

19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 555 056**

51 Int. Cl.:

**C12N 15/52** (2006.01)

**C12N 9/02** (2006.01)

**A61K 38/36** (2006.01)

**A61K 38/44** (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 10182740 (0)**

97 Fecha y número de publicación de la concesión europea: **02.09.2015 EP 2325301**

54 Título: **Diagnóstico y tratamiento de deficiencia múltiple de sulfatasas y otras usando una enzima generadora de formilglicina (FGE)**

30 Prioridad:

**11.02.2003 US 447747 P**

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

**28.12.2015**

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, DR. y  
COSMA, MARIA PIA**

74 Agente/Representante:

**IZQUIERDO BLANCO, María Alicia**

ES 2 555 056 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).

## DESCRIPCIÓN

**Diagnóstico y tratamiento de deficiencia múltiple de sulfatasas y otras usando una enzima generadora de formilglucina (FGE)****Campo de la invención**

La presente invención se refiere a usos médicos para el tratamiento de deficiencia múltiple de sulfatasas (MSD) y otras deficiencias de sulfatasas, además de a métodos de diagnóstico o pronóstico de tales deficiencias.

**Antecedentes de la invención**

Las sulfatasas son miembros de una familia de genes altamente conservada, que comparten una amplia homología de secuencia (Franco, B., et al., Cell, 1995, 81:15-25; Parenti, G., et al., Curr. Opin. Gen. Dev., 1997, 7:386-391), un alto grado de similitud estructural (Bond, C.S., et al., Structure, 1997, 5:277-289; Lukatela, G., et al., Biochemistry, 1998, 37:3654-64) y una modificación postraduccional única que es esencial para la escisión del éster de sulfato (Schmidt, B., et al., Cell, 1995, 82: 271-278; Selmer, T., et al., Eur. J. Biochem., 1996, 238:341-345). La modificación postraduccional implica la oxidación de un residuo conservado de cisteína (en eucariotas) o serina (en determinados procariontes), en C<sub>β</sub>, proporcionando L-C<sub>α</sub>-formilglucina (también conocida como *FGly*, ácido 2-amino-3-oxopropanoico) en el que un grupo aldehído sustituye al grupo tiometilo de la cadena lateral. El aldehído es una parte esencial del sitio catalítico de la sulfatasa y probablemente actúa como hidrato de aldehído. Uno de los grupos hidroxilo geminales acepta el sulfato durante la escisión del éster de sulfato conduciendo a la formación de un producto intermedio de enzima covalentemente sulfatada. El otro hidroxilo se requiere para la eliminación posterior del sulfato y regeneración del grupo aldehído. Esta modificación se produce en el retículo endoplásmico durante, o poco después de, la importación del polipéptido de sulfatasa naciente y se dirige por una secuencia lineal corta que rodea al residuo de cisteína (o serina) que va a modificarse. Esta secuencia altamente conservada es el hexapéptido L/V-C(S)-X-P-S-R (SEC ID N°: 32), presente en la región del extremo N de todas las sulfatasas eucariotas y lo más frecuentemente lleva un grupo hidroxilo o tiol en el residuo X (Dierks, T., et al., Proc. Natl. Acad. Sci. U. S. A., 1997, 94:11963-11968).

El documento WO 01/60991 trata proteínas cinasas humanas (PKIN) y polinucleótidos que identifican y codifican PKIN. También trata vectores de expresión adecuados, células huésped, anticuerpos, agonistas y antagonistas, además de métodos de diagnóstico, tratamiento o prevención de trastornos asociados a expresión aberrante de PKIN.

Fey et al. (J. Biol. Chem., 276, No 50, 47021-47028 [2001]) trata la caracterización de la formación postraduccional de formilglucina por componentes luminales del retículo endoplásmico.

Dierks et al. (PNAS, 94, 11963-11968 [1997]) trata la conversión de cisteína en formilglucina y la modificación de proteínas en sulfatasas eucariotas.

Dierks et al. (EMBO Journal, 18, No 8, 2084-2091 [1999]) trata los determinantes de secuencia que dirigen la conversión de cisteína en formilglucina en sulfatasas eucariotas.

Hasta la fecha se han identificado trece genes de sulfatasa en seres humanos. Codifican enzimas con diferente especificidad de sustrato y localización subcelular tales como lisosomas, aparato de Golgi y RE. Cuatro de esos genes, *ARSC*, *ARSD*, *ARSE* y *ARSF*, que codifican arilsulfatasa C, D, E y F, respectivamente, están localizados dentro de la misma región cromosómica (Xp22.3). Comparten una similitud de secuencia significativa y una organización genómica casi idéntica, que indica que surgen de eventos de duplicación que se produjeron recientemente durante la evolución (Franco B, et al., Cell, 1995, 81:15-25; Meroni G, et al., Hum Mol Genet, 1996, 5:423-31).

La importancia de las sulfatasas en el metabolismo humano se enfatiza por la identificación de al menos ocho enfermedades monogénicas humanas provocadas por la deficiencia de actividades de sulfatasa individuales. La mayoría de estas afecciones son trastornos de almacenamiento lisosómico en los que las consecuencias fenotípicas se derivan del tipo y la distribución tisular del material almacenado. Entre ellos hay cinco tipos diferentes de mucopolisacaridosis (tipos IIIA, IIID, IVA y VI de MPS) debido a deficiencias de sulfatasas que actúan sobre el catabolismo de glicosaminoglicanos (Neufeld y Muenzer, 2001, The mucopolysaccharidoses, en The Metabolic and Molecular Bases of Inherited Disease, C.R. Scriver, A.L. Beaudet, W.S. Sly, D. Valle, B. Childs, K.W. Kinzler y B. Vogelstein, eds. Nueva York: Me Graw-Hill, págs. 3421-3452) y leucodistrofia metacromática (MLD), que se caracteriza por el almacenamiento de sulfolípidos en los sistemas nerviosos central y periférico conduciendo a deterioro neurológico grave y progresivo. Dos enfermedades humanas adicionales están provocadas por deficiencias de sulfatasas no lisosómicas. Incluyen ictiosis ligada al cromosoma X, un trastorno cutáneo debido a deficiencia de sulfatasa esteroidea (STS/ARSC), y condrodiplosia punctata, un trastorno que afecta al hueso y cartílago debido a deficiencia de arilsulfatasa E (ARSE). Las sulfatasas también participan en síndromes de malformación humanos

inducidos por fármacos, tales como embriopatía por warfarina, provocada por la inhibición de la actividad ARSE debido a la exposición *in utero* a warfarina durante el embarazo.

5 En un trastorno monogénico humano intrigante, la deficiencia múltiple de sulfatasas (MSD), todas las actividades de sulfatasa son defectuosas simultáneamente. Por consiguiente, el fenotipo de esta enfermedad multisistémica grave combina las características observadas en deficiencias de sulfatasas individuales. Las células de pacientes con MSD son deficientes en actividades de sulfatasa incluso tras la transfección con ADNc que codifican 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., USA, 1992, 89:2561-2565). Se encontró que la modificación postraducciona

10 sulfatasas era defectuosa en un paciente con MSD, sugiriendo que este trastorno está provocado por una mutación en un gen, o genes, que participan en la maquinaria de conversión de cisteína en formilglicina (Schmidt, B., et al., Cell, 1995, 82:271-278). A pesar del intenso interés biológico y médico, los esfuerzos dirigidos a la identificación de este/estos gen(es) se han visto dificultados por la rareza de pacientes con MSD y la consiguiente falta de casos familiares adecuados para realizar un mapeo genético.

15 **Resumen de la invención**

La presente invención proporciona usos médicos para el tratamiento de deficiencia múltiple de sulfatasas (MIM 272200) y otras deficiencias de sulfatasas como se expone en las reivindicaciones 1 a 12 y 15, además de métodos de diagnóstico o pronóstico como se exponen en las reivindicaciones 13 y 14.

20

Los presentes inventores han identificado un gen que codifica la enzima generadora de formilglicina (FGE), una enzima responsable de la modificación postraducciona

25 Se ha descubierto, inesperadamente, que mutaciones en el gen de FGE conducen al desarrollo de deficiencia múltiple de sulfatasas (MSD) en sujetos. También se ha descubierto, inesperadamente, que la FGE potencia la actividad de sulfatasas, incluyendo, pero sin limitarse a, iduronato 2-sulfatasa, sulfamidasa, N-acetilgalactosamina 6-sulfatasa, N-acetilglucosamina 6-sulfatasa, arilsulfatasa A, arilsulfatasa B, arilsulfatasa C, arilsulfatasa D, arilsulfatasa E, arilsulfatasa F, arilsulfatasa G, HSulf-1, HSulf-2, HSulf-3, HSulf-4, HSulf-5 y HSulf-6. En vista de estos descubrimientos, las moléculas de la presente invención pueden usarse en el diagnóstico y tratamiento de deficiencia múltiple de sulfatasas, así como otras deficiencias de sulfatasa.

30

**Breve descripción de las secuencias**

35 SEC ID N°: 1 es la secuencia de nucleótidos del ADNc de FGE humana.

SEC ID N°: 2 es la secuencia de aminoácidos predicha del producto de traducción de ADNc de FGE humana (SEC ID N°: 1).

40 SEC ID N°: 3 es la secuencia de nucleótidos del ADNc de FGE humana que codifica el polipéptido de SEC ID N°: 2 (es decir, nucleótidos 20-1141 de SEC ID N°: 1).

SEC ID N°: 4 es la secuencia de nucleótidos con N° de acceso de GenBank AK075459.

45 SEC ID N°: 5 es la secuencia de aminoácidos predicha del producto de traducción de SEC ID N°: 4, un producto de proteína sin nombre que tiene el N° de acceso de GenBank BAC11634.

SEC ID N°: 6 es la secuencia de nucleótidos del ADNc de iduronato 2-sulfatasa humana (N° de acceso de GenBank M58342).

50 SEC ID N°: 7 es la secuencia de aminoácidos predicha del producto de traducción de ADNc de iduronato 2-sulfatasa humana (SEC ID N°: 6).

SEC ID N°: 8 es la secuencia de nucleótidos del ADNc de sulfamidasa humana (N° de acceso de GenBank U30894).

55 SEC ID N°: 9 es la secuencia de aminoácidos predicha del producto de traducción de ADNc de sulfamidasa humana (SEC ID N°: 8).

SEC ID N°: 10 es la secuencia de nucleótidos del ADNc de N-acetilgalactosamina 6-sulfatasa humana (N° de acceso de GenBank U06088).

60 SEC ID N°: 11 es la secuencia de aminoácidos predicha del producto de traducción de ADNc de N-acetilgalactosamina 6-sulfatasa humana (SEC ID N°: 10).

65 SEC ID N°: 12 es la secuencia de nucleótidos del ADNc de N-acetilglucosamina 6-sulfatasa humana (N° de acceso de GenBank Z12173).

## ES 2 555 056 T3

SEC ID N°: 13 es la secuencia de aminoácidos predicha del producto de traducción de ADNc de N-acetilglucosamina 6-sulfatasa humana (SEC ID N°: 12).

5 SEC ID N°: 14 es la secuencia de nucleótidos del ADNc de arilsulfatasa A humana (N° de acceso de GenBank X52151).

SEC ID N°: 15 es la secuencia de aminoácidos predicha del producto de traducción de ADNc de arilsulfatasa A humana (SEC ID N°: 14).

10 SEC ID N°: 16 es la secuencia de nucleótidos del ADNc de arilsulfatasa B humana (N° de acceso de GenBank J05225).

15 SEC ID N°: 17 es la secuencia de aminoácidos predicha del producto de traducción de ADNc de arilsulfatasa B humana (SEC ID N°: 16).

SEC ID N°: 18 es la secuencia de nucleótidos del ADNc de arilsulfatasa C humana (N° de acceso de GenBank J04964).

20 SEC ID N°: 19 es la secuencia de aminoácidos predicha del producto de traducción de ADNc de arilsulfatasa C humana (SEC ID N°: 18).

SEC ID N°: 20 es la secuencia de nucleótidos del ADNc de arilsulfatasa D humana (N° de acceso de GenBank X83572).

25 SEC ID N°: 21 es la secuencia de aminoácidos predicha del producto de traducción de ADNc de arilsulfatasa D humana (SEC ID N°: 20).

30 SEC ID N°: 22 es la secuencia de nucleótidos del ADNc de arilsulfatasa E humana (N° de acceso de GenBank X83573).

SEC ID N°: 23 es la secuencia de aminoácidos predicha del producto de traducción de ADNc de arilsulfatasa E humana (SEC ID N°: 22).

35 SEC ID N°: 24 es la secuencia de nucleótidos del ADNc de arilsulfatasa F humana (N° de acceso de GenBank X97868).

SEC ID N°: 25 es la secuencia de aminoácidos predicha del producto de traducción de ADNc de arilsulfatasa F humana (SEC ID N°: 24).

40 SEC ID N°: 26 es la secuencia de nucleótidos del ADNc de arilsulfatasa G humana (N° de acceso de GenBank BC012375).

45 SEC ID N°: 27 es la secuencia de aminoácidos predicha del producto de traducción de la arilsulfatasa G humana (SEC ID NO: 26).

SEC ID N°: 28 es la secuencia de nucleótidos del ADNc de HSulf-1 (N° de acceso de GenBank AY101175).

50 SEC ID N°: 29 es la secuencia de aminoácidos predicha del producto de traducción de ADNc de HSulf-1 (SEC ID N°: 28).

SEC ID N°: 30 es la secuencia de nucleótidos del ADNc de HSulf-2 (N° de acceso de GenBank AY101176).

55 SEC ID N°: 31 es la secuencia de aminoácidos predicha del producto de traducción de ADNc de HSulf-2 (SEC ID N°: 30).

SEC ID N°: 32 es el hexapéptido altamente conservado L/V-FGly-X-P-S-R presente en sulfatasas.

60 SEC ID N°: 33 es un sustrato de formación de *FGly* sintético; su secuencia primaria se deriva de arilsulfatasa A humana.

SEC ID N°: 34 es el oligopéptido reorganizado al azar PVSLPTRSCAALLTGR.

SEC ID N°: 35 es el oligopéptido Ser69 PVSLSTPSRAALLTGR.

65 SEC ID N°: 36 es un cebador específico para FGE humana 1199nc.

## ES 2 555 056 T3

- SEC ID N°: 37 es un cebador directo específico para FGE humana 1c.
- 5 SEC ID N°: 38 es un cebador inverso específico para FGE humana 1182c.
- SEC ID N°: 39 es un cebador específico para 5'-FGE humana que contiene un sitio de EcoRI.
- SEC ID N°: 40 es un cebador específico para HA.
- 10 SEC ID N°: 41 es un cebador específico para c-myc.
- SEC ID N°: 42 es un cebador específico para RGS-His<sub>6</sub>.
- SEC ID N°: 43 es un oligopéptido tríptico SQNTPDSSASNLGFR de una preparación de FGE humana.
- 15 SEC ID N°: 44 es un oligopéptido tríptico MVPIPAGVFTMGTDPPQIK de una preparación de FGE humana.
- SEC ID N°: 45 es la secuencia de nucleótidos del parálogo de FGE2 humana (GenBank GI: 24308053).
- 20 SEC ID N°: 46 es la secuencia de aminoácidos predicha del producto de traducción del parálogo de FGE2 humana (SEC ID N°: 45).
- SEC ID N°: 47 es la secuencia de nucleótidos del parálogo de FGE de ratón (GenBank GI: 26344956).
- 25 SEC ID N°: 48 es la secuencia de aminoácidos predicha del producto de traducción del parálogo de FGE de ratón (SEC ID N°: 47).
- SEC ID N°: 49 es la secuencia de nucleótidos del ortólogo de FGE de ratón (GenBank GI: 22122361).
- 30 SEC ID N°: 50 es la secuencia de aminoácidos predicha del producto de traducción del ortólogo de FGE de ratón (SEC ID N°: 49).
- SEC ID N°: 51 es la secuencia de nucleótidos del ortólogo de FGE de mosca de la fruta (GenBank GI: 20130397).
- 35 SEC ID N°: 52 es la secuencia de aminoácidos predicha del producto de traducción del ortólogo de FGE de mosca de la fruta (SEC ID N°: 51).
- SEC ID N°: 53 es la secuencia de nucleótidos del ortólogo de FGE de mosquito (GenBank GI: 21289310).
- 40 SEC ID N°: 54 es la secuencia de aminoácidos predicha del producto de traducción del ortólogo de FGE de mosquito (SEC ID N°: 53).
- SEC ID N°: 55 es la secuencia de nucleótidos del ortólogo de FGE de *S. coelicolor* estrechamente relacionado (GenBank GI: 21225812).
- 45 SEC ID N°: 56 es la secuencia de aminoácidos predicha del producto de traducción del ortólogo de FGE de *S. coelicolor* (SEC ID N°: 55).
- SEC ID N°: 57 es la secuencia de nucleótidos del ortólogo de FGE de *C. efficiens* estrechamente relacionado (GenBank GI: 25028125).
- 50 SEC ID N°: 58 es la secuencia de aminoácidos predicha del producto de traducción del ortólogo de FGE de *C. efficiens* (SEC ID N°: 57).
- 55 SEC ID N°: 59 es la secuencia de nucleótidos del ortólogo de FGE de *N. aromaticivorans* (GenBank GI: 23108562).
- SEC ID N°: 60 es la secuencia de aminoácidos predicha del producto de traducción del ortólogo de FGE de *N. aromaticivorans* (SEC ID N°: 59).
- 60 SEC ID N°: 61 es la secuencia de nucleótidos del ortólogo de FGE de *M. loti* (GenBank GI: 13474559).
- SEC ID N°: 62 es la secuencia de aminoácidos predicha del producto de traducción del ortólogo de FGE de *M. loti* (SEC ID N°: 61).
- 65 SEC ID N°: 63 es la secuencia de nucleótidos del ortólogo de FGE de *B. fungorum* (GenBank GI: 22988809).

## ES 2 555 056 T3

- SEC ID N°: 64 es la secuencia de aminoácidos predicha del producto de traducción del ortólogo de FGE de *B. fungorum* (SEC ID N°: 63).
- 5 SEC ID N°: 65 es la secuencia de nucleótidos del ortólogo de FGE de *S. meliloti* (GenBank GI: 16264068).
- SEC ID N°: 66 es la secuencia de aminoácidos predicha del producto de traducción del ortólogo de FGE de *S. meliloti* (SEC ID N°: 65).
- 10 SEC ID N°: 67 es la secuencia de nucleótidos del ortólogo de FGE de *Microscilla sp.* (GenBank GI: 14518334).
- SEC ID N°: 68 es la secuencia de aminoácidos predicha del producto de traducción del ortólogo de FGE de *Microscilla sp.* (SEC ID N°: 67).
- 15 SEC ID N°: 69 es la secuencia de nucleótidos del ortólogo de FGE de *P. putida* KT2440 (GenBank GI: 26990068).
- SEC ID N°: 70 es la secuencia de aminoácidos predicha del producto de traducción del ortólogo de FGE de *P. putida* KT2440 (SEC ID N°: 69).
- 20 SEC ID N°: 71 es la secuencia de nucleótidos del ortólogo de FGE de *R. metallidurans* (GenBank GI: 22975289).
- SEC ID N°: 72 es la secuencia de aminoácidos predicha del producto de traducción del ortólogo de FGE de *R. metallidurans* (SEC ID N°: 71).
- 25 SEC ID N°: 73 es la secuencia de nucleótidos del ortólogo de FGE de *P. marinus* (GenBank GI: 23132010).
- SEC ID N°: 74 es la secuencia de aminoácidos predicha del producto de traducción del ortólogo de FGE de *P. marinus* (SEC ID N°: 73).
- 30 SEC ID N°: 75 es la secuencia de nucleótidos del ortólogo de FGE de *C. crescentus* CB15 (GenBank GI: 16125425).
- SEC ID N°: 76 es la secuencia de aminoácidos predicha del producto de traducción del ortólogo de FGE de *C. crescentus* CB 15 (SEC ID N°: 75).
- 35 SEC ID N°: 77 es la secuencia de nucleótidos del ortólogo de FGE de *M. tuberculosis* Ht37Rv (GenBank GI: 15607852).
- SEC ID N°: 78 es la secuencia de aminoácidos predicha del producto de traducción del ortólogo de FGE de *M. tuberculosis* Ht37Rv (SEC ID N°: 77).
- 40 SEC ID N°: 79 es el heptapéptido altamente conservado presente en el subdominio 3 de parálogos y ortólogos de FGE.
- SEC ID N°: 80 es la secuencia de nucleótidos del fragmento EST ortólogo de FGE que tiene N° de acceso de GenBank: CA379852.
- 45 SEC ID N°: 81 es la secuencia de nucleótidos del fragmento EST ortólogo de FGE que tiene N° de acceso de GenBank: AI721440.
- SEC ID N°: 82 es la secuencia de nucleótidos del fragmento EST ortólogo de FGE que tiene N° de acceso de GenBank: BJ505402.
- 50 SEC ID N°: 83 es la secuencia de nucleótidos del fragmento EST ortólogo de FGE que tiene N° de acceso de GenBank: BJ054666.
- 55 SEC ID N°: 84 es la secuencia de nucleótidos del fragmento EST ortólogo de FGE que tiene N° de acceso de GenBank: AL892419.
- SEC ID N°: 85 es la secuencia de nucleótidos del fragmento EST ortólogo de FGE que tiene N° de acceso de GenBank: CA064079.
- 60 SEC ID N°: 86 es la secuencia de nucleótidos del fragmento EST ortólogo de FGE que tiene N° de acceso de GenBank: BF189614.
- SEC ID N°: 87 es la secuencia de nucleótidos del fragmento EST ortólogo de FGE que tiene N° de acceso de GenBank: AV609121.
- 65

SEC ID N°: 88 es la secuencia de nucleótidos del ADNc de HSulf-3.

SEC ID N°: 89 es la secuencia de aminoácidos predicha del producto de traducción de ADNc de HSulf-3 (SEC ID N°: 88).

5 SEC ID N°: 90 es la secuencia de nucleótidos del ADNc de HSulf-4.

SEC ID N°: 91 es la secuencia de aminoácidos predicha del producto de traducción de ADNc de HSulf-4 (SEC ID N°: 90).

10 SEC ID N°: 92 es la secuencia de nucleótidos del ADNc de HSulf-5.

SEC ID N°: 93 es la secuencia de aminoácidos predicha del producto de traducción de ADNc de HSulf-5 (SEC ID N°: 92).

15 SEC ID N°: 94 es la secuencia de nucleótidos del ADNc de HSulf-6.

SEC ID N°: 95 es la secuencia de aminoácidos predicha del producto de traducción de ADNc de HSulf-6 (SEC ID N°: 94).

## 20 **Breve descripción de los dibujos**

**Fig. 1:** Un esquema de espectro de masas MALDI-TOF de P23 tras la incubación en ausencia (A) o presencia (B) de un extracto soluble de microsomas de testículos bovinos.

25 **Fig. 2:** Un árbol filogenético derivado de un alineamiento de FGE humana y 21 proteínas de la semilla DUF323 en PFAM.

**Fig. 3:** Organización del locus del gen de FGE humana y murina. Los exones se muestran a escala como recuadros y recuadros claros (locus murino). Los números encima de las líneas de intrón indican el tamaño de los intrones en kilobases.

**Fig. 4:** Diagrama que muestra un mapa de plásmido pXMG.1.3 de expresión de FGE.

35 **Fig. 5:** Gráfico de barras que representa actividad N-acetilgalactosamina 6-sulfatasa en células 36F transfectadas de manera transitoria con plásmido de expresión de FGE.

**Fig. 6:** Gráfico de barras que representa actividad específica N-acetilgalactosamina 6-sulfatasa en células 36F transfectadas de manera transitoria con plásmido de expresión de FGE.

40 **Fig. 7:** Gráfico de barras que representa la producción de N-acetilgalactosamina 6-sulfatasa en células 36F transfectadas de manera transitoria con plásmido de expresión de FGE.

**Fig. 8:** Gráfico que representa actividad iduronato 2-sulfatasa en células 3006 transfectadas de manera transitoria con plásmido de expresión de FGE.

45 **Fig. 9:** Representa un kit que incluye características de la presente invención.

## **Descripción detallada de la invención**

50 La invención se explica en las reivindicaciones 1 a 15. Se ha descubierto el gen que codifica enzima generadora de formilglicina (FGE), una enzima responsable de la modificación postraduccional única que se produce en sulfatasas que es esencial para la función de sulfatasa: la formación de L-C<sub>α</sub>-formilglicina (también conocida como *FGly* y/o ácido 2-amino-3-oxopropanoico). Se ha descubierto, inesperadamente, que mutaciones en el gen de FGE conducen al desarrollo de deficiencia múltiple de sulfatasas (MSD) en sujetos. También se ha descubierto, inesperadamente,

55 que la FGE potencia la actividad de sulfatasas, incluyendo, pero sin limitarse a, iduronato 2-sulfatasa, sulfamidasa, N-acetilgalactosamina 6-sulfatasa, N-acetilglucosamina 6-sulfatasa, arilsulfatasa A, arilsulfatasa B, arilsulfatasa C, arilsulfatasa D, arilsulfatasa E, arilsulfatasa F, arilsulfatasa G, HSulf-1, HSulf-2, HSulf-3, HSulf-4, HSulf-5 y HSulf-6, y sulfatasas descritas en las solicitudes provisionales de EE.UU. con números de publicación 20030073118, 20030147875, 20030148920, 20030162279 y 20030166283.

60 “Actividad de generación de C<sub>α</sub>-formilglicina” se refiere a la capacidad de una molécula para formar, o potenciar, la formación de *FGly* en un sustrato. El sustrato puede ser una sulfatasa como se describe en otra parte en el presente documento, o un oligopéptido sintético (véase, por ejemplo, SEC ID N°: 33, y los ejemplos). El sustrato contiene preferiblemente el hexapéptido conservado de SEC ID N°: 32 [L/V-C(S)-X-P-S-R]. Métodos de ensayo de la

65 formación de *FGly* son como se describen en la técnica (véase, por ejemplo, Dierks, T., et al., Proc. Natl. Acad. Sci. U.S.A., 1997, 94:11963-11968), y en otra parte en el presente documento (véanse, por ejemplo, los ejemplos). Una

“molécula”, como se usa en el presente documento, engloba tanto “ácidos nucleicos” como “polipéptidos”. Las moléculas de FGE son capaces de formar, o potenciar/aumentar, la formación de *FGly* tanto *in vivo* como *in vitro*.

La “potenciación (o “aumento”)” de la actividad de generación de  $C_{\alpha}$ -formilglicina, como se usa en el presente documento, se refiere normalmente a una expresión aumentada de FGE y/o su polipéptido codificado. La expresión aumentada se refiere a aumentar (es decir, hasta un grado detectable) la replicación, transcripción y/o traducción de cualquiera de los ácidos nucleicos (ácidos nucleicos de FGE como se describen en otra parte en el presente documento), ya que la regulación por incremento de cualquiera de estos procesos produce un aumento de la concentración/cantidad del polipéptido codificado por el gen (ácido nucleico). La potenciación (o aumento) de la actividad de generación de  $C_{\alpha}$ -formilglicina también se refiere a prevenir o inhibir la degradación de FGE (por ejemplo, mediante el aumento de la ubiquitinización), regulación por disminución, etc., produciendo, por ejemplo, una  $t_{1/2}$  (semivida) de molécula de FGE estable o aumentada en comparación con un control. La regulación por disminución o expresión reducida se refiere a la expresión reducida de un gen y/o su polipéptido codificado. La regulación por incremento o regulación por disminución de la expresión génica puede determinarse directamente detectando un aumento o disminución, respectivamente, en el nivel de ARNm para el gen (por ejemplo, FGE), o el nivel de expresión de proteína del polipéptido codificado por el gen, usando cualquier medio adecuado conocido en la técnica, tal como hibridación de ácido nucleico o métodos de detección con anticuerpos, respectivamente, y en comparación con controles. La regulación por incremento o regulación por disminución de la expresión del gen de FGE también puede determinarse indirectamente detectando un cambio en la actividad de generación de  $C_{\alpha}$ -formilglicina.

“Expresión”, como se usa en el presente documento, se refiere a expresión de ácido nucleico y/o polipéptido, así como a actividad de la molécula de polipéptido (por ejemplo, actividad de generación de  $C_{\alpha}$ -formilglicina de la molécula).

Como se usa en el presente documento, un sujeto es un mamífero o un mamífero no humano. En todas las realizaciones se prefieren FGE humana y sujetos humanos.

Como se usa en el presente documento con respecto a ácidos nucleicos, el término “aislado” significa: (i) amplificado *in vitro* mediante, por ejemplo, reacción en cadena de la polimerasa (PCR); (ii) producido de manera recombinante mediante clonación; (iii) purificado, como mediante escisión y separación en gel; o (iv) sintetizado mediante, por ejemplo, síntesis química. Un ácido nucleico aislado es uno que se manipula fácilmente mediante técnicas de ADN recombinante bien conocidas en la técnica. Por tanto, una secuencia de nucleótidos contenida en un vector en el que se conocen sitios de restricción en 5' y 3' o para la que se han divulgado secuencias de cebadores para reacción en cadena de la polimerasa (PCR) se considera aislada, pero una secuencia de ácido nucleico que existe en su estado nativo en su huésped natural no. Un ácido nucleico aislado puede estar sustancialmente purificado, pero no necesita estarlo. Por ejemplo, un ácido nucleico que está aislado dentro de un vector de clonación o expresión no es puro porque puede comprender solo un diminuto porcentaje del material en la célula en la que reside. Sin embargo, un ácido nucleico de este tipo está aislado, como se usa el término en el presente documento, porque se manipula fácilmente mediante técnicas convencionales conocidas por los expertos en la materia.

Como se usa en el presente documento con respecto a polipéptidos, el término “aislado” significa separado de su entorno nativo en una forma suficientemente pura de modo que puede manipularse o usarse para uno cualquiera de los fines de la invención descrita en el presente documento. Por tanto, aislado significa suficientemente puro para usarse (i) para preparar y/o aislar anticuerpos, (ii) como reactivo en un ensayo, (iii) para secuenciar, (iv) como producto terapéutico, etc.

Según la invención, moléculas de ácido nucleico aisladas que codifican un polipéptido de FGE que tiene actividad de generación de  $C_{\alpha}$ -formilglicina incluyen: (a) moléculas de ácido nucleico que se hibridan en condiciones rigurosas con una molécula que consiste en un ácido nucleico de SEC ID N°: 1 y que codifican un polipéptido de FGE que tiene actividad de generación de  $C_{\alpha}$ -formilglicina, (b) deleciones, adiciones y sustituciones de (a) que codifican un polipéptido de FGE respectivo que tiene actividad de generación de  $C_{\alpha}$ -formilglicina, (c) moléculas de ácido nucleico que se diferencian de las moléculas de ácido nucleico de (a) o (b) en secuencia de codón debido a la degeneración del código genético, y (d) complementos de (a), (b) o (c). “Complementos”, como se usa en el presente documento, incluye “cadenas complementarias de longitud completa o cadenas complementarias al 100 % de (a), (b) o (c)”.

Pueden proporcionarse homólogos y alelos de los ácidos nucleicos de FGE descritos en el presente documento que también tienen actividad de generación de  $C_{\alpha}$ -formilglicina. Los homólogos, como se describen en el presente documento, incluyen las moléculas identificadas en otra parte en el presente documento (véanse, por ejemplo, SEC ID N°: 4, 5, 45-78 y 80-87), es decir, ortólogos y parálogos. Pueden identificarse homólogos adicionales siguiendo las enseñanzas de la presente invención, así como mediante técnicas convencionales. Como los homólogos de FGE descritos en el presente documento comparten todos actividad de generación de  $C_{\alpha}$ -formilglicina, pueden usarse de manera intercambiable con la molécula de FGE humana en todos los aspectos de la invención.

En general, los homólogos y alelos compartirán normalmente al menos el 40 % de identidad de nucleótidos y/o al menos el 50 % de identidad de aminoácidos con SEC ID N°: 1 y SEC ID N°: 2, respectivamente, en algunos casos



compartirán al menos el 50 % de identidad de nucleótidos y/o al menos el 65 % de identidad de aminoácidos y en todavía otros casos compartirán al menos el 60 % de identidad de nucleótidos y/o al menos el 75 % de identidad de aminoácidos. En casos adicionales, los homólogos y alelos compartirán normalmente al menos el 90 %, el 95 %, o incluso el 99 % de identidad de nucleótidos y/o al menos el 95 %, el 98 %, o incluso el 99 % de identidad de aminoácidos con SEC ID N°: 1 y SEC ID N°: 2, respectivamente. La homología puede calcularse usando diversas herramientas de software disponibles para el público desarrolladas por NCBI (Bethesda, Maryland). Herramientas a modo de ejemplo incluyen el algoritmo heurístico de Altschul SF, et al., (J Mol Biol, 1990, 215: 403-410), también conocido como BLAST. Pueden obtenerse alineamientos por emparejamiento y de ClustalW (configuración de matriz BLOSUM30), así como análisis hidropático de Kyte-Doolittle usando herramientas públicas (EMBL, Heidelberg, Alemania) y comerciales (por ejemplo, el software de análisis de secuencias MacVector de Oxford Molecular Group/enetics Computer Group, Madison, WI). También pueden usarse los complementos de Watson-Crick de los ácidos nucleicos anteriores.

En el cribado de genes relacionados con FGE, tales como homólogos y alelos de FGE, puede realizarse una transferencia Southern usando las condiciones anteriores, junto con una sonda radiactiva. Después de lavar la membrana a la que se transfiere finalmente el ADN, puede colocarse la membrana contra una película de rayos X o una placa de sistema de detección y cuantificación de la radiactividad para detectar la señal radiactiva.

Dadas las enseñanzas en el presente documento de un clon de ADNc de FGE humana de longitud completa, pueden aislarse otras secuencias de mamífero tales como el clon de ADNc de ratón correspondiente al gen de FGE humana a partir de una biblioteca de ADNc, usando técnicas de hibridación de colonias convencionales.

Pueden proporcionarse ácidos nucleicos degenerados que incluyen codones alternativos a los presentes en los materiales nativos. Por ejemplo, se codifican residuos de serina por los codones TCA, AGT, TCC, TCG, TCT y AGC. Por tanto, será evidente para un experto en la materia que puede emplearse cualquiera de los tripletes de nucleótidos que codifican serina para dirigir el aparato de síntesis de proteínas, *in vitro* o *in vivo*, para incorporar un residuo de serina en un polipéptido de FGE que está alargándose. Similarmente, tripletes de secuencias de nucleótidos que codifican otros residuos de aminoácido incluyen, pero no se limitan a: CCA, CCC, CCG y CCT (codones de prolina); CGA, CGC, CGG, CGT, AGA y AGG (codones de arginina); ACA, ACC, ACG y ACT (codones de treonina); AAC y AAT (codones de asparagina); y ATA, ATC y ATT (codones de isoleucina). Otros residuos de aminoácido pueden codificarse similarmente por múltiples secuencias de nucleótidos. Por tanto, pueden usarse ácidos nucleicos que se diferencian de los ácidos nucleicos biológicamente aislados en la secuencia de codones debido a la degeneración del código genético.

Pueden aislarse fragmentos únicos de SEC ID N°: 1 o SEC ID N°: 3 o complementos de los mismos. Un fragmento único es uno que es una "firma" para el ácido nucleico más grande. Por ejemplo, el fragmento único es lo bastante largo para garantizar que su secuencia precisa no se encuentra en moléculas dentro del genoma humano fuera de los ácidos nucleicos de FGE definidos anteriormente (y alelos humanos). Los expertos en la materia pueden aplicar procedimientos simplemente de rutina para determinar si un fragmento es único dentro del genoma humano. Sin embargo, los fragmentos únicos excluyen fragmentos compuestos completamente de las secuencias de nucleótidos seleccionadas del grupo que consiste en SEC ID N°: 4, y/u otras secuencias previamente publicadas a partir de la fecha de presentación de la presente solicitud.

Un fragmento que está completamente compuesto de la secuencia descrita en los depósitos de GenBank anteriores es una que no incluye ninguno de los nucleótidos únicos para las secuencias de la invención. Por tanto, un fragmento único según la invención debe contener una secuencia de nucleótidos distinta de la secuencia exacta de esos en los depósitos de GenBank o fragmentos de la misma. La diferencia puede ser una adición, delección o sustitución con respecto a la secuencia de GenBank o puede ser una secuencia completamente separada de la secuencia de GenBank.

Pueden usarse fragmentos únicos como sondas en ensayos de transferencia Southern y Northern para identificar tales ácidos nucleicos, o pueden usarse en ensayos de amplificación tales como los que emplean PCR. Como saben los expertos en la materia, se prefieren sondas grandes tales como 200, 250, 300 o más nucleótidos para determinados usos tales como transferencias Southern y Northern, mientras que se preferirán fragmentos más pequeños para usos tales como PCR. También pueden usarse fragmentos únicos para producir proteínas de fusión para generar anticuerpos o determinar la unión de los fragmentos de polipéptido, como se demuestra en los ejemplos, o para generar componentes de inmunoensayo. Asimismo, pueden emplearse fragmentos únicos para producir fragmentos no fusionados de los polipéptidos de FGE, útiles, por ejemplo, en la preparación de anticuerpos, inmunoensayos o aplicaciones terapéuticas. Pueden usarse adicionalmente fragmentos únicos como moléculas antisentido para inhibir la expresión de ácidos nucleicos y polipéptidos de FGE respectivamente.

Como se reconocerá por aquellos expertos en la materia, el tamaño del fragmento único dependerá de su conservación en el código genético. Por tanto, algunas regiones de SEC ID N°: 1 o SEC ID N°: 3 y complementos requerirán que segmentos más largos sean únicos, mientras que otras solo requerirán segmentos cortos, normalmente de entre 12 y 32 nucleótidos de longitud (por ejemplo de 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 divulgada. Como se ha

mencionado anteriormente, la presente divulgación pretende englobar todos y cada uno de los fragmentos de cada secuencia, comenzando en el imprimación nucleótido, el segundo nucleótido, etcétera, hasta 8 nucleótidos del final, y terminando en cualquier lugar desde el nucleótido número 8, 9, 10, etcétera, para cada secuencia, hasta el último nucleótido (siempre que la secuencia sea única como se ha descrito anteriormente). Prácticamente cualquier segmento de la región de SEC ID N°: 1 comenzando en el nucleótido 1 y terminando en el nucleótido 1180, o SEC ID N°: 3 comenzando en el nucleótido 1 y terminando en el nucleótido 1122, o complementos de los mismos, que tenga 20 o más nucleótidos de longitud será único. Los expertos en la materia están bien versados en los métodos para seleccionar tales secuencias, normalmente basándose en la capacidad del fragmento único para distinguir selectivamente la secuencia de interés de otras secuencias en el genoma humano del fragmento. Normalmente, para aquellas en bases de datos conocidas es todo lo que se necesita, aunque pueden realizarse hibridación confirmatoria *in vitro* y análisis de secuenciación.

