
          465          470          475          480

Pro Ala Ser Gly Gly Cys Tyr Val Thr Ser Leu Cys Arg Cys Phe Gly
          485          490          495

Glu Gln Val Thr Tyr His Asn Pro Pro Leu Leu Phe Asp Leu Ser Arg
          500          505          510

Asp Pro Ser Glu Ser Thr Pro Leu Thr Pro Ala Thr Glu Pro Leu Tyr
          515          520          525

Asp Phe Val Ile Lys Lys Val Ala Asn Ala Leu Lys Glu His Gln Glu
          530          535          540

Thr Ile Val Pro Val Thr Tyr Gln Leu Ser Glu Leu Asn Gln Gly Arg
          545          550          555          560

Thr Trp Leu Lys Pro Cys Cys Gly Val Phe Pro Phe Cys Leu Cys Asp
          565          570          575

Lys Glu Glu Glu Val Ser Gln Pro Arg Gly Pro Asn Glu Lys Arg
          580          585          590

```

<210> 26

<211> 1578

<212> DNA

<213> *Homo sapiens*

<400> 26

atgggctggc	tttttctaaa	ggttttgttg	gcgggagtga	gtttctcagg	atttctttat	60
cctcttgtgg	atttttgcat	cagtgggaaa	acaagaggac	agaagccaaa	ctttgtgatt	120
attttggccg	atgacatggg	gtgggggtgac	ctgggagcaa	actgggcaga	aacaaaggac	180
actgccaacc	ttgataagat	ggcttcggag	ggaatgaggt	ttgtggattt	ccatgcagct	240
gcctccacct	gtcaccctc	ccgggcttcc	ttgtctaccg	gccggcttgg	ccttcgcaat	300
ggagtccacac	gcaactttgc	agtcacttct	gtgggaggcc	ttccgctcaa	cgagaccacc	360
ttggcagagg	tgctgcagca	ggcgggttac	gtcactggga	taataggcaa	atggcatctt	420
ggacaccacg	gtcttbatca	ccccaaattc	cgtgggtttg	attactactt	tggaaatcca	480
tatagccatg	atatgggctg	tactgatact	ccaggctaca	accaccctcc	ttgtccagcg	540
tgctccacagg	gtgatggacc	atcaaggaac	cttcaaagag	actgttacac	tgacgtggcc	600
ctccctcttt	atgaaaacct	caacattgtg	gagcagccgg	tgaacttgag	cagccttgcc	660
cagaagtatg	ctgagaaagc	aaccagttc	atccagcgtg	caagcaccag	cgggaggccc	720
tctctgtctt	atgtggctct	ggcccacatg	cacgtgccct	tacctgtgac	tcagctacca	780
gcagcgccac	ggggcagaag	cctgtatggt	gcagggtctt	gggagatgga	cagtctggtg	840
ggccagatca	aggacaaagt	tgaccacaca	gtgaaggaaa	acacattcct	ctggtttaca	900
ggagacaatg	gcccgtgggc	tcagaagtgt	gagctagcgg	gcagtgtggg	tcccttact	960
ggatttttggc	aaactcgtca	agggggaagt	ccagccaagc	agacgacctg	ggaaggaggg	1020
caccgggtcc	cagcactggc	ttactggcct	ggcagagttc	cagttaatgt	caccagcact	1080
gccttgttaa	gcgtgctgga	cattttttcca	actgtggtag	cctgggcca	ggccagctta	1140
cctcaaggac	ggcgctttga	tggtgtggac	gtctccgagg	tgctcttttg	ccggtcacag	1200
cctgggcaca	gggtgtgtt	ccaccccaac	agcggggcag	ctggagagtt	tggagccctg	1260
cagactgtcc	gcctggagcg	ttacaaggcc	ttctacatta	ccggtggagc	cagggcggtg	1320
gatgggagca	cggggcctga	gctgcagcat	aagtttcctc	tgattttcaa	cctggaagac	1380
gataccgcag	aagctgtgcc	cctagaaaga	ggtggtgcgg	agtaccaggc	tgtgtgtccc	1440
gaggtcagaa	aggttcttgc	agacgtcctc	caagacattg	ccaacgacaa	catctccagc	1500
gcagattaca	ctcaggaccc	ttcagtaact	cctgtctgta	atccctacca	aattgcctgc	1560
cgctgtcaag	ccgcataa					1578

<210> 27

<211> 525

<212> PRT

<213> *Homo sapiens*

<400> 27

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

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

```

      Gln His Lys Phe Pro Leu Ile Phe Asn Leu Glu Asp Asp Thr Ala Glu
      450                               455                               460
Ala Val Pro Leu Glu Arg Gly Gly Ala Glu Tyr Gln Ala Val Leu Pro
465                               470                               475                               480

Glu Val Arg Lys Val Leu Ala Asp Val Leu Gln Asp Ile Ala Asn Asp
      485                               490                               495

Asn Ile Ser Ser Ala Asp Tyr Thr Gln Asp Pro Ser Val Thr Pro Cys
      500                               505                               510

Cys Asn Pro Tyr Gln Ile Ala Cys Arg Cys Gln Ala Ala
      515                               520                               525

```

<210> 28

<211> 4669

<212> DNA

<213> *Homo sapiens*

<400> 28

cgagaccgt cgctaataa tcttggggcc ggtgtcgggc cggggcggct tgateggcaa	60
ctaggaaacc ccaggcgag agggcaggag cgagggcagc gaggatcaga ggccaggcct	120
tcccggctgc cggcgctcct cggaggtcag ggcagatgag gaacatgact ctcccccttc	180
ggaggaggaa ggaagtcccg ctgccacctt atctctgctc ctctgcctcc tccctgttcc	240
cagagctttt tctctagaga agattttgaa ggcggctttt gtgctgacgg ccacccacca	300
tcattctaaag aagataaact tggcaaatga catgcagggt cttcaaggca gaataattgc	360
agaaaatctt caaaggaccc tatctgcaga tgttctgaat acctctgaga atagagattg	420
attattcaac caggatacct aattcaagaa ctccagaaat caggagacgg agacattttg	480
tcagttttgc aacattggac caaatacaat gaagtattct tgctgtgctc tggttttggc	540
tgtcctgggc acagaattgc tgggaagcct ctgttcgact gtcagatccc cgagggtcag	600
aggacggata cagcaggaaac gaaaaaacat ccgacccaac attattcttg tgcttaccga	660
tgatcaagat gtggagctgg ggtccctgca agtcatgaac aaaacgagaa agattatgga	720
acatgggggg gccaccttca tcaatgcctt tgtgactaca cccatgtgct gcccgtaacg	780
gtcctccatg ctccacggga agtatgtgca caatcacaat gtctacacca acaacgagaa	840
ctgctcttcc cctcgtggc agggcatgca tgagcctcgg acttttgctg tatactctaa	900
caacactggc tacagaacag ccttttttgg aaaatacctc aatgaatata atggcagcta	960
catccccctt ggggtggcag aatggcttgg attaatcaag aattctcgct tctataatta	1020
cactgtttgt cgcaatggca tcaaagaaaa gcatggattt gattatgcaa aggactactt	1080
cacagactta atcactaacg agagcattaa ttacttcaa atgtctaaga gaatgtatcc	1140
ccataggccc gttatgatgg tgatcagcca cgctgcgccc cagggccccg aggactcagc	1200
cccacagttt tctaaactgt accccaatgc ttcccaacac ataactccta gttataacta	1260
tgcaccaa atggataaac actggattat gcagtacaca ggaccaatgc tgcccatcca	1320
catggaattt acaaacattc tacagcgcaa aaggctccag actttgatgt cagtggatga	1380
ttctgtggag aggtgtgata acatgctcgt ggagacgggg gagctggaga atacttacat	1440
catttacacc gccgaccatg gttaccatat tgggcagttt ggactgggtca aggggaaatc	1500
catgccatat gactttgata ttcgtgtgcc tttttttatt cgtgggtccaa gtgtagaacc	1560

aggatcaata gtcccacaga tcgttctcaa cattgacttg gccccacga tcctggatat 1620
 tgctgggctc gacacacctc ctgatgtgga cggcaagtct gtctcaaac ttctggaccc 1680
 agaaaagcca ggtaacaggt ttccaacaaa caagaaggcc aaaatttggc gtgatacatt 1740
 cctagtggaa agaggcaaat ttctacgtaa gaaggaagaa tccagcaaga atatccaaca 1800
 gtcaaatcac ttgcccaaat atgaacgggt caaagaacta tgccagcagg ccaggtaacca 1860
 gacagcctgt gaacaaccgg ggcagaagtg gcaatgcatt gaggatacat ctggcaagct 1920
 tcgaattcac aagtgtaaag gaccagtgga cctgctcaca gtccggcaga gcacgcggaa 1980
 cctctacgct cgcggcttcc atgacaaaga caaagagtgc agttgtaggg agtctggtta 2040
 ccgtgccagc agaagccaaa gaaagagtca acggcaattc ttgagaaacc aggggactcc 2100
 aaagtacaag ccagatttg tccatactcg gcagacagct tccttgctcg tcgaatttga 2160
 aggtgaaata tatgacataa atctggaaga agaagaagaa ttgcaagtgt tgcaaccaag 2220
 aaacattgct aagcgtcatg atgaaggcca caaggggcca agagatctcc aggcctccag 2280
 tgggtggcaac agggggcagga tgctggcaga tagcagcaac gccgtgggcc cacctaccac 2340
 tgctcgagtg acacacaagt gttttattct tcccaatgac tctatccatt gtgagagaga 2400
 actgtaccaa tcggccagag cgtggaagga ccataaggca tacattgaca aagagattga 2460
 agctctgcaa gataaaatta agaatttaag agaagtgaga ggacatctga agagaaggaa 2520
 gcctgaggaa tgtagctgca gtaaacaaaag ctattacaat aaagagaaag gtgtaaaaaa 2580
 gcaagagaaa ttaaagagcc atcttcaccc attcaaggag gctgctcagg aagtagatag 2640
 caaactgcaa cttttcaagg agaacaaccg taggaggaag aaggagagga aggagaagag 2700
 acggcagagg aagggggaag agtgacagct gcctggcctc acttgcttca cgcattgaaa 2760
 caaccactgg cagacagccc cgttctggaa cctgggatct ttctgtgctt gcacgagttc 2820
 taacaataac acctactggt gtttgcgtac agttaatgag acgcataatt ttcttttctg 2880
 tgagtttgct actggctttt tggagtattt tgatatgaat acagatcctt atcagctcac 2940
 aaatacagtg cacacggtag aacgaggcat ttgaaatcag ctacacgtac aactaatgga 3000
 gctcagaagc tgtcaaggat ataagcagtg caacccaaga cctaagaatc ttgatgttgg 3060
 aaataaagat ggaggaagct atgacctaca cagaggacag ttatgggatg gatgggaagg 3120
 ttaatcagcc ccgtctcact gcagacatca actggcaagg cctagaggag ctacacagtg 3180
 tgaatgaaaa catctatgag tacagacaaa actacagact tagtctgggtg gactggacta 3240
 attacttgaa ggatttagat agagtatttg cactgctgaa gagtcactat gagcaaaata 3300
 aaacaataaa gactcaaaact gctcaaagtg acgggttctt ggttgctctt gctgagcacg 3360
 ctgtgtcaat ggagatggcc tctgctgact cagatgaaga cccaaggcat aaggttggga 3420
 aaacacctca tttagacctg ccagctgacc ttcaaaccct gcatttgaac cgaccaacat 3480
 taagtcacaga gagtaaaact gaatggaata acgacattcc agaagttaat catttgaatt 3540
 ctgaacactg gagaaaaacc gaaaaatgga cggggcatga agagactaat catctggaaa 3600
 ccgatttcag tggcgatggc atgacagagc tagagctcgg gccagcccc aggctgcagc 3660
 ccattcacag gcacccgaaa gaacttcccc agtatgggtg tcctggaaaag gacatttttg 3720
 aagatcaact atatcttctc gtgcattccg atggaatttc agttcatcag atgtteacca 3780
 tggccaccgc agaaccaccg agtaattcca gcatagcggg gaagatgttg accaaggtgg 3840
 agaagaatca cgaaggagg aagtcacagc acctagaagg cagcacctcc tcttctctct 3900

```

cctctgatta gatgaaactg ttaccttacc ctaaacacag tatttctttt taactttttt 3960
atttgtaaac taataaagggt aatcacagcc accaacattc caagctaccc tgggtacett 4020
tgtgcagtag aagctagtga gcatgtgagc aagcgggtgt cacacggaga etcatcgta 4080
taatttacta tctgccaaga gtagaaagaa aggctgggga tatttgggtt ggcttggttt 4140
tgattttttg ctgttttgtt tgttttgtac taaaacagta ttatcttttg aatatcgtag 4200
ggacataagt atatacatgt tatccaatca agatggctag aatggtgcct ttctgagtgt 4260
ctaaaacttg acacccttgg taaatctttc aacacacttc cactgcctgc gtaatgaagt 4320
tttgattcat ttttaaccac tgggaattttt caatgccgtc attttcagtt agatgatttt 4380
gcactttgag attaaaatgc catgtctatt tgattagtct tattttttta tttttacagg 4440
cttatcagtc tcactgttgg ctgtcattgt gacaaagtca aataaacccc caaggacgac 4500
acacagtatg gatcacatat tgtttgacat taagcttttg ccagaaaatg ttgcatgtgt 4560
tttacctcga ctgtctaaaa tcgattagca gaaaggcatg gctaataatg ttggtggtga 4620
aaataaataa ataagtaaat gaaaaaaaaa aaaaaaaaaa aaaaaaaaaa 4669

```

<210> 29

<211> 871

<212> PRT

<213> *Homo sapiens*

<400> 29

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

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

Phe His Asp Lys Asp Lys Glu Cys Ser Cys Arg Glu Ser Gly Tyr Arg
 500 505 510
 Ala Ser Arg Ser Gln Arg Lys Ser Gln Arg Gln Phe Leu Arg Asn Gln
 515 520 525
 Gly Thr Pro Lys Tyr Lys Pro Arg Phe Val His Thr Arg Gln Thr Arg
 530 535 540
 Ser Leu Ser Val Glu Phe Glu Gly Glu Ile Tyr Asp Ile Asn Leu Glu
 545 550 555 560
 Glu Glu Glu Glu Leu Gln Val Leu Gln Pro Arg Asn Ile Ala Lys Arg
 565 570 575
 His Asp Glu Gly His Lys Gly Pro Arg Asp Leu Gln Ala Ser Ser Gly
 580 585 590
 Gly Asn Arg Gly Arg Met Leu Ala Asp Ser Ser Asn Ala Val Gly Pro
 595 600 605
 Pro Thr Thr Val Arg Val Thr His Lys Cys Phe Ile Leu Pro Asn Asp
 610 615 620
 Ser Ile His Cys Glu Arg Glu Leu Tyr Gln Ser Ala Arg Ala Trp Lys
 625 630 635 640
 Asp His Lys Ala Tyr Ile Asp Lys Glu Ile Glu Ala Leu Gln Asp Lys
 645 650 655
 Ile Lys Asn Leu Arg Glu Val Arg Gly His Leu Lys Arg Arg Lys Pro
 660 665 670
 Glu Glu Cys Ser Cys Ser Lys Gln Ser Tyr Tyr Asn Lys Glu Lys Gly
 675 680 685
 Val Lys Lys Gln Glu Lys Leu Lys Ser His Leu His Pro Phe Lys Glu
 690 695 700
 Ala Ala Gln Glu Val Asp Ser Lys Leu Gln Leu Phe Lys Glu Asn Asn
 705 710 715 720
 Arg Arg Arg Lys Lys Glu Arg Lys Glu Lys Arg Arg Gln Arg Lys Gly
 725 730 735
 Glu Glu Cys Ser Leu Pro Gly Leu Thr Cys Phe Thr His Asp Asn Asn
 740 745 750
 His Trp Gln Thr Ala Pro Phe Trp Asn Leu Gly Ser Phe Cys Ala Cys
 755 760 765
 Thr Ser Ser Asn Asn Asn Thr Tyr Trp Cys Leu Arg Thr Val Asn Glu
 770 775 780
 Thr His Asn Phe Leu Phe Cys Glu Phe Ala Thr Gly Phe Leu Glu Tyr
 785 790 795 800

```

Phe Asp Met Asn Thr Asp Pro Tyr Gln Leu Thr Asn Thr Val His Thr
      805                      810                      815

Val Glu Arg Gly Ile Leu Asn Gln Leu His Val Gln Leu Met Glu Leu
      820                      825                      830

Arg Ser Cys Gln Gly Tyr Lys Gln Cys Asn Pro Arg Pro Lys Asn Leu
      835                      840                      845

Asp Val Gly Asn Lys Asp Gly Gly Ser Tyr Asp Leu His Arg Gly Gln
      850                      855                      860

