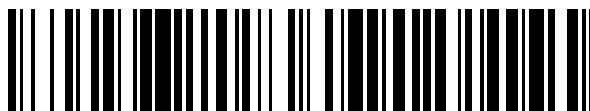


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 566 641**

51 Int. Cl.:

C12N 9/02 (2006.01)

C12N 15/52 (2006.01)

C12Q 1/68 (2006.01)

A61K 38/36 (2006.01)

A61K 38/44 (2006.01)

A61K 31/7088 (2006.01)

G01N 33/68 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **10.02.2004 E 10182644 (4)**

97 Fecha y número de publicación de la concesión europea: **06.01.2016 EP 2325302**

54 Título: **Células que co-expresan una enzima que genera a una sulfatasa y una C-formilglicina y sus métodos y usos**

30 Prioridad:

11.02.2003 US 447747 P

45 Fecha de publicación y mención en BOPI de la traducción de la patente:

14.04.2016

73 Titular/es:

**SHIRE HUMAN GENETIC THERAPIES, INC.
(100.0%)
300 Shire Way
Lexington MA 02421, US**

72 Inventor/es:

**VON FIGURA, KURT;
SCHMIDT, BERNHARD;
DIERKS, THOMAS;
HEARTLEIN, MICHAEL W;
BALLABIO, ANDREA y
COSMA, MARIA PIA**

74 Agente/Representante:

IZQUIERDO BLANCO, María Alicia

ES 2 566 641 T3

Aviso: En el plazo de nueve meses a contar desde la fecha de publicación en el Boletín europeo de patentes, de la mención de concesión de la patente europea, cualquier persona podrá oponerse ante la Oficina Europea de Patentes a la patente concedida. La oposición deberá formularse por escrito y estar motivada; sólo se considerará como formulada una vez que se haya realizado el pago de la tasa de oposición (art. 99.1 del Convenio sobre concesión de Patentes Europeas).

Células que co-expresan una enzima que genera a una sulfatasa y una C-formilglucina y sus métodos y usos**Descripción****5 Área del Invento**

Este invento se refiere a células que co-expresan una enzima que genera a una sulfatasa y una C-formilglucina y sus métodos y usos.

10 Antecedentes del Invento

Las sulfatasas son miembros de una familia genética altamente conservada, que comparte una extensa homología secuencial (Franco, B., et al., Cell (Célula), 1995, 81:15-25; Parenti, G., et al, Curr. Opin. Gen. Dev., 1997, 7:386-391), un alto nivel de similitud estructural (Bond, C.S., et al., Structure (Estructura), 1997, 5:277-289; Lukatela, G., et al., Biochemistry (Bioquímica), 1998, 37:3654-64), y una modificación post-interpretación única que es esencial para la división de ésteres de sulfatos (Schmidt, B., et al, Cell, 1995, 82:271-278; Selmer, T., et al, Eur. J. Biochem., 1996, 238:341-345). La modificación post-interpretación involucra la oxidación de ciertos residuos conservados de la cisteína (en eucariotas) o de la serina (en ciertas procariotas), a C β , que genera de L-C α -formilglucina (también conocida como *FGly*; *ácido 2-amino-3-oxopropanóico*) en el cual un grupo aldehído reemplaza al grupo tiometilo de la cadena lateral. El aldehído es una parte esencial del lado catalítico de las sulfatasas y asimismo actúa como un hidrato de aldehído. Uno de los grupos germinales hidroxilos acepta es un sulfato durante la división del éster del sulfato conllevando a la formación de un intermediador enzimático covalentemente sulfatado. El otro hidroxilo es requerido para la eliminación subsiguiente del sulfato y la regeneración del grupo aldehído. Esta modificación ocurre en el retículo endoplásmico mientras se realiza, o un poquito después, de la importación del polipéptido de la sulfatasa emergente y desorientada por una secuencia lineal corta que rodea al residuo de cisteína (o de serina) para que sea modificado. Esta secuencia altamente conservada es el hexapéptido L/V-C(S)-X-P-S-R (IDENTIFICACIÓN SECUENCIAL NÚMERO: 32), presente en la región de la terminal N de todas las sulfatasas eucariotas y los portadores más frecuentes llevan a un grupo hidroxilo o tiol en el residuo X (Dierks, T., et al, Proc. Natl. Acad. Sci. EE.UU., 1997, 94:11963-11968).

WO 01/60991 presenta quinasas humanas (PKIN) y polinucleótidos humanos que identifican y codifican a los PKINs. Además cubre vectores de expresión, células anfitrionas, anticuerpos, agonistas y antagonistas adecuados, así como métodos para diagnosticar, tratar o prevenir enfermedades asociadas con la expresión aberrante de los PKINs.

Dierks et al ([EMBO Journal (Revista EMBO), 18, 1, 2084-2091(1999)] explica que las sulfatasas portan en su lugar catalítico una modificación post-interpretación única, un residuo de alfa-formilglucina que es esencial para la actividad enzimática. La formación de la arilsulfatasa A fue estudiada in vitro y se descubrió que esta era dirigida por una secuencia lineal corta, CTPSR, empezando con la modificación de la cisteína. Análisis de mutaciones demostraron que la cisteína, la prolina y la arginina son residuos claves dentro de este motivo, mientras que la formación de formilglucina toleró una sustitución individual, pero no una sustitución simultánea de la treonina o de la serina. El motivo CTPSR fue transferido a una proteína heteróloga conllevando a la formación de una formilglucina de baja eficiencia. Se dice que la eficiencia alcanzó los valores de control cuando 7 residuos adicionales (AALLTGR) que seguían directamente al motivo CTPSR en la arilsulfatasa A están presentes. La mutación de hasta cuatro residuos simultáneamente dentro de esta secuencia de heptámeros quiere decir que se ha inhibido la modificación solo moderadamente. Se concluyó que el AALLTGR podría tener una función auxiliar de presentar al motivo esencial a la enzima modificadora y que dentro de los 2 motivos, los residuos clave están completamente conservados, y otros residuos están altamente conservados entre todos los miembros conocidos de la familia sulfatasa.

A la fecha 13 genes de sulfatasas han sido identificados en los humanos. Ellos codifican a enzimas con una especificidad y una localización sub-celular de sustratos diferentes tales como los lisosomas, Golgi y ER. 4 de estos genes, ARSC, ARSD, ARSE, y ARSF, que codifican a las arilsulfatasas C, D, E y F, respectivamente, están ubicados dentro de la región cromosómica (Xp22.3). Estas comparten una similitud secuencial significativa y una organización genómica que es casi idéntica, indicando que surgieron de eventos de duplicación que ocurrieron recientemente durante la evolución (Franco B, et al, Cell (Célula), 1995, 81:15-25; Meroni G, et al., Hum Mol Genet, 1996, 5:423-31).

La importancia de las sulfatasas en el metabolismo humano es resaltado de la identificación de por lo menos 8 enfermedades monogénicas humanas causadas por la deficiencia de las actividades individuales de las sulfatasas. La mayoría de estas condiciones son enfermedades de almacenamiento de lisosomas en los cuales las consecuencias fenotípicas se derivan del tipo y de la distribución de los tejidos del material almacenado. Entre estas existen 5 tipos diferentes de mucopolisacaridosis (MPS tipos II, IIIA, IIID, IVA, y VI) debido a deficiencias de las sulfatasas que actúan en el catabolismo de los years de glicosaminoglicanos (Neufeld y Muenzer, 2001, The mucopolysaccharidoses, In The Metabolic and Molecular Bases of Inherited Disease (las mucopolisacaridosis, en las bases metabólicas y moleculares de las enfermedades hereditarias) CR. Scriver, A.L. Beaudet, W.S. Sly, D. Nalle, B. Childs, K.W. Kinzler y B. Nogelstein, eds. Nueva York: Mc Graw-Hill, pp. 3421-3452), y la leucodistrofia

metacromática (MLD - metachromatic leukodystrophy), la cual es caracterizada por el almacenamiento de sulfolípidos en el sistema nervioso central y periférico conllevando a una determinación neurológica severa y progresiva. 2 enfermedades humanas adicionales son causadas por deficiencias de sulfatasas de lisosomas. Éstas incluyen ictiosis ligada al cromosoma X, una enfermedad de la piel debido a una deficiencia de esteroides (STS/ARSC), condrodisplasia punctata, una enfermedad que afecta a los huesos y a los cartílagos debido a una deficiencia de la arilsulfatasa E (ARSE). Las sulfatasas también están implicadas en los síndromes de malformaciones humanas inducidas por las drogas, tales como la embriopatía de Warfarina, causada por la inhibición de la actividad ARSE debido a una exposición en el útero a la warfarina durante el embarazo.

En una enfermedad monogénica humana intrigante, la deficiencia de sulfatasas múltiples (MSD - multiple sulfatase deficiency), todas las actividades de las sulfatasas son defectivas simultáneamente. Consecuentemente, el fenotipo de esta enfermedad multisistémica severa combina las características observadas en deficiencias individuales de las sulfatasas. Las células de los pacientes con MSD tienen deficiencias de las actividades de las sulfatasas aún después de su transfección con ADNcs que codifican a sulfatasas humanas, sugiriendo la presencia de un mecanismo común requerido para la actividad de todas las sulfatasas (Rommerskirch y von Figura, Proc. Natl. Acad. Sci., Estados Unidos, 1992, 89:2561-2565). La modificación post-interpretación de las sulfatasas demostró ser deficiente en un paciente con MSD, sugiriendo que esta enfermedad es causada por una mutación en un gen o genes, implicados en la maquinaria de la conversión de la cisteína a formilglicina (Schmidt, B., et al., Cell (Célula), 1995, 82:271-278). A pesar del intenso interés médico y biológico, esfuerzos dirigidos a la identificación de este gen o genes han sido obstaculizados por la rareza de los pacientes de MSD y la consecuente falta de casos familiares adecuados para realizar mapas genéticos.

Resumen del invento

Este invento suministra una célula de acuerdo a las reivindicaciones 1 a la 15, un método de acuerdo a cualquiera de las reivindicaciones 1 a la 23, o un uso médico de acuerdo a las reivindicaciones 24 o 25.

Hemos identificado un gen que codifica a la enzima generadora de Formilglicina (FGE - Formylglycine Generating Enzyme), una enzima responsable de la única modificación post - interpretación que ocurre en las sulfatasas la cual es esencial para la función de las sulfatasas (la formación de la L-Cα-formilglicina; también conocida como FGly y/o ácido 2-amino-3-oxopropanóico). Se ha descubierto, inesperadamente, que mutaciones en el gen FGE conllevan al desarrollo de una Deficiencia de Sulfatasas Múltiples (MSD - Multiple Sulfatase Deficiency) en sujetos. También se ha descubierto, inesperadamente, que la FGE mejora la actividad de las sulfatasas, incluyendo, pero sin limitarse a, la 2-Sulfatasa de Iduronato, la Sulfamidasa, la 6-Sulfatasa de N-Acetilgalactosamina, la 6-Sulfatasa de N-Acetilglucosamina, la Arilsulfatasa A, la Arilsulfatasa B, la Arilsulfatasa C, la Arilsulfatasa D, la Arilsulfatasa E, la Arilsulfatasa F, la Arilsulfatasa G, la HSulf-1, la HSulf-2, la HSulf-3, la HSulf-4, la HSulf-5, y la HSulf-6. En vista de estos descubrimientos, las moléculas de este invento pueden ser utilizadas para el diagnóstico y el tratamiento de la deficiencia de sulfatasas múltiples así como otras deficiencias de sulfatasas.

De acuerdo con el aspecto del invento, una célula que produce sulfatasas en la cual la tasa de sulfatasas activas en comparación con las sulfatasas totales producidas por la célula se incrementa, tal como se estableció en la reivindicación 1. La célula podría incluir: (i) una sulfatasa con una expresión incrementada, y (ii) una enzima generadora de Formilglicina con una expresión incrementada, donde la tasa de las sulfatasas activas en comparación de las sulfatasas totales (es decir, la actividad específica de las sulfatasas) producidas por la célula se incrementa en por lo menos un 5% sobre la tasa de las sulfatasas activas en comparación de las sulfatasas producidas por la célula en la ausencia de la enzima generadora de Formilglicina. En ciertas secciones, la tasa de las sulfatasas activas en comparación a las sulfatasas totales producidas por la célula se incrementa en por lo menos el 10%, el 15%, el 20%, el 50%, el 100%, el 200%, el 500%, el 1000%, sobre la tasa de sulfatasas activas en comparación a las sulfatasas totales producidas por la célula en la ausencia de la enzima generadora de Formilglicina.

Breve descripción de las secuencias

La IDENTIFICACIÓN SECUENCIAL NÚMERO: 1 es la secuencia de nucleótidos de la ADNc de la FGE humana.

La IDENTIFICACIÓN SECUENCIAL NÚMERO 2 es la secuencia pronosticada de aminoácidos producto de la interpretación de la ADNc de la FGE humana (IDENTIFICACIÓN SECUENCIAL NÚMERO: 1).

La IDENTIFICACIÓN SECUENCIAL NÚMERO: 3 es la secuencia de nucleótidos de la ADNc de la FGE humana que codifica al polipéptido de la IDENTIFICACIÓN SECUENCIAL NÚMERO: 2 (es decir, los nucleótidos 20-1141 de la IDENTIFICACIÓN SECUENCIAL NÚMERO: 1).

La IDENTIFICACIÓN SECUENCIAL NÚMERO: 4 es la secuencia de nucleótidos de la cuenta número AK075459 del GenBank (Banco Genético).

La IDENTIFICACIÓN SECUENCIAL NÚMERO: 5 es la secuencia pronosticada de aminoácidos producto de la interpretación de la IDENTIFICACIÓN SECUENCIAL NÚMERO: 4, un producto proteínico que no tiene nombre que tiene la cuenta número BAC11634 del GenBank (Banco Genético).

La IDENTIFICACIÓN SECUENCIAL NÚMERO: 6 es la secuencia de nucleótidos de la ADNc de la 2-sulfatasa de Iduronato (cuenta del GenBank (Banco Genético) número M58342).

La IDENTIFICACIÓN SECUENCIAL NÚMERO: 7 es la secuencia pronosticada de aminoácidos producto de la interpretación de la ADNc de la 2-sulfatasa de Iduronato (la IDENTIFICACIÓN SECUENCIAL NÚMERO: 6).

La IDENTIFICACIÓN SECUENCIAL NÚMERO: 8 es la secuencia de nucleótidos de la ADNc de la Sulfamidasa humana (cuenta del GenBank (Banco Genético) número U30894).

5 La IDENTIFICACIÓN SECUENCIAL NÚMERO: 9 es la secuencia pronosticada de aminoácidos producto de interpretación de la ADNc de Sulfamidases humanas (la IDENTIFICACIÓN SECUENCIAL NÚMERO: 8).

La IDENTIFICACIÓN SECUENCIAL NÚMERO: 10 es la secuencia de nucleótidos de la ADNc de la 6-sulfatasa de N-Acetilgalactosamina humana (cuenta del GenBank (Banco Genético) número U06088).

10 La IDENTIFICACIÓN SECUENCIAL NÚMERO: 11 es la secuencia pronosticada aminoácidos producto de la interpretación de la ADNc de la 6-sulfatasa de N-Acetilgalactosamina humana (la IDENTIFICACIÓN SECUENCIAL NÚMERO: 10).

La IDENTIFICACIÓN SECUENCIAL NÚMERO: 12 es la secuencia de nucleótidos de la ADNc de la 6-sulfatasa de la N-Acetilglucosamina humana (cuenta del GenBank (Banco Genético) número Z12173).

15 La IDENTIFICACIÓN SECUENCIAL NÚMERO: 13 es la secuencia pronosticada de aminoácidos producto de la interpretación de la ADNc de la 6-sulfatasa de N-Acetilglucosamina humana (la IDENTIFICACIÓN SECUENCIAL NÚMERO: 12).

La IDENTIFICACIÓN SECUENCIAL NÚMERO: 14 es la secuencia de nucleótidos de la ADNc de arilsulfatasa a humana (la cuenta del GenBank (Banco Genético) número X52151).

20 La IDENTIFICACIÓN SECUENCIAL NÚMERO: 15 es la secuencia pronosticada de aminoácidos producto de la interpretación de la ADNc de la arilsulfatasa A humana (la IDENTIFICACIÓN SECUENCIAL NÚMERO: 14).

La IDENTIFICACIÓN SECUENCIAL NÚMERO: 16 es la secuencia de nucleótidos de la ADNc de la arilsulfatasa B humana (la cuenta del GenBank (Banco Genético) número J05225).

La IDENTIFICACIÓN SECUENCIAL NÚMERO: 17 es la secuencia pronosticada de aminoácidos producto de la interpretación de la ADNc de la arilsulfatasa B humana (la IDENTIFICACIÓN SECUENCIAL NÚMERO: 16).

25 La IDENTIFICACIÓN SECUENCIAL NÚMERO: 18 es la secuencia de nucleótidos de la ADNc de la arilsulfatasa C humana (la cuenta del GenBank (Banco Genético) número J04964).

La IDENTIFICACIÓN SECUENCIAL NÚMERO: 19 es la secuencia pronosticada de aminoácidos producto de la interpretación de la ADNc de arilsulfatasa C humana (la IDENTIFICACIÓN SECUENCIAL NÚMERO: 18).

30 La IDENTIFICACIÓN SECUENCIAL NÚMERO: 20 es la secuencia de nucleótidos de la ADNc de la arilsulfatasa D humana (cuenta del GenBank (Banco Genético) número X83572).

La IDENTIFICACIÓN SECUENCIAL NÚMERO: 21 es la secuencia pronosticada de aminoácidos producto de la interpretación de la ADNc de la arilsulfatasa D humana (la IDENTIFICACIÓN SECUENCIAL NÚMERO: 20).

La IDENTIFICACIÓN SECUENCIAL NÚMERO: 22 es la secuencia de nucleótidos de la ADNc de la arilsulfatasa E humana (la cuenta del GenBank (Banco Genético) número X83573).

35 La IDENTIFICACIÓN SECUENCIAL NÚMERO: 23 es la secuencia pronosticada de aminoácidos producto de la interpretación de la ADNc de arilsulfatasa E humana (la IDENTIFICACIÓN SECUENCIAL NÚMERO: 22).

La IDENTIFICACIÓN SECUENCIAL NÚMERO: 24 es la secuencia de nucleótidos de la ADNc de la arilsulfatasa F humana (la cuenta del GenBank (Banco Genético) número X97868).

40 La IDENTIFICACIÓN SECUENCIAL NÚMERO: 25 es la secuencia pronosticada de aminoácidos producto de la interpretación de la ADNc de la arilsulfatasa F humana (la IDENTIFICACIÓN SECUENCIAL NÚMERO: 24).

La IDENTIFICACIÓN SECUENCIAL NÚMERO: 26 de la secuencia de nucleótidos de la ADNc de la arilsulfatasa G humana (la cuenta del GenBank (Banco Genético) número BC012375).

45 La IDENTIFICACIÓN SECUENCIAL NÚMERO: 27 es la secuencia pronosticada de aminoácidos producto de la interpretación de la arilsulfatasa G humana (IDENTIFICACIÓN SECUENCIAL NÚMERO: 26).

La IDENTIFICACIÓN SECUENCIAL NÚMERO: 28 es la secuencia de nucleótidos de la ADNc de la HSulf-1 (cuenta del GenBank (Banco Genético) AY101175).

La IDENTIFICACIÓN SECUENCIAL NÚMERO: 29 es la secuencia pronosticada de aminoácidos producto de la interpretación de la ADNc de la HSulf-1 (la IDENTIFICACIÓN SECUENCIAL NÚMERO: 28).

50 La IDENTIFICACIÓN SECUENCIAL NÚMERO: 30 es la secuencia de nucleótidos de la ADNc de la HSulf-2 (la cuenta del GenBank (Banco Genético) número AY101176).

La IDENTIFICACIÓN SECUENCIAL NÚMERO: 31 es la secuencia pronosticada de aminoácidos producto de la interpretación de la ADNc de la HSulf-2 (IDENTIFICACIÓN SECUENCIAL NÚMERO: 30).

La IDENTIFICACIÓN SECUENCIAL NÚMERO: 32 es el hexapéptido altamente conservado L/N-FGly-X-P-S-R presente en las sulfatasas.

55 La IDENTIFICACIÓN SECUENCIAL NÚMERO: 33 es un sustrato sintético de la formación de FGly; su secuencia primaria es derivada de la arilsulfatasa A humana.

La IDENTIFICACIÓN SECUENCIAL NÚMERO: 34 es un oligopéptido mezclado PVSLPTRSCAALLTGR.

La IDENTIFICACIÓN SECUENCIAL NÚMERO: 35 es un oligopéptido de Ser69 PVSLSTPSRAALLTGR.

La IDENTIFICACIÓN SECUENCIAL NÚMERO: 36 es un cebador 1199nc específico para FGEs humanas.

60 La IDENTIFICACIÓN SECUENCIAL NÚMERO: 37 es un cebador hacia adelante 1c específico para FGEs humanos.

La IDENTIFICACIÓN SECUENCIAL NÚMERO: 38 es un cebador en reversa 1182c específico para FGEs humanos.

65 La IDENTIFICACIÓN SECUENCIAL NÚMERO: 39 es un sitio EcoRI que contiene un cebador específico de la FGEs humanos de 5'.

La IDENTIFICACIÓN SECUENCIAL NÚMERO: 40 es un cebador específico de HAS.

La IDENTIFICACIÓN SECUENCIAL NÚMERO: 41 es un cebador específico de c-myc.
 La IDENTIFICACIÓN SECUENCIAL NÚMERO: 42 es un cebador específico de RGS-His6.
 La IDENTIFICACIÓN SECUENCIAL NÚMERO: 43 es un oligopéptido tríptico SQNTPDSSASNLGFR
 proveniente de una preparación de la FGEs humanas.
 5 La IDENTIFICACIÓN SECUENCIAL NÚMERO: 44 es un oligopéptido tríptico MVPIPAGVFTMGTDPPQIK
 proveniente de una preparación de la FGEs humanas.
 La IDENTIFICACIÓN SECUENCIAL NÚMERO: 45 es la secuencia de nucleótidos del parólogo de la FGE2
 humano (GI del GenBank (Banco Genético): 24308053).
 La IDENTIFICACIÓN SECUENCIAL NÚMERO: 46 es la secuencia pronosticada de aminoácidos producto de
 10 la interpretación del parólogo de la FGE2 humano (IDENTIFICACIÓN SECUENCIAL NÚMERO: 45).
 La IDENTIFICACIÓN SECUENCIAL NÚMERO: 47 es la secuencia de nucleótidos del parólogo de la FGE de
 ratón (GI del GenBank (Banco Genético): 26344956).
 La IDENTIFICACIÓN SECUENCIAL NÚMERO: 48 es la secuencia pronosticada de aminoácidos producto de
 la interpretación del parólogo de la FGE de ratón (IDENTIFICACIÓN SECUENCIAL NÚMERO: 47).
 15 La IDENTIFICACIÓN SECUENCIAL NÚMERO: 49 es la secuencia de nucleótidos del ortólogo de la FGE de
 ratón (GI del GenBank (Banco Genético): 22122361).
 La IDENTIFICACIÓN SECUENCIAL NÚMERO: 50 es la secuencia pronosticada de aminoácidos producto de
 la interpretación del ortólogo de la FGEs de ratón (la IDENTIFICACIÓN SECUENCIAL NÚMERO: 49).
 La IDENTIFICACIÓN SECUENCIAL NÚMERO: 51 es la secuencia de nucleótidos del ortólogo de la FGE de
 20 mosca de fruta (GI del GenBank (Banco Genético): 20130397).
 La IDENTIFICACIÓN SECUENCIAL NÚMERO: 52 es la secuencia pronosticada de aminoácidos del producto
 de interpretación del ortólogo de la FGE de la mosca de frutas (IDENTIFICACIÓN SECUENCIAL NÚMERO:
 51).
 La IDENTIFICACIÓN SECUENCIAL NÚMERO: 53 es la secuencia de nucleótidos del ortólogo de la FGE del
 25 mosquito (GI del GenBank (Banco Genético): 21289310).
 La IDENTIFICACIÓN SECUENCIAL NÚMERO: 54 es la secuencia pronosticada de aminoácidos producto de
 la interpretación del ortólogo de las FGEs de mosquito (la IDENTIFICACIÓN SECUENCIAL NÚMERO: 53).
 La IDENTIFICACIÓN SECUENCIAL NÚMERO: 55 de la secuencia de nucleótidos del ortólogo de la FGEs
 del *S. coelicolor* que es muy relacionado (GI del GenBank (Banco Genético): 21225812).
 30 La IDENTIFICACIÓN SECUENCIAL NÚMERO: 56 es la secuencia pronosticada de aminoácidos producto de
 la interpretación del ortólogo de la FGEs del *S. coelicolor* (IDENTIFICACIÓN SECUENCIAL NÚMERO: 55).
 La IDENTIFICACIÓN SECUENCIAL NÚMERO: 57 es la secuencia de nucleótidos del ortólogo de la FGEs
 relacionado de cerca con el *C. efficiens* (GI del GenBank (Banco Genético): 25028125).
 La IDENTIFICACIÓN SECUENCIAL NÚMERO: 58 es la secuencia pronosticada de aminoácidos producto de
 35 la interpretación del ortólogo de la FGEs del *C. efficiens* (IDENTIFICACIÓN SECUENCIAL NÚMERO: 57).
 La IDENTIFICACIÓN SECUENCIAL NÚMERO: 59 es la secuencia de nucleótidos del ortólogo de la FGE del
N. aromaticivorans (GI del GenBank (Banco Genético): 23108562).
 La IDENTIFICACIÓN SECUENCIAL NÚMERO: 60 es la secuencia pronosticada de aminoácidos producto de
 la interpretación del ortólogo de la FGE del *N. aromaticivorans* (IDENTIFICACIÓN SECUENCIAL NÚMERO:
 40 59).
 La IDENTIFICACIÓN SECUENCIAL NÚMERO: 61 es la secuencia de nucleótidos del ortólogo de la FGE del
M. loti (GI del GenBank (Banco Genético): 13474559).
 La IDENTIFICACIÓN SECUENCIAL NÚMERO: 62 es la secuencia pronosticada de aminoácidos producto de
 la interpretación del ortólogo de la FGEs del *M. Loti* (IDENTIFICACIÓN SECUENCIAL NÚMERO: 61).
 45 La IDENTIFICACIÓN SECUENCIAL NÚMERO: 63 es la secuencia de nucleótidos del ortólogo de la FGE del
B. fungorum (GI del GenBank (Banco Genético): 22988809).
 La IDENTIFICACIÓN SECUENCIAL NÚMERO: 64 es la secuencia pronosticada de aminoácidos producto de
 la interpretación del ortólogo de la FGE del *B. fungorum* (la IDENTIFICACIÓN SECUENCIAL NÚMERO: 63).
 La IDENTIFICACIÓN SECUENCIAL NÚMERO: 65 es la secuencia de nucleótidos del ortólogo de la FGE del
 50 *S. meliloti* (GI del GenBank (Banco Genético): 16264068).
 La IDENTIFICACIÓN SECUENCIAL NÚMERO: 66 es la secuencia pronosticada de aminoácidos producto de
 la interpretación del ortólogo de la FGE del *S. meliloti* (IDENTIFICACIÓN SECUENCIAL NÚMERO: 65).
 La IDENTIFICACIÓN SECUENCIAL NÚMERO: 67 es la secuencia de nucleótidos del ortólogo de la FGE de
 la *Microscilla* sp. (GI del GenBank (Banco Genético): 14518334).
 55 La IDENTIFICACIÓN SECUENCIAL NÚMERO: 68 es la secuencia pronosticada de aminoácidos producto de
 la interpretación del ortólogo de la FGE de la *Microscilla* sp. (IDENTIFICACIÓN SECUENCIAL NÚMERO: 67).
 La IDENTIFICACIÓN SECUENCIAL NÚMERO: 69 es la secuencia de nucleótidos del ortólogo de la FGE del
P. putida KT2440 (GI del GenBank (Banco Genético): 26990068).
 La IDENTIFICACIÓN SECUENCIAL NÚMERO: 70 es la secuencia pronosticada de aminoácidos producto de
 60 la interpretación del ortólogo de la FGE de la *P. putida* KT2440 (IDENTIFICACIÓN SECUENCIAL NÚMERO:
 69).
 La IDENTIFICACIÓN SECUENCIAL NÚMERO: 71 es la secuencia de nucleótidos del ortólogo de la FGE del
R. metallidurans (GI del GenBank (Banco Genético): 22975289).
 La IDENTIFICACIÓN SECUENCIAL NÚMERO: 72 es la secuencia pronosticada de aminoácidos producto de
 65 la interpretación del ortólogo de la FGE del *R. metallidurans* (la IDENTIFICACIÓN SECUENCIAL NÚMERO:
 71).

La IDENTIFICACIÓN SECUENCIAL NÚMERO: 73 es la secuencia de nucleótidos del ortólogo de la FGE del *P. marinus* (GI del GenBank (Banco Genético): 23132010).

La IDENTIFICACIÓN SECUENCIAL NÚMERO: 74 es la secuencia pronosticada de aminoácidos producto de la interpretación del ortólogo de la FGE del *P. marinus* (IDENTIFICACIÓN SECUENCIAL NÚMERO: 73).

La IDENTIFICACIÓN SECUENCIAL NÚMERO: 75 es la secuencia de nucleótidos del ortólogo de la FGE del *C. crescentus* CB15 (GI del GenBank (Banco Genético): 16125425).

La IDENTIFICACIÓN SECUENCIAL NÚMERO: 76 es la secuencia pronosticada de aminoácidos producto de la interpretación del ortólogo de la FGE del *C. crescentus* CB15 (IDENTIFICACIÓN SECUENCIAL NÚMERO: 75).

La IDENTIFICACIÓN SECUENCIAL NÚMERO: 77 es la secuencia de nucleótidos del ortólogo de la FGE de la *M. tuberculosis* Ht37Rv (GI del GenBank (Banco Genético): 15607852).

La IDENTIFICACIÓN SECUENCIAL NÚMERO: 78 es la secuencia pronosticada de aminoácidos producto de la interpretación del ortólogo de la FGE de la *M. tuberculosis* Ht37Rv (IDENTIFICACIÓN SECUENCIAL NÚMERO: 77).

La IDENTIFICACIÓN SECUENCIAL NÚMERO: 79 es el heptapéptido altamente conservado que está presente en el sub-dominio 3 de los ortólogos y parálogos de la FGE.

La IDENTIFICACIÓN SECUENCIAL NÚMERO: 80 es la secuencia de nucleótidos del fragmento de EST del ortólogo de la FGE que tiene la cuenta del GenBank (Banco Genético) número: CA379852.

La IDENTIFICACIÓN SECUENCIAL NÚMERO: 81 es la secuencia de nucleótidos del fragmento de EST del ortólogo de la FGE que tiene la cuenta del GenBank (Banco Genético) número: AI721440.

La IDENTIFICACIÓN SECUENCIAL NÚMERO: 82 es la secuencia de nucleótidos del fragmento de EST del ortólogo de la FGE que tiene la cuenta del GenBank (Banco Genético) número: BJ505402.

La IDENTIFICACIÓN SECUENCIAL NÚMERO: 83 es la secuencia de nucleótidos del fragmento de EST del ortólogo de la FGE que tiene la cuenta del GenBank (Banco Genético) número: BJ054666.

La IDENTIFICACIÓN SECUENCIAL NÚMERO: 84 es la secuencia de nucleótidos del fragmento de EST del ortólogo de la FGE que tiene la cuenta del GenBank (Banco Genético) número: AL892419.

La IDENTIFICACIÓN SECUENCIAL NÚMERO: 85 es la secuencia de nucleótidos del fragmento de EST del ortólogo de la FGE que tiene la cuenta del GenBank (Banco Genético) número: CA064079.

La IDENTIFICACIÓN SECUENCIAL NÚMERO: 86 es la secuencia de nucleótidos del fragmento de EST del ortólogo de la FGE que tiene la cuenta del GenBank (Banco Genético) número: BF189614.

La IDENTIFICACIÓN SECUENCIAL NÚMERO: 87 es la secuencia de nucleótidos del fragmento de EST del ortólogo de la FGE que tiene la cuenta del GenBank (Banco Genético) número: AV609121.

La IDENTIFICACIÓN SECUENCIAL NÚMERO: 88 es la secuencia de nucleótidos de la ADNc de la HSulf-3.

La IDENTIFICACIÓN SECUENCIAL NÚMERO: 89 es la secuencia pronosticada de aminoácidos producto de la interpretación de la ADNc de la HSulf-3 (la IDENTIFICACIÓN SECUENCIAL NÚMERO: 88).

La IDENTIFICACIÓN SECUENCIAL NÚMERO: 90 es la secuencia de nucleótidos de la ADNc de la HSulf-4.

La IDENTIFICACIÓN SECUENCIAL NÚMERO: 91 es la secuencia pronosticada de aminoácidos producto de la interpretación de la ADNc de la HSulf-4 (IDENTIFICACIÓN SECUENCIAL NÚMERO: 90).

La IDENTIFICACIÓN SECUENCIAL NÚMERO: 92 es la secuencia de nucleótidos de la ADNc de la HSulf-5.

La IDENTIFICACIÓN SECUENCIAL NÚMERO: 93 es la secuencia pronosticada de aminoácidos producto de la interpretación de la ADNc de la HSulf-5 (IDENTIFICACIÓN SECUENCIAL NÚMERO: 92).

La IDENTIFICACIÓN SECUENCIAL NÚMERO: 94 es la secuencia de nucleótidos de la ADNc de la HSulf-6.

La IDENTIFICACIÓN SECUENCIAL NÚMERO: 95 es la secuencia pronosticada de aminoácidos producto de la interpretación de la ADNc de la HSulf-6 (IDENTIFICACIÓN SECUENCIAL NÚMERO: 94).

Descripción breve de los esquemas

Figura 1: Un espectro esquemático de masa MALDI-TOF de P23 después de la incubación en la ausencia de (A) o la presencia de (B) de un extracto soluble proveniente de miocitos de testículos bovinos.

Figura 2: Un árbol filogenético derivado de una alineación de la FGE humana y 21 proteínas de la semilla MALDI-TOF.

Figura 3: La organización del locus genético de la FGE humana y de ratón. Las secciones se muestran a escala como cuadros y cuadros brillantes (locus de ratones). Los números que están sobre las líneas de los intrones indican el tamaño de los intrones en las kilobases.

Figura 4: Un diagrama que muestra un mapa del plásmido pXMG.1.3 y su expresión de la FGE.

Figura 5: Un gráfico de barras que muestra la actividad de la 6-sulfatasa de N-Acetilgalactosamina en las células 36F transfectadas transitoriamente con un plásmido con expresión de las FGEs.

Figura 6: Un gráfico de barras que muestra la actividad específica de la 6-sulfatasa de N-Acetilgalactosamina en las células 36F transfectadas transitoriamente con el plásmido con expresión de la FGE.

Figura 7: Un gráfico de barras que muestra la producción de la 6-sulfatasa de N-Acetilgalactosamina en las células 36F transfectadas transitoriamente con un plásmido con expresión de la FGE.

Figura 8: Un gráfico que muestra la actividad de la 2-sulfatasa de Iduronato en las células 30C6 transfectadas transitoriamente con un plásmido con expresión de FGEs.

Figura 9: Muestra un botiquín que tiene las características de este invento.

Descripción detallada del invento

El invento se establece en las reivindicaciones 1 a la 25. Se basa en el descubrimiento de un gen que codifica a la enzima generadora de Formilglicina (FGE – Formylglycine Generating Enzyme), una enzima responsable de la única modificación post-interpretación que ocurre en las sulfatasas la cual es esencial para la función de la sulfatasas: la formación de la L-C α -formilglicina (también conocida como FGly y/o el ácido 2-amino-3-oxopropanóico). Se ha descubierto, inesperadamente, que las mutaciones del gen FGE conllevan al desarrollo de una Deficiencia de Sulfatasas Múltiples (MSD - Multiple Sulfatase Deficiency) en sujetos. También se ha descubierto, inesperadamente, que la FGE mejora la actividad de las sulfatasas, incluyendo, pero sin limitarse a, la 2-sulfatasa de Iduronato, la sulfamidasa, la 6-sulfatasa de N-Acetilgalactosamina, la 6-sulfatasa de N-Acetilgalactosamina, la arilsulfatasa A, la arilsulfatasa B, la arilsulfatasa C, la arilsulfatasa D, la arilsulfatasa E, la arilsulfatasa F, la arilsulfatasa G, la HSulf-1, la HSulf-2, la HSulf-3, la HSulf-4, la HSulf-5, y la HSulf-6, y las sulfatasas descritas en las aplicaciones provisionales de Estados Unidos con los números de publicación 20030073118, 20030147875, 20030148920, 20030162279, y 20030166283.

La “actividad generadora de C α -formilglicina” se refiere a la capacidad de una molécula para formar, o mejorar la formación de, FGly en un sustrato. El sustrato podría ser una sulfatasa tal como se describe en otras secciones de este documento, o un oligopéptido sintético (refiérase, por ejemplo, a la IDENTIFICACIÓN SECUENCIAL NÚMERO: 33 y a los ejemplos). El sustrato contiene preferiblemente al hexapéptido conservado de la IDENTIFICACIÓN SECUENCIAL NÚMERO: 32 [L/V-C(S)-X-P-S-R]. Métodos para los ensayos de la formación de FGly son tal como se han descrito en la industria (refiérase, por ejemplo, a Dierks, T., et al., Proc. Natl. Acad. Sci. Estados Unidos de América, 1997, 94:11963-11968), y en otras secciones de este documento (refiérase, por ejemplo, a los ejemplos). Una “molécula”, tal como se utiliza este documento, abarca a los “ácidos nucleicos” y a los “polipéptidos”. Las moléculas FGE son capaces de formar, o mejorar/incrementar la formación de, FGly in vivo e in vitro.

“Mejorar (o “incrementar”)” la actividad generadora de C α -formilglicina, tal como se utiliza en este documento, se refiere típicamente a la expresión incrementada de la FGE y/o su polipéptido codificado. Una “expresión incrementada” se refiere a incrementar (es decir, a una medida detectable) la replicación, transcripción y/o interpretación de cualquiera de los ácidos nucleicos del invento (ácidos nucleicos de la FGE tal como se describen en otras secciones de este documento), puesto que la regulación/incremento de cualquiera de estos procesos resulta en un incremento en la concentración/monto del polipéptido codificado por el gen (ácido nucleico). Mejorar (o incrementar) la actividad generadora de la C α -formilglicina también se refiere a la prevención o inhibición de la degradación de la FGE (por ejemplo, por medio de una ubiquitinación), una regulación/reducción, etcétera, que resulta, por ejemplo, en una molécula t_{1/2} de la FGE incrementada o estable (vida-media) cuando se compara a un control. La regulación-reducción o expresión reducida se refiere a una expresión reducida de un gen y/o su polipéptido codificado. La regulación-incremento o regulación-reducción de la expresión genética puede determinarse directamente al detectar un incremento o reducción, respectivamente, en el nivel de ARNm para el gen (por ejemplo, la FGE), o el nivel de expresión proteínica del polipéptido codificado por el gen, usando cualquier medio adecuado conocido en la industria, tal como la hibridación del ácido nucleico o métodos de detección de anticuerpos, respectivamente, y en comparación a los controles. La regulación-incremento o la regulación-reducción de la expresión genética de la FGE también puede determinarse y directamente al detectar un cambio en la actividad generadora de C α -formilglicina.

La “expresión”, tal como se utiliza en este documento, se refiere a la expresión del ácido nucleico y/o del polipéptido, así como a la actividad de la molécula del polipéptido (por ejemplo, la actividad generadora de C α -formilglicina de la molécula).

Tal como se utiliza en este documento, un sujeto es un mamífero o un mamífero no humano. En todas las secciones las FGE humanas y los sujetos humanos son preferidos.

Tal como se utiliza en este documento en relación a los ácidos nucleicos, el término “aislados” se refiere a: (i) amplificado in vitro, por ejemplo, por la reacción en cadena de polimerasas (PCR - polymerase chain reaction); (ii) producida recombinantemente por medio de clonación; (iii) purificada, como por medio de una división y de separación de gel; o (iv) sintetizada, por ejemplo, por medio de una síntesis química. Un ácido nucleico aislado es uno que es manipulado fácilmente por las técnicas de ADN recombinante bien conocidas en la industria. Por lo tanto, una secuencia de nucleótidos se considera aislada si está contenida en un vector en el cual los lugares de restricción 5' y 3' son conocidos o para los cuales las secuencias incitadoras de reacciones en cadena de polimerasas (PCR - polymerase chain reaction) se han presentado, pero una secuencia de ácidos nucleicos

existentes en su estado nativo en su anfitrión natural no se considera aislada. Un ácido nucleico aislado puede ser sustancialmente purificado, pero no necesita serlo. Por ejemplo, un ácido nucleico que es aislado sin un vector de clonación o de expresión no es puro puesto que podría comprender únicamente un porcentaje diminuto del material en la célula en la cual reside. Tal ácido nucleico es aislado, sin embargo, tal como se utiliza el término en este documento puede ser manipulado fácilmente por medio de técnicas estándar conocidas para aquellas personas con conocimiento en la industria.

Tal como se utiliza en este documento en relación a los polipéptidos, el término "aislados" significa que está separado de su entorno natural en una forma suficientemente pura para que pueda manipularse o usarse para cualquiera de los propósitos aquí descritos. Por lo tanto, aislado significa suficientemente puro como para ser utilizado (i) para facilitar y/o aislar anticuerpos, (ii) como un reactivo en un ensayo, (iii) para secuenciaciones, (iv) como un elemento terapéutico, etcétera.

Homólogos y alelos de los ácidos nucleicos de la FGEs aquí descritos que también tienen la actividad generadora de Cα-formilglicina pueden ser suministrados. Los homólogos, tal como se describen en este documento, incluyen las moléculas identificadas en otras secciones de este documento (refiérase, por ejemplo, a las IDENTIFICACIONES SECUENCIALES NÚMEROS: 4, 5, 45-78 y 80-87) es decir, ortólogos y parálogos. Además, los homólogos pueden ser identificados siguiendo las enseñanzas de este invento, así como por medio de técnicas convencionales. Puesto que todos los homólogos de la FGE descritos en este documento comparten la actividad generadora de Cα-formilglicina, estos pueden ser utilizados intercambiabilmente con la molécula de la FGE humana en todos los aspectos de este invento.

En general, los homólogos y los alelos comúnmente compartirán por lo menos el 40% de su identidad de nucleótidos y/o por lo menos el 50% de la identidad de aminoácidos de la IDENTIFICACIÓN SECUENCIAL NÚMERO: 1 y la IDENTIFICACIÓN SECUENCIAL NÚMERO: 2, respectivamente. En algunas instancias compartirán por lo menos el 50% de la identidad de nucleótidos y/o por lo menos el 65% de identidad de aminoácidos y aún más, en otras instancias compartirán por lo menos el 60% de la identidad de nucleótidos y/o por lo menos el 75% de la identidad de aminoácidos. En instancias adicionales, los homólogos y los alelos comúnmente compartirán por lo menos el 90% (el 95%, o incluso el 99%), de la identidad de nucleótidos y/o por lo menos el 95%, 9SΨo, o incluso el 99% de la identidad de aminoácidos de la IDENTIFICACIÓN SECUENCIAL NÚMERO: 1 y la IDENTIFICACIÓN SECUENCIAL NÚMERO: 2, respectivamente. La homología puede ser calculada utilizando varias herramientas de software que son disponibles públicamente desarrolladas por NCBI (Bethesda, Maryland). Las herramientas de ejemplo incluyen al algoritmo heurístico de Altschul SF, et al., (J Mol Biol, 1990, 215:403-410), también conocido como BLAST. Las alineaciones por parejas y por ClustalW (configuración de la matriz BLOSUM30) así como el análisis hidropático de Kyte-Doolittle puede ser obtenido utilizando softwares públicos de análisis secuencial (EMBL, Heidelberg, Alemania) y comerciales (por ejemplo, MacNector de Oxford Molecular Group/enetics Computer Group, Madison, WI). Los complementos de Watson-Crick de los ácidos nucleicos que se acaban de mencionar también pueden ser usados.

Para detectar los genes relacionados con la FGE, tal como los homólogos y los alelos de la FGE, una prueba Southern blot puede ser realizada utilizando las condiciones ya mencionadas, junto con una sonda radiactiva. Después de lavar la membrana a la cual el ADN se ha transferido finalmente, la membrana puede ser colocada en contra de una lámina de rayos X o una placa de fósforo fotoestimulable para detectar la señal radiactiva.

Dadas las enseñanzas en este documento de un clon de longitud completa de la ADNc de la FGE humana, otras frecuencias de mamíferos tales como el clon de la ADNc de ratón correspondiente al gen de la FGE humana puede aislarse de una biblioteca de ADNcs, utilizando técnicas estándar de hibridación de colonias.

Ácidos nucleicos degenerados que incluyen a codones alternos a aquellos presentes en los materiales nativos pueden ser suministrados. Por ejemplo, los residuos de serina son codificados por los codones TCA, AGT, TCC, TCG, TCT y AGC. Por lo tanto, será evidente para una persona con conocimiento normal en la industria que cualquiera de los tripletes de nucleótidos que codifican a la serina podría ser utilizado para dirigir el aparato de síntesis proteínica, in vitro o in vivo, para incorporar un residuo de serina a un polipéptido alargador de la FGE. Asimismo, los tripletes de secuencias de nucleótidos que codifican a otros residuos de aminoácidos incluyen, pero no se limitan a: CCA, CCC, CCG y CCT (los codones de la prolina); CGA, CGC, CGG, CGT, AGA y AGG (los codones de la arginina); ACA, ACC, ACG y ACT (los codones de la treonina); AAC y AAT (los codones de la asparagina); y ATA, ATC y ATT (los codones de la isoleucina). Otros residuos de aminoácidos podrían ser codificados similarmente por secuencias múltiples de nucleótidos. Por lo tanto, podrían facilitarse ácidos nucleicos degenerados que difieren de los ácidos nucleicos aislados biológicamente en las secuencias de codones debido a la degeneración de los códigos genéticos.

Pueden aislarse a los fragmentos únicos de la IDENTIFICACIÓN SECUENCIAL NÚMERO: 1 o de la IDENTIFICACIÓN SECUENCIAL NÚMERO: 3 o sus complementos. Un fragmento único es uno que es una 'firma' para un ácido nucleico más grande. Por ejemplo, el fragmento único es lo suficientemente largo como para asegurar que es una secuencia precisa y no se encuentra en las moléculas dentro del genoma humano fuera de los ácidos nucleicos de la FGE definidos anteriormente (y alelos humanos). Aquellas personas con un conocimiento normal en

la industria podrían aplicar procedimientos rutinarios para determinar si un fragmento es único dentro del genoma humano. Los fragmentos únicos, sin embargo, excluyen fragmentos que están compuestos completamente de las secuencias de los nucleótidos seleccionados de un grupo que consiste de la IDENTIFICACIÓN SECUENCIAL NÚMERO: 4 y/u otras secuencias publicadas previamente a la fecha de presentación de esta aplicación.

Un fragmento que es compuesto completamente de las secuencias descritas en los depósitos anteriores del GenBnk (Banco Genético) es uno que no incluye cualquiera de los nucleótidos únicos de las secuencias del invento. Por lo tanto, un fragmento único de acuerdo al invento debe contener una secuencia de nucleótidos que no sea la secuencia exacta de aquellos en los depósitos del GenBank (Banco Genético) o sus fragmentos. La diferencia puede darse en una adición, eliminación o sustitución con respecto a la secuencia del GenBank (Banco Genético) o podría ser una secuencia completamente separada de la secuencia del GenBank (Banco Genético).

Los fragmentos únicos pueden ser utilizados como sondas en los ensayos Southern y Western blots para identificar a aquellos ácidos nucleicos, o pueden ser utilizados en ensayos de amplificación tales como aquellos que utilizan al PCR. Tal como es conocido por aquellas personas con conocimiento en la industria, sondas grandes tales como de 200, 250, 300 o más nucleótidos son preferidas para ciertos usos tales como las pruebas de Southern y Northern blots, mientras que fragmentos más pequeños serán preferidos para usos tales como el PCR. Fragmentos únicos también pueden ser utilizados para producir proteínas de fusión para generar anticuerpos o determinar los enlaces de los fragmentos de polipéptidos, tal como se demostró en los ejemplos, o para generar componentes de inmunoensayos. Asimismo, fragmentos únicos pueden ser utilizados para producir fragmentos no fusionados de los polipéptidos de la FGE, útiles, por ejemplo, para la preparación de anticuerpos, inmunoensayos o aplicaciones terapéuticas. Fragmentos únicos pueden ser utilizados además como moléculas antisentido para inhibir la expresión de ácidos nucleicos y polipéptidos de la FGE respectivamente.

Tal como será reconocido por aquellas personas con conocimiento en la industria, el tamaño de los fragmentos únicos dependerá de su conservación del código genético. Por lo tanto, algunas regiones de la IDENTIFICACIÓN SECUENCIAL NÚMERO: 1 o de la IDENTIFICACIÓN SECUENCIAL NÚMERO: 3 y de sus complementos requerirán que los segmentos más largos sean únicos mientras que otros sólo requerirán segmentos cortos, comúnmente, entre 12 a 32 nucleótidos de largo (por ejemplo, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31 y 32 bases) o más, hasta la longitud completa de la secuencia presentada. Tal como se mencionó anteriormente, esta presentación tiene el propósito de abarcar cada uno y todos los fragmentos de cada secuencia, empezando en el primer nucleótido, el 2º nucleótido y así sucesivamente, hasta los 8 nucleótidos antes de llegar al final, y finalizando en cualquier parte desde el nucleótido número 8, 9, 10 y así sucesivamente para cada secuencia, hasta el último nucleótido, (siempre y cuando la secuencia sea única tal como se describió anteriormente). Virtualmente cualquier segmento de la región de la IDENTIFICACIÓN SECUENCIAL NÚMERO: 1 que empiece en el nucleótido 1 y finaliza en el nucleótido 1180, o la IDENTIFICACIÓN SECUENCIAL NÚMERO: 3 que empiece en el nucleótido 1 y finaliza en el nucleótido 1122, o sus complementos, eso es 20 o más nucleótidos de largo serán únicos. Aquellas personas con conocimiento en la industria que tengan experiencia en los métodos para seleccionar a aquellas secuencias, comúnmente en base a la capacidad del fragmento único para distinguir selectivamente a la secuencia de interés de otras secuencias en el genoma humano del fragmento a aquellas en bases de datos conocidas, comúnmente es todo lo que se necesita, aunque una hibridación de confirmación in vitro y un análisis de secuencias también puede realizarse.

Las células del invento pueden ser suministradas por métodos para incrementar la actividad generadora de la Cα-formilglicina en una célula. En secciones importantes, esto se logra al utilizar vectores ("vectores de expresión" y/o "vectores de dirección").

Los "vectores", tal como se utilizan en este documento, puede ser cualquiera de un número de ácidos nucleicos en los cuales una secuencia deseada puede ser insertada por medio de restricción y ligación para su transportación entre diferentes entornos genéticos o para la expresión en una célula anfitriona. Los vectores son compuestos comúnmente de ADN aunque vectores de ARN también están disponibles. Los vectores incluyen, pero no se limitan a, plásmidos, fagémidos, y genoma del virus. Un vector de clonación es uno que puede replicarse en una célula anfitriona, y que se caracteriza además por uno o más lugares de restricción de endonucleasas en los cuales el vector podría ser cortado en una forma determinable y en los cuales una secuencia deseada de ADN podría ligarse para que el nuevo vector recombinante retenga su capacidad de replicarse en la célula anfitriona. En el caso de los plásmidos, la replicación de la secuencia deseada podría ocurrir muchas veces cuando se incrementan los plásmidos en un número de copias dentro de la bacteria anfitriona o sólo una única vez por el anfitrión antes de que el anfitrión se reproduzca por medio de mitosis. En el caso de los fagos, una replicación podría ocurrir activamente durante una fase lítica o pasivamente durante una fase lisogénica. Un "vector de expresión" es uno en el cual una secuencia deseada de ADN (por ejemplo, la ADNc de la FGE de la IDENTIFICACIÓN SECUENCIAL NÚMERO: 3) podría insertarse por medio de restricción y ligación de tal forma que es unida operativamente a secuencias regulatorias y puede expresarse como una transcripción de ARN. Los vectores pueden contener adicionalmente una o más secuencias marcadoras adecuadas para su uso en la identificación de células que han sido o no han sido transformadas o transfectadas con el vector. Los marcadores incluyen, por ejemplo, a proteínas que codifican a genes que incrementan o reducen la resistencia o la sensibilidad a antibióticos u otros compuestos, genes que codifican a enzimas cuyas actividades son detectables por medio de ensayos estándar conocidos en la industria (por

ejemplo, β -galactosidasa o fosfatasa alcalina), y genes que afectan visualmente al fenotipo o a las células, anfitriones, colonias o placas transformadas o transfectadas (por ejemplo, proteínas fluorescentes verdes).