“Vectores”, como se usa en el presente documento, puede ser cualquiera de varios ácidos nucleicos en los que puede insertarse una secuencia deseada mediante restricción y ligación para su transporte entre diferentes entornos genéticos o para su expresión en una célula huésped. Los vectores están compuestos normalmente de ADN, aunque también están disponibles vectores de ARN. Los vectores incluyen, pero no se limitan a, plásmidos, fagémidos y genomas de virus. Un vector de clonación es uno que es capaz de replicarse en una célula huésped, y que se caracteriza además por uno o más sitios de restricción de endonucleasas en los que puede cortarse el vector de una manera determinable y en los que puede ligarse una secuencia de ADN deseada de tal manera que el nuevo vector recombinante conserva su capacidad para replicarse en la célula huésped. En el caso de plásmidos, puede producirse la replicación de la secuencia deseada muchas veces ya que el plásmido aumenta el número de copias dentro de la bacteria huésped o solo una única vez por huésped antes de que el huésped se reproduzca por mitosis. En el caso de fago, la replicación puede producirse de manera activa durante una fase lítica o de manera pasiva durante una fase lisogénica. Un “vector de expresión” es uno en el que puede insertarse una secuencia de ADN deseada (por ejemplo, el ADNc de FGE de SEC ID N°: 3) mediante restricción y ligación de tal manera que está operativamente unido a secuencias reguladoras y puede expresarse como un transcrito de ARN. Los vectores pueden contener además una o más secuencias marcadoras adecuadas para su uso en la identificación de células que se han transformado o transfectado o no con el vector. Los marcadores incluyen, por ejemplo, genes que codifican proteínas que aumentan o disminuyen o tanto la resistencia como la sensibilidad a antibióticos u otros compuestos, genes que codifican enzimas cuyas actividades son detectables por ensayos convencionales conocidos en la técnica (por ejemplo,  $\beta$ -galactosidasa  $\beta$ -galactosidasa o fosfatasa alcalina), y genes que afectan visiblemente el fenotipo de células, huéspedes, colonias o placas transformadas o transfectadas (por ejemplo, proteína verde fluorescente).

Un “vector de direccionamiento” es uno que normalmente contiene construcciones/secuencias de direccionamiento que se usan, por ejemplo, para insertar una secuencia reguladora dentro de un gen endógeno (por ejemplo, dentro de las secuencias de un exón y/o intrón), dentro de las secuencias promotoras del gen endógeno, o en la dirección 5' de las secuencias promotoras del gen endógeno. En otro ejemplo, un vector de direccionamiento puede contener el gen de interés (por ejemplo, codificado por el ADNc de SEC ID N°: 1) y otras secuencias necesarias para el direccionamiento del gen a una localización preferida en el genoma (por ejemplo, una localización transcripcionalmente activa, por ejemplo, en la dirección 3' de un promotor endógeno de un gen no relacionado). La construcción de vectores y construcciones de direccionamiento se describe en detalle en las patentes de EE.UU. 5.641.670 y 6.270.989.

Pueden transformarse células procariotas o eucariotas con ADN o ARN heterólogo y que puede hacerse crecer o mantenerse en cultivo. Ejemplos incluyen células bacterianas tales como *Escherichia coli*, células de insecto y células de mamífero tales como ser humano, ratón, hámster, cerdo, cabra, primate, etc. Pueden ser cepas de células primarias o secundarias (que muestran un número finito de duplicaciones de población media en cultivo y no están inmortalizadas) y líneas celulares inmortalizadas (que muestran un periodo de vida aparentemente ilimitado en cultivo). Las células primarias y secundarias incluyen, por ejemplo, fibroblastos, queratinocitos, células epiteliales (por ejemplo, células epiteliales de mama, células epiteliales intestinales), células endoteliales, células de la glía, células neurales, elementos formados de la sangre (por ejemplo, linfocitos, células de la médula ósea), células de músculo y precursores de esos tipos de células somáticas incluyendo células madre embrionarias. Cuando las células van a usarse en terapia génica, se obtienen preferiblemente células primarias del individuo al que se administran las células manipuladas. Sin embargo, pueden obtenerse células primarias de un donante (distinto del receptor) de la misma especie. Ejemplos de líneas de células humanas inmortalizadas que pueden usarse con las construcciones de ADN y métodos de la presente invención incluyen, pero no se limitan a, células HT-1080 (ATCC CCL 121), células HeLa y derivados de células HeLa (ATCC CCL 2, 2.1 y 2.2), células de cáncer de mama MCF-7 (ATCC BTH 22), células de leucemia K-562 (ATCC CCL 243), células de carcinoma KB (ATCC CCL 17), células de carcinoma de ovario 2780AD (Van der Blick, A. M. at al., Cancer Res, 48:5927-5932 (1988), células Raji (ATCC CCL 86), células de adenocarcinoma de colon WiDr (ATCC CCL 218), células de adenocarcinoma de colon SW620 (ATCC CCL 227), células Jurkat (ATCC TIB 152), células Namalwa (ATCC CRL1432), células HL-60 (ATCC CCL 240), células Daudi (ATCC CCL 213), células RPMI 8226 (ATCC CCL 155), células U-937 (ATCC CRL 1593), células de melanoma de Bowes (ATCC CRL 9607), células 2R4 de la sublínea WI-38VA13 (ATCC CLL 75.1) y células MOLT-4 (ATCC CRL 1582), células CHO y células COS, así como células de heterohibridoma producidas por la fusión de células humanas y células de otra especie. También pueden usarse cepas de fibroblastos humanos

secundarios, tales como WI-38 (ATCC CCL 75) y MRC-5 (ATCC CCL 171). Se proporciona discusión adicional de los tipos de células que pueden usarse en las patentes de EE.UU. 5.641.670 y 6.270.989. También pueden usarse sistemas de transcripción libres de células en lugar de células.

5 Las células se mantienen en condiciones, como se conocen en la técnica, que producen la expresión de la proteína FGE o fragmentos funcionales de la misma. Las proteínas expresadas usando los métodos descritos pueden purificarse a partir de lisados celulares o sobrenadantes celulares. Pueden prepararse proteínas preparadas según este método como una formulación farmacéuticamente útil y administrarse a un ser humano o animal no humano por vías farmacéuticas convencionales como se conoce en la técnica (por ejemplo, oral, intravenosa, intramuscular, intranasal, intratraqueal o subcutánea). Como se describe en otra parte en el presente documento, las células recombinantes pueden ser células inmortalizadas, primarias o secundarias, preferiblemente humanas. El uso de células de otra especie puede ser deseable en casos en los que las células no humanas sean ventajosas para fines de producción de proteína en los que la FGE no humana producida es terapéuticamente útil.

15 Como se usa en el presente documento, se dice que una secuencia codificante y secuencias reguladoras están "operativamente" unidas cuando están covalentemente enlazadas de tal manera que se coloca la expresión o transcripción de la secuencia codificante bajo la influencia o el control de las secuencias reguladoras. Si se desea que las secuencias codificantes se traduzcan en una proteína funcional, se dice que dos secuencias de ADN están operativamente unidas si la inducción de un promotor en las secuencias reguladoras en 5' produce la transcripción de la secuencia codificante y si la naturaleza del enlace entre las dos secuencias de ADN no (1) produce la introducción de una mutación del marco de lectura, (2) interfiere con la capacidad de la región promotora para dirigir la transcripción de las secuencias codificantes, o (3) interfiere con la capacidad del transcrito de ARN correspondiente para traducirse en una proteína. Por tanto, una región promotora estaría operativamente enlazada a una secuencia codificante si la región promotora fuera capaz de efectuar la transcripción de esa secuencia de ADN de tal manera que el transcrito resultante pudiera traducirse en la proteína o polipéptido deseado.

La naturaleza precisa de las secuencias reguladoras necesarias para la expresión génica puede variar entre tipos de especies o células, pero en general incluirán, según sea necesario, secuencias no transcritas en 5' y no traducidas en 5' que participan en la iniciación de la transcripción y la traducción respectivamente, tal como una caja TATA, secuencia de ocupación, secuencia CAAT, y similares. Especialmente, tales secuencias reguladoras no transcritas en 5' incluirán una región promotora que incluye una secuencia promotora para el control transcripcional del gen operativamente unido. Las secuencias reguladoras también pueden incluir secuencias potenciadoras o secuencias activadoras en la dirección 5' según se desee. Los vectores de la invención pueden incluir opcionalmente secuencias señal o líder en 5'. La elección y el diseño de un vector apropiado están dentro de la capacidad y discreción de un experto habitual en la materia.

40 Vectores de expresión que contienen todos los elementos necesarios para la expresión están comercialmente disponibles y son conocidos por los expertos en la materia. Véase, por ejemplo, Sambrook et al., *Molecular Cloning: A Laboratory Manual*, segunda edición, Cold Spring Harbor Laboratory Press, 1989. Se diseñan mediante ingeniería genética células mediante la introducción en las células de ADN (ARN) heterólogo que codifica polipéptido de FGE o fragmento o VARIANTEe del mismo. Ese ADN (ARN) heterólogo se coloca bajo el control operativo de elementos transcripcionales para permitir la expresión del ADN heterólogo en la célula huésped.

45 Sistemas preferidos para la expresión de ARNm en células de mamífero son aquellos tales como pRc/CMV (disponible de Invitrogen, Carlsbad, CA) que contienen un marcador de selección tal como un gen que confiere resistencia a G418 (que facilita la selección de líneas celulares establemente transfectadas) y las secuencias de potenciador-promotor del citomegalovirus (CMV) humano. Adicionalmente, es adecuado para la expresión en líneas celulares de primate o canino el vector pCEP4 (Invitrogen, Carlsbad, CA), que contiene un origen de replicación de virus de Epstein Barr (EBV), que facilita el mantenimiento del plásmido como elemento extracromosómico de múltiples copias. Otro vector de expresión es el plásmido pEF-BOS que contiene el promotor de factor de elongación 1 $\alpha$  de polipéptido, que estimula eficazmente la transcripción *in vitro*. El plásmido se describe por Mishizuma y Nagata (*Nuc. Acids Res.* 18:5322, 1990), y su uso en experimentos de transfección se divulga, por ejemplo, por Demoulin (*Mol. Cell. Biol.* 16:4710-4716, 1996). Todavía otro vector de expresión preferido es un adenovirus, descrito por Stratford-Perricaudet, que es defectuoso para proteínas E1 y E3 (*J. Clin. Invest.* 90:626-630, 1992). El uso del adenovirus como recombinante Adeno.P1 A se divulga por Warnier et al., en la inyección intradérmica en ratones para inmunización frente a P1A (*Int. J. Cancer*, 67:303-310, 1996).

60 Los kits de expresión permiten al experto en la materia preparar un vector o vectores de expresión deseados. Tales kits de expresión incluyen al menos porciones separadas de cada una de las secuencias codificantes anteriormente tratadas. Pueden añadirse otros componentes, según se desee, siempre que incluyan las secuencias previamente mencionadas, que se requieren.

65 Pueden usarse vectores de expresión que contienen la secuencia de ADNc de FGE descrita anteriormente para transfectar líneas celulares y células huésped, ya sean procariontas (por ejemplo, *Escherichia coli*) o eucariotas (por ejemplo, células CHO, células COS, sistemas de expresión en levadura y expresión de baculovirus recombinante en células de insecto). Son especialmente útiles células de mamífero tales como ser humano, ratón, hámster, cerdo,

cabra, primate, etc. Pueden ser de una amplia variedad de tipos de tejido, e incluir células primarias y líneas celulares inmortalizadas como se describe en otra parte en el presente documento. Ejemplos específicos incluyen células HT-1080, células CHO, células dendríticas, células U293, leucocitos de sangre periférica, células madre de médula ósea, células madre embrionarias y células de insecto.

5 Pueden proporcionarse polipéptidos aislados (incluyendo proteínas completas y proteínas parciales), codificados por los ácidos nucleicos de FGE anteriores, e incluir el polipéptido de SEC ID N°: 2 y fragmentos únicos del mismo. Tales polipéptidos son útiles, por ejemplo, solos o como parte de proteínas de fusión para generar anticuerpos, como componentes de un inmunoensayo, etc. Pueden aislarse polipéptidos a partir de muestras biológicas que incluyen  
10 homogeneizados tisulares o celulares, y también pueden expresarse de manera recombinante en una variedad de sistemas de expresión procariotas y eucariotas construyendo un vector de expresión apropiado para el sistema de expresión, introduciendo el vector de expresión en el sistema de expresión y aislando la proteína expresada de manera recombinante. También pueden sintetizarse químicamente polipéptidos cortos, incluyendo péptidos  
15 antigénicos (tal como se presentan por moléculas del MHC sobre la superficie de una célula para su reconocimiento inmunitario) usando métodos bien establecidos de síntesis de péptido.

Un fragmento único de un polipéptido de FGE, en general, tiene los rasgos y las características de fragmentos  
20 únicos tal como se trató anteriormente en relación con ácidos nucleicos. Como se reconocerá por aquellos expertos en la materia, el tamaño del fragmento único dependerá de factores tales como si el fragmento constituyera una parte de un dominio de proteína conservado. Por tanto, algunas regiones de SEC ID N°: 2 requerirán segmentos más largos para ser únicas, mientras que otras solo requerirán segmentos cortos, normalmente de entre 5 y 12 aminoácidos (por ejemplo 5, 6, 7, 8, 9, 10, 11 y 12 aminoácidos de longitud o más, incluyendo cada número entero hasta la longitud completa, 287 aminoácidos de longitud).

25 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 interacción con anticuerpos, interacción con otros polipéptidos o fragmentos de los mismos, interacción con otras moléculas, etc. Una actividad importante es la capacidad para actuar como una firma para  
30 identificar el polipéptido. Aquellos expertos en la materia están bien versados en métodos de selección de secuencias de aminoácidos únicas, normalmente basándose en la capacidad del fragmento único para distinguir selectivamente la secuencia de interés de otras que miembros no de la familia. Una comparación de la secuencia del fragmento con las de bases de datos conocidas es normalmente todo lo que se necesita.

Pueden proporcionarse VARIANTEes de los polipéptidos de FGE descritos anteriormente. Como se usa en el  
35 presente documento, una "VARIANTEe" de un polipéptido de FGE es un polipéptido que contiene una o más modificaciones a la secuencia de aminoácidos primaria de un polipéptido de FGE. Las modificaciones que crean una VARIANTEe de polipéptido de FGE se hacen normalmente al ácido nucleico que codifica el polipéptido de FGE, y pueden incluir delecciones, mutaciones puntuales, truncaciones, sustituciones de aminoácidos y adición de aminoácidos o restos distintos de aminoácido para: 1) reducir o eliminar una actividad de un polipéptido de FGE; 2)  
40 potenciar una propiedad de un polipéptido de FGE, tal como estabilidad de la proteína en un sistema de expresión o la estabilidad de la unión proteína-ligando; 3) proporcionar una actividad o propiedad novedosa a un polipéptido de FGE, tal como adición de un epítipo antigénico o adición de un resto detectable; o 4) proporcionar una unión equivalente o mejor a un receptor de polipéptido de FGE u otra molécula. Alternativamente, pueden hacerse modificaciones directamente al polipéptido, tal como por escisión, adición de una molécula conectora, adición de un  
45 resto detectable, tal como biotina, adición de un ácido graso, y similares. Las modificaciones también engloban proteínas de fusión que comprenden la totalidad o parte de la secuencia de aminoácidos de FGE. Un experto en la materia estará familiarizado con métodos de predicción del efecto sobre la conformación de la proteína de un cambio en la secuencia de la proteína y, por tanto, puede "diseñar" un polipéptido de FGE de VARIANTEe según métodos conocidos. Un ejemplo de un método tal se describe por Dahiyat y Mayo en Science 278:82-87, 1997, mediante el  
50 cual pueden diseñarse proteínas *de novo*. El método puede aplicarse a una proteína conocida para variar solo una porción de la secuencia del polipéptido. Aplicando los métodos computacionales de Dahiyat y Mayo, pueden proponerse VARIANTEes específicas del polipéptido de FGE y probarse para determinar si la VARIANTEe retiene una conformación deseada.

55 Las VARIANTEes pueden incluir polipéptidos de FGE que se modifican, específicamente para alterar una característica del polipéptido no relacionada con su actividad fisiológica. Por ejemplo, pueden sustituirse o deleccionarse residuos de cisteína para prevenir enlaces disulfuro no deseados. Similarmente, pueden cambiarse determinados aminoácidos para potenciar la expresión de un polipéptido de FGE eliminando la proteólisis mediante  
60 proteasas en un sistema de expresión (por ejemplo, residuos de aminoácido dibásico en sistemas de expresión de levadura en los que está presente actividad de proteasa KEX2).

Las mutaciones de un ácido nucleico que codifica un polipéptido de FGE conservan preferiblemente el marco de  
lectura de aminoácidos de la secuencia codificante, y preferiblemente no crean regiones en el ácido nucleico que es probable que se hibriden para formar estructuras secundarias, tales como horquillas o bucles, que pueden ser  
65 perjudiciales para la expresión del polipéptido VARIANTEe.

Pueden hacerse mutaciones seleccionando una sustitución de aminoácidos, o mediante mutagénesis al azar de un sitio seleccionado en un ácido nucleico que codifica el polipéptido. Entonces se expresan polipéptidos VARIANTEes y se prueban para una o más actividades para determinar qué mutación proporciona un polipéptido VARIANTEe con las propiedades deseadas. Pueden hacerse mutaciones adicionales a VARIANTEes (o en polipéptidos de FGE no VARIANTEes) que son silenciosas en cuanto a la secuencia de aminoácidos del polipéptido, pero que proporcionan codones preferidos para la traducción en un huésped particular, o alteran la estructura del ARNm para, por ejemplo, potenciar la estabilidad y/o expresión. Los codones preferidos para la traducción de un ácido nucleico en, por ejemplo, *Escherichia coli*, células de mamífero, etc., son bien conocidos para aquellos expertos en la materia. Pueden hacerse todavía otras mutaciones a las secuencias no codificantes de un gen de FGE o clon de ADNc para potenciar la expresión del polipéptido.

El experto en la materia sabrá que pueden hacerse sustituciones de aminoácidos conservativas en polipéptidos de FGE para proporcionar VARIANTEes funcionalmente equivalentes de los polipéptidos anteriores, es decir, las VARIANTEes retienen las capacidades funcionales de los polipéptidos de FGE. Como se usa en el presente documento, una "sustitución de aminoácidos conservativa" se refiere a una sustitución de aminoácidos que no altera significativamente la estructura terciaria y/o actividad del polipéptido. Pueden prepararse VARIANTEes según métodos de alteración de la secuencia del polipéptido conocidos para un experto habitual en la materia, e incluyen los que se encuentran en referencias que compilan tales métodos, por ejemplo Molecular Cloning: A Laboratory Manual, J. Sambrook, *et al.*, eds., segunda edición, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, Nueva York, 1989, o Current Protocols in Molecular Biology, F.M. Ausubel, *et al.*, eds., John Wiley & Sons, Inc., Nueva York. VARIANTEes funcionalmente equivalentes a modo de ejemplo de los polipéptidos de FGE incluyen sustituciones de aminoácidos conservativas de SEC ID N°: 2. Las sustituciones conservativas de aminoácidos incluyen sustituciones hechas entre 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 tanto, se contemplan por la invención VARIANTEes funcionalmente equivalentes de polipéptidos de FGE, es decir, VARIANTEes de polipéptidos de FGE que retienen la función de los polipéptidos de FGE naturales. Se hacen normalmente sustituciones de aminoácidos conservativas en la secuencia de aminoácidos de polipéptidos de FGE para producir VARIANTEes funcionalmente equivalentes de polipéptidos de FGE mediante alteración de un ácido nucleico que codifica polipéptidos de FGE (SEC ID N°: 1, 3). Tales sustituciones pueden hacerse mediante una variedad de métodos conocidos por un experto habitual en la materia. Por ejemplo, pueden hacerse sustituciones de aminoácidos mediante mutación dirigida por PCR, mutagénesis dirigida al sitio según el método de Kunkel (Kunkel, Proc. Natl. Acad. Sci. U.S.A. 82: 488-492, 1985), o mediante síntesis química de un gen que codifica un polipéptido de FGE. La actividad de fragmentos funcionalmente equivalentes de polipéptidos de FGE puede probarse clonando el gen que codifica el polipéptido de FGE alterado en un vector de expresión bacteriano o de mamífero, introduciendo el vector en una célula huésped apropiada, expresando el polipéptido de FGE alterado, y probando para una capacidad funcional de los polipéptidos de FGE como se divulga en el presente documento (por ejemplo, actividad de generación de C<sub>α</sub>-formilglicina, etc.).

Puede utilizarse una variedad de metodologías bien conocidas por el experto en la materia para obtener moléculas de FGE aisladas. El polipéptido puede purificarse a partir de células que producen de manera natural el polipéptido mediante medios cromatográficos o reconocimiento inmunológico. Alternativamente, puede introducirse un vector de expresión en células para provocar la producción del polipéptido. En otro método, pueden microinyectarse transcritos de ARNm o introducirse de otro modo en células para provocar la producción del polipéptido codificado. También puede usarse la traducción de ARNm de FGE en extractos libres de células tales como el sistema de lisado de reticulocitos para producir polipéptidos de FGE. Aquellos expertos en la materia también pueden seguir fácilmente métodos conocidos de aislamiento de polipéptidos de FGE. Éstos incluyen, pero no se limitan a, inmunocromatografía, HPLC, cromatografía de exclusión molecular, cromatografía de intercambio iónico y cromatografía de inmutafinidad.

El aislamiento del ADNc de FGE también hace posible que el experto en la materia diagnostique un trastorno caracterizado por una expresión aberrante de FGE. Estos métodos implican determinar la expresión del gen de FGE, y/o polipéptidos de FGE derivados del mismo. En la imprimación situación, tales determinaciones pueden llevarse a cabo mediante cualquier ensayo de determinación de ácido nucleico convencional, incluyendo la reacción en cadena de la polimerasa, o ensayando con sondas de hibridación marcadas como se ejemplifica a continuación. En la última situación, tal determinación puede llevarse a cabo mediante cualquier ensayo inmunológico convencional usando, por ejemplo, anticuerpos que se unen a la proteína FGE secretada. Un trastorno preferido que puede diagnosticarse según la invención es deficiencia múltiple de sulfatasas.

Pueden proporcionarse agentes de unión a péptido aislados que pueden ser, por ejemplo, anticuerpos o fragmentos de anticuerpos ("polipéptidos de unión"), que tienen la capacidad de unirse selectivamente a polipéptidos de FGE. Los anticuerpos incluyen anticuerpos policlonales y monoclonales, preparados según metodología convencional.

Significativamente, como es muy conocido en la técnica, solo una pequeña porción de una molécula de anticuerpo, el parátipe, participa en la unión del anticuerpo a su epítipe (véanse, en general, Clark, W.R. (1986) The Experimental Foundations of Modern Immunology, Wiley & Sons, Inc., Nueva York; Roitt, L (1991) Essential

Immunology, 7<sup>a</sup> ed., Blackwell Scientific Publications, Oxford). Las regiones pFc' y Fc, por ejemplo, son efectores de la cascada del complemento, pero no participan en la unión a antígeno. Un anticuerpo del que se ha escindido enzimáticamente la región pFc', o que se ha producido sin la región pFc', denominado un fragmento F(ab')<sub>2</sub>, retiene ambos sitios de unión a antígeno de un anticuerpo intacto. Similarmente, un anticuerpo del que se ha escindido enzimáticamente la región Fc, o que se ha producido sin la región Fc, denominado un fragmento Fab, retiene uno de los sitios de unión a antígeno de una molécula de anticuerpo intacta. Avanzando adicionalmente, los fragmentos Fab consisten en una cadena ligera de anticuerpo unida covalentemente y una porción de la cadena pesada de anticuerpo denominada Fd. Los fragmentos Fd son los determinantes principales de la especificidad del anticuerpo (un fragmento Fd individual puede estar asociado con hasta diez cadenas ligeras diferentes si alterar la especificidad de anticuerpo) y los fragmentos Fd retienen la capacidad de unión a epítoto en aislamiento.

Dentro de la porción de unión a antígeno de un anticuerpo, como es muy conocido en la técnica, hay regiones determinantes de la complementariedad (CDR), que interaccionan directamente con el epítoto del antígeno, y regiones estructurales (FR), que mantienen la estructura terciaria del parátoto (véanse, en general, Clark, 1986; Roitt, 1991). En tanto el fragmento Fd de cadena pesada como la cadena ligera de inmunoglobulinas IgG, hay cuatro regiones estructurales (de FR1 a FR4) separadas respectivamente por tres regiones determinantes de la complementariedad (de CDR1 a CDR3). Las CDR, y en particular las regiones CDR3, y más particularmente la CDR3 de cadena pesada, son ampliamente responsables de la especificidad del anticuerpo.

Ahora está bien establecido en la técnica que las regiones distintas de CDR de un anticuerpo de mamífero pueden sustituirse por regiones similares de anticuerpos coespecíficos o heteroespecíficos mientras que retengan la especificidad epítotica del anticuerpo original. Esto se manifiesta de la manera más clara en el desarrollo y el uso de anticuerpos "humanizados" en los que se unen covalentemente CDR no humanas a regiones FR y/o Fc/pFc' humanas para producir un anticuerpo funcional. Véanse, por ejemplo, las patentes de EE.UU. 4.816.567, 5.225.539, 5.585.089, 5.693.762 y 5.859.205. Por tanto, por ejemplo, la publicación internacional PCT número WO 92/04381 enseña la producción y el uso de anticuerpos frente al VSR murinos humanizados en los que al menos una porción de las regiones FR murinas se ha sustituido por regiones FR de origen humano. Tales anticuerpos, incluyendo fragmentos de anticuerpos intactos con capacidad de unión a antígeno, se denominan frecuentemente anticuerpos "quiméricos".

Por tanto, como será evidente para un experto habitual en la materia, puede proporcionarse lo siguiente: fragmentos F(ab')<sub>2</sub>, Fab, Fv y Fd; anticuerpos quiméricos en los que las regiones Fc y/o FR y/o CDR1 y/o CDR2 y/o CDR3 de cadena ligera se han sustituido por secuencias humanas o no humanas homólogas; anticuerpos de fragmento F(ab')<sub>2</sub> quiméricos en los que las regiones FR y/o CDR1 y/o CDR2 y/o CDR3 de cadena ligera se han sustituido por secuencias humanas o no humanas homólogas; anticuerpos de fragmento Fab quimérico en los que las regiones FR y/o CDR1 y/o CDR2 y/o CDR3 de cadena ligera se han sustituido por secuencias humanas o no humanas homólogas; y anticuerpos de fragmento Fd quimérico en los que las regiones FR y/o CDR1 y/o CDR2 se han sustituido por secuencias humanas o no humanas homólogas. También pueden proporcionarse los denominados anticuerpos monocatenarios.

Por tanto, pueden proporcionarse polipéptidos de numerosos tamaños y tipos que se unen específicamente a polipéptidos de FGE, y complejos tanto de polipéptidos de FGE como de sus componentes de unión. Estos polipéptidos también pueden derivarse de fuentes distintas de tecnología de anticuerpos. Por ejemplo, tales agentes de unión a polipéptido pueden proporcionarse mediante bibliotecas de péptidos degenerados que pueden prepararse fácilmente en disolución, en forma inmovilizada, como bibliotecas de presentación de péptidos en flagelos bacterianos o como bibliotecas de presentación en fagos. También pueden sintetizarse bibliotecas combinatorias de péptidos que contienen uno o más aminoácidos. Pueden sintetizarse adicionalmente bibliotecas de péptidos y restos sintéticos distintos de péptidos.

La presentación en fagos puede ser particularmente eficaz en la identificación de péptidos de unión útiles en la invención. Brevemente, se prepara una biblioteca de fagos (usando, por ejemplo, fago m13, fd o lambda), que presenta insertos de 4 a aproximadamente 80 residuos de aminoácido usando procedimientos convencionales. Los insertos pueden representar, por ejemplo, una matriz completamente degenerada o sesgada. Entonces pueden seleccionarse insertos que llevan fago que se unen al polipéptido de FGE o un complejo de FGE y un componente de unión. Este proceso puede repetirse a lo largo de varios ciclos de re-selección de fago que se unen al polipéptido de FGE o complejo. Ciclos repetidos conducen a un enriquecimiento de fago que lleva secuencias particulares. Puede realizarse un análisis de secuencia de ADN para identificar las secuencias de los polipéptidos expresados. Puede determinarse la porción lineal mínima de la secuencia que se une al polipéptido de FGE o complejo. Puede repetirse el procedimiento usando una biblioteca sesgada que contiene insertos que contienen una parte o la totalidad de la porción lineal mínima más uno o más residuos degenerados adicionales en la dirección 5' o en la dirección 3' de la misma. También pueden usarse métodos de selección de dos híbridos de levadura para identificar polipéptidos que se unen a los polipéptidos de FGE. Por tanto, los polipéptidos de FGE de la invención, o un fragmento de los mismos, o complejos de FGE y un componente de unión, pueden usarse para cribar bibliotecas de péptidos, incluyendo bibliotecas de presentación en fagos, para identificar y seleccionar componentes de unión a péptidos de los polipéptidos de FGE de la invención. Tales moléculas pueden usarse, como se ha descrito, para ensayos de cribado, para protocolos de purificación, para interferir directamente con el funcionamiento de FGE y

para otros fines que serán evidentes para aquellos expertos habituales en la materia.

También puede usarse un polipéptido de FGE, o un fragmento del mismo, para aislar sus componentes de unión nativos. El aislamiento de componentes de unión puede realizarse según métodos bien conocidos. Por ejemplo, pueden unirse polipéptidos de FGE aislados a un sustrato, y después puede aplicarse una disolución que se sospecha que contiene un componente de unión a FGE al sustrato. Si la pareja de unión para polipéptidos de FGE está presente en la disolución, entonces se unirá al polipéptido de FGE unido al sustrato. Entonces puede aislarse el componente de unión. Otras proteínas que son componentes de unión para FGE pueden aislarse mediante métodos similares sin experimentación excesiva. Un componente de unión preferido es una sulfatasa.

Preferiblemente se producen polipéptidos de FGE de manera recombinante, aunque tales polipéptidos pueden aislarse de extractos biológicos. Polipéptidos de FGE producidos de manera recombinante incluyen proteínas quiméricas que comprenden una fusión de una proteína FGE con otro polipéptido, por ejemplo, un polipéptido capaz de proporcionar o potenciar la unión proteína-proteína, unión de ácido nucleico específica de secuencia (tal como GAL4), potenciar la estabilidad del polipéptido de FGE en condiciones de ensayo, o proporcionar un resto detectable, tal como proteína verde fluorescente. Un polipéptido fusionado a un polipéptido o fragmento de FGE también puede proporcionar medios para detectar fácilmente la proteína de fusión, por ejemplo, mediante reconocimiento inmunológico o por marcado fluorescente.

La invención puede usarse en terapia génica. El procedimiento para realizar terapia génica *ex vivo* se explica brevemente en la patente de EE.UU. 5.399.346 y en documentos presentados en la historia de presentación de esa patente, todos los cuales son documentos disponibles para el público. En general, implica la introducción *in vitro* de una copia funcional de un gen en una célula(s) de un sujeto que contiene una copia defectuosa del gen, y devolver la(s) célula(s) modificada(s) mediante ingeniería genética al sujeto. La copia funcional del gen está bajo el control operativo de elementos reguladores que permiten la expresión del gen en la(s) célula(s) modificada(s) mediante ingeniería genética. Aquellos expertos habituales en la técnica conocen bien numerosas técnicas de transfección y transducción, así como vectores de expresión apropiados, algunos de los cuales se describen en la solicitud PCT W095/00654. También se contempla terapia génica *in vivo* usando vectores tales como adenovirus, retrovirus, virus del herpes y liposomas seleccionados como diana.

Agentes de unión específicos para FGE, métodos de identificación y preparación de tales agentes, y su uso en diagnóstico, terapia y desarrollo farmacéutico. Por ejemplo, los agentes farmacológicos específicos para FGE son útiles en una variedad de aplicaciones de diagnóstico y terapéuticas, especialmente cuando la enfermedad o el pronóstico de la enfermedad está asociado con características de unión a FGE alteradas tales como en deficiencia múltiple de sulfatasas. Agentes de unión específicos para FGE novedosos incluyen anticuerpos específicos para FGE, receptores de superficie celular, y otros agentes de unión naturales intracelulares y extracelulares identificados con ensayos tales como cribados de dos híbridos, y agentes de unión no naturales intracelulares y extracelulares identificados en cribados de bibliotecas químicas y similares.

En general, la especificidad de la unión de FGE a una molécula específica se determina mediante constantes de equilibrio de unión. Las dianas que son capaces de unirse selectivamente a un polipéptido de FGE tienen preferiblemente constantes de equilibrio de unión de al menos aproximadamente  $10^7 M^{-1}$ , más preferiblemente al menos aproximadamente  $10^8 M^{-1}$ , y lo más preferiblemente al menos aproximadamente  $10^9 M^{-1}$ . Puede usarse una amplia variedad de ensayos basados en células y libres de células para demostrar la unión específica para FGE. Los ensayos basados en células incluyen uno, dos y tres cribados híbridos, ensayos en los que se inhibe o se incrementa la transcripción mediada por FGE, etc. Ensayos libres de células incluyen ensayos de unión a proteína FGE, inmunoensayos, etc. Otros ensayos útiles para cribar agentes que se unen a polipéptidos de FGE incluyen transferencia de energía de fluorescencia por resonancia (FRET), y análisis de desplazamiento de movilidad electroforética (EMSA).

Puede proporcionarse un método de diagnóstico de un trastorno caracterizado por una expresión aberrante de una molécula de ácido nucleico, un producto de expresión de la misma, o un fragmento de un producto de expresión de la misma. El método implica poner en contacto una muestra biológica aislada de un sujeto con un agente que se une específicamente a la molécula de ácido nucleico, un producto de expresión de la misma, o un fragmento de un producto de expresión de la misma, y determinar la interacción entre el agente y la molécula de ácido nucleico o el producto de expresión como una determinación del trastorno, en el que la molécula de ácido nucleico es una molécula de FGE. El trastorno es deficiencia múltiple de sulfatasas. Las mutaciones en el gen de FGE que provocan la expresión aberrante de moléculas de FGE produce los siguientes cambios de aminoácido en SEC ID N°: 2: Met1Arg; Met1Val; Leu20Phe; Ser155Pro; Ala177Pro; Cys218Tyr; Arg224Trp; Asn259Ile; Pro266Leu; Ala279Val; Arg327Stop; Cys336Arg; Arg345Cys; Ala348Pro; Arg349Gln; Arg349Trp; Arg349Trp; Ser359Stop; o una combinación de los mismos.

En el caso en el que la molécula sea una molécula de ácido nucleico, tales determinaciones pueden llevarse a cabo mediante cualquier ensayo de determinación de ácido nucleico estándar, incluyendo la reacción en cadena de la polimerasa, o ensayando con sondas de hibridación marcadas como se ejemplifica en el presente documento. En el caso en el que la molécula sea un producto de expresión de la molécula de ácido nucleico, o un fragmento de un

producto de expresión de la molécula de ácido nucleico, tal determinación puede llevarse a cabo mediante cualquier ensayo inmunológico estándar usando, por ejemplo, anticuerpos que se unen a cualquiera de los productos de expresión de polipéptido.

5 “Expresión aberrante” se refiere a expresión disminuida (subexpresión) o expresión aumentada (sobreexpresión) de moléculas de FGE (ácidos nucleicos y/o polipéptidos) en comparación con un control (es decir, expresión de la misma molécula en un sujeto sano o “normal”). Un “sujeto sano”, como se usa en el presente documento, se refiere a un sujeto que, según patrones médicos estándar, no tiene ni está en riesgo de desarrollar una deficiencia múltiple de sulfatasas. Los sujetos sanos tampoco muestran de otro modo síntomas de enfermedad. En otras palabras, tales  
10 sujetos, si se examinan por un profesional médico, se caracterizarían como sanos y libres de síntomas de una deficiencia múltiple de sulfatasas. Éstos incluyen características de leucodistrofia metacromática y de una mucopolisacaridosis, tales como cantidades aumentadas de mucopolisacáridos ácidos en diversos tejidos, “gargolismo” leve, rápido deterioro neurológico, presencia excesiva de mucopolisacárido y sulfatida en la orina, proteína de líquido cefalorraquídeo aumentada y degeneración metacromática de mielina en nervios periféricos.

15 Pueden usarse kits novedosos para medir los niveles de los ácidos nucleicos, o productos de expresión de los mismos.

20 En una realización, un kit comprende un envase que contiene un agente que se une selectivamente a cualquiera de los ácidos nucleicos aislados de FGE anteriores, o productos de expresión de los mismos, y un control para comparar con un valor medido de unión de dicho agente a cualquiera de los ácidos nucleicos aislados de FGE anteriores o productos de expresión de los mismos. En algunas realizaciones, el control es un valor predeterminado para comparar con el valor medido. En determinadas realizaciones, el control comprende un epítipo del producto de expresión de cualquiera de los ácidos nucleicos aislados de FGE anteriores. En una realización, el kit comprende  
25 además un segundo agente que se une selectivamente a un polipéptido seleccionado del grupo que consiste en iduronato 2-sulfatasa, sulfamidasa, N-acetilgalactosamina 6-sulfatasa, N-acetilglucosamina 6-sulfatasa, arilsulfatasa A, arilsulfatasa B, arilsulfatasa C, arilsulfatasa D, arilsulfatasa E, arilsulfatasa F, arilsulfatasa G, HSulf-1, HSulf-2, HSulf-3, HSulf-4, HSulf-5 y HSulf-6, o un péptido de los mismos, y un control para comparar con un valor medido de unión de dicho segundo agente a dicho polipéptido o péptido del mismo.

30 En el caso de detección de ácido nucleico, pueden incluirse pares de cebadores para amplificar una molécula de ácido nucleico de la invención. Los kits preferidos incluirían controles tales como cantidades conocidas de sondas de ácido nucleico, epítopes (tales como iduronato 2-sulfatasa, sulfamidasa, N-acetilgalactosamina 6-sulfatasa, N-acetilglucosamina 6-sulfatasa, arilsulfatasa A, arilsulfatasa B, arilsulfatasa C, arilsulfatasa D, arilsulfatasa E,  
35 arilsulfatasa F, arilsulfatasa G, HSulf-1, HSulf-2, HSulf-3, HSulf-4, HSulf-5 y HSulf-6, productos de expresión) o anticuerpos anti-epítipo, así como instrucciones u otro material impreso. En determinadas realizaciones, el material impreso puede caracterizar el riesgo de desarrollar una afección de deficiencia de sulfatasa basándose en el resultado del ensayo. Los reactivos pueden envasarse en recipientes y/o recubrirse sobre pocillos en cantidades predeterminadas, y los kits pueden incluir materiales estándar tales como reactivos inmunológicos marcados (tales como anticuerpos anti-IgG marcados) y similares. Un kit es una placa de microtitulación de poliestireno envasada recubierta con proteína FGE y un recipiente que contiene anticuerpos anti-IgG humana marcados. Se pone en contacto un pocillo de la placa con, por ejemplo, un líquido biológico, se lava y después se pone en contacto con el anticuerpo anti-IgG. Entonces se detecta el marcador. En la Figura 25 se ilustra un kit que es ejemplo de las características de la presente invención, generalmente designado con el número 11. El kit 11 comprende los  
40 siguientes elementos principales: un envase 15, un agente 17 como se describe en el presente documento, un agente 19 de control e instrucciones 21. El envase 15 es una estructura de tipo caja para contener un vial (o varios viales) que contiene un agente 17 de la invención, un vial (o varios viales) que contiene un agente de control 19, e instrucciones 21. Los expertos en la materia pueden modificar fácilmente el envase 15 para adaptarlo a necesidades individuales.

50 Un método para tratar deficiencia múltiple de sulfatasas en un sujeto implica administrar a un sujeto en necesidad de tal tratamiento un agente que modula la actividad de generación de  $C_{\alpha}$ -formilglicina, en una cantidad eficaz para aumentar la actividad de generación de  $C_{\alpha}$ -formilglicina en el sujeto. En algunas realizaciones, el método comprende además co-administrar un agente seleccionado del grupo que consiste en una molécula de ácido nucleico que codifica iduronato 2-sulfatasa, sulfamidasa, N-acetilgalactosamina 6-sulfatasa, N-acetilglucosamina 6-sulfatasa,  
55 arilsulfatasa A, arilsulfatasa B, arilsulfatasa C, arilsulfatasa D, arilsulfatasa E, arilsulfatasa F, arilsulfatasa G, HSulf-1, HSulf-2, HSulf-3, HSulf-4, HSulf-5 y HSulf-6, un producto de expresión de la molécula de ácido nucleico, y/o un fragmento del producto de expresión de la molécula de ácido nucleico.