Leu Trp Asp Gly Trp Glu Gly
      865                      870

```

<210> 30

<211> 4279

<212> DNA

<213> *Homo sapiens*

<400> 30

gggccatttc	tggacaacag	ctgctatatt	cacttgagcc	caagttaatt	tctcggggag	60
ttctcggggc	cgcacaggca	gctcgggttg	ccctcgcgatt	gagctgcggg	tcgcggccgg	120
cgccggcctc	tccaatggca	aatgtgtgtg	gctggaggcg	agcgcgaggc	tttcggcaaa	180
ggcagtcgag	tgtttgca	ccggggcgag	tcctgtgaaa	gcagataaaa	gaaaacattt	240
attaacgtgt	cattacgagg	ggagcgcccg	gccggggctg	tcgcactccc	cgcggaacat	300
ttggctccct	ccagctccta	gagaggagaa	gaagaaagcg	gaaaagaggc	agattcacgt	360
cgtttccagc	caagtggacc	tgatcgatgg	ccctcctgaa	tttatcacga	tatttgattt	420
attagegatg	ccccctgggt	tgtgtgttac	gcacacacac	gtgcacacaa	ggctctggct	480
cgttcctcct	cctcgtttcc	agctcctggg	cgaatccca	atctgtttca	actctccgcc	540
gagggcgagc	aggagcgaga	gtgtgtcgaa	tctgcgagtg	aagagggacg	agggaaaaga	600
aacaaagcca	cagacgcaac	ttgagactcc	cgcaccccaa	aagaagcacc	agatcagcaa	660
aaaaagaaga	tgggcccccc	gagcctcgtg	ctgtgcttgc	tgtccgcaac	tgtgttctcc	720
ctgctgggtg	gaagctcggc	cttcctgtcg	caccaccgcc	tgaaaggcag	gtttcagagg	780
gaccgcagga	acatccgcc	caacatcata	ctggtgctga	cggacgacca	ggatgtggag	840
ctgggttcca	tgcaggtgat	gaacaagacc	cggcgcatca	tggagcaggg	cggggcgcac	900
ttcatcaacg	ccttcgtgac	cacacccatg	tgtgcccct	cacgtcctc	catcctcacc	960
ggcaagtacg	tccacaacca	caacacctac	accaacaatg	agaactgctc	ctcgccctcc	1020
tggcaggcac	agcacgagag	ccgcaccttt	gccgtgtacc	tcaatagcac	tggtaccgg	1080
acagctttct	tcgggaagta	tcttaatgaa	tacaacggct	cctacgtgcc	acccggctgg	1140
aaggagtggg	tcggactcct	taaaaactcc	cgtttttata	actacacgct	gtgtcggaac	1200
ggggtgaaag	agaagcacgg	ctccgactac	tccaaggatt	acctcacaga	cctcatcacc	1260
aatgacagcg	tgagcttctt	ccgcacgtcc	aagaagatgt	acccgcacag	gccagtcctc	1320
atgggtcatca	gccatgcagc	ccccacggc	cctgaggatt	cagccccaca	atattcacgc	1380
ctcttcccaa	acgcatctca	gcacatcacg	ccgagctaca	actacgcgcc	caaccggac	1440
aaacactgga	tcatgcgcta	cacggggccc	atgaagccca	tccacatgga	attcaccaac	1500
atgctccagc	ggaagcgctt	gcagaccctc	atgtcgggtg	acgactccat	ggagacgatt	1560

tacaacatgc	tggttgagac	gggcgagctg	gacaacacgt	acatcgtata	caccgcccac	1620
cacggttacc	acatcggcca	gtttggcctg	gtgaaagga	aatccatgcc	atatgagttt	1680
gacatcaggg	tcccgttcta	cgtgaggggc	cccaacgtgg	aagccggctg	tctgaatccc	1740
cacatcgtcc	tcaacattga	cctggccccc	accatcctgg	acattgcagg	cctggacata	1800
cctgcggata	tggacgggaa	atccatcctc	aagctgctgg	acacggagcg	gccggtgaat	1860
cggtttcact	tgaaaaagaa	gatgagggtc	tggcgggact	ccttcttggg	ggagagaggc	1920
aagctgctac	acaagagaga	caatgacaag	gtggacgccc	aggaggagaa	ctttctgccc	1980
aagtaccagc	gtgtgaagga	cctgtgtcag	cgtgctgagt	accagacggc	gtgtgagcag	2040
ctgggacaga	agtggcagtg	tgtggaggac	gccacgggga	agctgaagct	gcataagtgc	2100
aagggcccca	tgcggctggg	cggcagcaga	gccctctcca	acctcgtgcc	caagtactac	2160
gggcagggca	gcgaggcctg	cacctgtgac	agcggggact	acaagctcag	cctggccgga	2220
cgcggaaaa	aactcttcaa	gaagaagtac	aaggccagct	atgtccgcag	tcgctccatc	2280
cgtcagtggt	ccatcgaggt	ggacggcagg	gtgtaccacg	taggcctggg	tgatgccgcc	2340
cagcccgaa	acctcaccaa	gcggcactgg	ccaggggccc	ctgaggacca	agatgacaag	2400
gatggtgggg	acttcagtggt	cactggaggc	cttcccgaact	actcagccgc	caacccatt	2460
aaagtgcac	atcggtgcta	catcctagag	aacgcacacg	tccagtgtga	cctggacctg	2520
tacaagtccc	tgcaggcctg	gaaagaccac	aagctgcaca	tcgaccacga	gattgaaacc	2580
ctgcagaaca	aaattaagaa	cctgagggaa	gtccgaggtc	acctgaagaa	aaagcggcca	2640
gaagaatgtg	actgtcaca	aatcagctac	cacaccacgc	acaaaggccg	cctcaagcac	2700
agaggtcca	gtctgcatcc	tttcaggaag	ggcctgcaag	agaaggacaa	ggtgtggctg	2760
ttgcgggagc	agaagcgcaa	gaagaaactc	cgaagctgc	tcaagcgctc	gcagaacaac	2820
gacacgtgca	gcatgccagg	cctcacgtgc	ttcacccacg	acaaccagca	ctggcagacg	2880
gcgcctttct	ggacactggg	gcctttctgt	gcctgcacca	gcgccacaaa	taacacgtac	2940
tgggtcatga	ggaccatcaa	tgagactcac	aatttctctc	tctgtgaatt	tgcaactggc	3000
ttcctagagt	actttgatct	caacacagac	ccctaccagc	tgatgaatgc	agtgaacaca	3060
ctggacaggg	atgtcctcaa	ccagctacac	gtacagctca	tggagctgag	gagctgcaag	3120
ggtttacaagc	agtgtaaccc	ccggactcga	aacatggacc	tgggacttaa	agatggagga	3180
agctatgagc	aatacaggca	gtttcagcgt	cgaagtggc	cagaaatgaa	gagaccttct	3240
tccaaatcac	tgggacaact	gtgggaaggc	tgggaagggt	aagaacaac	agaggtggac	3300
ctccaaaaac	atagaggcat	cacctgactg	cacaggcaat	gaaaaaccat	gtgggtgatt	3360
tccagcagac	ctgtgctatt	ggccaggagg	cctgagaaag	caagcacgca	ctctcagtca	3420
acatgacaga	ttctggagga	taaccagcag	gagcagagat	aacttcagga	agtccatttt	3480
tgcctctgct	tttgccttgg	attatacctc	accagctgca	caaaatgcat	tttttcgtat	3540
caaaaagtca	ccactaacc	tccccagaa	gtccacaaag	gaaaacggag	agagcgagcg	3600
agagagattt	ccttggaat	ttctcccaag	ggcgaagtc	attggaattt	ttaaatcata	3660
ggggaaaagc	agtcctgttc	taaatcctct	tattcttttg	gtttgtcaca	aagaaggaa	3720
taagaagcag	gacagaggca	acgtggagag	gctgaaaaca	gtgcagagac	gtttgacaat	3780
gagtcatgag	cacaaaagag	atgacattta	cctagcatat	aaacctgggt	tgcctctgaa	3840
gaaactgcct	tcattgtata	tatgtgacta	tttacatgta	atcaacatgg	gaacttttag	3900

gggaacctaa taagaaatcc caattttcag gagtgggtgt gtcaataaac gctctgtggc	3960
cagtgtaaaa gaaaaaaaaa aaaaattgtg gacatttctg ttctgtcca gataccattt	4020
ctcctagtat ttctttgtta tgtcccagaa ctgatgtttt tttttaagg tactgaaaag	4080
aatgaagtt gatgtatgtc ccaagttttg atgaaactgt atttgtaaaa aaaattttgt	4140
agtttaagta ttgtcataca gtgttcaaaa cccagccaa tgaccagcag ttggtatgaa	4200
gaacctttga cattttgtaa aaggccattt cttggggaaa aaaaaaaaaa aaaaaaaaaa	4260
aaaaaaaaa aaaaaaaaaa	4279

- <210> 31
- <211> 870
- <212> PRT
- <213> *Homo sapiens*

- <400> 31

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

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

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

Lys Trp Pro Glu Met Lys Arg Pro Ser Ser Lys Ser Leu Gly Gln Leu
 850 855 860

Trp Glu Gly Trp Glu Gly
 865 870

<210> 32

<211> 6

<212> PRT

<213> *Homo sapiens*

<220>

<221> VARIANTE

<222> (1)..(1)

<223> Leu OU Val

<220>

<221> característica_misturada

<222> (1)..(3)

<223> Xaa pode ser qualquer aminoácido de ocorrência natural

<220>

<221> VARIANTE

<222> (2)..(2)

<223> Cys OU Ser

<220>

<221> VARIANTE

<222> (3)..(3)

<223> Qualquer Aminoácido

<400> 32

Xaa Xaa Xaa Pro Ser Arg
 1 5

<210> 33

<211> 23

<212> PRT

<213> Artificial

<220>

<223> Sequência derivada da Arilsulfatase A humana

<220>

<221> PÉPTIDO

<222> (1)..(23)

<223> substrato de formação de FGly sintética; sequência primária da Arilsulfatase A humana

<400> 33

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

<210> 34

<211> 16

<212> PRT

<213> Artificial

<220>

<223> uma variante do péptido ASA65-80, na qual os resíduos Cys69, Pro71 e Arg73, críticos para a formação de FGly, foram misturados

<220>

<221> CARACTERÍSTICA_MISTURADA

<222> (1)..(16)

<223> oligopéptido misturado

<400> 34

Pro	Val	Ser	Leu	Pro	Thr	Arg	Ser	Cys	Ala	Ala	Leu	Leu	Thr	Gly	Arg
1				5					10					15	

<210> 35

<211> 16

<212> PRT

<213> Artificial

<220>

<223> uma variante do péptido ASA65-80, em que Cys69 foi substituído por uma Serina

<220>

<221> CARACTERÍSTICA_MISTURADA

<222> (1)..(16)

<223> oligopéptido Ser69

<400> 35

Pro	Val	Ser	Leu	Ser	Thr	Pro	Ser	Arg	Ala	Ala	Leu	Leu	Thr	Gly	Arg
1				5					10					15	

<210> 36

<211> 19

<212> DNA

<213> Artificial

<220>

<223> "primer" de PCR específico para FGE humana

<220>

<221> característica_misturada

<222> (1)..(19)

<223> "primer" de PCR específico para FGE humana 1199nc

<400> 36

ccaatgtagg tcagacaag 19

<210> 37

<211> 16

<212> DNA

<213> Artificial

<220>

<223> "primer" de PCR específico para FGE humana

<220>

<221> característica_misturada

<222> (1)..(16)

<223> "primer" direto de PCR específico para FGE humana 1c

<400> 37

acaatggcag cgggac 16

<210> 38

<211> 19

<212> DNA

<213> Artificial

<220>

<223> "primer" de PCR específico para FGE humana

<220>

<221> característica_misturada

<222> (1)..(19)

<223> "primer" reverso de PCR específico para FGE humana
1182c

<400> 38

cgacigctcc tggactgg 19

<210> 39

<211> 24

<212> DNA

<213> Artificial

<220>

<223> "primer" de PCR específico para FGE humana

<220>

<221> característica_misturada

<222> (1)..(24)

<223> "primer" de PCR 5' específico para FGE humana
contendo EcoRI

<400> 39

ggacttcggg acacacatggc tggg 24

<210> 40

<211> 54

<212> DNA

<213> Artificial

<220>

<223> "primer" específico para HA

<220>

<221> característica_misturada

<222> (1)..(54)

<223> "primer" específico para HA

<400> 40

~~ccacagctta tgcgtagctca ggcacatcat accgatagtc catggtgggc aggc~~ 54

<210> 41

<211> 57

<212> DNA

<213> Artificial

<220>

<223> "primer" específico para c-myc

<220>

<221> característica_misturada

<222> (1)..(57)

<223> "primer" específico para c-myc

<400> 41

~~ccacagctta caggctctct tcagaaatca gctttagtc gaccatggtg ggcaggc~~ 57

<210> 42

<211> 54

<212> DNA

<213> Artificial

<220>

<223> "primer" específico para RGS-His6

<220>

<221> característica_misturada

<222> (1)..(54)

<223> "primer" específico para RGS-His6

<400> 42

~~ccasgctta ctgatggta tgggatgag atctctgto catgtgggc aggc~~ 54

<210> 43

<211> 15

<212> PRT

<213> Artificial

<220>

<223> oligopéptido triptico de uma preparação de FGE humana

<220>

<221> CARACTERÍSTICA_MISTURADA

<222> (1)..(15)

<223> oligopéptido triptico de uma preparação de FGE humana

<400> 43

Ser	Gln	Asn	Thr	Pro	Asp	Ser	Ser	Ala	Ser	Asn	Leu	Gly	Phe	Arg
1				5					10					15

<210> 44

<211> 19

<212> PRT

<213> Artificial

<220>

<223> oligopéptido triptico de uma preparação de FGE humana

<220>

<221> CARACTERÍSTICA_MISTURADA

<222> (1)..(19)

<223> oligopéptido triptico de uma preparação de FGE humana

<400> 44

Met	Val	Pro	Ile	Pro	Ala	Gly	Val	Phe	Thr	Met	Gly	Thr	Asp	Asp	Pro
1				5					10					15	
Gln Ile Lys															

<210> 45

<211> 906

<212> DNA

<213> *Homo sapiens*

<400> 45

atggcccggc	atgggttacc	gctgctgcc	ctgctgtcgc	tcttggtcgg	cgcgtggctc	60
aagctaggaa	atggacaggc	tactagcatg	gtccaactgc	aggtggggag	attcctgatg	120
ggaacaaatt	ctccagacag	cagagatggg	gaagggcctg	tgcgggaggc	gacagtgaaa	180
ccctttgcc	tgcacatatt	tctgttcacc	aacaaagatt	tcagggattt	tgtcagggag	240
aaaaagtatc	ggacagaagc	tgagatgttt	ggatggagct	ttgtctttga	ggactttgtc	300
tctgatgagc	tgagaaacaa	agccaccag	ccaatgaagt	ctgtactctg	gtggcttcca	360
gtggaaaagg	cattttggag	gcagcctgca	ggtcctggct	ctggcatccg	agagagactg	420
gagcaccag	tgttacacgt	gagctggaat	gacgccctg	cctactgtgc	ttggcgggga	480
aaacgactgc	ccacggagga	agagtgggag	tttgccgcc	gagggggctt	gaagggtcaa	540
gtttacccat	gggggaactg	gttccagcca	aaccgcacca	acctgtggca	gggaaagtcc	600
cccaaggagg	acaaagctga	ggatggcttc	catggagtct	ccccagtga	tgctttcccc	660
gccagaaca	actacgggct	ctatgacctc	ctggggaacg	tgtgggagtg	gacagcatca	720
cgtaccagg	ctgctgagca	ggacatgcgc	gtcctccggg	gggcatectg	gacgacaca	780
gctgatggct	ctgccaatca	ccgggcccg	gtcaccacca	ggatgggcaa	cactccagat	840
tcagcctcag	acaacctcgg	tttccgtgt	gctgcagacg	caggccggcc	gccaggggag	900

ctgtaa 906

<210> 46

<211> 301

<212> PRT

<213> *Homo sapiens*

<400> 46

```

Met Ala Arg His Gly Leu Pro Leu Leu Pro Leu Leu Ser Leu Leu Val
 1          5          10          15

Gly Ala Trp Leu Lys Leu Gly Asn Gly Gln Ala Thr Ser Met Val Gln
 20          25          30

Leu Gln Gly Gly Arg Phe Leu Met Gly Thr Asn Ser Pro Asp Ser Arg
 35          40          45

Asp Gly Glu Gly Pro Val Arg Glu Ala Thr Val Lys Pro Phe Ala Ile
 50          55          60

Asp Ile Phe Pro Val Thr Asn Lys Asp Phe Arg Asp Phe Val Arg Glu
 65          70          75          80

Lys Lys Tyr Arg Thr Glu Ala Glu Met Phe Gly Trp Ser Phe Val Phe
 85          90          95

Glu Asp Phe Val Ser Asp Glu Leu Arg Asn Lys Ala Thr Gln Pro Met
100          105          110

Lys Ser Val Leu Trp Trp Leu Pro Val Glu Lys Ala Phe Trp Arg Gln
115          120          125

Pro Ala Gly Pro Gly Ser Gly Ile Arg Glu Arg Leu Glu His Pro Val
130          135          140

Leu His Val Ser Trp Asn Asp Ala Arg Ala Tyr Cys Ala Trp Arg Gly
145          150          155          160

Lys Arg Leu Pro Thr Glu Glu Glu Trp Glu Phe Ala Ala Arg Gly Gly
165          170          175

Leu Lys Gly Gln Val Tyr Pro Trp Gly Asn Trp Phe Gln Pro Asn Arg
180          185          190

Thr Asn Leu Trp Gln Gly Lys Phe Pro Lys Gly Asp Lys Ala Glu Asp
195          200          205

Gly Phe His Gly Val Ser Pro Val Asn Ala Phe Pro Ala Gln Asn Asn
210          215          220

Tyr Gly Leu Tyr Asp Leu Leu Gly Asn Val Trp Glu Trp Thr Ala Ser
225          230          235          240

Pro Tyr Gln Ala Ala Glu Gln Asp Met Arg Val Leu Arg Gly Ala Ser
245          250          255

Trp Ile Asp Thr Ala Asp Gly Ser Ala Asn His Arg Ala Arg Val Thr
260          265          270

```

Thr Arg Met Gly Asn Thr Pro Asp Ser Ala Ser Asp Asn Leu Gly Phe
 275 280 285

Arg Cys Ala Ala Asp Ala Gly Arg Pro Pro Gly Glu Leu
 290 295 300

<210> 47

<211> 927

<212> DNA

<213> *Mus musculus*

<400> 47

```

atgcgctctg agttctgggt cccagcatg ggttccttgc tccctcgggt gttgctgctg      60
aggtcctctg cctgccccag gcttcagcta ggacatgcc aggatcctgc catggtgcat      120
ctgccagggt gccggtttct gatggggaca gacgctccag atggcagaga cggatgaagg      180
cctgcccggg aagtgcagct aaaacccttt gccatcgaca tatttcagct caccaataaa      240
gacttcaggg agtttctcag ggagaagaag taccagactg aagccgaggc attcgggtgg      300
agcttcgtct ttgaggatct tgtctccctt gagctcagaa agcaagaaaa tctgatgccg      360
gctgttcact ggtggcagcc agtgccaaag gcattttgga ggcagcctgc aggtcccggc      420
tctggcatcc gagagaaact ggagcttccc gtggtacacg tgagctggaa cgacgctggt      480
gcttactgct catggcgggg gagacgcttg cccacagaag aggagtggga gtttgcagcc      540
cgaggggggt tgaagggtca ggtttatcca tgggggaacc ggttcagcc aaaccgcacc      600
aacttatggc agggaaagtt ccccaaaggt gacaaagctg aagatggttt tcatggactg      660
tcaccagtga acgctttccc cccacagaac aactacggac tgtatgacct catgggcaat      720
gtgtgggagt ggacagcgtc cacataccaa cctgctggcc aggacatgct tgcctccgg      780
ggggcatcat ggatcgacac cgcagacggc tctgctaata acagggctcg ggtcaccacc      840
aggatgggaa acactccaga ctcagcctca gacaacctgg gcttcgctg cgctccagct      900
gcaggccgac cgaaggagga cctgtga                                     927

```

<210> 48

<211> 308

<212> PRT

<213> *Mus musculus*

<400> 48

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

Ala Phe Gly Trp Ser Phe Val Phe Glu Asp Phe Val Ser Pro Glu Leu
 100 105 110
 Arg Lys Gln Glu Asn Leu Met Pro Ala Val His Trp Trp Gln Pro Val
 115 120 125
 Pro Lys Ala Phe Trp Arg Gln Pro Ala Gly Pro Gly Ser Gly Ile Arg
 130 135 140
 Glu Lys Leu Glu Leu Pro Val Val His Val Ser Trp Asn Asp Ala Gly
 145 150 155 160
 Ala Tyr Cys Ala Trp Arg Gly Arg Arg Leu Pro Thr Glu Glu Glu Trp
 165 170 175
 Glu Phe Ala Ala Arg Gly Gly Leu Lys Gly Gln Val Tyr Pro Trp Gly
 180 185 190
 Asn Arg Phe Gln Pro Asn Arg Thr Asn Leu Trp Gln Gly Lys Phe Pro
 195 200 205
 Lys Gly Asp Lys Ala Glu Asp Gly Phe His Gly Leu Ser Pro Val Asn
 210 215 220
 Ala Phe Pro Pro Gln Asn Asn Tyr Gly Leu Tyr Asp Leu Met Gly Asn
 225 230 235 240
 Val Trp Glu Trp Thr Ala Ser Thr Tyr Gln Pro Ala Gly Gln Asp Met
 245 250 255
 Arg Val Leu Arg Gly Ala Ser Trp Ile Asp Thr Ala Asp Gly Ser Ala
 260 265 270
 Asn His Arg Ala Arg Val Thr Thr Arg Met Gly Asn Thr Pro Asp Ser
 275 280 285
 Ala Ser Asp Asn Leu Gly Phe Arg Cys Ala Ser Ser Ala Gly Arg Pro
 290 295 300
 Lys Glu Asp Leu
 305

<210> 49

<211> 855

<212> DNA

<213> *Mus musculus*

<400> 49

```

atggteccca ttccctgctgg agtattcaca atgggactg atgacccca gatcaggcag    60
gatggagaag cccctgccag gagagtcact gttgatggct ttacatgga cgcctatgaa    120
gtcagcaatg cggattttga gaagtttggt aactcgactg gctatttgac agaggctgag    180
aagtttgag actctttcgt ctttgaaggc atgttgagcg agcaagtga aacgcataac    240
caccaggcag ttgcagctgc tccatgggtg ttgcctgtca agggagctaa ttggagacac    300
ccagaggggc cggactccag tattctgcac aggtcaaac atccggttct ccatgtttcc    360
tggaaacgat ctgttgccca ctgcacatgg gcgggcaaga ggttgccctac tgaggcagag    420
tgggaataca gctgtagagg aggcctgcag aacaggcttt tccctgggg caacaaactg    480

cagcccaaag gacagcatta tgccaacac tggcagggca agtttctgt gagcaacact    540
ggcgaggatg gcttccaagg aactgcccc gttgatgctt ttcctccaa tggctatggc    600
ttatacaaca tagtggggaa tgtgtgggag tggacctcag actggtggac tgttcacat    660
tctgttgagg aaacgttcaa cccaagggt cccacttctg ggaaagaccg agtgaagaag    720
ggtggatcct acatgtgcca taagtcctat tgctataggt accgctgtgc agctcgaagc    780
cagaacacac cagatagctc tgcacccaac ctgggattcc gatgtgcagc cgaccacctg    840
cccaccgcag actga                                                    855

```

<210> 50

<211> 284

<212> PRT

<213> *Mus musculus*

<400> 50

```

Met Val Pro Ile Pro Ala Gly Val Phe Thr Met Gly Thr Asp Asp Pro
1      5      10      15

Gln Ile Arg Gln Asp Gly Glu Ala Pro Ala Arg Arg Val Thr Val Asp
20      25      30

Gly Phe Tyr Met Asp Ala Tyr Glu Val Ser Asn Ala Asp Phe Glu Lys
35      40      45

Phe Val Asn Ser Thr Gly Tyr Leu Thr Glu Ala Glu Lys Phe Gly Asp
50      55      60

Ser Phe Val Phe Glu Gly Met Leu Ser Glu Gln Val Lys Thr His Ile
65      70      75      80

His Gln Ala Val Ala Ala Ala Pro Trp Trp Leu Pro Val Lys Gly Ala
85      90      95

Asn Trp Arg His Pro Glu Gly Pro Asp Ser Ser Ile Leu His Arg Ser
100     105     110

Asn His Pro Val Leu His Val Ser Trp Asn Asp Ala Val Ala Tyr Cys
115     120     125

Thr Trp Ala Gly Lys Arg Leu Pro Thr Glu Ala Glu Trp Glu Tyr Ser
130     135     140