Un "vector de dirección" es uno que contiene comúnmente a estructuras/secuencias de dirección que son utilizadas, por ejemplo, para insertar una secuencia regulatoria dentro del gen endógeno (por ejemplo, dentro de las secuencias de un exón y/o un intrón), dentro de las secuencias promotoras del gen endógeno, o corriente arriba de las secuencias promotoras del gen endógeno. En otro ejemplo, un vector de dirección podría contener al gen de interés (por ejemplo, codificado por la ADNc de la IDENTIFICACIÓN SECUENCIAL NÚMERO: 1) y otras secuencias necesarias para la dirección del gen a una ubicación preferida en el genoma (por ejemplo, una ubicación activa de transcripciones, por ejemplo, corriente abajo de un promotor endógeno de un gen no relacionado). La construcción de estructuras y vectores de dirección se describe en detalle en las patentes de Estados Unidos 5,641,670 y 6,270,989.

Virtualmente cualquier célula, procariota o eucariota, que puede transformarse con ADNs o ARNs heterólogos y que puede cultivarse o mantenerse en un cultivo, podría utilizarse en la práctica de este invento. Ejemplos incluyen a células bacterianas tales como el *Escherichia coli*, células de insectos y células de mamíferos tales como humanas, de ratones, de conejillos de indias, de puercos, de cabras, de primates, etcétera. Éstas podrían ser cepas celulares primarias o secundarias (que exhiben un número finito de doblajes medios poblacionales en cultivos y que no son inmortalizados) y líneas celulares inmortalizadas (que muestran una vida útil aparentemente ilimitada en los cultivos). Células binarias y secundarias incluyen, por ejemplo, a fibroblastos, queratinocitos, células epiteliales (por ejemplo, células epiteliales mamarias, células epiteliales intestinales), células endoteliales, células gliales, células neurales, elementos formados de la sangre (por ejemplo, linfocitos, células de la médula ósea), células musculares y precursores de estos tipos celulares somáticos incluyendo a células madres embrionarias. En los casos en que las células van a ser utilizadas en la terapia genética, las células primarias son obtenidas preferiblemente del individuo al cual las células manipuladas van a ser administradas. Sin embargo, las células primarias podrían obtenerse de un donante (otra persona que no sea el recipiente) de la misma especie. Ejemplos de líneas celulares humanas inmortalizadas que podrían ser utilizadas con las estructuras y métodos de ADN de este invento incluyen, pero no se limitan a, las células HT-1080 (ATCC CCL 121), las células HeLa y los derivados de las células HeLa (ATCC CCL 2, 2.1 y 2.2), las células MCF-7 de cáncer de la mama (ATCC BTH 22), las células K-562 de leucemia (ATCC CCL 243), las células KB de carcinoma (ATCC CCL 17), las células 2780AD de carcinoma de los ovarios (Van der Buck, A. M. et al., Cancer (Cáncer) Res, 48:5927-5932 (1988), las células Raji (ATCC CCL 86), las células WiDr de adenocarcinoma del colon (ATCC CCL 218), las células SW620 de adenocarcinoma del colon (ATCC CCL 227), las células Jurkat (ATCC TD3 152), las células Namalwa (ATCC CRL1432), las células HL-60 (ATCC CCL 240), las células Daudi (ATCC CCL 213), las células RPMI 8226 (ATCC CCL 155), las células U-937 (ATCC CRL 1593), las células Bowes de melanoma (ATCC CRL 9607), las células 2R4 de la sub - línea WI-38VA13 (ATCC CLL 75.1), y las células MOLT-4 (ATCC CRL 1582), las células CHO y las células COS, así como células del heterohibridoma producidas por la fusión de células humanas y células de otras especies. Cepas secundarias de fibroblastos humanos, tales como WI-38 (ATCC CCL 75) y MRC-5 (ATCC CCL 171) también pueden ser utilizadas. Información adicional acerca de los tipos de células que pueden ser utilizadas para practicar los métodos de este invento se describen en las patentes de Estados Unidos 5,641,670 y 6,270,989. Los sistemas de transcripción libres de células también pueden ser utilizados en lugar de células.

Las células del invento son mantenidas bajo condiciones, tal como se conoce en la industria, que resultan en la expresión de la proteína de la FGE o sus fragmentos funcionales. Las proteínas expresadas utilizando los métodos descritos pueden purificarse a partir de lisados celulares o sobrenadantes. Las proteínas hechas de acuerdo a este método pueden prepararse como una formulación farmacéuticamente útil y puede entregarse a un humano o a un animal no humano por rutas convencionales farmacéuticas tales como es conocido en la industria (por ejemplo, por vía oral, intravenosa, intramuscular, intranasal, intratráquea o subcutánea). Tal como se describe en otras secciones de este documento, las células recombinantes pueden ser células inmortalizadas, primarias, o secundarias, preferiblemente humanas. El uso de células de otras especies podría ser deseable en los casos en los cuales las células no humanas son ventajosas para fines de producción proteínica en los casos en los que la FGE humana producida es terapéuticamente útil.

Tal como se utiliza en este documento, una secuencia de codificación y las secuencias regulatorias se dice que están "operativamente" juntas cuando están enlazadas covalentemente para colocar a la expresión o a la transcripción de la secuencia codificadora bajo la influencia o el control de las secuencias regulatorias. Si se desea que las secuencias codificadoras se interpreten a una proteína funcional, 2 secuencias de ADN se dicen que están unidas operativamente si la inducción de un promotor en las secuencias regulatorias de 5' resulta en la transcripción de la secuencia codificadora y si la naturaleza de la vinculación entre las 2 secuencias de ADN no (1) resulta en la introducción de una mutación con desplazamiento del marco de lectura, (2) interfiera con la capacidad de la región promotora para orientar la transcripción de las secuencias codificadoras, o (3) interfiera con la capacidad de las transcripciones correspondientes de ARN para hacer interpretados en una proteína. Por lo tanto, una región promotora estaría unida operativamente a su secuencia codificadora si la región promotora fuese capaz de efectuar transcripciones de aquella secuencia de ADN para que la transcripción resultante pudiese ser interpretada a la proteína o al polipéptido deseado.

La naturaleza precisa de las secuencias regulatorias necesarias para la expresión genética podría variar entre tipos de especies celulares, pero, en general, deberá incluir, como necesario, secuencias de 5' no transcritas y no interpretadas involucradas con la iniciación de la transcripción y de la interpretación respectivamente, tales como la caja TATA, la secuencia de nivelación, la secuencia CAAT, y similares. Especialmente, aquellas secuencias regulatorias no transcritas de 5' incluirán a una región promotora que incluye a una secuencia promotora para el control de las transcripciones del gen unido operativamente. Las secuencias regulatorias también podrían incluir a secuencias mejoradoras o secuencias activadoras corriente arriba tal como sea deseado. Los vectores podrían incluir opcionalmente a secuencias de liderazgo o de señalización de 5'. La opción y el diseño de un vector apropiado caen dentro de la capacidad y discreción de una persona con conocimiento normal en la industria.

Los vectores de expresión que contienen todos los elementos necesarios para la expresión están comercialmente disponibles y son conocidos para aquellas personas con conocimiento en la industria. Refiérase a, por ejemplo, Sambrook et al, *Molecular Cloning: A Laboratory Manual* (Clonación Molecular: Un Manual de Laboratorio), 2ª edición, Cold Spring Harbor Laboratory Press, 1989. Las células son diseñadas genéticamente por la introducción a las células de polipéptidos o fragmentos heterólogos de la FGE que codifican al ADN (ARN) o sus variantes. Aquel ADN (ARN) heterólogo es ubicado bajo el control operativo de los elementos de transcripción para permitir la expresión del ADN heterólogo en la célula anfitriona.

Sistemas preferidos de la expresión del ARNm en células de mamíferos son aquellos tales como pRc/CMV (disponible de Invitrogen, Carlsbad, CA) que contienen un marcador seleccionable tal como un gen que otorga una resistencia al G418 (que facilita la selección de líneas celulares transfectadas establemente) y las secuencias promotoras del mejorador del citomegalovirus (CMV) humano. Adicionalmente, adecuado para la expresión en líneas celulares de primates o de caninos es el vector pCEP4 (Invitrogen, Carlsbad, CA), que contiene un virus Epstein barr (EBV - Epstein Barr virus) originado de una replicación, facilitando el mantenimiento del plásmido como un elemento extra cromosómico de multi copias. Otro vector de expresión es el plásmido pEF-BOS que contiene al promotor del factor de alargamiento de polipéptidos 1 α , que estimula la transcripción eficiente in vitro. El plásmido es descrito por Mishizuma y Nagata (*Nuc. Acids Res.* 18:5322, 1990), y su uso en experimentos de transfección se presenta, por ejemplo, por Demoulin (*Mol. Cell. Biol.* 16:4710-4716, 1996). Otro vector preferido de expresión es un adenovirus, descrito por StratfordPerricaudet, el cual es deficiente para las proteínas E1 y E3 (*J. Clin. Invest.* 90:626-630, 1992). El uso del adenovirus como un recombinante Adeno.PIA se presenta en Warmer et al., en inyecciones intradérmicas en ratones para su inmunización en contra de P1A (*Int. J. Cancer* (cáncer), 67:303-310, 1996).

Pueden suministrarse botiquines de expresión, que le permitirán a la persona con conocimiento preparar un vector o varios vectores deseados de expresión. Aquellos botiquines de expresión incluyen por lo menos porciones separadas de cada una de las secuencias de codificación ya descritas. Otros componentes podrían ser agregados, tal como fuese deseado, mientras se incluyan las secuencias mencionadas anteriormente, las cuales son requeridas.

La ya descrita, secuencia de ADNc de la FGE que contiene a los vectores de expresión puede ser utilizada para realizar una transfección de las células anfitrionas y de líneas celulares, siendo estas procariotas (por ejemplo, el *Escherichia coli*), o Eucariotas (por ejemplo, las células CHO, las células COS, los sistemas de expresión de levadura y la expresión recombinante del báculo virus en células de insectos). Especialmente útiles son las células de mamíferos tales como las de humanos, de ratones, de conejillos de indias, de cerdos, de cabras, de primates, etcétera. Podrían ser una amplia variedad de tipos de tejidos, incluyendo a células primarias y líneas celulares inmortalizadas tal como se describe en otras secciones de este documento. Ejemplos específicos incluyen a células HT-1080, células CHO, células dendríticas, células U293, leucocitos sanguíneos periféricos, células madre de médula ósea, células madres embrionarias y células de insectos.

Pueden suministrarse polipéptidos aislados (incluyendo proteínas completas y proteínas parciales) codificadas por el ácido nucleico de la FGE ya mencionado, e incluir al polipéptido de la IDENTIFICACIÓN SECUENCIAL NÚMERO: 2 y a sus fragmentos únicos. Aquellos polipéptidos son útiles, por ejemplo, individualmente o como parte de proteínas de fusión para generar anticuerpos, tal como componentes de un inmunoensayo, etcétera. Los polipéptidos pueden ser aislados de muestras biológicas incluyendo homogeneizados de tejidos o de células, y también pueden expresarse recombinantemente en una variedad de sistemas de expresión procariotas y eucariotas al construir un vector de expresión apropiado para el sistema de expresión, introduciendo al vector de expresión en el sistema de expresión, y aislando a la proteína expresada recombinantemente. Polipéptidos cortos, incluyendo a péptidos antigénicos (tales como aquellos presentados por las moléculas MHC en la superficie de una célula para reconocimiento inmunológico) también pueden sintetizarse químicamente utilizando métodos bien establecidos para la síntesis de péptidos.

Un fragmento único de un polipéptido de la FGE, en general, tienen las características y propiedades de fragmentos únicos tal como se mencionó anteriormente en conexión con los ácidos nucleicos. Tal como será reconocido por aquellas personas con conocimiento en la industria, el tamaño del fragmento único dependerá de factores tales como si el fragmento constituyera una porción de un dominio conservado de proteínas. Por lo tanto, algunas regiones de la IDENTIFICACIÓN SECUENCIAL NÚMERO: 2 requerirán que segmentos más largos sean únicos mientras que otros requerirán solamente segmentos cortos, comúnmente entre 5 y 12 aminoácidos (por ejemplo, 5,

6, 7, 8, 9, 10, 11 y 12 aminoácidos de largo o más, incluyendo cada número entero hasta la longitud completa, 287 aminoácidos de largo).

Fragmentos únicos de un polipéptido son, preferiblemente, aquellos fragmentos que retienen una capacidad funcional distinta del polipéptido. Las capacidades funcionales que pueden ser retenidas en un fragmento único de un polipéptido incluyen la interacción con anticuerpos, la interacción con otros polipéptidos o sus fragmentos, la interacción con otras moléculas, etcétera. Una actividad importante es la capacidad de actuar como una firma para identificar al polipéptido. Aquellas personas con conocimiento en la industria tienen un buen conocimiento de métodos para seleccionar secuencias únicas de aminoácidos, comúnmente en base a la capacidad de un fragmento único para distinguir selectivamente las secuencias de interés de miembros que no pertenecen a la familia. Una comparación de la secuencia del fragmento con aquellas en bases de datos conocidas comúnmente es todo lo que se necesita.

Pueden suministrarse variaciones de los polipéptidos de la FGE ya descritos. Tal como se utiliza en este documento, una "variación" de un polipéptido de la FGE es un polipéptido que contiene una o más modificaciones a la secuencia primaria de aminoácidos de un polipéptido de la FGE. Las modificaciones que crean a una variación del polipéptido de la FGE son comúnmente hechas al ácido nucleico que codifica al polipéptido de la FGE, y puede incluir eliminaciones, mutaciones de puntos, truncamientos, sustituciones de aminoácidos y adiciones de aminoácidos o de partículas que no son aminoácidos para: 1) reducir o eliminar una actividad de un polipéptido de la FGE; 2) mejorar una propiedad de un polipéptido de la FGE, tal como la estabilidad proteínica en un sistema de expresión o la estabilidad del enlace entre la proteína y el ligando; 3) suministrar una actividad o propiedad nueva a un polipéptido de la FGE, tal como la adición de un epítipo antigénico o la adición de una partícula detectable; o 4) para suministrar enlaces equivalentes o mejorados a un receptor del polipéptido de la FGE u otra molécula. Alternamente, se pueden realizar modificaciones directamente al polipéptido, tal como por la división, adición de una molécula enlazadora, la adición de una partícula detectable, tal como la biotina, la adición de un ácido graso, y similares. Las modificaciones también abarcan a las proteínas de fusión que comprenden a toda o a parte de la secuencia de aminoácidos de la FGE. Una persona con conocimiento en la industria estará familiarizada con métodos para predecir el efecto en una conformación proteínica de un cambio en la secuencia proteínica, y puede, por lo tanto, "diseñar" un polipéptido FGE variado de acuerdo a los métodos conocidos: un ejemplo de un método como estos es descrito por Dahiyat y Mayo en *Science (Ciencia)* 278:82-87, 1997, donde las proteínas pueden ser diseñadas desde cero. El método puede ser aplicado a una proteína conocida para variar solo una porción de la secuencia del polipéptido. Al aplicar los métodos computacionales de Dahiyat y Mayo, variaciones específicas del polipéptido de la FGE pueden ser propuestas y probadas para determinar si la variación retiene una conformación deseada.

Las variaciones pueden incluir a polipéptidos de la FGE que son modificados específicamente para alterar una característica del polipéptido que no se relaciona a su actividad fisiológica. Por ejemplo, los residuos de cisteína pueden ser sustituidos o eliminados para prevenir enlaces de disulfuros no deseados. Asimismo, ciertos aminoácidos pueden ser cambiados para mejorar la expresión de un polipéptido de la FGE al eliminar la proteólisis de las proteasas en un sistema de expresión (por ejemplo, los residuos de aminoácidos dibásicos en los sistemas de expresión de la levadura en los cuales la actividad de las proteasas KEX2 está presente).

Las mutaciones de un ácido nucleico que codifica a un polipéptido de la FGE conserva, preferiblemente, el marco de lectura de los aminoácidos de la secuencia codificadora, y preferiblemente no crea regiones en el ácido nucleico que posiblemente se hibridarán para formar estructuras secundarias, tales como horquillas o circuitos, que podrían eliminar las expresiones de un polipéptido variante.

Se pueden hacer mutaciones al seleccionar una sustitución de aminoácidos, o por medio de una mutagénesis aleatoria de un lugar seleccionado en un ácido nucleico que codifica al polipéptido. Los polipéptidos variantes son expresados y probados entonces para detectar una o más actividades para determinar qué mutación suministra una variación del polipéptido con las propiedades deseadas. Mutaciones adicionales pueden hacerse a variantes (o a polipéptidos de la FGE no variantes) que son silenciosos en lo que se refiere a la secuencia de aminoácidos del polipéptido, pero que suministran codones preferidos para la interpretación en un anfitrión específico, o alteran la estructura del ARNm, por ejemplo, para mejorar la estabilidad y/o la expresión. Los codones preferidos para la interpretación de un ácido nucleico en, por ejemplo, la *Escherichia coli*, las células de mamíferos, etcétera son bien conocidos por aquellas personas con un conocimiento normal en la industria. Otras mutaciones pueden ser hechas a las secuencias no calificadoras del gen de la FGE o al clon de la ADNc para mejorar la expresión del polipéptido.

Una persona con conocimiento en la industria se da cuenta que sustituciones conservadoras de aminoácidos pueden hacerse en polipéptidos de la FGE para suministrar variantes equivalentes funcionalmente de los polipéptidos ya mencionados, es decir, las variantes retienen las capacidades funcionales de los polipéptidos de la FGE. Tal como se utiliza en este documento, una "sustitución conservadora de aminoácidos" se refiere a una sustitución de aminoácidos que no altera significativamente la estructura terciaria y/o la actividad del polipéptido. Las variantes pueden ser preparadas de acuerdo a métodos de alteración de la secuencia de los polipéptidos que son conocidos para personas con conocimiento normal en la industria, e incluyen aquellas que se encuentran en referencias que recopilan aquellos métodos, por ejemplo, *Molecular Cloning: A Laboratory Manual (Clonación Molecular: un Manual de Laboratorio, J. Sambrook, et al., eds., 2ª edición, Cold Spring Harbor Laboratory Press,*

Cold Spring Harbor, Nueva York, 1989, o Current Protocols in Molecular Biology (Protocolos Actuales en Biología Molecular), F.M. Ausubel, et al., eds., John Wiley & Sons, Inc., Nueva York. Variantes equivalentes funcionalmente de ejemplo de los polipéptidos de la FGE incluyen sustituciones conservadoras de aminoácidos de la IDENTIFICACIÓN SECUENCIAL NÚMERO: 2. Sustituciones conservadoras de aminoácidos incluyen sustituciones hechas a aminoácidos dentro de los siguientes grupos: (a) M, I, L, V; (b) F, Y, W; (c) K, R, H; (d) A, G; (e) S, T; (f) Q, N; y (g) E, D.

Por lo tanto, las variantes funcionalmente equivalentes de los polipéptidos de la FGE, es decir, las variantes de los polipéptidos de la FGE que retienen la función de los polipéptidos naturales de la FGE, están contempladas por el invento. Sustituciones conservadoras de aminoácidos en la secuencia de aminoácidos de los polipéptidos de la FGE para producir variantes equivalentes funcionalmente de los polipéptidos de la FGE se realizan comúnmente por medio de la alteración de un ácido nucleico que codifica a los polipéptidos de la FGE (IDENTIFICACIONES SECUENCIALES NÚMEROS: 1, 3). Aquellas sustituciones pueden realizarse por medio de una variedad de métodos conocidos para una persona con conocimiento normal en la industria. Por ejemplo, las sustituciones de aminoácidos pueden realizarse por medio de mutaciones dirigidas por PCR, mutagénesis dirigida por lugares de acuerdo al método de Kunkel (Kunkel, Proc. Not. Acad. Sci. Estados Unidos de América 82: 488-492, 1985), o por medio de síntesis química de un gen que codifica a un polipéptido de la FGE. La actividad de fragmentos funcionalmente equivalentes de los polipéptidos de la FGE pueden probarse al clonar al gen que codifica al polipéptido de la FGE alterado en un vector de expresión bacteriana o de un mamífero, introduciendo al vector en una célula anfitriona apropiada, expresando al polipéptido alterado de la FGE, y probando para detectar una capacidad funcional de los polipéptidos de la FGE tal como se presenta en este documento (por ejemplo, la actividad generadora de C α -formilglutamina, etc.).

Aquellas personas con conocimiento en la industria podrán seguir fácilmente métodos conocidos para aislar los polipéptidos de la FGE. Estos incluyen, pero no se limitan a, inmuno - cromatografía, HPLC, cromatografía por medio de exclusión de tamaños, cromatografía por medio de intercambio de iones y cromatografía por medio de afinidad inmunológica.

Los polipéptidos de la FGE preferiblemente son producidos recombinantemente, aunque aquellos polipéptidos podrían ser aislados de extractos biológicos. Polipéptidos de la FGE producidos recombinantemente incluyen a proteínas quiméricas que comprenden una fusión de una proteína de la FGE con otro polipéptido, por ejemplo, un polipéptido capaz de suministrar o mejorar los enlaces proteína-proteína, los enlaces de ácidos nucleicos de secuencias específicas (tales como GAL4), mejorar la estabilidad del polipéptido de la FGE bajo condiciones de ensayo, o suministrar una partícula eliminable, tal como la proteína verde fluorescente. Un polipéptido fusionado a un polipéptido o fragmento de la FGE. También pueden suministrar medios de detección fácil de la proteína de fusión, por ejemplo, por medio de reconocimiento inmunológico o por medio de marcación fluorescente.

Un método para incrementar la actividad generadora de C α -formilglutamina en una célula involucra contactar a la célula con una molécula aislada de ácido nucleico del invento (por ejemplo, un ácido nucleico de la IDENTIFICACIÓN SECUENCIAL NÚMERO 1), o uno de sus productos de expresión (por ejemplo, un péptido de la IDENTIFICACIÓN SECUENCIAL NÚMERO 2), en un monto efectivo para incrementar la actividad generadora de C α -formilglutamina en la célula. En secciones importantes, el método incluye activar el gen endógeno de la FGE para incrementar la actividad generadora de C α -formilglutamina en la célula.

En cualquiera de las secciones anteriores, el ácido nucleico puede estar acoplado operativamente a una secuencia de expresión genética que dirige la expresión de la molécula de ácido nucleico dentro de una célula eucariota tal como una célula HT-1080. La "secuencia de expresión genética" es cualquier secuencia regulatoria de nucleótidos, tal como una secuencia promotora o una combinación promotora-mejoradora, que facilita la transcripción e interpretación eficientes del ácido nucleico con el cual está enlazado operativamente. La secuencia de expresión genética puede, por ejemplo, ser un promotor mamífero o viral, tal como un promotor constitutivo o inducible. Los promotores constitutivos mamíferos incluyen, pero no se limitan a, los promotores para los siguientes genes: la hipoxantina fosforribosiltransferasa (HPTR), la adenosina desaminasa, la piruvato quinasa, el promotor de α -actina y otros promotores constitutivos. Promotores virales de ejemplo que funcionan constitutivamente en células eucariotas incluyen, por ejemplo, promotores provenientes del virus símico, el virus de papiloma, el adenovirus, el virus de inmunodeficiencia humana (VIH), el virus del sarcoma de Rous, el citomegalovirus, las repeticiones terminales largas (LTR - long terminal repeat) del virus de leucemia de moloney y otros retrovirus, y el promotor de quinasa de timidina del virus de herpes simple. Otros promotores constitutivos son conocidos para aquellas personas con un conocimiento normal en la industria. Los promotores útiles como secuencias de expresión genética del invento también incluyen a promotores inducibles. Los promotores inducibles son activados en la presencia de un agente de inducción. Por ejemplo, el promotor de metalotioneína es activado para incrementar la transcripción e interpretación en la presencia de ciertos iones metálicos. Otros promotores inducibles son conocidos por aquellas personas con un conocimiento normal en la industria.

En general, las secuencias de expresión genética deberán incluir, tal como fuese necesario, a secuencias de 5' que no transcriben y de 5' que no interpretan involucradas con la iniciación de la transcripción e interpretación, respectivamente, tales como una caja TATA, la secuencia de nivelación, la secuencia CAAT, y similares.

Especialmente, aquellas secuencias de 5' que no transcriben incluirán una región promotora que incluye una secuencia promotora para el control de transcripción del ácido nucleico unido operativamente. Las secuencias de expresión genética incluyen opcionalmente a secuencias mejoradoras o secuencias activadoras corriente arriba tal como sea deseado.

Preferiblemente, cualquiera de las moléculas de ácidos nucleicos de la FGEs aquí descritas es enlazada a una secuencia de expresión genética que permite la expresión de la molécula de ácido nucleico en una célula de un linaje celular específico, por ejemplo, una neurona. Una secuencia que permite la expresión de la molécula de ácido nucleico en una célula tal como una neurona, es una que es activa selectivamente en ese tipo celular, causando, por lo tanto, la expresión de la molécula de ácido nucleico en estas células. El promotor sinapsina-1, por ejemplo, puede ser utilizado para expresar a cualquiera de las moléculas ya mencionadas de ácido nucleico del invento en una neurona; y el promotor genético del factor von Willebrand, por ejemplo, puede ser utilizado para expresar una molécula de ácido nucleico en una célula endotelial vascular. Aquellas personas con conocimiento normal en la industria identificarán fácilmente a promotores alternativos que sean capaces de expresar una molécula de ácido nucleico en cualquiera de las células preferidas del invento.

La secuencia de ácido nucleico y la secuencia de expresión genética, se dice que están "enlazadas operativamente" cuando están enlazadas covalentemente para colocar la transcripción y/o la interpretación de la secuencia que codifica al ácido nucleico (por ejemplo, en el caso de la FGE, IDENTIFICACIÓN SECUENCIAL NÚMERO 3) bajo la influencia o el control de la secuencia de expresión genética. Si fuese deseado que la secuencia del ácido nucleico sea interpretada a una proteína funcional, 2 secuencias de ADN se consideran como enlazadas operativamente si la inducción de una proteína en la secuencia de expresión genética de 5' resulta en la transcripción de la secuencia de ácido nucleico y la naturaleza del enlace entre las 2 secuencias de ADN no (1) resulta en la inducción de una mutación de traslado de marco, (2) interfiere con la capacidad de la región promotora para dirigir la transcripción de la secuencia de ácido nucleico, y/o (3) interfiere con la capacidad de la transcripción correspondiente de ARN para ser interpretada a la proteína. Por lo tanto, una secuencia de expresión genética estaría enlazada operativamente a una secuencia de ácido nucleico si la secuencia de expresión genética fuese capaz de efectuar transcripciones de aquella secuencia de ácidos nucleicos para que la transcripción resultante pudiese ser interpretada en la proteína o polipéptido deseado.

Las moléculas aquí descritas pueden ser entregadas a los tipos de células preferidas del invento individualmente o en combinación con un vector (refiérase a la sección anterior de vectores). En el sentido más amplio (y consistentemente con la descripción de la expresión y de los vectores de dirección descritos en otras secciones de este documento), un "vector" es cualquier portador capaz de facilitar: (1) la entrega de una molécula a una célula objetivo y/o (2) la absorción de la molécula por una célula objetivo. Preferiblemente, los vectores de entrega transportan a la molécula a la célula objetivo con una degradación reducida en relación a la magnitud de la degradación que resultaría con la ausencia del vector. Opcionalmente, un "ligando de dirección" puede adherirse al vector para entregar selectivamente al vector a una célula que expresa en su superficie el receptor cognado para el ligando de dirección. De esta forma, el vector (que contiene un ácido nucleico o una proteína) puede ser entregado selectivamente a una neurona. Metodologías para dirigir incluyen comprobaciones, tales como aquellas descritas en la patente de Estados Unidos 5,391,723 otorgada a Priest. Otro ejemplo de un portador de dirección bien conocido es un liposoma. Los liposomas están comercialmente disponibles de Gibco BRL. Muchos métodos están publicados para elaborar liposomas dirigidos.

En general, vectores útiles incluyen, pero no se limitan a, plásmidos, fagémidos, virus, otros vehículos derivados de fuentes virales o bacterianas que han sido manipulados por medio de la inserción o incorporación de las secuencias de ácidos nucleicos aquí descritas y fragmentos adicionales de ácidos nucleicos (por ejemplo, mejoradores, promotores) que pueden adherirse a la secuencia de ácidos nucleicos. Los vectores virales son el tipo preferido de vectores e incluyen, pero no se limitan a, secuencias de ácidos nucleicos de los siguientes virus: el adenovirus; el virus asociado con el adeno; retrovirus, tal como el virus de leucemia de ratones de moloney; el virus del sarcoma de ratones de harvey; el virus del tumor de las mamas de ratones; el virus del sarcoma de rouse; los virus de tipo SV40; los virus de polio; los virus de Epstein-Barr; los virus de papiloma; los virus del herpes; los virus de vaccinia; el virus de la polio; y virus de ARN tales como un retrovirus. Una persona puede utilizar fácilmente otros vectores no nombrados pero que son conocidos en la industria.

Un virus particularmente preferido para ciertas aplicaciones es el virus asociado con el adeno, un virus de ADN de doble cepa. El virus asociado con el adeno es capaz de infectar una amplia gama de tipos celulares y especies y puede ser diseñado para que tenga una replicación deficiente. Tiene ventajas adicionales, tales como estabilidad de solventes calientes y lípidos, frecuencias de interpretación alta en células de diversos linajes, incluyendo células hematopoyéticas, y la falta de inhibición de súper - infección permitiendo, por lo tanto, series múltiples de transducciones. Se ha reportado que el virus asociado con el adeno puede integrarse al ADN celular humano en lugares específicos, por lo tanto, minimizando la posibilidad de una mutagénesis por inserciones y la variabilidad de la expresión genética insertada. Adicionalmente, infecciones del virus asociado con el adeno de tipo silvestre han sido seguidas en cultivos celulares por más de 100 pases en la ausencia de presión selectiva, implicando que la integración genómica del virus asociado con el adeno es un evento relativamente estable. El virus asociado con el adeno también puede funcionar en una forma extra cromosómica.

En general, otros vectores virales preferidos se basan en virus eucariotas no citopáticos en los cuales genes no esenciales han sido reemplazados con el gen de interés. Los virus no citopáticos incluyen retrovirus, que tienen una vida útil que involucra la transcripción en reversa del ARN viral genómico a ADN con una integración pro - viral subsiguiente en un ADN celular anfitrión. Los adenovirus y los retrovirus han sido aprobados para ensayos de terapia genética humana. En general, los retrovirus tienen una deficiencia de replicación (es decir, son capaces de dirigir la síntesis de las proteínas deseadas, pero no son capaces de fabricar una partícula infecciosa). Aquellos vectores de expresión retroviral genéticamente alterados tienen una utilidad general para la transducción de alta eficiencia en genes in vivo. Protocolos estándar para producir retrovirus con deficiencias de replicación (incluyendo los pasos de incorporación del material genético exógeno en un plásmido, la transfección de una célula de empaque alineada con plásmidos, la producción de retrovirus recombinantes por la línea celular de empaque, la recolección de partículas virales de medio de cultivo de tejidos, y la infección de las células de dirección con partículas virales) se suministran en Kriegler, M., "Gene Transfer and Expression, A Laboratory Manual" ("Transferencia y Expresión Genética. Un Manual de Laboratorio"), W.H. Freeman CO., Nueva York (1990) y Murry, E.J. Ed. "Methods in Molecular Biology" ("Métodos en Biología Molecular") vol. 7, Humana Press, Inc., Clifton, Nueva Jersey (1991).

Otro vector retroviral preferido es el vector derivado del virus de leucemia de ratones de moloney, tal como es descrito en Nabel, E.G., et al., Science (Ciencia), 1990, 249:1285-1288. Se ha reportado que los vectores fueron efectivos para la entrega de genes a las 3 capas de la pared arterial, incluyendo al medio. Otros vectores preferidos se presentan en Flugelman, et al., Circulation (Circulación), 1992, 85:1110-1117. Vectores adicionales que son útiles para la entrega de moléculas son descritos en este documento también se describen en la patente de Estados Unidos número 5,674,722 de Mulligan, et. al..

Adicionalmente a los vectores ya mencionados, otros métodos de entrega podrían ser utilizados para entregar una molécula del invento a una célula tal como una neurona, al hígado, a fibroblastos y/o a una célula endotelial vascular, y facilitar de esa forma su absorción.

Un método de entrega preferido del invento es un sistema de dispersión coloidal. Los sistemas de dispersión coloidales incluyen sistemas que se basan en lípidos incluyendo emulsiones, micelas mezcladas y liposomas de aceites en agua. Un sistema coloidal preferido del invento es un liposoma. Los liposomas son portadores de membranas artificiales que son útiles como un vector de entrega in vivo o in vitro. Se ha demostrado que portadores unilamelares grandes (LUV - large unilamellar vessels), que varían en tamaño desde 0.2-4.0 micrómetros pueden encapsular macromoléculas grandes. ARN ADN y viriones intactos pueden ser encapsulados dentro del interior acuoso y pueden ser entregados a células en una forma biológicamente activa (Fraleley, et al., Trends (Tendencias) Biochem. Sci, 1981,6:77). Para qué un liposoma sea un vector de transferencia genética eficiente, una o más de las siguientes características deben estar presentes: (1) la encapsulación del gen de interés a una eficiencia alta con la retención de la actividad biológica; (2) enlaces preferenciales y sustanciales a una célula objetivo en comparación con las células que no son objetivo; (3) la entrega de los contenidos acuosos de la vesícula al citoplasma celular objetivo con una alta eficiencia; y (4) una expresión precisa y efectiva de la información genética.

Los liposomas pueden ser dirigidos a un tejido específico, tal como el miocardio o la pared celular vascular, al acoplar al liposoma a un ligando específico tal como un anticuerpo monoclonal, azúcar, glicolípidos o proteínas. Los ligandos que pueden ser útiles para dirigir a un liposoma a la pared vascular incluyen, pero no se limitan a, la proteína de cobertura viral del virus de Sendai. Adicionalmente, el vector podría acoplarse a un péptido de dirección nuclear, que dirigirá al ácido nucleico al núcleo de la célula anfitriona.

Los liposomas están disponibles comercialmente de Gibco BRL, por ejemplo, como LIPOFECTIN™ y LEPOFECTACE™, que son formados de lípidos catiónicos tales como el cloruro de N-[1-(2, 3 dioleiloxy)-propil]-N, N, N-trimetilamonio (DOTMA) y el bromuro de dioctadecilamonio de dimetilo (DDAB - dimethyl dioctadecylammonium bromide). Métodos para elaborar liposomas son bien conocidos en la industria y han sido descritos en muchas publicaciones. Los liposomas también han sido revisados por Gregoriadis, G. en Trends in Biotechnology (Tendencias en Biotecnología), V. 3, p. 235-241 (1985). Liposomas nuevos para la entrega intracelular de macromoléculas, incluyendo ácidos nucleicos, también se describen en la aplicación internacional PCT número PCT/US96/07572 (publicación número WO 96/40060, titulada "Intracellular Delivery of Macromolecules" ("Entrega Intracelular de Macromoléculas")).

Agentes de compactación también pueden ser utilizados en combinación con un vector. Un "agente de compactación", tal como se utiliza en este documento, se refiere a un agente, tal como la histona, que neutraliza a las cargas negativas en el ácido nucleico y, por lo tanto, permite la compactación del ácido nucleico en un gránulo fino. La compactación del ácido nucleico facilita la absorción del ácido nucleico por la célula objetivo. Los agentes de compactación pueden ser utilizados individualmente, es decir, para entregar un ácido nucleico aislado del invento en una forma que sea absorbida más eficientemente por la célula o, más preferiblemente, en combinación con uno o más de los vectores ya descritos.

Otras composiciones de ejemplo que pueden ser utilizadas para facilitar la absorción por una célula objetivo de los ácidos nucleicos del invento incluyendo al fosfato de calcio y otros mediadores químicos de transporte intracelular, composiciones de microinyección, y electroporación.

Métodos para incrementar la actividad de sulfatasas en la célula pueden involucrar contactar una célula que expresa una sulfatasa con una molécula aislada de ácidos nucleicos que codifican a una FGE. "Incrementar" la actividad de la sulfatasas, tal como se utiliza en este documento, se refiere a una afinidad incrementada para, y/o conversión de, el sustrato específico para la sulfatasa, comúnmente el resultado de un incremento en la formación de FGly en la molécula sulfatasa. En una sección, la célula expresa una sulfatasa a niveles más altos que aquellos de las células de tipo silvestre. El término "incrementar la actividad de la sulfatasas en una célula" también se refiere a incrementar la actividad de una sulfatasa que es secretada por la célula. La célula podría expresar a una sulfatasa endógena y/o exógena. Aquel contacto de la molécula de la FGE también se refiere a activar al gen de la FGE endógeno de la célula. En secciones importantes, la sulfatasa endógena es activada. En ciertas secciones, la sulfatasa es la 2-sulfatasa de Iduronato, la sulfamidasa, la 6-sulfatasa de N-Acetilgalactosamina, la 6-sulfatasa de N-Acetilglucosamina, la arilsulfatasa A, la arilsulfatasa B, la arilsulfatasa C, la arilsulfatasa D, la arilsulfatasa E, la arilsulfatasa F, la arilsulfatasa G, la HSulf-1, la HSulf-2, la HSulf-3, la HSulf-4, la HSulf-5, y/o HSulf-6. En ciertas secciones la célula es una célula de mamíferos.

Una composición farmacéutica podría ser suministrada. La composición puede comprender una sulfatasa activada producida por una célula de este invento, en un monto farmacéuticamente efectivo para tratar una deficiencia de sulfatasas, y un portador farmacéuticamente aceptable. En secciones importantes, la sulfatasa es expresada a niveles más altos que las células normales / de control.

La célula que produce sulfatasas puede comprender: (i) una sulfatasa con una actividad incrementada en comparación a un control, y (ii) una enzima generadora de Formilglicina con una actividad incrementada en comparación a un control, donde la tasa de la sulfatasa activa en comparación a la sulfatasa total producida por la célula es incrementada en por lo menos un 5% sobre la tasa de la sulfatasa activa en comparación a la sulfatasa producida por la célula en la ausencia de la enzima generadora de Formilglicina. Se conoce en la industria que la sobreexpresión de sulfatasas pueden reducir la actividad de las sulfatasas endógenas (Anson et al., Biochem. J., 1993, 294:657-662). Además, sólo una fracción de las sulfatasas recombinantes es activa. Hemos descubierto, inesperadamente, que la actividad/expresión incrementada de la FGE en una célula con una actividad/expresión incrementada de sulfatasas resulta en la producción de una sulfatasa que es más activa. Puesto que la presencia de FGly en una molécula sulfatasa es asociada con la actividad de las sulfatasas, las "sulfatasas activas" pueden ser cuantificadas al determinar la presencia de FGly en el producto celular de sulfatasas utilizando la espectrometría de masa MALDI-TOF, tal como se describe en otras secciones de este documento. La tasa con las sulfatasas totales puede determinarse entonces fácilmente.

La célula que produce sulfatasas o las sulfatasas producidas pueden ser utilizadas para la terapia de deficiencias de sulfatasas. Aquellas enfermedades incluyen, pero no se limitan a, la deficiencia de sulfatasas múltiples, la mucopolisacaridosis II (MPS II; el síndrome de Hunter), la mucopolisacaridosis IIA (MPS IIA; el síndrome de Sanfilippo A), la mucopolisacaridosis VIII (MPS VIII), la mucopolisacaridosis IVA (MPS IVA; el síndrome de Morquio A), la mucopolisacaridosis VI (MPS VI; el síndrome de Maroteaux-Lamy), la leucodistrofia metacromática (MLD - Metachromatic Leukodystrophy), la condrodistrofia punctata recesiva vinculada al cromosoma X 1, y la ictiosis vinculada al cromosoma X (deficiencia de sulfatasas de esteroides).

Puede realizarse el tratamiento agudo y profiláctico de cualquiera de las condiciones que se acaban de mencionar. Tal como se utiliza en este documento, un tratamiento agudo se refiere al tratamiento de sujetos que tienen una condición específica. El tratamiento profiláctico se refiere al tratamiento de sujetos que tienen el riesgo de tener la condición, pero que en ese momento no la tienen o no experimentan los síntomas de la condición.

En su sentido más amplio, los términos "tratamiento" o "tratar" se refieren a los tratamientos agudos y profilácticos. Si el sujeto tiene la necesidad del tratamiento o está experimentando una condición (o ha o está teniendo una condición específica), entonces tratar la condición se refiere a aliviar, reducir o eliminar la condición de uno o más síntomas que surjan de la condición. En algunas secciones importantes, tratar la condición se refiere a aliviar, reducir o eliminar un síntoma específico o un subconjunto específico de síntomas asociados con la condición. Si el sujeto tiene la necesidad del tratamiento es una persona que tiene el riesgo de tener una condición, entonces tratar al sujeto se refiere a reducir el riesgo de que el sujeto tenga la condición.

La modalidad de administración y la dosis de un agente terapéutico del invento variará con la etapa específica de la condición que se está tratando, la edad y la condición física del sujeto que está siendo tratado, la duración del tratamiento, la naturaleza de la terapia concurrente (si existiese), la ruta específica de administración, y factores similares dentro del conocimiento y experiencia especializada del proveedor de salud.

Tal como se describe en este documento, los agentes pueden ser administrados en montos efectivos para tratar cualquiera de las deficiencias de sulfatasas ya mencionadas. En general, un monto efectivo es cualquier monto que puede causar un cambio beneficioso en un tejido deseado de un sujeto. Preferiblemente, un monto efectivo que es un monto suficiente para causar un cambio fenotípico favorable en una condición específica tal como una reducción, un alivio o una eliminación de un síntoma o de una condición completa.

En general, un monto efectivo es aquel monto de una preparación farmacéutica que individualmente, o junto con

otras dosis, produce la respuesta deseada. Esto podría involucrar solamente desacelerar temporalmente el progreso de la condición, aunque más preferiblemente, involucra detener el progreso de la condición permanentemente o retrasar la ocurrencia de o evitar que ocurra la condición. Esto puede ser monitoreado por métodos rutinarios. Generalmente, las dosis de compuestos activos variarían de entre alrededor de 0.01 miligramos/kilogramo por día a 1000 mg/kilogramos por día. Se espera que las dosis que varían desde 50 µg-500 µg/kilogramo sean adecuadas, preferiblemente oralmente y con una o varias administraciones diariamente.

Aquellos montos dependerán, desde luego, de la condición específica que se está tratando, la gravedad de la condición, los parámetros individuales del paciente incluyendo su edad, su condición física, su tamaño y su peso, la duración del tratamiento, la naturaleza de la terapia concurrente (si existiese), la ruta específica de administración y factores similares dentro del conocimiento y experiencia especializada del proveedor de salud. Dosis más bajas resultarán de ciertas formas de administración, tales como una administración intravenosa. En el caso de que una respuesta del sujeto sea insuficiente con las dosis iniciales aplicadas, dosis más altas (o dosis efectivamente más altas por medio de una ruta de entrega diferente más localizada) podrían utilizarse dependiendo de la magnitud permitida por la tolerancia del paciente. Varias dosis diarias están contempladas para lograr niveles sistémicos apropiados de los compuestos. Es preferible, generalmente, que una dosis máxima sea utilizada, es decir, la dosis segura más alta de acuerdo a un juicio médico sólido. Se entenderá por aquellas personas con conocimiento en la industria, sin embargo, que un paciente podría insistir en una dosis más baja o una dosis tolerable por razones médicas, razones psicológicas o virtualmente cualquier otra razón.

Los agentes pueden ser combinados, opcionalmente, con un portador farmacéuticamente aceptable para formar una preparación farmacéutica.

El término "portador farmacéuticamente aceptable" tal como se utilizan este documento significa una o más sustancias de relleno, de dilución o de encapsulamiento que sean adecuadas para su administración a un humano.

El término "portador" denota un ingrediente orgánico o inorgánico, natural o sintético, con el cual el ingrediente activo es combinado para facilitar su aplicación. Los componentes de las composiciones farmacéuticas también son capaces de ser combinadas con las moléculas de este invento, y entre sí, en una forma para que no exista una interacción que deshabilite sustancialmente la eficacia farmacéutica deseada. En algunos aspectos, las preparaciones farmacéuticas comprenden un agente en un monto efectivo para tratar una enfermedad.

Las preparaciones farmacéuticas podrían contener agentes adecuados de amortiguación, incluyendo: al ácido acético en una sal; al ácido cítrico en una sal; al ácido bórico en una sal; o al ácido fosfórico en una sal. Las composiciones farmacéuticas también podrían contener, opcionalmente, conservadores adecuados, tal como: el cloruro de bezalconio; el clorobutanol; el parabeno o el timerosal.

Está disponible una variedad de rutas de administración. La modalidad particular seleccionada dependerá, desde luego, del medicamento específico seleccionado, de la gravedad de la condición que se está tratando y de la dosis requerida para una eficacia terapéutica.

Los métodos de tratamiento, hablando en general, pueden practicarse utilizando cualquier modalidad de administración que sea médicamente aceptable, refiriéndose a cualquier modalidad que produzca niveles efectivos de los compuestos activos sin causar efectos colaterales adversos clínicamente inaceptables.

Aquellas modalidades de administración incluyen las rutas orales, rectales, tópicas, nasales, intradérmicas, transdérmicas, o parenterales. El término "parenteral" incluyen rutas subcutáneas, intravenosas, intra-oméntales, intramusculares o por medio de infusiones. Las rutas intravenosas o intramusculares no son particularmente adecuadas para terapias y profilaxis a largo plazo. Por ejemplo, composiciones farmacéuticas para el tratamiento agudo de sujetos que tienen un dolor de cabeza tipo migraña pueden formularse en una variedad de formas diferentes y para una variedad de modalidades de administración incluyendo a tabletas, cápsulas, polvos, supositorios, inyecciones y aerosoles nasales.

Las preparaciones farmacéuticas pueden presentarse convenientemente en formas de dosis unitarias y pueden prepararse por cualquiera de los métodos bien conocidos en la industria farmacéutica.

Todos los métodos incluyen el paso de asociar al agente activo con un portador que constituye uno o más ingredientes tipo accesorio. En general, las composiciones son preparadas al asociar uniformemente e íntimamente al compuesto activo con un portador líquido, un portador sólido dividido finamente, o ambos, y entonces, si fuese necesario, darle forma al producto.

Las composiciones adecuadas para la administración oral pueden presentarse en forma de unidades discretas, tales como cápsulas, tabletas, pastillas, cada una conteniendo un monto predeterminado de un compuesto activo. Otras composiciones incluyen suspensiones en líquidos acuosos o líquidos no acuosos tales como un jarabe, un elixir o una emulsión.

Composiciones adecuadas para la administración parenteral comprenden convenientemente una preparación acuosa estéril de un agente del invento, que es preferiblemente isotónico a la sangre del recipiente.

Esta preparación acuosa puede ser formulada de acuerdo a métodos conocidos utilizando agentes de dispersión o humectantes adecuados y agentes de suspensión. La preparación inyectable estéril también puede ser una solución o suspensión inyectable estéril en un diluyente o solventes aceptables que no sean tóxicos, por ejemplo, una solución en diol de 1,3-butano.

Entre los portadores y solventes aceptables que pueden ser utilizados se encuentran el agua, la solución de Ringer, y una solución isotónica de cloruro de sodio. Adicionalmente, aceites estériles y fijos son utilizados convencionalmente como medios solventes o de suspensión. Con este fin, cualquier aceite fijo blando puede ser utilizado incluyéndolos tipo mono- o di-. Adicionalmente, ácidos grasos tales como el ácido oleico pueden ser utilizados en la preparación de inyectables.

Formulaciones adecuadas para administraciones orales, subcutáneas, intravenosas, intramusculares, etcétera pueden encontrarse en Remington's Pharmaceutical Sciences (Ciencias Farmacéuticas de Remington), Mack Publishing Co., Easton, PA.

Puede facilitarse un método para incrementar la actividad generadora de la C α -formilglicina en una célula. El método involucra contactar a la célula con una molécula aislada de ácidos nucleicos del invento (por ejemplo, un ácido nucleico de la IDENTIFICACIÓN SECUENCIAL NÚMERO 1), o uno de sus productos de expresión (por ejemplo, un péptido de la IDENTIFICACIÓN SECUENCIAL NÚMERO 2), en un monto efectivo para incrementar la actividad generadora de C α -formilglicina en la célula. En secciones importantes, el método involucra la activación del gen endógeno de la FGE para incrementar la actividad generadora de C α -formilglicina en la célula. En algunas secciones, el contacto es localizado bajo condiciones que permiten la entrada de una molécula del invento a la célula.

El término "permitir la entrada" de una molécula a una célula tiene los siguientes significados dependiendo de la naturaleza de la molécula. Para un ácido nucleico aislado tiene el objetivo de describir la entrada del ácido nucleico a través de la membrana celular y en el núcleo de la célula, donde el "transgén de ácido nucleico" puede utilizar la maquinaria celular para producir polipéptidos funcionales codificados por el ácido nucleico. El término "transgén de ácido nucleico" tiene el propósito de describir a todos los ácidos nucleicos del invento, con o sin vectores asociados. Para un polipéptido, tiene el propósito de describir la entrada del polipéptido a través de la membrana celular y adentro del citoplasma celular, y si fuese necesario, la utilización de la maquinaria citoplásmica celular para modificar funcionalmente al polipéptido (por ejemplo, a una forma activa).

Varias técnicas pueden ser utilizadas para introducir ácidos nucleicos a las células, dependiendo de si los ácidos nucleicos son introducidos in vitro o in vivo en un anfitrión. Aquellas técnicas incluyen la transfección de precipitaciones del ácido nucleico- CaPO₄, la transfección de ácidos nucleicos asociados con la DEAE, la transfección con un retrovirus incluyendo al ácido nucleico de interés, la transfección mediada por liposomas, y similares.

Para ciertos usos, es preferible dirigir al ácido nucleico a células específicas. En aquellas instancias, un vehículo utilizado para la entrega de un ácido nucleico del invento a una célula (por ejemplo, un retrovirus, u otro virus; un liposoma) puede tener una molécula de dirección adherida a este.

Por ejemplo, una molécula tal como un anticuerpo específico a una proteína de la superficie de la membrana en la célula objetivo o un ligando para un receptor en la célula objetivo puede enlazarse a, o incorporarse dentro de, el portador de entrega del ácido nucleico.

Por ejemplo, en los casos en los que se utilizan a liposomas para entregar a los ácidos nucleicos del invento, proteínas que se enlazan a una proteína de la superficie de la membrana con endocitosis podría incorporarse a la formulación de un liposoma para dirigir y/o facilitar la absorción.

Aquellas proteínas incluyen a proteínas cápside o sus fragmentos trópicos para un tipo celular específico, anticuerpos para proteínas que experimentan una internalización en ciclos, proteínas que apuntan a una localización intracelular y que mejoran la vida media intracelular, y similares. Los sistemas de entrega polimérica también han sido exitosos para entregar a ácidos nucleicos a células, tal como es conocido por aquellas personas con conocimiento en la industria. Aquellos sistemas incluso permiten una entrega oral de ácidos nucleicos.

Tal como se describió anteriormente, este invento permite la provisión de una célula productora de sulfatasas donde la tasa de sulfatasas activas en comparación a las sulfatasas producidas (es decir, la actividad específica) por la célula es incrementada.

La célula puede comprender: (i) una sulfatasa con una expresión incrementada, y (ii) una enzima generadora de formilglicina con una expresión incrementada, donde la tasa de la sulfatasas activas en comparación a la sulfatasa

total producida por la célula es incrementada en por lo menos el 5% sobre la tasa de la sulfatasas activas en comparación de las sulfatasas totales producidas por la célula en la ausencia de la enzima generadora de formilglicina.

- 5 Una "sulfatasa con una expresión incrementada", tal como se utiliza en este documento, comúnmente se refiere a una expresión incrementada de una sulfatasa y/o su polipéptido codificado en comparación a un control.