60 “Agentes que modulan la expresión” de un ácido nucleico o un polipéptido, como se usa en el presente documento, se conocen en la técnica, y se refieren a ácidos nucleicos sentido y antisentido, ácidos nucleicos negativos dominantes, anticuerpos para los polipéptidos, y similares. Es útil cualquier agente que module la expresión de una molécula (y como se describe en el presente documento, module su actividad). En determinadas realizaciones, el agente que modula la actividad de generación de  $C_{\alpha}$ -formilglicina es una molécula de ácido nucleico aislada (por ejemplo, un ácido nucleico de SEC ID N°: 3). En realizaciones importantes, el agente que modula la actividad de generación de  $C_{\alpha}$ -formilglicina es un péptido descrito en el presente documento (por ejemplo, un péptido de SEC ID  
65



Nº: 2). En algunas realizaciones, el agente que modula la actividad de generación de C<sub>α</sub>-formilglicina es un ácido nucleico sentido.

Se proporciona un método para aumentar la actividad de generación de C<sub>α</sub>-formilglicina en un sujeto. El método implica administrar una molécula de ácido nucleico de FGE aislada de la invención, y/o un producto de expresión de la misma, a un sujeto, en una cantidad eficaz para aumentar la actividad de generación de C<sub>α</sub>-formilglicina en el sujeto.

Se proporciona un método de aumento de la actividad de generación de C<sub>α</sub>-formilglicina en una célula. El método implica poner en contacto la célula con una molécula de ácido nucleico aislada descrita en el presente documento (por ejemplo, un ácido nucleico de SEC ID Nº: 1), o un producto de expresión de la misma (por ejemplo, un péptido de SEC ID Nº: 2), en una cantidad eficaz para aumentar la actividad de generación de C<sub>α</sub>-formilglicina en la célula. En realizaciones importantes, el método implica activar el gen de FGE endógeno para aumentar la actividad de generación de C<sub>α</sub>-formilglicina en la célula.

En cualquiera de las realizaciones anteriores, el ácido nucleico puede estar operativamente acoplado a una secuencia de expresión génica que dirige la expresión de la molécula de ácido nucleico dentro de una célula eucariota, como una célula HT-1080. La "secuencia de expresión génica" es cualquier secuencia de nucleótidos reguladora, tal como una secuencia promotora o combinación de promotor-potenciador, que facilita la transcripción y traducción eficaces del ácido nucleico al que está operativamente unida. La secuencia de expresión génica puede ser, por ejemplo, un promotor de mamífero o viral, tal como un promotor constitutivo o inducible. Promotores de mamífero constitutivos incluyen, pero no se limitan a, los promotores para los siguientes genes: hipoxantina fosforibosil transferasa (HFTR), adenosina desaminasa, piruvato cinasa, promotor de α-actina y otros promotores constitutivos. Promotores virales a modo de ejemplo que funcionan constitutivamente en células eucariotas incluyen, por ejemplo, promotores del virus simio, virus del papiloma, adenovirus, virus de inmunodeficiencia humana (VIH), virus del sarcoma de Rous, citomegalovirus, las repeticiones terminales largas (LTR) del virus de leucemia de Moloney y otros retrovirus, y el promotor de timidina cinasa del virus del herpes simple. Otros promotores constitutivos son conocidos por aquellos expertos habituales en la técnica. Los promotores útiles como secuencias de expresión génica de la invención también incluyen promotores inducibles. Los promotores inducibles se activan en presencia de un agente de inducción. Por ejemplo, el promotor de metalotioneína se activa para aumentar la transcripción y traducción en presencia de determinados iones metálicos. Otros promotores inducibles son conocidos por aquellos expertos habituales en la técnica.

En general, la secuencia de expresión génica debe incluir, según sea necesario, secuencias no transcritas en 5' y no traducidas en 5' que participan en la iniciación de la transcripción y traducción, respectivamente, tales como una caja TATA, secuencia de ocupación, secuencia CAAT y similares. Especialmente, tales secuencias no transcritas en 5' incluirán una región promotora que incluye una secuencia promotora para el control transcripcional del ácido nucleico operativamente unido. Las secuencias de expresión génica incluyen opcionalmente secuencias potenciadoras o secuencias activadoras en la dirección 5', según se desee.

Preferiblemente, cualquiera de las moléculas de ácido nucleico de FGE descritas en el presente documento está enlazada a una secuencia de expresión génica 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 selectivamente activa en un tipo de célula tal, provocando así la expresión de la molécula de ácido nucleico en esas células. El promotor de sinapsina 1, por ejemplo, puede usarse para expresar cualquiera de las moléculas de ácido nucleico anteriores de la invención en una neurona; y el promotor del gen del factor de von Willebrand, por ejemplo, puede usarse para expresar una molécula de ácido nucleico en una célula endotelial vascular. Aquellos expertos habituales en la técnica serán capaces de identificar fácilmente promotores alternativos que son capaces de expresar una molécula de ácido nucleico en cualquiera de las células preferidas.

Se dice que la secuencia de ácido nucleico y la secuencia de expresión génica están "operativamente enlazadas" cuando están unidas covalentemente de tal manera que se coloca la transcripción y/o traducción de la secuencia codificante de ácido nucleico (por ejemplo, en el caso de FGE, SEC ID Nº: 3) bajo la influencia o el control de la secuencia de expresión génica. Si se desea que la secuencia de ácido nucleico se traduzca en una proteína funcional, se dice que dos secuencias de ADN están operativamente enlazadas si la inducción de un promotor en la secuencia de expresión génica en 5' produce la transcripción de la secuencia de ácido nucleico y si la naturaleza del enlace entre las dos secuencias de ADN no (1) produce la introducción de una mutación del marco de lectura, (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 del transcrito de ARN correspondiente para traducirse en una proteína. Por tanto, una secuencia de expresión génica estaría operativamente enlazada a una secuencia de ácido nucleico si la secuencia de expresión génica fuera capaz de efectuar la transcripción de esa secuencia de ácido nucleico de tal manera que el transcrito resultante pudiera traducirse en la proteína o polipéptido deseado.

Las moléculas descritas en el presente documento pueden administrarse a los tipos de célula preferidos solas o en asociación con un vector (véase también la discusión anterior sobre vectores). En su sentido más amplio (y de manera coherente con la descripción de vectores de expresión y de direccionamiento en otra parte en el presente

documento), un "vector" es cualquier vehículo capaz de facilitar: (1) la administración de una molécula a una célula diana y/o (2) la captación de la molécula por una célula diana. Preferiblemente, los vectores de administración transportan la molécula al interior de la célula diana con degradación reducida con respecto al grado de degradación que se produciría en ausencia del vector. Opcionalmente, puede unirse un "ligando de direccionamiento" al vector para administrar selectivamente el vector a una célula que expresa sobre su superficie el receptor relacionado para el ligando de direccionamiento. De esta manera, el vector (que contiene un ácido nucleico o una proteína) puede administrarse selectivamente a una neurona. Metodologías para el direccionamiento incluyen conjugados, tales como los descritos en la patente de EE.UU. 5.391.723 concedida a Priest. Otro ejemplo de un vehículo de direccionamiento bien conocido es un liposoma. Los liposomas están comercialmente disponibles de Gibco BRL. Están publicados numerosos métodos para preparar liposomas de direccionamiento.

En general, los vectores útiles en la invención incluyen, pero no se limitan a, plásmidos, fagémidos, virus, otros vehículos derivados de cepas virales o bacterianas que se han manipulado mediante la inserción o incorporación de las secuencias de ácido nucleico descritas en el presente documento, y fragmentos de ácido nucleico adicionales (por ejemplo, potenciadores, promotores) que pueden unirse a las secuencias de ácido nucleico descritas en el presente documento. Vectores virales son un tipo preferido de vector e incluyen, pero no se limitan a, secuencias de ácido nucleico de los siguientes virus: adenovirus; virus adeno-asociado; retrovirus, tales como virus de la leucemia murina de Moloney; virus del sarcoma murino de Harvey; virus de tumor de mama murino; virus del sarcoma de Rous; virus de tipo SV40; virus del polio; virus de Epstein-Barr; virus del papiloma; virus del herpes; virus de la variolovacuna; virus de la polio; y virus de ARN tales como un retrovirus. Pueden emplearse fácilmente otros vectores no mencionados, pero conocidos en la técnica.

Un virus particularmente preferido para determinadas aplicaciones es el virus adeno-asociado, un virus de ADN bicatenario. El virus adeno-asociado es capaz de infectar una amplia gama de tipos de células y especies y puede diseñarse mediante ingeniería para ser deficiente para la replicación. Además tiene ventajas, tales como estabilidad al calor y a disolventes de lípidos, altas frecuencias de transducción en células de diversos linajes, incluyendo células hematopoyéticas, y falta de inhibición de superinfección, permitiendo así múltiples series de transducciones. Se notifica que el virus adeno-asociado puede integrarse en el ADN celular humano de una manera específica del sitio, minimizando así la posibilidad de mutagénesis por inserción y la variabilidad de la expresión del gen insertado. Además, se ha realizado un seguimiento de infecciones por virus adeno-asociado no mutante en cultivo tisular durante más de 100 pasos en ausencia de presión selectiva, que implica que la integración genómica del virus adeno-asociado es un acontecimiento relativamente estable. El virus adeno-asociado también puede funcionar de una manera extracromosómica.

En general, otros vectores virales preferidos se basan en virus eucariotas no citopáticos en los que se han sustituido genes no esenciales por el gen de interés. Virus no citopáticos incluyen retrovirus, cuyo ciclo de vida implica la transcripción inversa de ARN viral genómico en ADN con la posterior integración proviral en el ADN celular huésped. Se han aprobado adenovirus y retrovirus para ensayos de terapia génica humana. En general, los retrovirus son deficientes para la replicación (es decir, capaces de dirigir la síntesis de las proteínas deseadas, pero incapaces de fabricar una partícula infecciosa). Tales vectores de expresión retrovirales genéticamente alterados tienen utilidad general para la transducción de alta eficacia de genes *in vivo*. Se proporcionan protocolos estándar para producir retrovirus deficientes para la replicación (incluyendo las etapas de incorporación de material genético exógeno en un plásmido, transfección de una línea celular de empaquetamiento con plásmido, producción de retrovirus recombinantes mediante la línea celular de empaquetamiento, recogida de partículas virales de medios de cultivo tisular, e infección de las células diana con partículas virales) en Kriegler, M., "Gene Transfer and Expression, A Laboratory Manual", W.H. Freeman C.O., Nueva York (1990) y Murry, E.J. Ed. "Methods in Molecular Biology", vol. 7, Humana Press, Inc., Clifton, Nueva Jersey (1991).

Otro vector retroviral preferido es el vector derivado del virus de la leucemia murina de Moloney, como se describe Nabel, E.G., et al., Science, 1990, 249:1285-1288. Se notifica que estos vectores fueron eficaces para la administración de genes a las tres capas de la pared arterial, incluyendo la media. Otros vectores preferidos se divulgan en Flugelman, et al., Circulation, 1992, 85:1110-1117. Vectores adicionales que son útiles para administrar moléculas de la invención descritas en el presente documento se describen en la patente de EE.UU. Nº 5.674.722 de Mulligan, et al.

Además de los vectores anteriores, pueden usarse otros métodos de administración para suministrar una molécula descrita en el presente documento a una célula tal como una neurona, hígado, fibroblasto y/o una célula endotelial vascular, y facilitar la captación por la misma.

Un método de administración tal preferido de la invención es un sistema de dispersión coloidal. Los sistemas de dispersión coloidal incluyen sistemas basados en lípidos incluyendo emulsiones de aceite en agua, micelas, micelas mixtas y liposomas. Un sistema coloidal preferido es un liposoma. Los liposomas son vasos de membrana artificiales que son útiles como vector de administración *in vivo* o *in vitro*. Se ha mostrado que vasos unilaminares grandes (LUV), cuyo tamaño oscila de 0,2 - 4,0  $\mu\text{m}$ , pueden encapsular macromoléculas grandes. Pueden encapsularse ARN, ADN y viriones intactos dentro del interior acuoso y administrarse a las células en una forma biológicamente activa (Fraleley, et al., Trends Biochem. Sci., 1981, 6:77). Con el fin de que un liposoma sea un vector de transferencia

génica eficaz, debe estar presente una o más de las siguientes características: (1) encapsulación del gen de interés con alta eficacia con retención de actividad biológica; (2) unión preferible y sustancial a una célula diana en comparación con células no diana; (3) administración del contenido acuoso de la vesícula al citoplasma de la célula diana con alta eficacia; y (4) expresión precisa y eficaz de la información genética.

Pueden dirigirse liposomas a un tejido particular, tal como el miocardio o la pared de células vasculares, acoplando el liposoma a un ligando específico tal como un anticuerpo monoclonal, azúcar, glicolípido o proteína. Ligandos que pueden ser útiles para dirigir un liposoma a la pared vascular incluyen, pero no se limitan a, la proteína viral de la cubierta del virus hemaglutinante de Japón. Adicionalmente, el vector puede acoplarse a un péptido de direccionamiento nuclear, que dirigirá el ácido nucleico al núcleo de la célula huésped.

Los liposomas están comercialmente disponibles de Gibco BRL, por ejemplo, como LIPOFECTIN™ y LIPOFECTACE™, que están formados por lípidos catiónicos tales como cloruro de N-[1-(2,3-dioleiloxi)-propil]-N,N,N-trimetilamonio (DOTMA) y bromuro de dimetildioctadecilamonio (DDAB). Métodos de preparación de liposomas se conocen bien en la técnica y se han descrito en muchas publicaciones. También se han revisado liposomas por Gregoriadis, G. en Trends in Biotechnology. V. 3, págs. 235-241 (1985). También se describen liposomas novedosos para la administración intracelular de macromoléculas, incluyendo ácidos nucleicos, en la solicitud internacional PCT nº PCT/US96/07572 (publicación Nº WO 96/40060, titulada "Intracellular Delivery of Macromolecules").

En una realización particular, el vehículo preferido es una micropartícula biocompatible o implante que es adecuado para su implantación en el receptor mamífero. Implantes bioerosionables a modo de ejemplo que son útiles según este método se describen en la solicitud internacional PCT nº PCT/US/03307 (publicación Nº WO 95/24929, titulada "Polymeric Gene Delivery System", que reivindica prioridad a la solicitud de patente de EE.UU. nº de serie 213.668, presentada el 15 de marzo de 1994). El documento PCT/US/0307 describe una matriz polimérica biocompatible, preferiblemente biodegradable, para contener un gen exógeno bajo el control de un promotor apropiado. La matriz polimérica se usa para lograr la liberación sostenida del gen exógeno en el paciente. Los ácidos nucleicos descritos en el presente documento pueden encapsularse o dispersarse dentro de la matriz polimérica biocompatible, preferiblemente biodegradable, divulgada en el documento PCT/US/03307. La matriz polimérica está preferiblemente en forma de una micropartícula, tal como una microesfera (en la que un ácido nucleico se dispersa por la totalidad de una matriz polimérica sólida) o una microcápsula (en la que un ácido nucleico está almacenado en el núcleo de una carcasa polimérica). Otras formas de la matriz polimérica para contener los ácidos nucleicos de la invención incluyen películas, recubrimientos, geles, implantes y prótesis endovasculares. El tamaño y la composición del dispositivo de matriz polimérica se seleccionan para producir una cinética de liberación favorable en el tejido en el que se implanta el dispositivo de matriz. El tamaño del dispositivo de matriz polimérica se selecciona además según el método de administración que va a usarse, normalmente inyección en un tejido o administración de una suspensión mediante aerosol en las áreas nasal y/o pulmonar. La composición de matriz polimérica puede seleccionarse para que tenga tanto tasas de degradación favorables como también para estar formada de un material que es bioadhesivo, para aumentar adicionalmente la eficacia de transferencia cuando el dispositivo se administra a una superficie vascular. La composición de matriz también puede seleccionarse para no degradarse, sino más bien para liberarse mediante difusión a lo largo de un periodo de tiempo prolongado.

Pueden usarse matrices poliméricas tanto biodegradables como no biodegradables para administrar los ácidos de la invención al sujeto. Se prefieren matrices biodegradables. Tales polímeros pueden ser polímeros naturales o sintéticos. Se prefieren polímeros sintéticos. El polímero se selecciona basándose en el periodo de tiempo a lo largo del cual se desea la liberación, generalmente en el orden de algunas horas a un año o más. Normalmente, la liberación a lo largo de un periodo que oscila entre algunas horas y de tres a doce meses es lo más deseable. El polímero está opcionalmente en forma de un hidrogel que puede absorber hasta aproximadamente el 90 % de su peso en agua y más, opcionalmente está reticulado con iones multivalentes u otros polímeros.

En general, los ácidos nucleicos descritos en el presente documento pueden administrarse usando el implante bioerosionable mediante difusión, o más preferiblemente, mediante degradación de la matriz polimérica. Polímeros sintéticos a modo de ejemplo que pueden usarse para formar el sistema de administración biodegradable incluyen: poliamidas, policarbonatos, polialquilenos, polialquilenglicoles, poli(óxidos de alquileno), poli(tereftalatos de alquileno), poli(alcoholes vinílicos), poliviniléteres, poli(ésteres vinílicos), poli(haluros de vinilo), polivinilpirrolidona, poliglicólidos, polisiloxanos, poliuretanos y copolímeros de los mismos, alquicelulosa, hidroxialquicelulosas, éteres de celulosa, ésteres de celulosa, nitrocelulosas, polímeros de ésteres acrílicos y metacrílicos, metilcelulosa, etilcelulosa, hidroxipropilcelulosa, hipromelosa, hidroxibutilmetilcelulosa, acetato de celulosa, propionato de celulosa, acetato-butilato de celulosa, acetato-ftalato de celulosa, carboxietilcelulosa, triacetato de celulosa, sal sódica de sulfato de celulosa, poli(metacrilato de metilo), poli(metacrilato de etilo), poli(metacrilato de butilo), poli(metacrilato de isobutilo), poli(metacrilato de hexilo), poli(metacrilato de isodecilo), poli(metacrilato de laurilo), poli(metacrilato de fenilo), poli(acrilato de metilo), poli(acrilato de isopropilo), poli(acrilato de isobutilo), poli(acrilato de octadecilo), polietileno, polipropileno, polietilenglicol, poli(óxido de etileno), poli(tereftalato de etileno), poli(alcoholes vinílicos), poli(acetato de vinilo), poli(cloruro de vinilo), poliestireno y polivinilpirrolidona.

Ejemplos de polímeros no biodegradables incluyen etileno-acetato de vinilo, poli(ácido (met)acrílico), poliamidas, copolímeros y mezclas de los mismos.

Ejemplos de polímeros biodegradables incluyen polímeros sintéticos tales como polímeros de ácido láctico y ácido glicólico, polianhídridos, poli(orto)ésteres, poliuretanos, poli(ácido butírico), poli(ácido valérico) y poli(lactida-caprolactona), y polímeros naturales tales como alginato y otros polisacáridos incluyendo dextrano y celulosa, colágeno, derivados químicos de los mismos (sustituciones, adiciones de grupos químicos, por ejemplo, alquilo, alquileo, hidroxilaciones, oxidaciones y otras modificaciones realizadas rutinariamente por aquellos expertos en la materia), albúmina y otras proteínas hidrófilas, zeína y otras prolaminas y proteínas hidrófobas, copolímeros y mezclas de los mismos. En general, estos materiales se degradan o bien mediante hidrólisis enzimática o bien mediante exposición a agua *in vivo*, mediante erosión superficial o en masa.

Polímeros bioadhesivos de interés particular incluyen hidrogeles bioerosionables descritos por H.S. Sawhney, C.P. Pathak y J.A. Hubell en *Macromolecules*, 1993, 26, 581-587, poli(ácidos hialurónicos), caseína, gelatina, glutina, polianhídridos, poli(ácido acrílico), alginato, quitosano, poli(metacrilatos de metilo), poli(metacrilatos de etilo), poli(metacrilato de butilo), poli(metacrilato de isobutilo), poli(metacrilato de hexilo), poli(metacrilato de isodecilo), poli(metacrilato de laurilo), poli(metacrilato de fenilo), poli(acrilato de metilo) poli(acrilato de isopropilo), poli(acrilato de isobutilo) y poli(acrilato de octadecilo). Por tanto, puede proporcionarse una composición de las moléculas anteriormente descritas de la invención para su uso como medicamento, métodos de preparación del medicamento y métodos para la liberación sostenida del medicamento *in vivo*.

También pueden usarse agentes de compactación en combinación con un vector descrito en el presente documento. Un "agente de compactación", como se usa en el presente documento, se refiere a un agente, tal como una histona, que neutraliza las cargas negativas en el ácido nucleico y así permite la compactación del ácido nucleico en un gránulo fino. La compactación del ácido nucleico facilita la captación del ácido nucleico por la célula diana. Los agentes de compactación pueden usarse solos, es decir, para administrar un ácido nucleico aislado descrito en el presente documento en una forma que se capta más eficazmente por la célula o, más preferiblemente, en combinación con uno o más de los vectores descritos anteriormente.

Otras composiciones a modo de ejemplo que pueden usarse para facilitar la captación por una célula diana de los ácidos nucleicos descritos en el presente documento incluyen fosfato de calcio y otros mediadores químicos de transporte intracelular, composiciones de microinyección y electroporación.

Pueden proporcionarse métodos de aumento de la actividad de sulfatasa en una célula. Tales métodos implican poner en contacto una célula que expresa una sulfatasa con una molécula de ácido nucleico aislada descrita en el presente documento (por ejemplo, una molécula de ácido nucleico de FGE que tiene una secuencia seleccionada del grupo que consiste en SEC ID N°: 1, 3, 4, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77 y 80-87), o un producto de expresión de la misma (por ejemplo, un polipéptido que tiene una secuencia seleccionada del grupo que consiste en SEC ID N°: 2, 5, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76 y 78), en una cantidad eficaz para aumentar la actividad de sulfatasa en la célula. "Aumentar" la actividad de sulfatasa, como se usa en el presente documento, se refiere a aumentar la afinidad por, y/o conversión de, el sustrato específico para la sulfatasa, normalmente el resultado de un aumento en la formación de *FGly* en la molécula de sulfatasa. En una realización, la célula expresa una sulfatasa a niveles superiores a los de células no mutantes. Por "aumentar la actividad de sulfatasa en una célula" también se refiere a aumentar la actividad de una sulfatasa que se secreta por la célula. La célula puede expresar una sulfatasa endógena y/o exógena. Dicho poner en contacto la molécula de FGE también se refiere a activar el gen de FGE endógeno de las células. En realizaciones importantes, se activa la sulfatasa endógena. En determinadas realizaciones, la sulfatasa es iduronato 2-sulfatasa, sulfamidasa, N-acetilgalactosamina 6-sulfatasa, N-acetilglucosamina 6-sulfatasa, arilsulfatasa A, arilsulfatasa B, arilsulfatasa C, arilsulfatasa D, arilsulfatasa E, arilsulfatasa F, arilsulfatasa G, HSulf-1, HSulf-2, HSulf-3, HSulf-4, HSulf-5 y/o HSulf-6. En determinadas realizaciones la célula es una célula de mamífero.

Puede proporcionarse una composición farmacéutica. La composición puede comprender una sulfatasa que se produce por una célula, en una cantidad farmacéuticamente eficaz para tratar una deficiencia de sulfatasa, y un vehículo, en el que dicha célula se ha puesto en contacto con un agente que comprende una molécula de ácido nucleico aislada que codifica una FGE (por ejemplo, una molécula de ácido nucleico que tiene una secuencia seleccionada del grupo que consiste en la SEC ID N°: 1, 3, 4, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77 y 80-87), o un producto de expresión de la misma (por ejemplo, un péptido seleccionado del grupo que consiste en la SEC ID N°: 2, 5, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76 y 78). En realizaciones importantes, la sulfatasa se expresa a niveles superiores que en células normales/control.

También puede proporcionarse una célula que produce sulfatasa; en la que se aumenta la razón de sulfatasa activa con respecto a sulfatasa total producida por la célula. La célula comprende: (i) una sulfatasa con una actividad aumentada en comparación con un control, y (ii) una enzima generadora de formilglucina con una actividad aumentada en comparación con un control, en la que la razón de sulfatasa activa con respecto a sulfatasa total producida por la célula se aumenta en al menos el 5 % con respecto a la razón de sulfatasa activa con respecto a sulfatasa total producida por la célula en ausencia de la enzima generadora de formilglucina. Se conoce en la técnica que la sobreexpresión de sulfatasas puede disminuir la actividad de sulfatasas endógenas (Anson et al., *Biochem. J.*, 1993, 294:657-662). Además, solo una fracción de las sulfatasas recombinantes es activa. Los presentes inventores

han descubierto, inesperadamente, que la expresión/actividad aumentada de FGE en una célula con expresión/actividad aumentada de una sulfatasa produce la producción de una sulfatasa que es más activa. Como la presencia de *FGly* en una molécula de sulfatasa está asociada con actividad de sulfatasa, puede cuantificarse la "sulfatasa activa" determinando la presencia de *FGly* en el producto celular de sulfatasa usando espectrometría de masas MALDI-TOF, como se describe en otra parte en el presente documento. Entonces puede determinarse fácilmente la razón con sulfatasa total.

Pueden proporcionarse métodos para el diagnóstico y la terapia de deficiencias de sulfatasa. Tales trastornos incluyen, pero no se limitan a, deficiencia múltiple de sulfatasas, mucopolisacaridosis II (MPS II; síndrome de Hunter), mucopolisacaridosis IIIA (MPS IIIA; síndrome de Sanfilippo A), mucopolisacaridosis VIII (MPS VIII), mucopolisacaridosis IVA (MPS IVA; síndrome de Morquio A), mucopolisacaridosis VI (MPS VI; síndrome de Maroteaux-Lamy), leucodistrofia metacromática (MLD), condrodisplasia punctata 1 recesiva ligada al cromosoma X e ictiosis ligada al cromosoma X (deficiencia de sulfatasa esteroidea).

Los métodos descritos en el presente documento son útiles en el tratamiento tanto agudo como profiláctico de cualquiera de las afecciones anteriores. Como se usa en el presente documento, un tratamiento agudo se refiere al tratamiento de sujetos que tienen una afección particular. Tratamiento profiláctico se refiere al tratamiento de sujetos en riesgo de tener la afección, pero que actualmente no tienen o experimentan los síntomas de la afección.

En su sentido más amplio, los términos "tratamiento" o "tratar" se refieren a tratamientos tanto agudos como profilácticos. Si el sujeto en necesidad de tratamiento está experimentando una afección (o tiene o está teniendo una afección particular), entonces tratar la afección se refiere a mejorar, reducir o eliminar la afección o uno o más síntomas que surgen de la afección. En algunas realizaciones preferidas, tratar la afección se refiere a mejorar, reducir o eliminar un síntoma específico o un subconjunto específico de síntomas asociados a la afección. Si el sujeto en necesidad de tratamiento es uno que está en riesgo de tener una afección, entonces tratar al sujeto se refiere a reducir el riesgo de que el sujeto tenga la afección.

El modo de administración y la dosificación de un agente terapéutico descrito en el presente documento variarán con el estadio particular de la afección que está tratándose, la edad y la condición física del sujeto que está tratándose, la duración del tratamiento, la naturaleza de la terapia concurrente (si la hay), la vía específica de administración, y factores similares dentro del conocimiento y la experiencia del profesional sanitario.

Como se describe en el presente documento, se administran agentes en cantidades eficaces para tratar cualquiera de las deficiencias de sulfatasa anteriores. En general, una cantidad eficaz es cualquier cantidad que puede provocar un cambio beneficioso en un tejido deseado de un sujeto. Preferiblemente, una cantidad eficaz es la cantidad suficiente para provocar un cambio fenotípico favorable en una afección particular tal como una reducción, alivio o eliminación de un síntoma o de una afección en su totalidad.

En general, una cantidad eficaz es la cantidad de una preparación farmacéutica que sola, o junto con dosis adicionales, produce la respuesta deseada. Esto puede implicar solo ralentizar la progresión de la afección temporalmente, aunque más preferiblemente, implica detener la progresión de la afección permanentemente o retrasar la aparición de, o prevenir que se produzca, la afección. Esto puede monitorizarse mediante métodos rutinarios. Generalmente, las dosis de compuestos activos serían de aproximadamente 0,01 mg/kg por día a 1000 mg/kg por día. Se espera que dosis que oscilan de 50 µg-500 mg/kg sean adecuadas, preferiblemente por vía oral y en una o varias administraciones por día.

Tales cantidades dependerán, por supuesto, de la afección particular que está tratándose, la gravedad de la afección, los parámetros del paciente individual incluyendo edad, condición física, tamaño y peso, la duración del tratamiento, la naturaleza de la terapia concurrente (si la hay), la vía específica de administración y factores similares dentro del conocimiento y la experiencia del profesional sanitario. Dosis inferiores resultarán de determinadas formas de administración, tales como administración intravenosa. En el caso de que una respuesta en un sujeto sea insuficiente a las dosis iniciales aplicadas, pueden emplearse dosis superiores (o dosis eficazmente superiores mediante una vía de administración diferente, más localizada) hasta el grado que lo permita la tolerancia del paciente. Se contemplan múltiples dosis por día para lograr niveles sistémicos apropiados de compuestos. Se prefiere generalmente que se use una dosis máxima, es decir, la dosis segura más alta según el criterio médico sensato. Sin embargo, aquellos expertos habituales en la materia entenderán que un paciente puede insistir en una dosis inferior o dosis tolerable por motivos médicos, motivos psicológicos o por prácticamente cualquier otro motivo.

Los agentes descritos en el presente documento pueden combinarse, opcionalmente, con un vehículo farmacéuticamente aceptable para formar una preparación farmacéutica. La expresión "vehículo farmacéuticamente aceptable", como se usa en el presente documento, significa una o más cargas, diluyentes o sustancias de encapsulación sólidas o líquidas compatibles, que son adecuadas para su administración en un ser humano. El término "vehículo" indica un componente orgánico o inorgánico, natural o sintético, con el que se combina el principio activo para facilitar la aplicación. Los componentes de las composiciones farmacéuticas también pueden estar entremezclados con las moléculas de la presente invención, y entre sí, de una manera tal que no haya interacción que afectara sustancialmente la eficacia farmacéutica deseada. En algunos aspectos, las preparaciones

farmacéuticas comprenden un agente descrito en el presente documento en una cantidad eficaz para tratar un trastorno.

Las preparaciones farmacéuticas pueden contener agentes de tamponamiento adecuados, incluyendo: ácido acético en una sal; ácido cítrico en una sal; ácido bórico en una sal; o ácido fosfórico en una sal. Las composiciones farmacéuticas también pueden contener, opcionalmente, conservantes adecuados, tales como: cloruro de benzalconio; clorobutanol; parabenos o timerosal.

Están disponibles una variedad de vías de administración. El modo particular seleccionado dependerá, por supuesto, del fármaco particular seleccionado, la gravedad de la afección que está tratándose y la dosificación requerida para la eficacia terapéutica. Los métodos descritos en el presente documento, de manera general, pueden ponerse en práctica usando cualquier modo de administración que sea médicamente aceptable, que significa cualquier modo que produzca niveles eficaces de los compuestos activos sin provocar efectos adversos clínicamente inaceptables. Tales modos de administración incluyen vías oral, rectal, tópica, nasal, intradérmica, transdérmica o parenteral. El término "parenteral" incluye subcutáneo, intravenoso, intraocular, intramuscular o infusión. Las vías intravenosa o intramuscular no son particularmente adecuadas para terapia y profilaxis a largo plazo. Como un ejemplo, pueden formularse composiciones farmacéuticas para el tratamiento agudo de sujetos que tienen una cefalea de tipo migraña de una variedad de formas diferentes y para una variedad de modos de administración incluyendo comprimidos, cápsulas, polvos, supositorios, inyecciones y sprays nasales.

Las preparaciones farmacéuticas pueden presentarse convenientemente en forma de dosificación unitaria y pueden prepararse mediante cualquiera de los métodos bien conocidos en la técnica de la farmacia. Todos los métodos incluyen la etapa de poner el principio activo en asociación con un vehículo que constituye uno o más componentes auxiliares. En general, las composiciones se preparan poniendo de manera uniforme e íntima el compuesto activo en asociación con un vehículo líquido, un vehículo sólido finamente dividido, o ambos, y luego, si es necesario, moldear el producto.

Pueden presentarse composiciones adecuadas para administración oral como unidades diferenciadas, tales como cápsulas, comprimidos, pastillas para chupar, conteniendo cada una una cantidad predeterminada del compuesto activo. Otras composiciones incluyen suspensiones en líquidos acuosos o líquidos no acuosos tales como un jarabe, elixir o una emulsión.

Composiciones adecuadas para administración parenteral comprenden convenientemente una preparación acuosa estéril de un agente de la invención, que es preferiblemente isotónica con la sangre del receptor. Esta preparación acuosa puede formularse según métodos conocidos usando agentes de dispersión y humectantes adecuados y agentes de suspensión. La preparación inyectable estéril también puede ser una disolución o suspensión inyectable estéril en un diluyente o disolvente no tóxico parenteralmente aceptable, por ejemplo, como una disolución en 1,3-butanodiol. Entre los vehículos y disolventes aceptables que pueden emplearse están agua, disolución de Ringer y disolución isotónica de cloruro sódico. Además, se emplean convencionalmente aceites no volátiles estériles como medio disolvente o de suspensión. Para este fin puede emplearse cualquier aceite no volátil suave incluyendo mono o di-glicéridos sintéticos. Además, pueden usarse ácidos grasos tales como ácido oleico en la preparación de inyectables. Pueden encontrarse formulaciones adecuadas para administraciones oral, subcutánea, intravenosa, intramuscular, etc., en Remington's Pharmaceutical Sciences, Mack Publishing Co., Easton, PA.

Puede proporcionarse un método para aumentar la actividad de generación de  $C_{\alpha}$ -formilglicina en una célula. El método implica poner en contacto la célula con una molécula de ácido nucleico aislada (por ejemplo, un ácido nucleico de SEC ID N°: 1), o un producto de expresión de la misma (por ejemplo, un péptido de SEC ID N°: 2), en una cantidad eficaz para aumentar la actividad de generación de  $C_{\alpha}$ -formilglicina en la célula. En realizaciones importantes, el método implica activar el gen de FGE endógeno para aumentar la actividad de generación de  $C_{\alpha}$ -formilglicina en la célula. En algunas realizaciones, la puesta en contacto se realiza en condiciones que permiten la entrada de una molécula descrita en el presente documento en la célula.

La expresión "permitir la entrada" de una molécula en una célula tiene los siguientes significados dependiendo de la naturaleza de la molécula. Para un ácido nucleico aislado pretende describir la entrada del ácido nucleico a través de la membrana celular y en el núcleo de la célula, tras lo cual el "transgén de ácido nucleico" puede utilizar la maquinaria de la célula para producir polipéptidos funcionales codificados por el ácido nucleico. Por "transgén de ácido nucleico" se pretende describir todos los ácidos nucleicos descritos en el presente documento con o sin los vectores asociados. Para un polipéptido, se pretende describir la entrada del polipéptido a través de la membrana celular y al interior del citoplasma celular, y si es necesario, la utilización de la maquinaria citoplasmática de la célula para modificar funcionalmente el polipéptido (por ejemplo, para dar una forma activa).

Pueden emplearse diversas técnicas para introducir los ácidos nucleicos descritos en el presente documento en células, dependiendo de si los ácidos nucleicos se introducen *in vitro* o *in vivo* en un huésped. Tales técnicas incluyen transfección de precipitados de ácido nucleico- $CaPO_4$ , transfección de ácidos nucleicos asociados con DEAE, transfección con un retrovirus que incluye el ácido nucleico de interés, transfección mediada por liposoma y similares. Para determinados usos, se prefiere dirigir el ácido nucleico a células particulares. En tales casos, un vehículo usado para administrar un ácido nucleico descrito en el presente documento al interior de una célula (por

ejemplo, un retrovirus, u otro virus; un liposoma) puede tener una molécula de direccionamiento unida al mismo. Por ejemplo, una molécula tal como un anticuerpo específico para una proteína de membrana superficial en la célula diana o un ligando para un receptor en la célula diana puede unirse al, o incorporarse dentro del, vehículo de administración de ácido nucleico. Por ejemplo, cuando se emplean liposomas para suministrar los ácidos nucleicos descritos en el presente documento, pueden incorporarse proteínas que se unen a una proteína de membrana superficial asociada con la endocitosis en la formulación de liposoma para dirigir y/o para facilitar la captación. Tales proteínas incluyen proteínas de la cápside o fragmentos de las mismas tróficos para un tipo de célula particular, anticuerpos para proteínas que experimentan internalización en la ciclación, proteínas que seleccionan como diana una localización intracelular y potencian la semivida intracelular, y similares. También se han usado satisfactoriamente sistemas de administración poliméricos para administrar ácidos nucleicos al interior de células, como conocen aquellos expertos en la materia. Tales sistemas permiten incluso la administración oral de ácidos nucleicos.

Otros sistemas de administración pueden incluir sistemas de administración de liberación en el tiempo, liberación retrasada o liberación sostenida. Tales sistemas pueden evitar administraciones repetidas de un agente de la presente invención, aumentando la conveniencia para el sujeto y el médico. Están disponibles muchos tipos de sistemas de administración de liberación y son conocidos por aquellos expertos habituales en la técnica. Incluyen sistemas basados en polímero tales como poli(lactida-glicolida), copolioxalatos, policaprolactonas, poliesteramidas, poliortoésteres, poli(ácido hidroxibutírico) y polianhídridos. Se describen microcápsulas de los polímeros anteriores que contienen fármacos, por ejemplo, en la patente de EE.UU. 5.075.109. Los sistemas de administración también incluyen sistemas no poliméricos que son: lípidos incluyendo esteroides tales como colesterol, ésteres de colesterol y ácidos grasos o grasas neutras tales como mono-, di- y tri-glicéridos; sistemas de liberación de hidrogel; sistemas silásticos; sistemas basados en péptidos; recubrimientos de cera; comprimidos usando aglutinantes y excipientes convencionales; implantes parcialmente fundidos; y similares. Ejemplos específicos incluyen, pero no se limitan a: (a) sistemas de erosión en los que un agente descrito en el presente documento está contenido en una forma dentro de una matriz tal como aquellas descritas en las patentes de EE.UU. N° 4.452.775, 4.675.189 y 5.736.152, y (b) sistemas de difusión en los que un componente activo penetra a una tasa controlada desde un polímero tal como se describe en las patentes de EE.UU. N° 3.854.480, 5.133.974 y 5.407.686. Además, pueden usarse sistemas de administración de hardware basados en bombas, algunos de los cuales están adaptados para implantación.

Puede ser deseable el uso de un implante de liberación sostenida a largo plazo. La liberación a largo plazo, como se usa en el presente documento, significa que el implante se construye y dispone para administrar niveles terapéuticos del principio activo durante al menos 30 días, y preferiblemente 60 días. Aquellos expertos habituales en la técnica conocen bien implantes de liberación sostenida a largo plazo e incluyen algunos de los sistemas de liberación descritos anteriormente. Ejemplos específicos incluyen, pero no se limitan a, implantes de liberación sostenida a largo plazo descritos en la patente de EE.UU. N° 4.748.024, y la patente canadiense N° 1330939.

Administración, y en algunas realizaciones, la coadministración de agentes distintos de las moléculas de FGE descritas en el presente documento que cuando se administran en cantidades eficaces pueden actuar de manera cooperativa, aditiva o sinérgica con una molécula de FGE descrita en el presente documento para: (i) modular la actividad de generación de C<sub>α</sub>-formilglicina, y (ii) tratar cualquiera de las afecciones en las que participa la actividad de generación de C<sub>α</sub>-formilglicina de una molécula de la invención (por ejemplo, una deficiencia de sulfatasa incluyendo MSD). Agentes distintos de las moléculas de FGE descritas en el presente documento incluyen iduronato 2-sulfatasa, sulfamidasa, N-acetilgalactosamina 6-sulfatasa, N-acetilglucosamina 6-sulfatasa, arilsulfatasa A, arilsulfatasa B, arilsulfatasa C, arilsulfatasa D, arilsulfatasa E, arilsulfatasa F, arilsulfatasa G, HSulf-1, HSulf-2, HSulf-3, HSulf-4, HSulf-5 o HSulf-6, (ácidos nucleicos y polipéptidos, y/o fragmentos de los mismos), y/o combinaciones de los mismos.