Cys Arg Gly Gly Leu Gln Asn Arg Leu Phe Pro Trp Gly Asn Lys Leu
145     150     155     160

Gln Pro Lys Gly Gln His Tyr Ala Asn Ile Trp Gln Gly Lys Phe Pro
165     170     175

Val Ser Asn Thr Gly Glu Asp Gly Phe Gln Gly Thr Ala Pro Val Asp
180     185     190

Ala Phe Pro Pro Asn Gly Tyr Gly Leu Tyr Asn Ile Val Gly Asn Val
195     200     205

Trp Glu Trp Thr Ser Asp Trp Trp Thr Val His His Ser Val Glu Glu
210     215     220

Thr Phe Asn Pro Lys Gly Pro Thr Ser Gly Lys Asp Arg Val Lys Lys
225     230     235     240

Gly Gly Ser Tyr Met Cys His Lys Ser Tyr Cys Tyr Arg Tyr Arg Cys
245     250     255

Ala Ala Arg Ser Gln Asn Thr Pro Asp Ser Ser Ala Ser Asn Leu Gly
260     265     270

Phe Arg Cys Ala Ala Asp His Leu Pro Thr Ala Asp
275     280

```

<210> 51

<211> 1011

<212> DNA

<213> *Drosophila melanogaster*

<400> 51

```

atgacaacaa ttatattagt cctctttatt tggatagttt tattcaatga cgtatccagc    60
gactgtggct gccaaaagct cgaccggaag gccccggata tgccgtccat ttccggacaa    120
gtgtgccagc aacgagcaca ggggtgcacac agccactacc gggattacta tggcgaaactg    180
gagccaaata ttgcggacat gtcactgctt cggggaggca cggtttacat ggggtactgac    240
aaaccgcact ttccggccga ccgcgaggct ccggaacggc aggtgaagct gaatgacttc    300
tacatcgaca agtatgaggt ttccaacgaa gcctttgcga agtttgttct gcacactaac    360
tacaccacgg aggctgagcg atatggcgac agttttctgt ttaagagcct tttgagccca    420
ttggagcaga agaacctaga ggacttccga gtggcgagcg ctgtctggtg gtacaaagtg    480
gccggcgtga actggcgaca tccaaatggc gtggacagcg atatagacca cttaggccga    540
caccggtagt tgcacgtatc gtggcgcgac gctgtggagt actgtaagtg ggccggcaag    600
cggttgcccc gcgaggcgga gtgggaggcg gcttgcaggg gcggcaagga gcgcaaaactg    660
tttccttggg gcaacaagct gatgccaagg aatgaacatt ggctgaacat ctggcaggga    720
gactttcccc atggcaacct ggctgaagat gggtttgagt acaccagccc cgtggatgcc    780
ttccgacaga atatttacga cctgcacaac atggtgggca acgtctggga gtggacggca    840
gatctgtggg acgtaaatga cgtttagcgt aatccaaatc gggccaagaa gggcggttct    900
tatctgtgtc acaagtcta ctgtacagg tacaggtgcg cggcacgctc gcagaacaca    960
gaagacagtt cagccggtaa cctgggtttt cggtgcgcca agaatgcgtg a          1011

```

<210> 52

<211> 336

<212> PRT

<213> *Drosophila melanogaster*

<400> 52

```

Met Thr Thr Ile Ile Leu Val Leu Phe Ile Trp Ile Val Leu Phe Asn
 1              5              10              15

Asp Val Ser Ser Asp Cys Gly Cys Gln Lys Leu Asp Arg Lys Ala Pro
      20              25              30

Asp Met Pro Ser Ile Ser Gly Gln Val Cys Gln Gln Arg Ala Gln Gly
      35              40              45

Ala His Ser His Tyr Arg Asp Tyr Tyr Gly Glu Leu Glu Pro Asn Ile

```

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

<210> 53

<211> 870

<212> DNA

<213> *Anopheles gambiae*

<400> 53

ccggagagct tgctcgatct ggtggaacat tccaagcggc tcgaagacat gaggccttatc 60

ccaggagggtg aatatgtaat cggcacaaat gaacctatct tcgtcaagga tcgcgaatca	120
ccggccccggc ccgcgacgat ccgcgacttt taectcgacc agtacgaagt ctccaacgca	180
cagttcaagg cattcgtcga ccagacgggc tacgtcacgg aggcggaaaa gtttggcgac	240
agcttcgtct tccagcagct gctcagcgaa ccgggtgcgc agcagtacga agatttcgc	300
gtggcggcgg cgccctgggtg gtacaaggta cgtggagcct cctggcagca tccggaagggt	360
gatgtgtcac gtgatataag cgaccgattg gaccatccgg tgggtgcacgt gtcctggaa	420
gatgcgggtcg cgtactgcgc ctggaaaggg aagcgcctgc cgacggaagc ggaatgggaa	480
gcggcctgcc ggggcgggtcg caagcagaag ctgttccctt ggggtaacaa gctgatgcc	540
aaggagcagc acatgatgaa catatggcag ggcgagttcc cggacagcaa tctgaaggag	600
gatggctacg agaccacctg cccggtgacg tccttcgcc agaaccggtt cgagctgtac	660
aacatcgttg gcaacgtgtg ggagtggacg gcggatcttt gggacgcgaa ggatgcggcc	720
atcgagcgca agccgggcag cgatccaccg aatcgggtga aaaaggggtg ctcataacctg	780
tgtaacgaat cgtactgcta tcgctatcgc tgtgcggctc gatcgagaa caccgaggac	840
agttcggcgg gcaatctggg cttccggtgc	870

<210> 54

<211> 290

<212> PRT

<213> *Anopheles gambiae*

<400> 54

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


```

gtggcgtgg ccgccccgtc ccccgcggc gccgcggagc cggggcccgc cgcgcgtccg    60
cgctcgaccn gcggacaggt gcgcctgccg ggcggtgagt tcgcgatggg ggacgccttc    120
ggggagggat atccggccga cggcgagaca ccgctgcaca cggctgcgct gcggcccttc    180
cacatcgacg agacggccgt caccaacgcc cggttcgccg ccttcgtcaa ggcgaccggc    240
catgtgacgg acgcggaacg ctccggctcc tcggccgtct tccacctggt cgtcgccgcc    300
ccggacgccg acgtccctcg cagcgccgcc ggcgccccct ggtggatcaa cgtgcggggc    360
gcccactggc gccgccccga gggcgccgcg tccgacatca ccggccggcc gaaccatccg    420
gtcgtccacg tctcctggaa cgatgccacc gcctacgcgc ggtgggcccg caagcgctg    480
cccaccgagg ccgaatggga gtacgcgccg cgcgggggac tggccggccg ccgctacgcc    540
tggggcgacg agctgacccc gggcgccggg tggcgctgca acatctggca gggccgcttc    600
ccgcacgtca acacggccga ggacgggca ctagcaccg caccggtcaa gtcctaccgg    660
cccaacggcc acggcctgtg gaacaccgcg ggcaacgtgt gggaatggtg ctccgactgg    720
ttctcgccca cctactacgc cgaatcacc accgtcgacc cgcacggccc cgggaccggg    780
ggggcacggg tgctgcggcg cggctcctac ctgtgccacg actcctactg caaccgctac    840
cgggtcgccg cccgctcctc caacacccc gactcctcgt ccggcaacct cggattccgc    900
tgcgccaacg acgcggacct cacgtccgga tcagccgctg agtga                      945

```

<210> 56

<211> 314

<212> PRT

<213> *Streptomyces coelicolor*

<400> 56

```

Met Ala Val Ala Ala Pro Ser Pro Ala Ala Ala Glu Pro Gly Pro
1          5          10          15

Ala Ala Arg Pro Arg Ser Thr Arg Gly Gln Val Arg Leu Pro Gly Gly
20          25          30

Glu Phe Ala Met Gly Asp Ala Phe Gly Glu Gly Tyr Pro Ala Asp Gly
35          40          45

Glu Thr Pro Val His Thr Val Arg Leu Arg Pro Phe His Ile Asp Glu
50          55          60

Thr Ala Val Thr Asn Ala Arg Phe Ala Ala Phe Val Lys Ala Thr Gly
65          70          75          80

His Val Thr Asp Ala Glu Arg Phe Gly Ser Ser Ala Val Phe His Leu
85          90          95

Val Val Ala Ala Pro Asp Ala Asp Val Leu Gly Ser Ala Ala Gly Ala
100         105         110

Pro Trp Trp Ile Asn Val Arg Gly Ala His Trp Arg Arg Pro Glu Gly
115         120         125

Ala Arg Ser Asp Ile Thr Gly Arg Pro Asn His Pro Val Val His Val
130         135         140

Ser Trp Asn Asp Ala Thr Ala Tyr Ala Arg Trp Ala Gly Lys Arg Leu
145         150         155         160

Pro Thr Glu Ala Glu Trp Glu Tyr Ala Ala Arg Gly Gly Leu Ala Gly
165         170         175

Arg Arg Tyr Ala Trp Gly Asp Glu Leu Thr Pro Gly Gly Arg Trp Arg
180         185         190

Cys Asn Ile Trp Gln Gly Arg Phe Pro His Val Asn Thr Ala Glu Asp
195         200         205

Gly His Leu Ser Thr Ala Pro Val Lys Ser Tyr Arg Pro Asn Gly His
210         215         220

Gly Leu Trp Asn Thr Ala Gly Asn Val Trp Glu Trp Cys Ser Asp Trp
225         230         235         240

Phe Ser Pro Thr Tyr Tyr Ala Glu Ser Pro Thr Val Asp Pro His Gly
245         250         255

Pro Gly Thr Gly Ala Ala Arg Val Leu Arg Gly Gly Ser Tyr Leu Cys
260         265         270

His Asp Ser Tyr Cys Asn Arg Tyr Arg Val Ala Ala Arg Ser Ser Asn
275         280         285

Thr Pro Asp Ser Ser Ser Gly Asn Leu Gly Phe Arg Cys Ala Asn Asp
290         295         300

```

Ala Asp Leu Thr Ser Gly Ser Ala Ala Glu
305 310

<210> 57

<211> 1005

<212> DNA

<213> *Corynebacterium efficiens*

<400> 57

```

gtggttcgcc atcgactggg ccaccggccc tgcacactga ggattacgtc catgagtaac    60
tgctgctccc cgtcaagcgc acaatggcgt accactaccc gggatttata agatcctgtc    120
aatcccacca ctccatgcaa cccggaacaa tcccgcgatg ctgtgacact gccgggtgga    180
gctttccaca tgggcatca tcacggggag ggtaccgg cggaaggga ggggccagta    240
catgagggtc acctcgcccc ctccggcatt aatgtcacca cggtcacgaa tgccgagttc    300
ggacgattta ttgaagccac agggatatac acgacagcgg aacgctacgg tgtctcggct    360
gtattctacg cagcgttcca agggcaacgc gctgacattc ttccgccagg ttccggcggtg    420
cctgggtggc tggcgggtcaa ggggtggaac tggcagcgtc ccaacggccc cggatccacc    480
ctggacgggc ttgaggacca ccccgctcgtt cactgttctt gggatgatgc cgttgccctac    540
tgcaacctgg ctggcggtcg tctgccacc gaagccgagt gggaatacgc cgcccggggt    600
ggactgcagg ggcacagata tgcctggggg gataacctcg ccctagacgg gaggtggaac    660
tgcaatatct ggcagggggg ctccccatg gagaacacgc ccgcggatgg ttacctcacc    720
actgcaccgg tgaagacct caccgccaat ggatacggtc tgtggcagat ggcagggaat    780
gtatgggaat ggtgccagga ctggtttgat gcggagtact actcccgtag ttccctccatc    840
aaccgcggg gaccggatac cgggtgcgcgc cgggtgatgc gcggaggtgc gtatctctgc    900
catgattcct actgcaacag ataccgggtg gccgcccgc attcgaacac cccgatttcc    960
acctcgggga ataccggttt ccggtgcgtt ttcatagatc cttga                    1005

```

<210> 58

<211> 334

<212> PRT

<213> *Corynebacterium efficiens*

<400> 58

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


```

atggcgcaac cattccgac gacggcggcc agtcgtacaa gtattgaacg ccattctcga 60
cccaattgca ggagcacgtc gogaatggtc gaacgccccg gcatgcgcct gatcgaaggc 120
ggcactttca ccatgggctc ggaagccttc taccgggagg aagcgccgct tcgccgggtg 180
aaggtagaca gcttctggat cgatgaagcg ccggtgacga acgcacagtt cgcgcgattc 240
gtggaggcca cgggatacgt cactgtggcc gagatcgagc cggatcccaa ggactacccc 300
ggcatgctcc cgggcatgga ccgcgcggga tcgctgggtg tccagaaaac agcagggccg 360
gtcgacatgg cggatgcgtc caactggtgg cactttacct ttggcgctg ctggaagcat 420

```

```

ccacttggac cgggcagttc catcgatggg atcgaggacc atcccgctgt tcacgtcgcc 480
tatgccgatg ccgaggccta tgccaaatgg gcgggcaagg atctgccgac cgaagccgag 540
ttcgaatatg ctgcgcgcgg cgggttggac ggttcgaat ttctctgggg agacgaactc 600
gcacctgaag gccggatgat ggccaactac tggcaaggcc tgtttccctt cgccaaccag 660
tgctcgatg gctgggaacg gacatcgccc gtccgcaact tcccgcccaa cggctatggt 720
ctttacgaca tgatcgggaa cacgtgggag tggacctgcg attggtgggc cgacaagccg 780
ctgactccgc aaaggaaatc ggcattgctgc gcgatcagca atccgcgcgg cggcaagctc 840
aaggacagct tcgaccgctc gcaacccgca atgcgcacgc gccggaaggt cataaagggc 900
ggttcgacc tgtgtcggc caattactgc cagcgctatc gccccgcagc acgccatcct 960
gaaatggttg ataccgcgac gacgcacatc ggcttcagggt gtgtggtgcg gccctga 1017

```

<210> 60

<211> 338

<212> PRT

<213> *Novosphingobium aromaticivorans*

<400> 60

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

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

<210> 61

<211> 1119

<212> DNA

<213> *Mesorhizobium loti*

<400> 61

```

atgggcccac gaggtcgagg tcaaaaaccg catgaaaggc gacgcggtca tgttcgacat    60
tgccgggaag ttctagccga tagcgggtgg gcggctgatg gagatgagca cgccgtgtca    120
tttcgggata ttctgatgaa cgcccctgcc gaagtcttctg agcgcgctgc agccgaacgg    180
tcgtaecccc gaatgggtctg gatccccggc ggtaccttcc tgatgggtct agacaaccac    240
tatccggagg aggcaccggc ccaccgggtc agggtcgacg gcttctggat ggacaaattc    300
accgtctcca accgcgactt cgaacgcttc gttgcggcga caggacatgt cactcttgcc    360
gagaaacccg ccaatcccg aactatccc ggtgccttac ccgatctgct ggctccgtcc    420
tcgatgatgt tcaggaagcc ggccggccct gtcgaccttg gcaatcacta caattgggtg    480
gtctatgtcc gcggcgccaa ctggcgccat ccacgcgggc cggaagtac aatcaagaag    540
gttcgagatc atccgggtct gcatgtggcc tacgaggatg tcgtggccta tgccaactgg    600
gcaggcaagg aacttccca cgaaggccgag tgggaattcg cggcgcgagg cggcctcgat    660
gccgccgaat acgtctgggg caacgagctt acgcgggccc ggaagcacat ggccaacatc    720
tggaaggag actttcccta ccggaatact gtcgacgacg gttacgaata tacggcccca    780
gtaggctctg tcccgccaa cgactacggt ctctacgaca tggccggcaa tgtctggcaa    840
tggacgaccg actggtacca ggaccacaag gcgacgaca gcccgctgctg caccgctgtc    900
aatccgcgtg gcggccatcg cgaagcgagc tatgacaccc ggctacctga cgttaagatc    960
cctcgcaagg tcaccaaggg tggctcccat ctgtgcgcgc cgaactactg tcggcgctac   1020

cggcccgcg gcggaatgg gcaacccgtc gacactgcaa tctcccatct cggctttctg   1080
tgcatcgtgc gaaggaaaat ggaattgaac gcgcagtaa                               1119

```

<210> 62

<211> 372

<212> PRT

<213> *Mesorhizobium loti*

<400> 62

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


```

atgaagagtg aaagagatcg agagcccgca aagtcgtccc gctcgaacgg gtcgggtcgca    60
gcaacccaaa cgcgcgccgg tcgctgctgc aaactaatgt tgtggggcgc cctgctcgtc    120
atactgcccg cctgtgtcgg cgccgcggtc agttgggect tcacgcccga cgcacccgct    180
cacccgcaaa tcgttttcgg cgacggcacg catggtccgc tcggcatggc gtgggtgccc    240
ggcgccagtg tcctcatggg cagcgacgcc aaacaggcgc aaccgaacga acgccccgcg    300
cacaagggtc aggtgcacgg cttctggatg gaccgccatc acgtgacca cgcgcaattc    360
cgccgcttcg tcgaagcgac cggtacgtc accacggcgg agaagaaacc cgactgggag    420
accctgaaag tccagttgcc gcccggcacg ccgcgcccg cagagagcgc gatggtggcg    480
ggtgcaatgg tgttcgtcgg caccagccgt cccgtgccgc tagacgacta ttcgcagtgg    540
tggcgctatg tgcctggcgc taactggcgt catccagccg ggcctgagag caacatcacc    600
ggtaaagatg atcaccgccg ggttcaagtg tcctacgaag atgcgcaggc ttatgcgaaa    660
tgggcgggca agcgtctgcc gaccgaagcc gaatgggaat tcgccgcgcg cggcggcctc    720
gaacaggcca cgtatgcgtg gggcgatcag ttctctccca acggcaaaca gatggccaac    780
gtctggcagg gccagcagcc gcagtcttcc cccgttgtca acccgaaagc ggggtggcgg    840
ctcgggtacaa gtcgggtggg tactttcccg gccaacggct acggccttcc cgacatgacc    900
ggcaacgcct ggcagtggtg tgccgactgg tatcgcgcg atcagttcag gcgtgaggcg    960
gtaagcacca gcgcgatcga caatccggtg ggcccagcgc agtcgtggga ccccgagac   1020
cagggcgtgc ccgtcaacgc gcccaagcgt gtcacacgcg gcggttcgtt cctctgcaac   1080
gaaatctatt gcctgagcta ccggcccagc gcgagacgcg gcaccgatcc ctacaacagc   1140
atgtcgcatc tgggcttccg gctggtgatg gacgaagaca cctggaaaga agccggtgct   1200
cgccaggctt cggcgaaagc tgccggcgcg cctggaaacc ctggcggtta g           1251

```

<210> 64

<211> 416

<212> PRT

<213> *Burkholderia fungorum*

<400> 64

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

275	280	285
Phe Pro Ala Asn Gly Tyr Gly Leu Ser Asp Met Thr Gly Asn Ala Trp 290 295 300		
Gln Trp Val Ala Asp Trp Tyr Arg Ala Asp Gln Phe Arg Arg Glu Ala 305 310 315 320		
Val Ser Thr Ser Ala Ile Asp Asn Pro Val Gly Pro Ser Glu Ser Trp 325 330 335		
Asp Pro Ala Asp Gln Gly Val Pro Val Asn Ala Pro Lys Arg Val Thr 340 345 350		
Arg Gly Gly Ser Phe Leu Cys Asn Glu Ile Tyr Cys Leu Ser Tyr Arg 355 360 365		
Pro Ser Ala Arg Arg Gly Thr Asp Pro Tyr Asn Ser Met Ser His Leu 370 375 380		
Gly Phe Arg Leu Val Met Asp Glu Asp Thr Trp Lys Glu Ala Gly Ala 385 390 395 400		
Arg Gln Ala Ser Ala Lys Ala Ala Gly Ala Pro Gly Thr Pro Gly Gly 405 410 415		

<210> 65

<211> 912

<212> DNA

<213> *Sinorhizobium meliloti*

<400> 65