La expresión incrementada se refiere a incrementar (es decir, en una magnitud detectable) la replicación, la transcripción y/o la interpretación de cualquiera de los ácidos nucleicos de sulfatasas (los ácidos nucleicos de sulfatasas tal como son descritos en otras secciones de este documento), puesto que la regulación-incremento de cualquiera de estos procesos resulta en un incremento de concentración/monto del polipéptido codificado por el gen (ácido nucleico). Esto puede lograrse utilizando a varios métodos conocidos en la industria, también descritos en otras secciones de este documento, tal como la transfección de una célula con una ADNc y/o un ADN genómico de una sulfatasa que abarca el locus de la sulfatasa, activar el gen endógeno de la sulfatasa al colocar, por ejemplo, un elemento promotor fuerte corriente arriba del locus genómico del gen endógeno de la sulfatasa utilizando una recombinación homóloga (refiérase, por ejemplo, a la tecnología de activación genética descrita en detalle en las patentes de Estados Unidos números 5,733,761, 6,270,989, y 6,565,844), etcétera. Un control típico sería una célula idéntica transfectada con un plásmido o varios plásmidos vectores. Mejorar (o incrementar) la actividad de la sulfatasas también se refiere a prevenir o inhibir la degradación de las sulfatasas (por ejemplo, por medio de una ubiquitinación incrementada), regulación-reducción, etcétera, resultando, por ejemplo, en una molécula ty2 sulfatasa incrementada o estable (vida media) cuando se compara con un control. La regulación-reducción o expresión reducida se refiere a la expresión reducida de un gen y/o su polipéptido codificado. La regulación-incremento o regulación-reducción de la expresión genética puede determinarse directamente al detectar un incremento o una reducción, respectivamente, en el nivel de ARNm para el gen (por ejemplo, una sulfatasa), o el nivel de expresión proteínica del polipéptido que codifica al gen, utilizando cualquier medio adecuado conocido en la industria, tal como la hibridación de ácidos nucleicos o métodos para la detección de anticuerpos, respectivamente, en comparación con los controles. La regulación-incremento o la regulación-reducción de la expresión genética de la sulfatasas también puede determinarse indirectamente al detectar un cambio en la actividad de la sulfatasas.

Asimismo, una "enzima generadora de formilglicina con una expresión incrementada", tal como se utiliza en este documento, se refiere, comúnmente, a una expresión incrementada de un ácido nucleico de la FGE y/o su polipéptido codificado en comparación con un control. La expresión incrementada se refiere a incrementar (es decir, en una magnitud detectable) la replicación, la transcripción y/o la interpretación de cualquiera de los ácidos nucleicos de la FGE (tal como se describe en otras secciones de este documento), puesto que la regulación-incremento de cualquiera de estos procesos resulta en incrementar la concentración/monto del polipéptido codificado por el gen (ácido nucleico). Esto puede lograrse utilizando los métodos descritos anteriormente (para las sulfatasas), y en otras secciones de este documento.

En ciertas secciones, la tasa de las sulfatasas activas en comparación de las sulfatasas totales producidas por la célula se incrementa en por lo menos el 10%, el 15%, el 20%, el 50%, el 100%, el 200%, el 500%, el 1000%, sobre la tasa de sulfatasas activas en comparación de la sulfatasas totales producidas por la célula en la ausencia de la enzima generadora de formilglicina.

Puede facilitarse un método mejorado para tratar una deficiencia de sulfatasas en un sujeto. El método involucra la administración a un sujeto en necesidad de aquel tratamiento de una sulfatasa en un monto efectivo para tratar la deficiencia de sulfatasas en el sujeto, donde se contacta a la sulfatasa con la enzima generadora de formilglicina en un monto efectivo para incrementar la actividad específica de las sulfatasas. Tal como se describió en otras secciones de este documento, la "actividad específica" se refiere a la tasa de las sulfatasas activas en comparación a las sulfatasas totales producidas. El término "contactado" tal como se utiliza en este documento, se refiere a la FGE que modifica post - interpretación a las sulfatasas tal como se describe en otras secciones de este documento. Sería aparente para una persona con conocimiento normal en la industria que la FGE puede contactar a una sulfatasa y modificarla si los ácidos nucleicos que codifican a la FGE y una sulfatasa están siendo co- expresadas en una célula, o incluso si un polipéptido aislado de la FGE contacta a un polipéptido aislado de la sulfatasa in vivo o in vitro. Aunque un polipéptido aislado de la FGE pueda ser co- administrado con un polipéptido aislado de la sulfatasa a un sujeto para tratar una deficiencia de sulfatasas en el sujeto, es preferible que el contacto entre la FGE y la sulfatasa ocurra in vitro antes de la administración de la sulfatasa al sujeto. Este método mejorado de tratamiento es beneficioso a un sujeto puesto que montos más pequeños de la sulfatasa necesitan administrarse, con menor frecuencia, puesto que la sulfatasa tiene una actividad específica superior.

El invento será mejor y más entendido haciendo referencia a los siguientes ejemplos que no deben ser considerados como limitantes de la cobertura del invento. (Si cualquier ejemplo o cualquiera de sus partes no están dentro del enfoque de las reivindicaciones estos son proveídos para propósitos informativos).

Ejemplos

Ejemplo 1:

La deficiencia de sulfatasas múltiples es causada por mutaciones en los genes que codifican a la enzima generadora de Ca-formilglutamina

Procedimientos Experimentales

Materiales y Métodos

Ensayo in vitro para la FGE

Para monitorear la actividad de las FGEs, el péptido P23 23mer N-acetilado y C-amidado (MTDFYVPVSLCTPSRAALLTGRS) (IDENTIFICACIÓN SECUENCIAL NÚMERO: 33) fue utilizado como sustrato. La conversión del residuo de cisteína en la posición 11 de la FGly fue monitoreada por medio de una espectrometría de masa MALDI-TOF. Se preparó una solución de almacenamiento de 6 μ M de P23 en un 30% de acetonitrilo y 0.1 por ciento de ácido trifluoroacético (TFA - trifluoroacetic acid). Bajo condiciones estándar se incubaron 6 pmol de P23 a 37 °C con hasta 10 μ l de enzimas en un volumen final de 30 μ l de 50 mM de Tris/HCl, pH 9.0, que contenían 67 mM de NaCl, 15 mM de CaCl₂, 2 mM de DTT, y 0.33 mg/ml de albúmina sérica bovina. Para detener a la reacción enzimática se agregaron 1.5 microlitros de un 10% de TFA. Se enlazó entonces al P23 al ZipTip C18 (Millipore), se enjuagó con un 0.1 por ciento de TFA y se colocó en 3 μ l de un 50% de acetonitrilo, 0.1 por ciento de TFA. 0.5 microlitros del eluido se mezcló con 0.5 microlitros de la solución matriz (5 μ g/mililitros de ácido a-ciano-4-hidroxi-cinámico (Bruker Daltonics, Billerica, MA) en un 50% de acetonitrilo, 0.1 por ciento de TFA) en un objetivo de acero inoxidable. Se realizó una espectrometría de masa MALDI-TOF con un Reflex III (Bruker Daltonics) utilizando una modalidad de reflectrón y energía láser justo por encima del límite de desorción/ionización. Todos los espectros fueron promedios de 200-300 imágenes de varios lugares del objetivo. El eje de masa fue calibrado utilizando péptidos de masas moleculares que variaban desde 1000 a 3000 Da como estándares externos. El MH⁺ monoisotópico de P23 es 2526.28 y el producto que contiene a la FGly es de 2508.29. La actividad (pmol del producto/h) fue calculada en base a la altura pico del producto dividida para la suma de las alturas pico de P23 y del producto.

Purificación de la FGE proveniente de testículos bovinos

Se obtuvieron testículos bovinos del matadero local y se almacenaron hasta por 20 horas en hielo. El parénquima fue liberado del tejido conectivo y homogeneizado en una licuadora y por 3 rondas con un motor de alfarería. Se realizó la preparación de los microsomas brutos (RM - rough microsomes) por medio de un fraccionamiento celular de la homogeneización obtenida tal como fue descrito (Meyer et al., J. Biol. Chem., 2000, 275:14550-14557) con las siguientes modificaciones. 3 pasos diferenciales de centrifugación, 20 minutos cada uno a 4 °C, fueron realizados a 500 g (rotor JA10), 3 mg (JA10) y 10 mg (JA10). A partir de la última precipitación, las membranas RM fueron sedimentadas (125 mg, rotor Ti45, 45 minutos, 4 °C), homogeneizadas usando un motor de alfarería y divididas en capas en un colchón de sacarosa (50 mM de Hepes, pH 7.6, 50mM de KAc, 6 mM de MgAc₂, 1 mM de EDTA, 1.3 M de sacarosa, 5 mM β -mercaptoetanol). Los RMs fueron recuperados del pellet después de rotar durante 210 minutos a 4500 rpm en un rotor Ti45 a 4 °C. Usualmente entre 10000-15000 RM equivalentes, tal como se define por Walter y Blobel (Methods Enzymol (Métodos en Enzimológicos), 1983, 96:84-93), se obtuvieron de 1 kg de tejido de testículos. El reticuloplasma, es decir, el contenido luminoso del RM, fue obtenido por medio de extracción diferencial a concentraciones bajas de deoxi Big Chap (Gran Agrietamiento), tal como se ha descrito (Fey et al., J. Biol. Chem., 2001, 276:47021-47028). Para la purificación de la FGE, se dializaron 95 ml de reticuloplasma durante 20 horas a 4 °C en contra de 20mM de Tris/HCl, pH 8.0, 2.5 mM de DTT, y se despejaron por medio de centrifugación en 125000 g durante una hora. Porciones de 32 ml del reticuloplasma despejado fueron cargadas en una columna MonoQ HR10/10 (Amersham Biosciences, Piscataway, NJ) a la temperatura del cuarto, se lavaron y se eluyeron a 2 ml/minuto con un gradiente lineal de cero a 0.75 M de NaCl en 80 ml del amortiguador Tris. Las fracciones que contenían a la actividad FGE, eluyéndose a 50-165 mM de NaCl, de 3 ejecuciones fueron agrupadas (42 ml) y mezcladas con 2 ml de concanavalina A-sefarosa (Amersham Biosciences) que se habían lavado con 50 mM de amortiguador Hepes, pH 7.4, que contenía 0.5 M de KCl, 1mM de MgCl₂, 1 mM de MnCl₂, 1 mM de CaCl₂, y 2.5 mM de DTT. Después de una incubación durante 16 horas a 4 °C, la concanavalina A-Sefarosa fue recolectada en una columna y lavada con 6 ml del mismo amortiguador Hepes. El material enlazado fue eluido al incubar a la columna durante una hora a la temperatura del cuarto con 6 ml de 0.5 M de a-metilmanósido en 50 mM de Hepes, pH 7.4, 2.5 mM de DTT. La elución fue repetida con 4 ml del mismo eluyente. Los eluidos combinados (10 ml) provenientes de concanavalina A-Sefarosa fueron ajustados a un pH de 8.0 con 0.5 M de Tris/HCl, pH 9.0, y mezclados con 2 ml de Affigel 10 (Bio-Rad Laboratories, Hercules, CA) que se derivó con 10 mg de péptidos mezclados (PVSLPTRSCAALLTGR) (IDENTIFICACIÓN SECUENCIAL NÚMERO: 34) y lavada con el amortiguador A (50mM de Hepes, pH 8.0, que contenía 0.15 M de acetato de potasio, 0.125 M de sacarosa, 1 mM de MgCl₂, y 2.5 mM de DTT). Después de una incubación durante 3 horas a 4 °C la matriz de afinidad fue recolectada en una columna. El flujo y una fracción de lavado con 4 ml del amortiguador A fueron recolectados, combinados y mezclados con 2 ml de Affigel 10 que había sido sustituido con 10 mg del péptido Ser69 (PVSLSTPSRAALLTGR) (IDENTIFICACIÓN SECUENCIAL NÚMERO: 35) y lavado con el amortiguador A punto después de su incubación durante la noche a 4 °C, la matriz de afinidad fue recolectada en una columna, lavada 3 veces con 6 ml del amortiguador B (el amortiguador A que contenía 2 M de NaCl y una mezcla de los 20 aminoácidos proteinogénicos,

cada uno 50 mg/mililitro). El material enlazado fue eluido a partir de la matriz de afinidad al incubarlo al Affigel 2 veces durante 90 minutos cada uno con 6 ml del amortiguador B que contenía 25 mM del péptido Ser69. Una porción del eluido fue sustituida con 1 mg/mililitro de albúmina sérica bovina, dializadas en contra del amortiguador A y analizada para detectar su actividad. La parte restante de la actividad (11.8 mililitros) se concentró en un concentrador Vivaspín 500 (Vivascience AG, Hannover, Alemania), y se solubilizó a 95 °C en un amortiguador de muestra de Laemmli SDS. La composición de polipéptidos del material y de las preparaciones de inicio obtenidas después de los pasos de cromatografía fue monitoreada por medio de SDSPAGE (15% de acrilamida, 0.16 por ciento de bisacrilamida) y manchadas con SYPRO Ruby (Bio-Rad Laboratories).

Identificación de la FGE por medio de espectrometría de masa

Para un análisis de huellas peptídicas, los polipéptidos purificados fueron digeridos en gel con tripsina (Shevchenko et al., Anal. Chem., 1996,68:850-855), se desalaron en C18 ZipTip y fueron analizados por medio de una espectrometría de masa MALDI-TOF utilizando ácido dihidrobenzóico como matriz y 2 péptidos autolíticos de tripsina (m/z 842.51 y 2211.10) como estándares internos. Para el conjunto un análisis de espectrometría de masa seleccionó péptidos para una espectrometría de masa de decaimiento post-fuente MALDI-TOF. Sus iones correspondientes cargados doblemente fueron aislados y fragmentados por medio de una espectrometría de masa de trampa iónica nano-ESI fuera de línea. Los datos de las espectrometrías de masa fueron utilizados por el algoritmo de búsqueda Mascot (Mascota) para una identificación proteínica en la base de datos de proteínas NCBI y la base de datos de nucleótidos NCBI EST.

Informática biológica

Los péptidos de señalización y los lugares de división fueron descritos con el método de von Heijne (von Heijne, Nucleic Acids (ácidos nucleicos) Res., 1986, 14:4683-90) implementados en EMBOSS (Rice et al., Trends in Genetics (tendencias en genética), 2000, 16:276-277). Lugares de N-glicosilación fueron pronosticados utilizando el algoritmo de Brunak (Gupta y Brunak, Pac. Symp. Biocomput., 2002, 310-22). Se detectaron dominios funcionales usando los modelos PFAM-Hidden(escondido)-Markov (versión 7.8) (Sonnhammer et al., Nucleic Acids (Ácidos Nucleicos) Res., 1998, 26:320-322). Para buscar a homólogos de la FGE, las bases de datos del National Center for Biotechnology Information (Centro Nacional para Información Biotecnológica) (Wheeler et al., Nucleic Acids (Ácidos Nucleicos) Res., 2002, 20:13-16) fueron consultados con BLAST (Altschul et al., Nucleic Acids (Ácidos Nucleicos) Res., 1997, 25:3389-3402). Las similitudes secuenciales fueron calculadas utilizando herramientas estándar de EMBOSS. La organización y la ubicación de los lugares genómicos se determinaron utilizando los recursos genómicos de humanos y de ratones de NCBI y el mapa de homología de humanos-ratones también conforman a NCBI, Bethesda, MD).

Clonación de ADNc de la FGE humana

ARN completo, preparado de fibroblastos humanos utilizando el mini botiquín RNEASY™ (Qiagen, Inc., Valencia, CA) fue transcrito en forma reversa utilizando el botiquín OMNISCRIPT RT™ (Qiagen, Inc., Valencia, CA) y un cebador de oligo(dT) o el cebador específico de la FGE 1199nc (CCAATGTAGGTCAGACACG) (IDENTIFICACIÓN SECUENCIAL NÚMERO: 36). La primera cepa de ADNc fue amplificada por medio de PCR utilizando al cebador delantero 1c (ACATGGCCCGCGGGAC) (IDENTIFICACIÓN SECUENCIAL NÚMERO: 37) y, como cebador en reversa, a 1199nc o a 1182nc (CGACTGCTCCTTGACTGG), (IDENTIFICACIÓN SECUENCIAL NÚMERO: 38). Los productos PCR fueron clonados directamente al vector pCR4-TOPO™ (Invitrogen Corporation, Carlsbad, CA). Al secuenciar varios de los productos PCR clonados, que fueron obtenidos de varios individuos y de reacciones independientes de RT y PCR, la secuencia codificadora de la ADNc de la FGE fue determinada (IDENTIFICACIONES SECUENCIALES NÚMEROS: 1 y 3).

Detección de mutación, secuencia genómica, mutagénesis dirigida a lugares y análisis de Northern blot

Protocolos estándar utilizados en este estudio fueron esencialmente como se describen en Lübke et al. (Nat. Gen., 2001, 28:73-76) y Hansske et al. (J. Clin. Invest., 2002, 109:725-733). Los Northern blots fueron hibridados con una sonda de ADNc cubriendo toda la región codificadora y una sonda de ADNc β-actuadora como un control para cargar al ARN.

Líneas celulares y cultivo celular

Los fibroblastos de los pacientes de MSD 1-6 fueron obtenidos de E. Christenson (Rigshospitalet Copenhagen), M. Beck (Universitätskinderklinik Mainz), A. Kohlschütter (Universitätskrankenhaus Eppendorf, Hamburgo), E. Zammarchi (Meyer Hospital, University of Florence), K. Harzer (Institut für Hirnforschung, Universität Tübingen), y A. Fensom (Guy's Hospital, Londres), respectivamente. Los fibroblastos de piel humana, las células HT-1080, BHK21 y CHO fueron mantenidas a 37 °C bajo un 5% de CO₂ en el medio Eagle (águila) modificado de Dulbecco que contenía un 10% de suero fetal de becerro.

Transfección, inmunofluorescencia indirecta, análisis de Western blots y detección de la actividad FGE

La ADNc de la FGE fue equipada con un lugar 5' EcoRI y con una secuencia de marcación 3' HA-, c-Myc o RGS-His6, seguido por un codón de parada y un lugar HindIII, por medio de un PCR de alargamiento utilizando una polimerasa Pfu (Stratagene, La Jolla, CA) y los siguientes cebadores: GGAATTCGGGACAACATGGCTGCG (EcoRI) (IDENTIFICACIÓN SECUENCIAL NÚMERO: 39), CCCAAGCTTATGC GTAGTCAGGCACATCATACGGATAGTCCATGGTGGGCAGGC(HA) (IDENTIFICACIÓN SECUENCIAL NÚMERO: 40), CCCAAGCTTACAGGTCTTCTTCAGAAATCAGCTTITGTTTCGTCCATGGTGGGCAG GC (c-Myc) (IDENTIFICACIÓN SECUENCIAL NÚMERO: 41), CCCAAGCTTAGTGATGGTGATGGTGATGCGATC CTCTGTCCATGGTGGGCAGGC (RGS-His6) (IDENTIFICACIÓN SECUENCIAL NÚMERO: 42). Los productos resultantes de PCR fueron clonados como fragmentos de EcoRI/HindIII a pMPSVEH (Artelt et al., Gene, 1988, 68:213-219). Los plásmidos obtenidos fueron transfectados transitoriamente a células HT-1080, BHK21 y CHO cultivadas en fundas cobertores, utilizando EFFECTENE™ (Qiagen) como un reactivo de transfección. 48 horas después de la transfección, las células fueron analizadas por medio de inmunofluorescencia indirecta tal como se describió previamente (Lübke et al., Nat. Gen., 2001, 28:73-76; Hansske et al., J. Clin. Invest., 2002, 109:725-733), utilizando anticuerpos monoclonales IgG1 en contra de HA (Berkeley Antibody Company, Richmond, CA), cMyc (Santa Cruz Biotechnology, Inc., Santa Cruz, CA) o RGS-His (Qiagen), anticuerpos primarios. La proteína marcadora del retículo endoplásmico, la proteína disulfuro-isomerasa (PDI) fue detectada con un anticuerpo monoclonal de un subtipo diferente (IgG2A, Stressgen Biotech., Victoria BC, Canadá). Los anticuerpos primarios fueron detectados con anticuerpos secundarios de cabra de isotipos específicos acoplados a CY2 o CY3, respectivamente (Molecular Probes, Inc., Eugene, OR). Imágenes de inmunofluorescencia fueron obtenidos en un microscopio de detección láser Leica TCS Sp2 AOBS. Para un análisis de Western blots los mismos anticuerpos monoclonales y un IgG anti ratón conjugado con HRP como un anticuerpo secundario fueron utilizados. Para la determinación de la actividad FGE, se lavaron a células tripnizadas con una solución salina amortiguada con fosfato que contenía una mezcla de inhibidores de proteasas (208 µM de hidrocloreuro de fluoruro de sulfonilo de 4-(2- aminoetil)benceno, 0.16 µM de aprotinina, 4.2 µM de leupeptina, 7.2 µM de bestatina, 3 µM de peptastina A, 2.8 µM de E-64), solo utilizado en 10 mM de Tris, pH 8.0, que contenía 2.5 mM de DTT, los inhibidores de proteasas y 1% de Tritón X-100, y se despejó por medio de centrifugación a 125 mg durante una hora. El sobrenadante fue expuesto a cromatografía en una columna MonoQ PC 1.6/5 utilizando las condiciones descritas anteriormente. Las fracciones que se estaban eluyendo a 50-200 mM de NaCl fueron agrupadas, liofilizadas y reconstituidas en un décimo del volumen de agrupación original antes de su determinación de la actividad FGE con el péptido P23.

Transducción retroviral

Los ADNcs de interés fueron clonados en los vectores pLPCX y pLNCX2 que se basan en el virus de leucemia de ratones de Moloney (BD Biosciences Clontech, Palo Alto, CA). La transfección de las células ecotrópicas FNX-Eco (ATCC, Manassas, VA) y la transducción de las células anfotrópicas RETROPACK™ PT67 (BD Biosciences Clontech) y los fibroblastos humanos se realizó tal como fue descrito (Lübke et al., Nat. Gen., 2001, 28:73-76; Thiel et al., Biochem. J., 2002, 376, 195-201). Para algunos experimentos las células PT67 transducidas con pLPCX fueron seleccionadas con puomicina antes de la determinación de las actividades de las sulfatasas.

Ensayos de sulfatasas

La actividad de ASA, STS y GalNAc6S fue terminada tal como se describió en Rommerskirch y von Figura, Proc. Natl. Acad. Sci., Estados Unidos de América, 1992, 89:2561-2565; Glössl y Kresse, Clin. Chim. Acta, 1978, 88:111-119.

Resultados

Un ensayo rápido que se basa en péptidos para detectar la actividad de la FGE

Hemos desarrollado un ensayo para determinar la actividad FGE en extractos de microsomas utilizando como sustratos a fragmentos [³⁵S] ASA sintetizados in vivo. Los fragmentos fueron agregados a la mezcla del ensayo como complejos de cadenas emergentes asociadas con ribosomas. La cuantificación del producto incluyó la digestión crítica, la separación de los péptidos por medio de RP-HPLC y la identificación y cuantificación del FGly etiquetado como [³⁵S] que contiene a péptidos tripticos por medio de una combinación de una derivación química a hidrazonas, la separación RP-HPLC y un recuento de centelleo líquido (Fey et al., J. Biol. Cliem., 2001, 276:47021-47028). Para monitorear la actividad de la enzima durante su purificación, este procedimiento complicado necesitó ser modificado. Un péptido 16mer sintético correspondiente a los residuos ASA 75-80 y que contenía el motivo secuencial requerido para la formación de FGly inhibió a la actividad FGE en el ensayo in vitro. Esto sugirió que péptidos tales como ASA65-80 podrían servir como sustratos para FGE. Sintetizamos al 23mer péptido P23 (IDENTIFICACIÓN SECUENCIAL NÚMERO: 33), que corresponde a los residuos ASA 60-80 con un residuo de metionina N-acetilada y de serina C-amidada para proteger a las terminales N y C, respectivamente. La cisteína y la FGly que contenía formas de P23 pudieron ser identificadas y cuantificadas por una espectrometría de masa del tiempo de fuga de la desorción/ionización láser asistida por una matriz (MALDI-TOF - matrixassisted laser desorption/ionisation time of flight). La presencia del residuo de FGly en la posición 11 del P23 fue verificada por medio de una espectrofotometría de masa de post-decaimiento de la fuente MALDI-TOF (refiérase a Peng et al., J. Mass Spec., 2003, 38:80-86). La incubación de P23 con extractos de microsomas de páncreas bovinos o testículos

bovinos tuvo una conversión de hasta el 95% del péptido a un derivado que contenía a FGly (figura 1). Bajo condiciones estándar, la reacción fue proporcional al monto de enzimas y al tiempo de incubación. Alrededor del 50% de sustrato fue consumido y el período de incubación no excedió las 24 horas. El K_m para P23 fue de 13 nM. Los efectos del glutatión oxidado, el Ca^{2+} y el pH fueron comparables con aquellos vistos en el ensayo utilizando como sustratos a complejos de cadenas emergentes asociadas con ribosomas (Fey et al., J. Biol. Chem., 2001, 276:47021-47028).

Purificación de las FGEs

Para purificación de la FGE, la fracción soluble (el reticuloplasma) de microsomas de testículos bovinos sirvió como el material inicial. La actividad específica de la FGE fue de 10-20 veces más alta que aquella en el retículo-plasma de microsomas de páncreas bovinos (Fey et al., J. Biol. Chem., 2001, 276:47021-47028). La purificación de la FGE se logró por medio de una combinación de 4 pasos de cromatografía. Los primeros 2 pasos fueron la cromatografía en un intercambiador de aniones MonoQ y en la Concanavalina A-Sefarosa. A un pH de 8, la actividad de la FGE se enlazó al MonoQ y se eluyó a 50-165 mM de NaCl con una recuperación del 60-90%. Cuando esta fracción fue mezclada con concanavalina A-Sefarosa, la FGE fue enlazada. Del 30-40% de la actividad inicial pudo ser eluida con 0.5 M de manósido de α -metilo. Los 2 pasos de purificación finales fueron una cromatografía en derivaciones de matrices de afinidad con los péptidos 16mer. La primera matriz de afinidad fue Affigel 10 sustituida con una variación del péptido ASA65-80, en la cual los residuos Cys69, Pro71 y Arg73, críticos para la formación de FGly, fueron mezclados (péptido mezclado PVSLPTRSCAALLTGR -IDENTIFICACIÓN SECUENCIAL NÚMERO: 34). Éste péptido no inhibió la actividad FGE cuando se agregó a una concentración de 10 mM en el ensayo in vitro y, cuando fue inmovilizado a Affigel 10, no retuvo la actividad de la FGE. Cromatografía en la matriz de afinidad del péptido mezclado removió las proteínas de enlace del péptido incluyendo a las personas del retículo endoplásmico. La 2ª matriz de afinidad fue Affigel 10 sustituida con una variación del péptido ASA65-80, en el cual el Cys69 fue reemplazado por una serina (péptido de Ser69 PVSLSTPSRAALLTGR -IDENTIFICACIÓN SECUENCIAL NÚMERO: 35). La matriz de afinidad del péptido Ser69 enlazó eficientemente a la FGE. La actividad FGE pudo ser eluida con 2M de KSCN o 25 mM del péptido Ser69 con una recuperación del 20-40%. Antes de la determinación de la actividad el KSCN o el péptido Ser69 tuvo que ser removido por medio de diálisis. La sustitución de Cys69 por serina fue crucial para la elución de la FGE activa. Affigel 10 sustituido con el péptido de tipo silvestre ASA65-80 enlazó a la FGE eficientemente. Sin embargo, casi ninguna actividad pudo ser recuperada en eluidos con sales caotrópicas (KSCN, $MgCl_2$), péptidos (el péptido ASA65-80 o Ser69) o amortiguadores con un pH bajo o alto. En la figura 2, se muestra el patrón de los polipéptidos del material de inicio y de las fracciones activas obtenido después de los 4 pasos de cromatografía de una purificación típica. En la fracción final del 5% de la actividad FGE de inicio y el 0.0006 por ciento de la proteína de inicio fueron recuperadas (una purificación de 8333 veces).

Los polipéptidos purificados de 39.5 y 41.5 kDa son codificados en un solo gen

Los polipéptidos de 39.5 y 41.5 kDa en la preparación purificada de la FGE fueron sujetos a un análisis de huellas peptídicas. El espectro de masa de los péptidos tripticos de los 2 polipéptidos obtenidos por medio de una espectrometría de masa MALDI-TOF fueron superpuestos en una gran magnitud, sugiriendo que las 2 proteínas se originan del mismo gen. Entre los péptidos tripticos de ambos polipéptidos (2 péptidos abundantes MH^+ 1580.73, SQNTPDSSASNLGFR (IDENTIFICACIÓN SECUENCIAL NÚMERO: 43), y MH^+ 2049.91, MVPIPVAGVFTMGTDTPQIK -IDENTIFICACIÓN SECUENCIAL NÚMERO: 44 con 2 oscilaciones de metionina) fueron encontrados, que coincidieron con la proteína codificada por una ADNc con la cuenta del GenBank (Banco Genético) número AK075459 (IDENTIFICACIÓN SECUENCIAL NÚMERO: 4). La secuencia de aminoácidos de los 2 péptidos fue confirmada por el espectro post-decaimiento de la fuente MALDI-TOF y por medio de un análisis MS/MS utilizando una espectrometría de masa de una trampa iónica de ionización por medio de un nano-electro aerosol (ESI - electrospray ionisation). Una secuencia EST del ortólogo bovino de la ADNc humana que cubre a la parte de la terminal de C de la FGE y que coincide con las secuencias de ambos péptidos suministro información secuencial adicional para la FGE bovina.

Conservación evolucionaria y estructura del dominio de la FGE

El gen para la FGE humana es codificada por la ADNc de (las IDENTIFICACIONES SECUENCIALES NÚMEROS: 1 y/o 3) y ubicada en el cromosoma 3p26. Se expande por ~105 kb y la secuencia codificadora es distribuida por 9 exones. 3 ortólogos del gen de la FGE humana se encontraron en el ratón (87% de identidad), en la *Drosophila melanogaster* (48% de identidad), y en la *Anopheles gambiae* (47% identidad). Las secuencias de EST ortólogas se encuentran para 8 especies adicionales incluyendo a la vaca, al cerdo, al *Xenopus laevis*, al *Silurana tropicalis*, a la cebra, al pez, al salmón y a otras especies de peces (para detalles refiérase al ejemplo 2). La estructura de exones-intrones entre el gen de los humanos y de los ratones es conservado en el gen de los ratones en el cromosoma 6E2 que está ubicado dentro de la región sintética en el cromosoma humano 3p26. Los genomas de *S. cerevisiae* y de *C. elegans* no tienen los homólogos de la FGE. En procariotas 12 homólogos de la FGE humana fueron encontrados. La ADNc para la FGE humana se pronostica que codifica a una proteína de 374 residuos (la figura 3 y la IDENTIFICACIÓN SECUENCIAL NÚMERO: 2). La proteína contiene una secuencia divisible de señalización de 33 residuos, que indica la translocación de la FGE en el retículo endoplásmico, y contiene un solo lugar de N-glicosilación en Asn141. Los enlaces de la FGE a la concanavalina A sugieren que este lugar de N-glicosilación es

utilizado. Los residuos 87-367 de la FGE están listados en la base de datos de motivos proteínicos PFAM como un dominio de una función desconocida (PFAM: DUF323). Un análisis de comparación secuencial de la FGE humana y sus ortólogos procariotas identificados en las bases de datos indican que este dominio es compuesto de 3 sub - dominios diferentes.

Es un dominio de la terminal de N (los residuos 91-154 en la FGE humana) tiene una identificación secuencial del 46% y una similitud del 79% dentro de los 4 ortólogos de la FGE eucariota conocidos. En la FGE humana, este dominio porta el lugar de N- glicosilación en el Asn 141, que es conservado en otros ortólogos. La parte media de la FGE (los residuos 179-308 en la FGE humana) es representada por un sub - dominio rico en triptófano (12 triptófanos por 129 residuos). La identidad de los ortólogos procariotas dentro de este sub - dominio es el 57%, la similitud es del 82%. El sub - dominio de la terminal C (residuos 327-366 en la FGE humana) es la secuencia más altamente conservada dentro de la familia de las FGEs. La identidad secuencial del sub - dominio de la terminal C humana con los ortólogos eucariotas (3 secuencias de longitud completa y 8 ESTs) es del 85%, la similitud es del 97%. Dentro de los 40 residuos del sub - dominio 3, 4 residuos de cisteína son conservados completamente. 3 de las cisteínas también son conservadas en los ortólogos FGE procariotas. Los 12 miembros procariotas de la familia de la FGEs (para detalles refiérase al ejemplo 2) comparten la estructura del sub - dominio con FGEs procariotas. Los límites entre los 3 sub - dominios son más evidentes en la familia de las FGEs procariotas debido a que las secuencias no conservadas de longitudes variables que separan a los sub - dominios entre sí. Los genomas de humanos y de ratones codifican 2 homólogos que se relacionan de cerca a la FGE (IDENTIFICACIONES SECUENCIALES NÚMEROS: 43 y 44, cuenta del GenBank (Banco Genético) número NM_015411, en el hombre, y las IDENTIFICACIONES SECUENCIALES NÚMEROS: 45 y 46, cuenta del GenBank (Banco Genético) número AK076022, en el ratón). Los 2 parálogos son 86% idénticos. Sus genes están ubicados en regiones sintéticas cromosómicas (7q11 en humanos, 5G1 en ratones). Los 2 parálogos comparten con los ortólogos de la FGE la estructura del sub - dominio y son 35% idénticos y tienen un 47% de similitud a la FGE humana. En el 3^{er} sub - dominio, que es 100% idéntico en ambos homólogos, falta la cisteína que contiene a una secuencia de un oligómero de 11 subunidades (undecamer) del sub - dominio 3.

Expresión, localización sub - celular y formas moleculares

Una sola transcripción de 2.1 kb es detectable por el análisis de Northern blot del ARN total de los fibroblastos de la piel y de poli A+ ARN del corazón, del cerebro, de la placenta, de los pulmones, del hígado, del músculo esquelético, del riñón y del páncreas. En relación al ARN de β -actina la abundancia varía por una orden de magnitud y es más alta en el páncreas y en el riñón y más baja en el cerebro. Varias líneas celulares procariotas o que expresan transitoriamente a la ADNc de la FGE humana o de derivados de la FGE extendidos en la terminal C por una marcación HA-, Myc- o His6 fueron puestos en ensayos para detectar su actividad FGE y su localización sub - celular de la FGE. La expresión transitoria de la FGE marcada y no marcada incrementó la actividad de la FGE en unas 1.6-3.9 veces. La expresión estable de la FGE en las células PT67 incrementó la actividad de la FGE en alrededor de 100 veces. La detección de la forma marcada de la FGE por una inmunofluorescencia indirecta en células BHK 21, CHO, y HT1080 mostró una co - localización de varias formas de la FGE marcadas con isomerasas de proteindisulfuro, una proteína luminosa del retículo endoplásmico. Un análisis de Western blot de los extractos de las células BHK 21 transfectadas transitoriamente con la ADNc que codifica a formas marcadas de la FGE mostró una sola banda inmuno - reactiva con un tamaño aparente entre los 42 y los 44 kDa.

El gen de la FGE porta mutaciones en MSD

La MSD es causada por una deficiencia para generar residuos de *FGly* en la sulfatasas (Schmidt, B., et al., Cell (Célula), 1995, 82:271-278). El gen de la FGE es, por lo tanto, un gen candidato para la MSD. Amplificamos y secuenciamos a la ADNc que codifica a la FGE de 7 pacientes de MSD y encontramos 10 mutaciones diferentes que fueron confirmadas al secuenciar al ADN genómico (tabla 1).

Tabla 1: mutaciones en pacientes de MSD

Mutación	Efecto en la Proteína	Observaciones	Paciente
1076C>A	S359X	Truncamiento de 16 residuos de la terminal C	1*
IVS3+5-8 del	Eliminación de los residuos 149-173	Eliminación en el marco del exón 3	1,2
979C>T	R327X	Pérdida del sub - dominio 3	2
1045C>T	R349W	Sustitución de un residuo conservado en el sub - dominio 3	3.7
1046G>A	R349Q	Sustitución de un residuo conservado en el sub - dominio 3	4
1006T>C	C336R	Sustitución de un residuo conservado en el sub - dominio 3	4
836C>T	A279V	Sustitución de un residuo conservado en el sub - dominio 2	5
243delC	Cambio de marco y truncamiento	Pérdida de todos los 3 sub - dominios	5
661delG	Cambio de marco y truncamiento	Pérdida de la terminal C de la FGE incluyendo a sub - dominio 3	6**
IVS6-1G>A	Eliminación de los residuos 281-318	Eliminación en el marco del exón 7	5
*El paciente 1 es el paciente con MSD en Schmidt, B., et al., Cell (célula), 1995, 82:271-278 y Rommerskirch y von Proc. Natl. Acad. Sci., Estados Unidos de América, 1992, 89:2561-2565. Figura, **El paciente 6 es el paciente con MSD reportado por Burk et al., J. Pediatr., 1984, 104:574-578. Los otros pacientes representan casos no publicados.			

Para confirmar la deficiencia de las FGEs como la causa de la inactividad de sulfatasas sintetizadas en MSD, expresamos a la ADNc de las FGE en fibroblastos MSD utilizando una transferencia genética retroviral. Como un control, traducimos al vector retroviral sin una inserción de la ADNc. Para monitorear la complementación del defecto metabólico de la actividad de ASA, de la sulfatasa esteroide (STS - steroid sulfatase) y de la 6-sulfatasa N-acetilgalactosamina (GalNAc6S - N-acetylgalactosamine 6-sulfatase) fueron medidas en los fibroblastos transducidos antes o después de la selección. La transducción de la FGE de tipo silvestre restauró parcialmente la actividad catalítica de las 3 sulfatasas en 2 líneas celulares de MSD (tabla 2) y para STS en una 3ª línea celular de MSD. Debe tomarse en cuenta que para ASA y GalNAc6S la restauración fue únicamente parcial después de la selección de los fibroblastos reaccionando a un 20-50% de su actividad normal. Para STS, se descubrió que la actividad fue restaurada a los niveles de los fibroblastos de control después de la selección. La selección incrementó la actividad de ASA y de STS por un 50 a 80% lo cual es compatible con la observación mencionada anteriormente de que un 15 a 50% de los fibroblastos se vuelven transducidos (Lübke et al., Nat. Gen., 2001, 28:73-76). Las actividades de las sulfatasas en los fibroblastos de MSD transducidos con el vector retroviral individualmente (tabla 2) fueron comparables a aquellos en los fibroblastos de MSD que no fueron transducidos (no se muestra). La transducción de la ADNc de la FGE que porta a la mutación IVS3+5-8del no pudo restaurar las actividades de la sulfatasas (tabla 2).

Tabla 2: Complementación de los fibroblastos de MSD por medio de una transducción de la ADNc de la FGE de tipo silvestre o mutante

Fibroblastos	Inserción FGE	Sulfatasa		
		ASA1	STS1	Go]NAc6S1
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 control		58 ± 11	66 ± 31	828 ± 426
¹ los valores dan una tasa de entre ASA (mU/miligramos de proteína celular), STS (μU/miligramos de proteína celular), GalNAc6S (μU/mg de proteína celular) y aquella de β-hexosaminidasa (U/mg de proteína celular). Para los fibroblastos de control la media y la variación de las líneas celulares 6-11 está dada. En los casos en que se indica, el rango de 2 cultivos traducidos en paralelo es dado para los fibroblastos de MSD. ▪ El número de fibroblastos de MSD se refiere a aquel del paciente en la tabla 1. + Determinación de la actividad previo a la selección. ++ Determinación de la actividad después de la selección. n.d.: No fue determinado				

Debate**La FGE es una glicoproteína altamente conservada del retículo endoplásmico**

La purificación de la FGE de testículos bovinos produjo 2 polipéptidos de 39.5 y 41.5 kDa que se originaron del mismo gen. La expresión de 3 versiones marcadas diferentemente de la FGE en 3 diferentes líneas celulares eucariotas como una sola forma sugiere que una de las 2 formas observadas en la preparación purificada de la FGE proveniente de testículos bovinos pudo haber sido generada por una proteólisis limitada durante la purificación. La sustitución de Cys69 en el péptido ASA65-80 por la serina fue crítico para la purificación de la FGE por medio de cromatografía de afinidad. La FGE tiene una secuencia de señalización divisible que media la translocación a lo largo de la membrana del retículo endoplásmico. La mayor parte de la proteína madura (275 residuos de 340) define un dominio único, el cual muy posiblemente está compuesto de 3 sub - dominios (refiérase al ejemplo 2), puesto que ninguno de los 3 homólogos de los dominios existe en proteínas con una función conocida. El reconocimiento del motivo linear de la modificación de la *FGly* en polipéptidos de sulfatasas recientemente sintetizados (Dierks et al., EMBO J., 1999, 18:2084-2091) podría ser la función de un sub - dominio de la FGE. El dominio catalítico podría catalizar la formación de la *FGly* en algunas formas. Se ha propuesto que los electrones abstractos de la FGE del grupo tiol de la cisteína y los transfiere a un aceptador. El tialdehído resultante se hidrolisaría espontáneamente a la *FGly* y H₂S (Schmidt, B., et al., Cell (célula), 1995, 82:271-278). Alternamente, la FGE podría actuar como una oxigenasa de función mixta (monoxigenasa) introduciendo un átomo de O₂ a la cisteína y el otro en H₂O con la ayuda de un electrón donante tal como el FADH₂. El resultado del hidrato de tialdehído derivado de la cisteína fraccionaria espontáneamente a *FGly* and H₂S. Experimentos preliminares con una preparación de la FGE parcialmente purificada mostraron una dependencia crítica de la formación de *FGly* en el oxígeno molecular. Esto sugeriría que la FGE actúa como una oxigenasa de función múltiple. La conservación particular alta del sub - dominio 3 y la presencia de 3 residuos conservados de cisteína allí presentes hacen a este sub - dominio un posible candidato para el lugar catalítico. Será interesante ver si los elementos estructurales que median el reconocimiento del motivo *FGly* y los enlaces de un aceptador de electrones o de un donante de electrones tienen una correlación con la estructura del dominio de la FGE.

La FGE recombinante está ubicada en el retículo endoplásmico, que es compatible con el lugar propuesto de su acción. Los residuos de *FGly* son generados en sulfatasas recientemente sintetizadas durante el transcurso o un poquito después de su translocación al retículo endoplásmico (Dierks et al., Proc. Natl. Acad. Sci. Estados Unidos de América, 1997, 94:11963-11968; Dierks et al., FEBS Lett., 1998, 423:61-65). La FGE en sí no contiene una señal de retención de ER del tipo KDEL. Su retención en el retículo endoplásmico podría, por lo tanto, ser mediada por la interacción con otras proteínas ER. Componentes de la maquinaria de translocación/ N- glicosilación son candidatos atractivos para aquellos compañeros de interacción.

Las mutaciones en las FGEs causan la MSD

Hemos demostrado que las mutaciones en el gen que codifica a la FGE causan la MSD. La FGE también podría interactuar con otros componentes, y defectos en los genes que la codifican y podrían causar igualmente a la MSD. En 7 pacientes con MSD, encontramos, en efecto, 10 mutaciones diferentes en el gen de la FGE. Todas las mutaciones tenían efectos severos en la proteína FGE al reemplazar residuos altamente conservados en el sub - dominio 3 (3 mutaciones) o en el sub - dominio 2 (una mutación) o mutaciones de la terminal C de varias longitudes (4 mutaciones) o grandes eliminaciones adentro del marco (2 mutaciones). Para las líneas celulares de MSD y en una de las mutaciones de MSD se demostró que la transducción del tipo silvestre, pero no de la ADNc de la FGE mutante, restauró parcialmente las actividades de la sulfatasas. Esto identifica claramente al gen de la FGE como el lugar de mutación y la naturaleza causante de la enfermedad de la mutación. MSD es clínicamente y bioquímicamente heterogénea. Se ha diferenciado una forma rara neonatal que se presenta en el nacimiento y desarrolla una hidrocefalia, una forma común que parece inicialmente una leucodistrofia metacromática infantil y subsiguientemente se desarrollaron características similares a la ictiosis y a la mucopolisacaridosis, y una forma leve menos frecuente en la cual las características clínicas de una mucopolisacaridosis. Bioquímicamente, es característico que una actividad residual de sulfatasas puede detectarse, que para la mayoría de casos en fibroblastos cultivados de la piel es por debajo del 10% de los controles (Burch et al., Clin. Genet., 1986, 30:409-15; Basner et al., Pediatr. Res., 1979, 13:1316-1318). Sin embargo, en algunas líneas celulares de MSD la actividad de sulfatasas seleccionadas pueden alcanzar un rango normal (Yutaka et al., Clin. Genet., 1981, 20:296-303). Además, se ha reportado que la actividad residual es sujeto a variaciones dependiendo de las condiciones de los cultivos celulares y de factores desconocidos. Bioquímicamente, la MSD ha sido clasificada en 2 grupos. En el grupo 1, la actividad residual de la sulfatasas es por debajo del 15% incluyendo a aquella de ASB. En el grupo 2, la actividad residual de la sulfatasas es mayor y particularmente aquella de ASB podría alcanzar valores de hasta 50-100% del control. Todos los pacientes que se reportaron aquí caen en el grupo I excepto el paciente 5, que cae en el grupo II (actividad ASB en el rango del control) del fenotipo bioquímico. Basándose en los criterios clínicos, los pacientes 1 y 6 son casos neonatales, mientras que los pacientes 2-4 y 7 tienen el MSD mientras que el paciente 5 tiene una forma similar a la forma de mucopolisacaridosis de la MSD.

La homogeneidad fenotípica sugiere que las diferentes mutaciones en los pacientes MSD son asociadas con diferentes actividades residuales de la FGE. Datos preliminares de las células PT67 que expresan establemente a

FGE IVS3+5-8del indican que la eliminación en el marco del exón 3 elimina completamente a la actividad de las FGEs. Esta caracterización de las mutaciones en la MSD, de las propiedades bioquímicas de la FGE mutante y el contenido residual de FGly en la sulfatasas utilizando un método de espectrometría de masa altamente sensible que ha sido desarrollado recientemente (Peng et al., J. Mass Spec., 2003, 38:80-86) suministrará un mejor entendimiento de la correlación entre el genotipo-fenotipo en la MSD.

Ejemplo 2:

El gen de la FGE humana define una nueva familia genética que modifica a la sulfatasas que es conservada de procariotas a eucariotas

Información biológica

Los péptidos de señalización y los lugares de división fueron descritos con el método de von Heijne (Nucleic Acids (Ácidos Nucleicos) Res., 1986, 14:4683) implementado en EMBOSS (Rice et al., Trends in Genetics (Tendencias en Genética), 2000, 16:276-277), y el método de Nielsen et al. (Protein Engineering (Ingeniería de Proteínas), 1997, 10:1-6). Los lugares de N-glicosilación fueron pronosticados utilizando el algoritmo de Brunak (Gupta y Brunak, Pac. Symp. Biocomput., 2002, 310-22).

Se detectaron los dominios funcionales usando los modelos PFAM-Hidden(Escondido)-Markov (Versión 7.8) (Sonnhammer et al., Nucleic Acids (Ácidos Nucleicos) Res., 1998, 26:320-322). Las secuencias de la siembra PFAM DUF323 fueron obtenidas de TrEMBL (Bairoch, A. and Apweiler, R., Nucleic Acids (Ácidos Nucleicos) Res., 2000, 28:45-48). Múltiples alineaciones y construcciones de árboles filogenéticos fueron realizadas con Clustal W (Thompson, J., et al., Nucleic Acids (Ácidos Nucleicos) Res., 1994, 22:4673-4680). Para la computación de los árboles filogenéticos, se excluyeron las posiciones de vacíos y múltiples sustituciones fueron corregidas. La construcción de los árboles fue realizada para obtener resultados significativos. Los árboles fueron visualizados utilizando Njplot (Perriere, G. y Gouy, M., Biochimie, 1996, 78:364-369). Las alineaciones fueron dibujadas utilizando el comando pret- typlot de EMBOSS.

Para buscar los homólogos de la FGE, las bases de datos NR, NT and EST del National Center for Biotechnology Information (NCBI-Centro Nacional de Información Biotecnológica) (Wheeler et al., Nucleic Acids (Ácidos Nucleicos) Res., 2002, 20:13-16), fueron consultadas con BLAST (Altschul et al., Nucleic Acids (Ácidos Nucleicos) Res., 1997, 25:3389-3402). Para las secuencias proteínicas, la búsqueda fue realizada utilizando el Psi-Blast convergente iterativo contrastándolo con la versión actual de la base de datos NR usando un corte de valor esperado de 10^{-40} , y los parámetros predeterminados. La convergencia fue alcanzada después de 5 iteraciones. Para las secuencias de los nucleótidos, la búsqueda fue realizada con Psi-Tblastn: usando NR y la secuencia proteínica de la FGE humana como dato ingresado, una matriz de puntaje para hFGE fue construida con Psi-Blast de convergencia iterativa. Esta matriz fue utilizada como dato ingresado para blastall para consultar a las bases de datos de nucleótidos NT y EST. Para ambos pasos, se utilizó un corte de valor esperado de 10^{-20} .

Una predicción de estructura secundaria proteínica fue realizada utilizando Psipred (Jones, D., J Mol Biol., 1999, 292:1950-202; McGuffin, L., et al., Bioinformatics (informática biológica), 2000, 16:404-405).

Puntajes similares de los dominios de alineaciones fueron computados utilizando el algoritmo cons de EMBOSS con parámetros predeterminados. Las alineaciones meta fueron generadas al alinear secuencias de consenso de los grupos de la familia de la FGE. La organización y localización del locus genómico fueron determinadas utilizando los recursos de genoma humanos y de ratones de NCBI en NCBI (Bethesda, MD) y la localización de humano-ratón-rata de Softberry (Mount Kisco, NY). Las secuencias de genoma bacteriano fueron descargadas del servidor NCBI-FTP. La anotación genómica microbiana de NCBI fue utilizada para obtener una visión general del locus genómico de los genes bacterianos de la FGE.

Resultados y Debate

Características y motivos básicos de la FGE humana y sus proteínas relacionadas

El gen de la FGE humana (IDENTIFICACIONES SECUENCIALES NÚMEROS: 1, 3) codifica a la proteína de la FGE (IDENTIFICACIÓN SECUENCIAL NÚMERO: 2) que se pronostica que tiene 374 residuos. Una señal de división entre los residuos 22-33 (un puntaje de Heijne de 15.29) y un puntaje de hidroterapia (Kyte, J. y Doolittle, R., J Mol Biol., 1982, 157:105-132) de los residuos 17-29 entre 1.7 y 3.3 indica que los 33 residuos de la terminal N están divididos después de la translocación de ER. Sin embargo, con el algoritmo de Nielsen et al. (Protein Engineering (Ingeniería de Proteínas), 1997, 10:1-6), la división de la frecuencia de señalización es pronosticada después del residuo 34. La proteína tiene un solo lugar potencial de N-glicosilación en Asn 141.

Una búsqueda con la secuencia proteínica de la FGE en contraste con la base de datos de motivos proteínicos de PFAM (Sonnhammer et al., Nucleic Acids (Ácidos Nucleicos) Res., 1998, 26:320-322) reveló que los residuos 87-367 de la FGE humana pueden clasificarse como el dominio proteínico DUF323 ("dominio de una función

desconocida", PF03781) con un valor esperado altamente significativo de $7.9 \cdot 10^{-114}$. La semilla PFAM que define a DUF323 consiste de 25 secuencias proteínicas, de las cuales la mayoría son proteínas hipotéticas derivadas de datos de secuenciación. Para analizar la relación entre la FGE humana y DUF323, se realizó una alineación múltiple de la FGEs con las secuencias de la semilla DUF323. Basándose en esto, un árbol filogenético fue construido y completado. 4 de las secuencias hipotéticas (TrEMBL-IDs Q9CK12, Q9I761, Q94632 y Q9Y405) tuvieron una divergencia tan fuerte de los otros miembros de la siembra que evitaron una construcción exitosa y tuvieron que ser removidas del conjunto. La figura 2 muestra el árbol construido mostrando la relación entre la FGE humana y las 21 proteínas de semillas restantes de DUF323. El árbol puede ser utilizado para subdividir a los miembros de la siembra en 2 categorías: homólogos relacionados muy de cerca con la FGE humana y el resto, los genes menos relacionados.