“Co-administrar”, como se usa en el presente documento, se refiere a administrar simultáneamente dos o más compuestos de la invención (por ejemplo, un ácido nucleico y/o polipéptido de FGE, y un agente que se sabe que es beneficioso en el tratamiento de, por ejemplo, una deficiencia de sulfatasa (por ejemplo, iduronato 2-sulfatasa en el tratamiento de MPSII), como mezcla en una única composición, o secuencialmente, lo bastante próximos en el tiempo para que los compuestos puedan ejercer un efecto aditivo o incluso sinérgico.

Pueden proporcionarse matrices de moléculas de ácido nucleico en fase sólida. La matriz consiste esencialmente en un conjunto de moléculas de ácido nucleico, productos de expresión de las mismas, o fragmentos (o bien del ácido nucleico o bien de la molécula de polipéptido) de los mismos, seleccionándose cada molécula de ácido nucleico del grupo que consiste en FGE, iduronato 2-sulfatasa, sulfamidasa, N-acetilgalactosamina 6-sulfatasa, N-acetilglucosamina 6-sulfatasa, arilsulfatasa A, arilsulfatasa B, arilsulfatasa C, arilsulfatasa D, arilsulfatasa E, arilsulfatasa F, arilsulfatasa G, HSulf-1, HSulf-2, HSulf-3, HSulf-4, HSulf-5 y HSulf-6, fijado a un sustrato sólido. En algunas realizaciones, la matriz en fase sólida comprende además al menos una molécula de ácido nucleico de control. En determinadas realizaciones, el conjunto de moléculas de ácido nucleico comprende al menos una, al menos dos, al menos tres, al menos cuatro, o incluso al menos cinco moléculas de ácido nucleico, seleccionándose cada una del grupo que consiste en FGE, iduronato 2-sulfatasa, sulfamidasa, N-acetilgalactosamina 6-sulfatasa, N-acetilglucosamina 6-sulfatasa, arilsulfatasa A, arilsulfatasa B, arilsulfatasa C, arilsulfatasa D, arilsulfatasa E, arilsulfatasa F, arilsulfatasa G, HSulf-1, HSulf-2, HSulf-3, HSulf-4, HSulf-5 y HSulf-6. En realizaciones preferidas, el

conjunto de moléculas de ácido nucleico comprende un número máximo de 100 moléculas de ácido nucleico diferentes. En realizaciones importantes, el conjunto de moléculas de ácido nucleico comprende un número máximo de 10 moléculas de ácido nucleico diferentes.

5 Se utilizan técnicas de hibridación estándar de tecnología de micromatrices para evaluar patrones de expresión de ácido nucleico e identificar la expresión de ácido nucleico. La tecnología de micromatrices, que también se conoce por otros nombres incluyendo: tecnología de chip de ADN, tecnología de chip génico y tecnología de matriz de ácido nucleico en fase sólida, es muy conocida por aquellos expertos habituales en la técnica y se basa en, pero no se limita a, obtener una matriz de sondas de ácido nucleico identificadas (por ejemplo, moléculas descritas en otra parte  
10 en el presente documento tales como de FGE, iduronato 2-sulfatasa, sulfamidasa, N-acetilgalactosamina 6-sulfatasa, N-acetilglucosamina 6-sulfatasa, arilsulfatasa A, arilsulfatasa B, arilsulfatasa C, arilsulfatasa D, arilsulfatasa E, arilsulfatasa P, arilsulfatasa G, HSulf-1, HSulf-2, HSulf-3, HSulf-4, HSulf-5 y/o HSulf-6) sobre un sustrato fijo, marcar moléculas diana con moléculas indicadoras (por ejemplo, mascas radiactivas, quimioluminiscentes o fluorescentes tales como fluoresceína, Cy3-dUTP o Cy5-dUTP), hibridar ácidos nucleicos  
15 diana con las sondas, y evaluar la hibridación diana-sonda. En general, una sonda con una secuencia de ácido nucleico que se corresponde perfectamente con la secuencia diana producirá la detección de una señal de molécula indicadora más fuerte que sondas con correspondencias menos perfectas. Se presentan muchos componentes y técnicas utilizadas en la tecnología de micromatrices de ácido nucleico en The Chipping Forecast, Nature Genetics, Vol. 21, enero de 1999.

20 Los sustratos de micromatriz pueden incluir, pero no se limitan a, vidrio, sílice, aluminosilicatos, borosilicatos, óxidos metálicos tales como alúmina y óxido de níquel, diversas arcillas, nitrocelulosa o nailon. En todas las realizaciones se prefiere un sustrato de vidrio. Se seleccionan sondas del grupo de ácidos nucleicos incluyendo, pero sin limitarse a: ADN, ADN genómico, ADNc y oligonucleótidos; y pueden ser naturales o sintéticas. Las sondas de oligonucleótidos son preferiblemente oligonucleótidos 20 a 25 meros y las sondas de ADN/ADNc tienen preferiblemente 500 a 5000 bases de longitud, aunque pueden usarse otras longitudes. Un experto habitual en la materia puede determinar una longitud de sonda apropiada siguiendo procedimientos conocidos en la técnica. En una realización, sondas preferidas son conjuntos de dos o más de las moléculas de ácido nucleico expuestas como SEC ID N°: 1, 3, 4, 6, 8, 10 y/o 12. Pueden purificarse sondas para eliminar contaminantes usando métodos  
25 estándar conocidos por aquellos expertos habituales en la técnica tales como filtración en gel o precipitación.

30 En una realización, puede recubrirse el sustrato de micromatriz con un compuesto para potenciar la síntesis de la sonda sobre el sustrato. Tales compuestos incluyen, pero no se limitan a, oligoetilenglicoles. En otra realización, pueden usarse grupos o agentes de acoplamiento sobre el sustrato para enlazar covalentemente el imprimación nucleótido u oligonucleótido al sustrato. Estos agentes o grupos pueden incluir, pero no se limitan a: grupos amino, hidroxilo, bromo y carboxilo. Estos grupos reactivos se unen preferiblemente al sustrato mediante un radical hidrocarbilo tal como un radical divalente alquileo o fenileno, ocupándose una posición de valencia por la unión a la cadena y uniéndose la restante a los grupos reactivos. Estos grupos hidrocarbilo pueden contener hasta aproximadamente diez átomos de carbono, preferiblemente hasta aproximadamente seis átomos de carbono. Normalmente se prefieren radicales alquileo que contienen dos a cuatro átomos de carbono en la cadena principal. Estos y otros detalles del proceso se divulgan, por ejemplo, en la patente de EE.UU. 4.458.066.

35 En una realización, se sintetizan directamente sondas sobre el sustrato en un patrón de rejilla predeterminado usando métodos tales como síntesis química dirigida por luz, desprotección fotoquímica o administración de precursores de nucleótidos al sustrato y posterior producción de sonda.

40 En otra realización, el sustrato puede recubrirse con un compuesto para potenciar la unión de la sonda al sustrato. Tales compuestos incluyen, pero no se limitan a: polilisina, aminosilanos, silanos amino-reactivos (Chipping Forecast, 1999) o cromo (Gwynne y Page, 2000). En esta realización, se aplican sondas previamente sintetizadas al sustrato en un patrón de rejilla y volumen preciso, predeterminado, utilizando un robot controlado por ordenador para aplicar la sonda al sustrato de una manera de impresión por contacto o de una manera sin contacto tal como suministro por chorro de tinta o piezoeléctrico. Pueden enlazarse covalentemente sondas al sustrato con métodos que incluyen, pero no se limitan a, irradiación UV. En otra realización se enlazan sondas al sustrato con calor.

45 Las dianas son ácidos nucleicos seleccionados del grupo que incluye, pero no se limita a: ADN, ADN genómico, ADNc, ARN, ARNm y pueden ser naturales o sintéticas. En todas las realizaciones, se prefieren moléculas de ácido nucleico de sujetos que se sospecha que desarrollan o tienen una deficiencia de sulfatasa. En determinadas realizaciones, una o más moléculas de ácido nucleico de control se unen al sustrato. Preferiblemente, las moléculas de ácido nucleico de control permiten la determinación de factores incluyendo, pero sin limitarse a: calidad de ácido nucleico y características de unión; calidad y eficacia de reactivo; éxito de hibridación; y umbrales y éxito de análisis. Ácidos nucleicos de control pueden incluir, pero no se limitan a, productos de expresión de genes tales como genes de mantenimiento o fragmentos de los mismos.

50 Para seleccionar un conjunto de marcadores de enfermedad de deficiencia de sulfatasa, preferiblemente se analizan los datos de expresión generados, por ejemplo, por análisis de micromatrices de expresión génica, para determinar qué genes en diferentes categorías de pacientes (siendo cada categoría de pacientes un trastorno de deficiencia de



sulfatasa diferente) se expresan de manera significativamente diferenciada. La significación de la expresión génica puede determinarse usando el software informático Permax, aunque puede usarse cualquier paquete estadístico convencional que pueda distinguir diferencias significativas en la expresión. Permax realiza pruebas de la t de 2 muestras con permutación en grandes matrices de datos. Para vectores dimensionales altos de observaciones, el software Permax calcula datos estadísticos de la t para cada atributo, y evalúa la significación usando la distribución de permutación de los atributos globales máximo y mínimo. El uso principal es para determinar los atributos (genes) que son los más diferentes entre dos grupos (por ejemplo, sujeto sano de control y un sujeto con una deficiencia de sulfatasa particular), medir "lo más diferente" usando el valor de los datos estadísticos de la t, y sus niveles de significación.

También puede determinarse la expresión de moléculas de ácido nucleico relacionadas con una enfermedad de deficiencia de sulfatasa usando métodos de medición de proteínas para determinar la expresión de SEC ID N°: 2, por ejemplo, determinando la expresión de polipéptidos codificados por SEC ID N°: 1 y/o 3. Métodos preferidos para medir específica y cuantitativamente proteínas incluyen, pero no se limitan a: métodos basados en espectroscopía de masas tales como desorción/ionización por láser potenciada en superficie (SELDI; por ejemplo, sistema ProteinChip de CIPHERGEN), métodos no basados en espectroscopía de masas, y métodos basados en inmunohistoquímica tales como electroforesis en gel en dos dimensiones.

La metodología SELDI puede usarse, mediante procedimientos conocidos para aquellos expertos habituales en la técnica, para vaporizar cantidades microscópicas de proteína y para crear una "huella" de proteínas individuales, permitiendo así la medición simultánea de la abundancia de muchas proteínas en una única muestra. Preferiblemente pueden utilizarse ensayos basados en SELDI para caracterizar deficiencia múltiple de sulfatasas, así como estadios de tales afecciones. Tales ensayos incluyen preferiblemente, pero no se limitan a, los siguientes ejemplos. Pueden medirse selectivamente productos génicos descubiertos mediante micromatrices de ARN mediante captura específica (mediada por anticuerpos) en el disco de proteína de SELDI (por ejemplo, SELDI selectiva). Pueden resolverse productos génicos descubiertos mediante cribado de proteínas (por ejemplo, con geles en 2-D), mediante "SELDI de proteína total" optimizada para visualizar aquellos marcadores particulares de interés de entre SEC ID N°: 1, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26 y/o 28. Pueden utilizarse modelos predictivos de una deficiencia de sulfatasa específica de medición de SELDI de múltiples marcadores de entre SEC ID N°: 1, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26 y/o 28, para las estrategias de SELDI.

El uso de cualquiera de los métodos de micromatriz anteriores para determinar la expresión de ácidos nucleicos relacionados con una enfermedad de deficiencia de sulfatasa puede hacerse con métodos rutinarios conocidos por aquellos expertos habituales en la técnica y la expresión determinada mediante métodos de medición de proteínas puede correlacionarse con niveles predeterminados de un marcador usado como método de pronóstico para seleccionar estrategias de tratamiento para pacientes con enfermedad de deficiencia de sulfatasa.

Puede proporcionarse una célula productora de sulfatasa, en la que se aumenta la razón de sulfatasa activa con respecto a sulfatasa total producida (es decir, la actividad específica) por la célula. La célula comprende: (i) una sulfatasa con una expresión aumentada, y (ii) una enzima generadora de formilglicina con una expresión aumentada, en la que la razón de sulfatasa activa con respecto a sulfatasa total producida por la célula se aumenta en al menos el 5 % con respecto a la razón de sulfatasa activa con respecto a la sulfatasa total producida por la célula en ausencia de la enzima generadora de formilglicina.

Una "sulfatasa con expresión aumentada", como se usa en el presente documento, se refiere normalmente a expresión aumentada de una sulfatasa y/o su polipéptido codificado en comparación con un control. La expresión aumentada se refiere a aumentar (es decir, hasta un grado detectable) la replicación, transcripción y/o traducción de cualquiera de los ácidos nucleicos de sulfatasa (ácidos nucleicos de sulfatasa como se describen en otra parte en el presente documento), ya que la regulación por incremento de cualquiera de esos procesos produce un aumento de la concentración/cantidad del polipéptido codificado por el gen (ácido nucleico). Esto puede lograrse usando varios métodos conocidos en la técnica, también descritos en otra parte en el presente documento, tal como transfección de una célula con el ADNc de sulfatasa, y/o ADN genómico que engloba el locus de sulfatasa, activación del gen de sulfatasa endógeno colocando, por ejemplo, un elemento promotor fuerte en la dirección 5' del locus genómico del gen de sulfatasa endógeno usando recombinación homóloga (véase, por ejemplo, la tecnología de activación génica descrita en detalle en las patentes de EE.UU. N° 5.733.761, 6.270.989 y 6.565.844), etc. Un control típico sería una célula idéntica transfectada con un(os) plásmido(s) de vector. Potenciar (o aumentar) la actividad de sulfatasa también se refiere a prevenir o inhibir la degradación de sulfatasa (por ejemplo, mediante ubiquitinización aumentada), regulación por disminución, etc., dando como resultado, por ejemplo,  $t_{1/2}$  (semivida) de molécula de sulfatasa aumentada o estable en comparación con un control. Regulación por disminución o expresión disminuida se refiere a expresión disminuida de un gen y/o su polipéptido codificado. La regulación por incremento o regulación por disminución de la expresión génica puede determinarse directamente detectando un aumento o disminución, respectivamente, en el nivel de ARNm para el gen (por ejemplo, una sulfatasa), o el nivel de expresión de proteína del polipéptido codificado por el gen, usando cualquier medio adecuado conocido en la técnica, tal como hibridación de ácido nucleico o métodos de detección de anticuerpos, respectivamente, y en comparación con controles. La regulación por incremento o regulación por disminución de la expresión de genes de sulfatasa también puede determinarse indirectamente detectando un cambio en la actividad de sulfatasa.

Similarmente, una “enzima generadora de formilglicina con una expresión aumentada”, como se usa en el presente documento, se refiere normalmente a expresión aumentada de un ácido nucleico de FGE de la invención y/o su polipéptido codificado en comparación con un control. Expresión aumentada se refiere a aumentar (es decir, hasta un grado detectable) la replicación, transcripción y/o traducción de cualquiera de los ácidos nucleicos de FGE (como se describe en otra parte en el presente documento), ya que la regulación por incremento de cualquiera de esos procesos produce un aumento de la concentración/cantidad del polipéptido codificado por el gen (ácido nucleico). Esto puede lograrse usando los métodos descritos anteriormente (para las sulfatasas), y en otra parte en el presente documento.

En determinadas realizaciones, la razón de sulfatasa activa con respecto a sulfatasa total producida por la célula se aumenta en al menos el 10 %, el 15 %, el 20 %, el 50 %, el 100 %, el 200 %, el 500 %, el 1000 %, con respecto a la razón de sulfatasa activa con respecto a sulfatasa total producida por la célula en ausencia de la enzima generadora de formilglicina.

Puede proporcionarse un método mejorado para tratar una deficiencia de sulfatasa en un sujeto. El método implica administrar a un sujeto en necesidad de tal tratamiento una sulfatasa en una cantidad eficaz para tratar la deficiencia de sulfatasa en el sujeto, en el que la sulfatasa se pone en contacto con una enzima generadora de formilglicina en una cantidad eficaz para aumentar la actividad específica de la sulfatasa. Como se describe en otra parte en el presente documento, “actividad específica” se refiere a la razón de sulfatasa activa con respecto a sulfatasa total producida. “Poner en contacto”, tal como se usa en el presente documento, se refiere a modificar postraduccionalmente mediante FGE la sulfatasa como se describe en otra parte en el presente documento. Será evidente para un experto habitual en la materia que una FGE puede entrar en contacto con una sulfatasa y modificarla si los ácidos nucleicos que codifican FGE y una sulfatasa se co-expresan en una célula, o incluso si un polipéptido de FGE aislado entra en contacto con un polipéptido de sulfatasa aislado *in vivo* o *in vitro*. Aunque puede co-administrarse un polipéptido de FGE aislado con un polipéptido de sulfatasa aislado a un sujeto para tratar una deficiencia de sulfatasa en el sujeto, se prefiere que el contacto entre FGE y la sulfatasa tenga lugar *in vitro* antes de la administración de la sulfatasa al sujeto. Este método de tratamiento mejorado es beneficioso para un sujeto, ya que necesitan administrar cantidades menores de la sulfatasa, y/o con menos frecuencia, ya que la sulfatasa tiene una actividad específica superior.

La invención se entenderá más completamente mediante referencia a los siguientes ejemplos.

## Ejemplos

### Ejemplo 1:

*La deficiencia múltiple de sulfatasas está producida por mutaciones en el gen que codifica la enzima generadora de C $\alpha$ -formilglicina (FGE) humana*

### Procedimientos experimentales

#### Materiales y métodos

#### **Ensayo *in vitro* para FGE**

Para monitorizar la actividad FGE, se usó el péptido P23 23-mero N-acetilado y C-amidado (MTDFYVPVSLCTPSRAALLTGRS) (SEC ID N $^{\circ}$ : 33) como sustrato. Se monitorizó la conversión del residuo de cisteína en la posición 11 en *FGly* mediante espectrometría de masas MALDI-TOF. Se preparó una disolución madre 6  $\mu$ M de P23 en acetonitrilo al 30 % y ácido trifluoroacético al 0,1 % (TFA). En condiciones normales, se incubaron 6 pmoles de P23 a 37 °C con hasta 10  $\mu$ l de enzima en un volumen final de 30  $\mu$ l de Tris/HCl 50 mM, pH 9,0, que contenía NaCl 67 mM, CaCl $_2$  15  $\mu$ M, DTT 2 mM, y albúmina sérica bovina 0,33 mg/ml. Para detener la reacción enzimática, se añadieron 1,5  $\mu$ l de TFA al 10 %. Entonces P23 se unió a ZipTip C18 (Millipore), se lavó con TFA al 0,1 % y se eluyó en 3  $\mu$ l de acetonitrilo al 50 %, TFA al 0,1 %. Se mezclaron 0,5  $\mu$ l del eluato con 0,5  $\mu$ l de disolución de matriz (ácido a-ciano-4-hidroxi-cinámico 5 mg/ml (Bruker Daltonics, Billerica, MA) en acetonitrilo al 50 %, TFA al 0,1 %) en una diana de acero inoxidable. Se realizó la espectrometría de masas MALDI-TOF con un Reflex III (Bruker Daltonics) usando reflectrón y energía láser justo antes del umbral de desorción/ionización. Todos los espectros fueron promedios de 200-300 disparos de varios puntos en la diana. Se calibró el eje de masas usando péptidos de masas moleculares que oscilaban entre 1000 y 3000 Da como patrones externos. MH $^+$  monoisotrópico de P23 es 2526,28 y del producto que contiene *FGly*, 2508,29. Se calculó la actividad (pmol de producto / h) partiendo de la base de la altura pico del producto dividido por la suma de las alturas de pico de P23 y el producto.

#### **Purificación de FGE a partir de testículos bovinos**

Se obtuvieron testículos bovinos del matadero local y se almacenaron durante hasta 20 h en hielo. Se liberó el parénquima del tejido conjuntivo y se homogeneizó en una Waring Blendor y mediante tres rondas de licuado a

motor. Se realizó la preparación de microsomas rugosos (RM) mediante fraccionamiento celular del homogenizado obtenido como se describe (Meyer et al., J. Biol. Chem., 2000, 275:14550-14557) con las siguientes modificaciones. Se realizaron tres etapas de centrifugación diferencial, de 20 minutos cada una a 4 °C, a 500 g (rotor JA10), 3000 g (JA10) y 10000 g (JA20). A partir del último sobrenadante, se sedimentaron las membranas de RM (125000 g, rotor Tí45, 45 min, 4 °C), se homogeneizaron mediante licuado a motor y se estratificaron en un colchón de sacarosa (Hepes 50 mM, pH 7,6, KAc 50 mM, MgAc<sub>2</sub> 6 mM, EDTA 1 mM, sacarosa 1,3 M, β-mercaptoetanol 5 mM). Se recuperaron los RM del sedimento después de centrifugar durante 210 minutos a 45000 rpm en un rotor Tí45 a 4 °C. Normalmente, se obtuvieron 100000-150000 equivalentes de RM, como definen Walter y Blobel (Methods Enzymol., 1983, 96:84-93), de 1 kg de tejido testicular. Se obtuvo el reticuloplasma, es decir el contenido luminal del RM, mediante la extracción diferencial a concentraciones bajas de desoxi Big Chap, como se describe (Fey et al., J. Biol. Chem., 2001, 276:47021-47028). Para la purificación de FGE, se dializaron 95 ml de reticuloplasma durante 20 h a 4 °C frente a Tris/HCl 20 mM, pH 8,0, DTT 2,5 mM, y se aclararon mediante centrifugación a 125000 g durante 1 h. Se cargaron alícuotas de 32 ml del reticuloplasma aclarado en una columna MonoQ HR10/10 (Amersham Biosciences, Piscataway, NJ) a temperatura ambiente, se lavaron y se eluyeron a 2 ml/min con un gradiente lineal de NaCl de 0 a 0,75 M en 80 ml del tampón Tris. Se reunieron las fracciones que contenían actividad de FGE, que eluyeron a NaCl 50-165 mM, de tres rondas (42 ml) y se mezclaron con 2 ml de concanavalina A-Sepharose (Amersham Biosciences) que se había lavado con tampón Hepes 50 mM, pH 7,4, que contenía KCl 0,5 M, MgCl<sub>2</sub> 1 mM, MnCl<sub>2</sub> 1 mM, CaCl<sub>2</sub> 1 mM y DTT 2,5 mM. Después de la incubación durante 16 h a 4 °C, se recogió la concanavalina A-Sepharose en una columna y se lavó con 6 ml del mismo tampón Hepes. Se eluyó el material unido incubando la columna durante 1 h a temperatura ambiente con 6 ml de α-metilmanósido 0,5 M en Hepes 50 mM, pH 7,4, DTT 2,5 mM. Se repitió la elución con 4 ml del mismo eluyente. Se ajustaron los eluatos combinados (10 ml) de concanavalina A-Sepharose a pH 8,0 con Tris/HCl 0,5 M, pH 9,0, y se mezclaron con 2 ml de Affigel 10 (Bio-Rad Laboratories, Hercules, CA) que se había derivatizado con 10 mg del péptido reorganizado al azar (PVSLPTRSCAALLTGR) (SEC ID N°: 34) y se lavaron con tampón A (Hepes 50 mM, pH 8,0, que contenía acetato de potasio 0,15 M, sacarosa 0,125 M, MgCl<sub>2</sub> 1 mM y DTT 2,5 mM). Después de la incubación durante 3 h a 4 °C se recogió la matriz de afinidad en una columna. Se recogieron la fracción no retenida y una de lavado con 4 ml de tampón A, se combinaron y se mezclaron con 2 ml de Affigel 10 que se había sustituido con 10 mg del péptido Ser69 (PVSLSTPSRAALLTGR) (SEC ID N°: 35) y se lavaron con tampón A. Después de la incubación durante la noche a 4 °C, se recogió la matriz de afinidad en una columna, se lavó 3 veces con 6 ml de tampón B (el tampón A que contenía NaCl 2 M y una mezcla de los 20 aminoácidos proteínogénicos, cada uno a 50 mg/ml). Se eluyó el material unido de la matriz de afinidad incubando el Affigel dos veces durante 90 min cada una con 6 ml de tampón B que contenía el péptido Ser69 25 mM. Se sustituyó una alícuota del eluato con albúmina sérica bobina 1 mg/ml, se dializó frente al tampón A y se analizó para su actividad. Se concentró la parte restante de la actividad (11,8 ml) en un concentrador Vivaspín 500 (Vivascience AG, Hannover, Alemania), y se solubilizó a 95 °C en tampón de muestra Laemmli SDS. Se monitorizaron la composición del polipéptido del material de partida y las preparaciones obtenidas después de las etapas cromatográficas mediante SDS-PAGE (15 % de acrilamida, 0,16 % de bisacrilamida) y tinción con SYPRO Ruby (Bio-Rad Laboratories).

#### Identificación de FGE mediante espectrometría de masas

Para el análisis de huella de masa peptídica, los polipéptidos purificados se digirieron en gel con tripsina (Shevchenko et al., Anal. Chem., 1996, 68:850-855), se desalaron en C18 ZipTip y se analizaron mediante espectrometría de masas MALDI-TOF usando ácido dihidrobenzoico como matriz y dos péptidos autolíticos de tripsina (m/z 842,51 y 2211,10) como patrones internos. Para el análisis de espectrometría de masas en tándem se analizaron péptidos seleccionados mediante espectrometría de masas por descomposición metaestable - MALDI-TOF. Se aislaron sus iones doblemente cargados correspondientes y se fragmentaron mediante espectrometría de masas con trampa iónica con nano-ESI fuera de línea (EsquireLC, Bruker Daltonics). Se usaron los datos de la espectrometría de masas mediante el algoritmo de búsqueda Mascot para la identificación de proteínas en la base de datos de proteínas NCBI y la base de datos de nucleótidos NCBI EST.

#### Bioinformática

Se describieron péptidos señal y sitios de escisión con el método de von Heijne (von Heijne, Nucleic Acids Res., 1986, 14:4683-90) implementado en EMBOSS (Rice et al., Trends in Genetics, 2000, 16:276-277). Se predijeron los sitios de N-glicosilación usando el algoritmo de Brunak (Gupta y Brunak, Pac. Symp. Biocomput., 2002, 310-22). Se detectaron los dominios funcionales mediante la búsqueda de modelos ocultos de Markov en PFAM (versión 7.8) (Sonnhammer et al., Nucleic Acids Res., 1998, 26:320-322). Para buscar homólogos de FGE, se consultaron las bases de datos del National Center for Biotechnology Information (Wheeler et al., Nucleic Acids Res., 2002, 20:13-16) con BLAST (Altschul et al., Nucleic Acids Res., 1997, 25:3389-3402). Se calcularon las similitudes de secuencia usando herramientas estándar de EMBOSS. Se determinó la sintenia y la organización genómica de los loci usando fuentes de genoma humano y de ratón de NCBI y el mapa de homologías entre ser humano y ratón, también de NCBI, (Bethesda, MD).

#### Clonación de ADNc de FGE humana

Se sometió a transcripción inversa el ARN total, preparado a partir de fibroblastos humanos usando el kit RNEASY™

Mini (Qiagen, Inc., Valencia, CA) usando el kit OMNISCRIPRT RT™ (Qiagen, Inc., Valencia, CA) y o bien un cebador de oligo(dT) o bien el cebador específico de FGE, 1199nc (CCAATGTAGGTCAGACACG) (SEC ID N°: 36). Se amplificó el ADNc de imprimación hebra mediante PCR usando el cebador directo 1c (ACATGGCCCCGCGGGAC) (SEC ID N°: 37) y, como cebador inverso, o bien 1199nc o bien 1182nc (CGACTGCTCCTTGGACTGG) (SEC ID N°: 38). Se clonaron los productos de la PCR directamente en el vector pCR4-TOPO™ (Invitrogen Corporation, Carlsbad, CA). Mediante secuenciación múltiple de los productos clonados de la PCR, que se habían obtenido de diversos individuos y de reacciones de RT y PCR independientes, se determinó la secuencia codificante del ADNc de FGE (SEC ID N°: 1 y 3).

#### 10 **Detección de mutaciones, secuenciación genómica, mutagénesis dirigida al sitio y análisis de transferencia Northern**

Los protocolos convencionales 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). Las transferencias Northern se hibridaron con una sonda de ADNc que cubre toda la región codificante y una sonda de ADNc de  $\beta$ -actina como control para carga de ARN.

#### **Líneas celulares y cultivo celular**

20 Se obtuvieron fibroblastos de pacientes con MSD 1-6 de E. Christenson (Rigshospitalet Copenhagen), M. Beck (Universitätskinderklinik Mainz), A. Kohlschütter (Universitätskrankenhaus Eppendorf, Hamburgo), E. Zammarchi (Meyer Hospital, Universidad de Florencia), K. Harzer (Institut für Hirnforschung, Universität Tübingen) y A. Fensom (Guy's Hospital, Londres), respectivamente. Se mantuvieron los fibroblastos de piel humana, BHK21, HT-1080 y las células CHO a 37 °C en 5 % de CO<sub>2</sub> en medio Eagle modificado por Dulbecco que contenía 10 % de suero bovino fetal.

#### **Transfección, inmunofluorescencia indirecta, análisis de transferencia Western y detección de actividad FEG**

30 Se dotó el ADNc de FGE con un sitio de EcoRI en 5' y o bien una secuencia de marcador de HA, c-Myc o bien RGS-His<sub>6</sub> en 3', seguido por un codón de terminación y un sitio de HindIII, mediante PCR de adición usando Pfu polimerasa (Stratagene, La Jolla, CA) y los cebadores siguientes: GGAATTCGGGACAACATGGCTGCG (EcoRI) (SEC ID N°: 39), CCCAAGCTTATGCGTAGTCAGGCACATCATACGGATAGTCCATGGTGGGCAGGC (HA) (SEC ID N°: 40), CCCAAGCTTACAGGTCTTCTTCAGAAATCAGCTTTTGTTCGTCATGGTGGGCAG GC (c-Myc) (SEC ID N°: 41), CCCAAGCTT AGT GAT GGT GAT GGT GAT GCGAT C CTCTGTCCATGGTGGGCAGGC (RGS-His<sub>6</sub>) (SEC ID N°: 42). Se clonaron los productos resultantes de PCR como fragmentos de EcoRI/HindIII en pMPSVEH (Artelt et al., Gene, 1988, 68:213-219). Se transfectaron de manera transitoria los plásmidos obtenidos en células HT-1080, BHK21 y CHO, se hicieron crecer sobre portaobjetos, usando EFFECTENE™ (Qiagen) como reactivo de transfección. 48 después de la transfección, se analizaron las células mediante inmunofluorescencia indirecta como se ha descrito previamente (Lübke et al., Nat. Gen., 2001, 28:73-76; Hansske et al., J. Clin. Invest., 2002, 109:725-733), usando anticuerpos monoclonales IgG1 contra HA (Berkeley Antibody Company, Richmond, CA), c-Myc (Santa Cruz Biotechnology, Inc., Santa Cruz, CA) o RGS-His (Qiagen) como anticuerpos primarios. Se detectó la proteína marcadora de retículo endoplásmico, proteína disulfuro isomerasa (PDI) con un anticuerpo monoclonal de diferente subtipo (IgG2A, Stressgen Biotech., Victoria BC, Canadá). Se detectaron los anticuerpos primarios con anticuerpos secundarios de cabra específicos de isotipo acoplados a CY2 o CY3, respectivamente (Molecular Probes, Inc., Eugene, O). Se obtuvieron imágenes de inmunofluorescencia en un microscopio de barrido láser Leica TCS Sp2 AOBS. Para el análisis de transferencia Western se usaron los mismos anticuerpos monoclonales y una IgG anti-ratón conjugada con HRP como anticuerpo secundario. Para la determinación de la actividad de FEG, se lavaron células sometidas a tripsinización con solución salina tamponada con fosfato que contenía una mezcla de inhibidores de proteinasa (clorhidrato de fluoruro de 4-(2-aminoetil)benzenosulfonilo 208  $\mu$ M, aprotinina 0,16  $\mu$ M, leupeptina 4,2  $\mu$ M, bestatina 7,2  $\mu$ M, pepstatina A 3  $\mu$ M, E-64 22,8  $\mu$ M), se solubilizaron en Tris 10 mM, pH 8,0, que contenía DTT 2,5 mM, los inhibidores de proteinasa y Triton X-100 al 1 %, y se aclararon por centrifugación a 125.000 g durante 1 h. Se sometió el sobrenadante a cromatografía en una columna MonoQ PC 1.6/5 usando las condiciones descritas anteriormente. Se reunieron las fracciones que eluyeron en NaCl 50-200 mM, se liofilizaron y se reconstituyeron en un décimo del volumen reunido original antes de la determinación de la actividad de FEG con el péptido P23.

#### **Transducción retroviral**

60 Se clonaron los ADNc de interés en el vector pLPCX y pLNCX2 basado en el virus de la leucemia murina de Moloney (BD Biosciences Clontech, Palo Alto, CA). Se realizó la transfección de células FNX-Eco ecotrópicas (ATCC, Manassas, VA) y la transducción de células RETROPACK™ PT67 anfotróficas (BD Biosciences Clontech) y fibroblastos humanos como se describe (Lübke et al., Nat. Gen., 2001, 28:73-76; Thiel et al., Biochem. J., 2002, 376, 195-201). Para algunos experimentos se seleccionaron las células PT67 transducidas en pLPCX con puomicina antes de la determinación de las actividades de sulfatasa.

## Ensayos de sulfatasa

Se determinó la actividad de ASA, STS y GalNAc6S como se describe en Rommerskirch y von Figura, Proc. Natl. Acad. Sci., USA, 1992, 89:2561-2565; Glossi y Kresse, Clin. Chim. Acta, 1978, 88:111-119.

## Resultados

### Un ensayo basado en péptidos rápido para actividad FEG

Los presentes inventores han desarrollado un ensayo para determinar la actividad FEG en extractos de microsomas usando fragmentos de [<sup>35</sup>S] ASA sintetizados *in vitro* como sustrato. Se añadieron los fragmentos a la mezcla de ensayo como complejos de cadena naciente asociados a ribosoma. La cuantificación del producto incluyó digestión trípica, separación de los péptidos mediante RP-HPLC e identificación y cuantificación de péptido trípico que contenía *FGly* marcada con [<sup>35</sup>S] mediante una combinación de derivatización química a hidrazonas, separación mediante RP-HPLC y recuento mediante centelleo líquido (Fey et al., J. Biol. Chem., 2001, 276: 47021-47028). Para monitorizar la actividad enzimática durante la purificación, fue necesario modificar este procedimiento engorroso. Un péptido sintético 16-mero correspondiente a los residuos 65-80 de ASA y que contenía el motivo de secuencia requerido para la formación de *FGly* inhibió la actividad de FEG en el ensayo *in vitro*. Esto sugirió que péptidos tales como ASA65-80 pueden servir como sustratos para FGE. Los presentes inventores sintetizaron el péptido 23-mero P23 (SEC ID N°: 33), que se corresponde con los residuos 60-80 de ASA con un residuo de metionina N-acetilada y uno de serina C-amidada adicionales para proteger los extremos N y C, respectivamente. La cisteína y las formas que contienen *FGly* de P23 pudieron identificarse y cuantificarse mediante espectrometría de masas por desorción/ionización mediante láser asistida por matriz con analizador de tiempo de vuelo (MALDI-TOF). Se verificó la presencia del residuo de *FGly* en la posición 11 de P23 mediante espectrometría de masas por descomposición metaestable - MALDI-TOF (véase Peng et al., J. Mass Spec., 2003, 38:80-86). La incubación de P23 con extractos de microsomas de páncreas bovino o testículos bovinos convirtió hasta el 95 % del péptido en un derivado que contenía *FGly* (Figura 1). En condiciones normales, la reacción fue proporcional a la cantidad de enzima y el tiempo de incubación siempre que se hubiera consumido menos del 50 % del sustrato y el periodo de incubación no superara las 24 h. La  $k_m$  para P23 fue de 13 nM. Los efectos del glutatión reducido y oxidado,  $Ca^{2+}$  y pH fueron comparables a los observados en el ensayo usando complejos de cadena naciente asociados a ribosoma como sustrato (Fey et al., J. Biol. Chem., 2001, 276:47021-47028).

### Purificación de FGE

Para la purificación de FGE, la fracción soluble (reticuloplasma) de microsomas de testículos bovinos sirvió como material de partida. La actividad específica de FGE fue 10-20 veces superior a la que tiene en reticuloplasma de microsomas de páncreas bovino (Fey et al., J. Biol. Chem., 2001, 276:47021-47028). Se logró la purificación de FGE mediante una combinación de cuatro etapas cromatográficas. Las dos primeras etapas fueron cromatografía en un intercambiador aniónico MonoQ y concavalina A-Sepharose. A pH 8, se unió la actividad FEG a MonoQ y se eluyó en NaCl 50-165 mM con una recuperación del 60-90 %. Cuando esta fracción se mezcló con concavalina A-Sepharose, se unió FGE. Pudo eluirse el 30-40 % de la actividad de partida con a-metilmanósido 0,5 M. Las dos etapas de purificación finales fueron cromatografía en matrices de afinidad derivatizadas con péptidos 16-meros. La imprimación matriz de afinidad fue Affigel 10 sustituida con una VARIANTEe del péptido ASA65-80, en el que los residuos Cys69, Pro71 y Arg73, críticos para la formación de *FGly*, estaban reorganizados al azar (péptido reorganizado al azar PVSLPTRSCAALLTGR -SEC ID N°: 34). Este péptido no inhibió la actividad FEG cuando se añadió a una concentración 10 mM al ensayo *in vitro* y, cuando se inmovilizó en Affigel 10, no retuvo la actividad FEG. La cromatografía en la matriz de afinidad de péptido reorganizado eliminó las proteínas de unión al péptido incluyendo las chaperonas del retículo endoplásmico. La segunda matriz de afinidad fue Affigel 10 sustituida con una VARIANTEe del péptido ASA65-80, en el que se sustituyó la Cys69 por una serina (péptido Ser69 PVSLSTPSRAALLTGR-SEC ID N°: 35). La matriz de afinidad del péptido Ser69 unió de manera eficaz FGE. La actividad de FEG pudo eluirse o bien con KSCN 2 M o bien con péptido Ser69 25 mM con una recuperación del 20-40 %. Antes de la determinación de la actividad, tuvo que eliminarse el KSCN o el péptido Ser69 mediante diálisis. La sustitución de Cys69 por serina fue crucial para la elución de FGE activa. Affigel 10 sustituido con el péptido ASA65-80 no mutante unió FGE de manera eficaz. Sin embargo, casi no pudo recuperarse ninguna actividad en eluatos con sales caotrópicas (KSCN,  $MgCl_2$ ), péptidos (ASA65-80 o péptido Ser69) o tampones con pH bajo o alto. En la Fig. 2 se muestra el patrón de polipéptido del material de partida y de las fracciones activas obtenidas después de las cuatro etapas cromatográficas de una purificación típica. En la fracción final se recuperó el 5 % de la actividad FEG de partida y el 0,0006 % de la proteína de partida (purificación de 8333 veces).