```

atggtctggg ttcccgagc gaccttcacg atgggggtcga acgaccatta cccggaggaa    60
gcgcccgtgc atccggtaac cgtcgacgga ttctggatcg atgtgacacc ggtaacgaac    120
cgccagtttc tcgaattcgt aaatgcgacg gggcatgtga ccttcgcgga aagaaagccg    180
cgcgccgaag actatccggg cgctccgcca tccaatctaa gggccggttc gctcgtcttc    240
acacccccga agcgaccgct gcaggggaac gatatatcgc agtgggtggat attcacgctg    300
ggtgccaaact ggcggcaccg gctcggggcg aagagcagca tcggagcgat tctggatcat    360
ccggtcgtcc atgtcgctta cagcgacgca aaggcctatg ccgaatgggc cggcaaggac    420
ctcccgaccg agaccgagtg ggagctggcg gcccgcgcg gctcgcgatgg ggctgaattt    480
tcctggggcg gcgagcttgc gccgggcgga aatcacatgg ccaatacttg gcaggggaagt    540
tttccggtcg agaattctat ggacgatggt ttcgcgcgaa catcgccggt cagattttac    600
ccgccgaacg gctacggcct ctacgacatg atcggcaatg tgtgggagtg gaccacggat    660
tactggtccg tgcgccaccg ggaagcggcc gccaaagcctt gctgcattcc gagcaatccc    720
cgcaatgccg atgccgatgc gagtatcgat ccggcgcgga gcgtgaaagt tccgcgccgg    780
gtgctcaaag gtggatcgca tctctgcgcg ccgaactact gccggcggta ccgccctgcg    840
gcgaggcacc cccaggaat cgacacgacg accagccatg tcggtttccg atgtgtcagg    900
cgcgttcgat aa                                           912

```

<210> 66

<211> 303

<212> PRT

<213> *Sinorhizobium meliloti*

<400> 66

```

Met Val Trp Val Pro Gly Ala Thr Phe Met Met Gly Ser Asn Asp His
1      5      10      15

Tyr Pro Glu Glu Ala Pro Val His Pro Val Thr Val Asp Gly Phe Trp
      20      25      30

Ile Asp Val Thr Pro Val Thr Asn Arg Gln Phe Leu Glu Phe Val Asn
      35      40      45

Ala Thr Gly His Val Thr Phe Ala Glu Arg Lys Pro Arg Ala Glu Asp
      50      55      60

Tyr Pro Gly Ala Pro Pro Ser Asn Leu Arg Ala Gly Ser Leu Val Phe
      65      70      75      80

Thr Pro Pro Lys Arg Pro Leu Gln Gly Thr Asp Ile Ser Gln Trp Trp
      85      90      95

Ile Phe Thr Leu Gly Ala Asn Trp Arg His Pro Leu Gly Arg Lys Ser
      100      105      110

Ser Ile Gly Ala Ile Leu Asp His Pro Val Val His Val Ala Tyr Ser
      115      120      125

Asp Ala Lys Ala Tyr Ala Glu Trp Ala Gly Lys Asp Leu Pro Thr Glu
      130      135      140

Thr Glu Trp Glu Leu Ala Ala Arg Gly Gly Leu Asp Gly Ala Glu Phe
      145      150      155      160

Ser Trp Gly Gly Glu Leu Ala Pro Gly Gly Asn His Met Ala Asn Thr
      165      170      175

Trp Gln Gly Ser Phe Pro Val Glu Asn Ser Met Asp Asp Gly Phe Ala
      180      185      190

Arg Thr Ser Pro Val Arg Phe Tyr Pro Pro Asn Gly Tyr Gly Leu Tyr
      195      200      205

Asp Met Ile Gly Asn Val Trp Glu Trp Thr Thr Asp Tyr Trp Ser Val
      210      215      220

Arg His Pro Glu Ala Ala Ala Lys Pro Cys Cys Ile Pro Ser Asn Pro
      225      230      235      240

Arg Asn Ala Asp Ala Asp Ala Ser Ile Asp Pro Ala Ala Ser Val Lys
      245      250      255

Val Pro Arg Arg Val Leu Lys Gly Gly Ser His Leu Cys Ala Pro Asn
      260      265      270

Tyr Cys Arg Arg Tyr Arg Pro Ala Ala Arg His Ala Gln Glu Ile Asp
      275      280      285

Thr Thr Thr Ser His Val Gly Phe Arg Cys Val Arg Arg Val Arg
      290      295      300

```

<210> 67

<211> 1065

<212> DNA

<213> *Microscilla* sp.

<400> 67

```

atgaaataca tttttttagt tcttttctta tgggccttga cccgatgtac cggaaagtat    60
gaggacaaga gagtggaaac tgatacttcc agaccaaaag ccgaagcgtc agatataaaa    120
gttcccgaag gaatggctta tattcccgcg ggccagtaca tgatgggagg taaatcagac    180
caggcttata aggatgaata tccccgccat aacgtgaagg ttccggcttt ttatatggac    240
cttacagaag tgaccaatgc ggagtttaag cggttttagt acgaaacggg ctacgtgacc    300
attgctgaga aagatattga ctgggaagag ttaaagtctc aggtgccaca gggtagcccg    360
aagcctcctg attctgtgct tcaggcaggc tcaactgggt tcaagcagac agatgaaccc    420
gtttctctcc aggattatc acagtgggtg gaatggacta tcggagccaa ctggcgaaat    480
ccggagggtc caggtagtac gattgaggat cgtatggatc atccggtggt acacgtttcc    540
tttgaagatg tccaagcgta tgcggattgg gccggtaagc gcctgcctac tgaggcagaa    600
tgggaatggg ccgccatggg aggccaaaat gacgtgaaat atccatgggg aaatgaatcg    660
gtcgaacaag catccgataa agcaaacttt tggcagggga attttcaca tcaaaactat    720
gccctcgatg gattcgaacg caccgccctt gtacgctcct tcccagcgaa tgggtacggc    780
ctatatgata tggttggaac tgtgtgggaa tggtgccagg ataagtatga tgtcaatgct    840
tatgaaagct ataagcaaaa aggactgaca gaagaccca cgggttctga gcactacaac    900
gaccctaggg aaccgtatac tcctaagcat gtgatcagag ggggttcttt cctatgcaat    960
gacagctact gtagtgggta tcgtgtttca cgtcgtatga gttccagtag agattcaggc   1020
tttaatcata cgggattcag gtgtgtgaaa gatgtaaatg gatag                       1065

```

<210> 68

<211> 354

<212> PRT

<213> *Microscilla* sp.

<400> 68

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

```

Ser Gln Val Pro Gln Gly Thr Pro Lys Pro Pro Asp Ser Val Leu Gln
115 120 125

Ala Gly Ser Leu Val Phe Lys Gln Thr Asp Glu Pro Val Ser Leu Gln
130 135 140

Asp Tyr Ser Gln Trp Trp Glu Trp Thr Ile Gly Ala Asn Trp Arg Asn
145 150 155 160

Pro Glu Gly Pro Gly Ser Thr Ile Glu Asp Arg Met Asp His Pro Val
165 170 175

Val His Val Ser Phe Glu Asp Val Gln Ala Tyr Ala Asp Trp Ala Gly
180 185 190

Lys Arg Leu Pro Thr Glu Ala Glu Trp Glu Trp Ala Ala Met Gly Gly
195 200 205

Gln Asn Asp Val Lys Tyr Pro Trp Gly Asn Glu Ser Val Glu Gln Ala
210 215 220

Ser Asp Lys Ala Asn Phe Trp Gln Gly Asn Phe Pro His Gln Asn Tyr
225 230 235 240

Ala Leu Asp Gly Phe Glu Arg Thr Ala Pro Val Arg Ser Phe Pro Ala
245 250 255

Asn Gly Tyr Gly Leu Tyr Asp Met Ala Gly Asn Val Trp Glu Trp Cys
260 265 270

Gln Asp Lys Tyr Asp Val Asn Ala Tyr Glu Ser Tyr Lys Gln Lys Gly
275 280 285

Leu Thr Glu Asp Pro Thr Gly Ser Glu His Tyr Asn Asp Pro Arg Glu
290 295 300

Pro Tyr Thr Pro Lys His Val Ile Arg Gly Gly Ser Phe Leu Cys Asn
305 310 315 320

Asp Ser Tyr Cys Ser Gly Tyr Arg Val Ser Arg Arg Met Ser Ser Ser
325 330 335

Arg Asp Ser Gly Phe Asn His Thr Gly Phe Arg Cys Val Lys Asp Val
340 345 350

Asn Gly

```

<210> 69

<211> 876

<212> DNA

<213> *Pseudomonas putida* KT2440

<400> 69

```

atggtgcacg tgcggggcgg cgagttcagc ttggttcaa gccgcttta cgacgaagaa      60
ggcccgcttc accccgcca ggtgtccggc ttctggattg acgtgcatcc ggtcaccaac     120
gcccagttcg cgcgcttcgt caaggccacg gggatatgta cccatgccga gcgcggtacc     180
cgtgtcgagg acgaccctgc cctgcccagc gcgctgcgga taccgggtgc gatggtgttt     240

catcagggtg cggacgtgct cggccccggc tggcagttcg tgcgggcgc caactggcga      300
cacccgcaag ggcggggcag cagcctggcc gggctggaca accatccggt ggtgcagatc     360
gccctggaag atgccagggc ctatgccgcg tgggcaggcc gcgaactgcc cagcgaggcg     420
cagctggaat acgccatgcg cggcggcctg accgatgccg acttcagctg ggttaccacc     480
gagcagccca agggcaagct catggccaat acctggcagg gtcagttccc ttatcgcaat     540
gcggcgaagg atggttttac cgttacatcg ccggtggggt gcttcccggc caacggcttt     600
ggcctgttcg atgccggcgg caatgtctgg gagctgactc gcaegggcta tcggccaggc     660
catgacgcac agcgcgacgc caagctcgac ccttcaggcc cggccctgag tgacagcttc     720
gaccggcag accccggcgt gccggtggcg gtaatcaaag gcggctcgca cctgtgttcg     780
gcggaccgct gcatgcgcta ccgcccctcg gcacgccagc cgcagccggt gttcatgacg     840
acctcgcacg tgggtttcag aacgattcgg caatga                                876

```

<210> 70

<211> 291

<212> PRT

<213> *Pseudomonas putida* KT2440

<400> 70

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

<210> 71

<211> 780

<212> DNA

<213> *Ralstonia metallidurans*

<400> 71

```

atggtcgcgg gcgggatggg gttcgtcggc accaacagcc cggtgccgct gcgcgaatac    60
tggcgctggg ggcgcttcgt acctggcgcg gactggcgtc acccgaccgg cccgggcagt    120
tccatcgaag gcaaggacaa tcatcccgtc gtgcaggctc cgtatgaaga cgcgcaggcg    180
tacgccaagt gggccggcaa gcgtctgccc accgagggcg agtgggagtt tggcgcccgt    240
ggcggcctgg agcaggccac ctacgcctgg ggtgacaagt tcgcgccgga tggccggcag    300
atggcgaaatg tctggcaggg ccagcaggtg cagccggtcc cggtggtcag cgccaaggcg    360
ggcggcgcgg ctggcaccag tgctgtcggc acgttcccgg gcaatggcta tgggctctat    420
gacatgaccg gcaacgcctg gcagtggggt gccgactggg atcgegcgga ccagttccgc    480
cgcgaagcca cggtggcggc agtgctgcag aatccgaccg gcccgggcga ttcgtgggac    540
ccgaccgaac ctggcgtgcc ggtgtcggcg cccaagcggg tcacgcgcgg tggctcgttc    600
ctctgcaacg aggacttctg cctcagctac cgcccagtg cccggcgcgg taccgaccgg    660
tacaccagca tgtcgacact aggttccgg ctcgtgatgg atgacgccg ttgggcagaa    720
gttcgcaagc agccagccgt ggcaatggcc gcgggcgggc agcagaacgt gcagaaataa    780

```

<210> 72

<211> 259

<212> PRT

<213> *Ralstonia metallidurans*

<400> 72

```

Met Val Ala Gly Gly Met Val Phe Val Gly Thr Asn Ser Pro Val Pro
 1              5              10              15

Leu Arg Glu Tyr Trp Arg Trp Trp Arg Phe Val Pro Gly Ala Asp Trp
          20              25              30

Arg His Pro Thr Gly Pro Gly Ser Ser Ile Glu Gly Lys Asp Asn His

```

35	40	45
Pro Val Val Gln Val Ser Tyr Glu Asp Ala Gln Ala Tyr Ala Lys Trp		
50	55	60
Ala Gly Lys Arg Leu Pro Thr Glu Ala Glu Trp Glu Phe Ala Ala Arg		
65	70	75
Gly Gly Leu Glu Gln Ala Thr Tyr Ala Trp Gly Asp Lys Phe Ala Pro		
	85	90
Asp Gly Arg Gln Met Ala Asn Val Trp Gln Gly Gln Gln Val Gln Pro		
	100	105
Phe Pro Val Val Ser Ala Lys Ala Gly Gly Ala Ala Gly Thr Ser Ala		
	115	120
Val Gly Thr Phe Pro Gly Asn Gly Tyr Gly Leu Tyr Asp Met Thr Gly		
	130	135
Asn Ala Trp Gln Trp Val Ala Asp Trp Tyr Arg Ala Asp Gln Phe Arg		
145	150	155
Arg Glu Ala Thr Val Ala Ala Val Leu Gln Asn Pro Thr Gly Pro Ala		
	165	170
Asp Ser Trp Asp Pro Thr Glu Pro Gly Val Pro Val Ser Ala Pro Lys		
	180	185
Arg Val Thr Arg Gly Gly Ser Phe Leu Cys Asn Glu Asp Phe Cys Leu		
	195	200
Ser Tyr Arg Pro Ser Ala Arg Arg Gly Thr Asp Pro Tyr Thr Ser Met		
	210	215
Ser His Leu Gly Phe Arg Leu Val Met Asp Asp Ala Arg Trp Ala Glu		
225	230	235
Val Arg Lys Gln Pro Ala Val Ala Met Ala Ala Gly Gly Gln Gln Asn		
	245	250
Val Gln Lys		

<210> 73

<211> 876

<212> DNA

<213> *Prochlorococcus marinus*

<400> 73

```

gtgaccacat ctttgccagt agagatggta accatccccg cagggctcta tcgagttggc    60
tgtgatcgct gctatccgga tggttcagtt cgctgctatc cggaggaaac acccgcgaga    120
gaagtgcagc ttgactcatt ccagatcgac gtagggccag tcaccaatgc ccagttccga    180
gctttcgta ggcacacgca gcatttcaca gtctcggagc taccacctga tccaacgctc    240
tatcccgatc tagcgccga ggaacgcac cctgaatcag ttgtctttca accgcctcca    300
gcaacgggtg atcgagcaa acccttgagc tggtaggacc tcatggctgg ggctgattgg    360
cgtcacccc aaggaccga aagcacgac gatggcctt atgatcacc tgcgtgcat    420

```

```

gtcgctatg ccgacgccat cgcctatgcc cattgggctg gcaagcgtct cccctctgct    480
gaagagtggg aagtagccgc ccgcgggggt cttgtcgatg cccaatacgc ctgggggaat    540
gaactcactc ccaataaccg ctggatggcg aacatctggc aaggtccttt cccttggcac    600
aacgaggagc tagacggctg gttctggacc tcgcccgttg gcagctttcc tgccaacggc    660
tatggactct tggatgtttg cggcaatgtg tgggaatgga ccaactctgt ttatccctg    720
gcgtcaggcc accaggaacg gcgaactatc aaaggcggat cgtttctctg cgcagataat    780
tactgcgtac gttatcgacc ctctgcacta caaggccaga cagtagacac tgccacctgt    840
cacatgggct ttcgctgtgc aaaaggaggg ccttga    876

```

<210> 74

<211> 291

<212> PRT

<213> *Prochlorococcus marinus*

<400> 74

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

210	215	220
Asp Val Cys Gly Asn Val Trp Glu Trp Thr Asn Ser Val Tyr Pro Val		
225	230	235 240
Ala Ser Gly His Gln Glu Arg Arg Thr Ile Lys Gly Gly Ser Phe Leu		
	245	250 255
Cys Ala Asp Asn Tyr Cys Val Arg Tyr Arg Pro Ser Ala Leu Gln Gly		
	260	265 270
Gln Thr Val Asp Thr Ala Thr Cys His Met Gly Phe Arg Cys Ala Lys		
	275	280 285
Gly Gly Pro		
	290	

<210> 75

<211> 1017

<212> DNA

<213> *Caulobacter crescentus* CB15

<400> 75

```

ttgggaaaac tgacggcgct tcccgctcctg atgcttcttg cgctggccgg ctgcggccag    60
ccggcgccca aggcttgctt ggcggacctg ccggttcag atccccagaa ccgcaaggcg    120
ggatatggttc ggctggcggg cggcgacttc cagatgggcg ctgcgcgctt gcgtccggag    180
gagggaccgc cccagacggt caccggtccc cgttcttgga tcgatcagac agaggtcacc    240
aacgcgcgct tcgcgcgggt cgtcgaggcc acgggttatt gcaccgtggc cgagcgaccg    300
ctcgaccccg cgcgctacgc ccacgtaccg gcggcgacgc ggcgtccggc ctgcgtcgtc    360
ttcgtggggg cgaagggggg gaggtcggac gatccttccc aatggtggca ggtgatcccc    420
ggcgccgact ggcggcatcc cgaaggctcc ggctcgaaca tccggggcag ggacgcctgg    480
ccggtggtgc atatcgctg ggaggacgcc atggcctacg cccgctggct gggccgtgac    540
ctgcccacag aggcggaatg ggagtaagcc gcgcgcggcg ggctggttgg caagcgctac    600
acctggggcg accaggctca ggatcctgca aagccgcgcg ccaatacttg gcaaggcgtg    660
ttcccgcccc aggaccttgg caatgacggc ttcaaggcca agcccgcgcc ggtcggctgc    720
ttcccgcccc acggctatgg cctgcgcgac atggccggca atgtctggga gtggaccgcg    780
gactggttca agccgggcct ggatccggtc agcgtcctcg aaaccggcgg gccgcccgag    840
gccgcgcgcg tggatcccca ggaccggaac acgcccaagc acgtcgtgaa gggcggttcg    900
ttcctgtgcg ccgacgacta ctgcttccgc tategacctg cggcgcgaaac gccggggccg    960
ccggacagcg gcgcacgcga tgtcggtttc cgcaccgtgc tccgcgccga gcgctga    1017

```

<210> 76

<211> 338

<212> PRT

<213> *Caulobacter crescentus* CB15

<400> 76

Met	Gly	Lys	Leu	Thr	Ala	Leu	Pro	Val	Leu	Met	Leu	Leu	Ala	Leu	Ala
1				5					10					15	
Gly	Cys	Gly	Gln	Pro	Ala	Pro	Lys	Ala	Cys	Leu	Ala	Asp	Leu	Pro	Val
			20					25					30		

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

<210> 77

<211> 900

<212> DNA

<213> *Mycobacterium tuberculosis* H37Rv

<400> 77

```

gtgctgaccg agttggttga cctgcccggc ggatcggtcc gcatgggctc gacgcgcttc   60
taccgccgaag aagcgccgat tcataccgtg accgtgcgcg cctttgcggt agagcgacac   120
ccggtgacca acgcgcaatt tgccgaattc gtctccgcga caggctatgt gacggttgca   180
gaacaacccc ttgaccccg gctctacca ggagtggacg cagcagacct gtgtcccggt   240
gcgatggtgt tttgtccgac ggccggggccg gtcgacctgc gtgactggcg gcaatgggtg   300
gactgggtac ctggcgccctg ctggcgccat ccgtttggcc gggacagcga tatcgccgac   360
cgagccggcc acccggtcgt acaggtggcc tatccggacg ccgtggccta cgcacgatgg   420
gctgggtcgac gcctaccgac cgaggccgag tgggagtacg cggcccgtgg cgggaaccacg   480
gcaacctatg cgtggggcga ccaggagaag ccggggggca tgctcatggc gaacacctgg   540
cagggccggt ttcccttacg caacgacggt gcattgggct ggggtgggaac ctccccggtg   600
ggcaggttbc cggccaacgg gtttggcttg ctcgacatga tcggaaacgt ttgggagtgg   660
accaccaccg agttctatcc acaccatcgc atcgatccac cctcgacggc ctgctgcgca   720
ccggtcaagc tcgctacagc cgccgaccgg acgatcagcc agaccctcaa gggcgggctcg   780
cacctgtgcg cgccggagta ctgccaccgc taccgcccgg cggcgcgctc gccgcagtcg   840
caggacaccg cgaccacca tatcgggttc cggtgcgtag ccgacccggt gtccgggtag   900

```

<210> 78

<211> 299

<212> PRT

<213> *Mycobacterium tuberculosis* H37Rv

<400> 78

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

<210> 79

<211> 7

<212> PRT

<213> Artificial

<220>

<223> domínio conservado em procariotas e procariotas

<220>

<221> DOMÍNIO

<222> (1)..(7)

<223> domínio conservado

<220>

<221> CARACTERÍSTICA_MISTURADA

<222> (3)..(4)

<223> Qualquer aminoácido

<220>

<221> CARACTERÍSTICA_MISTURADA

<222> (6)..(6)

<223> Qualquer aminoácido

<220>

<221> CARACTERÍSTICA_MISTURADA

<222> (6)..(6)

<223> Gly ou Ala

<400> 79

Arg Val Xaa Xaa Gly Xaa Ser
1 5

<210> 80

<211> 630

<212> DNA

<213> *Oncorhynchus mykiss*

<400> 80