Las 7 proteínas superiores tienen una distancia filogenética de entre 0.41 y 0.73 a la FGE humana. Contienen un solo dominio, DUF323. La homología dentro de este grupo se extiende sobre toda la secuencia de aminoácidos, la mayor parte de la cual consiste del dominio DUF323. El dominio DUF323 es fuertemente conservado dentro de este grupo de homólogos, mientras que las otras 15 proteínas de la siembra son menos relacionadas a la FGE humana (la distancia citogenética es de entre 1.14 y 1.93). El sub - dominio DUF323 diverge considerablemente del dominio altamente conservado DUF323 del primer grupo (sección cf. "Sub - dominios de la FGE y mutaciones del gen de la FGE"). La mayoría de estas 15 proteínas son hipotéticas, 6 de ellas han sido investigadas más. Una de ellas, una quinasa de serina/treonina (TrEMBL:O84147) de *C. trachomatis* contiene otros dominios adicionalmente a DUF323: un dominio de enlaces ATP y un dominio de quinasas. Las secuencias de la *R. sphaeroides* (TrEMBL: Q9ALV8) y la *Pseudomonas* sp. (TrEMBL: 052577) codifican a la proteína NirV, un gen co- transcrito con la reductasa de nitrato que contienen cobre nirK (Jain, R. y Shapleigh, J., *Microbiology* (Microbiología), 2001, 147:2505-2515). CarC (TrEMBL: Q9XB56) es una oxigenasa involucrada en la síntesis de un antibiótico P- lactámico de la *E. carotovora* (McGowan, S., et al., *Mol Microbial* (Microbiana), 1996, 22:415-426; Khaleeli N, T. C., y Busby RW, *Biochemistry* (Bioquímica), 2000, 39:8666-8673). XylR (TrEMBL: 031397) y BH0900 (TrEMBL: Q9KEF2) son proteínas mejoradoras de enlaces involucradas en la regulación de la utilización de la pentosa (Rodionov, D., et al., *FEMS Microbiol Lett.*, 2001, 205:305-314) en la *bacillaceae* y la *clostridiaceae*. La comparación de la FGE y del DUF323 conllevó al establecimiento de un umbral homológico diferenciador de la familia de las FGEs de los homólogos distantes que contienen DUF323 con funciones diferentes. Estos incluyen una quinasa de serina/treonina y XylR, un mejorador de transcripciones, así como la FGE, una enzima generadora de FGly y CarC, una oxigenasa. Tal como se mencionó en otras secciones de este documento, la FGE podría ejercer además su función calificadora de cisteína como una oxigenasa, sugiriendo que los miembros de la FGE y los no miembros de la FGE de las semillas DUF323 podrían compartir una función de oxigenasa.

Homólogos de la FGE

La presencia de los homólogos relacionados de cerca a la FGE humana en las semillas DUF323 nos orientaron para buscar homólogos de la FGE humana en la base de datos NR de NCBI (Wheeler et al., *Nucleic Acids* (Ácidos Nucleicos) *Res.*, 2002, 20:13-16). El umbral de la búsqueda fue escogido de tal forma que todos los 6 homólogos presentes en la siembra de DUF323 y otros homólogos con una relación muy cercana fueron obtenidos sin encontrar los otros miembros de la siembra. Esta búsqueda conllevó a la identificación de 3 ortólogos de la FGE en eucariotas, 12 ortólogos en procariotas y 2 parálogos en el hombre y en el ratón (tabla 3).

Tabla 3: La familia genética de la FGEs en eucariotas y procariotas

IDENTIFICACIONES SECUENCIALES NÚMEROS: NA, AA [GI]	ESPECIES	LONGITUD [AA]	SUBGRUPO
1/3, 2	Homo sapiens	374	E1
49, 50 [22122361]	Mus musculus	372f	E1
51, 52 [20130397]	Drosophila melanogaster	336	E1
53, 54 [21289310]	Anopheles gambiae	290	E1
47, 48 [26344956]	Mus musculus	308	E2
45, 46 [24308053]	Homo sapiens	301	E2
55, 56 [21225812]	Streptomyces coelicolor A3(2)	314	P1
7, 58 [25028125]	Corynebacterium efficiens YS-314	334	P1
59, 60 [23108562]	Novosphingobium aromaticivorans	338	P2
61, 62 [13474559]	Mesorhizobium loti	372	P2
63, 64 [22988809]	Burkholderia fungorum	416	P2
65, 66 [16264068]	Sinorhizobium meliloti	303	P2
67, 68 [14518334]	Microscilla sp.	354	P2
69, 70 [26990068]	Pseudomonas putida KT2440	291	P2
71, 72 [22975289]	Ralstonia metallidurans	259	P2
73, 74 [23132010]	Prochlorococcus marinus	291	P2
75, 76 [16125425]	Caulobacter crescentus CB 15	338	P2
77, 78 [15607852]	Mycobacterium tuberculosis Ht37Rv	299	P2
GI- identificador proteínico del GenBank (Banco Genético protein identifier) NA- ácido nucleico (nucleic acid) AA – aminoácidos E1 - ortólogos eucariotas E2 – parálogos eucariotas P1 - ortólogos procariotas relacionados de cerca P2 - otros ortólogos procariotas f- secuencia proteínica pronosticada erróneamente en el GenBank (Banco Genético)			

Nótese que la secuencia de ratón GI 22122361 es pronosticada en el Gen Bank (Banco Genético) para codificar una proteína de 284 aa, aunque la secuencia de la ADNc NM 145937 codifica a una proteína de 372 residuos. Este pronóstico erróneo se basa en la omisión del primer exón del gen de la FGE de ratones. Todas las secuencias encontradas en la base de datos de NR son de eucariotas o procariotas de más alto nivel. Los homólogos de la FGE no fueron detectados en las arqueobacterias o las plantas. Búsquedas con umbrales aún más bajos en genomas completamente secuenciados de las *C. elegans* y *S. cerevisiae* y las bases de datos ORF relacionadas no revelaron ningún homólogo. Una búsqueda en las secuencias eucariotas de la base de datos de nucleótidos NT y EST conllevó a la identificación de 8 ESTs ortólogos adicionales de la FGE con fragmentos secuenciales de ADNc de terminal de 3' que muestran un nivel más alto de concentración en el nivel proteínico que no están listados en la base de datos NR. Estas secuencias no abarcan toda la parte calificadora de las ARNms y todas son de eucariotas más elevadas (tabla 4).

Tabla 4: Fragmentos EST ortólogos de la FGE en eucariotas

IDENTIFICACIONES SECUENCIALES NÚMEROS: NA [GB]	ESPECIES
80 [CA379852]	Oncorhynchus mykiss
81 [AI721440]	Danio rerio
82 [BJ505402]	Oryzias latipes
83 [BJ054666]	Xenopus laevis
84 [AL892419]	Silurana tropicalis
85 [CA064079]	Salmo salar
86 [BF189614]	Sus scrofa
87 [AV609121]	Bos taurus
GB - acceso al GenBank (Banco Genético) número; NA - ácido nucleico (nucleic acid)	

Alineaciones y construcciones múltiples de un árbol filogenético (utilizando ClustalW) de las secuencias codificadoras de la base de datos NR permitieron la definición de 4 subgrupos de homólogos: ortólogos eucariotas (FGEs humanos, de ratón, de mosquito y de mosca de frutas), parálogos eucariotas (parálogos de la FGE de humanos y de ratones), ortólogos procariotas relacionados de cerca a la FGE (Streptomyces y Corynebacterium) y otros ortólogos procariotas (Caulobacter, Pseudomonas, Mycobacterium, Prochlorococcus, Mesorhizobium, Sinorhizobium, Novosphingobium, Ralstonia, Burkholderia, y Microscilla). Los ortólogos eucariotas muestran una identidad general en contraste con las FGE humanas de un 87% (ratón), un 48% (mosca de fruta) y un 47% (el anopheles). Aunque los ortólogos de la FGE se encuentran en procariotas y en eucariotas de más alto nivel, estas no están presentes en los genomas completamente secuenciados de eucariotas de más bajo nivel situados genéticamente entre el *S. cerevisiae* y *D. melanogaster*. Adicionalmente, los homólogos de la FGE están ausentes

en los genomas completamente secuenciados del *E. Colli* y del pez globo.

Tal como se mencionó en otras secciones de este documento, los parálogos de la FGE encontrados en humanos y en ratones pueden tener una menor capacidad generadora de *FGly* y contribuyen a las actividades residuales de la sulfatasas encontradas en los pacientes de MSD.

Sub - dominios de la FGE

Los miembros de la familia genética de las FGEs tienen 3 partes/dominios altamente conservados (tal como se describió en otras secciones de este documento). Adicionalmente a las 2 secuencias no conservadas que separan a las que se acaban de mencionar, estas tienen extensiones no conservadas en las terminales N y C. Las 3 partes conservadas se consideran que representan sub - dominios del dominio DUF323 porque están separadas por partes no conservadas de longitudes que varían. La longitud de la parte que separa a los dominios 1 y 2 varía de entre 22 y 29 residuos y aquella que separa a los dominios 2 y 3 de entre 7 a 38 aminoácidos. Las partes no conservadas de las terminales N y C muestran una variación que es aún más grande en longitud (la terminal N: 0-90 AA, la terminal C: 0-28 AA). La secuencia para el gen de la FGE de *Ralstonia metallidurans* está probablemente incompleta puesto que le falta el primer sub - dominio.

Para verificar la factibilidad de definir los sub - dominios de DUF323, realizamos un pronóstico de la estructura secundaria de la proteína de la FGE humana usando Psipred. La señal hidrofóbica de ER (los residuos 1-33) se pronostica que contiene estructuras tipo hélice que conforman el pronóstico de señalización del algoritmo von-Heijne. La región no conservada de la terminal N (aa 34-89) y la región de separación entre los sub - dominios 2 y 3 (aa 308-327) contienen secciones enrolladas. La región separadora de los sub - dominios 1 y 2 contiene un rollo. La hélice α en los aa 65/66 tiene una seguridad baja de pronóstico y es probablemente un artefacto de pronóstico. Los bordes de los sub - dominios están situados dentro de rollos y no interrumpen a las hélices α ni a las hebras β . El primer sub - dominio está hecho de varias hebras β y una hélice α , el 2º sub - dominio contiene 2 hebras β y 4 hélices α . El 3º sub - dominio tiene una región de hélices α flanqueada por una lámina al inicio y al final del sub - dominio. En resumen, la estructura secundaria concuerda con la estructura y sub - dominios propuestos puesto que los límites de los sub - dominios están ubicados dentro de rollos y los sub - dominios contienen a los elementos estructurales de hélices α y hebras β .

Debe tomarse en cuenta que ninguno de los sub - dominios existe como un módulo aislado en secuencias listadas en la base de datos. Dentro de cada uno de los 4 subgrupos de la familia de las FGEs, los sub - dominios son altamente conservados, con el 3º sub - dominio mostrando la homología más alta (tabla 5). Este sub - dominio muestra además la homología más fuerte a lo largo de los subgrupos.

Tabla 5: Fomología (similitud porcentual) de los sub - dominios de la familia de la FGEs

Subfamilia	Miembros	Sub - dominio		
		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 eucariotas; E2 - parálogos eucariotas P1 - ortólogos procariotas relacionados de cerca; P2 - otros ortólogos procariotas				

El primer sub - dominio de la familia de las FGEs muestra la homología más débil a lo largo de los subgrupos. En los ortólogos eucariotas porta al sitio de N- glicosilación: en el residuo Asn 141 en humanos, en el Asn 139 en los ratones y en el Asn 120 en la mosca de la fruta. En el anofeles, no se encuentra aspargina en el residuo 130 homólogo al Asn 120 del *D. melanogaster*. Sin embargo, un cambio de los nucleótidos crearía un lugar de N- glicosilación en el Asn 130 del anofeles. Por lo tanto, la secuencia que abarca al residuo 130 necesita ser secuenciada nuevamente. El 2º sub - dominio rico en triptófanos con 12 Trp en 129 residuos de la FGE humana. 10 de los cuales están conservados en la familia de las FGEs.

La alta conservación del sub - dominio 3: sub - dominio 3 entre los ortólogos eucariotas son 100% similares y 90% idénticos. La importancia del 3º sub - dominio para la función de la proteína es resaltada por la observación que este sub - dominio es un lugar importante para mutaciones que causan enfermedades en los pacientes de MSD. 7 de 9 mutaciones identificadas en 6 MSD pacientes descritos en el ejemplo 1 están ubicadas en las secuencias que codifican a los 40 residuos del sub - dominio 3. Los residuos contienen 4 cisteína as, 3 de las cuales están conservadas entre los ortólogos procariotas y eucariotas. Los 2 parálogos eucariotas muestran la homología más baja de los otros miembros de la familia de la FGEs, por ejemplo, no tienen 2 de las 3 cisteínas conservadas del sub - dominio 3. Las características conservadas entre las secuencias de sub - dominio 3 de los ortólogos y de los parálogos son el motivo inicial RVXXGG(A)S (IDENTIFICACIÓN SECUENCIAL NÚMERO: 79), un heptámetro que contiene 3 argininas (los residuos 19-25 de la secuencia de consenso del sub - dominio) y el motivo GFR terminal. Una comparación con el dominio DUF323 de las 15 secuencias de siembra que no son homólogos cercanos de la

FGE, muestra diferencias secuenciales marcadas: las 15 secuencias de siembra tienen un primero y un 2º sub - dominio menos conservado, aunque la estructura general de sub - dominio también es visible. El sub - dominio 3, que es fuertemente conservado en la familia de la FGEs, es más corto y tiene una homología significativamente más débil en relación al sub - dominio 3 eucariota (similamente a alrededor del 20%) en comparación a los miembros procariotas de la familia de la FGE (una similitud de alrededor del 60%). Por lo que no tienen todos los residuos conservados de cisteína del sub - dominio 3. Las únicas características conservadas son el motivo inicial RVXXGG(A)S (IDENTIFICACIÓN SECUENCIAL NÚMERO: 79) y el motivo GFR terminal.

Organización genómica del gen de la FGE de humanos y ratones

El gen FGE humano está ubicado en el cromosoma 3p26. Abarca 105 kb y 9 exones para la secuencia interpretada. El gen de la FGE de ratones tiene una longitud de 80 Kb y está ubicado en el cromosoma 6E2. Los 9 exones del gen de la FGE de ratones tienen casi el mismo tamaño que los exones humanos (figura 3). Diferencias importantes entre el gen de humanos y el gen de ratones son la conservación más baja del 3'-UTR en el exón 9 y la longitud del exón 9, que es de 461 bp más larga en el gen de los ratones. El segmento 6E2 del cromosoma 6 de los ratones es altamente sinténico para el segmento cromosómico humano 3p26. Hacia el telómero, ambos el locus humano y el de ratones están flanqueados por los genes que codifican para LMCD1, KIAA0212, ITPR1, AXCAM, y IL5RA. En la dirección centromérica, ambos lugares de la FGE están flanqueados por el locus de CAV3 y OXTR.

Organización genómica de los genes de las FGEs procariotas

En los procariotas, las sulfatasas son clasificadas como sulfatasas de tipo cisteína o serina dependiendo del residuo que está siendo convertido a *FGly* en su centro activo (Miech, C., et al., J Biol Chem., 1998, 273:4835-4837; Dierks, T., et al., J Biol Chem., 1998, 273:25560-25564). En la *Klebsiella pneumoniae*, la *E. coli* y la *Yersinia pestis*, las sulfatasas de tipo serina son parte de un operón con *AtsB*, que codifica a la proteína citosólica que contiene motivos de agrupaciones de hierro-sulfuro y es crítico para la generación de *FGly* a partir de residuos de serina (Marquardt, C., et al., J Biol Chem., 2003, 278:2212-2218; Szameit, C., et al., J Biol Chem., 1999, 274:15375-15381).

Fue por lo tanto de interés examinar si los genes procariotas de la FGE están localizados en la proximidad de las sulfatasas de tipo cisteína que son los sustratos de la FGE. Entre los genes procariotas de la FGE mostrados en la tabla 3, 7 tienen genomas completamente secuenciados que permiten un análisis de las cercanías del locus de la FGE. En efecto, en 4 de los 7 genomas (*C. efficiens*: PID 25028125, *P. putida*: PID 26990068, *C. crescentus*: PID 16125425 y *M. tuberculosis*: PID 15607852) una sulfatasa de tipo cisteína se encuentra en cercanía directa de la FGE que es compatible con una co-transcripción de la FGE y de la sulfatasa. En 2 de ellas, (*C. efficiens* y *P. putida*), la FGE y la sulfatasa incluso tienen ORFs superpuestos, indicando fuertemente su co-expresión. Además, el vecindario genómico de los genes de la FGE y de las sulfatasas en 4 procariotas suministra evidencia adicional para asumir que las FGEs bacterianas son ortólogos funcionales.

El resto de los 3 organismos contienen sulfatasas de tipo cisteína (*S. coelicolor*: PID 24413927, *M. loti*: PID 13476324, *S. meliloti*: PIDs 16262963, 16263377, 15964702), sin embargo, los genes cercanos a las FGE en estos organismos no contienen una firma canónica de sulfatasas (Dierks, T., et al., J Biol Chem., 1998, 273:25560-25564) ni un dominio que pudiese indicar su función. En estos organismos la expresión de la FGE y de las sulfatasas de tipo cisteína es, por lo tanto, regulado muy posiblemente en tránsito.

Conclusiones

La identificación de la FGE humana cuya deficiencia causa la deficiencia de sulfatasas múltiples, enfermedad de almacenamiento de lisosomas transmitida recesivamente en una forma autosómica, permite la definición de una nueva familia genética que comprende a ortólogos de las FGEs provenientes de procariotas y eucariotas así como un parólogo en ratones y hombres. La FGE no se encuentra en genomas completamente secuenciados de *E. coli*, *S. cerevisiae*, *C. elegans* y *Fugu rubripes*. Adicionalmente, existe un vacío filogenético entre los procariotas y los eucariotas de más alto nivel con una falta de la FGEs en cualquier especie situada filogenéticamente entre los procariotas y el *D. melanogaster*. Sin embargo, algunos de estos eucariotas de un más bajo nivel, por ejemplo, el *C. elegans*, tienen genes de sulfatasas tipo cisteína. Esto señala la existencia de un 2º sistema generador de *FGly* que actúa en la sulfatasas tipo cisteína. Esta premisa es apoyada por la observación que el *E. coli*, que no tiene FGEs, puede generar *FGly* en sulfatasas tipo cisteína (Dierks, T., et al., J Biol Chem., 1998, 273:25560-25564).

Ejemplo 3:

La expresión de la FGE causa incrementos significativos en la actividad de sulfatasas en las líneas celulares que sobre expresan una sulfatasa

Deseamos examinar los efectos de la FGE en células que expresan/sobre-expresan una sulfatasa. Para este propósito, las células HT-1080 que expresan a la sulfatasas humanas, 2-sulfatasa de iduronato (I2S - Iduronate 2-Sulfatase) o 6-sulfatasa de N- Acetilgalactosamina (GALNS) fueron transfectadas por duplicado con una estructura

de expresión de la FGE, pXMG.1.3 (tabla 7 y figura 4) o un plásmido de control, pXMG.1.2 (FGE en orientación antisentido incapaz de producir FGEs funcionales, tabla 7). Las muestras en medios fueron cultivadas durante 24, 48 y 72 horas seguidas por 24 horas de un cambio de medio por post-electroporación. Las muestras del medio fueron probadas para detectar su actividad respectiva de sulfatasas por medio de un ensayo de actividad y un nivel proteínico total de sulfatasas estimado por Elisa específico para 2-sulfatasas de iduronato o sello sulfatasas de N-Acetilgalactosamina.

Tabla 6. Las Líneas Celulares Transfectadas que expresan sulfatasas utilizadas como sustratos de transfección

Cepa celular	Plásmido	Sulfatasa Expresada
36F	pXFM4A.1	6-sulfatasa de N-Acetilgalactosamina
30C6	pXI2S6	2-sulfatasa de iduronato

Tabla 7. FGEs y plásmidos de control utilizados para transfectar a las células HT-1080 que expresan a 2-sulfatasas de iduronato y 6-sulfatasas de N-Acetilgalactosamina

Plásmido	Configuración de elementos importantes de secuencias de ADN*
pXMG.1.3 (expresión de la FGE)	>1.6 kb CMV mejorador/promotor >1.1 kb de ADNc de la FGE > hGH3' secuencia no interpretada <amp < casete DHFR < casete Cdneo (fosfotransferasa de neomicina)
pXMG.1.2 (control, orientación reversa de la FGE)	>1.6 kb CMV mejorador/promotor >1.1 kb de ADNc de la FGE > hGH3' secuencia no interpretada <amp < casete DHFR < casete Cdneo (fosfotransferasa de neomicina)
* >Denota orientación 5' a 3'	

Procedimientos experimentales

Materiales y métodos

Transfección de células HT-1080 que producen 2-sulfatasa de iduronato y 6-sulfatasa de N-acetilgalactosamina

Las células HT-1080 fueron cultivadas para obtener $9-12 \times 10^6$ células para cada electroporación. 2 plásmidos fueron transfectados por duplicado: uno para ser probado (FGE) y un control; en este caso el plásmido de control contenía ADNc de la FGE clonado en orientación reversa con respecto al promotor CMV. Las células fueron centrifugadas a aproximadamente 1000 revoluciones por minuto durante 5 minutos. Las células fueron suspendidas en 1X PBS a 16×10^6 células/mililitro. Al fondo de la cubeta de electroporación, 100 µg de ADN plásmido fue agregado, 750 µl de suspensión celular (12×10^6 células) fueron agregados a la solución de ADN en la cubeta. Las células y el ADN se mezclaron suavemente con una pipeta de transferencia de plásmico, teniendo cuidado de no crear burbujas. Las células fueron electroporadas a 450 V, 250 µF (pulsador genético BioRad). La constante de tiempo fue registrada.

A las células electroporadas se les permitió permanecer sin estribaciones durante 10-30 minutos. 1.25 mililitros de DMEM/10% de suero de becerro fue agregado a cada cubeta, mezclado, y todas las células fueron transferidas a un frasco T 75 fresco que contenía 20 ml de DMEM/10. Después de 24 horas, el frasco fue re-alimentado con 20 ml de DMEM/10 para remover las células muertas. Después de 48-72 horas de la transfección, muestras del medio fueron recolectadas y las células fueron cultivadas de los frascos duplicados T75.

Preparación del medio

Células de 1 l de DMEM/10 (contiene: 23 ml de 2mM de glutamina L, 115 ml de suero de becerro) fueron transfectadas en un medio sin metotrexato (MTX). 24 horas después las células fueron re-alimentadas con un medio que contenía los montos apropiados de MTX (36F = 1.0 µM de MTX, 30C6 = 0.1M de MTX). El medio fue cultivado y las células fueron recolectadas 24, 48 y 72 horas después de la re-alimentación.

Ensayos de actividad

2-sulfatasa de iduronato (I2S - Iduronate 2-Sulfatase). Columnas de desalación NAP5 (Amersham Pharmacia Biotech AB, Uppsala, Suecia) fueron equilibradas con un amortiguador de diálisis (5 mM de acetato de sodio, 5 mM de tris, pH 7.0). La muestra que contenía I2S fue aplicada a la columna y se le permitió entrar en la cama. La muestra fue eluida en 1 ml de amortiguador de diálisis. Muestras desaladas fueron diluidas aún más a aproximadamente 100 ng/mililitros de I2S en un amortiguador de reacción (5 mM de acetato de sodio, 0.5

miligramos/litros de BSA, 0.1 por ciento de Tritón X-100, pH 4.5). 10 µl de cada muestra de I2S fue agregada a la fila superior de una placa Fluorométrica de 96 pozos (Perkin Elmer, Norwalk, CT) y pre-incubada durante 15 minutos a 37 °C. El sustrato fue preparado al disolver sulfato de 4-metil-umbelliferilo (Fluka, Buchs, Suiza) en un amortiguador de sustrato (5 mM de acetato de sodio, 0.5 miligramos/mililitros de BSA, pH 4.5) a una concentración final de 1.5 mg/mililitro. 100 µl de sustrato fueron agregados a cada pozo que contenía la muestra de I2S y la placa fue incubada durante una hora a 37 °C en la oscuridad. Después de su incubación 190 µl de amortiguador de parada (332.5 mM de lisina, 207.5 mM de carbonato de sodio, pH 10.7) se agregó a cada pozo que contenía la muestra. Se preparó 4-metilumbeliferona normal (4-MUF, Sigma, St. Louis, MO) como el producto estándar en agua de grado reactivo a una concentración final de 1 µM. 150 µl de 1 µM de 4-MUF normal y 150 µl de amortiguador de parada fueron agregados a un pozo de la fila superior en la placa. 150 µl de amortiguador de parada fueron agregados a cada pozo restante en la placa de 96 pozos. Diluciones seriales de 2 veces fueron hechas a partir de la fila superior de cada columna hasta la última fila de la placa. La placa fue leída en un analizador de micro platos de fusión universal (Packard, Meriden, CT) con una longitud de onda de un filtro de excitación de 330 nm y una longitud de onda de un filtro de emisión de 440 nm. Se generó una curva estándar de micromoles de 4-MUF normal vs. fluorescencia, y se extrapolaron muestras no conocidas a partir de esta curva. Los resultados son reportados como unidades/mililitros donde 1 U de actividad fue igual a 1 µmol de 4-MUF producido por minuto a 37 °C.

6-sulfatasa de N-Acetilgalactosamina (GALNS - N-Acetylglactosamine 6-Sulfatase). El ensayo de la actividad de GALNS aprovecha el sustrato fluorescente 4-metilumbelliferil-β-D-galactopiranosido-6-sulfato (Toronto Research Chemicals Inc., catálogo No. M33448). El ensayo fue conformado de 2 pasos. En el primer paso, 75 µl de 1.3 mM del sustrato preparado en reacción con el amortiguador (0.1M de acetato de sodio, 0.1 M de cloruro de sodio, pH 4.3) fue incubado durante 4 horas a 37 °C con 10 µl del medio/la muestra proteínica o sus diluciones correspondientes. La reacción fue detenida al agregar 5 µl de 2M de fosfato de sodio monobásico para inhibir la actividad de la GALNS. Después de la adición de aproximadamente 500 U de β-galactosidasa de *Aspergillus oryzae* (Sigma, Catalogue No. G5160), la mezcla de la reacción fue incubada a 37 °C durante una hora adicional para liberar la partícula fluorescente del sustrato. La 2ª reacción fue detenida al agregar 910 µl de una solución de parada (1% de glicina, 1% de carbonato de sodio, pH 10.7). La fluorescencia de la mezcla resultante fue medida utilizando una longitud de onda de medición de 359 nm y una longitud de onda referencial de 445 nm con 4-metilumbeliferona (sal de sodio de Sigma, Catalogue No. M1508) que sirvió como un estándar referencial. 1 U de la actividad corresponde a nmoles de la 4-metilumbeliferona liberada por hora.

Inmunoensayos (ELISA)

2-sulfatasa de iduronato (I2S). Una placa de fondo plano de 96 pozos fue cubierta con un anticuerpo anti-I2S monoclonal de ratón diluido a 10 µg/mililitros en 50 nM de bicarbonato de sodio, pH 9.26 durante una hora a 37 °C. El anticuerpo anti-I2S monoclonal de ratón fue desarrollado bajo contrato por Maine Biotechnology Services, Inc. (Portland, ME) a un polipéptido I2S humano de longitud completa producido recombinantemente y purificado utilizando tecnología estándar productora de hibridomas. La placa fue lavada 3 veces con 1X PBS que contenía 0.1 por ciento de Tween-20 y bloqueada durante una hora con un 2% de BSA en un amortiguador de lavado a 37 °C. El amortiguador de lavado con un 2% de BSA fue utilizado para diluir las muestras y los estándares. El estándar de I2S fue diluido y utilizado desde 100 ng/mililitros a 1.56 nanogramos/mililitros. Después de la remoción del amortiguador de bloqueo, las muestras y los estándares fueron aplicados a la placa e incubados durante una hora a 37 °C. El anticuerpo de detección, un anticuerpo anti-I2S de ratón conjugado con peroxidasa de rábano, se diluyó 0.15 microgramos/mililitros en un amortiguador de lavado con un 2% de BSA. La placa fue lavada 3 veces, detectando los anticuerpos agregados a la placa, y fue incubada durante 30 minutos a 37 °C. Para desarrollar la placa, se preparó sustrato de TMB (Bio-Rad, Hercules, CA). La placa fue lavada 3 veces, se agregaron 100 µl de sustrato a cada pozo y fue incubado durante 15 minutos a 37 °C. La reacción fue detenida con 2 N de ácido sulfúrico (100 µl/pozo) y la placa fue leída en un lector de placas de micro titulación a 45 nm, utilizando 655 nm como la longitud de onda referencial. **6-sulfatasa de N-Acetilgalactosamina (GALNS).** 2 anticuerpos anti-GALNS monoclonales suministraron la base para la ELISA de GALNS. Los anticuerpos anti-GALNS monoclonales de ratón también fueron desarrollados bajo contrato por Maine Biotechnology Services, Inc. (Portland, ME) a un polipéptido de GALNS humano de longitud completa producido recombinantemente y purificado utilizando tecnología estándar de producción de hibridomas. El primer anticuerpo, para la captura de GALNS fue utilizado para cubrir una placa F96 MaxiSorp Nunc-Immuno (Nalge Nunc, catálogo No. 442404) en un amortiguador de recubrimiento (50 mM de bicarbonato de sodio, pH. 6). Después de una incubación durante una hora a 37 °C y de lavarse con un amortiguador de lavado, la placa fue bloqueada con un amortiguador de bloqueo (PBS, 0.05% de Tween-20, 2% de BSA) durante una hora a 37 °C. Muestras experimentales y de control junto con los estándares de GALNS fueron cargados entonces en una placa e incubados aún más durante una hora a 37 °C. Después de lavarse con un amortiguador de lavado, el 2º anticuerpo de detección conjugado a HRP fue aplicado en un amortiguador de bloqueo seguido por 30 minutos de incubación a 37 °C. Después de lavarse la placa nuevamente, el reactivo del sustrato Bio-Rad TMB fue agregado e incubado durante 15 minutos. 2N de ácido sulfúrico fue agregado entonces para detener la reacción y los resultados fueron calificados utilizando espectrofotometría por medio de un dispositivo molecular lector de placas a una longitud de onda de 450 nm.

Debate

Efecto de la FGE en la actividad sulfatasa

GALNS. Un incremento de aproximadamente 50 veces en la actividad total de GALNS fue observada por sobre los niveles del control (figura 5). Este nivel de actividad incrementada fue observado con los 3 momentos de muestras del medio. Además, la actividad de GALNS fue acumulada linealmente a lo largo del tiempo con un incremento de 4 veces entre 24 y 48 horas y un incremento de 2 veces entre los momentos a las 48 horas y 72 horas.

I2S. Aunque de magnitud absoluta más pequeña, un efecto similar fue observado para la actividad total de I2S donde un incremento de aproximadamente 5 veces en la actividad total de I2S fue observado por sobre los niveles de control. Este nivel de actividad incrementada fue sostenido durante el experimento. La actividad de I2S acumulada linealmente en el medio a lo largo del tiempo, similar a los resultados vistos con GALNS (2.3 veces entre las 24 y las 48 horas, y 1.8 veces entre las 48 y las 72 horas).

Efecto de la FGE en la actividad específica de la sulfatasas

GALNS. La expresión de las FGE en las células 36F mejoró la actividad aparente específica de las GALNS (tasa de actividad enzimática en comparación de las enzimas totales estimadas por ELISA) por 40-60 veces sobre los niveles de control (figura 6). El incremento de la actividad específica fue sostenido a lo largo de 3 momentos en el estudio y pareció incrementarse a lo largo de 3 días de acumulación post-transfección.

I2S. Un efecto similar fue visto con I2S, donde un incremento de 6-7 veces en actividad específica (3-5 U/mg) fue observada por sobre los valores de control (0.5-0.7 U/mg).

Los valores ELISA para GALNS (figura 7) e I2S no fueron afectados significativamente por la transfección de la FGE. Esto indica que la expresión de FGE no deshabilita los senderos de interpretación y de secreción involucrados en la producción de sulfatasas.

En general, todos estos resultados para ambas sulfatasas indican que la expresión de FGE incrementa dramáticamente la actividad específica de sulfatasas en las líneas celulares que sobre-expresan a GALNS y a I2S.

Co- expresión de FGE (SUMF1) y otros genes de sulfatasas

Para probar el efecto de las FGE (SUMF1) en actividades adicionales de sulfatasas en células normales sobre -expresamos a las ADNc de ARSA (IDENTIFICACIÓN SECUENCIAL NÚMERO: 14), ARSV (IDENTIFICACIÓN SECUENCIAL NÚMERO: 18) y ARSE (IDENTIFICACIÓN SECUENCIAL NÚMERO: 22) en varias líneas celulares con y sin una co-transfección de la ADNc de FGE (SUMF1) y las actividades medidas de las sulfatasas. La sobreexpresión de ADNcs de sulfatasas en las células Cos-7 co-expresó en un incremento moderado de las actividades sulfatasas, y un incremento sinérgico importante (20 a 50 veces) en el gen de la sulfatasa y el gen de la FGE (SUMF1). Un efecto similar, aunque más bajo, fue observado en 3 líneas celulares adicionales, HepG2, LE293, y U2OS. Una sobreexpresión simultánea de ADNcs de sulfatasas múltiples resultó en un incremento más bajo de cada actividad de sulfatasas específica en comparación a la sobreexpresión de una sola sulfatasa, indicando la presencia de competición de las diferentes sulfatasas por la maquinaria de modificación.

Para probar la conservación funcional del gen de la FGE (SUMF1) durante la evolución sobre expresamos las ADNcs de ARSA, ARSC y ARSE en varias líneas celulares con y sin co-transfección de la ADNc de MSD y medimos las actividades de las sulfatasas. Los genes de la FGE (SUMG1) de los ratones y de la Drosophila estuvieron activos en las 3 sulfatasas humanas, siendo la FGE (SUMF1) de la Drosophila la menos eficiente. Éstos datos demuestran un alto grado de conservación funcional de la FGE (SUMF1) durante la evolución implicando una importancia biológica significativa para la función y supervivencia celular. Un efecto similar y consistente, aunque más débil, fue observado utilizando el gen FGE2 (SUMF2), sugiriendo que la proteína codificada por este gen también tiene una actividad codificadora de sulfatasas. Éstos datos demuestran que el monto de la proteína codificada con FGE (SUMF1) es un factor limitante para las actividades de la sulfatasas, un hallazgo con implicaciones importantes para la producción a larga escala de la sulfatasa vivas que serían utilizadas en la terapia de reemplazo enzimático.

Ejemplo 4:

Identificación del gen mutado en la MAS por medio de una complementación funcional utilizando transferencias cromosómicas mediadas con micros células.

En un experimento separado utilizando transferencias cromosómicas mediadas por micros células por medio de complementos funcionales confirmamos que el gen mutado en las MSD era de la FGE. Nuestros hallazgos suministran una mayor visión en un mecanismo biológico nuevo que afecta toda una familia de proteínas en organismos distantes relacionados. Adicionalmente al identificar la base molecular de una enfermedad genética rara, nuestra información confirma un efecto mejorador poderoso del producto del gen de la FGE en la actividad de las sulfatasas. El hallazgo posterior tiene implicaciones clínicas directas para la terapia de por lo menos 8 enfermedades humanas causadas por deficiencias de sulfatasas.

El gen para mapas de MSD al cromosoma 3p26

Para identificar la ubicación cromosómica del gen mutado en la MSD intentamos rescatar las enzimas con deficiencia de sulfatasas por medio de complementaciones funcionales vía transferencias cromosómicas mediadas por micro células. Un panel de líneas celulares híbridas humano/ratón que contenían cromosomas humanos normales individuales manchados con marcadores seleccionables dominantes, HyTK, que fueron utilizados como la fuente de los cromosomas humanos donantes y fusionados a una línea celular inmortalizada de un paciente con MSD. Todos los 22 autosomas fueron transferidos uno a uno a la línea celular del paciente y se seleccionaron híbridos en higromicina. Aproximadamente 25 colonias sobrevivientes fueron escogidas en cada uno de los 22 experimentos de transferencias. Éstas crecieron por separado y se cultivaron para sus siguientes pruebas enzimáticas. Las actividades de la arilsulfatasa A (ARSA) (IDENTIFICACIÓN SECUENCIAL NÚMERO: 15), la arilsulfatasa B (ARSB) (IDENTIFICACIÓN SECUENCIAL NÚMERO: 17), y la arilsulfatasa C (ARSC) (la IDENTIFICACIÓN SECUENCIAL NÚMERO: 19) fueron probadas para cada uno de los aproximadamente 440 clones (20 x 22). Éste análisis indicó claramente que las actividades de la sulfatasas de varios clones que se derivan de la transferencia cromosómica 3 fue significativamente más alto en comparación con aquella de todos los otros clones. Una variabilidad impresionante fue observada cuando se analizaron las actividades de cada uno de los clones individualmente de la transferencia cromosómica 3. Para verificar si cada clon tuvo un cromosoma 3 intacto de la línea celular donante, utilizamos un panel de 23 marcadores genéticos por tipos del cromosoma 3, distribuido uniformemente a lo largo del cromosoma y seleccionado previamente en base a diferentes alelos entre las líneas celulares del donante y del paciente. Esto nos permitió examinar para detectar la presencia del cromosoma donante e identificar una posible pérdida de regiones específicas debido a la ruptura cromosómica incidental. Cada clon que tenía una actividad enzimática alta retuvo a todo el cromosoma 3 de la línea celular donante, mientras que los clones con actividades bajas parecieron haber perdido a todo el cromosoma en base de la ausencia de los alelos del cromosoma 3 de la línea celular donante. Los clones posteriores probablemente retuvieron una región pequeña del cromosoma donante que contenía el gen marcador seleccionable que les permitía sobrevivir en el medio que contenía higromisina. Esta información indica que un cromosoma humano normal 3 era capaz de complementar el efecto observado en la línea celular de los pacientes con MSD.

Para determinar la región cromosómica específica que contiene al gen responsable de complementar la actividad, usamos híbridos del cromosoma 3 neo-marcados que descubrimos que habían perdido varias porciones del cromosoma. Adicionalmente, realizamos transferencias cromosómicas mediadas por micro células irradiadas de cromosomas 3 humanos marcados con HyTK. 115 híbridos irradiados del cromosoma 3 fueron probados para detectar sus actividades de sulfatasas y fueron clasificados de acuerdo a su genotipo utilizando un panel de 31 marcadores tipo microsatélites polimórficos que abarcaron a todo el cromosoma. Todos los clones que mostraron actividades enzimáticas altas parecieron haber retenido al cromosoma 3p26. Un análisis de alta resolución utilizando marcadores adicionales de esta región generó un mapa de la posible ubicación del gen de complemento entre los marcadores *D3S3630* y *D3S2397*.

Identificación del gen mutado en la MSD

Investigamos genes de la región genómica 3p26 para detectar mutaciones en los pacientes con MSD. Cada exón que incluía uniones de empalme fue amplificado por medio de PCR y analizado por medio de secuenciaciones directas. Los análisis de mutación fueron realizados en 12 individuos no relacionados; 5 pacientes con MSD descritos previamente y 7 casos no publicados. Algunas votaciones fueron identificadas a partir de nuestra cohorte con MSD en la marcación secuencial expresada (EST) AK075459 (IDENTIFICACIONES SECUENCIALES NÚMEROS: 4, 5), correspondiente a un gen de función desconocida, sugiriendo fuertemente que este era el gen involucrado en la MSD. Se descubrió que cada mutación estaba ausente en 100 individuos de control, excluyendo por lo tanto la presencia de un polimorfismo secuencial. Un análisis adicional de mutación confirmatoria fue realizado en ARNs de pacientes transcritos en reversa, particularmente en aquellos casos en los cuales el análisis genómico de ADN reveló la presencia de una mutación en, o cerca del lugar de empalme, afectando posiblemente a los empalmes. Mutaciones de movimiento de marco, sin sentido, de empalme y de cambio de sentido también fueron identificados, sugiriendo que la enfermedad es causada por una pérdida del mecanismo funcional, tal como se anticipó para una enfermedad recesiva. Estos exámenes concuerdan con la observación de que casi todas las mutaciones de cambio de sentido afectan a aminoácidos que son altamente conservados a través de la evolución (refiérase a continuación).

Tabla 8: Mutaciones identificadas adicionales de MSD

	Caso	referencia	fenotipo	exón	cambio de nucleótidos	cambio de aminoácidos
5	1. BA426.	Conary et al, 1988	Moderado	3	463T>C	S155P
				3	463T>C	S155P
	2. BA428	Burch et al, 1986	Neonatal severo	5	661delG	Cambio de marco
	10	3. BA431	Zenger et al, 1989	Moderado	1	2T>G
2					276delC	cambio de marco
4. BA799		Burk et al, 1981	Leve-moderado	3	463T>C	S155P
				3	463T>C	S155P
15	5. BA806	No publicado	Neonatal severo	9	1045T>C	R349W
	6. BA807	Schmidt et al, 1995	Desconocido	3	c519+4delGTAA ex 3	Saltando
					1076C>A	S359X
	7. BA809	Couchot et al, 1974	Leve-moderado	9	1A>G	MIV
1042G>C					A348P	
20	8. BA810	No publicado	Severo	1	1006T>C	C336R
				9	1046G>A	R349Q
	9. BA811	No publicado	Neonatal severo	3	c519+4delGTAAex 3	Saltando
				8	979C>T	R327X
25	10. BA815	No publicado	Moderado	5	c.603-6delC	Ex 6 saltando
				6	836C>T	A279V
	11. BA919	No publicado	Leve-moderado	9	1033C>T	R345C
				9	1033C>T	R345C
12. BA920	No publicado	Moderado	5	653G>A	C218Y	
			9	1033C>T	R345C	

Se identificaron mutaciones en cada paciente con MSD que fue examinado, excluyendo por lo tanto heterogeneidades de locus. Ninguna correlación obvia fue observada entre los tipos de mutaciones identificadas y la severidad del fenotipo reportado en los pacientes, sugiriendo que una variabilidad clínica no es causada por la heterogeneidad de los alelos. En 3 instancias en pacientes diferentes (caso 1 y 4, caso 6 y 9, y el caso 11 y 12 en la tabla 6) se descubrió que portaban la misma mutación. 2 de estos pacientes (el caso 11 y 12) tuvieron origen en el mismo pueblo en Sicilia, sugiriendo la presencia de un efecto fundador que fue confirmado en efecto por un análisis de haplotipos. Sorpresivamente, se descubrió que la mayoría de pacientes eran heterocigotos de compuestos, portando diferentes mutaciones de alelos, mientras que sólo unos pocos fueron homocigotos. Aunque fue consistente con la ausencia de consanguinidad reportada por los padres, este fue un hallazgo algo inesperado para una enfermedad recesiva muy rara tal como la MSD.

El gen y la proteína de la FGE

La secuencia de consenso de la ADNc de la FGE humana (también utilizados intercambiabilmente en este documento como SUMF1) (IDENTIFICACIÓN SECUENCIAL NÚMERO: 1) fue armada de algunos clones de marcaciones (EST) secuenciales expresadas y parcialmente de la secuencia genómica correspondiente. El gen contiene 9 exones y se expande aproximadamente por 105 kb (refiérase al ejemplo 1). La comparación secuencial también identificó la presencia de un parólogo genético de la FGE en el cromosoma humano 7 que lo designamos, FGE2 (también utilizado intercambiabilmente en este documento, SUMF2) (las IDENTIFICACIONES SECUENCIALES NÚMEROS: 45, 46).

Complementación funcional de las deficiencias de sulfatasas

Fibroblastos de los pacientes (caso 1 y 12 en la tabla 8) con MSD en quienes identificamos mutaciones del gen FGE (SUMF1) (líneas celulares BA426 y BA920) fueron infectadas con virus HSV que contenían formas de tipo silvestre y 2 formas mutadas de la ADNc de la FGE (SUMF1) (R327X y Δ ex3). Las actividades de ARSA, ARSB, y ARSC fueron probadas 72 horas después de la infección. La expresión de la ADNc de FGE (SUMF1) de tipo silvestre resultó en una complementación funcional de las 3 actividades, mientras que en las ADNcs de FGE (SUMF1) mutantes no ocurrió eso (tabla 9). Esta información suministró evidencia concluyente de la identidad de la FGE (SUMF1) puesto que el gen de la MSD y estas prueban una relevancia funcional de las mutaciones encontradas en los pacientes. Estas mutaciones asociadas con la enfermedad resultan en una deficiencia de sulfatasas, demostrando de esta forma que la FGE (SUMF1) es un factor esencial para la actividad de las sulfatasas.

Tabla 9: Complementación funcional de deficiencia de sulfatasas

Línea celular de MSD del receptor	Estructura	ARSA(1)	ARSB(1)	ARSC(1)
5	BA426	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)
10	BA920	16.6	11.3	0.15
	SUMF1-Δex3	17.2	14.4	0.07
	SUMF1-R327X	36.0	13.5	0.13
	SUMF1	73.5 (4.0 x)	21.6 (1.9 x)	0.42 (2.8 x)
Rango de control		123.7-294.6	50.6-60.7	1.80-1.58

(1) Todas las actividades enzimáticas son expresadas en nmoles de 4-metilumbeliferona mg de proteína liberada⁻¹ durante 3 horas. Las líneas celulares de MSD BA426 y BA920 fueron infectadas con el HSV amplicón individualmente, y con estructuras que portaban ADNcs de SUMF1 mutantes o de tipo silvestre. El incremento de actividades individuales de arilsulfatasas en fibroblastos infectados con el gen SUMF1 de tipo silvestre, en comparación con aquellas células infectadas con el vector individual, está indicado en paréntesis. Las actividades medidas en los fibroblastos de control que no fueron infectados también son indicadas.

Base molecular de la MSD

Basándonos en la hipótesis de que el gen de la enfermedad debería poder complementar la deficiencia enzimática en la línea celular de un paciente, realizamos transferencias cromosómicas mediadas por micro células a una línea celular inmortalizada de un paciente con MSD. Esta técnica ha sido usada exitosamente para la identificación de genes cuya función pronosticada puede ser evaluada en líneas celulares (por ejemplo, al medir la actividad enzimática o al detectar características morfológicas). Para abordar el problema de variabilidad estocástica de la actividad enzimática, medimos las actividades de 3 sulfatasa diferentes (ARSA, ARSB y ARSC) en el ensayo de complementación. Los resultados de la transferencia cromosómica indicaron claramente mapas del gen complementario del cromosoma 3. Los mapas subregionales fueron logrados al generar un panel híbrido de radiación para el cromosoma 3. Los clones híbridos individuales fueron caracterizados a un nivel genómico, al clasificar el tipo de 31 marcadores tipo microsatélite que mostraban diferentes alelos entre las líneas celulares de los donantes y de los receptores, y a nivel funcional al probar las actividades de las sulfatasas. El análisis de 130 de esos híbridos resultó en los mapas de la región complementaria del cromosoma 3p26.

Una vez que la región genómica crítica fue definida, el gen de la FGE (SUMF1) también fue identificado por medio de análisis de mutaciones en los ADN de los pacientes. Las mutaciones fueron encontradas en todos los pacientes que fueron examinados, suministrando un solo gen que está involucrado en la MSD. Las mutaciones encontradas fueron de diferentes tipos, la mayoría (por ejemplo, en el lugar de empalme, en el lugar de inicio, sin sentido, de cambio de marco) resultaron putativamente en una función perdida de la proteína codificada, tal como se esperaba para una enfermedad recesiva. La mayoría de mutaciones de cambio de sentido afectaron a los codones correspondiente de los aminoácidos que han sido altamente conservados a lo largo de la evolución, sugiriendo que estas mutaciones además causan una pérdida de función. No se pudieron concluir correlaciones entre el tipo de mutación y la gravedad del fenotipo, indicando que la gravedad del fenotipo es debido a factores no relacionados. Contrario a lo que se esperaría para una enfermedad genética rara, se descubrió que muchos pacientes fueron heterocigotos compuestos, portando 2 diferentes mutaciones. Sin embargo, un efecto fundador fue identificado para una mutación que se originó en un pueblo pequeño en Sicilia.

Función genética de la FGE (SUMF1)

La identidad del gen de la FGE (SUMF1) como el "factor complementario" fue demostrada definitivamente al rescatar la deficiencia enzimática de 4 diferentes sulfatasas en la expresión de la ADNc de la FGE (SUMF1) exógena, insertada en un vector viral, en 2 líneas diferentes de células de pacientes. En cada caso se observó una restauración consistente aunque parcial de todas las actividades sulfatasas examinadas, cuando se compararon con las líneas celulares de pacientes de control transfectadas a vectores vacíos. En promedio, el incremento de actividades enzimáticas varió de entre 1.7 a 4.9 veces y alcanzaron aproximadamente la mitad de los niveles observados en líneas celulares normales. La actividad enzimática tiene una correlación con el número de partículas de virus utilizadas en cada experimento y con la eficiencia de la infección tal como se probó por el análisis de proteínas marcadoras (GFP). En los mismos experimentos, vectores que contenían a ADNcs de la FGE (SUMF1) que portaban 2 de las mutaciones encontradas en los pacientes, R327X y Δex3, fueron utilizadas y ningún incremento significativo de la actividad enzimática fue observada, demostrando, entonces, la relevancia funcional de estas mutaciones.

Tal como se menciona en otras secciones de este documento, Schmidt et al. descubrió por primera vez que las sulfatasas atraviesan una modificación post-interpretación de una cisteína altamente conservada, que se encuentra en el lugar activo de la mayoría de sulfatasas, a Cα-formilglicina. También demostraron que esta modificación tuvo

deficiencias en la MSD (Schmidt, B., et al., Cell (Célula), 1995, 82:271-278). Nuestra información de mutaciones y de funciones suministra una evidencia fuerte de que la FGE (SUMF1) es responsable de esta modificación.

El gen de la FGE (SUMF1) muestra un nivel extremadamente alto de conservación secuencial a lo largo de todas las especies relacionadas distantemente que fueron analizadas, desde las bacterias hasta el hombre. Suministramos evidencia de que el homólogo de la Drosophila en el gen de la FGE (SUMF1) humana es capaz de activar sulfatasas humanas sobre-expresadas, probando que el alto nivel observado de similitud secuencial de los genes de la FGE (SUMF1) de especies relacionadas distantemente tiene una correlación con una conservación funcional impresionante. Una excepción notable es la levadura, que parece no tener al gen de la FGE (SUMF1) así como cualquier gen codificador de sulfatasas, indicando que la función de la sulfatasas no es requerida por este organismo y sugiriendo la presencia de una influencia recíproca en la evolución de los genes de la FGE (SUMF1) y de la sulfatasas.

Interesantemente, existen 2 genes homólogos, la FGE (SUMF1) y la FGE2 (SUMF2), en los genomas de todos los vertebrados analizados, incluyendo a los humanos. Tal como resulta evidente del árbol filogenético, el gen de la FGE2 (SUMF2) parece haber evolucionado independientemente del gen de la FGE (SUMF1). En nuestros ensayos el gen de la FGE (SUMF2) también puede activar sulfatasas, sin embargo, lo hace en una forma mucho menos eficiente en comparación con el gen de la FGE (SUMF1). Esto podría explicar la actividad residual de sulfatasas que se encuentra en los pacientes con MSD y sugiere que una deficiencia completa de sulfatasas sería letal. En este momento, no podemos descartar la posibilidad de que el gen de la FGE2 (SUMF2) tiene una función adicional pero que todavía desconocida.

Impacto de la terapia de enfermedades debido a las deficiencias de sulfatasas

Un fuerte incremento (de hasta 50 veces, de las actividades de sulfatasas fueron observadas en las células que sobre expresaban a la ADNc de la FGE (SUMF1) junto con los ADN es de ARSA, ARSC, o ARSE, en comparación con las células que sobre expresaban una sola sulfatasa individual. En todas las líneas celulares, un efecto sinérgico significativo fue encontrado, indicando que la FGE (SUMF1) es un factor limitante para la actividad de la sulfatasas. Sin embargo, se observó una variabilidad entre las diferentes sulfatasas, debido, posiblemente, a afinidades diferentes de la proteína codificada por la FGE (SUMF1) con las varias sulfatasas. También se observó una variabilidad entre diferentes líneas celulares que podrían tener diferentes niveles de enzimas generadoras de formilglicina endógena. Consistente con estas observaciones, hemos encontrado que la expresión del gen de la MSD varía entre diferentes tejidos, con niveles significativamente altos en los riñones y en el hígado. Esto podría tener implicaciones importantes puesto que los tejidos con bajos niveles de expresión genética de FGE (SUMF1) podrían ser menos capaces de modificar efectivamente a proteínas de sulfatasas entregadas exógenamente (refiérase a secciones posteriores de este documento). Toda esta información sugiere que la función del gen de la FGE (SUMF1) ha evolucionado para lograr un sistema regulatorio dual, en la que cada sulfatasa está siendo controlada por un mecanismo individual, responsable de los niveles de ARNm de cada gen de sulfatasas estructurales, y un mecanismo compartido y común de toda la sulfatasas. Adicionalmente, la FGE2 (SUMF2) suministra una redundancia parcial para la modificación de las sulfatasas.

Esta información tiene implicaciones profundas para la producción en masa de sulfatasas activas que serían utilizadas en las terapias de reemplazo enzimático. Los estudios de reemplazo enzimático han sido reportados en modelos animales de deficiencia sulfatasas, tales como un modelo felino de mucopolisacaridosis VI, que demostró ser efectivo para prevenir y curar algunos síntomas. Ensayos terapéuticos en humanos se están realizando actualmente para 2 enfermedades congénitas debido a deficiencias de sulfatasas, el MPSII (síndrome de Hunter) y MPSVI (el síndrome de Maroteaux-Lamy) y pronto se extenderá a un número grande de pacientes.