### Los polipéptidos purificados de 39,5 y 41,5 kDa están codificados por un único gen

Se sometieron a análisis de huella peptídica los polipéptidos de 39,5 y 41,5 kDa en la preparación de FGE purificada. Los espectros de masas de los péptidos tríplicos de los dos polipéptidos obtenidos mediante espectrometría de masas MALDI-TOF fueron en su mayor parte solapantes, sugiriendo que las dos proteínas se originan a partir del mismo gen. Entre los péptidos tríplicos de ambos polipéptidos se encontraron dos péptidos abundantes  $MH^+$  1580,73, SQNTPDSSASNLGFR (SEC ID N°: 43) y  $MH^+$  2049,91, MVPIPAGVFTMGTDPPQIK -

SEC ID N°: 44 más dos oxidaciones de metionina), que correspondían con la proteína codificada por un ADNc con N° de acceso de GenBank AK075459 (SEC ID N°: 4). Se confirmó la secuencia de aminoácidos de los dos péptidos mediante espectros de descomposición metaestable - MALDI-TOF y mediante análisis de EM/EM usando espectrometría de masas con trampa iónica con ionización por nano-electropulverización (ESI) fuera de línea. Una secuencia de EST del ortólogo bovino del ADNc humano que cubre la parte del extremo C de la FGE y que se corresponde con las secuencias de ambos péptidos proporcionó la información de secuencia adicional para la FGE bovina.

### Conservación evolutiva y estructura de dominio de FGE

El gen para la FGE humana está codificado por el ADNc de (SEC ID N°: 1 y/o 3) y está localizado en el cromosoma 3p26. Abarca ~105 kb y la secuencia codificante se distribuye a lo largo de 9 exones. Se encontraron tres ortólogos del gen de FGE humana en ratón (identidad del 87 %), *Drosophila melanogaster* (identidad del 48 %) y *Anopheles gambiae* (identidad del 47 %). Se encontraron secuencias EST ortólogas para 8 especies adicionales incluyendo vaca, cerdo, *Xenopus laevis*, *Silurana tropicalis*, pez cebra, salmón y otras especies de peces (para más detalles véase el Ejemplo 2). La estructura exón-intrón entre el gen humano y el de ratón está conservada y el gen de ratón en el cromosoma 6E2 está localizado dentro de una región sintética con respecto a la del cromosoma 3p26 humano. Los genomas de *S. cerevisiae* y *C. elegans* carecen de homólogos de FGE. En procariontes se encontraron 12 homólogos de FGE humana. Se predijo que el ADNc para la FGE humana codificaba para una proteína de 374 residuos (Fig. 3 y SEC ID N°: 2). La proteína contiene una secuencia señal escindible de 33 residuos, que indica translocación de FGE al interior del retículo endoplásmico, y contiene un único sitio de N-glicosilación en Asn141. La unión de FGE a concanavalina A sugiere que se utiliza este sitio de N-glicosilación. Los residuos 87-367 de FGE se enumeran en la base de datos de motivos de proteínas PFAM como un dominio de función desconocida (PFAM: DUF323). El análisis de comparación de secuencias de FGE humana y sus ortólogos eucariotas identificados en bases de datos indica que este dominio está compuesto por tres subdominios distintos.

El subdominio del extremo N (residuos 91-154 en FGE humana) tiene una identidad de secuencia del 46 % y una similitud del 79 % dentro de los cuatro ortólogos de FGE eucariotas conocidos. En la FGE humana, este dominio porta el sitio de N-glicosilación en Asn 141, que está conservado en los otros ortólogos. La parte media de FGE (residuos 179-308 en FGE humana) está representada por un subdominio rico en triptófano (12 triptófanos por 129 residuos). La identidad de los ortólogos eucariotas dentro de este subdominio es del 57 %, la similitud es del 82 %. El subdominio del extremo C (residuos 327-366 en FGE humana) es la secuencia más altamente conservada dentro de la familia de FGE. La identidad de secuencia del subdominio del extremo C humano con los ortólogos eucariotas (3 secuencias de longitud completa y 8 EST) es del 85 %, la similitud del 97 %. Dentro de los 40 residuos del subdominio 3, cuatro residuos de cisteína están completamente conservados. Tres de las cisteínas también están conservadas en los ortólogos de FGE procariontes. Los 12 miembros procariontes de la familia de FGE (para más detalles véase el Ejemplo 2) comparten la estructura de subdominios con los FGE eucariotas. Los límites entre los tres subdominios son más evidentes en la familia procarionte de FGE debido a secuencias no conservadas de longitudes variables que separan los subdominios entre sí. El genoma humano y de ratón codifican dos homólogos estrechamente relacionados de FGE (SEC ID N°: 43 y 44, N° de acceso de GenBank NM\_015411, en el hombre, y SEC ID N°: 45 y 46, N° de acceso de GenBank AK076022, en ratón). Los dos parálogos son idénticos en un 86 %. Sus genes están localizados en regiones sintéticas del cromosoma (7q11 en ser humano, 5G1 en ratón). Ambos parálogos comparten con los ortólogos de FGE la estructura de subdominios y son idénticos en un 35 % y similares en un 47 % con respecto a la FGE humana. En el tercer subdominio, que es idéntico en un 100 % en ambos homólogos, falta la secuencia undecamérica que contiene cisteína del subdominio 3.

### Expresión, localización subcelular y formas moleculares

Es detectarse un único transcrito de 2,1 kb mediante análisis de transferencia Northern del ARN total de fibroblastos de piel y ARN-poli A<sup>+</sup> de corazón, cerebro, placenta, pulmón, hígado, músculo esquelético, riñón y páncreas. En relación con el ARN de  $\beta$ -actina, la abundancia varía en un orden de magnitud y es la más alta en páncreas y riñón y la más baja en cerebro. Se ensayaron diversas líneas de células eucariotas que expresan de manera estable o transitoria el ADNc de FGE humana o derivados de FGE prolongados en el extremo C por un marcador de HA, Myc o His<sub>6</sub> para actividad de FEG y la localización subcelular de FGE. La expresión transitoria de FGE marcada y no marcada aumentó la actividad FEG 1,6 - 3,9 veces. La expresión estable de FGE en células PT67 aumentó la actividad FGE en aproximadamente 100 veces. La detección de la forma de FGE marcada mediante inmunofluorescencia indirecta en células BHK 21, CHO y HT1080 mostró una co-ubicación de las formas de FGE marcadas de manera diversa con la proteína disulfuro isomerasa, una proteína luminal del retículo endoplásmico. El análisis de transferencia Western de los extractos de células BHK 21 transfectadas de manera transitoria con ADNc que codifica las formas marcadas de FGE mostró una única banda inmunorreactiva con un tamaño aparente de entre 42 y 44 kDa.

### El gen de FGE porta mutaciones en MSD

La MSD se produce por una deficiencia para generar residuos de *FGly* en sulfatasas (Schmidt, B., et al., Cell, 1995, 82: 271-278). El gen de FGE es, por tanto, un gen candidato para la MSD. Los presentes inventores amplificaron y

secuenciaron el ADNc que codifica FGE de siete pacientes con MSD y encontraron diez mutaciones diferentes que se confirmaron mediante secuenciación del ADN genómico (Tabla 1).

**Tabla 1: Mutaciones en pacientes con MSD**

Mutación	Efecto en proteína	Observaciones	Paciente
1076C>A	S359X	El truncamiento de la C-terminal de 16 residuos	1*
IVS3+5-8 del	La delección de residuos 149-173	En marco delección del exón 3	1,2
979C>T	R327X	Pérdida de subdominio 3	2
1045C>T	R349W	La sustitución de un residuo conservado en subdominio 3	3,7
1046G>A	R349Q	La sustitución de un residuo conservado en subdominio 3	4
1006T>C	C336R	La sustitución de un residuo conservado en subdominio 3	4
836C>T	A279V	La sustitución de un residuo conservado en subdominio 3	5
243delC	desplazamiento del marco y el truncamiento	La pérdida de los tres subdominios	5
661delG	desplazamiento del marco y el truncamiento	Pérdida de la tercera C-terminal de FGE incluyendo subdominio 3	6**
IVS6-1G>A	La delección de residuos 281-318	En marco delección del exón 7	5
<p>* Paciente 1 es el paciente MSD Mo. en Schmidt, B., et al, Cell, 1 995, 82:. 271-278 y Rommerskirch y von Figura, Proc. Natl. Acad. Sci, EE.UU., 1992, 89:. 2561-2565.  ** Paciente 6 es el paciente MSD reportado por Burk et al, J. Pediatr, 1984, 104:... 574-578.</p>			

Los otros pacientes representan casos no publicados.

El imprimación paciente era heterocigoto para una sustitución 1076C>A que convierte el codón para serina 359 en un codón de terminación (S359X) y una mutación que produce la delección de los 25 residuos 149-173 que están codificados por el exón 3 y separan los dominios imprimación y segundo de la proteína. La secuenciación genómica reveló una delección de nucleótidos +5-8 del tercer intrón (IVS3+5-8 del), destruyéndose así el sitio donador de corte y empalme del intrón 3. El segundo paciente era heterocigoto para la mutación que produce la pérdida del exón 3 (IVS3+5-8 del) y una sustitución 979C>T que convierte el codón para arginina 327 en un codón de terminación (R327X). La FGE truncada codificada por el alelo 979C>T carece de la mayor parte del subdominio 3. El tercer paciente era homocigoto para una sustitución 1045T que sustituye la arginina conservada 349 en el subdominio 3 por triptófano (R349W). El cuarto paciente era heterocigoto para dos mutaciones de aminoácido que reemplazan residuos conservados en el dominio de FGE: una sustitución 1046>T que reemplaza la arginina 349 por glutamina (R349Q) y una sustitución 1006T>C que reemplaza la cisteína 336 por arginina (C336R). El quinto paciente era heterocigoto para una sustitución 836 C>T que reemplaza la alanina conservada 279 por valina (A279V). La segunda mutación es una delección de único nucleótido (243delC) que cambia la secuencia después de la prolina 81 y que produce una detención de la traducción después del residuo 139. El sexto paciente era heterocigoto para la delección de un único nucleótido (661delG) que cambia la secuencia de aminoácidos tras el residuo 220 y que introduce un codón de terminación tras el residuo 266. La segunda mutación es una mutación del sitio aceptor de corte y empalme del intrón 6 (IVS6-1G>A) que produce una delección en marco del exón 7 que codifica los residuos 281-318. En el séptimo paciente se encontró la misma sustitución 1045C>T que en el tercer paciente. Además se detectaron dos polimorfismos en la región codificante de 18 alelos de FGE de los controles y pacientes con MSD. El 22 % portaba una sustitución 188G>A, que reemplaza la serina 63 por asparagina (S63N) y el 28 % una sustitución silenciosa 1116C>T.

#### Transducción de fibroblastos con MSD con ADNc de FGE no mutante y mutante

Con el fin de confirmar la deficiencia de FGE como la causa de la inactividad de las sulfatasas sintetizadas en la MSD, se expresó el ADNc de FGE en fibroblastos con MSD utilizando transferencia de genes retrovirales. Como control, los presentes inventores transdujeron el vector retroviral sin inserto de ADNc. Para monitorizar la complementación del defecto metabólico con la actividad de ASA, se midieron la esteroide sulfatasa (STS) y la N-acetilgalactosamina 6-sulfatasa (GalNAc6S) en los fibroblastos transducidos antes o después de la selección. La transducción de la FGE no mutante restauró parcialmente la actividad catalítica de las tres sulfatasas en dos líneas de células con MSD (Tabla 2) y para la STS en una tercera línea de células con MSD. Debe observarse que para ASA y GalNAc6S la restauración fue solo parcial después de la selección de los fibroblastos que alcanzaban del 20 al 50 % de la actividad normal. Para la STS, se encontró que la actividad se restauraba a la de los fibroblastos de

control después de la selección. La selección aumentó la actividad de ASA y STS del 50 al 80 %, que es compatible con la observación anterior de que llegan a translucirse del 15 al 50 % de los fibroblastos (Lübke et al., Nat. Gen., 2001, 28:73-76). Las actividades de sulfatasa en los fibroblastos con MSD transducidos con el vector retroviral solo (Tabla 2) fueron comparables a las de los fibroblastos con MSD no transducidos (no mostradas). La transducción de ADNc de FGE que porta la mutación IVS3+5-8del no restauró las actividades de sulfatasa (Tabla 2).

**Tabla 2: Complementación de fibroblastos con MSD por transducción de ADNc de FGE no mutante o mutante**

Fibroblastos	Inserción FGE	ASA <sup>1</sup>	Sulfatasa STS <sup>1</sup>	Ga1NAc6S <sup>1</sup>
MSD 3 <sup>o</sup>	-	1.9 ± 0.2	<3	56.7 ± 32
	FGE <sup>+</sup>	7.9	13.5	n. d.
	FGE <sup>++</sup>	12.2 ± 0.2	75.2	283 ± 42
	FGE-IVS3+5-8del <sup>+</sup>	1.8	< 3	n. d.
	FGE-IVS3+5-8del <sup>++</sup>	2.1	< 3	98.5
MSD 4 <sup>o</sup>	-	1.1 ± 0.3	< 3	n. d.
	FGE <sup>+</sup>	4.7	17.0	n. d.
Control Fibroblastos		58 ± 11	66 ± 31	828 ± 426

1 Los valores dan la razón entre ASA (proteína celular mU / mg), STS (/ mg de proteína celular mU), Ga1NAc6S (mU / mg celular proteína) y la de β-hexosaminidasa (proteína de la célula U / mg). Para el control de fibroblastos la media y la variación de 6 a 11 líneas celulares se da. Donde se indica el rango de dos culturas transducidas en paralelo se da para fibroblastos MSD.  
<sup>o</sup> El número de fibroblastos MSD se refiere a la del paciente en la Tabla 1.  
<sup>+</sup> Determinación de la actividad antes de la selección.  
<sup>++</sup> Determinación de la actividad después de la selección.  
 N.D. ∴ no determinado

## Discusión

### FGE es una glicoproteína altamente conservada del retículo endoplásmico.

La purificación de FGE de testículos bovinos dio dos polipéptidos de 39,5 y 41,5 kDa que se originan a partir del mismo gen. La expresión de tres versiones de FGE marcadas de manera diferente en tres líneas de células eucariotas diferentes como una única forma sugiere que una de las dos formas observadas en la preparación de FGE purificada de testículos bovinos puede haberse generado por proteólisis limitada durante la purificación. La sustitución de Cys69 en el péptido ASA65-80 por serina fue crítica para la purificación de FGE mediante cromatografía de afinidad. La FGE tiene una secuencia señal escindible que media en la translocación a través de la membrana del retículo endoplásmico. La mayor parte de la proteína madura (275 residuos de 340) define un dominio único, que es probable que esté compuesto por tres subdominios (véase el Ejemplo 2), para ninguno de los tres subdominios existen homólogos en proteínas con función conocida. El reconocimiento del motivo de modificación *FGly* lineal en polipéptidos de sulfatasa recién sintetizados (Dierks et al., EMBO J., 1999, 18:2084-2091) podría ser la función de un subdominio de FGE. El dominio catalítico podría catalizar la formación de *FGly* de varias formas. Se ha propuesto que FGE extrae electrones del grupo tiol de la cisteína y los transfiere a un aceptor. El tialdehído resultante se hidrolizaría espontáneamente a *FGly* y H<sub>2</sub>S (Schmidt, B., et al., Cell, 1995, 82:271-278). Alternativamente, FGE podría actuar como una oxigenasa de función mixta (monooxigenasa) que introduce un átomo de O<sub>2</sub> en la cisteína y el otro en H<sub>2</sub>O con la ayuda de un donante de electrones tal como FADH<sub>2</sub>. El derivado de hidrato de tialdehído resultante de cisteína reaccionaría espontáneamente con *FGly* y H<sub>2</sub>S. Experimentos preliminares con una preparación de FGE parcialmente purificada mostraron una dependencia crítica de la formación de *FGly* del oxígeno molecular. Esto sugeriría que FGE actúa como una oxigenasa de función mixta. La alta conservación particular del subdominio 3 y la presencia de tres residuos de cisteína completamente conservados en el mismo hacen que este subdominio sea un candidato probable para el sitio catalítico. Será interesante ver si los elementos estructurales que median en el reconocimiento del motivo *FGly* y la unión de un aceptor de electrones o un donante de electrones se correlacionan con la estructura de dominio de FGE.

FGE recombinante se localiza en el retículo endoplásmico, que es compatible con el sitio propuesto de su acción. Los residuos de *FGly* se generan en sulfatasas recién sintetizadas durante o poco después de su translocación al retículo endoplásmico (Dierks et al., Proc. Natl. Acad. Sci. U.S.A., 1997, 94:11963-11968; Dierks et al., FEBS Lett.,



1998, 423:61-65). La propia FGE no contiene una señal de retención en el RE del tipo KDEL. Su retención en el retículo endoplásmico puede estar mediada, por tanto, por la interacción con otras proteínas del RE. Los componentes de la maquinaria de translocación/N-glicosilación son candidatos atractivos para tales parejas de interacción.

5

### Las mutaciones en FGE producen MSD

Los presentes inventores han mostrado que las mutaciones en el gen que codifica FGE producen MSD. La FGE también puede interactuar con otros componentes, y los defectos en genes que codifican estos últimos podrían producir MSD de igual forma. En siete pacientes con MSD los presentes inventores encontraron de hecho diez mutaciones diferentes en el gen de FGE. Todas las mutaciones tienen efectos graves en la proteína FGE sustituyendo residuos altamente conservados en el subdominio 3 (tres mutaciones) o en el subdominio 2 (una mutación) o truncamientos en el extremo C de diversas longitudes (cuatro mutaciones) o grandes deleciones en marco (dos mutaciones). Para dos líneas celulares con MSD y una de las mutaciones en MSD se demostró que la transducción del ADNc de FEG no mutante, pero no del mutante, restaura parcialmente las actividades de sulfatasas. Esto identifica claramente el gen de FGE como el sitio de mutación y la naturaleza de la mutación que produce la enfermedad. MSD es tanto clínica como bioquímicamente heterogénea. Se ha diferenciado una forma neonatal poco frecuente que se presenta al nacer y que desarrolla un hidrocéfalo, una forma común que se asemeja inicialmente a una leucodistrofia metacromática infantil y que posteriormente desarrolla características similares a las de la ictiosis y la mucopolisacaridosis, y una forma leve menos frecuente en la que prevalecen las características clínicas de una mucopolisacaridosis. Bioquímicamente, es característico que pueda detectarse una actividad de sulfatasas residual, que para la mayoría de los casos en fibroblastos de piel en cultivo es inferior al 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 con MSD, la actividad de sulfatasas seleccionadas puede alcanzar el intervalo normal (Yutaka et al., Clin. Genet., 1981, 20:296-303). Además, se ha informado que la actividad residual se somete a variaciones dependiendo de las condiciones del cultivo celular y de factores desconocidos. Bioquímicamente, la MSD se ha clasificado en dos grupos. En el grupo I, la actividad residual de sulfatasas es inferior al 15 % incluyendo la de ASB. En el grupo II, la actividad residual de sulfatasas es superior y particularmente la de ASB puede alcanzar valores de hasta el 50-100 % del control. Todos los pacientes informados en el presente documento se clasifican en el grupo I, excepto el paciente 5, que se clasifica en el grupo II (actividad de ASB en el intervalo control) del fenotipo bioquímico. Basándose en criterios clínicos, los pacientes 1 y 6 son casos de neonatos, mientras que los pacientes 2-4 y 7 tienen la forma de MSD común y el paciente 5 la forma similar a mucopolisacaridosis.

La heterogeneidad fenotípica sugiere que las diferentes mutaciones en pacientes con MSD están asociadas con diferentes actividades residuales de FGE. Los datos preliminares en células PT67 que expresan de manera estable FGE IVS3+5-8del indican que la delección en marco del exón 3 anula la actividad FEG completamente. La caracterización de las mutaciones en MSD, de las propiedades bioquímicas de FGE mutante y del contenido residual de *FGly* en sulfatasas usando un método de espectrometría de masas altamente sensible desarrollado recientemente (Peng et al., J. Mass Spec., 2003, 38:80-86) proporcionará un mejor entendimiento de la correlación genotipo-fenotipo en MSD.

### Ejemplo 2:

*El gen de FGE humana define una nueva familia de genes que modifican sulfatasas que está conservada de procariotas a eucariotas*

### Bioinformática

Se describieron los péptidos señal y los sitios de escisión con el método de von Heijne (Nucleic Acids Res., 1986, 14:4683) implementado en EMBOSS (Rice et al., Trends in Genetics, 2000, 16:276-277), y el método de Nielsen et al. (Protein Engineering, 1997, 10:1-6). Se predijeron los sitios de N-glicosilación usando el algoritmo de Brunak (Gupta y Brunak, Pac. Symp. Biocomput., 2002, 310-22).

Se detectaron los dominios funcionales buscando modelos ocultos de Markov en PFAM (versión 7.8) (Sonnhammer et al., Nucleic Acids Res., 1998, 26:320-322). Se obtuvieron secuencias de la semilla DUF323 en PFAM de TrEMBL (Bairoch, A. y Apweiler, R., Nucleic Acids Res., 2000, 28:45-48). Se realizaron alineamientos múltiples y construcciones de árbol filogenético con Clustal W (Thompson, J., et al., Nucleic Acids Res., 1994, 22:4673-4680). Para la computación de árbol filogenético, se excluyeron las posiciones de hueco y se corrigieron para sustituciones múltiples. Se realizó remuestreo en árbol para obtener resultados significativos. Se visualizaron los árboles usando Njplot (Perriere, G. y Gouy, M., Biochimie, 1996, 78:364-369). Los alineamientos se representaron gráficamente usando el comando prettyplot de EMBOSS.

Para buscar homólogos de FGE, se consultaron las bases de datos NR, NT y EST del National Center for Biotechnology Information (NCBI) (Wheeler et al., Nucleic Acids Res., 2002, 20:13-16) con BLAST (Altschul et al., Nucleic Acids Res., 1997, 25:3389-3402). Para las secuencias proteicas, se realizó la búsqueda usando Psi-Blast de convergencia iterativo frente a la versión actual de la base de datos NR que usa un punto de corte de valor probable

65

de  $10^{-40}$ , y parámetros por defecto. Se alcanzó la convergencia después de 5 iteraciones. Para las secuencias de nucleótidos, se realizó la búsqueda con Psi-TBlastn: usando NR y la secuencia de proteínas de FGE humana como entrada, se construyó una matriz de puntuación para hFGE con Psi-Blast de convergencia iterativo. Se usó esta matriz como entrada para Blastall para consultar las bases de datos de nucleótidos NT y EST. Para ambas etapas, se usó un punto de corte de valor probable de  $10^{-20}$ .

Se realizó la predicción de la estructura secundaria de las proteínas usando Psipred (Jones, D., J Mol Biol., 1999, 292:1950-202; McGuffin, L., et al., Bioinformatics, 2000, 16:404-405).

Se calcularon puntuaciones de similitud de los subdominios a partir de los alineamientos usando el algoritmo cons. de EMBOSS con parámetros por defecto. Se generaron los meta-alineamientos alineando secuencias consenso de subgrupos de la familia de FGE. Se determinaron la sintenia y la organización de loci genómicos usando recursos de genoma de ser humano y ratón de NCBI en NCBI (Bethesda, MD) y sintenia de ser humano-ratón-rata de Softberry (Mount Kisco, NY). Se descargaron secuencias genómicas bacterianas del servidor FTP de NCBI. Se usó la anotación genómica microbiana de NCBI para obtener una visión general de los loci genómicos de los genes bacterianos de FGE.

## **Resultados y discusión**

### **Características y motivos básicos de la FGE humana y proteínas relacionadas**

El gen de FGE humana (SEC ID N°: 1, 3) codifica la proteína FGE (SEC ID N°: 2) que se predice que tiene 374 residuos. Una señal de escisión entre los residuos 22-33 (Heijne-Score de 15,29) y un índice hidropático (Kyte, J. y Doolittle, R., J Mol Biol., 1982, 157:105-132) de los residuos 17-29 de entre 1,7 y 3,3 indican que los 33 residuos del extremo N se escinden después de la translocación al RE. Sin embargo con el algoritmo de Nielsen et al. (Protein Engineering, 1997, 10:1-6), se predice la escisión de la secuencia señal después del residuo 34. La proteína tiene un único sitio posible de N-glicosilación en Asn 141.

Una búsqueda con la secuencia de proteína FGE en la base de datos de motivos de proteínas PFAM (Sonnhammer et al., Nucleic Acids Res., 1998, 26:320-322) reveló que los residuos 87-367 de la FGE humana pueden clasificarse como el dominio de proteína DUF323 ("dominio de función desconocida", PF03781) con un valor probable altamente significativo de  $7:9 \times 10^{-114}$ . La semilla en PFAM que define DUF323 consiste en 25 secuencias de proteínas, de las que la mayoría son proteínas hipotéticas derivadas de datos de secuenciación. Para analizar la relación entre la FGE humana y DUF323, se realizó un alineamiento múltiple de FGE con las secuencias de la semilla DUF323. Basándose en esto, se construyó y remuestreo un árbol filogenético. Cuatro de las secuencias hipotéticas (TrEMBL-ID Q9CK12, Q9I761, Q94632 y Q9Y405) tuvieron una divergencia tan fuerte de los otros miembros de la semilla que evitaron el remuestreo satisfactorio y tuvieron que eliminarse del conjunto. La Figura 2 muestra el árbol de remuestreo que presenta la relación entre la FGE humana y las 21 proteínas de semilla DUF323 restantes. El árbol puede usarse para subdividir los miembros de semilla en dos categorías: homólogos estrechamente relacionados con FGE humana y los genes restantes menos relacionados.

Las 7 proteínas superiores tienen una distancia filogenético entre 0,41 y 0,73 con respecto a la FGE humana. Solo contienen un único dominio, DUF323. La homología dentro de este grupo se extiende por toda la secuencia de aminoácidos, consistiendo la mayor parte en el dominio DUF323. El dominio DUF323 está fuertemente conservado dentro de este grupo de homólogos, mientras que las otras 15 proteínas de la semilla están menos relacionadas con la FGE humana (distancia filogenética entre 1,14 y 1,93). Su dominio DUF323 diverge considerablemente del dominio DUF323 altamente conservado del imprimación grupo (véase la sección "Subdominios de FGE y mutaciones en el gen de FGE"). La mayoría de estas 15 proteínas son hipotéticas, habiéndose investigado adicionalmente seis de ellas. Una de ellas, una serina/treonina cinasa (TrEMBL:084147) de *C. trachomatis* contiene otros dominios, además de DUF323: un dominio de unión a ATP y un dominio cinasa. Las secuencias de *R. sphaeroides* (TrEMBL: Q9ALV8) y *Pseudomonas sp.* (TrEMBL: 052577) codifican la proteína NirV, un gen cotranscrito con la nitrito reductasa que contiene cobre nirK (Jain, R. y Shapleigh, J., Microbiology, 2001, 147:2505-2515). CarC (TrEMBL: Q9XB56) es una oxigenasa implicada en la síntesis de un antibiótico  $\beta$ -lactámico de *E. carotovora* (McGowan, S., et al., Mol Microbiol., 1996, 22:415-426; Khaleeli N, T. C. y Busby RW, Biochemistry, 2000, 39:8666-8673). XylR (TrEMBL: 031397) y BH0900 (TrEMBL: Q9KEF2) son proteínas potenciadoras de unión implicadas en la regulación de la utilización de pentosas (Rodionov, D., et al., FEMS Microbiol Lett., 2001, 205:305-314) en *Bacillaceae* y *Clostridiaceae*. La comparación de FGE y DUF323 condujo al establecimiento de un umbral de homología que diferencia la familia de FGE de homólogos que contienen DUF323 distantes con diferentes funciones. Estos últimos incluyen una serina/treonina cinasa y XylR, un potenciador de la transcripción, así como FGE, una enzima generadora de FGly y CarC, una oxigenasa. Como se trata en otra parte en el presente documento, la FGE también podría ejercer su función de modificación de cisteína como una oxigenasa, sugiriendo que la FGE y miembros distintos de FGE de la semilla DUF323 pueden compartir una función oxigenasa.

### **Homólogos de FGE**

La presencia de homólogos estrechamente relacionados de la FGE humana en la semilla DUF323 dirigió a los

presentes inventores a la búsqueda de homólogos de FGE humana en la base de datos NR de NCBI (Wheeler et al., Nucleic Acids Res., 2002, 20:13-16). El umbral de la búsqueda se eligió de tal forma que se obtuvieron los 6 homólogos presentes en la semilla DUF323 y otros homólogos estrechamente relacionados sin encontrar los otros miembros de semilla. Esta búsqueda condujo a la identificación de tres ortólogos de FGE en eucariotas, 12 ortólogos en procariontes y dos parálogos en el hombre y el ratón (Tabla 3).

**Tabla 3: La familia de genes de FGE en eucariotas y procariontes**

SEQ ID NOs: NA, AA [GI]	Especies	Longitud [AA]	Subgrupo
1/3,2	Homo sapiens	374	E1
49,50 [22122361]	Mus musculus	372 <sup>f</sup>	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
57,58 [25028125]	Corynebacterium efficiens YS-314	334	P1
59,60 [23108562]	Novosphingobium aromaticivorans	338	P2
61,62 [13474559]	Mesorhizobium loti	372	P2
63,64 [2298809]	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
Identificador proteína GI GenBank NA- AA ácido nucleico - aminoácidos, E1 - E2 - ortólogos eucariotas parálogos eucariotas P1 - estrechamente relacionado procarionte P2 ortólogos - otros ortólogos procariontes secuencia de la proteína f- mispredicted en GenBank			

Obsérvese que en GenBank se predice que la secuencia GI 22122361 de ratón codifica una proteína de 284 aa, aunque la secuencia NM 145937 de ADNc codifica una proteína de 372 residuos. Esta mala predicción se basa en la omisión del imprimación exón del gen murino de FGE. Todas las secuencias encontradas en la base de datos NR son de procariontes o eucariotas superiores. No se detectaron homólogos de FGE en *Archaeobacteriae* o plantas. Las búsquedas con umbrales incluso inferiores en los genomas completamente secuenciados de *C. elegans* y *S. cerevisiae* y las bases de datos relacionadas ORF no revelaron ningún homólogo. Una búsqueda en las secuencias eucariotas de las bases de datos de nucleótidos NT y EST condujo a la identificación de 8 EST de ortólogo de FGE adicionales, mostrando los fragmentos de secuencia de ADNc en el extremo 3' un alto grado de conservación al nivel de proteína que no se enumeran en la base de datos NR. Estas secuencias no engloban la parte codificante completa de los ARNm y todos son de eucariotas superiores (Tabla 4).

Tabla 4: Fragmentos EST ortólogos de FGE en eucariotas

SEQ ID NOs: 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- GenBank adhesión n; Ácido nucleico NA-

El alineamiento múltiple y la construcción de un árbol filogenético (usando ClustalW) de las secuencias codificantes a partir de la base de datos NR permitió la definición de cuatro subgrupos de homólogos: ortólogos eucariotas (FGE de ser humano, ratón, mosquito y mosca de la fruta, parálogos eucariotas (parálogo de FGE de ser humano y ratón), ortólogos procariotas estrechamente relacionados con 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 global con la FGE humana del 87 % (ratón), 48 % (mosca de la fruta) y 47 % (*Anopheles*). Aunque los ortólogos de FGE se encuentran en procariotas y eucariotas superiores, no aparecen en los genomas completamente secuenciados de eucariotas inferiores situados filogenéticamente entre *S. cerevisiae* y *D. melanogaster*. Además, los homólogos de FGE están ausentes en los genomas completamente secuenciados de *E. coli* y el pez globo.

Como se trata en otra parte en el presente documento, los parálogos de FGE encontrados en ser humano y ratón pueden tener una actividad de generación de *FGly* menor y contribuir a las actividades residuales de las sulfatasas encontradas en pacientes con MSD.

#### Subdominios de FGE

Los miembros de la familia de genes de FGE tienen tres partes/dominios altamente conservados (como se describe en otra parte en el presente documento). Además de las dos secuencias no conservadas que separan a la imprimación, tienen extensiones no conservadas en los extremos N y C. Se considera que las tres partes representan subdominios del dominio DUF323 debido a que están separadas por partes no conservadas de longitud variable. La longitud de la parte que separa los subdominios 1 y 2 varía entre 22 y 29 residuos y la que separa los subdominios 2 y 3 entre 7 y 38 aminoácidos. Las partes no conservadas de los extremos N y C muestran una variación incluso más fuerte en la longitud (extremo N: 0-90 AA, extremo C: 0-28 AA). La secuencia para el gen de FGE de *Ralstonia metallidurans* está probablemente incompleta, ya que carece del imprimación subdominio.

Para verificar la plausibilidad de definir subdominios de DUF323, presentes inventores realizaron la predicción de una estructura secundaria de la proteína FGE humana usando Psipred. Se predice que la señal de RE hidrófoba (residuos 1-33) contiene estructuras en hélice, que confirma la predicción de señal del algoritmo de von-Heijne. La región no conservada del extremo N (aa 34-89) y la región de separación entre los subdominios 2 y 3 (aa 308-327) contienen secciones helicoidales. La región que separa los subdominios 1 y 2 contiene una hélice. La hélice  $\alpha$  en los aa 65/66 tiene una baja confianza de predicción y es probablemente un artefacto de predicción. Los límites de subdominio están situados dentro de las hélices y no interrumpen las hélices  $\alpha$  o las hebras  $\beta$ . El imprimación subdominio está constituido por varias hebras  $\beta$  y una hélice  $\alpha$ , el segundo subdominio contiene dos hebras  $\beta$  y cuatro hélices  $\alpha$ . El tercer subdominio tiene una región de hélice  $\alpha$  flanqueada por una hoja al principio y al final del subdominio. En resumen, la estructura secundaria está de acuerdo con la estructura de subdominios propuesta ya que los límites de subdominio están situados dentro de hélices y los subdominios contienen los elementos estructurales hélices  $\alpha$  y hebras  $\beta$ .

Debe observarse que ninguno de los subdominios existe como un módulo aislado en las secuencias enumeradas en las bases de datos. Dentro de cada uno de los cuatro subgrupos de la familia de FGE, los subdominios están

altamente conservados, mostrando el tercer subdominio la mayor homología (Tabla 5). Este subdominio también muestra la homología más fuerte en todos los subgrupos.

5

10

15

Subfamilia	Miembros	Subdominio		
		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 estrechamente relacionados; P2 - otros ortólogos procariotas

20 El imprimación subdominio de la familia de FGE muestra la homología más débil en todos los subgrupos. En los ortólogos eucariotas lleva el sitio de N-glicosilación: en el residuo Asn 141 en ser humano, en Asn 139 en el ratón y Asn 120 en la mosca de la fruta. En *Anopheles*, no se encontró asparagina en el residuo 130 homólogo a Asn 120 de *D. melanogaster*. Sin embargo, un cambio de dos nucleótidos crearía un sitio de N-glicosilación Asn 130 en *Anopheles*. Por tanto, es necesario volver a secuenciar la secuencia que engloba el residuo 130. El segundo subdominio es rico en triptófanos con 12 Trp en 129 residuos de la FGE humana. Diez de estos triptófanos están conservados en la familia de FGE.

25

Alta conservación del subdominio 3: el subdominio 3 entre ortólogos eucariotas es similar en un 100 % e idéntico en un 90 %. La importancia del tercer subdominio para la función de la proteína está subrayada por la observación de que este subdominio no es un punto clave para las mutaciones causantes de la enfermedad en pacientes con MSD. Siete de las nueve mutaciones identificadas en seis pacientes con MSD descritos en el Ejemplo 1 están localizadas en secuencias que codifican los 40 residuos del subdominio 3. Los residuos contienen cuatro cisteínas, tres de las cuales están conservadas entre los ortólogos procariotas y eucariotas. Los dos parálogos eucariotas muestran la menor homología con los otros miembros de la familia de FGE, por ejemplo, carecen de dos de las tres cisteínas conservadas del subdominio 3. Las características conservadas entre las secuencias del subdominio 3 de ortólogos y parálogos son el motivo RVXXGG(A)S inicial (SEC ID N°: 79), un heptámero que contiene tres argininas (residuos 19-25 de la secuencia consenso del subdominio) y el motivo GFR terminal. Una comparación con el dominio DUF323 de las 15 secuencias semillas que no son homólogos estrechos de FGE muestra marcadas diferencias de secuencia: las 15 secuencias de semillas tienen un subdominio imprimación y segundo menos conservado, aunque la estructura de subdominios global también es visible. El subdominio 3, que está fuertemente conservado en la familia de FGE, es más corto y tiene una homología significativamente más débil en el subdominio 3 de eucariotas (similitud de aproximadamente el 20 %) en comparación con los miembros de la familia de FGE de procariotas (similitud de aproximadamente el 60 %). Por tanto, todos carecen de los residuos conservados de cisteína del subdominio 3. Las únicas características conservadas son el motivo RVXXGG(A)S inicial (SEC ID N°: 79) y el motivo GFR terminal.

30

35

40

45

#### Organización genómica del gen de FGE humana y murina

El gen de FGE humana está localizado en el cromosoma 3p26. Abarca 105 kb y 9 exones para la secuencia traducida. El gen de FGE murina tiene una longitud de 80 Kb y está localizado en el cromosoma 6E2. Los 9 exones del gen de FGE murina tienen casi el mismo tamaño que los exones humanos (Figura 3). Las principales diferencias entre el gen humano y de ratón son la menor conservación de UTR en 3' en el exón 9 y la longitud de exón 9, que es 461 pb más largo que en el gen murino. El segmento 6E2 del cromosoma 6 de ratón es altamente sinténico al segmento 3p26 del cromosoma humano. Hacia el telómero, tanto los loci de FGE humana como murina están flanqueados por los genes que codifican LMCD1, KIAA0212, ITPR1, AXCAM y IL5RA. En la dirección centromérica, ambos loci de FGE están flanqueados por los loci de CAV3 y OXTR.

50

55

#### Organización genómica de los genes de FGE de procariotas

En procariotas, las sulfatasas se clasifican o bien como sulfatasas de tipo cisteína o bien de tipo serina dependiendo del residuo que se convierte en *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 *Klebsiella pneumoniae*, *E. coli* y *Yersinia pestis*, las sulfatasas de tipo serina forman parte de un operón con *AtsB*, que codifica una proteína citosólica que contiene motivos de agrupamiento de hierro-azufre y son críticos 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).

60

65

Por tanto, fue de interés examinar si los genes de FGE procariotas estaban localizados en la proximidad de sulfatasas de tipo cisteína que son los sustratos de FGE. Entre los genes de FGE procariotas mostrados en la Tabla 3, siete tienen genomas completamente secuenciados, que permite un análisis de las proximidades de los loci de FGE. De hecho, en cuatro de los 7 genomas (*C. efficiens*: PID 25028125, *P. putida*: PID 26990068, *C. crescentus*: PID 16125425 y *M. tuberculosis*: PID 15607852) se encontró una sulfatasa de tipo cisteína en proximidad directa con FGE compatible con una cotranscripción de FGE y la sulfatasa. En dos de ellos (*C. efficiens* y *P. putida*), FGE y la sulfatasa tienen incluso ORF solapantes, que indica fuertemente su coexpresión. Además, la proximidad genómica de los genes de FGE y sulfatasa en cuatro procariotas proporciona una evidencia adicional para suponer que las FEG bacterianas son ortólogos funcionales.

Los tres microorganismos restantes sí contienen sulfatasas de tipo cisteína (*S. coelicolor*: PID 24413927, *M. loti*: PID 13476324, *S. meliloti*: PID 16262963, 16263377, 15964702), sin embargo, los genes próximos a FGE en estos microorganismos ni contienen una secuencia distintiva de sulfatasa canónica (Dierks, T., et al., J Biol Chem., 1998, 273:25560-25564) ni un dominio que indique su función. En estos organismos, la expresión de FGE y sulfatasas de tipo cisteína, por tanto, probablemente va a regularse en *trans*.