```

tcaggtggct gctgccccct ggtgggtgcc tgtcagagga gcagactgga ggcaccctga    60
gggccccgac tccagcatca cagacagget ggaccaccct gtgctgcatg tgtcatggca    120
ggacgctgtg gcctactget cctgggccta caagagacta cccacagagg ctgagtggga    180
gtacgcctgc agagggggcc tacaggagag actttaccgc tgggggaaca aactgaaacc    240
taaaggacag cactacgcca acctctggca gggaaagtgc cccacacaca actcagaaga    300
ggacgggtac actaaaacct caccagtga gtcatttcct gcaaatggct atggcctgta    360
caacatggta gggaatgcat gggagtggac atctgactgg tggactgtac accacaccac    420
agatgaacag cacaacccgc cagggtccacc atcaggcaca gaccgagtga agaaaggagg    480
ctctacatg tgccataagt catactgtta caggtagagg tgtgcagcac ggagtcagaa    540
caccctgac agctctgct ctaacctagg gttccgctgt gtctcccagg agcagccgta    600
accttccacc ctcgacctg acatgggtag                                     630

```

<210> 81

<211> 655

<212> DNA

<213> *Danio rerio*

<220>

<221> característica_misturada

<222> (590)..(590)

<223> n é a, c, g, ou t

<220>

<221> característica_misturada

<222> (626)..(626)

<223> n é a, c, g, ou t

<400> 81

```

caaatgggtt tatttacata aaaaaatcct cttagtgtga agtgtaagac agtgagatta    60
gtgatgtttg aggttatgga tcaacatcag aggcgcagcg gaagcccaag ttcgaggttg    120
aactgtccgg tgtgttctga ctgcgagcgg cacacctgta tctgtagcag taagacttgt    180
ggcacatgta ggatcctcct ttcttgactc tgtctgtccc tgattctggt ccccttgggt    240
taaactttgc ttctgcagtg tgatgcacag tccaccagtc tgccgtccac tcccacgcat    300
tccccaccat gtcatacagg ccaaagccat tgggaggaaa agacatcacc ggggatgtgt    360
tggcatagcc gtcctctgca gtgttgtgat tagggaaatc tccctgccac aggttagcat    420
agtgtgtccc tcttggcatt aatttatttc cccatgggta catcctgtcc tgtagtcttc    480
ctctacaggg caactcccat tcagcttctg taggaagtct gcgtttggcc cattgacagt    540
acgcccgtgc atcatcccat gaaacatgca gaggagggtg attcattctn gtgtgtatgg    600
ttgaatctgg tcctttctgg tgtctncagt ctgcaccttt cactggtgac cacca      655

```

<210> 82

<211> 773

<212> DNA

<213> *Oryzias latipes*

<220>

<221> característica_misturada

<222> (690)..(690)

<223> n é a, c, g, ou t

<220>

<221> característica_misturada

<222> (755)..(755)

<223> n é a, c, g, ou t

<400> 82

```

tctccttttt tccataaata acattagagt ccttacatcc tgcctttaça tacattgtca    60
gagacagtac aaaaaatctg cctttgtaaa attagagtta caaaaatata ttttagattt    120
gacttcttca gaattgtcgg tggcagcaaa agaatcggat tgatctcatg acaagagcgt    180
gagccagaag ttcttggatc aaactgattt ggttctgtca tcgtttctgt tcagcagcac    240
agcgaaaacc aagattggaa gcggagctgt ctggagtgtt ttggcttcga gcagcacatc    300
tgtacctgta acaataagac ttgtggcaca tgtacgagcc tcctttcttc accttatctg    360
tgcctgaagg aggaccggtt gggttgtgct gatggtctgt tgtgtggtgc acgctccacc    420
agtctgaggt ccactcccat gcgttcccca ccatgtcata cagacaaaa gcattgcctg    480
ggaaggacat caccggggag gttttagtgt agccatcctc tgcagagttg tgtgctggga    540
attccccctg ccagagggtg gcgtaatgct gtccctttgg gtttagcttg tttccccagg    600
ggtagagtct gtccttcagg ccgccccctg aggcaacctc ccactctgcc tcagtgggaa    660
gtctcttgtt gaccaggag cagtaagcch aggcattcatt ccagaaaacc tgaacgacgg    720
atgatccatc ctgtctgtga tgttgagtc tggancttca ggggtcttcc agt          773

```

<210> 83

<211> 566

<212> DNA

<213> *Xenopus laevis*

<220>

<221> característica_misturada

<222> (6)..(6)

<223> n é a, c, g, ou t

<220>

<221> característica_misturada

<222> (47)..(47)

<223> n é a, c, g, ou t

<220>

<221> característica_misturada

<222> (81)..(81)

<223> n é a, c, g, ou t

<400> 83

```

atatgnaact aaaggtaatg taattggaat gatggatttc acaaggnetg agagtccct      60
attgctcctg cttgtcgtgt nacaggtcac ggagccggcg ccacacagcg aaatcccagg    120
ttggaggccg agctgtcggg tgtattctga cttcgagcag cacagcgata cctgtagcaa    180
taggactcat ggcacatgta ggagcctcct ttcttcactc tatcatttcc cgtagaaggt    240
cctttcgggt tgtgaacctc atctgctgta tgatgagtgt cccaccaatc agatgtccac    300
tcccagcat ttcccaccat gttatataga ccataaccat tggctgggaa agcagttaca    360
gggtgaagtct gcacataacc atcctctcca gtgttttggg ttggaaaatc cccctgccag    420
acattcgcat aatgttctcc ctttgggtcc agcttgttcc cccatggaaa aatcctgttc    480
tcaagtcccc cggggcaggc gtattccac tcagcttcag ttggaaggcg tttacctgcc    540
caggtgcaga aagcagaagc atcatt                                     566

```

<210> 84

<211> 647

<212> DNA

<213> *Silurana tropicalis*

<400> 84

```

gccgcttttt tttttttttt tttttttttt catcacaaaa ataattttat taataaaata      60
ggattttgtg ttcattctta ttatgaagga caaggaatgt cattgaaatt tttgttttca    120
caaggtcttg ggagttcctt cctgctcagg tcatttttga gtggtcacgg agccgacgcc    180
acgcagcggg atcccagggt agaggccgag ctgtcaggtg tattctgact tcgagcagca    240
cagcgatacc tgtagcagta ggactcatgg cacatgtatg agcctccttt tttcaccttg    300
tcttttcccc taaaaggacc tttcgggttg taagtctcat ctgctgtatg atgagtgtcc    360
caccaatcgg atgtccactc ccaagcattt cccaccatgt tatataggct ataaccattg    420
gctgggaaag cggttacagg tgaagtctgc acatagccgt cctctccagt gttttgggtt    480
ggaaattccc cctgccagac attcgcataa tgttctcctt ttggttccag cttgttcccc    540
cacggaaaaa gcctgttctc aagtccccc cgggaggcat attcccactc agcttctgtc    600
ggaaggcgct taccgcccc ggtgcagaag gcagaagcat cgttcca                       647

```

<210> 85

<211> 636

<212> DNA

<213> *Salmo salar*

<400> 85

```

atagacattt tttaaatttt ttacaacaaa atatattcca taaatatcca catgtcatgc      60
ggtaatcctg catttcatga agaactactga catcactggc tgtatgaaga ggtgcacttg      120
atattgtttcg cctggcgggc aagataggca gagttagcac cctagactag agccaatggc      180
gaatggtaca aaaagggaaa agtcagacta cccatgtcag ggtcaagggt aaaagggttac      240
ggctgctcct gggagacaca gcggaaccct aggttagagg cagagctgtc aggggtgttc      300
tgactccgtg ctgcacacct gtacctgtaa cagtatgact tatggcacat gtaggagcct      360
cctttcttca ctgggtctgt gcctgatggg ggacctgccg ggttgtgccg ttcactctgtg      420
gtgtggtgta cagtccacca gtcagatgtc cactcccatg cattccctac catgttgtac      480
aggccatagc catttgagg aaatgacttc actggtgagg ttttggtgta cccgtcctct      540
tctgagttgt gtgtggggaa ctttccctgc cagagggttg cgtagtgtgt tccttttaggt      600
ttcagtttgt tccccacgg gtaaagtctg tcctgt                                     636

```

<210> 86

<211> 415

<212> DNA

<213> *Sus scrofa*

<400> 86

```

agtttctgt gaccaacacc ggagaggatg gcttccgagg aactgcgcct gttgatgcct      60
ttcctcccaa tggttatggc ctttacaata tagtagggaa cgcctgggaa tggacctcag      120
actggtggac cattcaccat gctgctgaag aaacaattaa cccatcaagt tcttctgtct      180
gcaccgaata acagagccgc cactacgtga tgaaagcaga gaaaggcccc ccttctggga      240
aagaccgggt gaagaaaggg ggatcctata tgtgccataa gtcctactgc tacagggtacc      300
gctgtgctgc tcgaagccag aacacgccgg acagctcggc ttcaaatctg gggttccgct      360

                                     --
gtgcagctga ccaccagccc accacagget gagtcaggaa gagtcttccc gaatc          415

```

<210> 87

<211> 595

<212> DNA

<213> *Bos taurus*

<400> 87

```

ccacgcgtcc gggggcaaca aactgcagcc gaaaggccag cattatagcc aacatcttgg      60
caaggcgagt ttctgtgac caacaccggg gaggacggct tccgagggaac cgcgcctgtt      120
gacgccttcc ctcccaatgg ttattggctt atacaatata gtagggaacg cctggggagtg      180
gacttcagac tgggtggactg ttcaccattc tgctgaagaa acgattaacc caaaaggccc      240
cccttctggg aaagaccggg tgaagaaagg tggatcctac atgtgccata aatcctattg      300
ctacaggtat cgctgtgctg ctggaagcca gaacacaccc gacagctctg cttcgaatct      360
gggattecgt tgtgcagctg accacctgcc caccacaggc taagagccaa aaagagcctt      420
cccgaaaccc agaagtcgtg tctactctgc acgcggcttc cctcagaagg ctgaacaacc      480
tgctgtgaag aattcccacc ccaagggtggg ttacatacct tgcccagtgg ccaaaggacc      540
tatggcaaga ccaaattgct gagctgatca gcatgtgcgc tttattgggg gatgg      595

```

<210> 88

<211> 1611

<212> DNA

<213> *Homo sapiens*

<220>

<221> CDS

<222> (1) .. (1608)

<223> hSULF3

<400> 88

atg cta ctg ctg tgg gtg tgg gtg gtc gca gcc ttg gcg ctg gcg gta Met Leu Leu Leu Trp Val Ser Val Val Ala Ala Leu Ala Leu Ala Val 1 5 10 15	48
ctg gcc ccc gga gca ggg gag cag agg cgg aga gca gcc aaa gcg ccc Leu Ala Pro Gly Ala Gly Glu Gln Arg Arg Ala Ala Lys Ala Pro 20 25 30	96
aat gtg gtg ctg gtc gtg agc gac tcc ttc gat gga agg tta aca ttt Asn Val Val Leu Val Val Ser Asp Ser Phe Asp Gly Arg Leu Thr Phe 35 40 45	144
cat cca gga agt cag gta gtg aaa ctt cct ttt atc aac ttt atg aag His Pro Gly Ser Gln Val Val Lys Leu Pro Phe Ile Asn Phe Met Lys 50 55 60	192
aca cgt ggg act tcc ttt ctg aat gcc tac aca aac tct cca att tgt Thr Arg Gly Thr Ser Phe Leu Asn Ala Tyr Thr Asn Ser Pro Ile Cys 65 70 75 80	240
tgc cca tca cgc gca gca atg tgg agt ggc ctc ttc act cac tta aca Cys Pro Ser Arg Ala Ala Met Trp Ser Gly Leu Phe Thr His Leu Thr 85 90 95	288
gaa tct tgg aat aat ttt aag ggt cta gat cca aat tat aca aca tgg Glu Ser Trp Asn Asn Phe Lys Gly Leu Asp Pro Asn Tyr Thr Thr Trp 100 105 110	336
atg gat gtc atg gag agg cat ggc tac cga aca cag aaa ttt ggg aaa Met Asp Val Met Glu Arg His Gly Tyr Arg Thr Gln Lys Phe Gly Lys 115 120 125	384
ctg gac tat act tca gga cat cac tcc att agt aat cgt gtg gaa gcg Leu Asp Tyr Thr Ser Gly His His Ser Ile Ser Asn Arg Val Glu Ala 130 135 140	432

tgg aca aga gat gtt gct ttc tta ctc aga caa gaa ggc agg ccc atg Trp Thr Arg Asp Val Ala Phe Leu Leu Arg Gln Glu Gly Arg Pro Met 145 150 155 160	480
gtt aat ctt atc cgt aac agg act aaa gtc aga gtg atg gaa agg gat Val Asn Leu Ile Arg Asn Arg Thr Lys Val Arg Val Met Glu Arg Asp 165 170 175	528
tgg cag aat aca gac aaa gca gta aac tgg tta aga aag gaa gca att Trp Gln Asn Thr Asp Lys Ala Val Asn Trp Leu Arg Lys Glu Ala Ile 180 185 190	576
aat tac act gaa cca ttt gtt att tac ttg gga tta aat tta cca cac Asn Tyr Thr Glu Pro Phe Val Ile Tyr Leu Gly Leu Asn Leu Pro His 195 200 205	624
cct tac cct tca cca tct tct gga gaa aat ttt gga tct tca aca ttt Pro Tyr Pro Ser Pro Ser Ser Gly Glu Asn Phe Gly Ser Ser Thr Phe 210 215 220	672
cac aca tct ctt tat tgg ctt gaa aaa gtg tct cat gat gcc atc aaa His Thr Ser Leu Tyr Trp Leu Glu Lys Val Ser His Asp Ala Ile Lys 225 230 235 240	720
atc cca aag tgg tca cct ttg tca gaa atg cac cct gta gat tat tac Ile Pro Lys Trp Ser Pro Leu Ser Glu Met His Pro Val Asp Tyr Tyr 245 250 255	768
tct tct tat aca aaa aac tgc act gga aga ttt aca aaa aaa gaa att Ser Ser Tyr Thr Lys Asn Cys Thr Gly Arg Phe Thr Lys Lys Glu Ile 260 265 270	816
aag aat att aga gca ttt tat tat gct atg tgt gct gag aca gat gcc Lys Asn Ile Arg Ala Phe Tyr Tyr Ala Met Cys Ala Glu Thr Asp Ala 275 280 285	864
atg ctt ggt gaa att att ttg gcc ctt cat caa tta gat ctt ctt cag Met Leu Gly Glu Ile Ile Leu Ala Leu His Gln Leu Asp Leu Leu Gln 290 295 300	912
aaa act att gtc ata tac tcc tca gac cat gga gag ctg gcc atg gaa Lys Thr Ile Val Ile Tyr Ser Ser Asp His Gly Glu Leu Ala Met Glu 305 310 315 320	960
cat cga cag ttt tat aaa atg agc atg tac gag gct agt gca cat gtt His Arg Gln Phe Tyr Lys Met Ser Met Tyr Glu Ala Ser Ala His Val 325 330 335	1008
ccg ctt ttg atg atg gga cca gga att aaa gcc gcc cta caa gta tca Pro Leu Leu Met Met Gly Pro Gly Ile Lys Ala Gly Leu Gln Val Ser 340 345 350	1056
aat gtg gtt tct ctt gtg gat att tac cct acc atg ctt gat att gct Asn Val Val Ser Leu Val Asp Ile Tyr Pro Thr Met Leu Asp Ile Ala 355 360 365	1104
gga att cct ctg cct cag aac ctg agt gga tac tct ttg ttg ccg tta Gly Ile Pro Leu Pro Gln Asn Leu Ser Gly Tyr Ser Leu Leu Pro Leu 370 375 380	1152
tca tca gaa aca ttt aag aat gaa cat aaa gtc aaa aac ctg cat cca Ser Ser Glu Thr Phe Lys Asn Glu His Lys Val Lys Asn Leu His Pro 385 390 395 400	1200
ccc tgg att ctg agt gaa ttc cat gga tgt aat gtg aat gcc tcc acc Pro Trp Ile Leu Ser Glu Phe His Gly Cys Asn Val Asn Ala Ser Thr 405 410 415	1248
tac atg ctt cga act aac cac tgg aaa tat ata gcc tat tgg gat ggt Tyr Met Leu Arg Thr Asn His Trp Lys Tyr Ile Ala Tyr Ser Asp Gly 420 425 430	1296
gca tca ata ttg cct caa ctc ttt gat ctt tcc tgg gat cca gat gaa Ala Ser Ile Leu Pro Gln Leu Phe Asp Leu Ser Ser Asp Pro Asp Glu 435 440 445	1344
tta aca aat gtt gct gta aaa ttt cca gaa att act tat tct ttg gat Leu Thr Asn Val Ala Val Lys Phe Pro Glu Ile Thr Tyr Ser Leu Asp 450 455	1392

```

      450              455              -- 460
cag aag ctt cat tcc att ata aac tac cct aaa gtt tct gct tct gtc 1440
Gln Lys Leu His Ser Ile Ile Asn Tyr Pro Lys Val Ser Ala Ser Val
465              470              475              480

cac cag tat aat aaa gag cag ttt atc aag tgg aaa caa agt ata gga 1488
His Gln Tyr Asn Lys Glu Gln Phe Ile Lys Trp Lys Gln Ser Ile Gly
485              490              495

cag aat tat tca aac gtt ata gca aat ctt agg tgg cac caa gac tgg 1536
Gln Asn Tyr Ser Asn Val Ile Ala Asn Leu Arg Trp His Gln Asp Trp
500              505              510

cag aag gaa cca agg aag tat gaa aat gca att gat cag tgg ctt aaa 1584
Gln Lys Glu Pro Arg Lys Tyr Glu Asn Ala Ile Asp Gln Trp Leu Lys
515              520              525