Ejemplo 5:

Terapia de reemplazo enzimático con GALNS activada por FGE para la enfermedad de Morquio MPS IVA

La causa principal de la patología esquelética en pacientes de la enfermedad de Morquio es la acumulación del queratán sulfato (KS - keratan sulfate) en los condrocitos disco fisis (cartílago de crecimiento) debido a deficiencias de las sulfatasas de los liposomas, GALNS. El objetivo principal de los estudios de investigación in vivo es determinar si es que la administración intravenosa (IV) de GALNS activadas por la FGE es capaz de penetrar a los condrocitos del ligamento de crecimiento así como otros tipos celulares apropiados en ratones normales. Sin importar, la falta general de anomalías esqueléticas, un modelo de ratón deficientes de GALNS (Morquio Knock-In -MKI, S. Tomatsu, St. Louis University, MO) también fue utilizado para demostrar la actividad bioquímica in vivo de una administración repetitiva de GALNS activados por la FGE. La falta de patología esquelética en modelos de ratones refleja el hecho de que la KS esquelética ha sido reducida en una forma importante o está ausente en los roedores (Venn G, & Mason RM., Biochem J., 1985, 228:443-450). Estos ratones, sin embargo, demostraron una acumulación detectable de GAG y otras anomalías celulares en varios órganos y tejidos. Por lo tanto, el objetivo general de los estudios fue el demostrar que las GALNS activadas por FGEs penetran al ligamento de crecimiento (estudio de bio-distribución) y muestran una actividad enzimática funcional de GALNS hacia la remoción de GAGs acumulados en los tejidos afectados (estudio fármaco - dinámico).

Los resultados de estos estudios demostraron que GALNS activados por FGEs inyectados intravenosamente fueron internalizados por los condrocitos del ligamento de crecimiento, a pesar de niveles relativamente bajos en comparación a otros tejidos. Adicionalmente, la inyección de GALNS activados por FGE durante el transcurso de 16 semanas en ratones MKI despejó efectivamente al GAG acumulado y redujo la coloración de los bio-marcadores de lisosomas en todos los tejidos suaves examinados. En general, los experimentos demostraron exitosamente la entrega de GALNS a los condrocitos del ligamento de crecimiento y demostraron una actividad bioquímica en términos del despeje de GAG en varios tejidos.

Estudio de bio - distribución

A ratones ICR (normales) de 4 semanas de edad se les administró una sola inyección intravenosa de 5 mg/kilogramos de GALNS activado por FGEs. El hígado, el fémur (hueso), el corazón, el riñón y el bazo fueron recaudados 2 horas después de la inyección y preparados para una examinación histológica. Un anticuerpo monoclonal de GALNS anti-humano fue utilizado para detectar la presencia de GALNS inyectados en los varios tejidos. GALNS fue detectado en todos los tejidos examinados en comparación a los controles portadores. Además, se observó fácilmente a GALNS en todos los tejidos examinados utilizando un sistema reportador más sensible de rábano-peroxidasa, con la excepción del hueso. La demostración de la absorción de GALNS en el ligamento de crecimiento requirió el uso de un sistema de informes más sensible con isotiocianato de fluoresceína (FTTC - fluorescein-isothiocyanate) indica que, aunque las GALNS penetran al ligamento de crecimiento, está menos disponible en los condrocitos del ligamento de crecimiento que en células de tejidos suaves. Sin importar el requerimiento de un método más sensible de detección presente, la entrega de GALNS a condrocitos óseos del ligamento de crecimiento fue observada en todas las secciones del ligamento de crecimiento examinadas en comparación a los controles portadores.

Estudió fármaco - dinámico en ratones MKI

A ratones MKI o de tipo silvestre de 4 semanas de edad se les administró inyecciones semanales intravenosas (n = 8 por grupo) hasta que tenían 20 semanas de edad. Cada inyección semanal consistió de 2 mg/kilogramos de GALNS activados por FGE o un portador de control. (Ninguna inyección para ratones de tipo silvestre). Todos los ratones fueron sacrificados para una examinación histológica a las 20 semanas de edad y se mancharon utilizando los siguientes métodos: hematoxilina y eosina para la morfología celular, azul alciano para la detección de GAGs.

El despeje de GAGs acumulados fue demostrado al reducir la coloración de azul alciano en todos los tejidos suaves examinados (el hígado, el corazón, el riñón y el bazo). Esto fue observado únicamente en los ratones inyectados con GALNS. Aunque el ligamento de crecimiento en los ratones MKI funcionó normalmente tal como se evidenció por la morfología esquelética normal, existieron anomalías celulares sutiles observadas (incluyendo la formación de vacuolas de condrocitos sin efectos patológicos aparentes). Los condrocitos con vacuolas de las zonas hipertróficas y proliferativas del ligamento de crecimiento no fueron afectados por la administración de GALNS. Esto contrastó con los condrocitos en la zona de calcificación del ligamento de crecimiento donde una reducción de la formación de vacuolas se observó en ratones inyectados con GALNS. La formación de vacuolas de los condrocitos y la acumulación de GAG que se presume que no tienen relación con KS en el ligamento de crecimiento en ratones MKI fue, en general, sorprendente e inesperada debido a la falta conocida de KS en el ligamento de crecimiento de los ratones. Estas observaciones particulares reflejan muy posiblemente el hecho que, en los ratones knock in, niveles altos de GALNS mutantes están presentes (lo contrario de los ratones knock-out en los cuales no existe ningún GALNS mutante residual, no existió ninguna formación de vacuolas en los condrocitos del ligamento de crecimiento y no existió ninguna acumulación de GAGs. Tomatsu S. et al., Human Molecular Genetics (Genética Molecular Humana), 2003, 12:3349-3358). El fenómeno de la formación de vacuolas en el ligamento de crecimiento podría ser una indicación de un efecto secundario en un subconjunto de células que expresan a las GALNS mutantes. Sin embargo, la inyección de enzimas durante el transcurso de 16 semanas demostró una fuerte evidencia de una entrega de GALNS activada por FGE a varios tejidos y una actividad enzimática in vivo.

Descripción detallada de los esquemas

Figura 1: espectros de masa MALDI-TOF de P23 después de la incubación en la ausencia de (A) o con la presencia de (B) de un extracto soluble de microsomas de testículos bovinos. 6 pmoles de P23 fueron incubados bajo condiciones estándar durante 10 minutos a 37 °C en la ausencia o en la presencia de 1 µl de extracto microsomal. Las muestras preparadas para espectrometría de masa de MALDI-TOF tal como se describió en los procedimientos experimentales. Las masas monoisotópicas MH⁺ de P23 (2526.28) y su derivado de FGly (2508.29) son indicadas.

Figura 2: Árbol filogenético derivado de una alineación de FGEs y 21 proteínas humanas de la semilla PFAM-DUF323. Los números de las ramas indican la distancia filogenética. Las proteínas son designadas de acuerdo a su número de identificación TrEMBL y el nombre de la especie. hFGE - FGE humana. Superior derecha: la escala de las distancias genéticas. Un asterisco indica que el gen ha sido

investigado aún más. Los 7 genes superiores son parte de la familia genética de las FGEs. **Figura 3: Organización del locus genético de las FGEs en humanos y ratones.** Los exones se muestran a escala como cajas oscuras (locus humano) y cajas claras (locus de los ratones). La barra en la esquina inferior izquierda muestra la escala. Las líneas entre los exones muestran los intrones (no a escala). Los números sobre las líneas de intrones indican el tamaño de los intrones en kilo bases.

Figura 4: Diagrama que muestra un mapa de plásmidos pXMG.1.3 con expresión de FGE.

Figura 5: Gráfico de barras que muestra la actividad de las-sulfatasas de N-Acetilgalactosamina en células 36F Transfectadas transitoriamente con plásmidos de expresión de FGE. Las células fueron transfectadas con un plásmido de control, con pXMG.1.2, con la ADNc de FGE en orientación reversa, o un plásmido de expresión FGE, pXMG.1.3 en un medio sin metotrexato (MTX). 24 horas después las células fueron re-alimentadas. La actividad de las 6-sulfatasas de N-Acetilgalactosamina se determinó por medio de un ensayo de actividad. Cada valor mostrado es el promedio de 2 transfecciones separadas con desviaciones estándar indicadas por medio de las barras de error.

Figura 6: Gráfico de barras que muestra la actividad específica de las 6-sulfatasas de N-Acetilgalactosamina en células Transfectadas transitoriamente con plásmidos de expresión de FGE. Las células fueron transfectadas con un plásmido de control, con pXMG.1.2, con una ADNc de FGE en orientación reversa, o con un plásmido de expresión de FGE, pXMG.1.3, en un medio sin metotrexato (MTX). 24 horas después las células fueron re-alimentadas con un medio que contenía 1.0 μ M de MTX. El medio fue de privado y las células fueron recolectadas 24, 48 y 72 horas después de la re-alimentación. La actividad específica de las 6-sulfatasas de N-Acetilgalactosamina fue determinada por medio de un ensayo de actividad y ELISA y está representada como una tasa de la actividad de las 6-sulfatasas de N-Acetilgalactosamina por miligramo de 6-sulfatasas de N-Acetilgalactosamina de reactivos ELISA. Cada valor mostrado es el promedio de 2 transfecciones separadas.

Figura 7: Gráfico de barras que muestra la producción de 6-sulfatasas de N-Acetilgalactosamina en las células 36F Transfectadas transitoriamente con plásmidos de expresión de FGE. Las células fueron transfectadas con un plásmido de control, pXMG.1.2, con la ADNc de FGE en una orientación reversa, o un plásmido con expresión de FGE, pXMG.1.3, en un medio sin metotrexato (MTX). 24 horas después las células fueron re - alimentadas con un medio que contenía 1.0 μ M de MTX. El medio fue cultivado y las células fueron recolectadas 24, 48 y 72 horas después de la reunión alimentación. Las proteínas totales de las 6-sulfatasas de N-Acetilgalactosamina se determinaron por medio de ELISA. Cada valor mostrado es el promedio de 2 transfecciones separadas con las desviaciones estándar indicadas por las barras de error.

Figura 8: Gráfico que muestra la actividad de 2-sulfatasas de iduronato en las células 30C6 Transfectadas transitoriamente con plásmidos de expresión de FGE. Las células fueron transfectadas con un plásmido de control, pXMG.1.2, con la ADNc de FGE en orientación reversa, o un plásmido de expresión de FGE, pXMG.1.3, en un medio sin metotrexato (MTX). 24 horas después las células fueron re - alimentadas con un medio que contenía 1.0 μ M de MTX. El medio fue cultivado y las células fueron recolectadas 24, 48 y 72 horas después de la reunión alimentación. La actividad de 2-sulfatasas de Iduronato fue determinada por medio de un ensayo de actividad. Cada valor mostrado es el promedio de 2 transfecciones separadas.

Figura 9: Muestra un botiquín que plasma las características de este invento.

LISTADO SECUENCIAL

<110> Transkaryotic Therapies, Inc. von Figura, Kurt Schmidt, Bernhard Dierks, Thomas Heartlein, Michael W.

Cosma, Maria P. Ballabio, Andrea

<120> DIAGNÓSTICO Y TRATAMIENTO DE DEFICIENCIAS DE MÚLTIPLES SULFATASAS Y OTRAS DEFICIENCIAS DE SULFATASAS

<130> 0403WO

<150> ESTADOS UNIDOS 60/447,747

<151> 2003-02-11

<160> 95

<170> Versión de la patente 3.2

<210> 1
 <211> 1180
 <212> ADN
 <213> Homo sapiens

<220>
 <221> CDS
 <222> (20)..(1141)
 <223> ADNc de FGE

<400> 1

```

acatggcccg cgggacaac atg get gcg ecc goa eta 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 tgg ctg      100
Arg Cys Pro Glu Leu Gly Leu Val Leu Leu Leu Leu Leu Ser Leu
          15          20          25

ctg tgt gga gcg goa ggg agc cag gag gcc ggg acc ggt gcg gcc 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 acc gcc ecc cag cgg cct gcc      196
Gly Ser Leu Ala Gly Ser Cys Gly Cys Gly Thr Pro Gln Arg Pro Gly
          45          50          55

gcc cat gcc agt tgg gca gcc get ccc cga tac tgg cgg gag get aac      244
Ala His Gly Ser Ser Ala Ala Ala His Arg Tyr Ser Arg Glu Ala Asn
          60          65          70          75

get ccg gcc ecc gta ecc gga gag cgg caa ctc gcg ccc tca aag atg      292
Ala Pro Gly Pro Val Pro Gly Glu Arg Gln Leu Ala His Ser Lys Met
          80          85          90

gtc ecc atc cct get gga gta ttt aca atg gcc 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 goa cct gcg egg 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 gcc tat ttg aca gag get gag aag ttt gcc 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 gcc 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 get get ecc tgg tgg tta cct gtg aaa gcc get aac      580

```

5 Gln Ala Val Ala Ala Ala Pro Trp Trp Leu Pro Val Lys Gly Ala Asn
175 180 185
628 tgg aga cac cca gaa ggg cct gac tct act att ctg cac agg ccg gat
Trp Arg His Pro Glu Gly Pro Asp Ser Thr Ile Leu His Arg Pro Asp
190 195 200
676 cat cca gtt ctc cat gtg tcc tgg aat gat gcg gtt gcc tac tgc act
His Pro Val Leu His Val Ser Trp Asn Asp Ala Val Ala Tyr Cys Thr
205 210 215
724 tgg gca ggg aag cgg ctg ccc acg gaa gct gag tgg gaa tac agc tgt
Trp Ala Gly Lys Arg Leu Pro Thr Glu Ala Glu Trp Glu Tyr Ser Cys
220 225 230 235
772 cga gga ggc ctg cat aat aga ctt ttc ccc tgg ggc aac aaa ctg cag
Arg Gly Gly Leu His Asn Arg Leu Phe Pro Trp Gly Asn Lys Leu Gln
240 245 250
820 ccc aas ggc cag cat tat gcc aac att tgg cag ggc gag ttt ccg gtg
Pro Lys Gly Gln His Tyr Ala Asn Ile Trp Gln Gly Glu Phe Pro Val
255 260 265
868 acc aad act ggt gag gat ggc ttc caa gga act gcg cct gtt gat gcc
Thr Asn Thr Gly Glu Asp Gly Phe Gln Gly Thr Ala Pro Val Asp Ala
270 275 280
916 ttc cct ccc aat ggt tat ggc tta tac aac ata gtg ggg aac gca tgg
Phe Pro Pro Asn Gly Tyr Gln Leu Tyr Asn Ile Val Gly Asn Ala Trp
285 290 295
964 gaa tgg act tca gac tgg tgg act gtt cat cat tct gtt gaa gaa acg
Glu Trp Thr Ser Asp Trp Trp Thr Val His His Ser Val Glu Glu Thr
300 305 310 315
1012 ctt aac cca aaa ggt ccc cct tct ggg aaa gac cga gtg aag aaa ggt
Leu Asn Pro Lys Gly Pro Pro Ser Gly Lys Asp Arg Val Lys Lys Gly
320 325 330
1060 gga tcc tac atg tgc cat agg tct tat tgt tac agg tat cgc tgt gct
Gly Ser Tyr Met Cys His Arg Ser Tyr Cys Tyr Arg Tyr Arg Cys Ala
335 340 345
1108 gct cgg agc cag aac aca cct gat agc tct gct tgg aat ctg gga ttc
Ala Arg Ser Gln Asn Thr Pro Asp Ser Ser Ala Ser Asn Leu Gly Phe
350 355 360
1161 cgc tgt gca gcc gac cgc ctg ccc acc atg gac tgacaaccaa gggtagtctt
Arg Cys Ala Ala Asp Arg Leu Pro Thr Met Asp
365 370
1180 cccagtcaca aggagcagt

<210> 2

<211> 374

<212> PRT

<213> Homo sapiens

<400> 2

50 Met Ala Ala Pro Ala Leu Gly Leu Val Cys Gly Arg Cys Pro Glu Leu
1 5 10 15
55 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
60 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

5
10
15
20
25
30
35
40
45
50
55
60

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
145 150 155 160

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 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
305 310 315 320

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> 3
<211> 1122
<212> ADN

<213> Homo sapiens

<400> 3

5

10

15

20

25

30

```

atggetgoge cegcactagg gctgggtgtt ggacgttgcc ctgagctggg tctcgtctct 60
ttgctgctgc tgcctctgct gctgtgtgga gggcaggga gccaggagge cgggaccggt 120
gggggcggg ggtcccttgc ggggtcttgc ggtgcggga cgcgccagcg gcttggcgcc 180
catggcagtt cggcagccgc tcaccgatac tcgcgggagg ctacgcctcc gggcccccga 240
ccgggagagc ggcaactcgc gcaactcaag atggteccca tccctgctgg agtattttaca 300
atgggcacag atgatoctca gataaagcag gatggggaag cactgcgcag gagagttact 360
attgatgcct ttacatgga tgcctatgaa gtcagtaata ctgaatttga gaagtttgtg 420
aactcaactg gctatttgac agagggtgag aagtttggcg actcctttgt ctttgaagga 480
atgttgagtg agcaagtga gaccaatatt caacaggcag ttgcagctgc tccctggttg 540
ttacctgtga aaggcgctaa ctggagacac ccagaagggc ctgactctac tattctgcac 600
agggcggatc atccagttct ccatgtgtcc tggaatgatg cggttgccta ctgcacttgg 660
gcagggaagc ggtgccccac ggaagctgag tgggaataca gctgtcgagg aggcctgcct 720
aatagacttt tccctgggg caacaaactg cagcccaag gccagcatta tgcacacatt 780
tggcaggcg agtttcgggt gaccaacact ggtgaggatg gcttccaagg aactgcgcct 840
gttgatgcct tccctcccaa tggttatgga ttatacaaca tagtggggaa cgcattggga 900
tggacttcag actggtggac tgttcacat tctgttgaag aaacgcttaa cccaaaaggt 960
cccccttctg ggaagacgc agtgaagaaa ggtggatcct acatgtgcca taggtcttat 1020
tgttacaggt atcgtgtgct tgcctggagc cagaacacac ctgatatgct tgcctcgaat 1080
ctgggattcc gctgtgcagc cgaaccgctg cccaccatgg ac 1122

```

<210> 4

<211> 2130

35

<212> ADN

<213> Homo sapiens

<400> 4

40

45

50

55

60

65

```

acatggcccg cgggacaaca tggctgcgcc cgcactaggg ctggtgtgtg gacgttgccc 60
tgagctgggt ctggtctctc tgcgtctgct gctctcgtct ctgtgtggag cggcaggag 120
ccaggaggcc ggyaccggtg cgggcgcggg gtcccttgcg ggttcttgcg gctgcggcac 180
gccccagcgg cctgggcgcc atggcagttc ggcagccgct caccgatact cgcgggagge 240
taacgctccg ggcgccgtac ccggagagcg gcaactcgcg caotcaaaga tgggtcccat 300
ccctgctgga gtatttaca tgggcacaga tgatcctcag ataaagcagg atgggggaagc 360
aectgcaggg agagttacta ttgatgcct ttacatggat gcttatgaag tcagtaatac 420
tgaatttgag aagtttgtga actcaactgg ctatttgaca gaggetgaga agtttggcga 480
ctcctttgtc tttgaaggca tgttgagtg gcaagtgaag accaatatc aacaggcagt 540
tgcagctgct cctggtggt tacctgtgaa aggcgctaac tggagacacc cagaagggcc 600
tgactctact attctgcaca ggcgggatac tccagttctc catgtgtcct ggaatgatgc 660
ggttgcctac tgcacttggg caggggaagc gctgccacg gaagctgagt ggggaatacag 720
ctgtcgagga ggcctgcata atagactttt cccctggggc aacaaactgc agcccaagg 780

```

5
10
15
20
25
30
35

```

ccagcattat gccaacattt ggcagggcga ttttcgggtg accaacactg gtgaggatgg      840
cttccaagga actgcgcctg ttgatgcctt cctcccaat ggttatggct tatacaacat      900
agtgggggaa gcatgggaat ggacttcaga ctgggtggact gttcatcatt ctggtgaaga      960
aacgcttaac ccaaaaggto ccccttcttg gaaagaccga gtgaagaaag gtggatccta     1020
catgtgccat aggtcttatt gttacaggta tgcctgtgct gctcggagcc agaacacacc     1080
tgatagctct gcttcgaato tgggattccg ctgtgcagcc gacgcctgc ccaccatgga     1140
ctgacaacca agggtagtct tcccagtc caggagcagt cgtgtctgac ctacattggg      1200
ctttcctcag aactttgaac gatcccatgc aaagaattcc caccctgagg tgggttacat      1260
aactgcccaa tggccaaagg aacgccttg tgagaccaa ttgctgacct gggtcagtgc      1320
atgtgcttta tgggtgtggt catctttgga gatcatcacc atattttact tttgagagtc      1380
tttaaagagg aaggggagtg gagggaacce tgagctagge ttcaggagge ccgcatacta      1440
cgcaggctct gccacagggg ttagaccca ggtccgacgc ttgaccttc tgggcctcaa      1500
gtgccctccc ctatcaaatg aaggaatgga cagcatgacc tctgggtgtc tctccaactc      1560
accagttcta aaaagggtat cagattctat tgtgacttca tagaatttat gatagattat      1620
tttttagcta ttttttccat gtgtgaacct tgagtgaac taatcatgta aagtaagagt      1680
tctcttatgt attatgttgc gaagaggggt gtggtgactc ctttatatc gtactgcact      1740
ttgtttttcc aaggaatca gtgtcttta cgttggttat atgaatcca catggggccg      1800
gtgatggtat gctgaagtc agccgttgaa cacataggaa tgtctgtggg gtgactctac      1860
tgtgctttat cttttaacat taagtgcctt tggttcagag gggcagtcac aagctctggt      1920
tccccctoto cccaaagcct tcagogaag tgaaatgtgc gctaaacggg gaaacctggt      1980
taattctaga tataggggaa aaggaacgag gaccttgaat gagctatat caggggatcc      2040
ggatatttgt aataggggaa aggaacctt gttggctgtg gaatatccga tgccttgaat      2100
catgcactgt gttgaataaa cgtatctgtc      2130

```

<210> 5
<211> 374
<212> PRT
<213> Homo sapiens

<400> 5

45
50
55
60
65

```

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 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> ADN
 <213> Homo sapiens
 <400> 6

5 eggctgtgtt ggcagttct catgggttcc cgaaggagg gtctctgtgg ctgcggcggc 60
 tgctaaactgc gccacctgct gcagcctgtc cccgcgcgc tgaagcggcc gcgtcgaagc 120
 cgaaatgcgc ccaccccgga ccggccgagg ccttctctgg ctgggtctgg ttctgagctc 180
 cgtctgcgtc gccctcggat ccgaaaagca ggcacactcg accacagatg ctctgaacgt 240
 tcttctcctc atcgtggatg acctgcgcgc ctcctcgggc tgttatgggg ataagctggt 300
 10 gaggtcccca aatattgac aactggcctc ccacagcctc ctctccaga atgcctttgc 360
 gcagcaagca gtgtgcgcgc cgagccgcgt ttctttctc actggcagga gacctgacac 420
 caccgcctg tacgaattca actcctactg gaggtgcac gctggaaact tctccacct 480
 ccccagtac ttcaaggaga atggctatgt gacctgtcg gtgggaaaag tctttcaccc 540
 15 tgggatatct tctaaccata ccgatgatc tccgtatagc tgggtcttct cacttatca 600
 tcttctctct gagaagtatg aaacactaa gacatgtcg gggccagatg gagaactcca 660
 tgccaaactg ctttgccctg tggatgtgct gcatgttccc gagggcacc tgcctgacaa 720
 20 acagagcact gagcaagcca tacagtgtt ggaagagatg aaacgtcag ccagtccttt 780
 cttcctggcc gttgggtatc ataagccaca catccctctc agatacccca aggaatttca 840
 gaagtgttat ccttggaga acatcaccct ggcctccgat ccgaggtcc ctgatggct 900
 acccctgtg gctacaacc cctggatgga catcaggcaa cgggaagacg tccaagcctt 960
 25 aaacatcagt gtgcgtatg gtccaatcc tgtggaactt cagcggaaa tccgccagag 1020
 ctaatttgc tctgtgctat atttggatac acaggtcggc cgcctcttga gtgctttgga 1080
 cgatcttcag ctggccaaca gacccatcat tgcatttacc tggatcatg ggtgggctct 1140
 30 aggtgaacat ggagaatgg ccaatacag caattttgat gttgctacc atgttccct 1200
 gatattctat gttcctgga ggaaggctc acttcaggag gcaggcgaga agcttttccc 1260
 ttacctcagc ctttttgatt ccgcctcaca gttgatggag ccaggcaggc aatccatgga 1320
 35 ccttgtgga cttgtgtctc ttttccac gcgtgctgga cttgcaggac tgcaggctcc 1380
 acctcctgc ccgcttctt catttccagt tgaetgtgc agagaaggca agaacctct 1440
 gaagcatttt cgattccgtg acttggaga ggatecgtac ctccttgga atccctgta 1500
 actgattgac tatagccagt atccccggcc ttcagacatc cctcagtga attctgacaa 1560
 40 gcaggttta aaagatataa agatcatggc ctattccata cgcaccatag actataggta 1620
 tactgtgtgg gttggcttca atcctgatga atttctagct aacttttctg acatccatgc 1680
 aggggaactg tattttgtgg attctgacc attgcaggat ccaaatatgt ataatgatc 1740
 45 ccaaggtgga gatcttttcc agttgttgat gccttgagtt ttgccaaca tggatggcaa 1800
 atgtgatgtg ctccttcca gctgggtgaga ggaggagta gagctggctg ttttgtgatt 1860
 acccataata ttggaagcag cctgagggt agttaatcca aacatgcctc aacaatttgg 1920
 cctgagaata tgaacagcc aaaccttctc gtttagtctt tattaatatt tataattggg 1980
 50 aattggacca gttttttttt taatttccct ctttttaaaa cagttacggc ttatttactg 2040
 aataaataca aagcaacaa actcaagtta tgcatacct ttggatacga agaccataca 2100
 taataaccaa acataacatt atacacaaag aatacttcca ttatttgtgg aatttagtgc 2160
 55 atttcaaaa gtaatcatat atcaactag gcacacact aggttctga ttattttgtt 2220
 tataatttaa taatatatct tatgagccct atatattcaa aatattatgt taacatgtaa 2280
 tccatgttcc ttttccc 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

```


5 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

10 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

15 Val Ala Thr His Val Pro Leu Ile Phe Tyr Val Pro Gly Arg Thr Ala
355 360 365

20 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

25 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

30 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

35 Gln Tyr Pro Arg Pro Ser Asp Ile Pro Gln Trp Asn Ser Asp Lys Pro
465 470 475 480

40 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

45 Asn Phe Ser Asp Ile His Ala Gly Glu Leu Tyr Phe Val Asp Ser Asp
515 520 525

50 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> 2857

<212> ADN

<213> Homo sapiens

<400> 8

gaattccggg ccattgagctg ccccgtagcc gcctgctgcg cgtgctgct agtccctgggg

60

	ctctgcgggg	egcgtccccc	gaaagcactg	ctgctcctcg	eggatgacgg	aggctttgag	120
	agtggcgegt	acaacaacag	cgcctcgcgc	accccgacac	tggacgcctt	ggcccgcgcg	180
5	agcctcctct	ttcgcaatgc	cttcacctcg	gtcagcagct	gctctcccag	cgcgcgcgcg	240
	ctctcactcg	gctgcccc	gcctcagaat	gggatgtacg	ggctgcacca	ggacgtgca	300
	cacttcacct	ccttcgacaa	gggtcgggag	ctgcogctgc	tgtcagcca	agctggtgtg	360
10	ggcacaggca	tcctcgggaa	gaagcagctg	gggcgggaga	cgtgtaccc	gtttgacttt	420
	ggttacacgg	aggagaatgg	ctcgtctctc	caggtggggc	ggacatcac	tgaatttaag	480
	ctgctcgtcc	ggaaattcct	gcagactcag	gatgaecggc	ctttcttctt	ctacgtcgcc	540
15	ttccacagac	cccacgcctg	tgggcactcc	cagccccagt	acggaaacct	ctgtgagaag	600
	tttggcaacg	gagagagcgg	cattgggtcgt	atcccagact	ggacccccca	ggcctacgac	660
	ccactggacg	tgtgtgtgac	ttacttcgtc	cccaacccc	cggcagcccg	agccgacctg	720
	gcogctcagt	acaaccaocg	cggcgcgcatg	gacaaaggag	ttggactggg	gctccaggag	780
20	ctggtgtgac	cgggtgtctt	gaacgacaca	ctggtgatct	tcacgtccga	caacgggac	840
	cccttcacca	gggcaggag	caacctgtac	tggcgggga	ctgttgacc	cttactgggtg	900
	tcctcccccg	agcaacccaa	acgtcggggc	cagtcagcg	aggcctacgt	gagcctccta	960
25	gacctcagcg	ccacatcttt	ggattggttc	togatccgt	accccagcta	cgcctctttt	1020
	ggctcgaaga	ccatccacct	cactggccgg	tcctcctg	cggcgttgga	ggccgagccc	1080
	ctctggggca	cgtcttttgg	cagccagagc	caacaagagg	taacctgtc	ctaccccatg	1140
30	cgctcgtg	agcaecggca	cttcgcctc	gtgacacac	taacttcaa	gatgcctttt	1200
	cccatcgacc	aggacttcta	cgtctcacc	accttcacgg	acctcctgaa	cgcacccaca	1260
	gctggtcagc	ccacgggctg	gtacaaggac	ctcgtctatt	actactacgg	ggcgcgctgg	1320
35	gagctctacg	acgggagccg	ggacccccac	gagacccaga	acctggccac	cgaaccgcgc	1380
	tttgcctcag	ttctggagat	gcttcgggac	cagctggcca	agtggcagtg	ggagacccac	1440
	gacccctggg	tgtgcgcgcc	cgaaggcgct	ctggaggaga	agctctctcc	ccagtgcacg	1500
40	ccctccaca	atgagctgtg	acctccccag	gaggcctgtg	cacacatccc	aggcatgtcc	1560
	cagacacatc	ccacaogtgt	cgtgtggccc	ggccagcctg	gggagttagtg	gcaacagccc	1620
	ttcgttcac	actcccatcc	aaggagggtt	cttctctctt	gtgggggtac	tcttgccatt	1680
	gcctggaggg	ggaccagagc	atgtgaccag	agcatgtgac	cagccctccc	accaccaggg	1740
45	gcactgcogt	cattggcagg	gacacagttg	tccttgtgtc	tgaacctatg	cccagcacgg	1800
	gaattctaga	catacgtggt	ctgcggacag	ggcagcgccc	ccagcccattg	acaaggaggt	1860
	cttgtttttt	ggcttgggtt	ggggacctgc	aaatgggagg	cctgaggccc	tcttcaggct	1920
50	ttggcagcca	cagatacttc	tgaaccttcc	acagagagca	ggcaggggct	tgggtgcgcg	1980
	gtgggcagta	cgcaggctcc	accgacactc	acctgggagc	acggcgccctg	gctcttacc	2040
	gcgtctggcc	tagagggaag	ctttgagcga	cccttgggca	ggtttctgct	tcttctgttt	2100
55	tgcctatggt	caagtccctg	ttccccaggg	aggtttcagc	tgattggcag	caggctccct	2160
	gagtgatgag	cttgaacctg	tgggtgtttt	gggcagaagc	ttatcttttt	tgagagtgtc	2220
	ggaagatgaa	ggcatggcga	tgcctgtctt	ctggcttggg	ttatctcttc	ggtgacactg	2280
60	gcattgctgg	gtggtgatgc	cgtctctctg	gcttgggtta	attcttcgggt	gacactggcg	2340
	ttgctgggtg	gcaatgcgcc	tcctctgggt	tgggttaatt	cttcgggtgac	actggcggtg	2400

```

ctgggtggcg atgcccgccc tctggcttgg gttattcttt ggatgacgtc ggcgttgctg 2460
ggagaatgtg ccgttcctgc cctgcctcca cccacctcgg gagcagaago cgggcctgga 2520
caccctcgg cctggacacc cctcgaagga gaggcgctt ccttgagtag gtgggctccc 2580
cttgccttc cctccctatc actccatact ggggtgggct ggaggaggcc acaggccagc 2640
tattgtaaaa gcttttt 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

```

5

10

15

20

25

30

35

40

45

50

55

60

65

Thr Thr Val Gly Arg Met Asp Gln Gly Val Gly Leu Val Leu Gln Glu
245 250 255
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
305 310 315 320
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> ADN
<213> Homo sapiens

<400> 10

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

5
10
15
20
25
30

```

gagtgaagact gtctcagaac agcaacaaca aaatgcccgc tgetgctggg tccagaagag 180
cttgaataac tgcattgtct tttctcaat tttcatttcc cagaactggg caectccggg 240
ctgtgaaaag ttagggaggt gtctgacacc tccagaatcc attcccaaga agtgcctctg 300
gtcccactag caectgogca gactcaggee aggcctagaa tctccagttg gccctgcaag 360
tgcctggagg aaggatggct ctggcctcgg tctcccccac accctgccca agccagacag 420
acagcacctg cagacgcagg gggactgcac aattccacct gcccaggacc tgaccctggc 480
gtgtgcttgg cctctctct cgcacagggc gcctcagatt tcaggacct cctcctcgcc 540
cacggcgccct cagacctcag gacctgcgc tctcagcct ttgtgaacc caaatatctg 600
agaccagtct cagtttattt tgccaagggt aaggatgcac ctgtgacagc ctcaggagggt 660
cctgacaaca ggtgcccgag gtggctgggg atacagtttg cctttataca tcttagggag 720
acacaagatc agtatgtgta tgggtacat tggttcagtg agccttccac tgaatacacg 780
attgagtctg gccagtgaa tccgatttt tatgtaaca gtaaggggac ggggcaatca 840
tataagcgtt tgtctcaggg gagccccaga gggatgaatt ccagttccgt ctgtcctttg 900
tccacaagga atttccttg ggcctaatta tgaggggagg gtgtagcttc ttatcattgt 960
agctatgta tttgaaata aaacgggagg caggtttgcc taattcccag gttg 1014

```

<210> 11
<211> 522
<212> PRT
<213> Homo sapiens

<400> 11

35
40
45
50
55
60
65

```

Met Ala Ala Val Val Ala Ala Thr Arg Trp Trp Gln Leu Leu Val 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

```

5
10
15
20
25
30
35
40
45
50
55
60
65

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> ADN
 <213> Homo sapiens
 <400> 12

ggaatttcagg tgggctcttc gcccttcagg tacctgtggg tccctcogtc cegtccectc 60
 cgggggtcac cccggagcct gtcctgatg cggctcctgc ctctagcccc aggtcggctc 120
 cggcggggca gcccccgcca cctgcctccc tgcagcccag cgtctctact gctgggtgctg 180
 ggccgctgcc tgggggtctt cgggggtggt gcgggaaccc ggaggcccaa cgtgggtgctg 240
 ctctcaccgg acgaccagga cgaagtgttc ggccgcatga caccactaaa gaaaaccaaa 300
 gctctcatcg gagagatggg gatgaatttt tccagtgttt atgtgccaa gctctctctg 360
 tgccccagca gagccagtat cctgacagga aagtacccac ataactatca cgttggtgac 420
 aacactctgg aggggaactg cagtgtgaag tcttgccaga agatccaa ga accaaatact 480
 ttcccagcaa ttctcagatc aatgtgtggt taccagacct tttttgcagg gaaatattta 540
 aatgagtacg gagccccaga tgcaggtgga ctagaacaag ttctctctgg ttggagttaa 600
 tgggtatgct tggaaaagaa ttctaagtat tataattaca cctgtctat caatgggaag 660
 gcacgggaag atggtgaaaa ctatagtgtg gactacctga cagatgtttt ggctaattgt 720
 tccttggaat ttctggacta caagtcacac tttagacct tottcatgat gatcgccact 780
 ccagcgcctc attcgcttg gacagctgca cctcagtaac agaaggcttt ccagaatgtc 840
 tttgcaccaa gaaacaagaa ettcacatc catggaacga acaagcactg gtttaattagg 900
 caagccaa ga ctccaatgac taattcttca atacagtttt tagataatgc atttaggaaa 960
 aggtggcaaa ctctctctc agttgatgac ctgtggaga aactggtcac gaggtggag 1020
 ttcaactggg agctcaacaa caettacatc ttctataact cagacaatgg ctatcacaca 1080
 ggacagtttt ccttgccaat agacaagaga cagctgtatg agtttgatat caaagttcca 1140
 ctgttggttc gaggacctgg gatcaacaa aatcagacaa gcaagatgct ggttgccaac 1200
 attgaacttg gtctactat tttggacatt gctggctacg acctaaataa gacacagatg 1260
 gatgggatgt ccttattgac ctttttgaga ggtgccagta acttgacctg gcatcagat 1320
 gtcttggtgg aataccaagg agaaggcctg aacgtcactg acccaacatg ccttccctg 1380
 agtctgggct tatctcaatg ctccacagac tgtgtatgtg aagatgctta taacsatacc 1440
 tatgctgtg tgaggacaat gtcagcattg tggattttgc agtattgaga gtttgatgac 1500
 caggaggtgt ttgtagaagt ctataatctg actgcagacc cagaccagat cactaacatt 1560
 gctaaaacca tagaccaga gcttttagga aagatgaact atcggttaat gatgttacag 1620
 tctgtttctg ggccaacctg tgcactcca ggggtttttg acccgggata caggtttgac 1680
 cccgtctca tgttcagcaa toggggcagt gtcaggactc gaagattttc caaacattct 1740

5
10
15
20

```

ctgtagegac ctcacacagc ctctgcagat ggatccctgc acgcctcttt ctgatgaagt 1800
gattgtagta ggtgtctgta gctagtcttc aagaccacac ctggaagagt ttctgggctg 1860
gctttaagtc ctgtttgaaa aagcaacca gtcagctgac ttcctcgtgc aatgtgttaa 1920
actgtgaact ctgcccatgt gtcaggagtg gctgtctctg gtctcttctt ttagctgaca 1980
aggacactcc tgagggtctt gttctcactg tttttttttt atcctggggc cacagttctt 2040
gattatctct cttgtgggta aagactgaat ttgtaaacco attcagatan atggcagtac 2100
tttaggacac acacaaacac acagatacac cttttgatat gtaagcttga cctaaagtca 2160
aaggacotgt gtagcatttc agattgagca ctccactatc aaaaatacta acatcacatg 2220
gcttgaagag taaccatcag agctgaatca tccaagtaag aacaagtacc attgttgatt 2280
gataagtaga gatacatttt ttatgatgtt catcacagtg tggttaagggt gcaaatccaa 2340
aacatgtcac ccaagctctg ttcatgtttt tgtgaattc 2379

```

<210> 13
<211> 552
<212> PRT
<213> Homo sapiens

<400> 13

30
35
40
45
50
55
60
65

```

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

```


5 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

 10 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

 15 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

 20 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

 30 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

 40 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

 50 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

 55 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

 60 Pro Glu Leu Leu Gly Lys Met Asn Tyr Arg Leu Met Met Leu Gln Ser
 500 505 510

 65

5 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

10 Arg Arg Phe Ser Lys His Leu Leu
545 550

<210> 14
<211> 2022
15 <212> ADN
<213> Homo sapiens

<400> 14

20 ccgggtaccgg ctctctctgg gctccctcta ggccttccc ccgggcccca ctgcctggtc 60
agcgccaagt gacttaagcc ccgaacctg agcccggaac gctaggcgag gaggatcaga 120
tctccgctcg agaactctaa ggtgccttg tctggagga gtcccgccc agcctcgagg 180
25 tctcccggtg ctgctcgccc cggccctctg gacttcagg agggggccgt cagggtcggg 240
gagtatttgg gtcgggggtc tcagggaagg gcggcgctg ggtctcggt atcgaaaga 300
gcctgctgga gccaaagtag cctccctctc ttgggacaga cccctcggtc ccatgtccat 360
30 gggggcaccg cggctccctc tctggcctt ggtgctggc ctggcgttg ccgctccgcc 420
caacatcgtg ctgactcttg ccgacgacct cggctatggg gacctgggt gctatgggca 480
ccccagctct accactccca acctggacca gctggcgggg ggagggtgc ggttcacaga 540
cttctacgtg cctgtgtctc tgtgcacacc ctctaggggc gccctcctga cgggcgggt 600
35 ccgggttgg atgggcatgt acctggggt cctgggtgcc agctccggg ggggcctgcc 660
cctggaggag gtgacctgg ccgaagtctt ggtgcccga ggtacctca caggaaatggc 720
cggaagtgg cactctgggg tggggcctga gggggccttc ctgcccccc atcagggtt 780
40 ccctcgattt ctaggcctc cgtactccca cgaaccaggc cctgcccaga acctgacctg 840
cttcccgccg gccactcctt ggaagggtg ctgtgaccag ggcctgggtc ccatccact 900
gttggccaac ctgtcctgg aggcgcagcc cccctgggtg ccggactag agggcgcta 960
45 catggttttc gccatgacc tcatggcca cgcacagcgc caggatcgcc ccttcttct 1020
gtactatgcc tctaccaca ccaactacc tcagttcagt ggcagagct ttgcagagcg 1080
ttcaggccgc gggccatttg gggactcct gatggagct gatgcagct tggggacct 1140
gatgacagcc ataggggacc tggggctgt tgaagagac ctggtcatct tcaactgcaga 1200
50 caatggacct gagaccatgc gtatgtccg aggcgggtgc tccggtctct tgggtgttg 1260
aaaggggacg acctacgagg ggggtgtcc agagcctgcc ttggccttct ggcagggtca 1320
tatcgctccc ggcgtgacc acgagctggc cagctccctg gacctgctgc ctacctggc 1380
55 agccttggct ggggccccac tgcacaatgt cacttggat ggctttgacc tcagccccct 1440
gctgctgggc acaggcaaga gccctggcca gtctctcttc ttctacctgt cctacctaga 1500
cgaggtccgt ggggtttttg ctgtgggac tggaaagta aggtctcact tcttaccaca 1560
60 gggctctgcc cacagtata ccaactgcga cctgctgct cgcgctcca gctctctgac 1620
tgctcatgag ccccgctgc tctatgacct gtccaaggac cctggtgaga actacaacct 1680
gctggggggt gtggccgggg caaccccaga ggtgctgcaa gccctgaac agcttcagct 1740

65

5
10
getcaaggcc cagttagacg cagctgtgac ctteggccce agccaggtgg cccggggcga 1800
ggaccccgcc ctgcagatct gctgtcctcc tggetgcacc ccccgcccag cttgtgtcca 1860
ttgcccagat ccccatgect gaggccctct cggetggcct gggcatgtga tggetcctca 1920
ctgggagcct gtgggggagg ctcaggtgtc tggagggggt ttgtgectga taacgtaata 1980
acaccagtgg agacttgac atctgaaaaa aaaaaaaaa aa 2022

15
20
25
30
35
40
45
50
55
60
65
<210> 15
<211> 507
<212> PRT
<213> Homo sapiens
<400> 15

20 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
25 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
30 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
40 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
45 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
50 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
55 Met Ala Asp Ala Gln Arg Gln Asp Arg Pro Phe Phe Leu Tyr Tyr Ala
210 215 220
60 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> ADN
 <213> Homo sapiens
 <400> 16

acaaggatgg gtcggcgagg cgaggagagc ttgcccagag gccccggacc tcggcggtg 60
 ctctcccccg tcgtctctcc gctgctgctg ctgctgttgt tggcgccgcc gggctgggc 120

5
 10
 15
 20
 25
 30
 35
 40
 45
 50

```

gccccggccc gccggccgcc ccacctgggc ttcttggctgg cagacgacct aggtctggaac 180
gaactggggt tccacgggtc cgcctccgc acgcgcgacc tggacgcgct ggcggccggc 240
ggggtgctcc tggacaacta ctacacgcag ccgctgtgca cgcgctcgcg gagccagctg 300
ctcactggcc gctaccagat ccgtacaggc ttacagcacc aaataatctg gccctgtcag 360
cccagctgtg ttctctctga tgaaaaactc ctgcgccagc tccctaaaga agcaggttat 420
actaccata tggtcggaaa atggcacctg ggaatgtacc ggaagaatg ccttccancc 480
cgcggaggat ttgataccta ctttggatat ctctgggta gtgaagatta ttatcccat 540
gaacgctgta cattaattga cgtctgcat gtacacgat gtgctcttga ttttcggat 600
ggcgaagaag ttgcacagg atataaaat atgtattcaa caacatatt ccccaaaagg 660
gctatagccc tcataactaa ccatccacca gagaagctc tgtttctota ccttgcctc 720
cagtctgtgc atgagccct tcaggccct gaggaatact tgaagccata tgactttatc 780
caagacaaga acaggcatca ctatgcagga atggtgtccc ttatggatga agcagtagga 840
aatgtcactg cagctttaaa aagcagtggt ctctggaaac acacggtgtt catctttct 900
acagataacg ggggcagac ttggccagg ggtataact ggcctctcg aggaagaaa 960
tggagctgt ggaaggagg cgtccagggt gtgggtttt tggcaagccc cttgtgaag 1020
cagaagggtg tgaagaaccg ggcgtcctc cacatctctg actggctgac aacactctg 1080
aagctggcca ggggacacac caatggcaca aagctctggt atggcttcca cgtgtggaaa 1140
accatcagtg aagyaagccc atccccaga attgagctgc tgcataatat tgacccaaac 1200
ttegtggact ctccacgtg tcccaggaae agcatggctc cagcaaggga tgaactctt 1260
cttcagaaat attcagctt taacacatct gtccatgctg caattagaca tggaaattg 1320
aaactctca cgggtaccc aggtgtggt tactggttcc ctccaccgtc tcaatacaat 1380
gtttctgaga taccctcctc agaccacca accaagacc cctggctctt tgatattgat 1440
cgggacctg aagaaagaca tgacctgtcc agagaatate ctacacatct cacaagctc 1500
ctgtcccgcc tacagttcta ccataaacac tcagtcctcg tgaacttccc tgcacaggac 1560
ccccgtgtg atccccagge caetggggtg tggggccctt ggaatgtagga ttccaggag 1620
gttagaaaaa ctttcaattg gaagtggac ctccagcctt ttctcagac tctgtctca 1680
tttgttatcc caacctgggt tcacttggcc cttctcttgc tcttaacca caccaggtg 1740
tctaatttca accctaatg catttaagaa gctgataaaa tctgcaaac tctgtgtgt 1800
ggctggagca tgtgtctaga ggtgggggtg gctgggttta tcccccttcc ctaagcttg 1860
ggacagctgg gaacttaact tgaantagga agttctcact gaatcctgga ggcctggaac 1920
gctggctctt ttgactcac aagtcagacg ttogattccc ctctgccat agccagtttt 1980
attggagtga atcacattt ttacgcaaat gaaggagca gacagtgtt aatgggtctg 2040
ttggccagg cttctccctg tgggtgaagg atcatgttca ggcactccaa gtgaaccacc 2100
ctcttgggt cacccttcc tcaattatct catcacagag cataaggccc attttgtgt 2160
tcaggtcaac agcaaatgg cctgcacat gactgtggt tttaaaaa agaatgtgt 2220
ttttatcg 2228
  
```

<210> 17
 <211> 533
 <212> PRT
 <213> Homo sapiens
 <400> 17

5 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 Leu
 20 25 30
 10 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
 15 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
 25 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
 30 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
 35 Ser His Glu Arg Cys Thr Leu Ile Asp Ala Leu Asn Val Thr Arg Cys
 180 185 190
 40 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
 45 Asn His Pro Pro Glu Lys Pro Leu Phe Leu Tyr Leu Ala Leu Gln Ser
 225 230 235 240
 50 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
 55 Met Asp Glu Ala Val Gly Asn Val Thr Ala Ala Leu Lys Ser Ser Gly
 275 280 285
 60 Leu Trp Asn Asn Thr Val Phe Ile Phe Ser Thr Asp Asn Gly Gly Gln
 290 295 300
 65

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> ADN
 <213> Homo sapiens
 <400> 18

gectccagca getgaaggga cccagctgta gtgaggttgc agtgattgag taggattggc 60
 ctgcttcaaa gcagaggttt ctcattggaa tatgcttatt aaactccac tgggcagaa 120
 accatgaaca gaggatgaac aagtgaagtt gcaatctcct ccatcacagc tcagttcccc 180
 aacaacagga tcacaagctg gagatgcctt taaggagat gaagatccct ttctctctac 240
 tgttttttct gtgggaagcc gagagccacg cagcatcaag gccgaacatc atctgggtga 300
 tggctgacga cctgggcatt ggagatcctg ggtgctatgg gaacaaaact atcaggactc 360

5 ccaatatoga cccgttggcc agtgggggag tgaactcac tcagcacctg gcagcatcac 420
 cgcgtgtcac accaagcagg gcagccttca tgaactggcg gtacccctgtc cgatcaggaa 480
 tggcatcttg gtcccgcact ggagttttcc tcttcacagc ctcttcggga ggacttccca 540
 ccgatgagat tacctttgct aagcttctga aggatcaagg ttattcaaca gcactgatag 600
 10 ggaaatggca ccttgggatg agctgtcaca gcaagactga cttctgtcac caccctttac 660
 atcaaggctt caattatttc tatgggatct ctttgacca tctgagagac tgcaagcccg 720
 gagagggcag tgtcttcacc aagggttcca agaggctggc ctctctcccc ctgcagatcg 780
 tgggggtcac cctccttacc cttgctgcac tcaattgtct ggggtactc caagtgcctc 840
 15 taggcgtttt ttccagcctt ctctctctag cagccctaact cctgaccctt ttcttgggct 900
 tcttccatta cttccggccc ctgaactgct tcatgatgag gaactacgag atcattcagc 960
 agcccatgtc ctatgacaat ctcaccaga ggctaacggc ggaggcggcc cagttccatc 1020
 20 agcggaaacac tgagactcgg tctctgcttg tcttgctcta cctccacgtg cacacagccc 1080
 tgttctcccg caagactttt gctggcaaaa gtcaaacagg agtctacggg gatgctgttg 1140
 aggaatgga ctggagtgtg gggcagatct tgaaccttct ggatgagctg agattggcta 1200
 atgataccct catctacttc acatcggacc agggagcaca tgtagaggag gtgtcttcca 1260
 25 aaggagaaat tcctggcgga agtaatggga tctataaagg aggaaaagca acaactggg 1320
 aaggaggat cccgggttcca ggcctccttc gttggcccag ggtgatacag gctggccaga 1380
 agattgatga gcccaactagc aacatggaca tatttctac agtagccaag ctggctggag 1440
 30 ctcccttgcc tgaggacagg atcattgatg gaagtgatct gatccccctg cttgaaggaa 1500
 aaagccaaag ctccgateat gagtttctct tccattactg caacgcctac ttaaattgctg 1560
 tgcgtgggca cctcagaaac agcacatcca tctggaagge cttttctctc acccccaact 1620
 35 tcaaccccggt ggggttccaa ggtatgcttg ccacacaagt gtgcttctgt ttccggagtt 1680
 atgtcaccca tcacgacca cctttactct ttgatatttc caaagatccc agagagagaa 1740
 acccaactaa tccagcatcc gagccccggc ttatgaaat cctcaaagtc atgcaggaa 1800
 ctggggacag acacaccag accctgccag aggtgccga tcagttttca tggaaacaat 1860
 40 ttctttggaa gccctggctt cagctgtgct gtccttccac cggcctgtct tgccagtgtg 1920
 atagagaaaa acaggataag agactgagcc gctagcagcg cctggggacc agacagacgc 1980
 atgtggcaaa gctcaccato ttcactacaa acacgcctga gagtggcact ggggaacat 2040
 45 aactccatct acaccttggg ttggactga ttctccattt tctacctga aggcttgggc 2100
 cagagctcaa cagctactca actggagggg tgagggggat aaggctctga gtatacagac 2160
 aggaagatgg taggtttatg ccttctgtgg ccagagtctt ggactcatgg aaatagaatg 2220
 50 aatagagggg cattcacaag gcacaccagt gcaagcagat gacaaaaagg tgcagaagga 2280
 aatcttaaaa cagaagggtg caggaggtac ettaactcac cctcagcaa atacctatgt 2340
 caacagtata agttaccatt tactctataa tctgcagtga tgcaataacc agcataataa 2400
 a 2401
 55 <210> 19
 <211> 583
 <212> PRT
 <213> Homo sapiens
 60 <400> 19
 65