### Conclusiones

La identificación de FGE humana cuya deficiencia produce la enfermedad de almacenamiento lisosómico transmitida de manera autonómica-recesiva, deficiencia múltiple de sulfatasas, permite la definición de una nueva familia de genes que comprende ortólogos de FGE de procariotas y eucariotas, así como un parálogo de FGE en ratón y ser humano. La FGE no se encuentra en genomas completamente secuenciados de *E. coli*, *S. cerevisiae*, *C. elegans* y *Fugu rubripes*. Además, hay un vacío filogenético entre procariotas y eucariotas superiores, careciéndose de FGE en algunas especies situadas filogenéticamente entre procariotas y *D. melanogaster*. Sin embargo, algunos de estos eucariotas inferiores, por ejemplo *C. elegans*, tienen genes de sulfatasa de tipo cisteína. Esto indica la existencia de un segundo sistema de generación de *FGly* que actúa sobre las sulfatasas de tipo cisteína. Esta suposición está apoyada por la observación de que *E. coli*, que carece de FGE, puede generar *FGly* en sulfatasas de tipo cisteína (Dierks, T., et al., J Biol Chem., 1998, 273:25560-25564).

### Ejemplo 3:

*La expresión de FGE produce aumentos significativos en la actividad de sulfatasa en líneas celulares que expresan en exceso una sulfatasa*

Los presentes inventores quisieron examinar los efectos de FGE sobre células que expresan/expresan en exceso una sulfatasa. Para este fin, se transfectaron células HT-1080 que expresan las sulfatasas humanas iduronato 2-sulfatasa (I2S) o N-acetilgalactosamina 6-sulfatasa (GALNS) por duplicado con o bien una construcción de expresión de FGE, pXMG.1.3 (Tabla 7 y Figura 4) o bien un plásmido de control, pXMG.1.2 (FGE en orientación antisentido no puede producir FGE funcional, Tabla 7). Se recogieron muestras de los medios 24, 48 y 72 horas tras un cambio de medio tras electroporación de 24 horas. Se probaron las muestras de medio para actividad de sulfatasa respectiva mediante ensayo de actividad y el nivel de proteína sulfatasa total estimado mediante ELISA específico para o bien iduronato 2-sulfatasa o bien N-acetilgalactosamina 6-sulfatasa.

**Tabla 6. Líneas celulares transfectadas que expresan sulfatasas usadas como sustratos para transfección**

Cepa celular	Plásmido	Sulfatasa Expresada
36F	pXFM4A.1	N-acetilgalactosamina 6-sulfatasa
30C6	pXI2S6	Iduronato 2-sulfatasa

**Tabla 7. Plásmidos de FGE y de control usados para transfectar células HT-1080 que expresan iduronato 2-sulfatasa y N-acetilgalactosamina 6-sulfatasa**

Plásmido	Configuración de los elementos de la secuencia de ADN Mayor *
pXMG.1.3 (expresión FGE)	> 1,6 kb CMV potenciador / promotor > 1,1 kb cADN FGE > secuencia no traducida hGH3 ' < amp < DHFR casete < casete Cdneo (neomicina fosfotransferasa)
pXMG.1.2 (control, FGE orientación inversa)	kb CMV potenciador / promotor de < 1,1 kb cADN FGE < hGH3 'no traducida secuencia < amp < DHFR casete < casete Cdneo (neomicina fosfotransferasa)
* > denotes orientation 5' to 3'	

Denota orientación 5' a 3'

## **Procedimientos experimentales**

### 5 **Materiales v métodos**

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

10 Se recogieron células HT-1080 para obtener  $9-12 \times 10^6$  células para cada electroporación. Se transfectaron dos plásmidos por duplicado: uno para probarse (FGE) y un control; en este caso, el plásmido de control contenía el ADNc de FGE clonado en la orientación inversa con respecto al promotor de CMV. Se centrifugaron las células a aproximadamente 1000 RPM durante 5 minutos. Se suspendieron las células en 1X PBS a  $16 \times 10^6$  células/ml. Se añadieron al fondo de la cubeta de electroporación 100 µg de ADN de plásmido, se añadieron 750 µl de suspensión celular ( $12 \times 10^6$  células) a la disolución de ADN en la cubeta. Se mezclaron suavemente las células y el ADN con una pipeta de transferencia de plástico, teniendo cuidado de no crear burbujas. Se sometieron las células a electroporación a 450 V, 250 µF (BioRad Gene Pulser). Se registró la constante de tiempo.

20 Se permitió que las células sometidas a electroporación se asentaran sin alteraciones durante 10-30 minutos. Entonces se añadieron 1,25 ml de DMEM/suero bovino al 10 % a cada cubeta, se mezclaron y se transfirieron todas las células a un matraz T75 nuevo que contenía 20 ml de DMEM/10. Después de 24 horas, se volvió a alimentar el matraz con 20 ml de DMEM/10 para eliminar las células muertas. 48-72 horas después de la transfección, se recogieron muestras de los medios y se recogieron las células de los matraces T75 por duplicado.

#### Preparación del medio

25 1 l de DMEM/10 (contiene: 23 ml de L-glutamina 2 mM, 115 ml de suero bovino).

30 Se transfectaron las células en medios sin metotrexato (MTX). 24 horas después, las células volvieron a alimentarse con medios que contenían las cantidades apropiadas de MTX (36F = MTX 1,0 µM, 30C6 = MTX 0,1 M). Se recogió el medio y se recogieron las células 24, 48 y 72 horas después de volver a alimentar. Ensayos de actividad

35 **Iduronato 2-sulfatasa (I2S)**. Se equilibraron columnas de desalación NAP5 (Amersham Pharmacia Biotech AB, Uppsala, Suecia) con tampón de diálisis (acetato sódico 5 mM, Tris 5 mM, pH 7,0). Se aplicó una muestra que contenía I2S a la columna y se permitió que entrara en el lecho. Se eluyó la muestra en 1 ml en tampón de diálisis. Se diluyeron adicionalmente las muestras desaladas hasta aproximadamente I2S 100 ng/ml tampón de reacción (acetato sódico 5 mM, BSA 0,5 mg/l, Triton X-100 al 0,1 %, pH 4,5). Se añadieron 10 µl de cada muestra de I2S a la fila superior de una placa fluorimétrica de 96 pocillos (Perkin Elmer, Norwalk, CT) y se preincubó durante 15 minutos a 37 °C. Se preparó el sustrato disolviendo sulfato de 4-metil-umbelíferilo (Fluka, Buchs, Suiza) en tampón de sustrato (acetato sódico 5 mM, BSA 0,5 mg/l, pH 4,5) a una concentración final de 1,5 mg/ml. Se añadieron 100 µl de sustrato a cada pocillo que contenía muestra de I2S y se incubó la placa durante 1 hora a 37 °C en la oscuridad. Después de la incubación, se añadieron 190 µl de tampón de parada (glicina 332,5 mM, carbonato sódico 207,5 mM, pH 10,7) a cada pocillo que contenía la muestra. Se preparó 4-metilumbeliferona como disolución madre (4-MUF, Sigma, St. Louis, MO) como patrón de producto en agua de calidad como reactivo hasta una concentración final de 1 µM. Se añadieron 150 de la disolución madre de 4-MUF 1 µM y 150 µl de tampón de parada a un pocillo de fila superior en la placa. Se añadieron 150 de tampón de parada a cada pocillo restante en la placa de 96 pocillos. Se realizaron diluciones sucesivas dobles desde la fila superior de cada columna hasta la última fila de la placa. Se leyó la placa en un analizador de microplacas universal Fusion (Packard, Meriden, CT) con una longitud de onda de filtro de excitación de 330 nm y una longitud de onda de filtro de emisión de 440 nm. Se generó una curva patrón de µmoles de disolución madre de 4-MUF frente a fluorescencia, y se extrapóla la fluorescencia de las muestras desconocidas a partir de esta curva. Los resultados se informan como unidades/ml en las que una unidad de actividad era igual a 1 µmol de 4-MUF producido por minuto a 37 °C.

55 **N-acetilgalactosamina 6-sulfatasa (GALNS)**. El ensayo de actividad de GALNS hace uso del sustrato fluorescente 4- metilumbeliferil-β-D-galactopiranosido-6-sulfato (Toronto Research Chemicals Inc., nº de catálogo M33448). El ensayo comprendió dos etapas. En la imprimación etapa, se incubaron 75 µL del sustrato 1,3 mM preparado en tampón de reacción (acetato sódico 0,1 M, cloruro sódico 0,1 M, pH 4,3) durante 4 horas a 37 °C con 10 µl de medio/muestra de proteína o sus diluciones correspondientes. Se detuvo la reacción mediante la adición de 5 µl de fosfato de sodio monobásico 2 M para inhibir la actividad de GALNS. Tras la adición de aproximadamente 500 U de β-galactosidasa de *Aspergillus oryzae* (Sigma, nº de catálogo G5160), se incubó la mezcla de reacción a 37 °C durante una hora adicional para liberar el resto fluorescente del sustrato. Se detuvo la segunda reacción mediante la adición de 910 µl de disolución de parada (glicina al 1 %, carbonato de sodio al 1 %, pH 10,7). Se midió la fluorescencia de la mezcla resultante usando una longitud de onda de medición de 359 nm y una longitud de onda de referencia de 445 nm con 4-metilumbeliferona (sal sódica de Sigma, nº de catálogo M1508) que sirve como patrón de referencia. Una unidad de la actividad se corresponde con nmoles de 4-metilumbeliferona liberados por hora.

65

**Inmunoensayos (ELISA)**

**Iduronato 2-sulfatasa (I2S).** Se recubrió una placa de fondo plano de 96 pocillos con un anticuerpo anti-I2S monoclonal de ratón diluido a 10 ng/ml en bicarbonato sódico 50 nM a pH 9,6 durante 1 hora a 37 °C. Se desarrolló el anticuerpo anti-I2S monoclonal de ratón por contrato por Maine Biotechnology Services, Inc. (Portland, ME) hasta dar lugar a un polipéptido de I2S humano, de longitud completa, producido de manera recombinante, purificado usando tecnología de producción de hibridomas estándar. Se lavó la placa 3 veces con 1X PBS que contenía Tween-20 al 0,1 % y se bloqueó durante 1 hora con BSA al 2 % en tampón de lavado a 37 °C. Se usó tampón de lavado con BSA al 2 % para diluir las muestras y los patrones. Se diluyó el patrón de I2S y se usó desde 100 ng/ml hasta 1,56 ng/ml. Después de la eliminación del tampón de bloqueo, se aplicaron las muestras y los patrones a la placa y se incubaron durante 1 hora a 37 °C. Se diluyó el anticuerpo de detección, anticuerpo anti-I2S de ratón conjugado con peroxidasa de rábano, a 0,15 µg/ml en tampón de lavado con BSA al 2 %. Se lavó la placa 3 veces, se añadió el anticuerpo de detección a la placa y se incubó durante 30 minutos a 37 °C. Para revelar la placa, se preparó el sustrato TMB (Bio-Rad, Hercules, CA). Se lavó la placa 3 veces, se añadieron 100 µl de sustrato a cada pocillo y se incubó durante 15 minutos a 37 °C. Se detuvo la reacción con ácido sulfúrico 2 N (100 µl/pocillo) y se leyó la placa en un lector de placas de microtitulación a 450 nm, usando 655 nm como la longitud de onda de referencia.

**N-acetilgalactosamina 6-sulfatasa (GALNS).** Dos anticuerpos anti-GALNS monoclonales de ratón proporcionaron la base del ELISA de GALNS. Los anticuerpos anti-GALNS monoclonales de ratón también se desarrollaron por contrato por Maine Biotechnology Services, Inc. (Portland, ME) hasta dar lugar a un polipéptido de GALNS humano, de longitud completa, producido de manera recombinante, purificado usando tecnología de producción de hibridomas estándar. Se usó el imprimación anticuerpo para la captura de GALNS para recubrir una placa F96 MaxiSorp Nunc-Immuno (Nalge Nunc, nº de catálogo 442404) en un tampón de recubrimiento (bicarbonato sódico 50 mM, pH 9,6). Después de la incubación durante una hora a 37 °C y el lavado con un tampón de lavado, se bloqueó la placa con tampón de bloqueo (PBS, Tween-20 al 0,05 %, BSA al 2 %) durante una hora a 37 °C. Entonces se cargaron en la placa las muestras experimentales y de control junto con los patrones de GALNS y se incubaron adicionalmente durante una hora a 37 °C. Después de lavar con un tampón de lavado, se aplicó el segundo anticuerpo de detección conjugado con HRP en tampón de bloqueo seguido por incubación de 30 minutos a 37 °C. Después de lavar la placa de nuevo, se añadió el reactivo de sustrato de TMB de Bio-Rad y se incubó durante 15 minutos. Entonces se añadió ácido sulfúrico 2 N para detener la reacción y los resultados se puntuaron espectrofotométricamente usando un lector de placas de Molecular Device a una longitud de onda de 450 nm.

**Discusión****Efecto de FGE sobre la actividad de sulfatasa**

**GALNS.** Se observó un aumento de aproximadamente 50 veces en la actividad de GALNS total con respecto a los niveles control (Figura 5). Se observó este nivel de actividad aumentada con los tres puntos de tiempo de muestreo de medio. Además, la actividad de GALNS se acumuló linealmente a lo largo del tiempo con un aumento de cuatro veces entre las 24 y las 48 horas y un aumento de dos veces entre los momentos de tiempo de 48 horas y 72 horas.

**I2S.** Aunque de menor magnitud absoluta, se observó un efecto similar para la actividad de I2S total, observándose un aumento de aproximadamente 5 veces en la actividad de I2S total con respecto a los niveles control. Este nivel de actividad aumentada se mantuvo durante la duración del experimento. La actividad de I2S se acumuló en el medio linealmente a lo largo del tiempo, similarmente a los resultados observados 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 FGE sobre la actividad de sulfatasa específica**

**GALNS.** La expresión de FGE en células 36F potenció la actividad específica aparente de GALNS (razón de actividad enzimática con respecto a enzima total estimada mediante ELISA) en 40-60 veces con respecto a los niveles control (Figura 6). El aumento en la actividad específica se mantuvo durante los tres momentos de tiempo en el estudio y pareció aumentar durante los tres días de acumulación tras la transfección.

**I2S.** Se observó un efecto similar con I2S, observándose un aumento de 6-7 veces en la actividad específica (3-5 U/mg) con respecto a los valores de control (0,5-0,7 U/mg).

Los valores del ELISA tanto para GALNS (Figura 7) como para I2S no resultaron significativamente afectados por la transfección de FGE. Esto indica que la expresión de FGE no altera las rutas de traducción y secreción que participan en la producción de sulfatasa.

En resumen, todos estos resultados para ambas sulfatasas indican que la expresión de FGE aumenta espectacularmente la actividad específica de sulfatasa en líneas celulares que expresan en exceso GALNS e I2S.



**Coexpresión de FGE (SUMF1) y otros genes de sulfatasa**

Para probar el efecto de FGE (SUMF1) sobre actividades de sulfatasa adicionales en células normales, los presentes inventores expresaron en exceso ADNc de ARSA (SEC ID N°: 14), ARSC (SEC ID N°: 18) y ARSE (SEC ID N°: 22) en diversas líneas celulares con y sin co-transfección del ADNc de FGE (SUMF1) y midieron las actividades de sulfatasa. La expresión en exceso de ADNc de sulfatasa en células Cos-7 produjo un aumento moderado de la actividad de sulfatasa, mientras que se observó un sorprendente aumento sinérgico (20 a 50 veces) cuando se co-expresaron tanto un gen de sulfatasa como un gen de FGE (SUMF1). Se observó un efecto similar, aunque inferior, en tres líneas celulares adicionales, HepG2, LE293 y U2OS. La expresión en exceso simultánea de múltiples ADNc de sulfatasa produjo un aumento inferior de cada actividad de sulfatasa específica en comparación con la expresión en exceso de una única sulfatasa, que indica la presencia de competición de las diferentes sulfatasas por la maquinaria de modificación.

Para probar la conservación funcional del gen de FGE (SUMF1) durante la evolución, los presentes inventores expresaron en exceso ADNc de ARSA, ARSC y ARSE en diversas líneas celulares con y sin cotransfección del ADNc de MSD y midieron las actividades de sulfatasa. Tanto los genes de FGE (SUMF1) murinos como los de *Drosophila* eran activos en las tres sulfatasas humanas, siendo la FGE (SUMF1) de *Drosophila* menos eficaz. Estos datos demuestran un alto grado de conservación funcional de FGE (SUMF1) durante la evolución, que implica una importancia biológica significativa para la función y supervivencia celulares. Se observó un efecto similar y constante, aunque mucho más débil, usando el gen de FGE2 (SUMF2), sugiriendo que la proteína codificada por este gen también tiene una actividad modificante de sulfatasa. Estos datos demuestran que la cantidad de la proteína FGE (SUMF1) codificada es un factor limitante para las actividades de sulfatasa, un hallazgo con importantes implicaciones para la producción a gran escala de sulfatasas activas que van a utilizarse en la terapia de sustitución de enzimas.

**Ejemplo 4:**

*Identificación del gen mutado en MSD por medio de complementación funcional usando transferencia de cromosomas mediada por microcélulas.*

En un experimento separado usando transferencia de cromosomas mediada por microcélulas por medio de complementación funcional, los presentes inventores confirmaron que el gen mutado en MSD es FGE. Los hallazgos de los inventores proporcionan un punto de vista adicional en un mecanismo biológico novedoso que afecta a una familia de proteínas entera en organismos relacionados de manera distante. Además de identificar la base molecular de una enfermedad genética poco frecuente, los datos de los presentes inventores confirman además un efecto de potenciación poderoso del producto génico de FGE sobre la actividad de sulfatasas. Este último hallazgo tiene implicaciones clínicas directas para la terapia de al menos ocho enfermedades humanas producidas por deficiencias de sulfatasa.

**El gen para MSD mapea en el cromosoma 3p26**

Para identificar la localización cromosómica del gen mutado en MSD, los presentes inventores intentaron rescatar las enzimas deficientes en sulfatasa mediante complementación funcional a través de transferencia de cromosomas mediada por microcélulas. Se usó un panel de líneas celulares híbridas humanas/de ratón, que contenían cromosomas humanos normales individuales marcados con el marcador de selección dominante HyTK, como fuente de cromosomas humanos donadores y se fusionaron a una línea celular inmortalizada de un paciente con MSD. Se transfirieron los 22 autosomas humanos uno por uno a la línea celular del paciente y se seleccionaron los híbridos en higromicina. Se recogieron aproximadamente 25 colonias supervivientes en cada uno de los 22 experimentos de transferencia. Se hicieron crecer éstas por separado y se recogieron para las pruebas enzimáticas posteriores. Se probaron las actividades de arilsulfatasa A (ARSA) (SEC ID N°: 15), arilsulfatasa B (ARSB) (SEC ID N°: 17) y arilsulfatasa C (ARSC) (SEC ID N°: 19) para cada uno de los aproximadamente 440 clones (20 x 22). Este análisis indicó claramente que las actividades de sulfatasa de varios clones que se derivaban de la transferencia del cromosoma 3 eran significativamente superiores en comparación con las de los otros clones. Se observó una variabilidad sorprendente cuando se analizaron las actividades de cada clon individual a partir de la transferencia del cromosoma. Para verificar si cada clon tenía un cromosoma 3 humano intacto procedente de la línea celular donante, los presentes inventores usaron un panel de 23 marcadores genéticos polimórficos del cromosoma 3, distribuidos de manera uniforme a lo largo de la longitud del cromosoma y seleccionados previamente basándose en que tienen diferentes alelos entre las líneas celulares de donante y del paciente. Esto permitió examinar la presencia del cromosoma donante e identificar la posible pérdida de regiones específicas debido a la rotura cromosómica imprevista. Cada clon que tenía alta actividad enzimática retuvo el cromosoma 3 completo de la línea celular donante, mientras que los clones con bajas actividades parecían haber perdido el cromosoma completo basándose en la ausencia de alelos del cromosoma 3 de la línea celular donante. Estos últimos clones probablemente retuvieron una región pequeña del cromosoma donante que contenía el gen marcador de selección que les permitía sobrevivir en medio que contenía higromicina. Estos datos indican que un cromosoma 3 humano normal era capaz de complementar el defecto observado en la línea celular del paciente con MSD.

Para determinar la región cromosómica específica que contenía el gen responsable de la actividad de

complementación, los presentes inventores usaron híbridos del cromosoma 3 marcados con Neo que se encontró que habían perdido varias porciones del cromosoma. Además, los presentes inventores realizaron transferencia de cromosomas mediada por microcélulas irradiadas de cromosomas 3 humanos marcados con HyTK. Se probaron ciento quince híbridos irradiados de cromosoma 3 para actividades de sulfatasa y se genotiparon usando un panel de 31 marcadores microsatélites polimórficos que abarcaban el cromosoma completo. Todos los clones que presentaban altas actividades enzimáticas parecían haber retenido el cromosoma 3p26. Un análisis de resolución superior usando marcadores adicionales de esta región mapeó la posible localización para el gen de complementación entre los marcadores *D3S3630* y *D3S2397*.

10 **Identificación del gen mutado en MSD**

Los presentes inventores investigaron los genes de la región genómica 3p26 para mutaciones en pacientes con MSD. Se amplificó por PCR cada exón incluyendo las uniones de corte y empalme y se analizó mediante secuenciación directa. Se realizó el análisis de mutación en doce individuos afectados no relacionados; cinco pacientes con MSD descritos previamente y siete casos no publicados. Se identificaron varias mutaciones de la cohorte de MSD de los presentes inventores en el marcador de secuencia expresada (EST) AK075459 (SEC ID N°: 4, 5), correspondientes a un gen de función desconocida, sugiriendo fuertemente que éste era el gen implicado en MSD. Se encontró que cada mutación estaba ausente en 100 individuos de control, excluyendo así la presencia de un polimorfismo de secuencia. Se realizó un análisis de mutación de confirmación adicional en ARN de pacientes sometidos a transcripción inversa, particularmente en aquellos casos en los que el análisis del ADN genómico reveló la presencia de una mutación en o cerca de un sitio de corte y empalme, que afecta posiblemente el corte y empalme. También se identificaron mutaciones del marco de lectura, terminadora, de corte y empalme y de aminoácido, sugiriendo que la enfermedad se produce por un mecanismo de pérdida de función, como se prevé para un trastorno recesivo. Esto también concuerda con la observación de que casi todas las mutaciones de aminoácido afectan a aminoácidos que están altamente conservados a lo largo de la evolución (véase a continuación).

30 **Tabla 8: Mutaciones en MSD adicionales identificadas**

Caso	Referencia	fenotipo	exon	Cambio nucleotido	Cambio aminoácido
1. BA426	Conary et al, 1988	moderado	3	463T>C	S155P
			3	463T>C	S155P
2. BA428	Burch et al, 1986	Neonato severo	5	661delG	desplazamiento
3. BA431	Zenger et al, 1989	moderado	1	2T>G	M1R
			2	276delC	desplazamiento
4. BA799	Burk et al, 1981	Leve a moderado	3	463T>C	S155P
			3	463T>C	S155P
5. BA806	No publicado	Neonato severo	9	1045T>C	R349W
6. BA807	Schmidt et al, 1995	desconocido	3	c519+4delGTAA	ex 3 skipping
7. BA809	Couchot et al, 1974	Leve a moderado	1	1A>G	M1V
			9	1042G>C	A348P
8. BA810	No publicado	severo	8	1006T>C	C336R
			9		
9. BA811	No publicado	Neonato severo	3	c519+4delGTAA	ex 3 skipping
			8	979C>T	R327X
10. BA815	No publicado	moderado	5	c.603-6delC	ex 6 skipping
			6	836C>T	A279V
11. BA919	No publicado	Leve a 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 probado, excluyendo así la heterogeneidad de locus. No se observó correlación obvia entre los tipos de mutaciones identificados y la gravedad del fenotipo informado en los pacientes, sugiriendo que la variabilidad clínica no está producida por heterogeneidad alélica. En tres casos, se encontró que diferentes pacientes (caso 1 y 4, caso 6 y 9, y caso 11 y 12 en la Tabla 6) portaban la misma mutación. Dos de estos pacientes (caso 11 y 12) proceden de la misma ciudad en Sicilia, sugiriendo la presencia de un efecto fundador que de hecho se confirmó mediante análisis de haplotipos. Sorprendentemente, se encontró que la

mayoría de los pacientes se componía de heterocigotos que portaban diferentes mutaciones alélicas, mientras que solo algunos eran homocigotos. Aunque de acuerdo con la ausencia de consanguineidad informada por los progenitores, este fue un hallazgo algo inesperado para un trastorno recesivo muy poco frecuente tal como MSD.

#### El gen y la proteína FGE

Se reunió la secuencia de ADNc consenso del ADNc de la FGE humana (también usada de manera intercambiable en el presente documento como SUMF1) (SEC ID N°: 1) a partir de varios clones de marcador de secuencia expresada (EST) y parcialmente a partir de la secuencia genómica correspondiente. El gen contiene nueve exones y abarca aproximadamente 105 kb (véase el Ejemplo 1). La comparación de secuencias también identificó la presencia de un gen de FGE parálogo localizado en el cromosoma 7 humano que los presentes inventores designaron FGE2 (también usado de manera intercambiable en el presente documento como SUMF2) (SEC ID N°: 45, 46).

#### Complementación funcional de deficiencias de sulfatasa

Se infectaron fibroblastos de dos pacientes (caso 1 y 12 en la Tabla 8) con MSD en los que los presentes inventores identificaron mutaciones del gen de FGE (SUMF1) (líneas celulares BA426 y BA920) con virus VHS que contenían las formas no mutada y dos formas mutada del ADNc de FGE (SUMF1) (R327X y  $\Delta$ ex3). Se probaron las actividades ARSA, ARSB y ARSC 72 h después de la infección. La expresión del ADNc de FGE no mutante (SUMF1) produjo la complementación funcional de las tres actividades, mientras que el ADNc de FGE (SUMF1) mutante no (Tabla 9). Estos datos proporcionan una evidencia concluyente para la identidad de FGE (SUMF1) como el gen en MSD y demuestran la relevancia funcional de las mutaciones encontradas en los pacientes. Las mutaciones asociadas con la enfermedad producen deficiencia de sulfatasa, demostrando así que FGE (SUMF1) es un factor esencial para la actividad de sulfatasa.

**Tabla 9: Complementación funcional de deficiencias de sulfatasa**

Recipiente Línea celular MSD	Contrucción	ARSA <sup>(1)</sup>	ARSB <sup>(1)</sup>	ARSC <sup>(1)</sup>
BA426	HSV amplicon	24.0	22.5	0.15
	SUMF1- $\Delta$ ex3	42.0	23.8	0.29
	SUMF1-R327X	33.6	24.2	0.16
	SUMF1	119.5 (4.9x)	37.8 (1.7x)	0.62 (4.1x)
BA920	HSV amplicon	16.6	11.3	0.15
	SUMF1- $\Delta$ ex3	17.2	14.4	0.07
	SUMF1-R327X	36.0	13.5	0.13
	SUMF1	66.5 (4.0x)	21.6 (1.9x)	0.42 (2.8x)
Rango control		<b>123.7-394.6</b>	<b>50.6-60.7</b>	<b>1.80-1.58</b>

(1) Todas las actividades enzimáticas se expresan como nmoles de 4 metilumbeliferona liberada 'de proteínas mg<sup>-1</sup>' 3 hrs. MSD líneas celulares BA426 y BA920 se infectaron con el amplicón HSV solo, y con construcciones que llevan ya sea mutante o de tipo salvaje SUMF1 ADNc. El aumento de las actividades de arilsulfatasa individuales en los fibroblastos infectados con el gen SUMF1 de tipo salvaje, en comparación con los de las células infectadas con el vector solo, se indica entre paréntesis. Actividades medidos en los fibroblastos control no infectadas se indican.

#### Base molecular de MSD

Basándose en la hipótesis de que el gen de la enfermedad debe ser capaz de complementar la deficiencia enzimática en una línea celular del paciente, los presentes inventores realizaron transferencia de cromosomas mediada por microcélulas a una línea celular inmortalizada de un paciente con MSD. Esta técnica se ha usado satisfactoriamente para la identificación de genes cuya función predicha podía evaluarse en líneas celulares (por ejemplo, midiendo la actividad enzimática o detectando características morfológicas). Para abordar el problema de la variabilidad estocástica de la actividad enzimática, los presentes inventores midieron las actividades de tres sulfatasas diferentes (ARSA, ARSB y ARSC) en el ensayo de complementación. Los resultados de la transferencia de cromosomas indicaron claramente el mapeo del gen de complementación en el cromosoma 3. Se logró un mapeo subregional generando un panel de híbridos de radiación para el cromosoma 3. Se caracterizaron clones híbridos individuales tanto a nivel genómico, tipificando 31 marcadores microsatélites que presentaban diferentes alelos entre las líneas celulares donantes y receptoras, como a nivel funcional probando las actividades de sulfatasa. El análisis de 130 de tales híbridos produjo el mapeo de la región de complementación en el cromosoma 3p26.

Una vez definida la región genómica crítica, también se identificó el gen de FGE (SUMF1) mediante análisis de mutación en ADN de los pacientes. Se encontraron mutaciones en todos los pacientes probados, que demuestra que un único gen está implicado en la MSD. Las mutaciones encontradas eran de diferentes tipos, produciendo la mayoría (por ejemplo, sitio de corte y empalme, sitio de inicio, terminadora, de cambio de marco) posiblemente una

función perdida de la proteína codificada, como era de esperar para una enfermedad recesiva. La mayoría de las mutaciones de aminoácido afectan a codones correspondientes a aminoácidos que están altamente conservados durante la evolución, sugiriendo que estas mutaciones también producen una pérdida de función. No pudieron establecerse correlaciones entre el tipo de mutación y la gravedad del fenotipo, que indica que este último se debe a factores no relacionados. Inesperadamente para una enfermedad genética poco frecuente, se encontró que muchos pacientes se componían de heterocigotos, que portaban dos mutaciones diferentes. Sin embargo, se identificó un efecto fundador para una mutación que se originó en una pequeña ciudad de Sicilia.

### Función del gen de FGE (SUMF1)

Se demostró de manera definitiva la identidad del gen de FGE (SUMF1) como el "factor de complementación" rescatando la deficiencia enzimática de cuatro sulfatasas diferentes con la expresión del ADNc de FGE (SUMF1) exógeno, insertado en un vector viral, en dos líneas celulares de pacientes diferentes. En cada caso se observó una restauración constante, aunque parcial, de todas las actividades probadas, en comparación con las líneas celulares de pacientes de control transfectadas con vectores vacíos. En promedio, el aumento de las actividades enzimáticas osciló entre 1,7 y 4,9 veces y alcanzó aproximadamente la mitad de los niveles observados en las líneas celulares normales. La actividad enzimática se correlaciona con el número de partículas virales usadas en cada experimento y con la eficacia de la infección como se prueba mediante un análisis con proteína marcadora (GFP). En los mismos experimentos, se usaron vectores que contenían ADNc de FGE (SUMF1) que portaban dos de las mutaciones encontradas en los pacientes, R327X y  $\Delta$ ex3, y no se observó aumento significativo de la actividad enzimática, demostrando así la relevancia funcional de estas mutaciones.

Como se menciona en otra parte en el presente documento, Schmidt et al. descubrieron en imprimación lugar que las sulfatasas experimentan una modificación postraduccional de una cisteína altamente conservada, que se encuentra en el sitio activo de la mayoría de las sulfatasas, a C $\alpha$ -formilglicina. También demostraron que esta modificación era defectuosa en MSD (Schmidt, B., et al., Cell, 1995, 82:271-278). Los datos funcionales y de mutación de los presentes inventores proporcionan una fuerte evidencia de que FGE (SUMF1) es responsable de esta modificación.

El gen de FGE (SUMF1) muestra un grado extremadamente alto de conservación de secuencia en todas las especies relacionadas de manera distante analizadas, desde las bacterias hasta el hombre. Los presentes inventores proporcionan evidencia de que el homólogo de *Drosophila* del gen de FGE humana (SUMF1) es capaz de activar sulfatasas humanas expresadas en exceso, demostrando que el alto nivel de similitud de secuencia observado de los genes de FGE (SUMF1) de especies relacionadas de manera distante se correlaciona con una conservación funcional sorprendente. Una excepción notable son las levaduras, que parecen carecer del gen de FGE (SUMF1), así como de cualquier gen que codifique sulfatasas, que indica que este organismo no requiere la función sulfatasa y sugiere la presencia de una influencia recíproca en la evolución de los genes de FGE (SUMF1) y sulfatasas.

De forma interesante, hay dos genes homólogos, FGE (SUMF1) y FGE2 (SUMF2), en los genomas de todos los vertebrados analizados, incluyendo los seres humanos. Como es evidente a partir del árbol filogenético, el gen de FGE2 (SUMF2) parece haber evolucionado independientemente del gen de FGE (SUMF1). En los ensayos de los presentes inventores, el gen de FGE2 (SUMF2) también es capaz de activar sulfatasas, sin embargo lo hace de una manera mucho menos eficaz en comparación con el gen de FGE (SUMF1). Esto puede explicar la actividad de sulfatasa residual encontrada en pacientes con MSD y sugiere que una deficiencia de sulfatasa completa sería letal. En este momento los presentes inventores no pueden descartar la posibilidad de que el gen de FGE2 (SUMF2) tenga una función adicional, todavía desconocida.

### Impacto sobre la terapia de enfermedades debidas a deficiencias de sulfatasas

Se observó un fuerte aumento, de hasta 50 veces, de las actividades de sulfatasa en células que expresan en exceso ADNc de FGE (SUMF1) junto con o bien ADNc de ARSA, ARSC, o bien ARSE, en comparación con células que expresan en exceso sulfatasas individuales solas. En todas las líneas celulares se encontró un efecto sinérgico significativo, que indica que FGE (SUMF1) es un factor limitante para la actividad de sulfatasa. Sin embargo, se observó variabilidad entre diferentes sulfatasas, posiblemente debido a la diferente afinidad de la proteína codificada por FGE (SUMF1) con las diversas sulfatasas. También se observó variabilidad entre diferentes líneas celulares que pueden tener diferentes niveles de enzima generadora de formilglicina endógena. De acuerdo con estas observaciones, los presentes inventores encontraron que la expresión del gen de MSD varía entre diferentes tejidos, con niveles significativamente altos en riñón e hígado. Esto puede tener implicaciones importantes, ya que los tejidos con bajos niveles de expresión del gen de FGE (SUMF1) pueden ser menos capaces de modificar de manera eficaz las proteínas sulfatasas administradas de manera exógena (véase más adelante). Estos datos juntos sugieren que la función del gen de FGE (SUMF1) ha evolucionado para lograr un sistema de regulación doble, controlándose cada sulfatasa tanto mediante un mecanismo individual, responsable de los niveles de ARNm de cada gen de sulfatasa estructural, como mediante un mecanismo común compartido por todas las sulfatasas. Además, FGE2 (SUMF2) proporciona redundancia parcial para la modificación de sulfatasas.

Estos datos tiene profundas implicaciones para la producción en masa de sulfatasas activas que van a utilizarse en la terapia de sustitución de enzimas. Se han informado estudios de sustitución de enzimas en modelos animales de deficiencias de sulfatasas, tales como un modelo fenilo de mucopolisacaridosis VI, y se ha demostrado que son eficaces para prevenir y curar diversos síntomas. En la actualidad se están realizando ensayos terapéuticos en seres humanos para dos trastornos congénitos debidos a deficiencias de sulfatasas, MPSII (síndrome de Hunter) y MPSVI (síndrome de Maroteaux-Lamy) y pronto se ampliará a un número mayor de pacientes.

#### **Ejemplo 5:**

10 *Terapia de sustitución de enzimas con GALNS activada por FGE para MPS IVA de la enfermedad de Morquio* La causa principal de la patología esquelética en pacientes con Morquio es la acumulación de queratán sulfato (KS) en condrocitos del disco epifisiario (placa de crecimiento) debido a la deficiencia de la sulfatasa lisosómica, GALNS. El objetivo primario de estudios de investigación *in vivo* fue determinar si GALNS activada por FGE administrada por vía intravenosa (i.v.) era capaz de penetrar en los condrocitos de la placa de crecimiento, así como en otros tipos celulares apropiados en ratones normales. Pese a la falta general de anomalías esqueléticas, también se usó un modelo de ratón deficiente en GALNS (Morquio Knock-In-MKI, S. Tomatsu, Universidad de St. Louis, MO) para demostrar la actividad bioquímica *in vivo* de GALNS activada por FGE administrada repetidamente. La falta de patología esquelética en los modelos de ratón refleja el hecho de que el KS esquelético o bien está enormemente reducido o bien está ausente en roedores (Venn G, & Mason RM., *Biochem J.*, 1985, 228:443-450). Sin embargo, estos ratones sí demostraron acumulación detectable de GAG y otras anomalías celulares en diversos órganos y tejidos. Por tanto, el objetivo global de los estudios fue demostrar que GALNS activada por FGE penetra en la placa de crecimiento (estudio de biodistribución) y muestran la actividad enzimática de GALNS funcional dirigida hacia la eliminación de GAG acumulados en los tejidos afectados (estudio farmacodinámico).

25 Los resultados de estos estudios demostraron que GALNS activada por FGE inyectada por vía i.v. se internalizó por los condrocitos de la placa de crecimiento, aunque a niveles relativamente bajos en comparación con otros tejidos. Además, la inyección de GALNS activada por FGE durante el transcurso de 16 semanas en ratones MKI produjo el aclaramiento eficaz de los GAG acumulados y redujo la tinción con biomarcador lisosómico en todos los tejidos blandos examinados. En resumen, los experimentos demostraron satisfactoriamente la administración de GALNS a los condrocitos de la placa de crecimiento y demostraron la actividad bioquímica en lo que se refiere al aclaramiento de GAG en múltiples tejidos.

#### **Estudio de biodistribución**

35 Se administró a ratones ICR (normales) de cuatro semanas de edad una única inyección i.v. de 5 mg/kg de GALNS activada por FGE. Se recogieron el hígado, fémur (hueso), corazón, riñón y bazo dos horas tras la inyección y se prepararon para el examen histológico. Se usó un anticuerpo monoclonal anti-GALNS humano para detectar la presencia de GALNS inyectada en los diversos tejidos. Se detectó GALNS en todos los tejidos examinados en comparación con los controles con vehículo. Además, GALNS se observó fácilmente en todos los tejidos examinados usando un sistema indicador con peroxidasa de rábano, con la excepción del hueso. La demostración de la captación de GALNS en la placa de crecimiento requirió el uso de un sistema indicador de isotiocianato de fluoresceína (FTTC) más sensible e indica que aunque GALNS penetra en la placa de crecimiento, está menos fácilmente disponible para los condrocitos de la placa de crecimiento que para las células de los tejidos blandos. Pese al requisito de un método de detección de fluorescencia más sensible, se observó la administración de GALNS a los condrocitos de la placa de crecimiento ósea en todas las secciones de placa de crecimiento examinadas en comparación con los controles con vehículo.

#### **Estudio farmacodinámico en ratones MKI**

50 Se administraron a ratones MKI o no mutantes de cuatro semanas de edad inyecciones i.v. administradas semanalmente (n = 8 por grupo) hasta las 20 semanas de edad. Cada inyección semanal consistió en o bien 2 mg/kg de GALNS activada por FGE o bien control de vehículo (sin inyección para los ratones no mutantes). Se sacrificaron todos los ratones para el examen histológico a las 20 semanas de edad y se tiñeron usando los siguientes métodos: hematoxilina y eosina para morfología celular, azul alciano para la detección de GAG.