acc cat atg aat cca aga gca gtt tga 1611
Thr His Met Asn Pro Arg Ala Val
530              535

```

<210> 89

<211> 536

<212> PRT

<213> *Homo sapiens*

<400> 89

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

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

Gln Lys Glu Pro Arg Lys Tyr Glu Asn Ala Ile Asp Gln Trp Leu Lys
515 520 525

Thr His Met Asn Pro Arg Ala Val
530 535

<210> 90

<211> 1722

<212> DNA

<213> *Homo sapiens*

<220>

<221> CDS

<222> (1)..(1719)

<223> hSULF4

<400> 90

atg ggg gcg ctg gca gga ttc tgg atc ctc tgc ctc ctc act tat ggt Met Gly Ala Leu Ala Gly Phe Trp Ile Leu Cys Leu Leu Thr Tyr Gly 1 5 10 15	48
tac ctg tcc tgg ggc cag gcc tta gaa gag gag gaa gaa ggg gcc tta Tyr Leu Ser Trp Gly Gln Ala Leu Glu Glu Glu Glu Gly Ala Leu 20 25 30	96
cta gct caa gct gga gag aaa cta gag ccc agc aca act tcc acc tcc Leu Ala Gln Ala Gly Glu Lys Leu Glu Pro Ser Thr Thr Ser 35 40 45	144
cag ccc cat ctc att ttc atc cta gcg gat gat cag gga ttt aga gat Gln Pro His Leu Ile Phe Ile Leu Ala Asp Asp Gln Gly Phe Arg Asp 50 55 60	192
gtg ggt tac cac gga tct gag att aaa aca cct act ctt gac aag ctc Val Gly Tyr His Gly Ser Glu Ile Lys Thr Pro Thr Leu Asp Lys Leu 65 70 75 80	240
gct gcc gaa gga gtt aaa ctg gag aac tac tat gtc cag cct att tgc Ala Ala Glu Gly Val Lys Leu Glu Asn Tyr Tyr Val Gln Pro Ile Cys 85 90 95	288
aca cca tcc agg agt cag ttt att act gga aag tat cag ata cac acc Thr Pro Ser Arg Ser Gln Phe Ile Thr Gly Lys Tyr Gln Ile His Thr 100 105 110	336
gga ctt caa cat tct atc ata aga cct acc caa ccc aac tgt tta cct Gly Leu Gln His Ser Ile Ile Arg Pro Thr Gln Pro Asn Cys Leu Pro 115 120 125	384
ctg gac aat gcc acc cta cct cag aaa ctg aag gag gtt gga tat tca Leu Asp Asn Ala Thr Leu Pro Gln Lys Leu Lys Glu Val Gly Tyr Ser 130 135 140	432
acg cat atg gtc gga aaa tgg cac ttg ggt ttt tac aga aaa gaa tgc Thr His Met Val Gly Lys Trp His Leu Gly Phe Tyr Arg Lys Glu Cys 145 150 155 160	480
atg ccc acc aga aga gga ttt gat acc ttt ttt ggt tcc ctt ttg gga Met Pro Thr Arg Arg Gly Phe Asp Thr Phe Phe Gly Ser Leu Leu Gly 165 170 175	528
agt ggg gat tac tat aca cac tac aaa tgt gac agt cct ggg atg tgt Ser Gly Asp Tyr Tyr Thr His Tyr Lys Cys Asp Ser Pro Gly Met Cys 180 185 190	576
ggc tat gac ttg tat gaa aac gac aat gct gcc tgg gac tat gac aat Gly Tyr Asp Leu Tyr Glu Asn Asp Asn Ala Ala Trp Asp Tyr Asp Asn 195 200 205	624
ggc ata tac tcc aca cag atg tac act cag aga gta cag caa atc tta Gly Ile Tyr Ser Thr Gln Met Tyr Thr Gln Arg Val Gln Gln Ile Leu 210 215 220	672

gct tcc cat aac ccc aca aag cct ata ttt tta tat att gcc tat caa Ala Ser His Asn Pro Thr Lys Pro Ile Phe Leu Tyr Ile Ala Tyr Gln 225 230 235 240	720
gct gtt cat tca cca ctg caa gct cct ggc agg tat ttc gaa cac tac Ala Val His Ser Pro Leu Gln Ala Pro Gly Arg Tyr Phe Glu His Tyr 245 250 255	768
cga tcc att atc aac ata aac agg agg aga tat gct gcc atg ctt tcc Arg Ser Ile ile Asn Ile Asn Arg Arg Tyr Ala Ala Met Leu Ser 260 265 270	816
tgc tta gat gaa gca atc aac aac gtg aca ttg gct cta aag act tat Cys Leu Asp Glu Ala Ile Asn Asn Val Thr Leu Ala Leu Lys Thr Tyr 275 280 285	864
ggc ttc tat aac aac agc att atc att tac tct tca gat aat ggt ggc Gly Phe Tyr Asn Asn Ser Ile Ile Ile Tyr Ser Ser Asp Asn Gly Gly 290 295 300	912
cag cct acg gca gga ggg agt aac tgg cct ctc aga ggt agc aaa gga Gln Pro Thr Ala Gly Gly Ser Asn Trp Pro Leu Arg Gly Ser Lys Gly 305 310 315 320	960
aca tat tgg gaa gga ggg atc cgg gct gta ggc ttt gtg cat agc cca Thr Tyr Trp Glu Gly Gly Ile Arg Ala Val Gly Phe Val His Ser Pro 325 330 335	1008
ctt ctg aaa aac aag gga aca gtg tgt aag gaa ctt gtg cac atc act Leu Leu Lys Asn Lys Gly Thr Val Cys Lys Glu Leu Val His Ile Thr 340 345 350	1056
gac tgg tac ccc act ctc att tca ctg gct gaa gga cag att gat gag Asp Trp Tyr Pro Thr Leu Ile Ser Leu Ala Glu Gly Gln Ile Asp Glu 355 360 365	1104
gac att caa cta gat ggc tat gat atc tgg gag acc ata agt gag ggt Asp Ile Gln Leu Asp Gly Tyr Asp Ile Trp Glu Thr Ile Ser Glu Gly 370 375 380	1152
ctt cgc tca ccc cga gta gat att ttg cat aac att gac ccc ata tac Leu Arg Ser Pro Arg Val Asp Ile Leu His Asn Ile Asp Pro Ile Tyr 385 390 395 400	1200
acc aag gca aaa aat ggc tcc tgg gca gca ggc tat ggg atc tgg aac Thr Lys Ala Lys Asn Gly Ser Trp Ala Ala Gly Tyr Gly Ile Trp Asn 405 410 415	1248
act gca atc cag tca gcc atc aga gtg cag cac tgg aaa ttg ctt aca Thr Ala Ile Gln Ser Ala Ile Arg Val Gln His Trp Lys Leu Leu Thr 420 425 430	1296
gga aat cct ggc tac agc gac tgg gtc ccc cct cag tct ttc agc aac Gly Asn Pro Gly Tyr Ser Asp Trp Val Pro Pro Gln Ser Phe Ser Asn 435 440 445	1344
ctg gga ccg aac cgg tgg cac aat gaa cgg atc acc ttg tca act ggc Leu Gly Pro Asn Arg Trp His Asn Glu Arg Ile Thr Leu Ser Thr Gly 450 455 460	1392
aaa agt gta tgg ctt ttc aac atc aca gcc gac cca tat gag agg gtg Lys Ser Val Trp Leu Phe Asn Ile Thr Ala Asp Pro Tyr Glu Arg Val 465 470 475 480	1440
gac cta tct aac agg tat cca gga atc gtg aag aag ctc cta cgg agg Asp Leu Ser Asn Arg Tyr Pro Gly Ile Val Lys Lys Leu Leu Arg Arg 485 490 495	1488
ctc tca cag ttc aac aaa act gca gtg ccg gtc agg tat ccc ccc aaa Leu Ser Gln Phe Asn Lys Thr Ala Val Pro Val Arg Tyr Pro Pro Lys 500 505 510	1536
gac ccc aga agt aac cct agg ctc aat gga ggg gtc tgg gga cca tgg Asp Pro Arg Ser Asn Pro Arg Leu Asn Gly Gly Val Trp Gly Pro Trp 515 520 525	1584
tat aaa gag gaa acc aag aaa aag aag cca agc aaa aat cag gct gag Tyr Lys Glu Glu Thr Lys Lys Lys Lys Pro Ser Lys Asn Gln Ala Glu 530 535 540 545 550	1632

530	535	540	
aaa aag caa aag aaa agc aaa aaa aag aag aag aaa cag cag aaa gca			1680
Lys Lys Gln Lys Lys Ser Lys Lys Lys Lys Lys Lys Gln Gln Lys Ala			
545	550	555	560
gtc tca ggt tca act tgc cat tca ggt gtt act tgt gga taa			1722
Val Ser Gly Ser Thr Cys His Ser Gly Val Thr Cys Gly			
	565	570	

<210> 91

<211> 573

<212> PRT

<213> *Homo sapiens*

<400> 91

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

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

Val Ser Gly Ser Thr Cys His Ser Gly Val Thr Cys Gly
565 570

<210> 92

<211> 1710

<212> DNA

<213> *Homo sapiens*

<220>

<221> CDS

<222> (1) .. (1707)

<223> hSULF5

<400> 92

atg	cac	acc	ctc	act	ggc	ttc	tcc	ctg	gtc	agc	ctg	ctc	agc	ttc	ggc	48
Met	His	Thr	Leu	Thr	Gly	Phe	Ser	Leu	Val	Ser	Leu	Leu	Ser	Phe	Gly	
1				5				10						15		
tac	ctg	tcc	tgg	gac	tgg	gcc	aag	ccg	agc	ttc	gtg	gcc	gac	ggg	ccc	96
Tyr	Leu	Ser	Trp	Asp	Trp	Ala	Lys	Pro	Ser	Phe	Val	Ala	Asp	Gly	Pro	
			20					25					30			
ggg	gag	gct	ggc	gag	cag	ccc	tcg	gcc	gct	ccg	ccc	cag	cct	ccc	cac	144
Gly	Glu	Ala	Gly	Glu	Gln	Pro	Ser	Ala	Ala	Pro	Pro	Gln	Pro	Pro	His	
		35					40					45				
atc	atc	ttc	atc	ctc	acg	gac	gac	caa	ggc	tac	cac	gac	gtg	ggc	tac	192
Ile	Ile	Phe	Ile	Leu	Thr	Asp	Asp	Gln	Gly	Tyr	His	Asp	Val	Gly	Tyr	
		50				55					60					
cat	ggt	tca	gat	atc	gag	acc	cct	acg	ctg	gac	agg	ctg	gcg	gcc	aag	240
His	Gly	Ser	Asp	Ile	Glu	Thr	Pro	Thr	Leu	Asp	Arg	Leu	Ala	Ala	Lys	
65					70					75					80	
ggg	gtc	aag	ttg	gag	aat	tat	tac	atc	cag	ccc	atc	tgc	acg	cct	tcg	288
Gly	Val	Lys	Leu	Glu	Asn	Tyr	Tyr	Ile	Gln	Pro	Ile	Cys	Thr	Pro	Ser	
				85					90					95		
cgg	agc	cag	ctc	ctc	act	ggc	agg	tac	cag	atc	cac	aca	gga	ctc	cag	336
Arg	Ser	Gln	Leu	Leu	Thr	Gly	Arg	Tyr	Gln	Ile	His	Thr	Gly	Leu	Gln	
			100					105					110			
cat	tcc	atc	atc	cgc	cca	cag	cag	ccc	aac	tgc	ctg	ccc	ctg	gac	cag	384
His	Ser	Ile	Ile	Arg	Pro	Gln	Gln	Pro	Asn	Cys	Leu	Pro	Leu	Asp	Gln	
		115				120						125				
gtg	aca	ctg	cca	cag	aag	ctg	cag	gag	gca	ggt	tat	tcc	acc	cat	atg	432
Val	Thr	Leu	Pro	Gln	Lys	Leu	Gln	Glu	Ala	Gly	Tyr	Ser	Thr	His	Met	
	130					135					140					
gtg	ggc	aag	tgg	cac	ctg	ggc	ttc	tac	cgg	aag	gag	tgt	ctg	ccc	acc	480
Val	Gly	Lys	Trp	His	Leu	Gly	Phe	Tyr	Arg	Lys	Glu	Cys	Leu	Pro	Thr	
145					150					155					160	
cgt	cgg	ggc	ttc	gac	acc	ttc	ctg	ggc	tgc	ctc	acg	ggc	aat	gtg	gac	528
Arg	Arg	Gly	Phe	Asp	Thr	Phe	Leu	Gly	Ser	Leu	Thr	Gly	Asn	Val	Asp	
				165					170					175		
tat	tac	acc	tat	gac	aac	tgt	gat	ggc	cca	ggc	gtg	tgc	ggc	ttc	gac	576
Tyr	Tyr	Thr	Tyr	Asp	Asn	Cys	Asp	Gly	Pro	Gly	Val	Cys	Gly	Phe	Asp	
			180					185					190			
ctg	cac	gag	ggt	gag	aat	gtg	gcc	tgg	ggg	ctc	agc	ggc	cag	tac	tcc	624
Leu	His	Glu	Gly	Glu	Asn	Val	Ala	Trp	Gly	Leu	Ser	Gly	Gln	Tyr	Ser	
		195				200						205				
act	atg	ctt	tac	gcc	cag	cgc	gcc	agc	cat	atc	ctg	gcc	agc	cac	agc	672
Thr	Met	Leu	Tyr	Ala	Gln	Arg	Ala	Ser	His	Ile	Leu	Ala	Ser	His	Ser	
	210					215					220					
cct	cag	cgt	ccc	ctc	ttc	ctc	tat	gtg	gcc	ttc	cag	gca	gta	cac	aca	720
Pro	Gln	Arg	Pro	Leu	Phe	Leu	Tyr	Val	Ala	Phe	Gln	Ala	Val	His	Thr	
225					230					235					240	

ccc ctg cag tcc cct cgt gag tac ctg tac cgc tac cgc acc atg ggc Pro Leu Gln Ser 245 Pro Arg Glu Tyr Leu Tyr Arg Tyr Arg Thr Met Gly 255	768
aat gtg gcc cgg cgg aag tac gcg gcc atg gtg acc tgc atg gat gag Asn Val Ala Arg Arg Lys Tyr Ala Ala Met Val Thr Cys Met Asp Glu 260 265 270	816
gct gtg cgc aac atc acc tgg gcc ctc aag cgc tac ggt ttc tac aac Ala Val Asn Ile Thr Trp Leu Lys Arg Tyr Gly Phe Tyr Asn 275 280 285	864
aac agt gtc atc atc ttc tcc agt gac aat ggt gcc cag act ttc tgc Asn Ser Val Ile Ile Phe Ser Ser Asp Asn Gly Gly Gln Thr Phe Ser 290 295 300	912
ggg gcc agc aac tgg ccg ctc cga gga cgc aag gcc act tat tgg gaa Gly Gly Ser Asn Trp Pro Leu Arg Gly Arg Lys Gly Thr Tyr Trp Glu 305 310 315 320	960
ggt gcc gtg cgg gcc cta gcc ttt gtc cac agt ccc ctg ctc aag cga Gly Gly Val Arg Gly Leu Gly Phe Val His Ser Pro Leu Leu Lys Arg 325 330 335	1008
aag caa cgg aca agc cgg gca ctg atg cac atc act gac tgg tac ccg Lys Gln Arg Thr Ser Arg Ala Leu Met His Ile Thr Asp Trp Tyr Pro 340 345 350	1056
acc ctg gtg ggt ctg gca ggt ggt acc acc tca gca gcc gat ggg cta Thr Leu Val Gly Leu Ala Gly Gly Thr Thr Ser Ala Ala Asp Gly Leu 355 360 365	1104
gat gcc tac gac gtg tgg ccg gcc atc agc gag gcc cgg gcc tca cca Asp Gly Tyr Asp Val Trp Pro Ala Ile Ser Glu Gly Arg Ala Ser Pro 370 375 380	1152
cgc acg gag atc ctg cac aac att gac cca ctc tac aac cat gcc cag Arg Thr Glu Ile Leu His Asn Ile Asp Pro Leu Tyr Asn His Ala Gln 385 390 395 400	1200
cat gcc tcc ctg gag gcc gcc ttt gcc atc tgg aac acc gcc gtg cag His Gly Ser Leu Glu Gly Gly Phe Gly Ile Trp Asn Thr Ala Val Gln 405 410 415	1248
gct gcc atc cgc gtg ggt gag tgg aag ctg ctg aca gga gac ccc gcc Ala Ala Ile Arg Val Gly Glu Trp Lys Leu Leu Thr Gly Asp Pro Gly 420 425 430	1296
tat gcc gat tgg atc cca ccg cag aca ctg gcc acc ttc ccg ggt agc Tyr Gly Asp Trp Ile Pro Pro Gln Thr Leu Ala Thr Phe Pro Gly Ser 435 440 445	1344
tgg tgg aac ctg gaa cga atg gcc agt gtc cgc cag gcc gtg tgg ctc Trp Trp Asn Leu Glu Arg Met Ala Ser Val Arg Gln Ala Val Trp Leu 450 455 460	1392
ttc aac atc agt gct gac cct tat gaa cgg gag gac ctg gct gcc cag Phe Asn Ile Ser Ala Asp Pro Tyr Glu Arg Glu Asp Leu Ala Gly Gln 465 470 475 480	1440
cgg cct gat gtg gtc cgc acc ctg ctg gct cgc ctg gcc gaa tat aac Arg Pro Asp Val Val Arg Thr Leu Leu Ala Arg Leu Ala Glu Tyr Asn 485 490 495	1488
cgc aca gcc atc ccg gta cgc tac cca gct gag aac ccc cgg gct cat Arg Thr Ala Ile Pro Val Arg Tyr Pro Ala Glu Asn Pro Arg Ala His 500 505 510	1536
cct gac ttt aat ggg ggt gct tgg ggg ccc tgg gcc agt gat gag gaa Pro Asp Phe Asn Gly Gly Ala Trp Gly Pro Trp Ala Ser Asp Glu Glu 515 520 525	1584
gag gag gaa gag gaa ggg agg gct cga agc ttc tcc cgg ggt cgt cgc Glu Glu Glu Glu Gly Arg Ala Arg Ser Phe Ser Arg Gly Arg Arg 530 535 540	1632
aag aaa aaa tgc aag att tgc aag ctt cga tcc ttt ttc cgt aaa ctc Lys Lys Lys Cys Lys Ile Cys Lys Leu Arg Ser Phe Phe Arg Lys Leu	1680