5 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

10 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

15 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

20 Ser Arg Thr Gly Val Phe Leu Phe Thr Ala Ser Ser Gly Gly Leu Pro
100 105 110

25 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

30 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

35 Val Phe Thr Thr Gly Phe Lys Arg Leu Val Phe Leu Pro Leu Gln Ile
180 185 190

40 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

45 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

50 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

55 Val His Thr Ala Leu Phe Ser Ser Lys Asp Phe Ala Gly Lys Ser Gln
290 295 300

60 His Gly Val Tyr Gly Asp Ala Val Glu Glu Met Asp Trp Ser Val Gly
305 310 315 320

5

10

15

20

25

30

35

40

45

50

55

60

65

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> ADN

<400> 20

ggaagccttg gcaetagegg cggccggggc cggagtgcgc agggcaaggt cctgcgetet 60

5
10
15
20
25
30
35
40
45

```

ggggcagcgc tggggcatgc gatcgcgcgc ggcgagggga cgcgcgcgcgc ccgcgcgcag 120
ggactcctttg ccggtgctac tgtttttatg cttgcttctg aagacgtgtg aacctaaaaa 180
tgcaaatgcc tttaaacca atactactact gatcatggcg gatgatctag gcactgggga 240
tctcggttgc tacgggaaca atacctgag aagccgaat attgaccagc ttgcagagga 300
agggtgtgagg ctcactcagc acctggcggc cgcgcgcgc tgcaccccaa gccagagctgc 360
attcctcaca gggagacatt ccttcagatc aggcctggac gccagcaatg gataccgggc 420
ccttcagtg gacgcaggct caggtggact ccttgagaa gaaaccaact ttgcaggaat 480
cttgacagcag catggetatg caacggcct cataggaaaa tggcaccagg gtgtgaattg 540
tgcatccgcg ggggataact gccaccacc cctgaacca ggatttgact atttctacgg 600
catgccttc acgtcacaa acgactgtga cccaggcagg ccccccgaag tggacgcgc 660
cctgagggcg cagctctggg gttacacca gttcctggcg ctggggattc tcacctggc 720
tgcgggcag acctgcggtt tcttctctgt ctccgcgaga gcagtcacgc gcctggcgcg 780
cgtgggctgc ctgtttttca tctcttggtc ctctccttc gggtttctgc gacgtggaa 840
ctgtatcctg atgagaaacc atgacgtcac ggagcaacce atggttctcg agaaaccagc 900
gagttctatg ctaaagggaag ctgtttccta tattgaaaga cacaagcatg ggcatttct 960
cctcttctct tctttgtgc atgtgcacat tcccttctg accacagatg catctctgg 1020
gaaaagtcag catgcttat atggtgataa tgtggaggag atggactggc tcataggtaa 1080
ggttcttaac gccatcgaag acaatggtt aaagaactca acattcagct atttcacctc 1140
tgaccatgga ggcattttag aggcagaga tggacacgc cagttagggg gatggaacgg 1200
aatttacaaa ggtgggaagg gcctgggagg atgggaagggt gggatccag tgcgcgggat 1260
cttccactgg ccgggggtgc tcccgcccg ccgagtgatt ggagagccca cagacctgat 1320
ggacgtgttc ctaactgttg tccagctggt ggggtggcgag gtgccccagg acagggtgat 1380
tgatggccac agcctggtac ccttctgca gggagctgag gcacgctcgg cacatgagtt 1440
cctgtttcat tactgtgggc agcatcttca cgcagcagc tggcaccaga aggacagtgg 1500
aagcgtctgg aaggttcatt acacagccgc gcagttccac cccgaggagc ggggctgct 1560
aacggcgcag gcgtctgcc atgtgaatg gggaggcgtg acccatcaca gaccccttt 1620
getctttgac ctctccaggg accctccga ggcacggccc ctgaccccg actccagccc 1680
cctgtaccac gccgtgatag caagggtagg tgcgcgggtg tggagcacc gccagacct 1740
gagtcctgtg cccagcagt tttccatgag caacatctg tggagccgt ggtgcagcc 1800
gtgtgcgga catttccgt tctgttcag ccacagggat ggggatggca cccctgaat 1860
gccaggactg tgagagagga tccaggagag cctgaactgc ttgcaacaa aattctcaa 1920
gettggttct atcttcagtc cggaa 1945

```

50
55

```

<210> 21
<211> 593
<212> PRT
<213> Homo sapiens

<400> 21

```

60
65

```

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

5 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
 10 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
 15 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
 20 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
 25 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
 30 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
 35 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
 40 Cys Gly His Phe Pro Phe Cys Ser Cys His Glu Asp Gly Asp Gly Thr
 580 585 590
 Pro

<210> 22

45 <211> 1858

<212> ADN

<213> Homo sapiens

50 <400> 22

55 ccttctctctt ettgateggg gattcaggaa ggagccagg agcagagga gtagagagag 60
 agacaacatg ttacatctgc accattcttg ttgtgtttc aggagctggc tgccagcgat 120

55

60

65

5
10
15
20
25
30
35
40
45

```

gctcgtgta ctgctaagtt tggcaccate agcttccage gacatttccg cctcccgacc 180
gaacatcctt cttctgatgg cggacgacct tggcattggg gacattgggt gctatggcaa 240
caacaccatg aggsctccga atattgaccg ccttgacagag gacggcggtg agctgacce 300
acacatctct gcgcctctt tgtgcacccc aagcagagcc gccttctca cgggcagata 360
ccttgctgga tcagggatgg tttccagcat tggttaccgt gttcttcagt ggaccggage 420
atctggaggt cttccaacaa atgagacaa ttttgcaaaa atactgaaag agaaaggcta 480
tgccactgga ctcatggaa aatggcatct ggtctcaac tgtgagtcag ccagtgatea 540
ttgccaccac cctctccate atggttttga gcatttctac ggaatgcctt tctccttgat 600
gggtgattgc gcccgctggg aactctcaga gaagcgtgtc aacctggaa aaaaactcaa 660
cttctcttct caagtctctg ccttgggttg cctcacactg gtacagggga agctcacaca 720
cctgataccc gtctcgtgga tgccggtcat ctggtcagcc ctctcgccg tctctctct 780
cgcaagctcc tattttgtgg gtgtctgat tgtccatgcc gattgcttto tgatgagaaa 840
ccacaccate accggagcgc ccagtgtctt ccaagaacg acacctta ttctgcagga 900
ggttgcttc tttctcaaaa ggaataagca tgggccttct cctctctttg tttcctttct 960
acacgttca atcctctta tcactatgga gaacttctc gggagagtc tccacgggt 1020
gtatggggac aacgtagagg agatggactg gatggtagga cggatccttg acactttgga 1080
cgtggagggg ttgagcaaca gcacctcat ttattttacg tcggatcaag gcggttccct 1140
agagaatcaa cttggaacaa ccagtatgg tggctggaat ggaattata aaggtgggaa 1200
gggcattgga gatgggaag gtgggatccg cgtgccggg atcttccgt ggcgggggt 1260
gtcccgccg gcccgagtga ttggcgagcc cagcagctg atggaagtg tccccacct 1320
ggtccggctg ggcggcgccg aggtgccca ggcagagtg attgacgcc aagaacttct 1380
gccttctct ctggggacag cccaacactc agaccagag tctctgatgc attattgtga 1440
gaggtttctg cagcagcca ggtggcatc acgggacaga ggaacaatgt ggaagttca 1500
cttctgtacg cctgtgttcc agccagagg agccggtgcc tgcctatgga gaaaggtctg 1560
ccgtgtctt ggggaaaaag tagtccacca cgtccacct ttgtctttg acctctcaag 1620
agaccttct gagaccaca tctctacac agcctcagag ccctgttct atcaggtgat 1680
ggaacagag cagcagggg tgtgggaaca ccagcgaca ctacgccag tctctctga 1740
gttgacagc ctgggcaaca tctggagacc gtggtgcag cctgtgtg gcccgttcc 1800
cctctgtg tgccctaggg aagatgacc acaataaat tctgcagtga aaagctgg 1858

```

<210> 23
<211> 589
<212> PRT
<213> Homo sapiens

<400> 23

55
60
65

```

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
 5 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
 10 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
 15 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
 20 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
 25 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
 30 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
 35 Phe Pro Leu Cys Trp Cys Leu Arg Glu Asp Asp Pro Gln
 580 585

<210> 24

<211> 1996

<212> ADN

<213> Homo sapiens

<400> 24

45 gggttctgct cctagacatt agagagataa tacggtgat agacacaaag aaggtattcc 60
 aagctgcaca atgaggccca ggagaccgtt ggtcttcattg tctttggtgt gtgcactctt 120
 gaacacatpg ccagggcaca caggggtgat gacgacaagg cctaatattg tectaatcat 180
 50 gggtgatgac ctgggtattg gagatctggg ctgctacggc antgacacca tgaggacgac 240
 tcacatcgac cgccttgcca ggaaggcgt ggaactgact cagcacctct ctgcgcctc 300
 cctctgcagg ccnagccggt ccggttctt gaagggaaga taccacatcc gatcaggtat 360
 55 gggttctagt ggtaatagac gtgtcatoca aattcttgca gtccccgcag gctccctct 420

5	taatgagaca	acatttgcag	cettgctaaa	gaagcaagga	tacagcacgg	ggcttatagg	480
	caaattggca	caaggcttga	actgcgactc	cgaagtgc	cagtgcaccc	atccatataa	540
	ttatgggttt	gactactact	atggcatgcc	gttccctctc	gttgacagct	gctggccgga	600
	ccctctctgt	aacacggaat	tagcctttga	gagtcagctc	tggctctgtg	tgcagctagt	660
10	tgcatttgc	atcctcacc	taacctttg	gaagctgagc	ggctgggtct	ctgttcctcg	720
	gtctctgate	ttctccatga	ttctgtttat	ttctctcttg	ggctatgctt	ggttctccag	780
	ccacacgtcc	cctttatact	gggactgcct	cctcatgogg	gggcacgaga	tcaaggagaa	840
	gccccatga	gctgaacgag	ctggatccat	tatggtgaag	gaagcgattt	cctttttaga	900
15	aaggcacagt	aaggaaactt	tcctctctct	ttctctcttt	cttcacgtgc	acacacctct	960
	ccccaccacg	gacgatttca	ctggcaccag	caagcatggc	ttgtatgggg	ataatgtgga	1020
	agagatggac	tccatgggtg	gcgaagattct	tgatgcctatc	gatgattttg	gcctaaaggaa	1080
	caacacccctt	gtctacttta	cctcagatca	cggaggggcat	ttggaagcta	ggcgagggca	1140
20	tgcctcaactt	ggtaggatga	atggaatata	caaaggttga	aaaggcatgg	ggggctggga	1200
	aggttgaatc	cgcgtcccag	gaattgtccg	atggccttga	aaggtaccag	ctggacgggt	1260
	gattaaggaa	cctacaagtt	taatggatat	tttaccaaact	gtgcctcag	tgtcaggagg	1320
	aagtctccct	caggacaggg	tcattgaagg	cagagacctc	atgcccttgc	tgcaggggca	1380
25	cgtcaggcac	toggagcatg	aatttctttt	ccactactgt	ggctctctacc	tgcacggcgt	1440
	gggttgatc	cccaaggacg	acagtgggtc	agttttggaag	gtcactatg	tgaacccgggt	1500
	attccagcca	ccagcttctg	gtggttgcct	tgtcaccctc	ttatgcagat	gtttcggaga	1560
	acaggttacc	taccacaacc	ccctctgtct	cttcgatctc	tccagggacc	cctcagagtc	1620
30	cacaccccctg	acacctgcca	cagagccctt	ctatgatttt	gtgattaaaa	aggtggccaa	1680
	cgccttgaag	gaacaccagg	aaacctatct	gcctgtgacc	taccaactct	cagaactgaa	1740
	tcagggcagg	acgtggctga	agccttgcct	tgggggtgtc	ccattttgtc	tgtgtgacaa	1800
	ggaagaggaa	gtctctcagc	ctgggggtcc	taacgagaag	agataattac	aatcaggcta	1860
40	ccagagggaag	ccttttggtcc	taacgagaag	agataattac	aatcaggcta	ccaaagggaag	1920
	cactaacttt	ggtagctttca	agttggcaag	gagtgcaatt	aatagtcaat	aaattcatct	1980
	acatttccag	attatt					1996

```

45      <210> 25
          <211> 591
          <212> PRT
          <213> Homo sapiens
50      <400> 25

```

55	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		
60	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				

5	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	
10	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	
15	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
20	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	
25	Cys Val Gln Leu Val Ala Ile Ala Ile Leu Thr Leu Thr Phe Gly Lys	195	200	205	
30	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
35	Pro Leu Tyr Trp Asp Cys Leu Leu Met Arg Gly His Glu Ile Thr Glu	245	250	255	
40	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	
45	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
50	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	
55	Ala Arg Arg Gly His Ala Gln Leu Gly Gly Trp Asn Gly Ile Tyr Lys	355	360	365	
60	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> ADN

<213> Homo sapiens

<400> 26

atgggctggc tttttctaaa ggttttgttg gggggagtga gtttctcagg atttctttat 60
 cctettgttg atttttgcat cagtgggaaa acaagaggac agaagccaaa ctttgtgatt 120
 attttggcgg atgacatggg gtgggggtgac ctggggagcaa actgggcaga aacaaggagc 180
 actgccaacc ttgataagat ggtctcggag ggaatgaggt ttgtggattt ccatgcagct 240
 gcctccaaat gctcaccctc ccgggcttcc ttgtcaccg gccggcttgg ccttcgcaat 300
 ggagtccac gcaactttgc agtcacttct gtgggaggcc ttccgtccaa cgagaccacc 360
 ttggcagagg tgcgcagca ggcgggttac gtcactggga taataggcaa atggcatctt 420
 ggacaccaag gctcttata ccccaacttc cgtgggtttg attactactt tggaaatcca 480
 tatagccatg atatgggctg tactgatact ccaggctaca accaccctcc ttgtccagcg 540

5
10
15
20
25

```

tgctccacagg gtgatggacc atcaaggaac ctccaagag actgttacac tgacgtggcc 600
ctccctcttt atgaaaacct caacattgtg gagcagccgg tgaacttgag cagcettgce 660
cagaagtatg ctgagaaagc aacccagttc atccagcgtg caagcaccag cgggaggccc 720
ttctgtctct atgtggctct gggccacatg cagtgccct tacctgtgac tcagctacca 780
gcagcgccac ggggcagaag cctgtatggt gcagggtctt gggagatgga cagtctggtg 840
ggccagatca aggacaaagt tgaccacaca gtgaaggaaa acacattcct ctgggtttaca 900
ggagacaatg gcccggtggc tcagaagtgt gagctagcgg gcagtgtggg tcccttcaact 960
ggattttggc aaactcgtca agggggaagt ccagccaagc agacgacctg ggaaggaggg 1020
caccgggtcc cagcactggc ttactggcct ggcagagttc cagttaatgt caccagcaact 1080
gccttggtta gctgtgtgga cttttttcca actgtggtag ccttggtcca ggcagotta 1140
cctcaaggac ggcgctttga tgggtgtggc gtctccgagg tgccttttgg cgggtcacag 1200
cctggggcaca ggggtgtgtt ccaccccaac agcggggcag ctggagagtt tggagccctg 1260
cagactgtcc gctggagcgt ttacaaggcc ttctacatta ccggtggagc cagggcgtgt 1320
gatgggagca cggggcctga gctgcagcat aagtctctc tgattttcaa cctggaagac 1380
gataccgcag aagctgtgcc cctagaaaga ggtggtgcgg agtaccaggc tctgtgtccc 1440
gaggtcagaa aggttcttgc agcgtctc caagacattg ccaacgacaa catctccage 1500
gcagattaca ctcaggaccc ttcagtaact cctgtgtgta atccctacca aattgcctgc 1560
cgctgtcaag ccgcataa 1578

```

30
35

```

<210> 27
<211> 525
<212> PRT
<213> Homo sapiens

<400> 27

```

40
45
50
55
60
65

```

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
 5 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
 10 Arg Asp Cys Tyr Thr Asp Val Ala Leu Pro Leu Tyr Glu Asn Leu Asn
 195 200 205
 15 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
 20 Phe Leu Leu Tyr Val Ala Leu Ala His Met His Val Pro Leu Pro Val
 245 250 255
 25 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
 30 His Thr Val Lys Glu Asn Thr Phe Leu Trp Phe Thr Gly Asp Asn Gly
 290 295 300
 35 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
 40 Trp Glu Gly Gly His Arg Val Pro Ala Leu Ala Tyr Trp Pro Gly Arg
 340 345 350
 45 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
 50 Arg Phe Asp Gly Val Asp Val Ser Glu Val Leu Phe Gly Arg Ser Gln
 385 390 395 400
 55 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
 60 Ile Thr Gly Gly Ala Arg Ala Cys Asp Gly Ser Thr Gly Pro Glu Leu
 435 440 445
 65

5 Gln His Lys Phe Pro Leu Ile Phe Asn Leu Glu Asp Asp Thr Ala Glu
 450 455 460
 10 Ala Val Pro Leu Glu Arg Gly Gly Ala Glu Tyr Gln Ala Val Leu Pro
 465 470 475 480
 15 Glu Val Arg Lys Val Leu Ala Asp Val Leu Gln Asp Ile Ala Asn Asp
 485 490 495
 20 Asn Ile Ser Ser Ala Asp Tyr Thr Gln Asp Pro Ser Val Thr Pro Cys
 500 505 510
 25 Cys Asn Pro Tyr Gln Ile Ala Cys Arg Cys Gln Ala Ala
 515 520 525

<210> 28

<211> 4669

<212> ADN

<213> Homo sapiens

<400> 28

25 gcagacccgt cgctaataaa tottggggcc ggtgtcgggc cggggcggtt tgatcgccaa 60
 ctaggaaacc ccaggcgcag aggccaggag cgagggcagc gaggatcaga ggccaggcct 120
 30 tcccgggtgc cggcgctcct cggaggtcag ggcagatgag gascctgact ccccccttc 180
 ggaggaggaa ggaagtcccg ctgccacctt atctctgtct ctctgcctcc tccctgttcc 240
 cagagctttt tctctagaga agattttgaa ggcggctttt gtgttgacgg ccaccaccca 300
 tcctctaagc aagataaact tggcaaatga catgcagggt cttcaaggca gaataattgc 360
 35 agaaaatctt caaaggaccc tatctgcaga tgtttgaat acctctgaga atagagattg 420
 attattcaac caggatacct aattcaagaa ctccagaaat caggagacgg agacattttg 480
 tcagttttgc aacattggac caaatacaat gaagtattct tgctgtgtct tggtttttgc 540
 40 tgtctcgggc acagaattgc tgggaagcct ctgttcgact gtcagatccc cgagggttcag 600
 aggacggata cagcaggaac gaaaaaacat ccgacccaac attattcttg tgcttaccga 660
 tgatcaagat gtggagctgg ggtccctgca agtcatgaa aacacgagaa agattatgga 720
 acatgggggg gccaccttca tcaatgcctt tgtgactaca cccatgtgct gcccgtaacg 780
 45 gtcctccatg ctccacggga agtatgtgca caatcacaa gtctcaccca acaacgagaa 840
 ctgtctcttc cctcgttggc aggcacatga tgagcctcgg acctttgtgt tatatcttaa 900
 caacactggc tacagaacag ccttttttgg aaataacctc aatgaatata atggcagcta 960
 50 cateccccc cgggtggcag aatggcttgg attaatcaag aattctcgtc totataatta 1020
 cactgtttgt cgcattggca tcaaaagaaa gcattggattt gattatgcaa aggactactt 1080
 cacagactta atcactaacg agagcattaa ttacttcaaa atgtctaaga gaattgtatc 1140
 55 ccataggccc gttatgatgg tgatcagcca cgtgcgcgcc caccggcccc aggaactcagc 1200
 cccacagttt tctaaaactgt accccaatgc ttcccaacac ataactccta gttataacta 1260
 tgcaccaaat atggataaac actggattat gcagtaacaa ggaccaatgc tgcccatcca 1320
 catggaattt acaaacattc tacagcgcaa aaggtctcag accttgatgt cagtggatga 1380
 60 ttctgtggag aggtgtata acatgtctgt ggagacgggg gagctggaga atacttacat 1440
 catttacacc gcagaccatg gttaccatat tgggcagttt ggactggta aggggaaatc 1500
 catgccatat gaatttgata ttctgtgtgc tttttttatt cgtgggtcaa gtgtagaacc 1560

	aggatcaata gteccacaga togtttctca cattgaattg gecccacaga tcoctggatat	1620
	tgetgggctc gacacacctc ctgatgtgga cggcaaglet gtectcaaac ttctggaccc	1680
5	agaaaagcca ggtaacaggt ttctgaacaa caagaaggcc aaattttggc gtgatacatt	1740
	cctagtggaa agaggcaaat ttctacgtaa gaaggaagaa tccagcaaga atatccaaac	1800
	gtcaaatcac ttgcccaaat atgaacgggt caaagaacta tgccagcagg ccaggtacca	1860
10	gacagcctgt gaacaacogg ggcagaagtg gcaatgcatt gaggatacat ctggcaagct	1920
	togaattcac aagtgttaag gaccagtgga cctgtctaca gtccggcaga gcaacoggaa	1980
	cctctacgtc cgcggcttcc atgacaaaga caaagagtgc agttgtaggg agtctggtta	2040
15	cogtgcagc agaagocaaa gaaagagtea acggcaatto ttgagaaac aggggactcc	2100
	aaagtacaag cccagatttt tccatactcg gcagacacgt tcoctgtccg tcgaatttga	2160
	aggtgaataa tatgacataa atctgggaaga agaagaaga ttgcaagtgt tgcaacccag	2220
20	aaacattgct aagcgtcatg atgaaggcca caagggggcca agagatctcc aggcctccag	2280
	tggtggcaac agggggcagga tgctggcaga tagcagcaac gcctggggcc cactatccac	2340
	tgtecgagtg acaacacagt gttttattct tcccaatgac tctatccatt gtgagagaga	2400
25	actgtaccaa toggccagag cgtggaagga cctaaggcca tacattgaca aagagattga	2460
	agctctgcaa gataaaatta agaatttaag agaagtgaga ggacatctga agagaaggaa	2520
	gectgaggaa tgtagctgca gtaaaccaag ctattacaat aaagagaaag gtgtaaaaaa	2580
30	gcaagagaaa ttaaagagcc atcttcaccc attcaaggag gctgtcagg aagtagatag	2640
	'cnaactgcaa cttttcaagg agaacaaccc taggaggag agggagagga aggagagag	2700
	acggcagagg aaggggggag agtcagacct gectggcctc acttgcttca cgcattgaca	2760
35	caaccactgg cagacagccc cgttctggaa cctgggatct ttctgtgctt gcacaggttc	2820
	taacataaac acctactggt gtttgcgtac agttaatgag acgcataatt ttcttttctg	2880
	tgagtttgct actggctttt tggagtattt tgatatgaat acagatccct atcagctcac	2940
40	aaatacagtg cacaaggtag aaagaggcat tttgaatcag ctacaogtac aaactaatga	3000
	gtcagaagc tgtcaaggat ataagcagtg caacccaaga cctaagaato ttgatgttg	3060
	aaataaagat ggaggaagct atgacctaca cagaggacag ttatgggatg gatgggaagg	3120
45	ttaatcagcc cgtctcact gcagacatca actggcaagg cctagaggag ctacacagtg	3180
	tgaatgaaaa catctatgag tacagacaaa actacagact tagtctgggt gactggacta	3240
	attacttgaa ggatttagat agagtatttg cactgctgaa gagtcaatat gagcaaaata	3300
50	aaacaaataa gactcaact gtcaaaagt acgggttctt ggttgtctct gctgagcag	3360
	ctgtgtcaat ggagatggcc tctgtgact cagatgaaga cccaaggcat aaggttggga	3420
	aaacacctca tttgacctg ccagctgacc ttcaaacctt gcatttgaa cgcaccaat	3480
55	taagtcagaa gagttaaact gaatggaata acgacattcc agaagttaat catttgaatt	3540
	ctgaacactg gagaaaaacc gaaaaatgga cggggcatga agagactaat catctggaaa	3600
	cagatttcag tggcgatggc atgacagagc tagagctcgg gccagcccc aggetgcagc	3660
60	ccattacag gcaccgaaa gaacttccc agtatggttg tcoctggaag gacatttttg	3720
	aagatcaact atatcttctt gtgcattccg atggaatttc agttcatcag atgttcacca	3780
	tggccaccgc agaacaccga agtaattcca gcatagcggg gaagatgttg accaaggtgg	3840
65	agaagaatca cgaagaggag aagtcacagc acctagaagg cagcacctcc tottcaactc	3900

5 cctctgatta gatgaaactg ttacettacc ctaaacacag tatttttttt taactttttt 3960
 atttgtaaac taataaagggt aatcacagcc accaacatte caagctaccc tgggtacett 4020
 tgtgcagtag aagctagtga gcattgtgagc aagcgggtgt cacacggaga ctcatcggtt 4080
 taatttaata tctgccaaga gtgaaagaa aggetgggga tatttgggtt ggettggtt 4140
 tgattttttg cttgtttgtt tgttttgtac taaaacagta ttatcttttg aatctcgtag 4200
 10 ggacataagt atatacatgt tatccaatca agatggetag aatggtgeet ttctgagtgt 4260
 ctaaaacttg acacccctgg taatctttt aacacacttc cactgectgc gtaatgaagt 4320
 ttgtattcat ttttaaccac tgggaatttt caatgcctgc attttcagtt agatgatttt 4380
 gcactttgag attaaaatgc catgtctatt tgattagtct tattttttta tttttacagg 4440
 15 cttatcctgc tcactgttgg ctgtcattgt gacaaagtc aataaaccac caaggacgac 4500
 acacagtatg gatcacatat tgtttgacat taagcttttg ccagaaaatg ttgcattgtt 4560
 tttaacctga cttgtataaa tegtattgca gaaaggcatg gctaataatg ttggttgtga 4620
 20 aataaataa ataagttaat gaaaaaaaa aaaaaaaaa aaaaaaaa 4680

<210> 29

<211> 871

<212> PRT

<213> Homo sapiens

<400> 29

30 Met Lys Tyr Ser Cys Cys Ala Leu Val Leu Ala Val Leu Gly Thr Glu
 1 5 10 15
 35 Leu Leu Gly Ser Leu Cys Ser Thr Val Arg Ser Pro Arg Phe Arg Gly
 20 25 30
 40 Arg Ile Gln Gln Glu Arg Lys Asn Ile Arg Pro Asn Ile Ile Leu Val
 35 40 45
 45 Leu Thr Asp Asp Gln Asp Val Glu Leu Gly Ser Leu Gln Val Met Asn
 50 55 60
 50 Lys Thr Arg Lys Ile Met Glu His Gly Gly Ala Thr Phe Ile Asn Ala
 65 70 75 80
 55 Phe Val Thr Thr Pro Met Cys Cys Pro Ser Arg Ser Ser Met Leu Thr
 85 90 95
 60 Gly Lys Tyr Val His Asn His Asn Val Tyr Thr Asn Asn Glu Asn Cys
 100 105 110
 65 Ser Ser Pro Ser Trp Gln Ala Met His Glu Pro Arg Thr Phe Ala Val
 115 120 125
 70 Tyr Leu Asn Asn Thr Gly Tyr Arg Thr Ala Phe Phe Gly Lys Tyr Leu
 130 135 140
 75 Asn Glu Tyr Asn Gly Ser Tyr Ile Pro Pro Gly Trp Arg Glu Trp Leu
 145 150 155 160
 80 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
 5 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
 10 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
 20 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
 25 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
 35 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
 45 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
 55 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
 5 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
 10 Ser Leu Ser Val Glu Phe Glu Gly Glu Ile Tyr Asp Ile Asn Leu Glu
 545 550 555 560
 15 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
 20 Gly Asn Arg Gly Arg Met Leu Ala Asp Ser Ser Asn Ala Val Gly Pro
 595 600 605
 25 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
 30 Asp His Lys Ala Tyr Ile Asp Lys Glu Ile Glu Ala Leu Gln Asp Lys
 645 650 655
 35 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
 40 Val Lys Lys Gln Glu Lys Leu Lys Ser His Leu His Pro Phe Lys Glu
 690 695 700
 45 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
 50 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
 55 Thr Ser Ser Asn Asn Asn Thr Tyr Trp Cys Leu Arg Thr Val Asn Glu
 770 775 780
 60 Thr His Asn Phe Leu Phe Cys Glu Phe Ala Thr Gly Phe Leu Glu Tyr
 785 790 795 800
 65

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> ADN
 <213> Homo sapiens
 <400> 30

gggccatttc tggacaacag ctgctatattt cacttgagcc caagttaatt tctcggggag 60
 ttctcggggcg cgcacaggca gctcggtttg ccttcgatt gagctcggg tcgcccggg 120
 cgcgggcctc tccaatggca aatgtgtgtg gctggaggcg agcgcgagcc ttctggcaca 180
 ggcagtcgag tgtttcgaga cggggcgag tctgtgaaa gcagataaaa gaaacattt 240
 attaacgtgt cattaagagg ggagcgccg gccggggctg tcgcactccc cgcggaaeat 300
 ttggtccct ccagctccta gagaggagaa gaggaaagcg gaaaagaggc agattcacgt 360
 cgtttccagc caagtggacc tgatcgatgg ccctctgaa ttatcaca tatttgattt 420
 attagcgatg cccctggtt tgtgtgttac gcacacacac gtgcacacaa ggctctggct 480
 cgttcctc cctcgtttcc agctcctggg cgaatccac atctgtttca actctccgc 540
 gaggcgagc aggagcgaga gtgtgtcgaa tctcgagtg aagaggagc agggaaaaga 600
 aacaaagcca cagaagcaac ttgagactcc cgcacccaa aagaagcacc agatcagcaa 660
 aaaaagaaga tgggcccccc gagcctcgtg ctgtgcttgc tgcccgcaac tgtgttctc 720
 ctgtcgggtg gaagctcggc ctctcgtgc caccacggc tgaaaggcag gtttcagagg 780
 gaccgcagga acatccgc ccaatcctc ctggtgctga cggacgacca ggatgtggag 840
 ctgggttcca tgcaggtgat gaacaagacc cggcgcatca tggagcaggg cggggcgcac 900
 ttcacaaag ccttcgtgac caccacctg tctgcccct cagctcctc catcctcacc 960
 ggcaagtacg tccacaacca caacacctac accaacaatg agaactgctc ctgcctctc 1020
 tggcaggcac agcagcagag ccgcacctt gccgtgtacc tcaatagcac tggctacagg 1080
 acagctttct tgggaagta tctaatgaa tacacggct cgtacgtgac acccggtgg 1140
 aaggagtggg tggactcct taaaaactc cgttttata actacagct gtgtcggaac 1200
 ggggtgaaag agaagcagcg ctccgactac tccaggtatt acctcacaga cctcctacc 1260
 aatgacagcg tgagcttctt ccgcagctc aagagatgt acccgcacag gccagtcctc 1320
 atggtcatca gccatgcag cccccaggg cctgaggatt cagccccaca atattcagc 1380
 ctcttcccaa acgcatctca gcacatcag ccgagctaca actacggcc caaccggac 1440
 aaacactgga tcatgccta cagggggccc atgaagcca tccacatgga attcaccac 1500
 atgtccagc ggaagcgctt gcagaccctc atgtgggtg agactccat ggagacgatt 1560

5
10
15
20
25
30
35
40
45
50
55
60
65

```

tacaacatgc tgyttgagac gggcgagctg gacaacacgt acatcgtata caccgcggac 1620
cacggttacc acatcgggcca gtttggcctg gtgaagggga aatccatgcc atatgagttt 1680
gacatcaggg tcccgtteta cgtgaggggc cccaacgtgg aagccggctg tctgaatecc 1740
cacatcgtec tcaacattga cctggccccc accatcctgg acattgcagg cctggacata 1800
cctgcggata tggacgggaa atccatcttc aagetgctgg acaeggagcg gccgggtgaat 1860
cggttttcact tgaataagaa gatgagggtc tggcgggact ctttcttggg ggagagaggg 1920
aagetgctac acaagagaga caatgacaag gtggacggcc aggaggagaa ctttctgccc 1980
aagtaccagc gtgtgaagga cctgtgtcag cgtgctgagt accagacggc gtgtgagcag 2040
ctgggacaga agtggcagtg tgtggaggac gccacgggga agctgaagct gcataagtcg 2100
aagggcccca tgcggctggg cggcagcaga gccctctcca acctcgtgcc caagtactac 2160
gggcaggcca gcgaggcctg cactgtgac agcggggact acaagctcag cctggccgga 2220
cgccggaaaa aactcttcaa gaagaagtao aaggccagct atgtccgag tgcctccate 2280
cgctcagtg ccatcgaggt ggaacggcagg gtgtaccacg taggcctggg tgatgcggc 2340
cagccccgaa acctcaccaa ggggcaactg ccaggggccc ctgaggacca agatgacaag 2400
gatggtgggg acctcagtg cactggaggc ctteccagct actcagccgc caacccatt 2460
aaagtgcac atcggtgcta catctagag aacgacacag tccagtgtga cctggacctg 2520
tacaagtccc tgcaggcctg gaagaccac aagetgcaca tgcaccagc gattgaacc 2580
ctgcagaaac aaattaagaa cctgagggaa gtccgaggtc acctgaagaa aaagcggcca 2640
gaagaatgtg actgtcaca aatcagctac cacacccagc acaaggccg cctcaagcac 2700
agaggctcca gtctgcatcc ttccaggaag ggcctgcag agaaggacaa ggtgtggctg 2760
ttgggggagc agaagcgcaa gaagaaactc cgcagctgc tcaagcgct gcagaacac 2820
gacacgtgca gcattgccagg cctcagctgc tccacccacg acaaccagca ctggcagacg 2880
gcgcctttct ggacactggg gcctttctgt gcctgcacca gcgccaacaa taacacgtac 2940
tggtgcabga ggaccatcaa tgagactcac aatttctct tctgtgaatt tgcaactggc 3000
ttctagagat accttgatct caacacagac cctaccagc tgatgaatgc agtgaacaca 3060
ctggacagg atgtectcaa ccagctacac gtacagctca tggagctgag gagctgcaag 3120
ggttacagc agtgaacccc ccggaactca aacatggacc tgggaactaa agatggagga 3180
agctatgagc aatacaggca gtttcagcgt cgaagtggc cagaantgaa gagacctct 3240
tccaaatcac tgggacaact gtgggaaggc tgggaagggt agaaaccaac agaggtggac 3300
ctccaaaaac atagaggcat cactgactg cacaggcaat gaaaaacat gtgggtgatt 3360
tccagcagac ctgtgctatt ggcaggagg cctgagaaag caagcaogca ctctcagtea 3420
acatgacaga ttctggagga taaccagcag gagcagagat aacttcagga agtccatttt 3480
tgccctgct tttgctttgg attatactc accagctgca caaatgcat ttttctgtat 3540
caaaaagtea ccaatacccc tccccagaa gctcacaag gaaaacggag agagcgagcg 3600
agagagattt ccttggaatt ttctcccaag ggogaagto attggaattt ttaaatcata 3660
ggggaasagc agtctgttc taaatctct tattcttttg gttgtcaca aagaaggaa 3720
taagaagcag gacagaggca aogtggagag gctgaaaaca gtgcagagac gtttgacaa 3780
gagtcagtag cacaasagag atgacattta cctagcatat aaacctggg tgcctctgaa 3840
gaaactgctt ccatgtata tatgtgacta ttacatgta atcaacatgg gaacttttag 3900

```

5
10
gggaacctaa taagaaatcc caatttttcag gagtggtggt gtcaataaac gctctgtggc 3960
cagtgtaaaa gaaaaaaaaa aaaaattgtg gacatttctg ttctgtcca gataccattt 4020
ctcctagtat ttctttgtta tgtcccgaa ctgatgtttt ttttttaag tactgaagg 4080
aaatgaagtt gatgtatgtc ccaagttttg atgaactgt atttgtaaa aaaattttgt 4140
agtttaagta ttgtcataca gtgttcasaa cccagccaa tgaccagcag ttggtatgan 4200
gaacctttga cattttgtaa aaggccattt cttgggggaa aaaaaaaaaa aaaaaaaaaa 4260
aaaaaaaaaa aaaaaaaaaa 4279

15
<210> 31
<211> 870
<212> PRT
<213> Homo sapiens
20
<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
 5 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
 10 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
 20 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
 30 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
 35 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
 45 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
 55 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
 60 Ala Ser Tyr Val Arg Ser Arg Ser Ile Arg Ser Val Ala Ile Glu Val
 530 535 540
 65

	Asp	Gly	Arg	Val	Tyr	His	Val	Gly	Leu	Gly	Asp	Ala	Ala	Gln	Pro	Arg	
	545						550				555					560	
5	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	
10				580					585					590			
	Ala	Ala	Asn	Pro	Ile	Lys	Val	Thr	His	Arg	Cys	Tyr	Ile	Leu	Glu	Asn	
			595					600					605				
15	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	
20	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	
25				660					665					670			
	Gly	Arg	Leu	Lys	His	Arg	Gly	Ser	Ser	Leu	His	Pro	Phe	Arg	Lys	Gly	
		675						680					685				
30	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	
35	705					710					715					720	
	Ser	Met	Pro	Gly	Leu	Thr	Cys	Phe	Thr	His	Asp	Asn	Gln	His	Trp	Gln	
				725						730					735		
40	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				
45	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	
50	785					790					795					800	
	Asp	Val	Leu	Asn	Gln	Leu	His	Val	Gln	Leu	Met	Glu	Leu	Arg	Ser	Cys	
				805						810					815		
55	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	
60		835						840					845				
	Lys	Trp	Pro	Glu	Met	Lys	Arg	Pro	Ser	Ser	Lys	Ser	Leu	Gly	Gln	Le	
		850															

<210> 32
 <211> 6
 <212> PRT
 <213> Homo sapiens
 <220>
 <221> VARIANTE
 <222> (1)..(1)
 <223> Leu O Val
 5
 <220>
 <221> característica_misc
 <222> (1)..(3)
 <223> Xaa puede ser cualquier aminoácido que ocurre naturalmente
 10
 <220>
 <221> VARIANTE
 <222> (2)..(2)
 <223> Cys O Ser
 15
 <220>
 <221> VARIANTE
 <222> (3)..(3)
 <223> cualquier aminoácido
 20
 <400> 32
 Xaa Xaa Xaa Pro Ser Arg
 1 5
 25
 <210> 33
 <211> 23
 <212> PRT
 <213> Artificial
 30
 <220>
 <223> secuencia derivada de la arilsulfatasa A humana
 35
 <220>
 <221> PÉPTIDO
 <222> (1)..(23)
 <223> sustrato sintético para la formación de FGly; secuencia primaria proveniente de la arilsulfatasa A
 40
 <400> 33
 Met Thr Asp Phe Tyr Val Pro Val Ser Leu Cys Thr Pro Ser Arg Ala
 1 5 10 15
 45
 Ala Leu Leu Thr Gly Arg Ser
 20
 50
 <210> 34
 <211> 16
 <212> PRT
 <213> Artificial
 <220>
 55
 <223> una variante del péptido ASA65-80, en el cual residen la Cys69, la Pro71 y la Arg73, críticas para la formación de FGly, y estuvieron mezcladas
 60
 <220>
 <221> CARACTERÍSTICA_MISC
 <222> (1)..(16)
 <223> oligopéptido mezclado

<400> 34

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

<210> 35
<211> 16
10 <212> PRT
<213> Artificial

<220>
15 <223> una variante del péptido ASA65-80, en el cual la Cys69 fue reemplazada por serina
<220>
<221> CARACTERÍSTICA _MISC
<222> (1)..(16)
<223> oligopéptido Ser69
20 <400> 35

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

25 <210> 36
<211> 19
<212> ADN
<213> Artificial

30 <220>
<223> cebador de PCR específico para FGEs humanas

<220>
35 <221> característica _misc
<222> (1)..(19)
<223> cebador 1199nc de PCR específico para FGEs humanas

<400> 36
40 ccaatgtagg tcagacacg 19

<210> 37
<211> 16
<212> ADN
45 <213> Artificial

<220>
<223> cebador de PCR específico para FGEs humanas

<220>
50 <221> característica _misc
<222> (1)..(16)
<223> cebador 1c de PCR específico para FGEs humanas

<400> 37
55 acatggcccg cgggac 16

<210> 38
<211> 19
<212> ADN
60 <213> Artificial

<220>
<223> cebador de PCR específico para FGEs humanas

65 <220>
<221> característica _misc

ES 2 566 641 T3

<222> (1)..(19)
 <223> cebador 1182c de PCR específico para FGEs humanas

5 <400> 38
 cgactgtccc ttggactgg 19

10 <210> 39
 <211> 24
 <212> ADN
 <213> Artificial

15 <220>
 <223> cebador de PCR específico para FGEs humanas

20 <220>
 <221> característica _misc
 <222> (1)..(24)
 <223> cebador que contiene EcoRI de PCR específico para FGEs humanas de 5'

25 <400> 39
 ggaattcggg acaacatggc tgcg 24

30 <210> 40
 <211> 54
 <212> ADN
 <213> Artificial

35 <220>
 <223> cebador específico de HA

40 <220>
 <221> característica _misc
 <222> (1)..(54)
 <223> cebador específico de HA

45 <400> 40
 cccaagctta tgcgtagtca ggcacatcat acgगतatgc catggtgggc aggc 54

50 <210> 41
 <211> 57
 <212> ADN
 <213> Artificial

55 <220>
 <223> cebador específico de c-myc

60 <220>
 <221> característica _misc
 <222> (1)..(57)
 <223> cebador específico de c-myc

65 <400> 41
 cccaagctta caggtcttct tcagaaatca gctttgttc gtccatggtg ggcaggc 57

70 <210> 42
 <211> 54
 <212> ADN
 <213> Artificial

75 <220>
 <223> cebador específico de RGS-His6

80 <220>
 <221> característica _misc
 <222> (1)..(54)
 <223> cebador específico de RGS-His6

ES 2 566 641 T3

<400> 42
 cccaagctta gtgatggtga tggatgatgcg atcctctgtc catggtgggc aggc 54

5 <210> 43
 <211> 15
 <212> PRT
 <213> Artificial

10 <220>
 <223> oligopéptido tríptico proveniente de la preparación de FGEs humanas

15 <220>
 <221> CARACTERÍSTICA _MISC
 <222> (1)..(15)
 <223> oligopéptido tríptico proveniente de la preparación de FGEs humanas

20 <400> 43

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

25 <210> 44
 <211> 19
 <212> PRT
 <213> Artificial

30 <220>
 <223> oligopéptido tríptico proveniente de la preparación de FGEs humanas

35 <220>
 <221> CARACTERÍSTICA _MISC
 <222> (1)..(19)
 <223> oligopéptido tríptico proveniente de la preparación de FGEs humanas
 <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 .

45 <210> 45
 <211> 906
 <212> ADN

50

55

60

65

<213> Homo sapiens

<400> 45

```

atggcccggc atgggttacc gctgctgccc ctgctgtcgc tctgggtcgg cgcgtggctc   60
aagctaggaa atggacaggc tactagcatg gtccaaactg aggggtggag attcctgatg   120
ggaacaaatt ctccagacag cagagatggt gaagggcctg tgcgggagcc gacagtgaaa   180
ccctttgcc a cgacatatt tctgtcacc aacaaagatt tcagggattt tgcagggag   240
aaaaagtac ggacagaagc tgagatgttt gcatggaget ttgtctttga ggaactttgtc   300
ctgatgagc tgagaacaa agccaccag ccaatgaagt ctgtactctg gtggcttcca   360
gtggaaaagg ctttttggag gcagcctgca ggtcctggct ctggcaccg agagagactg   420
gagcaccag tgttacacgt gagctggaat gacgccctg cctactgtgc ttggcgggga   480
aaacgaetgc ccacggaggc agagtggag ttgcgcgcc gagggggctt gaagggtcaa   540
gtttaccat gggggaaatg gttccagcca aaccgcacca acctgtggca gggaaagtgc   600
cccaaggag acaaaagctg gcatggcttc catggagtct cccagtgaa tgccttcccc   660
gccagaaca actacgggct ctatgacctc ctggggaaag tgtgggagtg gacagcatca   720
cgtaccagg ctgctgagca ggacatgcgc gtctccggg gggcactctg gatcgacaca   780
gtgatggct ctgccaatca cggggcccg gtcaccacca ggatgggcaa cactccagat   840
tcagcctcag acaacctcgg ttccgctgt gctgcagaag 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> AND

<213> Mus musculus

<400> 47

5
10
15
20
25
30
35
40
45
50
55
60
65

atgcgcctctg agttctgggt cccacagcatg ggttccttgc tcectccgggt gttgctgctg 60
aggtctctgt cctgccccag gtttcagcta ggacatgcc aggatcctgc catggtgcat 120
ctgccagggtg gccgggtttct gatggggaca gacgtccag atggcagaga cgggtgaaggg 180
cctgccccggg aagtgcacagt aaaacccttt gccatcgaca tatttccagt caccataaaa 240
gaattcaggg agttttgtcag ggagaagaag taccagactg aagccgaggc attcgggtgg 300
agcttcgtct ttgaggattt tgtctccct gagctcagaa agcaagaaaa tctgatgcog 360
gctgttcact ggtggcagcc agtgccaaag gcattttgga ggcagcctgc aggtcccggc 420
tctggcatcc gagagaaact ggagcttccc gtggtacacg tgagctggaa cgacgtggt 480
gcttactgcg catggcgggg gagacgttg cccacagaag aggagtggga gtttgcagcc 540
cgagggggct tgaagggtca ggtttatcca tgggggaacc ggttcagcc aaaccgcacc 600
aacttatggc agggaaagtt ccccaagggt gacaaagctg aagatgggtt tcatggactg 660
tcaccagtga acgtttccc cccacagAAC aactacggac tgtatgaact catgggcaat 720
gtgtgggagt ggacagcgtc cacataccaa cctgtggcc aggacatgcg tgtctccgg 780
ggggcatcat ggatcgacac cgcagacggo tctgctaate acagggctcg ggtcaccacc 840
aggatgggaa acactccaga ctcagctca gacaacctgg gcttcogctg cgcctccagt 900
gcaggccgac cgaaggagga cctgtga 927

<210> 48

<211> 308

<212> PRT

<213> Mus musculus

<400> 48

40
45
50
55
60
65

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

5 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

10 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

15 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

20 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

25 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

30 Arg Val Leu Arg Gly Ala Ser Trp Ile Asp Thr Ala Asp Gly Ser Ala
260 265 270

35 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

40 Lys Glu Asp Leu
305

<210> 49
<211> 855

45 <212> ADN
<213> Mus musculus

50 <400> 49

atgggcccca ttcctgctgg agtattcaca atgggcactg atgatactca gatcagggcag 60
gatggagaag cccctgccag gagagtcact gttgatggct ttacatgga cgcctatgaa 120
gtcagcaatg cggattttga gaagtttgtg aactcgactg gctatttgac agaggetgag 180
aagtttggag actctttcgt ctttgaaggc atgttgagcg agcaagtga aacgcatac 240
caccagggcag ttgcagctgc tccatgggtg ttgcctgtca agggagctaa ttggagacac 300
ccagaggggtc eggactccag tattctgcac aggtcaaatc atccggttct ccattgttcc 360
tggaacgatg ctgttgccca ctgcacatgg ggggcaaga ggttgccctac tgaggcagag 420
tggaataca gctgtagagg aggcctgcag aacaggettt tccctgggg caacaaactg 480

5 cagcccaag gacagcatta tgccaacatc tggcagggca agtttctgt gagcaacact 540
 ggcgaggatg gcttccaagg aactgcccc gttgatgect ttctcccaa tggetatggc 600
 ttatacaaca tagtggggaa tgtgtgggag tggacctcag actgggtggac tgttcaccat 660
 tctgttgagg aaacgttcaa cccaaagggc cccactctg ggaagaccg agtgaagaag 720
 10 ggtggatcct acatgtgcca taagtctat tgetataggt accgetgtgc agctcgaagc 780
 cagaacacac cagatagctc tgcattccaa ctgggattee gatgtgcagc cgaccacctg 840
 cccaccgcag actga 855

15 <210> 50
 <211> 284
 <212> PRT
 <213> Mus musculus
 20 <400> 50

25 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
 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
 35 Ser Phe Val Phe Glu Gly Met Leu Ser Glu Gln Val Lys Thr His Ile
 65 70 75 80
 40 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
 45 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
 50 Cys Arg Gly Gly Leu Gln Asn Arg Leu Phe Pro Trp Gly Asn Lys Leu
 145 150 155 160
 55 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
 60 Ala Phe Pro Pro Asn Gly Tyr Gly Leu Tyr Asn Ile Val Gly Asn Val
 195 200 205
 65 Trp Glu Trp Thr Ser Asp Trp Trp Thr Val His His Ser Val Glu Glu
 210 215 220

5 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

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

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

<210> 51

<211> 1011

<212> ADN

<213> *Drosophila melanogaster*

<400> 51

25 atgacaacaa ttatatattagt cctcttttatt tggatagttt tattcaatga cgtatccage 60

gactgtgget gccaaaagct cgaccggaag gccccggata tgcctgccat ttcgggacaa 120

30 gtgtgccage aacgagcaca ggggtgcacac agccactacc gggattacta tggcgaaactg 180

gagccaaata ttgcggacat gtcactgctt cggggaggca cggtttacat gggtaactgac 240

aaaccgcact ttcgggcoga ccgcgagget ccggaacggc aggtgaagct gaatgacttc 300

35 tacatcgaca agtatgaggt ttccaacgaa gccctttgoga agttttgtct gcacactaac 360

tacaaccagg aggtcgagcg atatggcgac agttttctgt ttaagagcct ttbgagccca 420

ttggagcaga agaacctaga ggaacttcga gtggcgagcg ctgtctgggt gtacaaagtg 480

40 gccggcgtga actggcgaca tccaaatggc gtggacagcg atatagacca cttaggcoga 540

cacccggtag tgcaagtatc gtggcgcgac gctgtggagt actgtaagtg gcccgccaag 600

cggttgccca gcgagggcga gtggggaggc gcttgcaagg gcggcaagga gcgcaaaactg 660

45 tttccctggg gcaacaagct gatgccaagg aatgaacatt ggctgaacat ctggcagggg 720

gactttcccg atggcaacct ggctgaagat gggtttgagt acaccagccc cgtggatgac 780

ttecgacaga atatttacga cctgcacaac atgggtggca acgtctggga gtggacggca 840

gatctgtggg acgtaaatga cgttagcgat aatccaaatc gggtaagaa gggcggttct 900

50 tatctgtgtc acaagtecta ctgctacagg tacaggtgag cggcacgctc gcagacaca 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 Asp Met Ser Leu Leu Pro Gly Gly Thr Val Tyr Met Gly Thr Asp
65 70 75 80

Lys Pro His Phe Pro Ala Asp Arg Glu Ala Pro Glu Arg Gln Val Lys
85 90 95

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

Ala Lys Phe Val Leu His Thr Asn Tyr Thr Thr Glu Ala Glu Arg Tyr
115 120 125

Gly Asp Ser Phe Leu Phe Lys Ser Leu Leu Ser Pro Leu Glu Gln Lys
130 135 140

Asn Leu Glu Asp Phe Arg Val Ala Ser Ala Val Trp Trp Tyr Lys Val
145 150 155 160

Ala Gly Val Asn Trp Arg His Pro Asn Gly Val Asp Ser Asp Ile Asp
165 170 175

His Leu Gly Arg His Pro Val Val His Val Ser Trp Arg Asp Ala Val
180 185 190

Glu Tyr Cys Lys Trp Ala Gly Lys Arg Leu Pro Ser Glu Ala Glu Trp
195 200 205

Glu Ala Ala Cys Arg Gly Gly Lys Glu Arg Lys Leu Phe Pro Trp Gly
210 215 220

Asn Lys Leu Met Pro Arg Asn Glu His Trp Leu Asn Ile Trp Gln Gly
225 230 235 240

Asp Phe Pro Asp Gly Asn Leu Ala Glu Asp Gly Phe Glu Tyr Thr Ser
245 250 255

Pro Val Asp Ala Phe Arg Gln Asn Ile Tyr Asp Leu His Asn Met Val
260 265 270

Gly Asn Val Trp Glu Trp Thr Ala Asp Leu Trp Asp Val Asn Asp Val
275 280 285

Ser Asp Asn Pro Asn Arg Val Lys Lys Gly Gly Ser Tyr Leu Cys His
290 295 300

Lys Ser Tyr Cys Tyr Arg Tyr Arg Cys Ala Ala Arg Ser Gln Asn Thr
305 310 315 320

Glu Asp Ser Ser Ala Gly Asn Leu Gly Phe Arg Cys Ala Lys Asn Ala
325 330 335

<210> 53
<211> 870

<212> ADN

5

<213> Anopheles gambiae

<400> 53

10

ccggagagct tgctcgatct ggtggaacat tccaagcggt tcgaagacat ggccttacc 60

15

ccaggagggt aatatgtaat cggcacaat gaacctatct tcgtcaagga tcgcgaatca 120
 cgggcccgcc cgcgcacgat cgcgcacttt tacctcgacc agtaagaagt ctccaacgca 180
 cagttcaagg cattcgatga ccagacgggc tacgtcacgg aggcgggaaa gtttgccgac 240
 agcttctgtct tccagcagct gctcagcgaa cgggtgcgcc agcagtaaga agatttcggc 300
 gtggcggcgg cgcctcgggt gtacaaggta cgtggagcct cctggcagca tccggaaggt 360
 gatgtgtcac gtgatataag cgcgcgattg gaccatccgg tgggtgcacgt gtcctgggac 420
 gatgcgggtcg cgtactgcgc ctggaaaggg aagcgcctgc cgcgcggaagc ggaatgggaa 480
 gcggcctgcc ggggcgggtcg caagcagaag ctgttccctt ggggtaacaa gctgatgcgc 540
 aaggagcagc acatgatgaa catatggcag ggcgagttcc cgcgcagcaa tctgaaggag 600
 gatggctaac agaccacctg cccggtgacg tcttccgcc agaacccgtt cgcgctgtac 660
 aacatcgttg gcaacgtgtg ggcgtggacg gcggatcttt gggacgcgaa ggcgtcggcc 720
 atcagcgcca agcggggcag cgcacacacg aatcgggtga aaaggggtgg ctcataacctg 780
 tgtaacgaat cgtactgcta tcgctatcgc tctcgggtgc gatcgacgaa caccgaggac 840
 agtccggcgg gcaatctggg ctcccggtgc 870