55 Se demostró el aclaramiento de los GAG acumulados mediante tinción con azul alciano reducida o ausente en todos los tejidos blandos examinados (hígado, corazón, riñón y bazo). Esto se observó solo en los ratones inyectados con GALNS. Aunque la placa de crecimiento en los ratones MKI funcionaba normalmente como se demuestra por una morfología esquelética normal, hubo más anomalías celulares sutiles observadas (incluyendo vacuolización de condrocitos sin efecto patológico aparente). Los condrocitos vacuolizados de las zonas hipertróficas y proliferativas de la placa de crecimiento no resultaron afectados por la administración de GALNS. Esto contrastó con los condrocitos en la zona de calcificación de la placa de crecimiento en los que se observó una reducción de la vacuolización en los ratones inyectados con GALNS. La vacuolización de los condrocitos y la acumulación de GAG supuestamente distintos de KS en la placa de crecimiento en los ratones MKI fueron, en general, sorprendentes e inesperadas debido a la falta conocida de KS en la placa de crecimiento de ratones. Estas observaciones particulares reflejan probablemente el hecho de que, en ratones inactivados, están presentes altos niveles de

GALNS mutante (a diferencia de los ratones silenciados en los que no hay GALNS mutante residual, ni vacuolización de condrocitos de la placa de crecimiento ni acumulación de GAG - Tomatsu S. et al., Human Molecular Genetics, 2003, 12:3349-3358). El fenómeno de vacuolización en la placa de crecimiento puede ser indicativo de un efecto secundario de un subconjunto de células que expresan GALNS mutante. No obstante, la inyección de enzima durante el transcurso de 16 semanas demostró una fuerte evidencia de la administración de GALNS activada por FGE a múltiples tejidos y de la actividad enzimática *in vivo*.

### Descripción detallada de los dibujos

**Fig. 1: Espectros de masas MALDI-TOF de P23 después de la incubación en ausencia (A) o presencia (B) de un extracto soluble de microsomas de testículos bovinos.** Se incubaron 6 pmoles de P23 en condiciones normales durante 10 min a 37 °C en ausencia o presencia de 1 µl de extracto microsómico. Se prepararon las muestras para espectrometría de masas MALDI-TOF como se describe en Procedimientos experimentales. Se indican las masas monoisotópicas  $MH^+$  de P23 (2526,28) y su derivado de *FGly* (2508,29).

**Fig. 2: Árbol filogenético derivado de un alineamiento de FGE humana y 21 proteínas de la semilla DUF323 en PFAM.** Los números en las ramas indican distancia filogenética. Las proteínas se designan mediante su número ID de TrEMBL y el nombre de la especie. hFGE - FGE humana. Parte superior derecha: escala de las distancias filogenéticas. Un asterisco indica que el gen se ha investigado adicionalmente. Los siete genes superiores forman parte de la familia de genes de FGE.

**Fig. 3: Organización del locus del gen de FGE humana y murina.** Los exones se muestran a escala como recuadros oscuros (locus humano) y recuadros claros (locus murino). La barra en la esquina inferior derecha muestra la escala. Las líneas entre los exones muestran los intrones (que no están a escala). Los números encima de las líneas de intrón indican el tamaño de los intrones en kilobases.

**Fig. 4: Diagrama que muestra un mapa de plásmido pXMG.1.3 de expresión de FGE**

**Fig. 5: Gráfico de barras que representa la actividad N-acetilgalactosamina 6-sulfatasa en células 36F transfectadas de manera transitoria con plásmido de expresión de FGE.** Se transfectaron las células con bien un plásmido control, pXMG.1.2, con el ADNc de FGE en la orientación inversa, o bien un plásmido de expresión de FGE, pXMG.1.3 en medios sin metrotrexato (MTX). Se volvió a alimentar a las células 24 horas más tarde con medios que contenían MTX 1,0 µM. Se recogió el medio y se recogieron las células 24, 48 y 72 horas después de volver a alimentar. Se determinó la actividad N-acetilgalactosamina 6-sulfatasa mediante el ensayo de actividad. Cada valor mostrado es el promedio de dos transfecciones separadas indicándose las desviaciones estándar mediante barras de error.

**Fig. 6: Gráfico de barras que representa la actividad específica N-acetilgalactosamina 6-sulfatasa en células 36F transfectadas de manera transitoria con plásmido de expresión de FGE.** Se transfectaron las células con o bien un plásmido de control, pXMG.1.2, con el ADNc de FGE en la orientación inversa, o bien un plásmido de expresión de FGE, pXMG.1.3 en medios sin metrotrexato (MTX). Se volvió a alimentar a las células 24 horas más tarde con medios que contenían MTX 1,0 µM. Se recogió el medio y se recogieron las células 24, 48 y 72 horas después de volver a alimentar. Se determinó la actividad específica N-acetilgalactosamina 6-sulfatasa mediante el ensayo de actividad y ELISA y se representa como la razón de actividad N-acetilgalactosamina 6-sulfatasa por mg de N-acetilgalactosamina 6-sulfatasa reactiva en ELISA. Cada valor mostrado es el promedio de dos transfecciones separadas.

**Fig. 7: Gráfico de barras que representa producción de N-acetilgalactosamina 6-sulfatasa en células 36F transfectadas de manera transitoria con plásmido de expresión de FGE.** Se transfectaron las células con o bien un plásmido control, pXMG.1.2, con el ADNc de FGE en la orientación inversa, o bien un plásmido de expresión de FGE, pXMG.1.3 en medios sin metrotrexato (MTX). Se volvió a alimentar a las células 24 horas más tarde con medios que contenían MTX 1,0 µM. Se recogió el medio y se recogieron las células 24, 48 y 72 horas después de volver a alimentar. Se determinó la proteína total N-acetilgalactosamina 6-sulfatasa mediante ELISA. Cada valor mostrado es el promedio de dos transfecciones separadas indicándose las desviaciones estándar mediante barras de error.

**Fig. 8: Gráfico que representa actividad iduronato 2-sulfatasa en células 3006 transfectadas de manera transitoria con plásmido de expresión de FGE.** Se transfectaron las células con o bien un plásmido control, pXMG.1.2, con el ADNc de FGE en la orientación inversa, o bien un plásmido de expresión de FGE, pXMG.1.3 en medios sin metrotrexato (MTX). Se volvió a alimentar a las células 24 horas más tarde con medios que contenían MTX 1,0 µM. Se recogió el medio y se recogieron las células 24, 48 y 72 horas tras volver a alimentar. Se determinó la actividad iduronato 2-actividad mediante el ensayo de actividad. Cada valor mostrado es el promedio de dos transfecciones separadas.

Fig. 9: Representa un kit que incluye características de la presente invención.

Todas las referencias dadas a conocer en el presente documento se incorporan como referencia en su totalidad. Lo que se reivindica se presenta a continuación y va seguido por una lista de secuencias.

5 LISTA DE SECUENCIAS

<110> Transkaryotic Therapies, Inc. von Figura, Kurt Schmidt, Bernhard Dierks, Thomas Heartlein, Michael W. Cosma, Maria P. Ballabio, Andrea

10 <120> DIAGNÓSTICO Y TRATAMIENTO DE MÚLTIPLES SULFATASA DEFICIENCIA Y OTRAS DEFICIENCIAS SULFATASA

15 <130> 0403WO

<150> US 60/447,747

<151> 2003-02-11

20 <160> 95

<170> PatentIn version 3.2

<210> 1

<211> 1180

25 <212> ADN

<213> Homo sapiens

<220>

<221> CDS

30 <222> (20)..(1141)

<223> FGE cADN

<400> 1

35

40

45

50

55

60

65

ES 2 555 056 T3

```

5      acatggcccg cgggacaac atg gct gcg ccc gca cta ggg ctg gtg tgt gga      52
      Met Ala Ala Pro Ala Leu Gly Leu Val Cys Gly
      1                      5                      10

10     cgt tgc cct gag ctg ggt ctc gtc ctc ttg ctg ctg ctg ctc tcg ctg      100
      Arg Cys Pro Glu Leu Gly Leu Val Leu Leu Leu Leu Leu Leu Ser Leu
      15                      20                      25

15     ctg tgt gga gcg gca ggg agc cag gag gcc ggg acc ggt gcg ggc gcg      148
      Leu Cys Gly Ala Ala Gly Ser Gln Glu Ala Gly Thr Gly Ala Gly Ala
      30                      35                      40

20     ggg tcc ctt gcg ggt tct tgc ggc tgc ggc acg ccc cag cgg cct gcc      196
      Gly Ser Leu Ala Gly Ser Cys Gly Cys Gly Thr Pro Gln Arg Pro Gly
      45                      50                      55

25     gcc cat ggc agt tcg gca gcc gct cac cga tac tcg cgg gag gct aac      244
      Ala His Gly Ser Ser Ala Ala Ala His Arg Tyr Ser Arg Glu Ala Asn
      60                      65                      70                      75

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

35     gtc ccc atc cct gct gga gta ttt aca atg ggc aca gat gat cct cag      340
      Val Pro Ile Pro Ala Gly Val Phe Thr Met Gly Thr Asp Asp Pro Gln
      95                      100                      105

40     ata aag cag gat ggg gaa gca cct gcg agg aga gtt act att gat gcc      388
      Ile Lys Gln Asp Gly Glu Ala Pro Ala Arg Arg Val Thr Ile Asp Ala
      110                      115                      120

45     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

50     gtg aac tca act ggc tat ttg aca gag gct gag aag ttt ggc gac tcc      484
      Val Asn Ser Thr Gly Tyr Leu Thr Glu Ala Glu Lys Phe Gly Asp Ser
      140                      145                      150                      155

55     ttt gtc ttt gaa ggc atg ttg agt gag caa gtg aag acc aat att caa      532
      Phe Val Phe Glu Gly Met Leu Ser Glu Gln Val Lys Thr Asn Ile Gln
      160                      165                      170

60     cag gca gtt gca gct gct ccc tgg tgg tta cct gtg aaa ggc gct aac      580

```

50

55

60

65



5 Gln Ala Val Ala Ala Ala Pro Trp Trp Leu Pro Val Lys Gly Ala Asn  
175 180 185

10 tgg aga cac cca gaa ggg cct gac tct act att ctg cac agg ccg gat 628  
Trp Arg His Pro Glu Gly Pro Asp Ser Thr Ile Leu His Arg Pro Asp  
190 200

15 cat cca gtt ctc cat gtg tcc tgg aat gat gcg gtt gcc tac tgc act 676  
His Pro Val Leu His Val Ser Trp Asn Asp Ala Val Ala Tyr Cys Thr  
205 210 215

20 tgg gca ggg aag cgg ctg ccc acg gaa gct gag tgg gaa tac agc tgt 724  
Trp Ala Gly Lys Arg Leu Pro Thr Glu Ala Glu Trp Glu Tyr Ser Cys  
220 225 230 235

25 cga gga ggc ctg cat aat aga ctt ttc ccc tgg ggc aac aaa ctg cag 772  
Arg Gly Gly Leu His Asn Arg Leu Phe Pro Trp Gly Asn Lys Leu Gln  
240 245 250

30 ccc aaa ggc cag cat tat gcc aac att tgg cag ggc gag ttt ccg gtg 820  
Pro Lys Gly Gln His Tyr Ala Asn Ile Trp Gln Gly Glu Phe Pro Val  
255 260 265

35 acc aac act ggt gag gat ggc ttc caa gga act gcg cct gtt gat gcc 868  
Thr Asn Thr Gly Glu Asp Gly Phe Gln Gly Thr Ala Pro Val Asp Ala  
270 275 280

40 ttc cct ccc aat ggt tat ggc tta tac aac ata gtg ggg aac gca tgg 916  
Phe Pro Pro Asn Gly Tyr Gly Leu Tyr Asn Ile Val Gly Asn Ala Trp  
285 290 295

45 gaa tgg act tca gac tgg tgg act gtt cat cat tct gtt gaa gaa acg 964  
Glu Trp Thr Ser Asp Trp Trp Thr Val His His Ser Val Glu Glu Thr  
300 305 310 315

50 ctt aac cca aaa ggt ccc cct tct ggg aaa gac cga gtg aag aaa ggt 1012  
Leu Asn Pro Lys Gly Pro Pro Ser Gly Lys Asp Arg Val Lys Lys Gly  
320 325 330

55 gga tcc tac atg tgc cat agg tct tat tgt tac agg tat cgc tgt gct 1060  
Gly Ser Tyr Met Cys His Arg Ser Tyr Cys Tyr Arg Tyr Arg Cys Ala  
335 340 345

60 gct cgg agc cag aac aca cct gat agc tct gct tcg aat ctg gga ttc 1108  
Ala Arg Ser Gln Asn Thr Pro Asp Ser Ser Ala Ser Asn Leu Gly Phe  
350 355 360

65 cgc tgt gca gcc gac cgc ctg ccc acc atg gac tgacaacca gggtagtott 1161  
Arg Cys Ala Ala Asp Arg Leu Pro Thr Met Asp  
365 370

70 cccagtcaca aggagcagt 1180

<210> 2  
<211> 374  
<212> PRT  
<213> Homo sapiens

<400> 2

ES 2 555 056 T3

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

ES 2 555 056 T3

5  
65    70    75    80

Pro Gly Glu Arg Gln Leu Ala His Ser Lys Met Val Pro Ile Pro Ala  
85    90

10 Gly Val Phe Thr Met Gly Thr Asp Asp Pro Gln Ile Lys Gln Asp Gly  
100    105

15 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

20 Tyr Leu Thr Glu Ala Glu Lys Phe Gly Asp Ser Phe Val Phe Glu Gly  
145    150    155    160

25 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

30 Gly Pro Asp Ser Thr Ile Leu His Arg Pro Asp His Pro Val Leu His  
195    200    205

35 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

40 Asn Arg Leu Phe Pro Trp Gly Asn Lys Leu Gln Pro Lys Gly Gln His  
245    250    255

45 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

50 Tyr Gly Leu Tyr Asn Ile Val Gly Asn Ala Trp Glu Trp Thr Ser Asp  
290    295    300

55 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

60 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

65 Arg Leu Pro Thr Met Asp  
370

<210> 3

ES 2 555 056 T3

5 <211> 1122  
 <212> ADN  
 <213> Homo sapiens

<400> 3

10 atggctgcgc ccgcactagg gctggtgtgt ggacgttgcc ctgagctggg tctcgtcctc 60  
 ttgctgctgc tgctctcgct gctgtgtgga gcggcagggg gccaggaggc cgggaccggg 120  
 15 gggggcgcgg ggtcccttgc gggttcttgc ggctgcggca cgtcccagcg gcctggcgcc 180  
 catggcagtt cggcagccgc tcaccgatac tcgctgggagg ctaacgctcc gggcccgtgta 240  
 cccggagagc ggcaactcgc gcaactcaaag atgggtcccca tcctctgtgg agtatttaca 300  
 20 atgggcacag atgatoctca gataaagcag gatggggaag cacctgcgag gagagttact 360  
 attgatgcct tttacatgga tgcctatgaa gtcagtaata ctgaatttga gaagtttgtg 420  
 aactcaactg gctatttgac agaggctgag aagtttggcg actcctttgt ctttgaaggc 480  
 25 atgttgagtg agcaagtgaa gaccaatatt caacaggcag ttgcagctgc tccttgggtgg 540  
 ttacctgtga aaggcgctaa ctggagacac ccagaagggc ctgactctac tattctgcac 600  
 30 aggccggatc atccagttct ccatgtgtcc tggaatgatg cggttgccta ctgcacttgg 660  
 gcaggggaagc ggctgcccac ggaagctgag tgggaataca gctgtcgagg aggcctgcat 720  
 aatagacttt tcccctgggg caacaaaactg cagcccaaag gccagcatta tgccaacatt 780  
 35 tggcagggcg agtttccggg gaccaactact ggtgaggatg gcttccaagg aactgcgcct 840  
 gttgatgcct tcctcccaa tggttatggc ttatacaaca tagtggggaa cgcattggaa 900  
 tggacttcag actgggtggac tgttcatcat tctgttgaag aaacgcttaa cccaaaaggt 960  
 40 cccccttctg ggaaagaccg agtgaagaaa ggtggatcct acatgtgcca taggtcttat 1020  
 tgttacaggt atcgctgtgc tgctoggagc cagaacacac ctgatagctc tgcttcgaat 1080  
 45 ctgggatcc gctgtgcagc cgaccgctg cccaccatgg ac 1122

<210> 4  
 <211> 2130  
 <212> ADN  
 <213> Homo sapiens

<400> 4

55

60

65

ES 2 555 056 T3

5     acatggcccg cgggacaaca tggctgcgcc cgcactaggg ctgggtgtgtg gacgttgccc     60  
       tgagctgggt ctcgctctct tgcctgctgt gctctcgctg ctgtgtggag cggcagggag     120  
 10     ccaggaggcc gggaccggtg cgggcgcggg gtcccttgcg ggttcttgcg gctgcggcac     180  
       gccccagcgg cctggcgccc atggcagttc ggcagccgct caccgatact cgcgggagggc     240  
       taacgctccg ggccccgtac ccggagagcg gcaactcgcg cactcaaaga tgggtcccat     300  
 15     ccctgctgga gtatttacia tgggcacaga tgatcctcag ataaagcagg atgggggaagc     360  
       acctgcgagg agagttacta ttgatgccct ttacatggat gcctatgaag tcagtaatac     420  
       tgaatttgag aagtttgtga actcaactgg ctatttgaca gaggctgaga agtttggcga     480  
 20     ctcctttgtc tttgaaggca tgttgagtga gcaagtgaag accaatattc aacaggcagt     540  
       tgcagctgct ccctgggtgt tacctgtgaa aggcgctaac tggagacacc cagaagggcc     600  
       tgactctact attctgcaca ggccggatca tccagttctc catgtgtcct ggaatgatgc     660  
 25     ggttgcttac tgcacttggg cagggaaagcg gctgcccacg gaagctgagt gggaaatacag     720  
       ctgtcgagga ggctgcata atagactttt ccctgggggc aacaaactgc agcccaaagg     780  
 30     ccagcattat gccaacattt ggcagggcga ttttccggtg accaacactg gtgaggatgg     840  
       cttccaagga actgcgcctg ttgatgcctt cctcccaat ggttatggct tatacaacat     900  
       agtggggaac gcatgggaat ggacttcaga ctgggtggact gttcatcatt ctgttgaaga     960  
 35     aacgcttaac ccaaaaggtc ccccttctgg gaaagaccga gtgaagaaag gtggatccta     1020  
       catgtgccat aggtcttatt gttacagta tcgctgtgct gctcggagcc agaacacacc     1080  
       tgatagctct gcttcgaatc tgggatccg ctgtgcagcc gaccgcctgc ccaccatgga     1140  
 40     ctgacaacca agggtagtct tccccagtc aaggagcagt cgtgtctgac ctacattggg     1200  
       ctttcctcag aactttgaac gatcccatgc aaagaattcc caccctgagg tgggttacat     1260  
       acctgccaa tggccaaagg aaccgccttg tgagacaaa ttgctgacct gggtcagtgc     1320  
 45     atgtgcttta tgggtgtgtg catctttgga gatcatcacc atattttact tttgagagtc     1380  
       tttaaagagg aaggggagtg gagggaacct tgagctaggg ttcaggaggg ccgcatccta     1440  
 50     cgcaggctct gccacagggg ttagacccca ggtccgacgc ttgaccttc tgggcctcaa     1500  
       gtgccctccc ctatcaaatg aaggaatgga cagcatgacc tctgggtgtc tctccaactc     1560  
       accagttcta aaaagggtat cagattctat tgtgacttca tagaatttat gatagattat     1620  
 55     tttttagcta ttttttccat gtgtgaacct tgagtgatac taatcatgta aagtaagagt     1680  
       tctcttatgt attatgttcg gaagaggggt gtgggtgactc ctttatattc gtactgcact     1740  
       ttgtttttcc aaggaaatca gtgtctttta cgttgttatg atgaatcca catggggccg     1800  
 60     gtgatgggat gctgaagtte agccgttgaa cacataggaa tgtctgtggg gtgactctac     1860  
       tgtgctttat cttttaacat taagtgcctt tggttcagag gggcagtcac aagctctggt     1920  
       tccccctctc cccaaagcct tcagcgaacg tgaaatgtgc gctaaacggg gaaacctggt     1980  
 65     taattctaga tatagggaaa aaggaacgag gaccttgaat gagctatatt cagggtatcc     2040  
       ggtattttgt aatagggaaat aggaaacctt gttggctgtg gaatatccga tgctttgaat     2100  
       catgcactgt gttgaataaa cgtatctgct                                     2130

ES 2 555 056 T3

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

10 <400> 5

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

5 Gly Val Phe Thr Met Gly Thr Asp Asp Pro Gln Ile Lys Gln Asp Gly  
100 105 110

10 Glu Ala Pro Ala Arg Arg Val Thr Ile Asp Ala Leu Tyr Met Asp Ala  
115 120 125

15 Tyr Glu Val Ser Asn Thr Glu Phe Glu Lys Phe Val Asn Ser Thr Gly  
130 135 140

20 Tyr Leu Thr Glu Ala Glu Lys Phe Gly Asp Ser Phe Val Phe Glu Gly  
145 150 155 160

25 Met Leu Ser Glu Gln Val Lys Thr Asn Ile Gln Gln Ala Val Ala Ala  
165 170 175

30 Ala Pro Trp Trp Leu Pro Val Lys Gly Ala Asn Trp Arg His Pro Glu  
180 185 190

35 Gly Pro Asp Ser Thr Ile Leu His Arg Pro Asp His Pro Val Leu His  
195 200 205

40 Val Ser Trp Asn Asp Ala Val Ala Tyr Cys Thr Trp Ala Gly Lys Arg  
210 215 220

45 Leu Pro Thr Glu Ala Glu Trp Glu Tyr Ser Cys Arg Gly Gly Leu His  
225 230 235 240

50 Asn Arg Leu Phe Pro Trp Gly Asn Lys Leu Gln Pro Lys Gly Gln His  
245 250 255

55 Tyr Ala Asn Ile Trp Gln Gly Asp Phe Pro Val Thr Asn Thr Gly Glu  
260 265 270

60 Asp Gly Phe Gln Gly Thr Ala Pro Val Asp Ala Phe Pro Pro Asn Gly  
275 280 285

65 Tyr Gly Leu Tyr Asn Ile Val Gly Asn Ala Trp Glu Trp Thr Ser Asp  
290 295 300

70 Trp Trp Thr Val His His Ser Val Glu Glu Thr Leu Asn Pro Lys Gly  
305 310 315 320

75 Pro Pro Ser Gly Lys Asp Arg Val Lys Lys Gly Gly Ser Tyr Met Cys  
325 330 335

80 His Arg Ser Tyr Cys Tyr Arg Tyr Arg Cys Ala Ala Arg Ser Gln Asn  
340 345 350

85 Thr Pro Asp Ser Ser Ala Ser Asn Leu Gly Phe Arg Cys Ala Ala Asp  
355 360 365

90 Arg Leu Pro Thr Met Asp  
370

ES 2 555 056 T3

5

<210> 6  
 <211> 2297  
 <212> ADN  
 <213> Homo sapiens

10

<400> 6

```

cggtgtgtt ggcagtcct catgggttcc cgacgaggag gtctctgtgg ctgctggcggc 60
tgctaactgc gccacctgct gcagcctgtc cccgcctctc tgaagcggcc gcgtcgaagc 120
cgaaatgccg ccaccccgga ccggccgagg ccttctctgg ctgggtctgg ttctgagctc 180
cgtctgcgtc gccctcggat ccgaaacgca ggccaactcg accacagatg ctctgaacgt 240
tcttctcctc atcgtggatg acctgcgccc ctccctgggc tgttatgggg ataagctggg 300
gaggtcccca aatattgacc aactggcctc ccacagcctc ctcttccaga atgcctttgc 360
gcagcaagca gtgtgcgccc cgagccgctt ttcttctctc actggcagga gacctgacac 420
caccgcctct tacgacttca actcctactg gaggtgtcac gctggaaact tctccaccat 480
ccccagtac ttcaaggaga atggctatgt gaccatgtcg gtgggaaaag tctttcacc 540
tgggatctct totaaccata ccgatgattc tccgtatagc tggctctttc cacttatca 600
tccttctctc gagaagtatg aaaacactaa gacatgtcga gggccagatg gagaactcca 660
tgccaacctg ctttgcctct tggatgtgct ggatgttccc gagggcactt tgctgacaa 720
acagagcact gagcaagcca tacagttggt ggaaaagatg aaaacgtcag ccagtccttt 780
cttctggccc gttgggtatc ataagccaca catccccttc agatacccca aggaatttca 840
gaagttgtat cccttggaga acatcaccct ggccccgat cccgaggtcc ctgatggcct 900
accccctgtg gcctacaacc cctggatgga catcaggcaa cgggaagacg tccaagcctt 960
aaacatcagt gtgccgtatg gtccaattcc tgtggacttt cagcggaaa tccgccagag 1020
ctactttgcc tctgtgtcat atttggatac acaggtgggc cgcctcttga gtgctttgga 1080
cgatcttcag ctggccaaca gcaccatcat tgcatttacc toggatcatg ggtgggctct 1140
aggtgaacat ggagaatggg ccaatacag caattttgat gttgctacct atgttcccct 1200
gatattctat gttcctggaa ggacggcttc acttccggag gcaggcgaga agcttttccc 1260
ttacctcgac ccttttgatt ccgcctcaca gttgatggag ccaggcaggc aatccatgga 1320
cctgttggaa cttgtgtctc tttttcccac gctggctgga cttgcaggac tgcaggttcc 1380
acctcgtctc cccgttccct catttcacgt tgagctgtgc agagaaggca agaaccttct 1440
gaagcatttt cgattccgtg acttggaaaga ggatccgtac ctccctggta atcccctga 1500
actgattgcc tatagccagt atccccggcc ttcagacatc cctcagtgga attctgacaa 1560
gccgagttta aaagatataa agatcatggg ctattccata cgcaccatag actataggta 1620
tactgtgtgg gttggcttca atcctgatga atttctagct aacttttctg acatccatgc 1680
aggggaactg tattttgtgg attctgacct attgcaggat cacaatatgt ataatgatc 1740
ccaaggtgga gatcttttcc agttgttgat gccttgagtt ttgccaacca tggatggcaa 1800
atgtgatgtg ctcccttcca gctgggtgaga ggaggagtta gagctggctc ttttgtgatt 1860
accataata ttggaagcag cctgagggct agttaatcca aacatgcctc aacaatttgg 1920
cctgagaata tgtaacagcc aaaccttttc gtttagtctt tattaaaatt tataattggt 1980
aattggacca gttttttttt taatttccct ctttttaaaa cagttacggc ttatttactg 2040
aataaataca aagcaaaaa actcaagtta tgtcatacct ttggatacga agaccataca 2100
taataaccaa acataacatt atacacaaag aatactttca ttatttggg aatttagtgc 2160
atttcaaaaa gtaatcatat atcaaacatg gcaccacact aagttcctga ttattttggt 2220
tataatttaa taatatatct tatgagccct atabatcaaa aatattatgt taacatgtaa 2280
tccatgttcc ttttttcc 2297
    
```



ES 2 555 056 T3

5 <210> 7  
 <211> 550  
 <212> PRT  
 <213> Homo sapiens

10 <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

ES 2 555 056 T3

5 Pro Val Asp Phe Gln Arg Lys Ile Arg Gln Ser Tyr Phe Ala Ser Val  
290 295 300

10 Ser Tyr Leu Asp Thr Gln Val Gly Arg Leu Leu Ser Ala Leu Asp Asp  
305 310 315 320

15 Leu Gln Leu Ala Asn Ser Thr Ile Ile Ala Phe Thr Ser Asp His Gly  
325 330 335

20 Trp Ala Leu Gly Glu His Gly Glu Trp Ala Lys Tyr Ser Asn Phe Asp  
340 345 350

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

30 Ser Leu Pro Glu Ala Gly Glu Lys Leu Phe Pro Tyr Leu Asp Pro Phe  
370 375 380

35 Asp Ser Ala Ser Gln Leu Met Glu Pro Gly Arg Gln Ser Met Asp Leu  
385 390 395 400

40 Val Glu Leu Val Ser Leu Phe Pro Thr Leu Ala Gly Leu Ala Gly Leu  
405 410 415

45 Gln Val Pro Pro Arg Cys Pro Val Pro Ser Phe His Val Glu Leu Cys  
420 425 430

50 Arg Glu Gly Lys Asn Leu Leu Lys His Phe Arg Phe Arg Asp Leu Glu  
435 440 445

55 Glu Asp Pro Tyr Leu Pro Gly Asn Pro Arg Glu Leu Ile Ala Tyr Ser  
450 455 460

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

65 Ser Leu Lys Asp Ile Lys Ile Met Gly Tyr Ser Ile Arg Thr Ile Asp  
485 490 495

70 Tyr Arg Tyr Thr Val Trp Val Gly Phe Asn Pro Asp Glu Phe Leu Ala  
500 505 510

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

80 Pro Leu Gln Asp His Asn Met Tyr Asn Asp Ser Gln Gly Gly Asp Leu  
530 535 540

85 Phe Gln Leu Leu Met Pro  
545 550

ES 2 555 056 T3

5 <210> 8  
 <211> 2657  
 <212> ADN  
 <213> Homo sapiens

10 <400> 8

	gaattccggg ccatgagctg ccccgtagcc gcctgctgcg cgctgctgct agtccctggg	60
	ctctgccggg cgcgtccccc gaacgcactg ctgctcctcg cggatgacgg aggcctttgag	120
15	agtggcgcgt acaacaacag cgcctatgcc accccgcacc tggacgcctt ggcccggcgc	180
	agcctcctct ttcgcaatgc cttcacctcg gtcagcagct gctctcccag ccgcgccagc	240
	ctcctcactg gcctgcccga gcatcagaat gggatgtacg ggctgcacca ggacgtgca	300
20	cacttcaact ccttcgacaa ggtgctgagc ctgccgctgc tgctcagcca agctgggtgtg	360
	cgacaggca tcatcgggaa gaagcacgtg gggccggaga ccgtgtacco gtttgacttt	420
	gcgtacacgg aggagaatgg ctccgtcctc cagggtggggc ggaacatcac tagaattaag	480
	ctgctcgtcc ggaaattcct gcagactcag gatgacgggc ctttcttctc ctacgtgcc	540
25	ttccacgacc ccacccgtg tgggcactcc cagcccagc acggaacctt ctgtgagaag	600
	tttggcaacg gagagagcgg catgggtcgt atcccagact ggacccccca ggctacgac	660
	ccactggacg tgcctgggtcc ttacttcgtc cccaacaccc cggcagcccg agccgacctg	720
30	gcccgtcagt acaccaccgt cggccgcatg gaccaaggag ttggactggt gctccaggag	780
	ctgctgtagc ccggtgtcct gaacgacaca ctgggtgatc tcaagtcgca caacgggatc	840
	cccttcccca gcggcaggac caacctgtac tggccgggca ctgctgaacc cttactggtg	900
35	tcatccccgg agcacccaaa acgctggggc caagtcagcg aggcctacgt gacccctcta	960
	gacctcaagc ccaccatctt ggattggttc tcgatccogt accccagcta cggcatcttt	1020
	ggctegaaga ccatccacct cactggccgg tccctcctgc cggcgtgga ggccgagccc	1080
	ctctgggcca ccgtctttgg cagccagagc caccacgagg tcaccatgtc ctacccccatg	1140
40	cgctccgtgc agcacccgca cttccgcctc gtgcacaacc tcaacttcaa gatgcccttt	1200
	cccatcgacc aggacttcta cgtctcacc accctccagg acctcctgaa ccgcaccaca	1260
	gctggtcagc ccacgggctg gtacaaggac ctccgtcatt actactaccg ggccgctg	1320
45	gagctctacg accggagccg ggacccccac gagaccaga acctggcca cagaccgccc	1380
	tttgctcagc ttctggagat gcttcgggac cagctggcca agtggcagtg ggagaccac	1440
	gaccctggg tgtgcgcccc cgacggcgtc ctggaggaga agctctctcc ccagtccag	1500
50	cccctccaca atgagctgtg accatcccag gaggcctgtg cacacatccc aggcattgcc	1560
	cagacacatc coacacgtgt ccgtgtggcc ggccagcctg gggagtagtg gcaacagccc	1620
	ttccgtccac actcccatcc aaggagggtt cttccttctc gtggggtcac tcttgccatt	1680
	gcctggaggg ggaccagagc atgtgaccag agcatgtgcc cagcccctcc accaccaggg	1740
55	gcactgccgt catggcaggg gacacagttg tccttgtgtc tgaacctgtt ccagcacagg	1800
	gaattctaga catacgtggt ctgcccagac ggcagcggcc ccagcccatt acaagggagt	1860
	cttgttttct ggcttggttt ggggacctgc aaatgggagg cctgaggccc tcttcaggct	1920
60	ttggcagcca cagatacttc tgaaccttc acagagagca ggcaggggct tcgggtgccgc	1980
	gtgggcagta cgcaggctcc accgacactc acctgggagc acggcgcctg gctottacca	2040
	gcgtctggcc tagaggaagc ctttgagcga cctttgggca ggtttctgct tcttctgttt	2100
	tgcccatggt caagtcctctg ttccccaggc aggtttcagc tgattggcag caggctccct	2160
65	gagtgatgag cttgaacctg tgggtgtttc gggcagaagc ttatcttttt tgagagtgtc	2220
	cgaagatgaa ggcattggcga tgcctcctct ctggcttggg ttaattcttc ggtgacactg	2280
	gcattgctgg gtgggtgatgc ccgtcctctg gcttgggtta attctctggg gacactggcg	2340
	ttgctgggtg gcaatgcccg tctctgggtc tgggttaatt cttcgggtgac actggcgttg	2400

ES 2 555 056 T3

5 ctgggtggcg atgcccgtcc tctggcttgg gttaattctt ggatgacgtc ggcgttgctg 2460  
 ggagaatgtg cegttcctgc cctgcctcca cccacctcgg gagcagaagc cgggcctgga 2520  
 caccctcctgg cctggacacc cctcgaagga gagggcgctt ccttgagtag gtgggctccc 2580  
 10 cttgcccttc cctccctatc actccatact ggggtgggct ggaggaggcc acaggccagc 2640  
 tattgtaaaa gcttttt 2657

15 <210> 9  
 <211> 502  
 <212> PRT  
 <213> Homo sapiens

20 <400> 9

Met Ser Cys Pro Val Pro Ala Cys Cys Ala Leu Leu Leu Val Leu Gly  
 1 5 10 15  
 25 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 35 40 45  
 His Leu Asp Ala Leu Ala Arg Arg Ser Leu Leu Phe Arg Asn Ala Phe  
 50 55 60  
 35 Thr Ser Val Ser Ser Cys Ser Pro Ser Arg Ala Ser Leu Leu Thr Gly  
 65 70 75 80  
 40 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  
 45 Gln Ala Gly Val Arg Thr Gly Ile Ile Gly Lys Lys His Val Gly Pro  
 115 120 125  
 50 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  
 55 Lys Phe Leu Gln Thr Gln Asp Asp Arg Pro Phe Phe Leu Tyr Val Ala  
 165 170 175  
 60 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  
 65 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

ES 2 555 056 T3

5

Thr Thr Val Gly Arg Met Asp Gln Gly Val Gly Leu Val Leu Gln Glu  
245 250 255

10

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

15

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

20

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

25

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

30

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

35

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

40

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

45

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

50

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

55

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

60

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

65

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

ES 2 555 056 T3

5 <210> 10  
 <211> 1014  
 <212> ADN  
 <213> Homo sapiens

10 <400> 10

cgTgcctgta atcccagcag ctactcactc aggaggctga ggcaggagaa tctcttgaac 60  
 15 ccggaaggca gaggttgacag tgagccaaga tggcgccact gaactccagc ctgggtgaca 120

gagtgagact gtctcagaac agcaacaaca aatgccccgc tgctgctggg tccagaagag 180  
 20 cttgaataac tgcattgtct tttctcaat tttcatttcc cagaactggg cacctccggg 240  
 ctgtgaaaag ttagggaagt gtctgacacc tccagaatcc attcccaga agtgccctctg 300  
 gtcccactag cacctgcgca gactcaggcc aggccataga tctccagttg gccctgcaag 360  
 25 tgccctggagg aaggatggct ctggcctcgg tctccccca accctgcca agccagacag 420  
 acagcacctg cagacgcagg gggactgcac aattccacct gcccaggacc tgaccctggc 480  
 gtgtgcttgg cctcctcct cgcaccggc gcctcagatt tcaggacct cctcctcgc 540  
 30 caaggcgct cagacctcag gaccctgccc tctcacgct ttgtgaacc caaatatctg 600  
 agaccagtct cagtttatt tgccaaggct aaggatgcac ctgtgacagc ctcaggaggt 660  
 35 cctgacaaca ggtgcccag gtggctgggg atacagtttg cctttataca tcttagggag 720  
 acacaagatc agtatgtgta tggcgtacat tggttcagtc agccttcac tgaatacacg 780  
 attgagtctg gcccagtga tccgcatttt tatgtaaaca gtaaggaac ggggcaatca 840  
 40 tataagcgtt tgtctcaggg gagccccaga gggatgaatt ccagttccgt ctgtcctttg 900  
 tccacaagga atttccctgg gcgctaatta tgagggaggc gtgtagcttc ttatcattgt 960  
 45 agctatgtta tttagaaata aaacgggagg caggtttgcc taattcccag gttg 1014