```
545          550          555          560
aac acc agg cta atg tcc caa cgg atc tga      1710
Asn Thr Arg Leu Met Ser Gln Arg Ile
565
```

<210> 93

<211> 569

<212> PRT

<213> *Homo sapiens*

<400> 93

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

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

<210> 94

<211> 2067

<212> DNA

<213> *Homo sapiens*

<220>

<221> CDS

<222> (1) .. (2064)

<223> hSULF6

<400> 94

atg cta att tca gga aga gaa gag aac caa ata gac ata tcc aag acc Met Leu Ile Ser Gly Arg Glu Glu Asn Gln Ile Asp Ile Ser Lys Thr 1 5 10 15	48
aca gag gta gat tgt ttt gtg gtt gaa tta gga agt cta cac aat cct Thr Glu Val Asp Cys Phe Val Val Glu Leu Gly Ser Leu His Asn Pro 20 25 30	96
aca cgg aac cca cag cga att ttc acc aag cac gtg gcc acc aag tca Thr Arg Asn Pro Gln Arg Ile Phe Thr Lys His Val Ala Thr Lys Ser 35 40 45	144
tcc agc tcc aaa tgt cag ctg gac caa ggt gga aaa agc ctg gtc cag Ser Ser Ser Lys Cys Gln Ser Asp Gln Gly Lys Ser Leu Val Gln 50 55 60	192
tgc att tta ccc aga tct tca aag ctc ctc tca ccc ttg tgt ctc ccc Cys Ile Leu Pro Arg Ser Lys Leu Leu Ser Pro Leu Cys Leu Pro 65 70 75 80	240
cat ccg tgt gga gct tta ctt ctg tat aga tcc tca gga atc gcc tct His Pro Cys Gly Ala Leu Leu Leu Tyr Arg Ser Ser Gly Ile Ala Ser 85 90 95	288
gct ctt gct gcc ttt aca gac tcc ctc tct agg agc tgc tgg ctg tca Ala Leu Ala Ala Phe Thr Asp Ser Leu Ser Arg Ser Cys Trp Leu Ser 100 105 110	336
gtg tcc ctg tgc tgt ttg ttt tgc ggt gtt gat ggc aca ttt atg aca Val Ser Leu Cys Cys Leu Phe Cys Gly Val Asp Gly Thr Phe Met Thr 115 120 125	384
aga aac gcc aga ccc aac att gtc ctg ctg atg gca gat gac ctt gga Arg Asn Ala Arg Pro Asn Ile Val Leu Leu Met Ala Asp Asp Leu Gly 130 135 140	432
gtg ggg gat ttg tgc tgc tac ggt aat aac tca gtg agc aca cct aat Val Gly Asp Leu Cys Cys Tyr Gly Asn Asn Ser Val Ser Thr Pro Asn 145 150 155 160	480
att gac cgc ctg gca agt gaa gga gtg agg ctt acc cag cat ctc gca Ile Asp Arg Leu Ala Ser Glu Gly Val Arg Leu Thr Gln His Leu Ala 165 170 175	528
gct gct tcc atg tgc acc cca agt cgg gct gcc ttc ctg acc ggc cgg Ala Ala Ser Met Cys Thr Pro Ser Arg Ala Ala Phe Leu Thr Gly Arg 180 185 190	576
tac ccc atc aga tca ggg atg gtg tct gcc tac aac ctg aac cgt gcc Tyr Pro Ile Arg Ser Gly Met Val Ser Ala Tyr Asn Leu Asn Arg Ala 195 200 205	624
ttc acg tgg ctt ggt ggg tca ggt ggt ctt ccc acc aat gaa acg act Phe Thr Trp Leu Gly Gly Ser Gly Gly Leu Pro Thr Asn Glu Thr Thr 210 215 220	672
ttt gcc aag ctg ctg cag cac cgt ggc tac cgc acg gga ctc ata ggc Phe Ala Lys Leu Leu Gln His Arg Gly Tyr Arg Thr Gly Leu Ile Gly 225 230 235 240	720
aaa tgg cac ctg ggt ttg agc tgc gcc tct cgg aat gat cac tgt tac Lys Trp His Leu Gly Leu Ser Cys Ala Ser Arg Asn Asp His Cys Tyr 245 250 255	768

cac ccg ctc aac cat ggt ttt cac tac ttt tac ggg gtg cct ttt gga His Pro Leu Asn His Gly Phe His Tyr Phe Tyr Gly Val Pro Phe Gly 260 265 270	816
ctt tta agc gac tgc cag gca tcc aag aca cca gaa ctg cac cgc tgg Leu Leu Ser Asp Cys Gln Ala Ser Lys Thr Pro Glu Leu His Arg Trp 275 280 285	864
ctc agg atc aaa ctg tgg atc tcc acg gta gcc ctt gcc ctg gtt cct Leu Arg Ile Lys Leu Trp Ile Ser Thr Val Ala Leu Ala Leu Val Pro 290 295 300	912
ttt ctg ctt ctc att ccc aag ttc gcc cgc tgg ttc tca gtg cca tgg Phe Leu Leu Leu Ile Pro Lys Phe Ala Arg Trp Phe Ser Val Pro Trp 305 310 315 320	960
aag gtc atc ttt gtc ttt gct ctc ctc gcc ttt ctg ttt ttc act tcc Lys Val Ile Phe Val Phe Ala Leu Leu Ala Phe Leu Phe Phe Thr Ser 325 330 335	1008
tgg tac tct agt tat gga ttt act cga cgt tgg aat tgc atc ctt atg Trp Tyr Ser Ser Tyr Gly Phe Thr Arg Arg Trp Asn Cys Ile Leu Met 340 345 350	1056
agg aac cat gaa att atc cag cag cca atg aaa gag gag aaa gta gct Arg Asn His Glu Ile Ile Gln Gln Pro Met Lys Glu Glu Lys Val Ala 355 360 365	1104
tcc ctc atg ctg aag gag gca ctt gct ttc att gaa agg tac aaa agg Ser Leu Met Leu Lys Glu Ala Leu Ala Phe Ile Glu Arg Tyr Lys Arg 370 375 380	1152
gaa cct ttt ctc ctc ttt ttt tcc ttc ctg cac gta cat act cca ctc Glu Pro Phe Leu Leu Phe Phe Ser Phe Leu His Val His Thr Pro Leu 385 390 400	1200
atc tcc aaa aag aag ttt gtt ggg cgc agt aaa tat ggc agg tat ggg Ile Ser Lys Lys Lys Phe Val Gly Arg Ser Lys Tyr Gly Arg Tyr Gly 405 410 415	1248
gac aat gta gaa gaa atg gat tgg atg gtg ggt aaa atc ctg gat gcc Asp Asn Val Glu Glu Met Asp Trp Met Val Gly Lys Ile Leu Asp Ala 420 425 430	1296
ctg gac cag gag cgc ctg gcc aac cac acc ttg gtg tac ttc acc tct Leu Asp Gln Glu Arg Leu Ala Asn His Thr Leu Val Tyr Phe Thr Ser 435 440 445	1344
gac aac ggg ggc cac ctg gag ccc ctg gac ggg gct gtt cag ctg ggt Asp Asn Gly Gly His Leu Glu Pro Leu Asp Gly Ala Val Gln Leu Gly 450 455 460	1392
ggc tgg aac ggg atc tac aaa ggt ggc aaa gga atg gga gga tgg gaa Gly Trp Asn Gly Ile Tyr Lys Gly Gly Lys Gly Met Gly Gly Trp Glu 465 470 475 480	1440
gga ggt atc cgt gtg cca ggg ata ttc cgg tgg ccg tca gtc ttg gag Gly Gly Ile Arg Val Pro Gly Ile Phe Arg Trp Pro Ser Val Leu Glu 485 490 495	1488
gct ggg aga gtg atc aat gag ccc acc agc tta atg gac atc tat ccg Ala Gly Arg Val Ile Asn Glu Pro Thr Ser Leu Met Asp Ile Tyr Pro 500 505 510	1536
acg ctg tct tat ata ggc gga ggg atc ttg tcc cag gac aga gtg att Thr Leu Ser Tyr Ile Gly Gly Gly Ile Leu Ser Gln Asp Arg Val Ile 515 520 525	1584
gac ggc cag aac cta atg ccc ctg ctg gaa gga agg gcg tcc cac tcc Asp Gly Gln Asn Leu Met Pro Leu Leu Glu Gly Arg Ala Ser His Ser 530 535 540	1632
gac cac gag ttc ctc ttc cac tac tgt ggg gtc tat ctg cac acg gtc Asp His Glu Phe Leu Phe His Tyr Cys Gly Val Tyr Leu His Thr Val 545 550 555 560	1680
agg tgg cat cag aag gac tgt gca act gtg tgg aaa gct cat tat gtg Arg Trp His Gln Lys Asp Cys Ala Thr Val Trp Lys Ala His Tyr Val	1728

	565	570	575	
act cct aaa ttc tac cct gaa gga aca ggt gcc tgc tat ggg agt gga				1776
Thr Pro Lys Phe Tyr Pro Glu Gly Thr Gly Ala Cys Tyr Gly Ser Gly	580	585	590	
ata tgt tca tgt tcg ggg gat gta acc tac cac gac cca cca ctc ctc				1824
Ile Cys Ser Cys Ser Gly Asp Val Thr Tyr His Asp Pro Pro Leu Leu	595	600	605	
ttt gac atc tca aga gac cct tca gaa gcc ctt cca ctg aac cct gac				1872
Phe Asp Ile Ser Arg Asp Pro Ser Glu Ala Leu Pro Leu Asn Pro Asp	610	615	620	
aat gag cca tta ttt gac tcc gtg atc aaa aag atg gag gca gcc ata				1920
Asn Glu Pro Leu Phe Asp Ser Val Ile Lys Lys Met Glu Ala Ala Ile	625	630	635	640
aga gag cat cgt agg aca cta aca cct gtc cca cag cag ttc tct gtg				1968
Arg Glu His Arg Arg Thr Leu Thr Pro Val Pro Gln Gln Phe Ser Val	645	650	655	
ttc aac aca att tgg aaa cca tgg ctg cag cct tgc tgt ggg acc ttc				2016
Phe Asn Thr Ile Trp Lys Pro Trp Leu Gln Pro Cys Cys Gly Thr Phe	660	665	670	
ccc ttc tgt ggg tgt gac aag gaa gat gac atc ctt ccc atg gct ccc				2064
Pro Phe Cys Gly Cys Asp Lys Glu Asp Asp Ile Leu Pro Met Ala Pro	675	680	685	
tga				2067

<210> 95

<211> 688

<212> PRT

<213> *Homo sapiens*

<400> 95

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

Ile Asp Arg Leu Ala Ser Glu Gly Val Arg Leu Thr Gln His Leu Ala
 165 170 175
 Ala Ala Ser Met Cys Thr Pro Ser Arg Ala Ala Phe Leu Thr Gly Arg
 180 185 190
 Tyr Pro Ile Arg Ser Gly Met Val Ser Ala Tyr Asn Leu Asn Arg Ala
 195 200 205
 Phe Thr Trp Leu Gly Gly Ser Gly Gly Leu Pro Thr Asn Glu Thr Thr
 210 215 220
 Phe Ala Lys Leu Leu Gln His Arg Gly Tyr Arg Thr Gly Leu Ile Gly
 225 230 235 240
 Lys Trp His Leu Gly Leu Ser Cys Ala Ser Arg Asn Asp His Cys Tyr
 245 250 255
 His Pro Leu Asn His Gly Phe His Tyr Phe Tyr Gly Val Pro Phe Gly
 260 265 270
 Leu Leu Ser Asp Cys Gln Ala Ser Lys Thr Pro Glu Leu His Arg Trp
 275 280 285
 Leu Arg Ile Lys Leu Trp Ile Ser Thr Val Ala Leu Ala Leu Val Pro
 290 295 300
 Phe Leu Leu Leu Ile Pro Lys Phe Ala Arg Trp Phe Ser Val Pro Trp
 305 310 315 320
 Lys Val Ile Phe Val Phe Ala Leu Leu Ala Phe Leu Phe Phe Thr Ser
 325 330 335
 Trp Tyr Ser Ser Tyr Gly Phe Thr Arg Arg Trp Asn Cys Ile Leu Met
 340 345 350
 Arg Asn His Glu Ile Ile Gln Gln Pro Met Lys Glu Glu Lys Val Ala
 355 360 365
 Ser Leu Met Leu Lys Glu Ala Leu Ala Phe Ile Glu Arg Tyr Lys Arg
 370 375 380
 Glu Pro Phe Leu Leu Phe Phe Ser Phe Leu His Val His Thr Pro Leu
 385 390 395 400
 Ile Ser Lys Lys Lys Phe Val Gly Arg Ser Lys Tyr Gly Arg Tyr Gly
 405 410 415
 Asp Asn Val Glu Glu Met Asp Trp Met Val Gly Lys Ile Leu Asp Ala
 420 425 430
 Leu Asp Gln Glu Arg Leu Ala Asn His Thr Leu Val Tyr Phe Thr Ser
 435 440 445
 Asp Asn Gly Gly His Leu Glu Pro Leu Asp Gly Ala Val Gln Leu Gly
 450 455 460

Gly Trp Asn Gly Ile Tyr Lys Gly Gly Lys Gly Met Gly Gly Trp Glu
 465 470 475 480
 Gly Gly Ile Arg Val Pro Gly Ile Phe Arg Trp Pro Ser Val Leu Glu
 485 490 495
 Ala Gly Arg Val Ile Asn Glu Pro Thr Ser Leu Met Asp Ile Tyr Pro ,
 500 505 510
 Thr Leu Ser Tyr Ile Gly Gly Gly Ile Leu Ser Gln Asp Arg Val Ile
 515 520 525
 Asp Gly Gln Asn Leu Met Pro Leu Leu Glu Gly Arg Ala Ser His Ser
 530 535 540
 Asp His Glu Phe Leu Phe His Tyr Cys Gly Val Tyr Leu His Thr Val
 545 550 555 560
 Arg Trp His Gln Lys Asp Cys Ala Thr Val Trp Lys Ala His Tyr Val
 565 570 575
 Thr Pro Lys Phe Tyr Pro Glu Gly Thr Gly Ala Cys Tyr Gly Ser Gly
 580 585 590
 Ile Cys Ser Cys Ser Gly Asp Val Thr Tyr His Asp Pro Pro Leu Leu
 595 600 605
 Phe Asp Ile Ser Arg Asp Pro Ser Glu Ala Leu Pro Leu Asn Pro Asp
 610 615 620
 Asn Glu Pro Leu Phe Asp Ser Val Ile Lys Lys Met Glu Ala Ala Ile
 625 630 635 640
 Arg Glu His Arg Arg Thr Leu Thr Pro Val Pro Gln Gln Phe Ser Val
 645 650 655
 Phe Asn Thr Ile Trp Lys Pro Trp Leu Gln Pro Cys Cys Gly Thr Phe
 660 665 670
 Pro Phe Cys Gly Cys Asp Lys Glu Asp Asp Ile Leu Pro Met Ala Pro
 675 680 685

Lisboa, 11 de Janeiro de 2012

REIVINDICAÇÕES

1. Método para a determinação do nível de expressão de enzima geradora de C α -formilglicina (FGE) num sujeito, cujo método compreende medir *in vitro*, numa amostra obtida de um sujeito, a atividade geradora de C α -formilglicina de um péptido ou polipéptido possuindo uma sequência selecionada do grupo que consiste em: aminoácidos 34-374 da SEQ ID NO: 2, SEQ ID NO: 2, 5, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76 e 78.

2. Método para a identificação de um agente útil na modulação da atividade geradora de C α -formilglicina, que compreende:

(a) contactar uma molécula possuindo atividade geradora de C α -formilglicina, ou uma molécula de ácido nucleico que codifica uma molécula com a referida atividade, com um agente candidato,

(b) medir a atividade geradora de C α -formilglicina da molécula, e

(c) comparar a atividade geradora de C α -formilglicina medida da molécula com um controlo, para determinar se o agente candidato modula a atividade geradora de C α -formilglicina da molécula;

em que a molécula é uma molécula de ácido nucleico possuindo a sequência de nucleótidos apresentada como SEQ ID NO: 1, ou é um respetivo produto de expressão.

3. Método para a determinação de uma deficiência de sulfatase num sujeito **caracterizada por** expressão aberrante de uma molécula de ácido nucleico ou um respetivo produto de expressão, que compreende:

monitorizar uma amostra de um paciente quanto a um parâmetro selecionado do grupo que consiste em:

- (i) uma molécula de ácido nucleico possuindo a sequência de nucleótidos apresentada como SEQ ID NO: 1,