30

<210> 54

<211> 290

<212> PRT

<213> Anopheles gambiae

35

<400> 54

40

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

65

5
10
15
20
25

	165	170	175
	Lys Leu Met Pro Lys Glu Gln His Met Met Asn Ile Trp Gln Gly Glu		
	180	185	190
	Phe Pro Asp Ser Asn Leu Lys Glu Asp Gly Tyr Glu Thr Thr Cys Pro		
	195	200	205
	Val Thr Ser Phe Arg Gln Asn Pro Phe Glu Leu Tyr Asn Ile Val Gly		
	210	215	220
	Asn Val Trp Glu Trp Thr Ala Asp Leu Trp Asp Ala Lys Asp Ala Ala		
	225	230	235
	Ile Glu Arg Lys Pro Gly Ser Asp Pro Pro Asn Arg Val Lys Lys Gly		
	245	250	255
	Gly Ser Tyr Leu Cys His Glu Ser Tyr Cys Tyr Arg Tyr Arg Cys Ala		
	260	265	270
	Ala Arg Ser Gln Asn Thr Glu Asp Ser Ser Ala Gly Asn Leu Gly Phe		
	275	280	285
	Arg Cys		
	290		

<210> 55
<211> 945
<212> ADN
<213> Streptomyces coelicolor
<400> 55

35
40
45
50
55
60

gtggccgtgg	ccgccccgtc	ccccgcggcc	gcgcgggagc	cggggcccgc	cgcccgctcc	60
cgctcgaccc	gcggacaggt	gcgcctgcgc	ggcggtgagt	tcgcgatggg	ggacgccttc	120
ggggagggat	atccggccga	cggcgagaca	cccgctgcaca	cggctgcgct	ggcgcccttc	180
cacatcgacg	agaccgcgct	caccacgcgc	cggctcgcgc	ccttcgtcaa	ggcgaccggc	240
catgtgaccg	acgcgcgaac	cttcggctcc	tgggcgtct	tccacctggt	cgtcgccgcc	300
cgggacgcgc	acgtcctcgg	cagcgcgcgc	ggcgccccc	ggtggatcaa	cgtcgggggc	360
gcccaactgg	gcgcgcccca	ggcgccccgc	tccgacatca	cgggcgggcc	gaaccatccg	420
gtcgctcacg	tctcctggaa	cgatgccacc	gcctacgcgc	ggtgggccc	caagcgctg	480
cccaccgagg	cgaatggga	gtaacgcgc	cgggggggac	tgccgggcgc	ccgctacgcc	540
tggggcgacg	agctgacccc	ggcgggccgc	tgccgctgca	acatctggca	gggcgccttc	600
ccgcacgtca	acacggccga	ggaaggccac	ctgagcaccg	caccggtcaa	gtcctacccg	660
cccacgggcc	acggcctgtg	gaacaccgcg	ggcaacgtgt	gggaatgggt	ctccgactgg	720
ttctgcacca	cctactacgc	cgaatcacc	acgctcgacc	cgcacggccc	cgggacgggg	780
ggggacgggg	tgctgcggcg	cggctcctac	ctgtgccacg	actcctactg	caaccgctac	840
cgggtgcgcg	ccgctcctc	caacacccc	gactcctcgt	ccggcaacct	cggatccgc	900
tgcgccaaag	acggggacct	cacgtccgga	tcggccgctg	agtgaa		945

<210> 56
<211> 314
<212> PRT
<213> Streptomyces coelicolor

<400> 56

5

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

10

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

15

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

20

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

25

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

30

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

35

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

40

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

45

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

50

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

55

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

60

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

65

<210> 57
<211> 1005
<212> ADN

<213> *Corynebacterium efficiens*

<400> 57

```

gtgggttcgcc atcgactggg ccacggggcc tgcacactga ggattacgto catgagtaac      60
tgetgctccc cgtaagcgc acaatggcgt accactaccc gggatttato agatcctgtc      120
aateccacca ctccatgcaa ccgggaacaa tcccgcatg ctgtgacact gccgggtgga      180
gctttccaca tggcgatca tcacggggag gsgtacccgg cggacgggga ggggccagta      240
catgagggttc acctcgccc cttcggcattaatgtcacca cggtcacgaa tgcgaggttc      300
ggacgattta ttgaagccac aggggtatacg acgacagcgg aacgctacgg tgtctcggtc      360
gtattctacg cagcgttcca agggcaacgc gctgacatlc ttgcgcaggt tccggcggtg      420
ccctgggtggc tggcggtcaa ggggtgcgaac tggcagcgtc ccaacggccc cggatccacc      480
ctggacgggc ttgaggacca ccccgtcgtt cagcttctct gggatgatgc cgttgccctac      540
tgcacctggg ctggcggtcg tctgccacc gaagccgagt gggaatacgc cggccggggt      600
ggactgcggg ggcacagata tgccctgggg gataacctcg cctagacggg gaggtggaac      660
tgcaatatct ggcagggggg ctcccccatg gagaacacgc cggcggtatg ttacctcacc      720
actgcacggg tgaagacctc cagccccaat ggatacggto tgtggcagat ggcaggggaat      780
gtatgggaat ggtgccagga ctggtttgat gggagtaact actcccgctc ttccctccatc      840
aaccgcgggg gaccggatac cgggtgcgcgc cgggtgatgc gggagggtc gtatctctgc      900
catgattcct actgcaacag ataccgggtg gccgcgcgca attcgaaacac cccggattcc      960
acctcgggga ataccgggtt cgggtgcgtt ttcgatagtc cttaga      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

```

[illegible]

◀210▶ 59

<211> 1017

<212> ADN

<213> Novosphingobium aromaticivorans

«400» 59

atggcgcaac cattccgac gacggcgcc agtcgtaca gtattgaacg ccattctcga	60
cccaattgca ggagcaccgtc gcgaatggtc gaacgcccg gcattgcgcct gatcgaaaggc	120
ggcattttca ccatgggctc ggaagccttc taccgcggag aagcgccgct tcgcgggggtg	180
aaggtagaca gcttctggat cgatgaagcg ccggtgacga acgcacagtt cgcgcgcatte	240
gtggaggcca cgggatacgt cactgtggcc gagatcgagc cggatcccaa ggaactacccc	300
ggcatgctcc cgggcatgga ccgcgcggga tcgctggtgt tccagaaaac agcagggccg	360
gtcgacatgg cggatgcgtc caactcgtgg cactttacct ttggcgccctg ctggaagcat	420

5 ccacttggac cgggcagttc catcgatggg atcgaggacc atcccgtcgt tcacgtcgcc 480
 tatgccgatg cggaggccta tgccaaatgg gggggcaagg atctgcccac cgaagccgag 540
 ttccaatatg ctgcgcgogg cgggttggac ggttcgaat ttctctgggg agacgaactc 600
 gcacctgaag gccggatgat ggccsactac tggcaaggcc tgtttccctt cgccaaccag 660
 10 tgccctgatg gctgggaacg gacatcgccc gtccgcaact tcccgcccaa cggtatgggt 720
 ctttaacaca tgatcgggaa cacgtgggag tggacctgag attggtgggc cgacaagccg 780
 ctgactccgc aaaggaaate ggcctgctgc ggcctcaga atccgcggcg cggcaagctc 840
 15 aaggacagct tcgaacctgc gcaacctgca atgcgcctcg gccggaaggt cataaagggc 900
 ggttcgcacc tgtgtgcggc caattactgc cagcgctatc gccctgcagc acgccatcct 960
 gaaatgggtg ataccgcgac gacgcacatc ggttcagggt gtgtggtgcg gccctga 1017

20 <210> 60
 <211> 338
 <212> PRT
 <213> *Novosphingobium aramaiscivorans*
 25 <400> 60

30 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
 35 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
 40 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
 45 Lys Asp Tyr Pro Gly Met Leu Pro Gly Met Asp Arg Ala Gly Ser Leu
 100 105 110
 50 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
 55 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
 60 Thr Glu Ala Glu Phe Glu Tyr Ala Ala Arg Gly Gly Leu Asp Gly Ser
 180 185 190
 65 Glu Phe Ser Trp Gly Asp Glu Leu Ala Pro Glu Gly Arg Met Met Ala

5		195	200	205	
		Asn Tyr Trp Gln Gly Leu Phe Pro Phe Ala Asn Gln Cys Leu Asp Gly			
		210	215	220	
10		Trp Glu Arg Thr Ser Pro Val Arg Asn Phe Pro Pro Asn Gly Tyr Gly			
		225	230	235	240
		Leu Tyr Asp Met Ile Gly Asn Thr Trp Glu Trp Thr Cys Asp Trp Trp			
			245	250	255
15		Ala Asp Lys Pro Leu Thr Pro Gln Arg Lys Ser Ala Cys Cys Ala Ile			
			260	265	270
		Ser Asn Pro Arg Gly Gly Lys Leu Lys Asp Ser Phe Asp Pro Ser Gln			
			275	280	285
20		Pro Ala Met Arg Ile Gly Arg Lys Val Ile Lys Gly Gly Ser His Leu			
			290	295	300
		Cys Ala Ala Asn Tyr Cys Gln Arg Tyr Arg Pro Ala Ala Arg His Pro			
			305	310	315
25		Glu Met Val Asp Thr Ala Thr Thr His Ile Gly Phe Arg Cys Val Val			
			325	330	335
		Arg Pro			
30		<210> 61			
		<211> 1119			
		<212> ADN			
		<213> Mesorhizobium loti			
35		<400> 61			
40		atgggcccac gaggtcgagg tcaaaaaccg catgaaaggg gaagcggtca tgttcgacat	60		
		tgccgggaag ttctagccga tagcggtgg gccgctgatg gagatgagca cgcggtgca	120		
		tttcgggata ttctgatgaa cgcctctgcc gaagtcttgc agcgcgctgc agccgaacgg	180		
45		tcttaccctg gaatggcttg gatcccgcc ggtaccttcc tgatgggtgc agacaaccac	240		
		tatccggagg aggcaccggc ccaccgggtc aggtcgagc gcttctggat ggacaaatcc	300		
		accgtctcca accgggactt cgaacgcttc gttcgggcga caggacatgt cactcttgc	360		
		gagaaaccgg ccaatccgca cgaactatcc ggtgccttac ccgatctgtt ggtcccgctc	420		
50		togatgatgt tcaggaagcc ggcggccct gtgcacctg gcaatcacta caatcggtgg	480		
		gtctatgtcc gggcgcccaa ctggcgccat ccacgggggc cggcaagtac aatcaagaag	540		
		gttgcagatc atccggctgt gcatgtggcc taagaggatg tcttggccta tgccaaactg	600		
55		gcaggcaagg aacttccac ccaggccgag tgggaattcg ccggcgaggg cggcctcgat	660		
		gcgcgcgaat acgtctgggg caacgagctt acgcggccg ggaagcacat ggccaacatc	720		
		tggcaaggag aatttcccta ccggaatact gtgcagagc gttacgaata taaggcccca	780		
60		gtaggtctgt tcccgcccaa cgaactacgt ctctacgaca tggccggcaa tgtctggcaa	840		
		tggacgacg actggtacca ggaccacaag gcgategaca gccgctgtg caccgctgtc	900		
		aatcccggtg gggccatcg cgaagcgagc tatgacacc ggtacctga cgttaagatc	960		
65		cctcgcaagg tcaccaaggg tggctcccat ctgtgcgcgc cgaactactg tggcgctac	1020		

5 eggeccggcg gggaatgga gcaaccggc gacactgca tctcccatct eggettgcg 1080
 tgcacgtgc gaaggaaat ggaattgaac ggcagtaa 1119

<210> 62
 <211> 372
 <212> PRT
 <213> Mesorhizobium loti

<400> 62

15 Met Gly Pro Arg Gly Arg Gly Gln Lys Pro His Glu Arg Arg Arg Gly
 1 5 10 15

20 His Val Arg His Cys Arg Glu Val Leu Ala Asp Ser Gly Trp Ala Ala
 20 25 30

25 Asp Gly Asp Glu His Ala Val Ser Phe Arg Asp Leu Ser Met Asn Ala
 35 40 45

30 Pro Ala Glu Val Phe Glu Arg Ala Ala Ala Glu Arg Ser Tyr Pro Gly
 50 55 60

35 Met Val Trp Ile Pro Gly Gly Thr Phe Leu Met Gly Ser Asp Asn His
 65 70 75 80

40 Tyr Pro Glu Glu Ala Pro Ala His Arg Val Arg Val Asp Gly Phe Trp
 85 90 95

45 Met Asp Lys Phe Thr Val Ser Asn Arg Asp Phe Glu Arg Phe Val Ala
 100 105 110

50 Ala Thr Gly His Val Thr Leu Ala Glu Lys Pro Ala Asn Pro Asp Asp
 115 120 125

55 Tyr Pro Gly Ala Leu Pro Asp Leu Leu Ala Pro Ser Ser Met Met Phe
 130 135 140

60 Arg Lys Pro Ala Gly Pro Val Asp Leu Gly Asn His Tyr Asn Trp Trp
 145 150 155 160

65 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

		260	265	270	
5		Asp Met Ala Gly Asn Val Trp Gln Trp Thr Thr Asp Trp Tyr Gln Asp 275 280 285			
		His Lys Ala Ile Asp Ser Pro Cys Cys Thr Ala Val Asn Pro Arg Gly 290 295 300			
10		Gly His Arg Glu Ala Ser Tyr Asp Thr Arg Leu Pro Asp Val Lys Ile 305 310 315 320			
		Pro Arg Lys Val Thr Lys Gly Gly Ser His Leu Cys Ala Pro Asn Tyr 325 330 335			
15		Cys Arg Arg Tyr Arg Pro Ala Ala Arg Met Ala Gln Pro Val Asp Thr 340 345 350			
		Ala Ile Ser His Leu Gly Phe Arg Cys Ile Val Arg Arg Lys Met Glu 355 360 365			
20		Leu Asn Ala Gln 370			
25					
	<210> 63				
	<211> 1251				
	<212> ADN				
	<213> Burkholderia fungorum				
30					
	<400> 63				
		atgaagagtg aaagagateg agagcccgca aagtcgtccc gctcgaaagg gtcggtcgca	60		
35		gcaacccaaa cgcgcgcggg tcgctgtcgc aaactaatgt tctgggggag cctgctctgt	120		
		atactgccg cctgtgtcgg cgcgcgggtc agttgggect tcacgcgcga tgcacccgct	180		
		caccgcgaaa tcgttttctg cgaaggcag catggtcctc tcggcatggc gtcgggtgcc	240		
40		ggcgcccgct tctctcatgg cagcgacgac aaacaggcgc aaccgaacga acgccccggc	300		
		cacaaggtea aggtgcacgg cttctggatg gaccgccatc acgtgaccaa cgcgcaatc	360		
		cgcgccttgc tcgaagcgac cggctacgtc accacggcgc agaagaaacc cgactgggag	420		
45		acccatgaag tcacgttgcc gcccggcag cgcgcgcgcg ccgagagcgc gatggtggcg	480		
		ggtgcgaatg tgttcgtcgg caacagccgt cccgtgcgc tagacgacta ttgcagtg	540		
		tggcgctatg tgcctggcgc taactggcgt catccagcgc ggctgagag caacatcctc	600		
50		ggtaaagatg atcaccocgt ggttcaagtg tctacgaag atgcgcagac ttatgcgaaa	660		
		tgggcgggca agcgtctgac gaccgaagcc gaatgggaat tcgcgcgcgc cggcgccctc	720		
		gaacaggcca cgtatgcgtg gggcgatcag ttctctccca acggcaaaac gatggccaac	780		
55		gtctggcagg gccagcagcc gcagttcttc cccgttgta acccgaaagc gggtaggcgc	840		
		ctgggtacaa gtccggtggg tactttcccg gccacggct acggccttct cgacatgacc	900		
		ggcaacgcct ggcagtggtg tgcgactgg tatcgccgcg atcagttcag gcgtgaggcg	960		
		gtaagcaaca gcgcgatcga caatccggtg ggcgcgagcg agtcgtggga cccgcagac	1020		
60		cagggcgctg ccgtcaagc gcccaagcgt gtcacaacgc cgggttcgtt cctctgcac	1080		
		gaaatctatt gcctgagcta cggccccagc gcgagacgcg gcaccgatcc ctacaacagc	1140		
		atgtgcctac tgggcttcgc gctggtgatg gacgaagaca cctgggaaga agccggtgct	1200		
65		cgcaggett cggcgaaagc tgcgggcgc cctggaaacc ctggcggtc 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
    
```

5		275	280	285	
		Phe Pro Ala Asn Gly Tyr Gly Leu Ser Asp Met Thr Gly Asn Ala Trp			
		290	295	300	
10		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	
15		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	
20		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
25		Arg Gln Ala Ser Ala Lys Ala Ala Gly Ala Pro Gly Thr Pro Gly Gly			
		405	410	415	
30	<210> 65				
	<211> 912				
	<212> ADN				
	<213> Sinorhizobium meliloti				
35	<400> 65				
		atggtctctggg ttcccgaggc gaccttcattg atgggggtcga acgaccatta ccgggaggaa	40		
		gcgcgcgtgc atccggtaac cgtcgacgga ttctggatcg atgtgacacc ggtaacgaac	120		
40		cgcagtttc tcgaattcgt aatcgagacg gggcatgtga ccttcgggga aagaaagccg	180		
		cgcgcgaag actatecggg cgtctcgcca tccaatctaa gggcgggttc gctcgtcttc	240		
		acacccccga agcgaccgct gcagggaaac gatatatcgc agtgggtggat attcaagctg	300		
45		ggtgccaaact ggccgcaccc gctcgggcgc aagagcagca tcggagcgat tctggatcat	360		
		cgggtcgtcc atgtcgttta cagcgacgca aaggcctatg ccgaatgggc cggcaaggac	420		
		ctcccgaccg agaccgagtg gaagctggcg gcccgcgcg gctcgatgg ggcctgaattt	480		
50		tcctggggcg gcagcttgc gccggggcga aatcacatgg ccaatacttg gcagggaaat	540		
		tttcgggtcg agaattctat ggaagatggt ttcccgcgaa catcgccggt cagattttac	600		
		cgcgcgaacg gctacggcct ctacgacatg atcggaatg tgtgggagtg gaccacggat	660		
55		tactggtecg tgcgcacccc ggaagcgccc gccagcctt gctgcattcc gagcaatccc	720		
		cgaatgcgc atgcgatgc gattatcgat ccggcgcgga gcgtgaaagt tccgcgcgg	780		
		gtgctcaagg gtggatcgca tctctggcgc ccgaactact gccggcggtta ccgcctggc	840		
		ggagggcagc ccaggaaat cgaacggacg accagccatg tcggtttccg atgtgtcagg	900		
60		cgggttcgat aa	912		
	<210> 66				
	<211> 303				
65	<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> ADN
<213> Microscilla sp.

<400> 67

atgaaataca ttttttttagt tttttttetta tgggcoettga cccgatgtac cggaaagtat 60
gaggacaaga gagtggaac tgatacttcc agaccaaag cogaagcgtc agatataaaa 120
gttcccgaaag gaatggetta tattcccgcg ggcaggtaca tgatggggagg taaatcagac 180
caggettata aggatgaata tccccecat aacgtgaagg tttcggtttt ttatatggac 240
cttacagaag tgaccaatgc ggagtttaag cggtttgtag acgaaacggg ctacgtgacc 300
attgctgaga aagatatga ctgggaagag ttaaagtctc aggtgccaca gggtaacctg 360
aagctctctg attctgtgct tcaggcaggt tcaactggtt tcaagcagac agatgaacco 420
gtttctctcc aggattatcc acagtgggtg gaatggacta tcggagccaa ctggcgaaat 480
ccggaggggc caggtagtag gattgaggat cgtatggatc atccggtggt acacgtttcc 540
tttgaagatg tccaagcgtc tgccgattgg gccggttaagc gcttgcttac tgaggcagaa 600
tgggaatggg cggccatggg aggcacaaat gacgtgaaat atccatgggg aaatgaatcg 660
gtcgaacaag catccgataa agcaacttt tggcagggga attttcacac tcaaaactat 720
gcctcgatg gattcgaaag caccgcccct gtaagctctc tccagcgaaa tgggtacggc 780
ctatatgata tggctggcaa tgbtgggaa tggtgccagg ataagtatga tgtcaatgct 840
tatgaaagct ataagcaaaa aggaactgac gaagacccca cgggttctga gcaactacaac 900
gacctaggg aacgttatcc tccaaagcat gtatcagag ggggttcttt cctatgcaat 960
gacagctact gtagtgggta tcgtgttcca cgtctatga gttccagtag agattcaggt 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

5 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

10 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

15 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

20 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

25 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

30 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

35 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

40 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> ADN

<213> Pseudomonas putida KT2440

<400> 69

atggtgcacg tgcggggcgg caggttcagc ttggttcaa gcgctttta cgacgaagaa 60
ggcccgccctc accccgccaa ggtgtccggc ttctggattg acgtgcctcc ggtcaccac 120
gcccagttcg cgcgttcgt caaggccacg gggatatgta cccatgcga gcgggtacc 180
cgtgtcgagg acgacctgc cctgccega gcgtgcga taccgggtgc gatggtgtt 240

5 catcaggggtg cggacgtggt cggccccggc tggcagttcg tgcocggggc caactgggga 300
 caccgcgaag ggcggggcag cagcctggcc gggtggaca accatccggg ggtgcagatc 360
 gccctggaag atgcccaggc ctatgccgcg tgggcaggcc ggcgaactgcc cagcgagggtg 420
 cagctggaat accccatggg cggcgggcctg accgatgctg acttcagctg gggtaaccac 480
 gagcagccca agggcaagct catggccaat acctggcagg gtcagttccc ctatcgcaat 540
 ggggcgaagg atggttttac cggtaacatg cccgtgggtt gcttcocggc caacggcttt 600
 ggcctgttcg atgcggggcg caatgtctgg gagctgactc gcacgggcta tcggccaggc 660
 catgagccac agcgcgagcg caagctcgac cctcaggcc cggccctgag tgacagcttc 720
 gaccgggcag aaccggggct gccggtggcg gtaatcaag gcggtcgca cctgtgttcg 780
 ggggaacgct gcatgcgcta ccgcccctcg gcacgcccgc cgcagccggt gttcatgacg 840
 acctgcacg tgggtttcag aacgattcgg caatga 876

<210> 70

<211> 291

<212> PRT

<213> Pseudomonas putida KT2440

<400> 70

30 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
 35 Ile Asp Val His Pro Val Thr Asn Ala Gln Phe Ala Arg Phe Val Lys
 35 40 45
 40 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
 45 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
 50 Asp Asn His Pro Val Val Gln Ile Ala Leu Glu Asp Ala Gln Ala Tyr
 115 120 125
 55 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
 60 Gln 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

5 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

10 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

15 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

20 Ile Arg Gln
290

<210> 71
<211> 780
<212> ADN
<213> *Ralstonia metallidurans*

<400> 71

atggtcgcgg ggggatggt gttcgtcgcc accaacagcc cggtgccgt ggcgaatac 60

tggcgtcgtt ggcgttcgt acctggcgcc gactggcgcc accgacccg ccggggcagt 120

35 tccatcgaag gcaaggacaa tcatcccgcc gtgcaggtct cgtatgaaga cgcgcaggcg 180

taegccaaat gggccggcaa gcgtctgcc accgagccg agtgggagtt tgcgcgccgt 240

ggcgccctgg agcaggccac ctacgcctgg ggtgacaagt tcgcgccgga tggccggcag 300

40 atggcgaatg tctggcaggg ccagcagggt cagccgttc cggtggtcag cggcaaggcg 360

ggcgcccgcc ctggcaccag tctgtcgcc acgttcgcc gcaatggcta tgggtcttat 420

gacatgaccg gcaacgcctg gcagtgggtg gccactggt atcgcgcgga ccagttccgc 480

45 cgcgaagcca cgggtggcgg agtgotgcag aatccgacc gcccgccga ttctgtgggac 540

ccgaccgaac ctggcggtgc ggtgtcgccg ccaagcggg tcacgcgcgg tggctcgttc 600

ctctgcaacg aggaattctg cctcagctac cgcgcgagt ccggcgccg taccgaccg 660

50 tacaccagca tgtcgccact aggettccgg ctctgatgg atgacgccg ttgggcagaa 720

gttcgcaagg agccagccgt ggcaatggcc gggggcgggc agcagaacct 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 80
 Gly Gly Leu Glu Gln Ala Thr Tyr Ala Trp Gly Asp Lys Phe Ala Pro
 85 90 95
 Asp Gly Arg Gln Met Ala Asn Val Trp Gln Gly Gln Gln Val Gln Pro
 100 105 110
 Phe Pro Val Val Ser Ala Lys Ala Gly Gly Ala Ala Gly Thr Ser Ala
 115 120 125
 Val Gly Thr Phe Pro Gly Asn Gly Tyr Gly Leu Tyr Asp Met Thr Gly
 130 135 140
 Asn Ala Trp Gln Trp Val Ala Asp Trp Tyr Arg Ala Asp Gln Phe Arg
 145 150 155 160
 Arg Glu Ala Thr Val Ala Ala Val Leu Gln Asn Pro Thr Gly Pro Ala
 165 170 175
 Asp Ser Trp Asp Pro Thr Glu Pro Gly Val Pro Val Ser Ala Pro Lys
 180 185 190
 Arg Val Thr Arg Gly Gly Ser Phe Leu Cys Asn Glu Asp Phe Cys Leu
 195 200 205
 Ser Tyr Arg Pro Ser Ala Arg Arg Gly Thr Asp Pro Tyr Thr Ser Met
 210 215 220
 Ser His Leu Gly Phe Arg Leu Val Met Asp Asp Ala Arg Trp Ala Glu
 225 230 235 240
 Val Arg Lys Gln Pro Ala Val Ala Met Ala Ala Gly Gly Gln Gln Asn
 245 250 255
 Val Gln Lys

<210> 73

<211> 876

<212> ADN

<213> *Prochlorococcus marinus*

<400> 73

5	gtgaccacat ctttgccagt agagatggta accatcccc cagggetcta tegagtggc	60
	tgtgatcgct gctatccgga tggttcagtt cgtgctato cggaggaaac aaccgcega	120
	gaagtgcagc ttgactcatt ccagatcgac gtagggccag tcaccaatgc ccagtccga	180
	gctttcgta gcgccagca gccttcaca gtctcgagc taaccaetga tccaaegtc	240
10	tatcccgatc tagcgccga ggaacgcat cctgaatcag ttgtcttca accgcctca	300
	gcaacgggtg atcgagcaa acccttgagc tgggtggacc tcatggctg ggetgattg	360
	cgctatccc aaggaccga aagcaegat gatggcctg atgatcccc tgtctgcat	420
15	gtcgctatg ccgacgccat cgcctatgc cattgggctg gcaagcgtct cccctctgt	480
	gaagagtggg aagtagccg ccgcgggggt cttgtcgat cccaatacgc ctgggggaat	540
	gaactcact ccaataacc ctggatggc aacatctggc aaggtcctt ccttgccac	600
20	aacgaggagc tagacggctg gttctggac tgcgccgtg gcagcttcc tgccaaaggc	660
	tatggactct tggatgttg cggcaatgt tgggaatgga ccaactctgt ttatccctg	720
	gcgtcaggcc accaggaac ggaactatc aaaggcggat cgtttctctg cgcagataat	780
25	tactgcgtac gttatcgac ctctgacta caaggccaga cagtagaac tgccaectgt	840
	cacatgggct ttgctgtgc aaaaggagg ccttga	876

<210> 74

<211> 291

<212> PRT

<213> *Prochlorococcus marinus*

<400> 74

5 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

10 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

15 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

20 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

25 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

30 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

35 Trp Thr Ser Pro Val Gly Ser Phe Pro Ala Asn Gly Tyr Gly Leu Leu
210 215 220

40 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

45 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

50 Gly Gly Pro
290

<210> 75
<211> 1017

<212> ADN
<213> Caulobacter crescentus CB15

5	ttgggaaaaa	tgacggcgct	tcccgctctg	atgctttctg	cgtgggcggg	ctggggccag	60
	cggcgcccc	aggtttgect	ggcggaectg	cgggttcacg	atccccagaa	cgcacgggog	120
	ggtatggttc	ggtgggggg	eggcgacttc	cagatggggg	ctgcgcgcct	gcgtccggag	180
10	gagggacgcg	cccagacggg	caeggtcccg	cgtttctgga	tccatccagac	agaggtccac	240
	aaagcgccct	tggcggggtt	cgtcgaggcc	aeggggttate	gaaacgttgg	cgagcgaccc	300
	ctcgaccceg	cggcgtaacg	ccacgtaccg	ggggcgacgc	ggcgccgggc	ctcgctcgte	360
15	ttcgtggggg	cgaagggggc	gaggtcggac	gacccctccc	aatggtggca	ggtgaccccc	420
	ggcgccgaet	ggcgccatcc	cgaaggctcc	ggctcgaaac	tccggggcag	ggacgcctgg	480
	cgggtggtgc	atategcgtg	ggaggacgcc	atggcctaag	ccgcctggtt	gggcctgac	540
20	ctgccacacg	aggccgaatg	ggagtaagcc	gcgcggggcg	ggctggttgg	caagcgctac	600
	acctggggcg	accaggtcca	ggatcctgca	aagccgcggc	ccaatacttg	gcaagggcgt	660
	ttcccgggcc	aggaccttgg	caatgaaggc	ttcaaggcca	agcccgccgc	ggtcggttgc	720
25	ttcccgcccc	acggtatatg	cctgcggcac	atggcgggca	atgtctggga	gtggaccgcg	780
	gactggttca	agccggggct	ggatccggte	agcgctctcg	aaacgggggg	gcgcgccgag	840
	gcccgcggcg	tggatcccca	ggacccgaa	acgcccaagc	acgtcgtgaa	gggcggttcg	900
	ttcctgtgcg	cgcacgacta	ctgcttccgc	tatcgacctg	cggcgcgaa	gcggggggcg	960
	cgggacagcg	ggcgatcgca	tgtcggtttc	cgcaccgctg	tccgcgcgca	gccttga	1017

210 76

211 338

<212> PRT

◀213▶ *Caulobacter crescentus* CB15

~~400~~ 76

40 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

5 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
 10 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
 15 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
 20 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
 25 Pro Val Val His Ile Ala Trp Glu Asp Ala Met Ala Tyr Ala Arg Trp
 165 170 175
 30 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
 35 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
 40 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
 45 Leu Glu Thr Gly Gly Pro Pro Glu Ala Arg Ala Leu Asp Pro Glu Asp
 275 280 285
 50 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
 55 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> ADN

<213> Mycobacterium tuberculosis H37Rv

<400> 77

```

gtgctgaccg agttgggtga cctgcccggc ggatcggttc gcatggggctc gacgcgcttc   60
tacccegaag aagcgccgat tcataccgtg accgtgcgcg cctttgcggt agagcgacac   120
cgggtgacca acgcgcaatt tgccgaatcc gtctccggga caggctatgt gacgggttga   180
gaacaacccc ttgaccccggt gctctaccca ggagtggacg cagcagacct gtgtcccggg   240
ggatgggtgt tttgtccgac ggcggggccg gtcgacctgc gtgactggcg gcaatgggtg   300
gaactgggtac ctggcgccgt ctggcgccat ccgtttggcc gggacagcga tctcgccgac   360
cgagccggcc acccggtcgt acaggtggcc tatccggacg ccgtggccta cgcacgatgg   420
gtgggtcgac gcctaccgac cgagggcgag tgggagtag cgcccggtgg cggaaaccag   480
gcaacctatg cgtggggcga ccaggagaag ccggggggca tgcctcatgc gaacacctgg   540
cagggcggtt tctctaccg caacgaagggt gcattgggct ggggtgggaa ctecccggtg   600
ggcaggttcc cggccaacgg gtttggcttg ctgacatga tcggaaacgt ttgggagtg   660
accaccaccg agttctatcc acaccatgcc atcgatccac cctcgacggc ctgctgcgca   720
cgggtcaagg tcgtacagc cgcggaccgg acgatcagcc agaccctcaa gggcggtcgc   780
cacctgtgcg cgcggagta ctgccaccgc taccgccggc cggcgcgctc gcgcagtcg   840
caggacaacc cgaccacca tatcggttcc cgggtgctgg ccgaccgggt gtcgggtag   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

```


5 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

10 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

15 Gly Trp Val Gly Thr Ser Pro Val Gly Arg Phe Pro Ala Asn Gly Phe
195 200 205

20 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

25 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

30 Pro Ala Ala Arg Ser Pro Gln Ser Gln Asp Thr Ala Thr Thr His Ile
275 280 285

35 Gly Phe Arg Cys Val Ala Asp Pro Val Ser Gly
290 295

<210> 79

<211> 7

<212> PRT

<213> Artificial

<220>

<223> dominio conservado en procariotas y procariotas

<220>

<221> DOMINIO

<222> (1)..(7)

<223> dominio conservado

<220>

<221> CARACTERÍSTICA_MISC

<222> (3)..(4)

<223> cualquier aminoácido

<220>

<221> CARACTERÍSTICA_MISC

<222> (6)..(6)

<223> cualquier aminoácido

<220>

<221> CARACTERÍSTICA_MISC

<222> (6)..(6)

<223> Gly o Ala

5

<400> 79

10

Arg Val Xaa Xaa Gly Xaa Ser
1 5

<210> 80

<211> 630

15

<212> ADN

<213> Oncorhynchus mykiss

<400> 80

20

tcagggtggt	getgccccct	ggtaggttgc	tgtagagga	gcagactgga	ggcaccctga	60
ggccccgac	tcagcatca	cagacaggt	ggaccacct	gtgtgcatg	tgcatggca	120
ggacgtgtg	gcctactgt	cctgggcta	caagagacta	cccacagagg	ctgagtggga	180
gtacgcctgc	agaggggggc	tacaggagag	actttaccgc	tgggggaaac	aactgaaacc	240
taaaggacag	cactaagcca	acctctggca	gggaaagtgc	cccacacaca	actcagaaga	300
ggacgggtac	actaaaacct	caccagtga	gtcatttctc	gcaaatggct	atggcctgta	360
caacatggta	gggaatgcac	gggagtggac	atctgactgg	tggactgtac	accacaccac	420
agatgaacag	cacaaccggg	caggtccacc	atcaggccac	gaccagtgta	agaaaggagg	480
ctcctacatg	tgcataagt	catactgta	caggtacagg	tgtagcagac	ggagtccaga	540
cacccctgac	agctctgct	ctaactagg	gttccgctgt	gtctccaggg	agcagccgta	600
acctttcaac	ctcgacctg	acatgggtag				630

35

<210> 81

<211> 655

<212> ADN

<213> Danio rerio

40

<220>

<221> característica _misc

<222> (590)..(590)

<223> n es a, c, g, o t

45

<220>

<221> característica _misc

<222> (626)..(626)

<223> n es a, c, g, o t

50

<400> 81

55

60

65

5 caaatgggttt tatttacata aaaaaatcct cttagtttga agtgtaagac agtgagatta 60
 gtgatgtttt aggttatgga tcaacatcag aggcgcagcg gaagcccaag ttgaggctg 120
 aactgtcagg tgtgttttga ctgcgagcgg cacacctgta tctgtagcag taagacttgt 180
 10 ggcacatgta ggatcctcct ttcttgactc tgtctgtccc tgattctggg ccttttgggt 240
 taactctgtc ttctgcagtg tgatgcacag tccaccagtc tgcggtccc tcccaagcat 300
 ttccacacat gtcatacagg ccaaaagccat tgggaggaaa agacatcac ggggatgtgt 360
 15 tggcatagcc gtctctgtca gtgttgtgat tagggaaatc tccctgcac aggttagcat 420
 agtgctgccc tcttggaatt aatttatctt cccatgggta cctcctgtcc tgtagtcttc 480
 ctctacagga caactccat tcaagctctg taggaagtct gcgtttggcc cattgcagc 540
 acgcccgtgc atcatccat gaacatgca ggcaggggtg attcattctc gtgtgtatgg 600
 20 ttgaatctgg tctttcttgg tctctcagc ctgcacctt cactgggtgac cacca 655

<210> 82

<211> 773

25 <212> ADN
 <213> *Oryzias latipes*

<220>

30 <221> característica _misc
 <222> (690)..(690)
 <223> n is a, c, g, o t

<220>

35 <221> característica _misc
 <222> (755)..(755)
 <223> n is a, c, g, o t

<400> 82

40 tctccttttt tccataaata acattagagt ccttacattc tgcctttaca tacattgtca 60
 gagacagtac aaaaaatctg cctttgtaaa attagagtta caaaatata ttttagattt 120
 gaattcttca gaattgtcgg tggcagcaa agaactcggat tgatctcatg acaagagcgt 180
 45 gaggcagaag ttcttggatc aaactgattt ggttctgtca tctttctgtt tcagcagcac 240
 agcgaatacc aagattggaa gcggagctgt ctggagtgtt ttggcttcca gcagccatc 300
 tgtacctgta acaataagac ttgtggcaca tgtacgagcc tctttcttct accttatctg 360
 50 tgcctgaagg aggaaccggt gggttgtgt gatgtctgt tgtgtgtgtg acgtccacc 420
 agtctgaggt ccactcccat gcgttccca ccatgtcata cagacaaaa gaattgcctg 480
 ggaaggacat cccgggggag gttttagtgt agcactctc tgcagagttg tgtgtgtgga 540
 attccccctg ccagaggttg gctaatgt gtcccttttg gtttagcttg tttccccagg 600
 55 ggttagagtc gtccctcagg cgcacctgc aggaacctc ccactctgcc tcagtgggaa 660
 gtctcttgtt gaccaggag cagtaagca aggcacatt cccagaaacc tgaacgagc 720
 atgatecacc ctgtctgtga tgttgagtc tgganctca ggtgtcttc agt 773

<210> 83

<211> 566

60 <212> ADN
 <213> *Xenopus laevis*

<220>

65

<221> característica _misc
<222> (6)..(6)
<223> n es a, c, g, o t

<220>
<221> característica _misc
<222> (47)..(47)
<223> n es a, c, g, o t

<220>
<221> característica _misc
<222> (81)..(81)
<223> n es a, c, g, o t

<400> 83

```

atatgnaact aaaggtaatg taattggaat gatggatttc acaaggnotg agagttoect      60
attgeteectg cttgtcgtgt nacaggtcac ggagcggcg ccacacagcg aatcccagg      120
ttggaggcgc agctgtcggg tgtattctga cttagagcg cacagcgata cctgtagcaa      180
taggaactcat ggcacatgta ggagcctect ttcttcaetc tatcatttcc cgtagaagggt      240
cctttcgggt tgtgaacctc atctgtctga tgatgagtgt cccaccaatc agatgtccac      300
tcccaagcat ttccaccat gttatataga ccataaccat tggtcggga agcagttaca      360
ggtgaagtct gcacataacc atctctcca gtgttttggg ttggaaaac cccctgccag      420
acattcgcac aatgttgtcc ctttgggtcc agcttgttcc cccatggaaa aatcctgttc      480
tcaagtcacc cgggcaggc gtattccac tcagcttcag ttggaaggcg tttacctgcc      540
caggtgcaga aagcagaagc atcatt                                         566

```

<210> 84

<211> 647

<212> ADN

<213> *Silurana tropicalis*

<400> 84

```

gcgcgttttt tttttttttt tttttttttt catcacaaaa ataattttat taataaaata      60
ggattttgtg ttcatcttta ttatgaagga caaggaaatg cattgaatt tttgttttca      120
caaggctctg ggagttcctt cctgtcagg tcatttttga gtggtcacgg agccgacgc      180
aegcagcgga atcccaggtt agaggccgag ctgtcaggtg tattctgaet togagcagea      240
cagcgatacc tgtagcagta ggactcatgg cacatgtatg agcctccttt ttccaccttg      300
tcttttcccg taaaaggacc ttctgggttg taagtctcat ctgtctgatg atgagtgtcc      360
caccaatcgg atgtccactc ccaagcattt cccaccatgt tatatagget ataaccattg      420
gctgggaaag cgtttacagg tgaagtctgc acatagccgt cctctccagt gttttgggtt      480
ggaaattccc cctgccagac attgcataa tgtctccct ttggttccag cttgttcccc      540
caccgaaaaa gcctgttctc aagtcccca cgggagccat attccactc agcttctgtc      600
ggaaggcgct taccgcaca ggtgcagaag gcagaagcat cgttcca                                         647

```

<210> 85

<211> 636

<212> ADN

<213> *Salmo salar*

<400> 85

```

atagacattt tttaatatt ttacaacaa atatattcca taaatatcca catgtcatgc      60
ggtaatcctg catttcattga agaacaetga catcaactggc tgtatgaaga. ggtgcacttg      120
atttgtttcg cctggcgggc aagataggca gagttagcac cctagaactag agccaatggc      180
gaatgggtaca aaaagggaaa agtcagaacta cccatgtcag ggtcaagggt aaaaggttac      240
ggetgtctct gggagacaca gcggaaacct aggttagagg cagagctgtc aggggtgttc      300
tgactcctgt ctgcacacct gtacctgtaa cagtatgact tatggccat gtaggagcct      360
cctttcttca ctgggtctgt gctgatggt ggacctgcgc ggttgtgcgc ttcattctgt      420
gtgtggtgta cagtcacaa gtcagatgtc cactcccatg caticcctac catgttgtac      480
agccatagc catttcagg aaatgaactc actggtaggg ttttggtgta ccgtctct      540
ctgagttgt gtgtgggaa ctttccctgc cagaggttgg cgtagtgtg tcttttaggt      600
tccagtttgt tccccacgg gtaagctgt tctgt      636

```

<210> 86

<211> 415

<212> ADN

<213> Sus scrofa

<400> 86

```

agtttctgt gaccaacacc ggagaggatg gcttcgagg aactgcgcct gttgatgcct      60
ttctcccaa tggttatggc ctttacaata tagtagggaa cgcctgggaa tggacctcag      120
actggtaggc cttcaccat gctgctgaag aaacaattaa cccatcaagt tcttctgtct      180
gcaccgaata acagagcgc cactacgtga tgaagcaga gaaaggeccc cttctggga      240
aagaccgggt gaagaaagg ggatctata tgtgccataa gtcctactgc tacaggtacc      300
gctgtgctgc tcgaagccag aacacgcgg acagctggc ttcantctg gggttccgct      360

gtgcagctga ccaccagccc accacaggt gagtcaggaa ggtcttccc gaate      415

```

<210> 87

<211> 595

<212> ADN

<213> Bos taurus

<400> 87

```

ccacgcgtcc gggggcaaca aactgcagcc gaagggccag cattatagcc aacatcttgg      60
caaggcgagt tctctgtgac caacacggg gaggaaggct tccaggggac cgcgcctgtt      120
gacgccttcc ctcccaatgg ttattggctt atacaantata gtaggggacg cctgggagtg      180
gacttcagac tggtagactg ttcaccattc tgetgaagaa acgattaaac caaaaggccc      240
cccttctggg aaagacggg tgaagaaagg tggatcctac atgtgcataa aatcctattg      300
ctacaggtat cgtgtgtctg ctogaagcca gaacacaccc gacagctctg ctogaatct      360
gggatccctg tgtgcagctg accacctgac caccacagge taagagccaa aaagagcctt      420
ccgaacccg agaagtctg tctactctgc acgaggcttc cctcagaagg ctgaacaaac      480
tgctgtgaag aattcccacc ccaagggtgg ttacatacct tgcccagtg ccaagggacc      540
tatggcaaga ccaattgtct gagctgatca gcagtgtgcg tttattgggg gatgg      595

```

<211> 1611

<212> ADN

<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	48
Met Leu Leu Leu Trp Val Ser Val Val Ala Ala Leu Ala Leu Ala Val	
1 5 10 15	
ctg gcc ccc gga gca ggg gag cag agg cgg aga gca gcc aaa gcg ccc	96
Leu Ala Pro Gly Ala Gly Glu Gln Arg Arg Arg Ala Ala Lys Ala Pro	
20 25 30	
aat gtg gtg ctg gtc gtg agc gac tcc ttc gat gga agg tta aca ttt	144
Asn Val Val Leu Val Val Ser Asp Ser Phe Asp Gly Arg Leu Thr Phe	
35 40 45	
cat cca gga agt cag gta gtg aaa ett cct ttt atc aac ttt atg aag	192
His Pro Gly Ser Gln Val Val Lys Leu Pro Phe Ile Asn Phe Met Lys	
50 55 60	
aca cgt ggg act tcc ttt ctg aat gcc tac aca aac tct cca att tgt	240
Thr Arg Gly Thr Ser Phe Leu Asn Ala Tyr Thr Asn Ser Pro Ile Cys	
65 70 75 80	
tgc cca tca cgc gca gca atg tgg agt ggc ctc ttc act cac tta aca	288
Cys Pro Ser Arg Ala Ala Met Trp Ser Gly Leu Phe Thr His Leu Thr	
85 90 95	
gaa tct tgg aat aat ttt aag ggt cta gat cca aat tat aca aca tgg	336
Glu Ser Trp Asn Asn Phe Lys Gly Leu Asp Pro Asn Tyr Thr Thr Trp	
100 105 110	
atg gat gtc atg gag agg cat gcc tac cga aca cag aaa ttt ggg aaa	384
Met Asp Val Met Glu Arg His Gly Tyr Arg Thr Gln Lys Phe Gly Lys	
115 120 125	
ctg gac tat act tca gga cat cac tcc att agt aat cgt gtg gaa gcg	432
Leu Asp Tyr Thr Ser Gly His His Ser Ile Ser Asn Arg Val Glu Ala	
130 135 140	

5

tgg aca aga gat gtt get ttc tta ctc aga caa gaa ggc agg ccc atg 480
Trp Thr Arg Asp Val Ala Phe Leu Leu Arg Gln Glu Gly Arg Pro Met
145 150 155 160

10

gtt aat ctt atc cgt aac agg act aaa gtc aga gtg atg gaa agg gat 528
Val Asn Leu Ile Arg Asn Arg Thr Lys Val Arg Val Met Glu Arg Asp
165 170 175

15

tgg cag aat aca gac aaa gca gta aac tgg tta aga aag gaa gca att 576
Trp Gln Asn Thr Asp Lys Ala Val Asn Trp Leu Arg Lys Glu Ala Ile
180 185 190

20

aat tac act gaa cca ttt gtt att tac ttg gga tta aat tta cca cac 624
Asn Tyr Thr Glu Pro Phe Val Ile Tyr Leu Gly Leu Asn Leu Pro His
195 200 205

25

cct tac cct tca cca tct tct gga gaa aat ttt gga tct tca aca ttt 672
Pro Tyr Pro Ser Pro Ser Ser Gly Glu Asn Phe Gly Ser Ser Thr Phe
210 215 220

30

cac aca tct ctt tat tgg ctt gaa aaa gtg tct cat gat gcc atc aaa 720
His Thr Ser Leu Tyr Trp Leu Glu Lys Val Ser His Asp Ala Ile Lys
225 230 235 240

35

atc cca aag tgg tca cct ttg tca gaa atg cac cct gta gat tat tac 768
Ile Pro Lys Trp Ser Pro Leu Ser Glu Met His Pro Val Asp Tyr Tyr
245 250 255

40

tct tct tat aca aaa aac tgc act gga aga ttt aca aaa aaa gaa att 816
Ser Ser Tyr Thr Lys Asn Cys Thr Gly Arg Phe Thr Lys Lys Glu Ile
260 265 270

45

aag aat att aga gca ttt tat tat gct atg tgt gct gag aca gat gcc 864
Lys Asn Ile Arg Ala Phe Tyr Tyr Ala Met Cys Ala Glu Thr Asp Ala
275 280 285

50

atg ctt ggt gaa att att ttg gcc ctt cat caa tta gat ctt ctt cag 912
Met Leu Gly Glu Ile Ile Leu Ala Leu His Gln Leu Asp Leu Leu Gln
290 295 300

55

aaa act att gtc ata tac tcc tca gac cat gga gag ctg gcc atg gaa 960
Lys Thr Ile Val Ile Tyr Ser Ser Asp His Gly Glu Leu Ala Met Glu
305 310 315 320

60

cat cga cag ttt tat aaa atg agc atg tac gag gct agt gca cat gtt 1008
His Arg Gln Phe Tyr Lys Met Ser Met Tyr Glu Ala Ser Ala His Val
325 330 335

65

cag ctt ttg atg atg gga cca gga att aaa gcc ggc cta caa gta tca 1056
Pro Leu Leu Met Met Gly Pro Gly Ile Lys Ala Gly Leu Gln Val Ser
340 345 350

70

aat gtg gtt tct ctt gtg gat att tac cct acc atg ctt gat att got 1104
Asn Val Val Ser Leu Val Asp Ile Tyr Pro Thr Met Leu Asp Ile Ala
355 360 365

75

gga att cct ctg cct cag aac ctg agt gga tac tot ttg ttg cag tta 1152
Gly Ile Pro Leu Pro Gln Asn Leu Ser Gly Tyr Ser Leu Leu Pro Leu
370 375 380

80

tca tca gaa aca ttt aag aat gaa cat aaa gtc aaa aac ctg cat cca 1200
Ser Ser Glu Thr Phe Lys Asn Glu His Lys Val Lys Asn Leu His Pro
385 390 395 400

85

ccc tgg att ctg agt gaa ttc cat gga tgt aat gtg aat gcc tcc acc 1248
Pro Trp Ile Leu Ser Glu Phe His Gly Cys Asn Val Asn Ala Ser Thr
405 410 415

90

tac atg ctt cga act aac cac tgg aaa tat ata gcc tat tct gat ggt 1296
Tyr Met Leu Arg Thr Asn His Trp Lys Tyr Ile Ala Tyr Ser Asp Gly
420 425 430

95

gca tca ata ttg cct caa ctc ttt gat ctt tcc tct gat cca gat gaa 1344
Ala Ser Ile Leu Pro Gln Leu Phe Asp Leu Ser Ser Asp Pro Asp Glu
435 440 445

100

tta aca aat gtt gct gta aaa ttt cca gaa att act tat tct ttg gat 1392
Leu Thr Asn Val Ala Val Lys Phe Pro Glu Ile Thr Tyr Ser Leu Asp

5

10

15

20

25

30

35

40

45

50

55

60

65