- (ii) um polipéptido codificado pela molécula de ácido nucleico,
- (iii) um péptido derivado do polipéptido, e
- (iv) um anticorpo que se liga seletivamente ao polipéptido ou péptido,

como determinação de uma deficiência de sulfatase no sujeito.

4. Método da reivindicação 3, em que a amostra é um fluido biológico ou um tecido.
5. Método da reivindicação 3, em que o passo de monitorização compreende contactar a amostra com um agente detetável selecionado do grupo que consiste em:

- (a) uma molécula de ácido nucleico que hibridiza seletivamente, em condições restritas, com a molécula de ácido nucleico de (i),
- (b) um anticorpo que se liga seletivamente ao polipéptido de (ii), ou ao péptido de (iii), e

(c) um polipéptido ou péptido que se liga ao anticorpo de (iv).

6. Método da reivindicação 5, em que o anticorpo, o polipéptido, o péptido ou o ácido nucleico é etiquetado com uma etiqueta radioativa ou uma enzima.
7. Método da reivindicação 3, que compreende avaliar a amostra quanto ao péptido.
8. Estojo compreendendo uma embalagem que contém um agente e um controlo;
em que o agente é um anticorpo ou respetivo fragmento que se liga seletivamente a um péptido ou polipéptido como descrito na reivindicação 1, ou é um ácido nucleico que se liga seletivamente a um ácido nucleico que codifica o referido péptido ou polipéptido;
e em que o controlo se destina a comparação com um valor medido de ligação do referido agente ao referido péptido, polipéptido ou ácido nucleico.
9. Estojo da reivindicação 8; em que o controlo consiste num valor predeterminado para comparação com o valor medido.
10. Estojo da reivindicação 8; em que o controlo compreende um epítipo do péptido ou polipéptido.
11. Utilização de uma molécula de ácido nucleico, péptido ou polipéptido que modula a atividade geradora de C_{α} -formilglicina na preparação de um medicamento para o tratamento de uma deficiência de sulfatases num sujeito; em que:

(i) a molécula de ácido nucleico é selecionada do grupo que consiste em:

(a) uma molécula de ácido nucleico de hibridização que hibridiza, em condições restritas, com uma molécula que consiste numa sequência de nucleótidos apresentada como SEQ ID NO: 1, ou um complemento da referida molécula de hibridização que codifica um polipéptido com atividade de FGE;

(b) uma molécula de ácido nucleico que difere de uma molécula de ácido nucleico de (a) devido à degenerescência do código genético; e

(c) um ácido nucleico possuindo uma sequência selecionada do grupo que consiste em: SEQ ID NO: 1, 3, 4, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77 e 80-87; e

(ii) o péptido ou o polipéptido é selecionado do grupo que consiste em:

(a) um péptido ou um polipéptido codificado por uma molécula de ácido nucleico como descrito acima; e

(b) um péptido ou um polipéptido que tem uma sequência selecionada do grupo que consiste em: aminoácidos 34-374 da SEQ ID NO: 2, SEQ ID NO: 2, 5, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76 e 78.

12. Utilização da reivindicação 11; em que o medicamento também compreende um agente selecionado do grupo que consiste em: uma molécula de ácido nucleico que codifica a Iduronato 2-Sulfatase, Sulfamidase, N-

Acetilgalactosamina 6-Sulfatase, N-Acetilglucosamina 6-Sulfatase, Arilsulfatase A, Arilsulfatase B, Arilsulfatase C, Arilsulfatase D, Arilsulfatase E, Arilsulfatase F, Arilsulfatase G, HSulf-1, HSulf-2, HSulf-3, HSulf-4, HSulf-5 ou HSulf-6; um produto de expressão da molécula de ácido nucleico; e um fragmento do produto de expressão da molécula de ácido nucleico.

13. Utilização da reivindicação 11; em que o agente que modula a atividade geradora de C_{α} -formilglicina é produzido por uma célula que expressa a referida molécula de ácido nucleico.
14. Utilização da reivindicação 13; em que a célula expressa uma molécula de ácido nucleico FGE exógena.
15. Utilização da reivindicação 13; em que a célula expressa uma molécula de ácido nucleico FGE endógena.
16. Molécula de ácido nucleico, péptido ou polipéptido que modula a atividade geradora de C_{α} -formilglicina; para utilização no tratamento de uma deficiência de sulfatases num sujeito; em que a molécula de ácido nucleico, péptido ou polipéptido é como descrita na reivindicação 11.
17. Composição farmacêutica que compreende uma molécula de ácido nucleico, péptido, ou polipéptido como descrita na reivindicação 11, e um transportador farmacêuticamente aceitável.

18. Método de identificação de um agente candidato útil no tratamento de uma deficiência de sulfatases, que compreende:

determinar a expressão de um conjunto de moléculas de ácidos nucleicos *in vitro* numa célula ou tecido em condições tais que, na ausência de um agente candidato, permitem uma primeira quantidade de expressão do conjunto de moléculas de ácidos nucleicos; em que o conjunto de moléculas de ácidos nucleicos compreende pelo menos uma molécula de ácido nucleico que é uma molécula de ácido nucleico como descrita na reivindicação 11;

contactar a célula ou tecido com o agente candidato *in vitro*, e

detetar uma quantidade de teste de expressão do conjunto de moléculas de ácidos nucleicos;

em que um aumento da quantidade de teste de expressão na presença do agente candidato relativamente à primeira quantidade de expressão indica que o agente candidato é útil no tratamento da deficiência de sulfatases.

19. Série de moléculas de ácidos nucleicos em fase sólida que consiste num conjunto de moléculas de ácidos nucleicos, respetivos produtos de expressão, ou respetivos fragmentos, quando fixados num substrato sólido; em que cada molécula de ácido nucleico codifica um polipéptido selecionado do grupo que consiste nos aminoácidos 34-374 da SEQ ID NO: 2, SEQ ID NO: 2, 5, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76 e 78, Iduronato 2-Sulfatase, Sulfamidase, N-Acetilgalactosamina 6-Sulfatase, N-Acetilglucosamina 6-

Sulfatase, Arilsulfatase A, Arilsulfatase B, Arilsulfatase C, Arilsulfatase D, Arilsulfatase E, Arilsulfatase F, Arilsulfatase G, HSulf-1, HSulf-2, HSulf-3, HSulf-4, HSulf-5, e HSulf-6.

20. Série de moléculas de ácidos nucleicos em fase sólida da reivindicação 19, que também compreende pelo menos uma molécula de ácido nucleico de controlo.

21. Série de moléculas de ácidos nucleicos em fase sólida da reivindicação 19; em que o conjunto de moléculas de ácidos nucleicos compreende pelo menos uma molécula de ácido nucleico que codifica um polipéptido selecionado do grupo que consiste em: aminoácidos 34-374 da SEQ ID NO: 2, SEQ ID NO: 2, 5, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76 e 78, Iduronato 2-Sulfatase, Sulfamidase, N-Acetilgalactosamina 6-Sulfatase, N-Acetilglucosamina 6-Sulfatase, Arilsulfatase A, Arilsulfatase B, Arilsulfatase C, Arilsulfatase D, Arilsulfatase E, Arilsulfatase F, Arilsulfatase G, HSulf-1, HSulf-2, HSulf-3, HSulf-4, HSulf-5, e HSulf-6.

22. Série de moléculas de ácidos nucleicos em fase sólida da reivindicação 19; em que o conjunto de moléculas de ácidos nucleicos compreende pelo menos duas moléculas de ácidos nucleicos, em que cada molécula de ácido nucleico codifica um polipéptido selecionado do grupo que consiste em: aminoácidos 34-374 da SEQ ID NO: 2, SEQ ID NO: 2, 5, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76 e 78, Iduronato 2-Sulfatase, Sulfamidase, N-Acetilgalactosamina 6-Sulfatase, N-Acetilglucosamina 6-Sulfatase, Arilsulfatase A,

Arilsulfatase B, Arilsulfatase C, Arilsulfatase D, Arilsulfatase E, Arilsulfatase F, Arilsulfatase G, HSulf-1, HSulf-2, HSulf-3, HSulf-4, HSulf-5, e HSulf-6.

23. Alelo variante isolado de um gene FGE humano que codifica um polipéptido FGE variante, em que o polipéptido compreende:

uma sequência de aminoácidos compreendendo pelo menos uma variação na SEQ ID NO: 2, em que a pelo menos uma variação compreende: Met1Arg; Met1Val; Ser155Pro; Cys218Tyr; Ala279Val; Arg327Stop; Cys336Arg; Arg345Cys; Arg349Trp; Arg349Gln; Ser359Stop; ou uma combinação destas.

24. Polipéptido FGE humano variante isolado que compreende uma sequência de aminoácidos com pelo menos uma variação na SEQ ID NO: 2, em que a pelo menos uma variação compreende: Met1Arg; Met1Val; Ser155Pro; Cys218Tyr; Ala279Val; Arg327Stop; Cys336Arg; Arg345Cys; Arg349Trp; Arg349Gln; Ser359Stop; ou uma combinação destas.

25. Anticorpo possuindo o polipéptido FGE humano variante da reivindicação 23 como imunogénio.

26. Anticorpo da reivindicação 25, que é um anticorpo policlonal.

27. Anticorpo da reivindicação 25, que é um anticorpo monoclonal.

28. Anticorpo da reivindicação 25, que é um anticorpo quimérico.
29. Anticorpo da reivindicação 25, etiquetado de forma detetável.
30. Anticorpo da reivindicação 29, em que a referida etiqueta detetável compreende um elemento radioativo, um químico fluorescente, ou uma enzima.
31. Método de produção de uma sulfatase ativada, em que o método compreende:
- (a) contactar a sulfatase com uma Enzima Geradora de C_{α} -Formililglicina (FGE) *in vitro*; em que a Enzima Geradora de C_{α} -Formililglicina é um polipéptido ou péptido como descrito na reivindicação 1; ou
 - (b) proporcionar uma célula que coexpressa uma sulfatase e uma Enzima Geradora de C_{α} -Formililglicina (FGE) de modo que uma sulfatase ativada é produzida na célula; em que a célula compreende DNA ou RNA heterólogo que resulta em expressão aumentada da sulfatase ativada, relativamente à que ocorreria na ausência do DNA ou RNA heterólogo; e em que a Enzima Geradora de C_{α} -Formililglicina é um polipéptido ou péptido como descrito na reivindicação 1.
32. Método da reivindicação 31, em que a sulfatase é selecionada do grupo que consiste em Iduronato 2-Sulfatase, Sulfamidase, N-Acetilgalactosamina 6-Sulfatase, N-Acetilglucosamina 6-Sulfatase, Arilsulfatase A, Arilsulfatase B, Arilsulfatase C, Arilsulfatase D, Arilsulfatase E, Arilsulfatase F,

Arilsulfatase G, HSulf-1, HSulf-2, HSulf-3, HSulf-4, HSulf-5 e HSulf-6.

33. Método de acordo com qualquer uma das reivindicações 3 até 7 ou 18, ou utilização de acordo com qualquer uma das reivindicações 11 até 16; em que a deficiência de sulfatases referida relativamente ao referido método ou utilização é selecionada do grupo que consiste em: Deficiência de Múltiplas Sulfatases, Mucopolissacaridose II, Mucopolissacaridose IIIA, Mucopolissacaridose IVA, Mucopolissacaridose VI, Mucopolissacaridose VIII, Leucodistrofia Metacromática, Condrodisplasia *Punctata* 1 Recessiva ligada ao cromossoma X e Ictiose ligada ao cromossoma X.

Lisboa, 11 de Janeiro de 2012

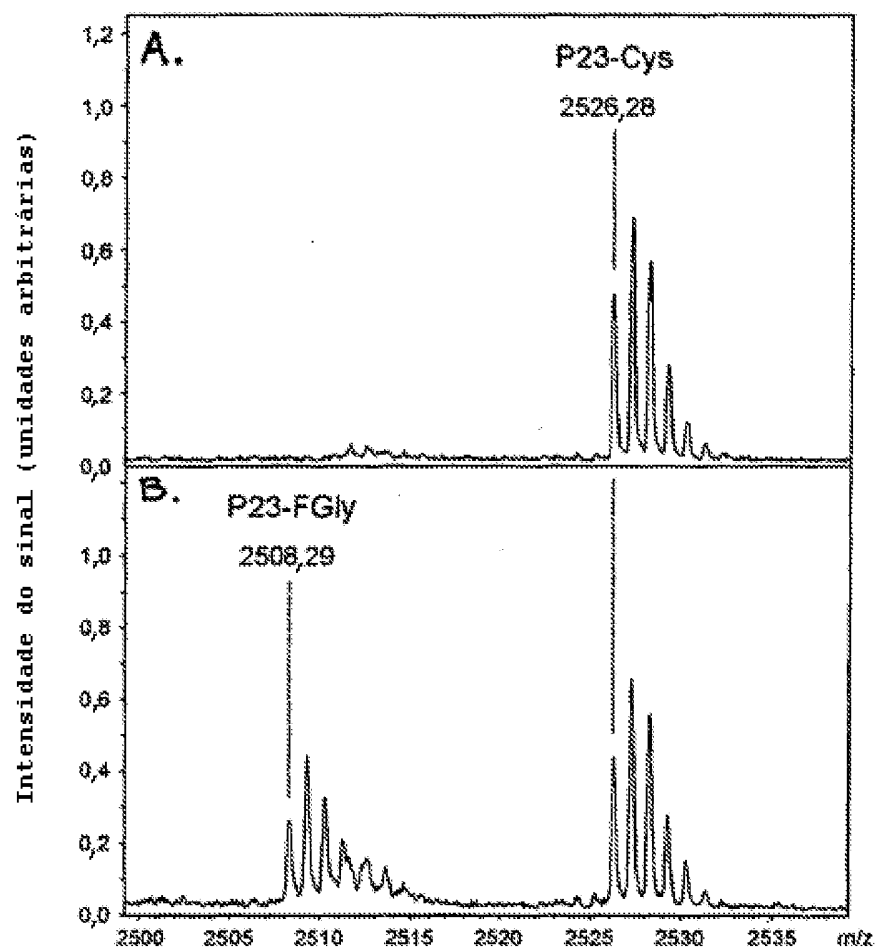


Fig. 1

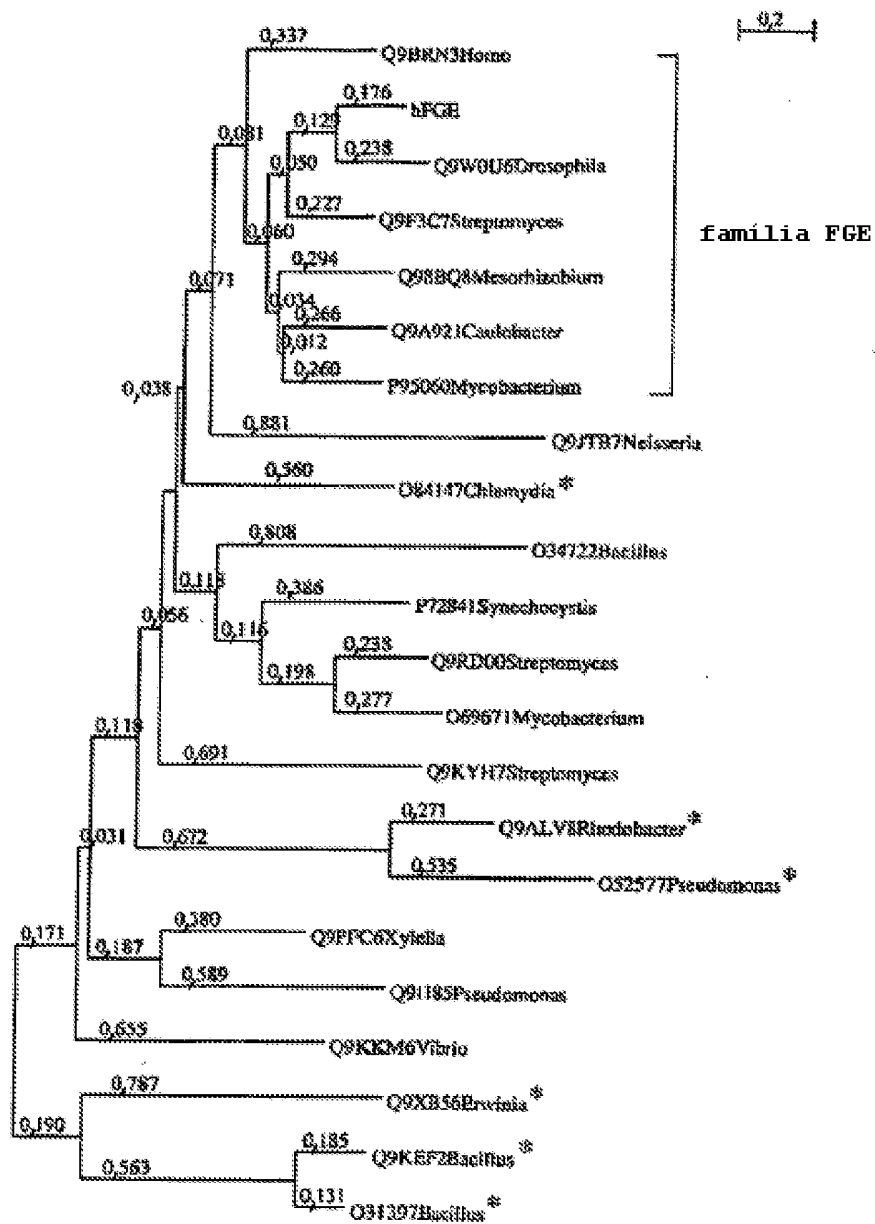


Fig. 2

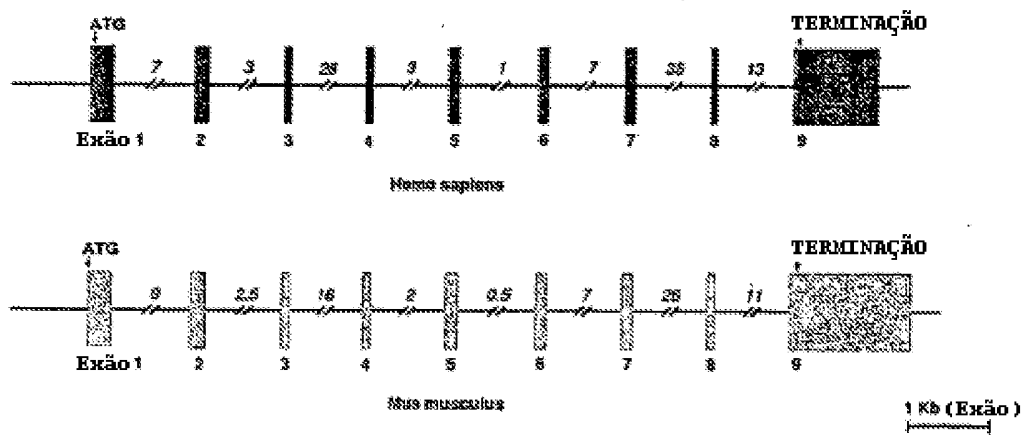


Fig. 3

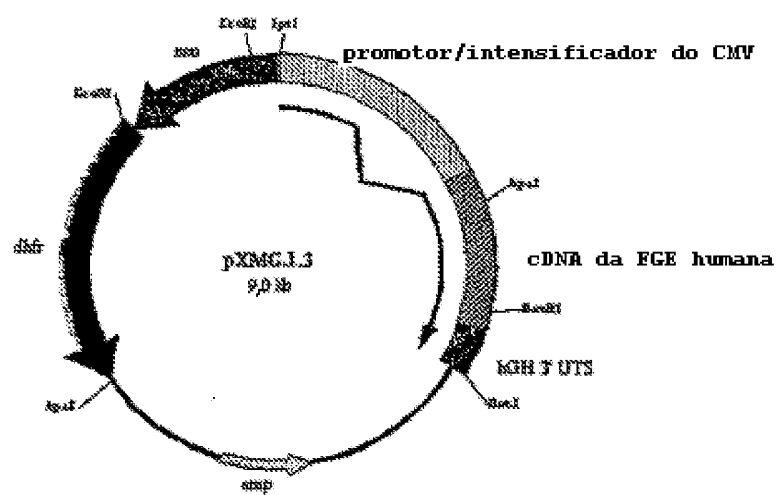


Fig. 4

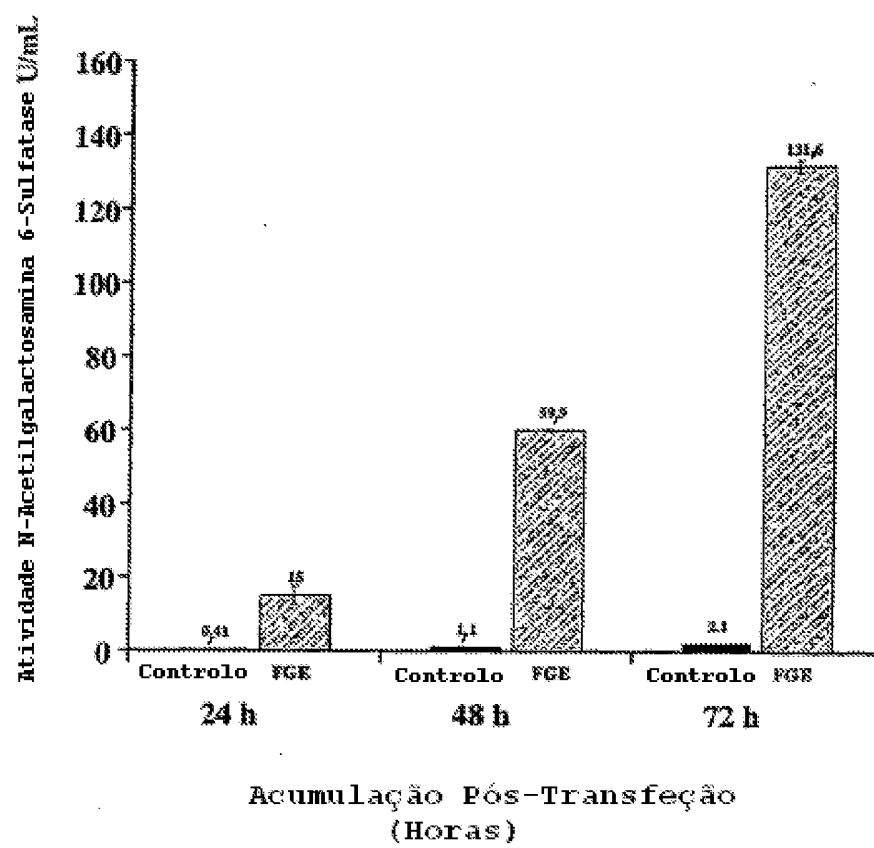


Fig. 5

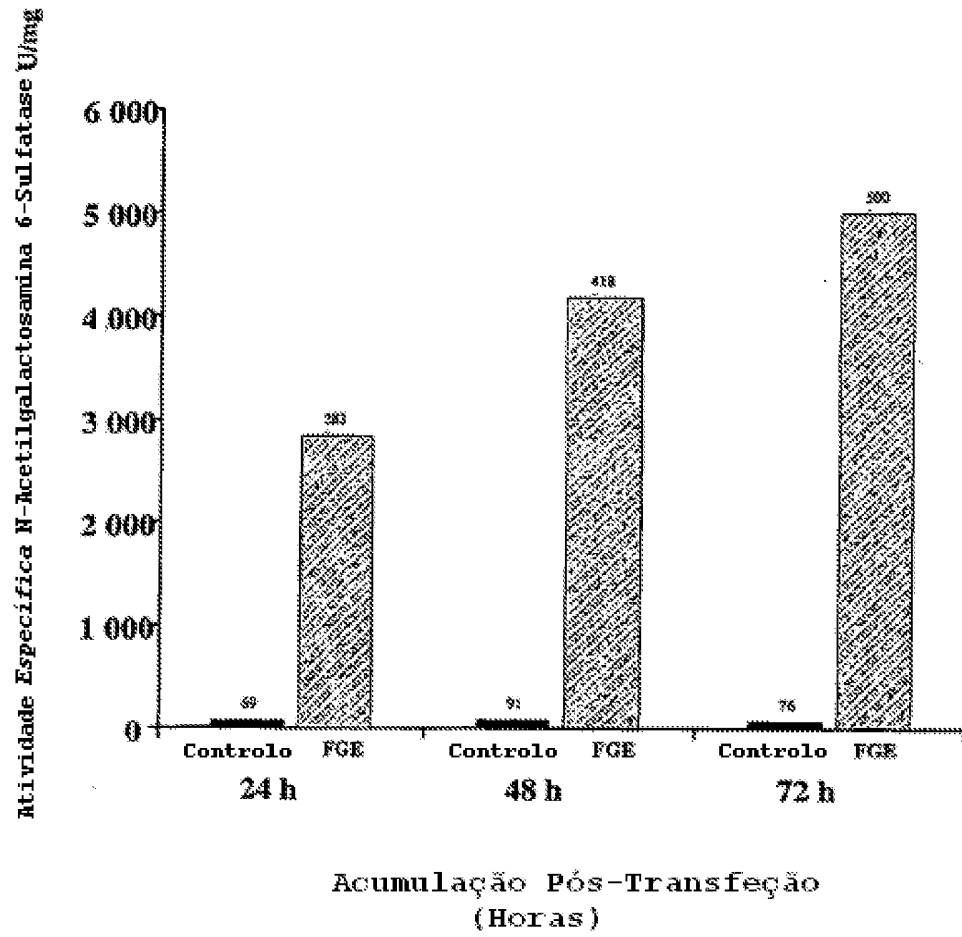


Fig. 6

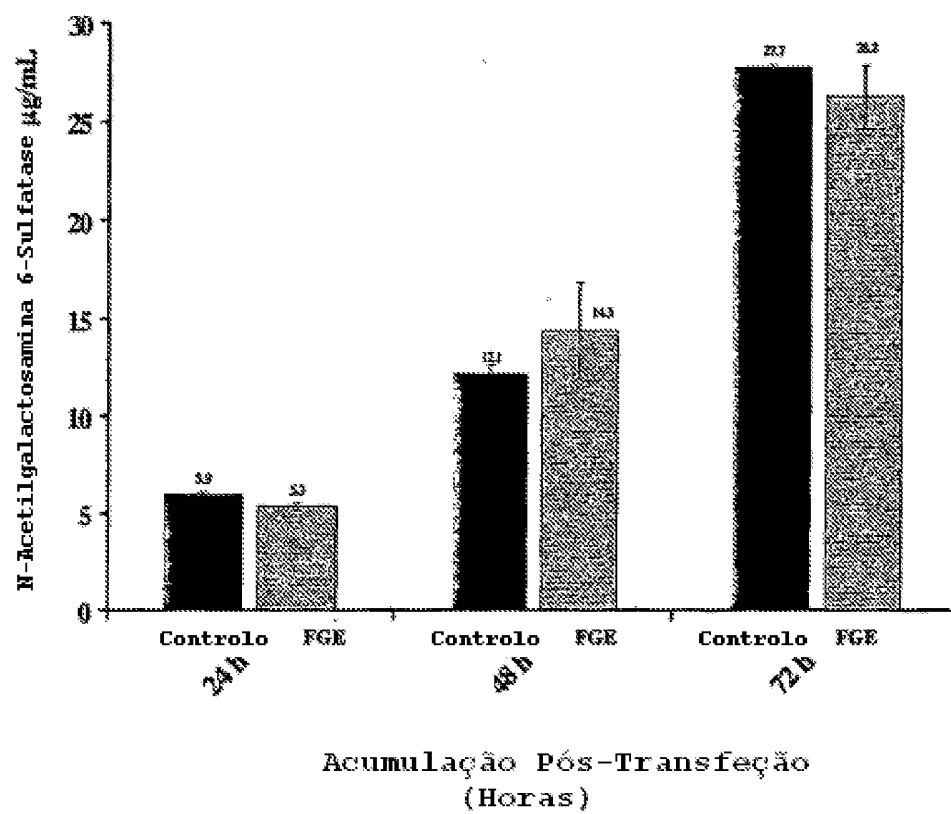


Fig. 7

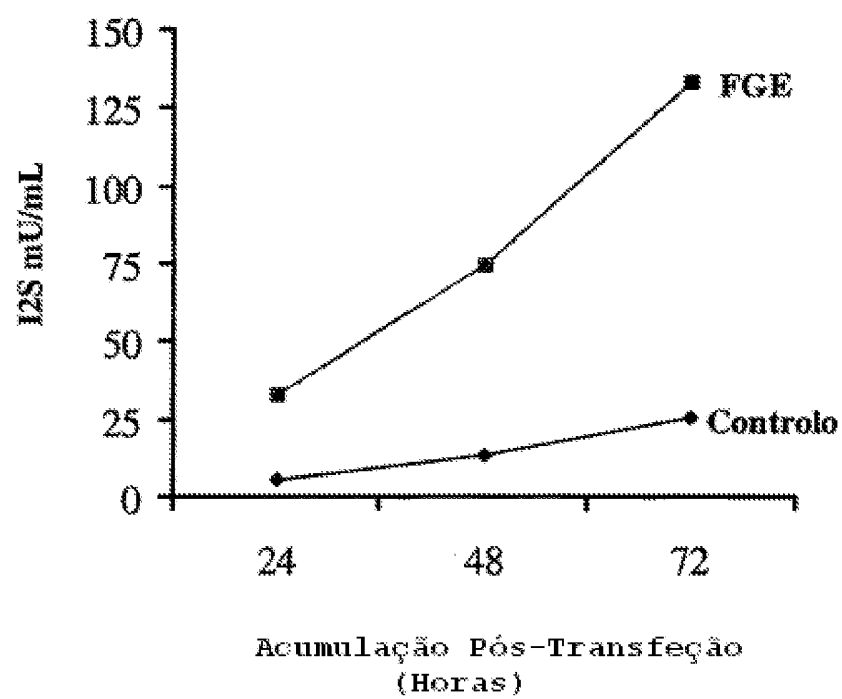


Fig. 8

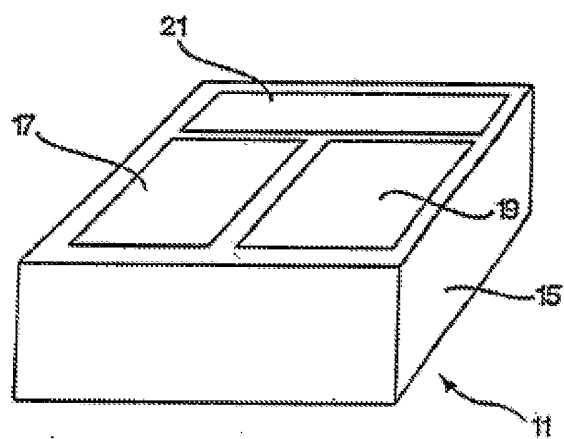


Fig. 9