```

      450      455      460
cag aag ctt cat tcc att ata aac tac cct aaa gtt tet 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

```


5

Asn Tyr Thr Glu Pro Phe Val Ile Tyr Leu Gly Leu Asn Leu Pro His
195 200 205

10

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

15

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

20

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

25

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

30

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

35

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

45

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

50

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

55

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

60

65

5 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

10 <210> 90
<211> 1722
<212> ADN
<213> Homo sapiens

15 <220>
<221> CDS
<222> (1)..(1719)
<223> hSULF4

20 <400> 90

25 atg ggg gcg ctg gca gga ttc tgg atc ctc tgc ctc ctc act tat ggt 48
Met Gly Ala Leu Ala Gly Phe Trp Ile Leu Cys Leu Leu Thr Tyr Gly
1 5 10 15

tac ctg tcc tgg ggc cag gcc tta gaa gag gag gaa gaa ggg gcc tta 96
Tyr Leu Ser Trp Gly Gln Ala Leu Glu Glu Glu Glu Gly Ala Leu
20 25 30

cta gct caa gct gga gag aaa cta gag ccc agc aca act tcc acc tcc 144
Leu Ala Gln Ala Gly Glu Lys Leu Glu Pro Ser Thr Thr Ser Thr Ser
35 40 45

cag ccc cat ctc att ttc atc cta gcg gat gat cag gga ttt aga gat 192
Gln Pro His Leu Ile Phe Ile Leu Ala Asp Asp Gln Gly Phe Arg Asp
50 55 60

gtg ggt tac cac gga tct gag att aaa aca cct act ctt gac aag ctc 240
Val Gly Tyr His Gly Ser Glu Ile Lys Thr Pro Thr Leu Asp Lys Leu
65 70 75 80

gct gcc gaa gga gtt aaa ctg gag aac tac tat gtc cag cct att tgc 288
Ala Ala Glu Gly Val Lys Leu Glu Asn Tyr Tyr Val Gln Pro Ile Cys
85 90 95

aca cca tcc agg agt cag ttt att act gga aag tat cag ata cac acc 336
Thr Pro Ser Arg Ser Gln Phe Ile Thr Gly Lys Tyr Gln Ile His Thr
100 105 110

gga ctt caa cat tct atc ata aga cct acc cca ccc aac tgt tta cct 384
Gly Leu Gln His Ser Ile Ile Arg Pro Thr Gln Pro Asn Cys Leu Pro
115 120 125

ctg gac aat gcc acc cta cct cag aaa ctg aag gag gtt gga tat tca 432
Leu Asp Asn Ala Thr Leu Pro Gln Lys Leu Lys Glu Val Gly Tyr Ser
130 135 140

acg cat atg gtc gga aaa tgg cac ttg ggt ttt tac aga aaa gaa tgc 480
Thr His Met Val Gly Lys Trp His Leu Gly Phe Tyr Arg Lys Glu Cys
145 150 155 160

atg ccc acc aga aga gga ttt gat acc ttt ttt ggt tcc ctt ttg gga 528
Met Pro Thr Arg Arg Gly Phe Asp Thr Phe Phe Gly Ser Leu Leu Gly
165 170 175

agt ggg gat tac tat aca cac tac aaa tgt gac agt cct ggg atg tgt 576
Ser Gly Asp Tyr Tyr Thr His Tyr Lys Cys Asp Ser Pro Gly Met Cys
180 185 190

ggc tat gac ttg tat gaa aac gac aat gct gcc tgg gac tat gac aat 624
Gly Tyr Asp Leu Tyr Glu Asn Asp Asn Ala Ala Trp Asp Tyr Asp Asn
195 200 205

ggc ata tac tcc aca cag atg tac act cag aga gta cag caa atc tta 672
Gly Ile Tyr Ser Thr Gln Met Tyr Thr Gln Arg Val Gln Gln Ile Leu
210 215 220

5	get tcc cat aac ccc aca aag cct ata ttt tta tat att gcc tat caa 720 Ala Ser His Asn Pro Thr Lys Pro Ile Phe Leu Tyr Ile Ala Tyr Gln 235 230 235 240
10	get gtt cat tca cca ctg caa gct cct gcc agg tat ttc gaa cac tac 768 Ala Val His Ser Pro Leu Gln Ala Pro Gly Arg Tyr Phe Glu His Tyr 245 250 255
15	cga tcc att atc aac ata aac agg agg aga tat gct gcc atg ctt tcc 816 Arg Ser Ile Ile Asn Ile Asn Arg Arg Arg Tyr Ala Ala Met Leu Ser 260 265 270
20	tgc tta gat gaa gca atc aac aac gtg aca ttg gct cta aag act tat 864 Cys Leu Asp Glu Ala Ile Asn Asn Val Thr Leu Ala Leu Lys Thr Tyr 275 280 285
25	ggt ttc tat aac aac agc att atc att tac tct tca gat aat ggt gcc 912 Gly Phe Tyr Asn Asn Ser Ile Ile Ile Tyr Ser Ser Asp Asn Gly Gly 290 295 300
30	cag cct acg gca gga ggg agt aac tgg cct ctc aga ggt agc aaa gga 960 Gln Pro Thr Ala Gly Gly Ser Asn Trp Pro Leu Arg Gly Ser Lys Gly 305 310 315 320
35	aca tat tgg gaa gga ggg atc cgg gct gta gcc ttt gtg cat agc cca 1008 Thr Tyr Trp Glu Gly Ile Arg Ala Val Gly Phe Val His Ser Pro 325 330 335
40	ctt ctg aaa aac aag gga aca gtg tgt aag gaa ctt gtg cac atc act 1056 Leu Leu Lys Asn Lys Gly Thr Val Cys Lys Glu Leu Val His Ile Thr 340 345 350
45	gac tgg tac ccc act ctc att tca ctg gct gaa gga cag att gat gag 1104 Asp Trp Tyr Pro Thr Leu Ile Ser Leu Ala Glu Gly Gln Ile Asp Glu 355 360 365
50	gac att caa cta gat gcc tat gat atc tgg gag acc ata agt gag ggt 1152 Asp Ile Gln Leu Asp Gly Tyr Asp Ile Trp Glu Thr Ile Ser Glu Gly 370 375 380
55	ctt cgc tca ccc cga gta gat att ttg cat aac att gac ccc ata tac 1200 Leu Arg Ser Pro Arg Val Asp Ile Leu His Asn Ile Asp Pro Ile Tyr 385 390 395 400
60	acc aag gca aaa aat gcc tcc tgg gca gca gcc tat ggg atc tgg aac 1248 Thr Lys Ala Lys Asn Gly Ser Trp Ala Ala Gly Tyr Gly Ile Trp Asn 405 410 415
65	act gca atc cag tca gcc atc aga gtg cag cac tgg aaa ttg ctt aca 1296 Thr Ala Ile Gln Ser Ala Ile Arg Val Gln His Trp Lys Leu Leu Thr 420 425 430
70	gga aat cct gcc tac agc gac tgg gtc ccc cct cag tct ttc agc aac 1344 Gly Asn Pro Gly Tyr Ser Asp Trp Val Pro Pro Gln Ser Phe Ser Asn 435 440 445
75	ctg gga ccg aac cgg tgg cac aat gaa cgg atc acc ttg tca act gcc 1392 Leu Gly Pro Asn Arg Trp His Asn Glu Arg Ile Thr Leu Ser Thr Gly 450 455 460
80	aaa agt gta tgg ctt ttc aac atc sca gcc gac cca tat gag agg gtg 1440 Lys Ser Val Trp Leu Phe Asn Ile Thr Ala Asp Pro Tyr Glu Arg Val 465 470 475 480
85	gac cta tct aac agg tat cca gga atc gtg aag aag ctc cta cgg agg 1488 Asp Leu Ser Asn Arg Tyr Pro Gly Ile Val Lys Lys Leu Leu Arg Arg 485 490 495
90	ctc tca cag ttc aac aaa act gca gtg ccg gtc agg tat ccc ccc aaa 1536 Leu Ser Gln Phe Asn Lys Thr Ala Val Pro Val Arg Tyr Pro Pro Lys 500 505 510
95	gac ccc aga agt aac cct agg ctc aat gga ggg gtc tgg gga cca tgg 1584 Asp Pro Arg Ser Asn Pro Arg Thr Asn Gly Gly Val Trp Gly Pro Trp 515 520 525
100	tat aaa gag gaa acc aag aaa aag aag cca agc aaa aat cag gct gag 1632 Tyr Lys Glu Glu Thr Lys Lys Lys Lys Pro Ser Lys Asn Gln Ala Glu

	530	535	540	
5	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
10	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			
15	<213> Homo sapiens			
	<400> 91			
20	Met Gly Ala Leu Ala Gly Phe Trp Ile Leu Cys Leu Leu Thr Tyr Gly			
	1	5	10	15
25	Tyr Leu Ser Trp Gly Gln Ala Leu Glu Glu Glu Glu Glu Gly Ala Leu			
		20	25	30
30	Leu Ala Gln Ala Gly Glu Lys Leu Glu Pro Ser Thr Thr Ser Thr Ser			
		35	40	45
35	Gln Pro His Leu Ile Phe Ile Leu Ala Asp Asp Gln Gly Phe Arg Asp			
		50	55	60
40	Val Gly Tyr His Gly Ser Glu Ile Lys Thr Pro Thr Leu Asp Lys Leu			
		65	70	75
45	Ala Ala Glu Gly Val Lys Leu Glu Asn Tyr Tyr Val Gln Pro Ile Cys			
		85	90	95
50	Thr Pro Ser Arg Ser Gln Phe Ile Thr Gly Lys Tyr Gln Ile His Thr			
		100	105	110
55	Gly Leu Gln His Ser Ile Ile Arg Pro Thr Gln Pro Asn Cys Leu Pro			
		115	120	125
60	Leu Asp Asn Ala Thr Leu Pro Gln Lys Leu Lys Glu Val Gly Tyr Ser			
		130	135	140
65	Thr His Met Val Gly Lys Trp His Leu Gly Phe Tyr Arg Lys Glu Cys			
		145	150	155
70	Met Pro Thr Arg Arg Gly Phe Asp Thr Phe Phe Gly Ser Leu Leu Gly			
		165	170	175
75	Ser Gly Asp Tyr Tyr Thr His Tyr Lys Cys Asp Ser Pro Gly Met Cys			
		180	185	190
80	Gly Tyr Asp Leu Tyr Glu Asn Asp Asn Ala Ala Trp Asp Tyr Asp Asn			
		195	200	205
85	Gly Ile Tyr Ser Thr Gln Met Tyr Thr Gln Arg Val Gln Gln Ile Leu			
		210	215	220
90	Ala Ser His Asn Pro Thr Lys Pro Ile Phe Leu Tyr Ile Ala Tyr Gln			
		225	230	235
95				240

5 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
 10 260 265 270
 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
 15 290 295 300
 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
 20 325 330 335
 Leu Leu Lys Asn Lys Gly Thr Val Cys Lys Glu Leu Val His Ile Thr
 25 340 345 350
 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
 30 370 375 380
 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
 35 405 410 415
 Thr Ala Ile Gln Ser Ala Ile Arg Val Gln His Trp Lys Leu Leu Thr
 40 420 425 430
 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
 45 450 455 460
 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
 50 485 490 495
 Leu Ser Gln Phe Asn Lys Thr Ala Val Pro Val Arg Tyr Pro Pro Lys
 55 500 505 510
 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
 60 530 535 540
 Lys Lys Gln Lys Lys Ser Lys Lys Lys Lys Lys Lys Gln Gln Lys Ala
 545 550 555 560
 65 Val Ser Gly Ser Thr Cys His Ser Gly Val Thr Cys Gly
 565 570

5

<210> 92

<211> 1710

<212> ADN

•213• *Homo sapiens*

10

220

«221» CDS

222 (1)-(1707)

<223> hSUMF5

15

~~400~~ 92

20

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	gac	aag	cag	agc	ttc	gtg	gac	gac	ggg	ccc	96
Tyr	Leu	Ser	Trp	Asp	Trp	Ala	Lys	Pro	Ser	Phe	Val	Ala	Asp	Gly	Pro	
			20					25					30			

25

ggg gag ggt ggc gag cag ccc tgg gcc gct cgg 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

30

atc atc ttc atc etc 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 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

35

ggg gtc aag ttg gag aat tat tac atc cag ccc atc tgc acg cct tgc 288
Gly Val Lys Leu Glu Asn Tyr Tyr Ile Glu Pro Ile Cys Thr Pro Ser
85 90 95

cgg agc gag ctc ctc act ggc agg tac cag atc cac aca gga ctc cag 136
 Arg Ser Gln Leu Leu Thr Gly Arg Tyr Gln Ile His Thr Gly Leu Gln
 100 105 110

40

cat tcc atc atc cgc cca gag gag ccc aac tgc ctg ccc ctg gac gag 384
His Ser Ile Ile Arg Pro Gln Gln Pro Asn Cys Leu Pro Leu Asp Gln
115 130 135

45

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 egg 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

50

cgt cgg ggc ttc gac acc ttc ctg ggc tgg 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

55

tat tac aac 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	gac	tgg	ggg	ctc	agg	ggc	cag	tac	tcg	624
Leu	His	Glu	Gly	Glu	Asn	Val	Ala	Trp	Gly	Leu	Ser	Gly	Gln	Tyr	Ser	
		195					200					205				

60

act atg ctt tac gcc gag 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 ctg 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

65

5	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 768 255
10	aat gtg gcc cgg cgg aag tac gcg gcc atg gtg acc tgc atg gat gag Asn Val Ala Arg Arg Lys Tyr Ala 265 Ala Met Val Thr Cys Met Asp Glu 816 260 270
15	gct gtg cgc aac atc acc tgg gcc ctc aag cgc tac ggt ttc tac aac Ala Val 275 Asn Ile Thr Trp 280 Ala Leu Lys Arg Tyr 285 Gly Phe Tyr Asn 864
20	aac agt gtc atc atc ttc tcc agt gac aat ggt ggc cag act ttc tgg Asn Ser Val Ile Ile Phe Ser Ser Asp Asn Gly Gly Gln Thr Phe Ser 912 290 295 300
25	ggg ggc agc aac tgg ccg ctc cga gga cgc aag ggc act tat tgg gaa Gly Gly Ser Asn Trp 310 Pro Leu Arg Gly Arg Lys Gly Thr Tyr Trp Glu 960 305 315 320
30	ggt ggc gtg cgg gcc cta gcc ttt gtc cac agt ccc ctg ctc aag cga Gly Gly Val Arg 325 Gly Leu Gly Phe Val 330 His Ser Pro Leu Leu Lys Arg 1008 335
35	aag caa cgg aca agc cgg gca ctg atg cac atc act gac tgg tac cgg Lys Gln Arg Thr Ser Arg Ala Leu Met His Ile Thr Asp Trp Tyr Pro 1056 340 345 350
40	acc ctg gtg ggt ctg gca ggt ggt acc acc tca gca gtc gat ggg cta Thr Leu Val Gly Leu Ala Gly Gly Thr Thr Ser Ala Ala Asp Gly Leu 1104 355 360 365
45	gat gcc tac gac gtg tgg ccg gcc atc agc gag ggc cgg gcc tca cca Asp Gly Tyr Asp Val Trp 375 Pro Ala Ile Ser Glu Gly Arg Ala Ser Pro 1152 370 380
50	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 1200 385 390 395 400
55	cat gcc tcc ctg gag gcc gcc ttt ggc atc tgg aac acc gcc gtg cag His Gly Ser Leu 405 Gly Gly Phe Gly Ile Trp Asn Thr Ala Val Gln 1248 410 415
60	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 1396 420 425 430
65	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 1344 435 440 445
70	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 1392 450 455 460
75	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 1440 465 470 475 480
80	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 1488 485 490 495
85	cgc aca gcc atc ccg gta cgc tac cca gct gag aac ccc cgg gct cat Arg Thr Ala Ile Pro Val Arg Tyr 505 Pro Ala Glu Asn Pro Arg Ala His 1536 500 510
90	cct gac ttt aat ggg ggt gct tgg ggg ccc tgg gcc agt gat gag gaa Pro Asp Phe Asn Gly Gly Ala 520 Trp Gly Pro Trp Ala Ser Asp Glu Glu 1584 515 525
95	gag gag gaa gag gaa ggg agg gct cga agc ttc tcc cgg ggt cgt cgc Glu Glu Glu Glu Glu Gly 535 Ala Arg Ser Phe Ser Arg Gly Arg Arg 1632 530 540
100	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

5

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

10

<210> 93

<211> 569

<212> PRT

15

<213> Homo sapiens

<400> 93

20

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

25

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

30

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

35

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

40

Val Thr Leu Pro Gln Lys Leu Gln Glu Ala Gly Tyr Ser Thr His Met

130 135 140

45

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

50

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

55

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

60

Pro Leu Gln Ser Pro Arg Glu Tyr Leu Tyr Arg Tyr Arg Thr Met Gly

245 250 255

65

5 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
 10 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
 15 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
 25 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
 30 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
 35 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
 40 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
 45 Arg Thr Ala Ile Pro Val Arg Tyr Pro Ala Glu Asn Pro Arg Ala His
 500 505 510
 50 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
 55 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> ADN

5 <213> Homo sapiens

<220>

<221> CDS

10 <222> (1)..(2064)

<223> hSULF6

<400> 94

15 atg cta att tca gga aga gaa gag aac caa ata gac ata tcc aag acc 48
Met Leu Ile Ser Gly Arg Glu Glu Asn Gln Ile Asp Ile Ser Lys Thr
1 5 10 15

20 aca gag gta gat tgt ttt gtg gtt gaa tta gga agt cta cac aat cct 96
Thr Glu Val Asp Cys Phe Val Val Glu Leu Gly Ser Leu His Asn Pro
20 25 30

25 aca egg aac cca cag cga att ttc acc aag cac gtg gcc acc aag tca 144
Thr Arg Asn Pro Gln Arg Ile Phe Thr Lys His Val Ala Thr Lys Ser
35 40 45

30 tcc agc tcc aaa tgt cag ctg gac caa ggt gga aaa agc ctg gtc cag 192
Ser Ser Ser Lys Cys Gln Leu Asp Gln Gly Gly Lys Ser Leu Val Gln
50 55 60

35 tgc att tta ccc aga tct tca aag ctc ctc tca ccc ttg tgt ctc ccc 240
Cys Ile Leu Pro Arg Ser Ser Lys Leu Leu Ser Pro Leu Cys Leu Pro
65 70 75 80

40 cat ccg tgt gga gct tta ctt ctg tat aga tcc tca gga atc gcc tct 288
His Pro Cys Gly Ala Leu Leu Leu Tyr Arg Ser Ser Gly Ile Ala Ser
85 90 95

45 gct ctt gct gcc ttt aca gac tcc ctc tct agg agc tgc tgg ctg tca 336
Ala Leu Ala Ala Phe Thr Asp Ser Leu Ser Arg Ser Cys Trp Leu Ser
100 105 110

50 gtg tcc ctg tgc tgt ttg ttt tgc ggt gtt gat gcc aca ttt atg aca 384
Val Ser Leu Cys Cys Leu Phe Cys Gly Val Asp Gly Thr Phe Met Thr
115 120 125

55 aga aac gcc aga ccc aac att gtc ctg ctg atg gca gat gac ctt gga 432
Arg Asn Ala Arg Pro Asn Ile Val Leu Leu Met Ala Asp Asp Leu Gly
130 135 140

60 gtg ggg gat ttg tgc tgc tac ggt aat aac tca gtg agc aca cct aat 480
Val Gly Asp Leu Cys Cys Tyr Gly Asn Asn Ser Val Ser Thr Pro Asn
145 150 155 160

65 att gac cgc ctg gca agt gaa gga gtg agg ctt acc cag cat ctc gca 528
Ile Asp Arg Leu Ala Ser Glu Gly Val Arg Leu Thr Gln His Leu Ala
165 170 175

70 gct gct tcc atg tgc acc cca agt cgg gct gcc ttc ctg acc ggc cgg 576
Ala Ala Ser Met Cys Thr Pro Ser Arg Ala Ala Phe Leu Thr Gly Arg
180 185 190

75 tac ccc atc aga tca ggg atg gtg tct gcc tac aac ctg aac cgt gcc 624
Tyr Pro Ile Arg Ser Gly Met Val Ser Ala Tyr Asn Leu Asn Arg Ala
195 200 205

80 ttc acg tgg ctt ggt ggg tca ggt ggt ctt ccc acc aat gaa acg act 672
Phe Thr Trp Leu Gly Gly Ser Gly Gly Leu Pro Thr Asn Glu Thr Thr
210 215 220

85 ttt gcc aag ctg ctg cag cac cgt gcc tac cgc acg gga ctc ata ggc 720
Phe Ala Lys Leu Leu Gln His Arg Gly Tyr Arg Thr Gly Leu Ile Gly
225 230 235 240

90 aaa tgg cac ctg ggt ttg agc tgc gcc tct cgg aat gat cac tgt tac 768
Lys Trp His Leu Gly Leu Ser Cys Ala Ser Arg Asn Asp His Cys Tyr
245 250 255

5		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	816
		260 265 270	
10		att tta agc gac tgc cag gca tcc aag aca cca gaa etg cac cgc tgg Leu Leu Ser Asp Cys Gln Ala Ser Lys Thr Pro Glu Leu His Arg Trp	864
		275 280 285	
15		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	912
		290 295 300	
20		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	960
		305 310 315 320	
25		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	1008
		325 330 335	
30		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	1056
		340 345 350	
35		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	1104
		355 360 365	
40		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	1152
		370 375 380	
45		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	1200
		385 390 395 400	
50		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	1248
		405 410 415	
55		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	1296
		420 425 430	
60		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	1344
		435 440 445	
65		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	1392
		450 455 460	
70		ggc tgg aac ggg atc tac aaa ggt gcc aaa gga atg gga gga tgg gaa Gly Trp Asn Gly Ile Tyr Lys Gly Gly Lys Gly Met Gly Gly Trp Glu	1440
		465 470 475 480	
75		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	1488
		485 490 495	
80		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	1536
		500 505 510	
85		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	1584
		515 520 525	
90		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	1632
		530 535 540	
95		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	1680
		545 550 555 560	
100		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

[illegible]

```

30      <210> 95
      <211> 688
      <212> PRT
      <213> Homo sapiens
35      <400> 95

```

	Met	Leu	Ile	Ser	Gly	Arg	Glu	Glu	Asn	Gln	Ile	Asp	Ile	Ser	Lys	Thr	
	1				5					10					15		
40	Thr	Glu	Val	Asp	Cys	Phe	Val	Val	Glu	Leu	Gly	Ser	Leu	His	Asn	Pro	
				20					25					30			
45	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					
50	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		
55	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				
60	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	
65	145					150					155					160	

5 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
 10 180 185 190
 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
 15 210 215 220
 Phe Ala Lys Leu Leu Gln His Arg Gly Tyr Arg Thr Gly Leu Ile Gly
 20 225 230 235 240
 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
 25 260 265 270
 Leu Leu Ser Asp Cys Gln Ala Ser Lys Thr Pro Glu Leu His Arg Trp
 30 275 280 285
 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
 35 305 310 315 320
 Lys Val Ile Phe Val Phe Ala Leu Leu Ala Phe Leu Phe Phe Thr Ser
 40 325 330 335
 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
 45 355 360 365
 Ser Leu Met Leu Lys Glu Ala Leu Ala Phe Ile Glu Arg Tyr Lys Arg
 50 370 375 380
 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
 55 405 410 415
 Asp Asn Val Glu Glu Met Asp Trp Met Val Gly Lys Ile Leu Asp Ala
 60 420 425 430
 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
 65 450 455 460

5 Gly Trp Asn Gly Ile Tyr Lys Gly Gly Lys Gly Met Gly Gly Trp Glu
465 470 475 480

10 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

15 Thr Leu Ser Tyr Ile Gly Gly Gly Ile Leu Ser Gln Asp Arg Val Ile
515 520 525

20 Asp Gly Gln Asn Leu Met Pro Leu Leu Glu Gly Arg Ala Ser His Ser
530 535 540

25 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

30 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

35 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

40 Arg Glu His Arg Arg Thr Leu Thr Pro Val Pro Gln Gln Phe Ser Val
645 650 655

45 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

Reivindicaciones

1. Una célula que expresa conjuntamente una sulfatasa y una enzima generadora de Cα-formilglicina (FGE - Cα-formylglycine generating enzyme) para que la sulfatasa activada sea producida; donde la célula comprende ADN o ARN heterólogo que resulta en una expresión incrementada de la sulfatasas activadas en relación a lo que ocurriría en la ausencia del ADN o ARN heterólogo;

Y donde el FGE es un polipéptido con una actividad generadora de Cα-formilglicina que:

- a) Tiene una secuencia seleccionada de un grupo que consiste de las IDENTIFICACIONES SECUENCIALES NÚMEROS 2, 5, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, o los aminoácidos 34-374 de la IDENTIFICACIÓN SECUENCIAL NÚMERO 2; o
 - b) Tiene por lo menos el 50% de identidad secuencial de la IDENTIFICACIÓN SECUENCIAL NÚMERO 2; o
 - c) Tiene una o más mutaciones conservadoras de aminoácidos en relación a un polipéptido tal como se describió en a) o b)
 - d) Es un fragmento de un polipéptido tal como se describió en desde a) a c); o
 - e) Es una proteína de fusión de cualquiera de la a) a la d).
2. Una célula de acuerdo a la reivindicación 1; donde la FGE es expresada a un nivel más alto en relación a lo que ocurriría en la ausencia del ADN o del ARN heterólogo.
 3. Una célula de acuerdo a la reivindicación 1 o la reivindicación 2; donde el nivel de la identidad secuencial en la parte b) de la reivindicación 1 que es por lo menos un 75%.
 4. Una célula de acuerdo a la reivindicación 3; donde el nivel de identidad secuencial en la parte b) de la reivindicación 1 que es por lo menos un 95%
 5. Una célula de acuerdo a la reivindicación 1; donde la FGE es un polipéptido tal como se describió en la parte a) de esa reivindicación.
 6. Una célula de acuerdo a cualquiera de las reivindicaciones 1 a 5; donde la FGE es un polipéptido que comprende la IDENTIFICACIÓN SECUENCIAL NÚMERO 2, o uno de sus fragmentos que tienen una actividad generadora de Cα-formilglicina.
 7. Una célula de acuerdo a la reivindicación 6, donde la FGE es un polipéptido que comprende a la IDENTIFICACIÓN SECUENCIAL NÚMERO 2 o a los aminoácidos 34-374 de la IDENTIFICACIÓN SECUENCIAL NÚMERO 2.
 8. Una célula de acuerdo a cualquiera de las reivindicaciones 1 a 5; donde la FGE es un polipéptido que tiene un sub dominio 3 que comprende por lo menos uno de los siguiente:
 - (i) un motivo GFR
 - (ii) un motivo RVXXGG(A)S
 - (iii) un heptámero que contiene 3 argininas
 - (iv) 3 residuos de cisteína.
 9. Una célula de acuerdo a cualquiera de las reivindicaciones anteriores; donde la sulfatasa es una sulfatasa eucariota, una sulfatasa de mamíferos o una sulfatasa humana.
 10. Una célula de acuerdo a cualquiera de las reivindicaciones anteriores; donde la sulfatasa es seleccionada de un grupo que consiste de una 2-sulfatasa de Iduronto, una Sulfamidasa, una 6-sulfatasa de N-Acetilgalactosamina, una 6-sulfatasa de N-Acetilglucosamina, una arilsulfatasa A, una arilsulfatasa B, una arilsulfatasa C, una arilsulfatasa D, una arilsulfatasa E, una arilsulfatasa F, una arilsulfatasa G, una HSulf-1, una HSulf-2, una HSulf-3, una HSulf-4, una HSulf-5, una HSulf-6, o uno de sus fragmentos que tenga una actividad de sulfatasas.
 11. Una célula de acuerdo a la reivindicación 10; donde la sulfatasa es 2-sulfatasa de Iduronato.
 12. Una célula de acuerdo a cualquiera de las reivindicaciones anteriores; donde la sulfatasa es exógena, en relación a la célula sin el ADN o ARN heterólogo.

13. Una célula de acuerdo a cualquiera de las reivindicaciones anteriores la cual es una célula eucariota.
14. Una célula de acuerdo a cualquiera de las reivindicaciones anteriores donde la célula ha sido transformada usando ADN o ARN heterólogo de la reivindicación 1 o es un cultivo celular derivado de esta.
15. Una célula de acuerdo a cualquiera de las reivindicaciones anteriores que es una célula primaria, o una célula secundaria, o es una célula proveniente de una línea celular inmortalizada.
16. Un método in vitro que comprende usar una o más células de acuerdo a cualquiera de las reivindicaciones 1 a la 15 para producir sulfatasas activadas.
17. Un método de acuerdo a la reivindicación 16 que es usado para producir sulfatasas con una proporción más alta de sulfatasas activadas en comparación con las sulfatasas totales que si fuese el caso para las sulfatasas producidas provenientes de dicha célula o células en la ausencia del ADN o ARN heterólogo.
18. Un método de acuerdo a la reivindicación 16 o 17, donde la sulfatasa es suministrada en una forma puede ser utilizada en una composición farmacéutica para el tratamiento de una deficiencia de sulfatasas.
19. Un método de acuerdo a cualquiera de las reivindicaciones 16 a 18 que comprende suministrar una composición farmacéutica que comprende sulfatasas activadas y un portador farmacéuticamente aceptable.
20. Un método de acuerdo a la reivindicación 18 o 19 donde la composición farmacéutica suministrada en una forma para administración por rutas orales, rectales, tópicas, nasales, intradérmicas, transdérmicas o parenterales.
21. Un método de acuerdo a cualquiera de las reivindicaciones 18 a 20, donde la composición farmacéutica es suministrada en una forma que es adecuada para su uso en el tratamiento de una deficiencia de sulfatasas.
22. Un método de acuerdo a las reivindicaciones 18 a 21, donde la composición farmacéutica es suministrada en una forma de dosis unitarias que es adecuada para su uso en el tratamiento de una deficiencia de sulfatasas.
23. Un método de acuerdo a cualquiera de las reivindicaciones 16 a 22 que comprende ensayos de la actividad de la sulfatasas.
24. Una célula de acuerdo a cualquiera de las reivindicaciones 1 a la 15, para su uso en un método para tratamiento de una deficiencia de sulfatasas.
25. Una célula de acuerdo a las reivindicaciones 1 a la 15, para su uso en un método para el tratamiento de una deficiencia de sulfatasas; donde la deficiencia de sulfatasas es seleccionada de un grupo que consiste de: la deficiencia de sulfatasas múltiples, la mucopolisacaridosis II, la mucopolisacaridosis IIIA, la mucopolisacaridosis IVA, la mucopolisacaridosis VI, la mucopolisacaridosis VIII, la leucodistrofia meta-cromática, la condrodysplasia punctata 1 recesiva vinculada con el cromosoma X y la ictiosis vinculada con el cromosoma X.
26. Una sulfatasa para su uso en un método para tratar una deficiencia de sulfatasas en un sujeto, donde la sulfatasa es producida por una célula con una expresión incrementada de enzimas generadoras de formilglicina (FGE - formylglycine generating enzyme) para incrementar la tasa de sulfatasas activas en comparación a las sulfatasas totales en un monto efectivo para incrementar la actividad específica de dichas sulfatasas, donde la FGE es un polipéptido con una actividad generadora de Cα-formilglicina que:
 - a) Tiene una secuencia seleccionada de un grupo que consiste de las IDENTIFICACIONES SECUENCIALES NÚMEROS 2, 5, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, o los aminoácidos 34-374 de la IDENTIFICACIÓN SECUENCIAL NÚMERO 2, o
 - b) Tiene por lo menos una identidad secuencial del 50% en relación a la IDENTIFICACIÓN SECUENCIAL NÚMERO 2; o
 - c) Tiene por lo menos una o más mutaciones conservadoras de aminoácidos en relación a un polipéptido tal como se describió en a) a c); o
 - d) Es un fragmento de un polipéptido tal como se describió en cualquiera de a) a c);
 - e) Es una proteína de fusión de cualquiera de a) a d)

27. La sulfatasa para su uso de acuerdo a la reivindicación 26, donde la sulfatasa es una 2-sulfatasa de Iduronato.

5

10

15

20

25

30

35

40

45

50

55

60

65

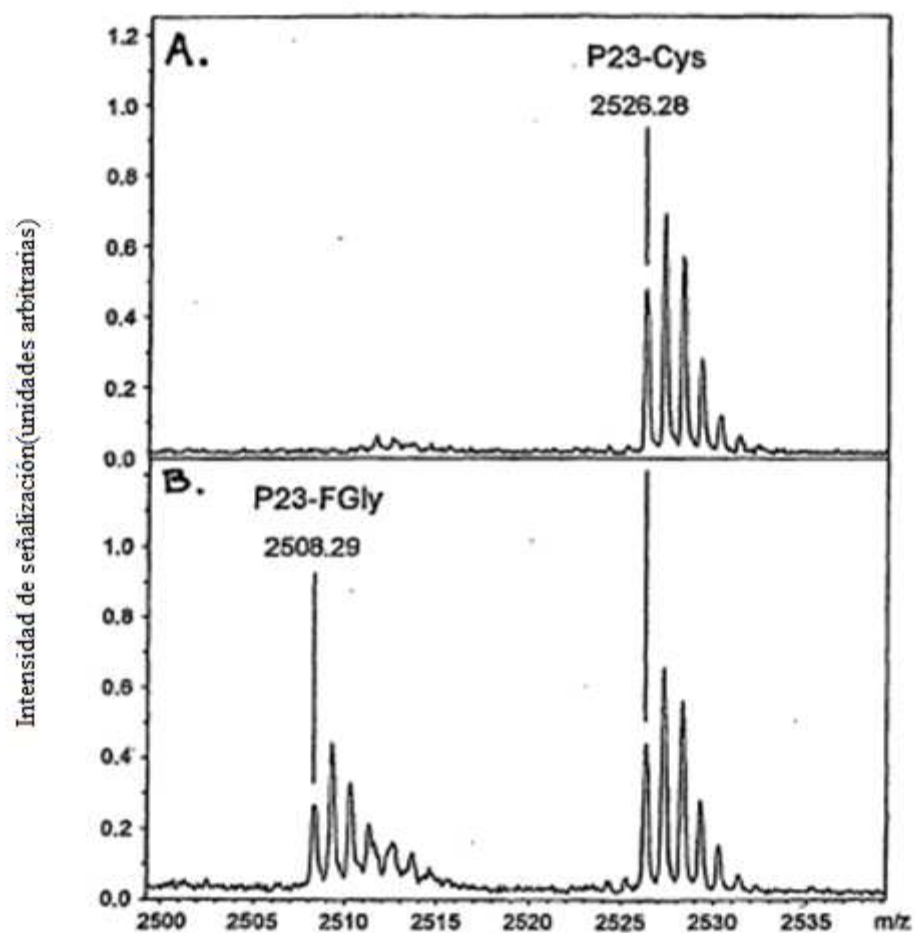


Fig. 1

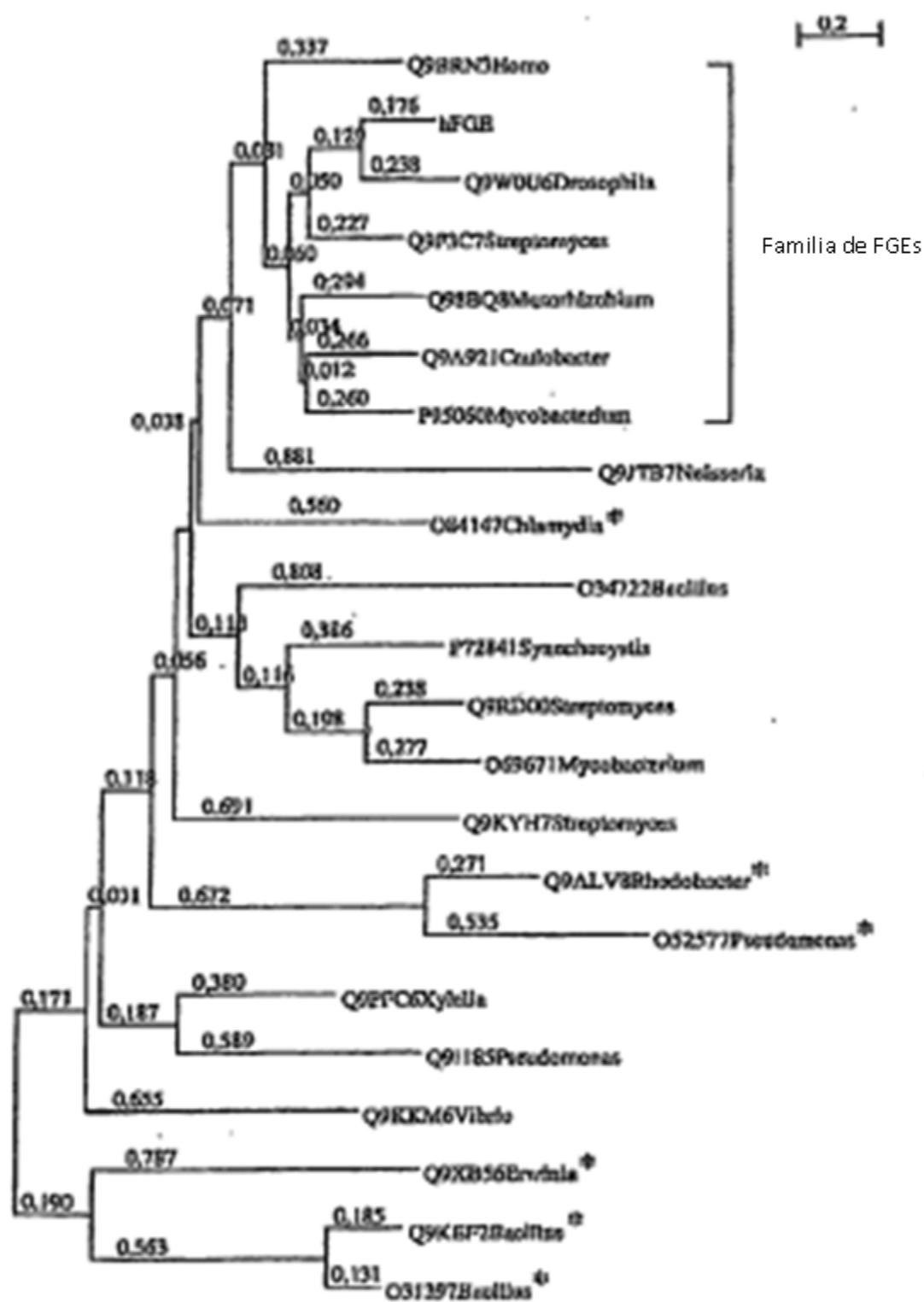


Fig. 2

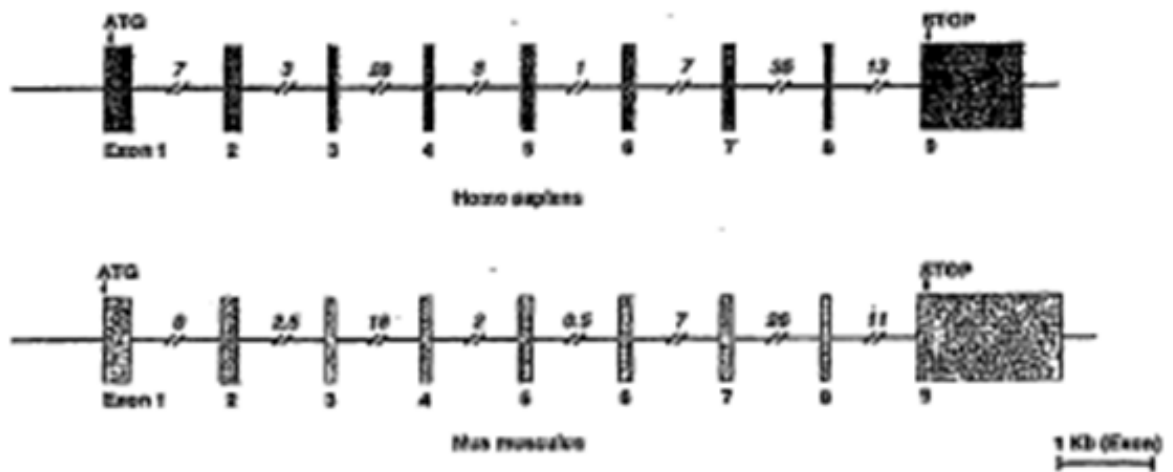


Fig. 3

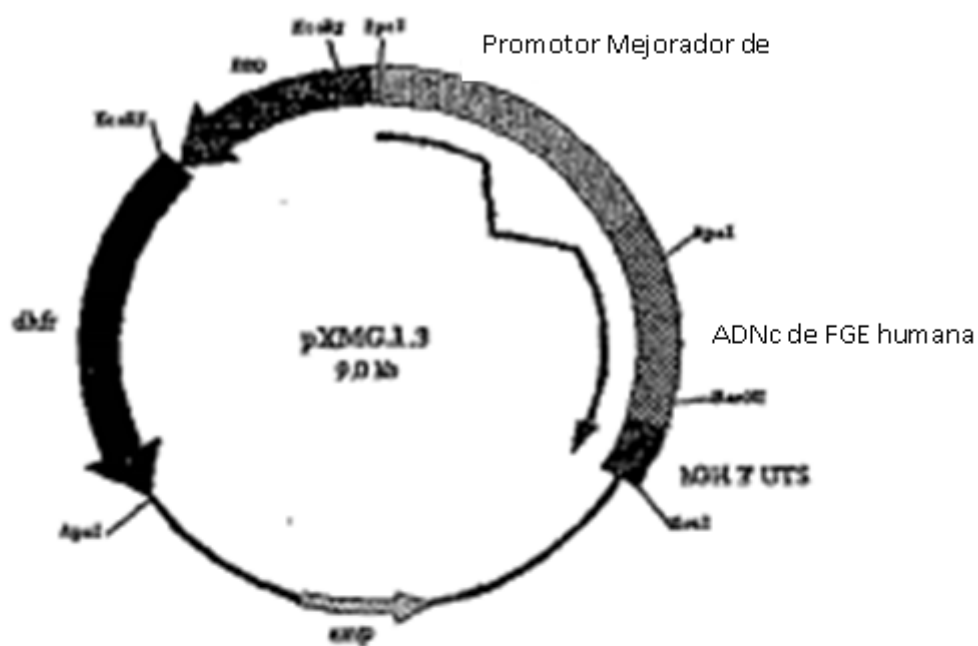


Fig. 4

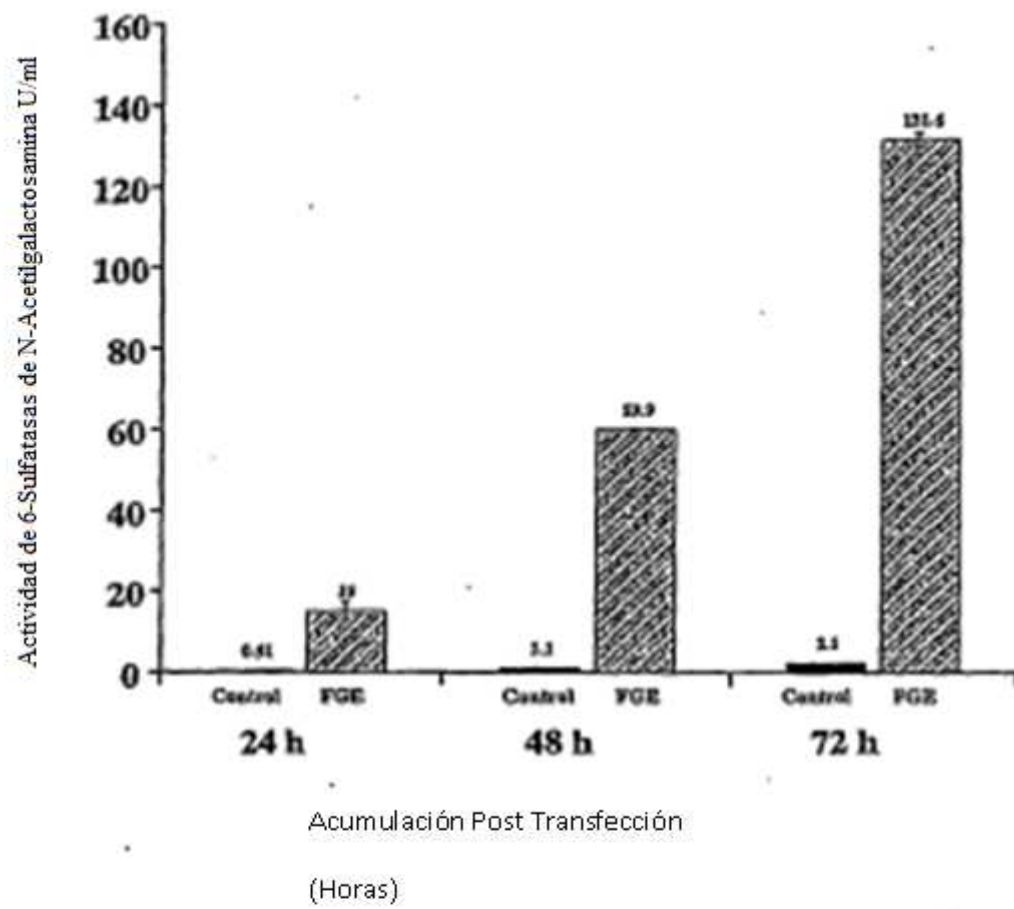
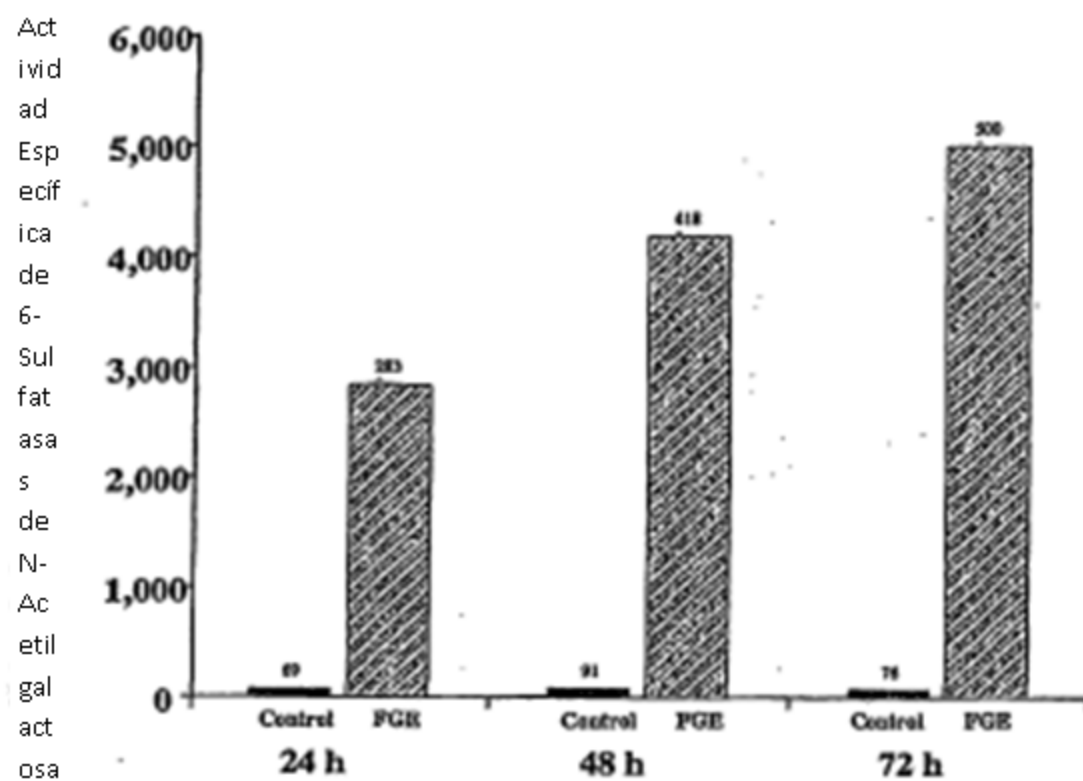


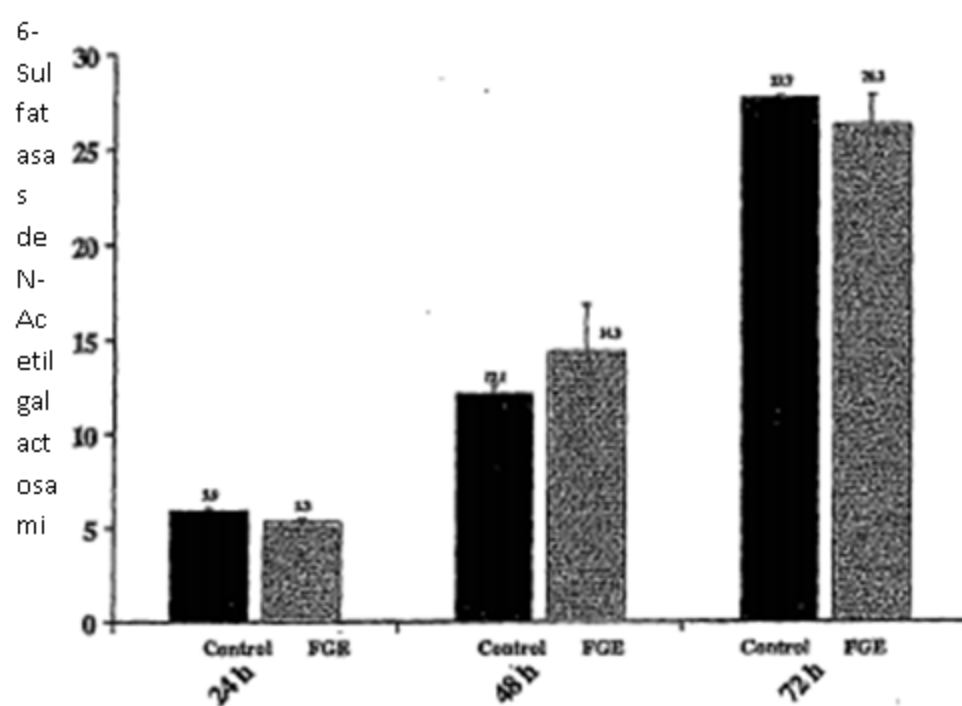
Fig. 5



Acumulación Post Transfección

(Horas)

Fig. 6



Acumulación Post Transfección
(Horas)

Fig. 7

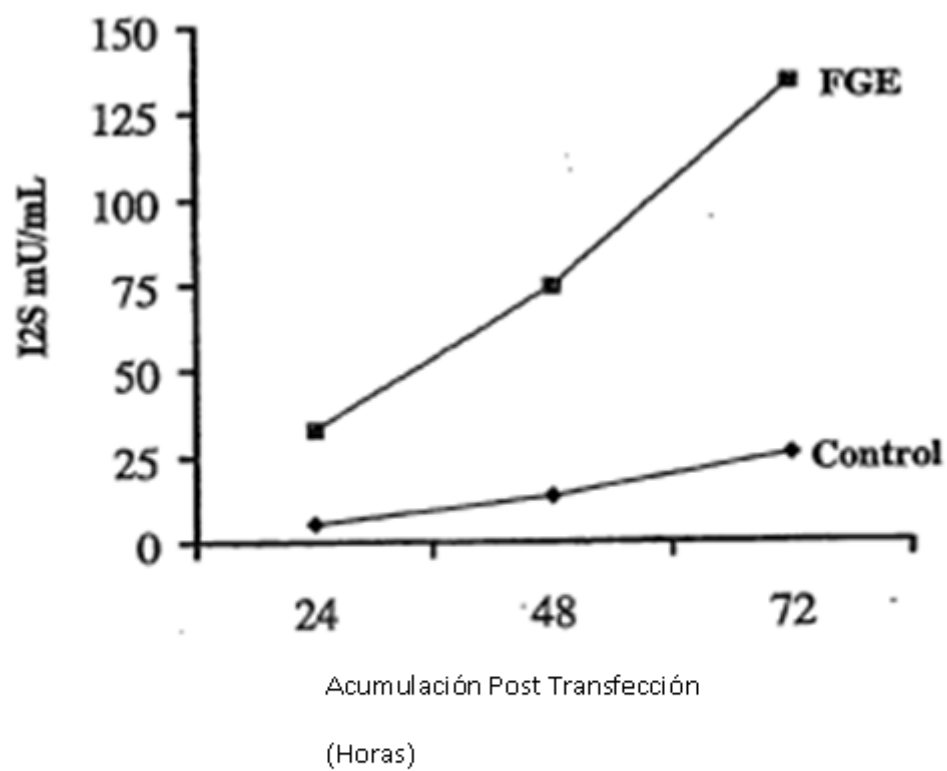


Fig. 8

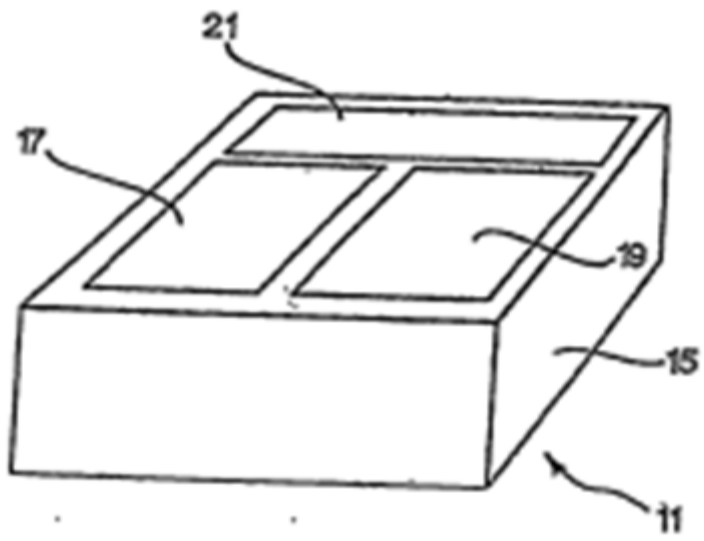


Fig. 9