50 <210> 11  
 <211> 522  
 <212> PRT  
 <213> Homo sapiens

<400> 11

55

60

65

5 Met Ala Ala Val Val Ala Ala Thr Arg Trp Trp Gln Leu Leu Leu Val  
1 5 10

10 Leu Ser Ala Ala Gly Met Gly Ala Ser Gly Ala Pro Gln Pro Pro Asn  
20 25 30

15 Ile Leu Leu Leu Leu Met Asp Asp Met Gly Trp Gly Asp Leu Gly Val  
35 40 45

20 Tyr Gly Glu Pro Ser Arg Glu Thr Pro Asn Leu Asp Arg Met Ala Ala  
50 55 60

25 Glu Gly Leu Leu Phe Pro Asn Phe Tyr Ser Ala Asn Pro Leu Cys Ser  
65 70 75 80

30 Pro Ser Arg Ala Ala Leu Leu Thr Gly Arg Leu Pro Ile Arg Asn Gly  
85 90 95

35 Phe Tyr Thr Thr Asn Ala His Ala Arg Asn Ala Tyr Thr Pro Gln Glu  
100 105 110

40 Ile Val Gly Gly Ile Pro Asp Ser Glu Gln Leu Leu Pro Glu Leu Leu  
115 120 125

45 Lys Lys Ala Gly Tyr Val Ser Lys Ile Val Gly Lys Trp His Leu Gly  
130 135 140

50 His Arg Pro Gln Phe His Pro Leu Lys His Gly Phe Asp Glu Trp Phe  
145 150 155 160

45

50

55

60

65

ES 2 555 056 T3

5 Gly Ser Pro Asn Cys His Phe Gly Pro Tyr Asp Asn Lys Ala Arg Pro  
165 170 175

10 Asn Ile Pro Val Tyr Arg Asp Trp Glu Met Val Gly Arg Tyr Tyr Glu  
180 185 190

15 Glu Phe Pro Ile Asn Leu Lys Thr Gly Glu Ala Asn Leu Thr Gln Ile  
195 200 205

20 Tyr Leu Gln Glu Ala Leu Asp Phe Ile Lys Arg Gln Ala Arg His His  
210 215 220

25 Pro Phe Phe Leu Tyr Trp Ala Val Asp Ala Thr His Ala Pro Val Tyr  
225 230 235 240

30 Ala Ser Lys Pro Phe Leu Gly Thr Ser Gln Arg Gly Arg Tyr Gly Asp  
245 250 255

35 Ala Val Arg Glu Ile Asp Asp Ser Ile Gly Lys Ile Leu Glu Leu Leu  
260 265 270

40 Gln Asp Leu His Val Ala Asp Asn Thr Phe Val Phe Phe Thr Ser Asp  
275 280 285

45 Asn Gly Ala Ala Leu Ile Ser Ala Pro Glu Gln Gly Gly Ser Asn Gly  
290 295 300

50 Pro Phe Leu Cys Gly Lys Gln Thr Thr Phe Glu Gly Gly Met Arg Glu  
305 310 315 320

55 Pro Ala Leu Ala Trp Trp Pro Gly His Val Thr Ala Gly Gln Val Ser  
325 330 335

60 His Gln Leu Gly Ser Ile Met Asp Leu Phe Thr Thr Ser Leu Ala Leu  
340 345 350

65 Ala Gly Leu Thr Pro Pro Ser Asp Arg Ala Ile Asp Gly Leu Asn Leu  
355 360 365

70 Leu Pro Thr Leu Leu Gln Gly Arg Leu Met Asp Arg Pro Ile Phe Tyr  
370 375 380

75 Tyr Arg Gly Asp Thr Leu Met Ala Ala Thr Leu Gly Gln His Lys Ala  
385 390 395 400

80 His Phe Trp Thr Trp Thr Asn Ser Trp Glu Asn Phe Arg Gln Gly Ile  
405 410 415

85 Asp Phe Cys Pro Gly Gln Asn Val Ser Gly Val Thr Thr His Asn Leu  
420 425 430

90 Glu Asp His Thr Lys Leu Pro Leu Ile Phe His Leu Gly Arg Asp Pro  
435 440 445

95 Gly Glu Arg Phe Pro Leu Ser Phe Ala Ser Ala Glu Tyr Gln Glu Ala  
450 455 460

100 Leu Ser Arg Ile Thr Ser Val Val Gln Gln His Gln Glu Ala Leu Val  
465 470 475 480



ES 2 555 056 T3

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

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

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

<210> 12  
 <211> 2379  
 <212> ADN  
 <213> Homo sapiens

<400> 12

25 ggaattccgg tgggcctctc gcccttcagc tacctgtgcg tccctccgtc ccgtcccgtc 60  
 ccgggggtcac cccggagcct gtcgctatg cggctcctgc ctctagcccc aggtcggctc 120  
 cggcggggca gcccccgcca cctgccctcc tgcagcccag cgctgctact gctggtgctg 180  
 ggcggctgcc tgggggtctt cggggtggct gcgggaaccc ggaggcccaa cgtggtgctg 240  
 30 ctctctcagc agcaccagga cgaagtgtc ggcgcatga caccactaaa gaaaaccaa 300  
 gctctcatcg gagagatggg gatgactttt tccagtgtt atgtgccaag tgctctctgc 360  
 tgccccagca gagccagtat cctgacagga aagtaccac ataatcatca cgttgtgaac 420  
 35 aacactctgg aggggaactg cagtagtaag tcctggcaga agatccaaga accaaatct 480  
 ttcccagcaa ttctcagatc aatgtgtggt tatcagacct tttttgcagg gaaatattta 540  
 aatgagtacg gagccccaga tgcaggtgga ctagaacacg ttctctggg ttggagttac 600  
 40 tggtatgcct tggaaaagaa ttctaagtat tataattaca ccctgtctat caatgggaag 660  
 gcacggaagc atggtgaaaa ctatagtgtg gactacctga cagatgtttt ggctaattgtc 720  
 tccttggact ttctggacta caagtccaac tttgagccct tcttcatgat gatcgccact 780  
 45 ccagcgccctc attgccttg gacagctgca cctcagtacc agaaggcttt ccagaatgtc 840  
 tttgcaccaa gaaacaagaa cttcaacatc catggaacga acaagcactg gttaattagg 900  
 caagccaaga ctccaatgac taattcttca atacagtttt tagataatgc atttaggaaa 960  
 50 aggtggcaaa ctctcctctc agttgatgac cttgtggaga aactggtcaa gaggtggag 1020  
 ttcactgggg agctcaacaa cacttacatc ttctatacct cagacaatgg ctatcacaca 1080  
 ggacagtttt ccttgccaat agacaagaga cagctgtatg agtttgatat caaagtcca 1140  
 55 ctgttggttc gaggacctgg gatcaaacca aatcagacaa gcaagatgct ggttgccaac 1200  
 attgacttgg gtctactat tttggacatt gctggtcagc acctaaataa gacacagatg 1260  
 gatgggatgt ccttattgcc cattttgaga ggtgccagta acttgacctg gcgatcagat 1320  
 60 gtctggttgg aataccaagg agaaggcctg aacgtcactg acccaacatg cccttccctg 1380  
 agtctctggc tatctcaatg cttcccagac tgtgtatgtg aagatgctta taacaatacc 1440  
 tatgcctgtg tgaggacaat gtcagcattg tggaatttgc agtattgca gtttgatgac 1500  
 65 caggaggtgt ttgtagaagt ctataatctg actgcagacc cagaccagat cactaacatt 1560  
 gctaaaacca tagaccaga gcttttagga aagatgaact atcggttaat gatgttacag 1620  
 tctgttctg ggccaacctg tcgcactcca ggggtttttg accccggata caggtttgac 1680  
 cccgctctca tgttcagcaa tcgcggcagt gtcaggactc gaagatttcc caaacatctt 1740

ES 2 555 056 T3

5 ctgtagcgac ctcacacagc ctctgcagat ggatccctgc acgcctcttt ctgatgaagt 1800  
 gattgtagta ggtgtctgta gctagtcttc aagaccacac ctggaagagt ttctgggctg 1860  
 10 gctttaagtc ctgtttgaaa aagcaacca gtcagctgac ttctctgtgc aatgtgtaa 1920  
 actgtgaact ctgcccattg gtcaggagt gctgtctctg gtctcttctt ttagctgaca 1980  
 aggacactcc tgaggtcttt gttctcactg tatttttttt atcctggggc cacagttctt 2040  
 15 gattattcct cttgtgggta aagactgaat ttgtaaacce attcagataa atggcagtac 2100  
 tttaggacac acacaaacac acagatacac cttttgatat gtaagcttga cctaaagtca 2160  
 aaggacctgt gtagcatttc agattgagca ctccactatc aaaaatacta acatcacatg 2220  
 20 gcttgaagag taaccatcag agctgaatca tccaagtaag aacaagtacc attgttgatt 2280  
 gataagtaga gatacatctt ttatgatggt catcacagtg tggtaagggt gcaaattcaa 2340  
 25 aacatgtcac ccaagctctg ttcattgttt tgtgaattc 2379

<210> 13  
 <211> 552  
 <212> PRT  
 30 <213> Homo sapiens  
 <400> 13

35

40

45

50

55

60

65

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

10 Arg His Leu Pro Ser Cys Ser Pro Ala Leu Leu Leu Leu Val Leu Gly  
20 25 30

15 Gly Cys Leu Gly Val Phe Gly Val Ala Ala Gly Thr Arg Arg Pro Asn  
35 40 45

20 Val Val Leu Leu Leu Thr Asp Asp Gln Asp Glu Val Leu Gly Gly Met  
50 55 60

25 Thr Pro Leu Lys Lys Thr Lys Ala Leu Ile Gly Glu Met Gly Met Thr  
65 70 75 80

30 Phe Ser Ser Ala Tyr Val Pro Ser Ala Leu Cys Cys Pro Ser Arg Ala  
85 90 95

35 Ser Ile Leu Thr Gly Lys Tyr Pro His Asn His His Val Val Asn Asn  
100 105 110

40 Thr Leu Glu Gly Asn Cys Ser Ser Lys Ser Trp Gln Lys Ile Gln Glu  
115 120 125

45 Pro Asn Thr Phe Pro Ala Ile Leu Arg Ser Met Cys Gly Tyr Gln Thr  
130 135 140

50 Phe Phe Ala Gly Lys Tyr Leu Asn Glu Tyr Gly Ala Pro Asp Ala Gly  
145 150 155 160

55 Gly Leu Glu His Val Pro Leu Gly Trp Ser Tyr Trp Tyr Ala Leu Glu  
165 170 175

60 Lys Asn Ser Lys Tyr Tyr Asn Tyr Thr Leu Ser Ile Asn Gly Lys Ala  
180 185 190

55

60

65

ES 2 555 056 T3

5 Arg Lys His Gly Glu Asn Tyr Ser Val Asp Tyr Leu Thr Asp Val Leu  
195 200 205

10 Ala Asn Val Ser Leu Asp Phe Leu Asp Tyr Lys Ser Asn Phe Glu Pro  
210 215 220

Phe Phe Met Met Ile Ala Thr Pro Ala Pro His Ser Pro Trp Thr Ala  
225 230 235 240

15 Ala Pro Gln Tyr Gln Lys Ala Phe Gln Asn Val Phe Ala Pro Arg Asn  
245 250 255

20 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

25 Phe Arg Lys Arg Trp Gln Thr Leu Leu Ser Val Asp Asp Leu Val Glu  
290 295 300

30 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

35 Pro Ile Asp Lys Arg Gln Leu Tyr Glu Phe Asp Ile Lys Val Pro Leu  
340 345 350

40 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

45 Asp Leu Asn Lys Thr Gln Met Asp Gly Met Ser Leu Leu Pro Ile Leu  
385 390 395 400

50 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

55 Pro Gly Val Ser Gln Cys Phe Pro Asp Cys Val Cys Glu Asp Ala Tyr  
435 440 445

60 Asn Asn Thr Tyr Ala Cys Val Arg Thr Met Ser Ala Leu Trp Asn Leu  
450 455 460

Gln Tyr Cys Glu Phe Asp Asp Gln Glu Val Phe Val Glu Val Tyr Asn  
465 470 475 480

65 Leu Thr Ala Asp Pro Asp Gln Ile Thr Asn Ile Ala Lys Thr Ile Asp  
485 490 495

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

ES 2 555 056 T3

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

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

15 Arg Arg Phe Ser Lys His Leu Leu  
 545 550

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

<400> 14

25 ccggtaccgg ctccctcctgg gctccctcta gcgccttccc cccggcccga ctgcctggtc 60  
 agcgccaagt gacttacgcc cccgaccctg agcccggacc gctaggcgag gaggatcaga 120  
 tctccgctcg agaatctgaa ggtgccctgg tcctggagga gttccgtccc agccctgcgg 180  
 tctcccggta ctgctcgccc cggccctctg gagcttcagg aggcggccgt cagggtcggg 240  
 30 gagtatttgg gtccggggtc tcagggaagg gcggcgccctg ggtctgcggg atcggaaaga 300  
 gcctgctgga gccaaagtagc cctccctctc ttgggacaga ccctcgggtc ccatgtccat 360  
 gggggcaccg cggtcctctc tcttgccctt ggtgctggtc ctggccgttg cccgtccgcc 420  
 35 caacatcgtg ctgatctttg ccgacgacct cggctatggg gacctgggct gctatgggca 480  
 cccagctct accactccca acctggacca gctggcggcg ggagggtgc ggttcacaga 540  
 cttctacgtg cctgtgtctc tgtgcacacc ctctagggcc gcctcctga ccggccggct 600  
 40 cccggttcgg atgggcatgt acctggcgt cctgggtgcc agctcccggg ggggcctgcc 660  
 cctggaggag gtgaccgtgg ccgaagtctt ggtgcgccga ggctacctca caggaatggc 720  
 cggcaagtgg caccttgggg tggggcctga gggggccttc ctgcccccc atcagggtct 780  
 45 ccatcgattt ctaggcatcc cgtactccca cgaccagggc cctgccaga acctgacctg 840  
 cttcccgcgg gccactcctt gcgacgggtg ctgtgaccag ggctgggtcc ccatcccact 900  
 gttggccaac ctgtccgtgg aggcgcagcc ccctggctg cccggactag aggcccgcta 960  
 50 catggctttc gcccatgacc tcatggccga cgcaccagcc caggatcgcc ccttcttctt 1020  
 gtactatgcc tctcaccaca cccactacce tcagttcagt gggcagagct ttgcagagcg 1080  
 ttcaggccgc gggccatttg gggactccct gatggagctg gatgcagctg tggggacctt 1140  
 55 gatgacagcc ataggggacc tggggctgct tgaagagacg ctggtcactt tcaactgcaga 1200  
 caatggacct gagaccatgc gtatgtcccg aggcggctgc tccggctctt tgcgggtgtg 1260  
 aaagggaaac acctacgagg gcggtgtccg agagcctgcc ttggccttct gcccaggtea 1320  
 60 tatcgctccc ggcgtgacct acgagctggc cagctccctg gacctgctgc ctaccctggc 1380  
 agccctggct ggggccccac tgcccaatgt caccttggat ggctttgacc tcagccccct 1440  
 gctgctgggc adaggycaaga gccctcggca gtctctcttc ttctaccctt cctaccacaga 1500  
 65 cgaggtccgt ggggtttttg ctgtgcggac tggaaagtac aaggctcaact tcttcaccca 1560  
 gggctctgcc cacagtgata ccaactgcaga ccctgcctgc caagcctcca gctctctgac 1620  
 tgctcatgag cccccgctgc tctatgacct gtccaaggac cctgggtgaga actacaacct 1680  
 gctggggggg gtggccgggg ccaccccaga ggtgctgcaa gcctgaaac agcttcagct 1740

ES 2 555 056 T3

5 gctcaaggcc cagttagacg cagctgtgac cttcggcccc agccagggtgg cccgggggcca 1800  
 ggacccccgcc ctgcagatct gctgtcatcc tggctgcacc ccccgcccag cttgctgcca 1860  
 10 ttgccagat ccccatgcct gagggcccct cggctggcct gggcatgtga tggctcctca 1920  
 ctgggagcct gtgggggagg ctcagggtgc tggaggggggt ttgtgcctga taacgtaata 1980  
 acaccagtgg agacttgcac atctgaaaaa aaaaaaaaaa aa 2022

15 <210> 15  
 <211> 507  
 <212> PRT  
 <213> Homo sapiens  
 20 <400> 15

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

5 Arg Ser Gly Arg Gly Pro Phe Gly Asp Ser Leu Met Glu Leu Asp Ala  
245 250 255

10 Ala Val Gly Thr Leu Met Thr Ala Ile Gly Asp Leu Gly Leu Leu Glu  
260 265 270

15 Glu Thr Leu Val Ile Phe Thr Ala Asp Asn Gly Pro Glu Thr Met Arg  
275 280 285

20 Met Ser Arg Gly Gly Cys Ser Gly Leu Leu Arg Cys Gly Lys Gly Thr  
290 295 300

25 Thr Tyr Glu Gly Gly Val Arg Glu Pro Ala Leu Ala Phe Trp Pro Gly  
305 310 315 320

30 His Ile Ala Pro Gly Val Thr His Glu Leu Ala Ser Ser Leu Asp Leu  
325 330 335

35 Leu Pro Thr Leu Ala Ala Leu Ala Gly Ala Pro Leu Pro Asn Val Thr  
340 345 350

40 Leu Asp Gly Phe Asp Leu Ser Pro Leu Leu Leu Gly Thr Gly Lys Ser  
355 360 365

45 Pro Arg Gln Ser Leu Phe Phe Tyr Pro Ser Tyr Pro Asp Glu Val Arg  
370 375 380

50 Gly Val Phe Ala Val Arg Thr Gly Lys Tyr Lys Ala His Phe Phe Thr  
385 390 395 400

55 Gln Gly Ser Ala His Ser Asp Thr Thr Ala Asp Pro Ala Cys His Ala  
405 410 415

60 Ser Ser Ser Leu Thr Ala His Glu Pro Pro Leu Leu Tyr Asp Leu Ser  
420 425 430

65 Lys Asp Pro Gly Glu Asn Tyr Asn Leu Leu Gly Gly Val Ala Gly Ala  
435 440 445

70 Thr Pro Glu Val Leu Gln Ala Leu Lys Gln Leu Gln Leu Leu Lys Ala  
450 455 460

75 Gln Leu Asp Ala Ala Val Thr Phe Gly Pro Ser Gln Val Ala Arg Gly  
465 470 475 480

80 Glu Asp Pro Ala Leu Gln Ile Cys Cys His Pro Gly Cys Thr Pro Arg  
485 490 495

85 Pro Ala Cys Cys His Cys Pro Asp Pro His Ala  
500 505

ES 2 555 056 T3

5 <210> 16  
 <211> 2228  
 <212> ADN  
 <213> Homo sapiens

10 <400> 16

	acaaggatgg gtcogcggg cgcggcgagc ttgccccgag gccccggacc tggcgggctg	60
	ctcctccccg tcgtcctccc gctgctgctg ctgctgttgt tggcgcggcc gggctcgggc	120
15	gcccggggcca gccggccgcc ccacctggtc ttcttgctgg cagacgacct aggctggaac	180
	gacgtcggct tccacggctc cgcacccgc acgcgcacc tggacgcgct ggcggccggc	240
20	ggggtgctcc tggacaacta ctacacgcag cgcctgtgca cgcctgcgcg gagccagctg	300
	ctcaactggcc gctaccagat ccgtacaggc ttacagcacc aaataatctg gccctgtcag	360
	cccagctgtg ttccctctgga tgaaaaactc ctgccccagc tcctaaaaga agcaggttat	420
	actaccata tggtcggaaa atggcacctg ggaatgtacc ggaaagaatg ccttccaacc	480
25	cgccgaggat ttgataccta ctttgatgat ctctgggta gtgaagatta ttattcccat	540
	gaacgctgta caltaattga cgtctgtaat gtcacacgat gtgctcttga ttttcgagat	600
	ggcgaagaag ttgcaacagg atataaaaat atgtattcaa caaacatatt caccaaaagg	660
30	gctatagccc tcataactaa ccattccacca gagaagcctc tgtttctcta ccttgctctc	720
	cagtctgtgc atgagcccct tcaggctcct gaggaatact tgaagccata tgactttatc	780
	caagacaaga acaggcatca ctatgcagga atgggtgccc ttatggatga agcagtagga	840
35	aatgtcactg cagctttaa aagcagtggg ctctggaaca acacggtgtt catctttctt	900
	acagataacg gagggcagac tttggcaggg ggaataaact ggccccctcg aggaagaaaa	960
	tggagcctgt ggggaaggagg cgtccgaggg gtgggctttg tggcaagccc cttgctgaag	1020
40	cagaagggcg tgaagaaccg ggagctcatc cacatctctg actggctgcc aacactcgtg	1080
	aagctggcca ggggacacac caatggcaca aagcctctgg atggcttoga cgtgtggaaa	1140
	accatcagtg aaggaagccc atccccaga attgagctgc tgcataatat tgaccctaac	1200
	ttcgtggact ctaccctgtg tcccaggaac agcatggctc cagcaaagga tgactcttct	1260
45	cttcagaat attcagcctt taacacatct gtccatgctg caattagaca tggaaattgg	1320
	aaactcctca cgggctaccc aggctgtggc tactggttcc ctccaccgtc tcaatacaat	1380
	gtttctgaga taccctcatc agaccacca accaagacc cctggctctt tgatattgat	1440
50	cgggaccctg aagaaagaca tgacctgtcc agagaatctc ctacatcgt cacaagctc	1500
	ctgtcccgc tacagttcta ccataaacac tcagtccccg tgtacttccc tgcacaggac	1560
	ccccgctgtg atcccaggc cactgggggtg tggggccctt ggatgtagga tttcagggag	1620
55	gctagaaaa ctttcaattg gaagtgggac ctcaggcctt ttctcagac tcttgtctca	1680
	tttgttatcc caacctgggt tcaactggcc ctctcttctg tcttaacca caccgaggtg	1740
	tctaatttca acccctaattg catttaagaa gctgataaaa tctgcaacac tcctgctgtt	1800
60	ggctggagca tgtgtctaga ggtgggggtg gctgggttta tcccccttcc ctaagccttg	1860
	ggacagctgg gaacttaact tgaaatagga agttctcact gaatcctgga ggctggaaca	1920
	gctggctctt ttagactcac aagtcagacg ttccgattccc ctctgccaat agccagtttt	1980
	attggagtga atcacattc ttacgcaaat gaagggagca gacagtgatt aatggttctg	2040
65	ttggccaagg cttctcccctg tcgggtgaagg atcatgttca ggcactccaa gtgaaccacc	2100
	cctcttgggt cacccttac tcaactatct catcacagag cataaggccc attttgttgt	2160
	tcaggccaac agcaaatgg cctgcacat gactgtggct tttaaaataa agaaatgtgt	2220
	ttttatcg	2228



ES 2 555 056 T3

5 <210> 17  
 <211> 533  
 <212> PRT  
 <213> Homo sapiens

10 <400> 17

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

ES 2 555 056 T3

5 <210> 18  
<211> 2401  
<212> ADN  
<213> Homo sapiens

10 <400> 18

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

25

30

35

40

45

50

55

60

65

ES 2 555 056 T3

5 ccaatatcga ceggttggcc agtgggggag tgaactcac tcagcacctg gcagcatcac 420  
 cgctgtgcac accaagcagg gcagccttca tgactggccg gtaccctgtc cgatcaggaa 480  
 10 tggcatcttg gtcccgcact ggagttttcc tcttcacagc ctcttcggga ggacttccca 540  
 cggatgagat tacctttgct aagcttctga aggatcaagg ttattcaaca gcaactgatag 600  
 ggaaatggca ccttgggatg agctgtcaca gcaagactga cttctgtcac caccctttac 660  
 15 atcacggctt caattatttc tatgggatct ctttgaccaa tctgagagac tgcaagcccg 720  
 gagagggcag tgtcttcacc acgggcttca agaggctggg ctctctcccc ctgcagatcg 780  
 tcggggtcac cctccttacc cttgctgcac tcaattgtct ggggctactc cacgtgcctc 840  
 20 taggcgtttt tttcagcctt ctcttctag cagccctaat cctgaccctt ttcttgggct 900  
 tccttcatta cttccggccc ctgaactgct tcatgatgag gaactacgag atcattcagc 960  
 25 agcccatgtc ctatgacaat ctcaccaga ggctaaccggg ggaggcggcc cagttcatac 1020  
 agcggaaacac tgagactccg ttctgtcttg tcttgtccta cctccacgtg cacacagccc 1080  
 tgtttotccag caaagacttt gctggcaaaa gtcaaacagg agtctacggg gatgctgttg 1140  
 30 aggaaatgga ctggagtgtg gggcagatct tgaaccttct ggatgagctg agattggcta 1200  
 atgataccct catctacttc acatcggacc agggagcaca tgtagaggag gtgtcttcca 1260  
 aaggagaaat tcatggcggg agtaatggga tctataaagg aggaaaagca aacaactggg 1320  
 35 aaggaggtat ccgggttcca ggcaccttc gttggcccag ggtgatacag gctggccaga 1380  
 agattgatga gccactagc aacatggaca tatttcctac agtagccaag ctggctggag 1440  
 40 ctcccttgcc tgaggacagg atcattgatg gacgtgatct gatgccctg cttgaaggaa 1500  
 aaagccaacg ctccgatcat gagtttctct tccattactg caacgcctac ttaaattgctg 1560  
 tgcgctggca ccctcagaac agcacatcca tctggaagge ctttttcttc acccccact 1620  
 45 tcaaccccggt gggttccaac ggatgctttg ccacacacgt gtgcttctgt ttccgggagt 1680  
 atgtcaccca tcacgacca cctttactct ttgatatttc caaagatccc agagagagaa 1740  
 acccactaac tccagcatcc gagccccggg tttatgaaat cctcaaagtc atgcaggaag 1800  
 50 ctgcccagacag acacaccag accctgccag aggtgcccga tcagttttca tggacaact 1860  
 ttctttggaa gccctggctt cagctgtgct gtccttccac cggcctgtct tgccagtgtg 1920  
 55 atagagaaaa acaggataag agactgagcc gctagcagcg cctggggacc agacagacgc 1980  
 atgtggcaaa gctcaccatc ttactataa acacgcctga gactggcact ggggaaacat 2040  
 aactccatct acaccttga tttggactga ttctcattt taccacctga aggcttgggc 2100  
 60 cagagctcaa cagctactca actggagggg tgagggggat aaggtctgta gtatacagac 2160  
 aggaagatgg taggtttatg ccttctgtgg ccagagtctt ggactcatgg aaatagaatg 2220  
 65 aatagagggg cattcacaag gcacaccagt gcaagcagat gacaaaagg tgcagaaggc 2280  
 aatcttaaaa cagaaagggtg caggaggtac cttaactcac ccctcagcaa atacctatgt 2340  
 caacagtata agttaccatt tactctataa tctgcagtga tgcaataacc agcataataa 2400  
 a 2401

ES 2 555 056 T3

5

<210> 19  
 <211> 583  
 <212> PRT  
 <213> Homo sapiens

10

<400> 19

Met Pro Leu Arg Lys Met Lys Ile Pro Phe Leu Leu Leu Phe Phe Leu  
 1 5 10 15

Trp Glu Ala Glu Ser His Ala Ala Ser Arg Pro Asn Ile Ile Leu Val  
 20 25 30

Met Ala Asp Asp Leu Gly Ile Gly Asp Pro Gly Cys Tyr Gly Asn Lys  
 35 40 45

Thr Ile Arg Thr Pro Asn Ile Asp Arg Leu Ala Ser Gly Gly Val Lys  
 50 55 60

Leu Thr Gln His Leu Ala Ala Ser Pro Leu Cys Thr Pro Ser Arg Ala  
 65 70 75 80

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

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

Thr Asp Glu Ile Thr Phe Ala Lys Leu Leu Lys Asp Gln Gly Tyr Ser  
 115 120 125

Thr Ala Leu Ile Gly Lys Trp His Leu Gly Met Ser Cys His Ser Lys  
 130 135 140

Thr Asp Phe Cys His His Pro Leu His His Gly Phe Asn Tyr Phe Tyr  
 145 150 155 160

Gly Ile Ser Leu Thr Asn Leu Arg Asp Cys Lys Pro Gly Glu Gly Ser  
 165 170 175

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

Val Gly Val Thr Leu Leu Thr Leu Ala Ala Leu Asn Cys Leu Gly Leu  
 195 200 205

Leu His Val Pro Leu Gly Val Phe Phe Ser Leu Leu Phe Leu Ala Ala  
 210 215 220

Leu Ile Leu Thr Leu Phe Leu Gly Phe Leu His Tyr Phe Arg Pro Leu  
 225 230 235 240

Asn Cys Phe Met Met Arg Asn Tyr Glu Ile Ile Gln Gln Pro Met Ser  
 245 250 255

Tyr Asp Asn Leu Thr Gln Arg Leu Thr Val Glu Ala Ala Gln Phe Ile  
 260 265 270

Gln Arg Asn Thr Glu Thr Pro Phe Leu Leu Val Leu Ser Tyr Leu His  
 275 280 285

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

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

5 Gln Ile Leu Asn Leu Leu Asp Glu Leu Arg Leu Ala Asn Asp Thr Leu  
325 330 335

10 Ile Tyr Phe Thr Ser Asp Gln Gly Ala His Val Glu Glu Val Ser Ser  
340 345 350

15 Lys Gly Glu Ile His Gly Gly Ser Asn Gly Ile Tyr Lys Gly Gly Lys  
355 360 365

20 Ala Asn Asn Trp Glu Gly Gly Ile Arg Val Pro Gly Ile Leu Arg Trp  
370 375 380

25 Pro Arg Val Ile Gln Ala Gly Gln Lys Ile Asp Glu Pro Thr Ser Asn  
385 390 395 400

30 Met Asp Ile Phe Pro Thr Val Ala Lys Leu Ala Gly Ala Pro Leu Pro  
405 410 415

35 Glu Asp Arg Ile Ile Asp Gly Arg Asp Leu Met Pro Leu Leu Glu Gly  
420 425 430

40 Lys Ser Gln Arg Ser Asp His Glu Phe Leu Phe His Tyr Cys Asn Ala  
435 440 445

45 Tyr Leu Asn Ala Val Arg Trp His Pro Gln Asn Ser Thr Ser Ile Trp  
450 455 460

50 Lys Ala Phe Phe Phe Thr Pro Asn Phe Asn Pro Val Gly Ser Asn Gly  
465 470 475 480

55 Cys Phe Ala Thr His Val Cys Phe Cys Phe Gly Ser Tyr Val Thr His  
485 490 495

60 His Asp Pro Pro Leu Leu Phe Asp Ile Ser Lys Asp Pro Arg Glu Arg  
500 505 510

65 Asn Pro Leu Thr Pro Ala Ser Glu Pro Arg Phe Tyr Glu Ile Leu Lys  
515 520 525

70 Val Met Gln Glu Ala Ala Asp Arg His Thr Gln Thr Leu Pro Glu Val  
530 535 540

75 Pro Asp Gln Phe Ser Trp Asn Asn Phe Leu Trp Lys Pro Trp Leu Gln  
545 550 555 560

80 Leu Cys Cys Pro Ser Thr Gly Leu Ser Cys Gln Cys Asp Arg Glu Lys  
565 570 575

85 Gln Asp Lys Arg Leu Ser Arg  
580

ES 2 555 056 T3

5

<210> 20  
 <211> 1945  
 <212> ADN  
 <213> Homo sapiens

10

<400> 20

	ggaagccttg gcactagcgg cgcccgggcg cggagtgcgc agggcaaggc cctgcgctct	60
15	gggccagcgc tcggccatgc gatccgccgc gcggagggga cgcgccgcgc ccgccgccag	120
	ggactctttg ccggtgctac tgtttttatg cttgcttctg aagacgtgtg aacctaaaac	180
	tgcaaatgcc tttaaaccaa atactact gatcatggcg gatgatctag gcactgggga	240
20	tctcggttgc tacgggaaca atactactgag aacgccgaat attgaccagc ttgcagagga	300
	aggtgtgagg ctactcagc acctggcggc cgccccgctc tgcaccccaa gccgagctgc	360
	attcctcaca gggagacatt ccttcagatc aggcattggac gccagcaatg gataccgggc	420
25	ccttcagtgg aacgcaggct caggtggact ccctgagaac gaaaccactt ttgcaagaat	480
	cttgcagcag catggctatg caaccggcct cataggaaaa tggcaccagg gtgtgaattg	540
	tgcattcccgc ggggatcact gccaccacc cctgaaccac ggatttgact atttctacgg	600
30	catgcccttc acgctcaca acgactgtga cccaggcagg cccccgaag tggacgccgc	660
	cctgagggcg cagctctggg gttacacca gttcctggcg ctggggatc tcacctggc	720
	tgccggccag acctgcgggt tcttctctgt ctccgcgaga gcagtcaccg gcatggccgg	780
35	cgtgggctgc ctgtttttca tctcttggtg ctctctcttc gggtttgtgc gacgctggaa	840
	ctgtatcctg atgagaaacc atgacgtcac ggagcaacc atggttctgg agaaaacagc	900
40	gagtcttatg ctaaaggaag ctgttttcta tattgaaaga cacaagcatg ggccatttct	960
	cctcttctt tctttgctgc atgtgcacat tcccctgtg accacgagtg cattcctggg	1020
	gaaaagtcag catggcttat atggtgataa tgtggaggag atggactggc tcataggtaa	1080
45	ggttcttaat gccatcgaag acaatggttt aaagaactca acattcacgt atttcacctc	1140
	tgaccatgga ggacatttag aggcaagaga tggacacagc cagttagggg gatggaacgg	1200
	aatttcaaaa ggtgggaagg gcatgggagg atgggaaggc gggatccgag tgcccgggat	1260
50	cttccactgg ccgggggtgc tcccggccgg ccgagtgatt ggagagcca cgagcctgat	1320
	ggacgtgttc cctactgtgg tccagctggt ggggtggcgag gtgccccagg acaggggtgat	1380
	tgatggccac agcctggtac ccttgcctga gggagctgag gcacgctcgg cacatgagtt	1440
55	cctgtttcat tactgtgggc agcatcttca cgcagcacgc tggcaccaga aggacagtgg	1500
	aagcgtctgg aaggttcatt acacgacccc gcagttcac cccgaggagc ggggcctgct	1560
60	aacggccgag gcgtctgcc atgctgaatg gggaggcgtg accatcaca gaccccttt	1620
	gctctttgac ctctccagg accctccga ggcacggccc ctgacccccg actccgagcc	1680
	cctgtaccac gccgtgatag caagggtagg tgccgcggtg tcggagcacc ggagaccct	1740
65	gagtcctgtg cccagcagt tttccatgag caacatcctg tggaaaccgt ggctgcagcc	1800
	gtgctgcgga catttccgt tctgttcatg ccacgaggat ggggatggca cccctgaat	1860
	gccaggactg tgagagagga tccaggagag cctgactgcy ttgcaaaaaa aattctccaa	1920
	gcttggttct atcttcagtc cggaa	1945

ES 2 555 056 T3

5 <210> 21  
 <211> 593  
 <212> PRT  
 <213> Homo sapiens

10 <400> 21

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

Ser Leu Pro Val Leu Leu Phe Leu Cys Leu Leu Leu Lys Thr Cys Glu  
 20 25 30

Pro Lys Thr Ala Asn Ala Phe Lys Pro Asn Ile Leu Leu Ile Met Ala  
 35 40 45

Asp Asp Leu Gly Thr Gly Asp Leu Gly Cys Tyr Gly Asn Asn Thr Leu  
 50 55 60

Arg Thr Pro Asn Ile Asp Gln Leu Ala Glu Glu Gly Val Arg Leu Thr  
 65 70 75 80

Gln His Leu Ala Ala Ala Pro Leu Cys Thr Pro Ser Arg Ala Ala Phe  
 85 90 95

Leu Thr Gly Arg His Ser Phe Arg Ser Gly Met Asp Ala Ser Asn Gly  
 100 105 110

Tyr Arg Ala Leu Gln Trp Asn Ala Gly Ser Gly Gly Leu Pro Glu Asn  
 115 120 125

Glu Thr Thr Phe Ala Arg Ile Leu Gln Gln His Gly Tyr Ala Thr Gly  
 130 135 140

Leu Ile Gly Lys Trp His Gln Gly Val Asn Cys Ala Ser Arg Gly Asp  
 145 150 155 160

His Cys His His Pro Leu Asn His Gly Phe Asp Tyr Phe Tyr Gly Met  
 165 170 175

Pro Phe Thr Leu Thr Asn Asp Cys Asp Pro Gly Arg Pro Pro Glu Val  
 180 185 190

Asp Ala Ala Leu Arg Ala Gln Leu Trp Gly Tyr Thr Gln Phe Leu Ala  
 195 200 205

Leu Gly Ile Leu Thr Leu Ala Ala Gly Gln Thr Cys Gly Phe Phe Ser  
 210 215 220

Val Ser Ala Arg Ala Val Thr Gly Met Ala Gly Val Gly Cys Leu Phe  
 225 230 235 240

Phe Ile Ser Trp Tyr Ser Ser Phe Gly Phe Val Arg Arg Trp Asn Cys  
 245 250 255

Ile Leu Met Arg Asn His Asp Val Thr Glu Gln Pro Met Val Leu Glu  
 260 265 270

Lys Thr Ala Ser Leu Met Leu Lys Glu Ala Val Ser Tyr Ile Glu Arg  
 275 280 285

His Lys His Gly Pro Phe Leu Leu Phe Leu Ser Leu Leu His Val His  
 290 295 300

Ile Pro Leu Val Thr Thr Ser Ala Phe Leu Gly Lys Ser Gln His Gly  
 305 310 315 320

Leu Tyr Gly Asp Asn Val Glu Glu Met Asp Trp Leu Ile Gly Lys Val  
 325 330 335

5 Leu Asn Ala Ile Glu Asp Asn Gly Leu Lys Asn Ser Thr Phe Thr Tyr  
340 345 350

10 Phe Thr Ser Asp His Gly Gly His Leu Glu Ala Arg Asp Gly His Ser  
355 360 365

15 Gln Leu Gly Gly Trp Asn Gly Ile Tyr Lys Gly Gly Lys Gly Met Gly  
370 375 380

20 Gly Trp Glu Gly Gly Ile Arg Val Pro Gly Ile Phe His Trp Pro Gly  
385 390 395 400

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

30 Val Phe Pro Thr Val Val Gln Leu Val Gly Gly Glu Val Pro Gln Asp  
420 425 430

35 Arg Val Ile Asp Gly His Ser Leu Val Pro Leu Leu Gln Gly Ala Glu  
435 440 445

40 Ala Arg Ser Ala His Glu Phe Leu Phe His Tyr Cys Gly Gln His Leu  
450 455 460

45 His Ala Ala Arg Trp His Gln Lys Asp Ser Gly Ser Val Trp Lys Val  
465 470 475 480

50 His Tyr Thr Thr Pro Gln Phe His Pro Glu Glu Arg Gly Leu Leu Thr  
485 490 495

55 Ala Glu Ala Ser Ala His Ala Glu Trp Gly Gly Val Thr His His Arg  
500 505 510

60 Pro Pro Leu Leu Phe Asp Leu Ser Arg Asp Pro Ser Glu Ala Arg Pro  
515 520 525

65 Leu Thr Pro Asp Ser Glu Pro Leu Tyr His Ala Val Ile Ala Arg Val  
530 535 540

70 Gly Ala Ala Val Ser Glu His Arg Gln Thr Leu Ser Pro Val Pro Gln  
545 550 555 560

75 Gln Phe Ser Met Ser Asn Ile Leu Trp Lys Pro Trp Leu Gln Pro Cys  
565 570 575

80 Cys Gly His Phe Pro Phe Cys Ser Cys His Glu Asp Gly Asp Gly Thr  
580 585 590

85 Pro



ES 2 555 056 T3

5 <210> 22  
 <211> 1858  
 <212> ADN  
 <213> Homo sapiens

10 <400> 22

ccttcctcctt cttgatcggg gattcaggaa ggagcccagg agcagaggaa gtagagagag 60  
 agacaacatg ttacatctgc accattcttg tttgtgtttc aggagctggc tgccagcgat 120  
 15 gctcgcctgta ctgctaagtt tggcaccatc agcttccagc gacatttccg cctcccagacc 180  
 gaacatcctt cttctgatgg cggacgacct tggcattggg gacattggct gctatggcaa 240  
 20 caacaccatg aggactccga atattgaccg ccttgcagag gacggcgtga agctgaccca 300  
 acacatctct gccgcatctt tgtgcacccc aagcagagcc gccttctca cgggcagata 360  
 25 ccctgtgcga tcagggatgg tttccagcat tggttaccgt gttcttcagt ggaccggagc 420  
 atctggaggt cttccaacaa atgagacaac ttttgcaaaa atactgaaag agaaaggcta 480  
 tgccactgga ctcatggaa aatggcatct gggctcaca tgtgagtcag ccagtgatca 540  
 30 ttgccaccac cctctccatc atggctttga gcatttctac ggaatgcctt tctccttgat 600  
 gggtgattgc gccgcctggg aactctcaga gaagcgtgtc aacctggaac aaaaactcaa 660  
 ctctctcttc caagtcctgg ccttgggtgc cctcacactg gtagcagggg agctcacaca 720  
 35 cctgataccc gtctcgtgga tgccggatcat ctggctcagcc ctttcggccg tctctctcct 780  
 cgcaagctcc tattttgtgg gtgctctgat tgtccatgcc gattgctttc tgatgagaaa 840  
 ccacaccatc acggagcagc ccatgtgctt ccaaagaacg acaccctta ttctgcagga 900  
 40 ggttgctctc tttctcaaaa ggaataagca tgggcctttc ctctctcttg tttcctttct 960  
 acacgttcac atccctctta tcaactatgga gaacttctc gggaaagagtc tccaagggct 1020  
 gtatggggac aacgtagagg agatggactg gatggtagga cggatccttg acactttgga 1080  
 45 cgtggagggg ttgagcaaca gcacctcat ttattttacg tcggatcacg ggggttccct 1140  
 agagaatcaa cttggaaaca ccagtatgg tggctggaat ggaatctata aaggtgggaa 1200  
 gggcatggga ggalgggaag gtgggatccg cgtgcccggg atcttccgct ggccccgggt 1260  
 50 gctcccggcc ggccagatga ttggcgagcc caccagctct atggacgtgt tccccaccgt 1320  
 ggtccggctg gggggcggcg aggtgccccg ggacagagtg attgacggcc aagacctct 1380  
 gcccttgctc ctggggacag cccaacaactc agaccacgag ttctctgatgc attattgtga 1440  
 55 gaggtttctg cacgcagcca ggtggcatca acgggacaga ggaacaatgt ggaaagtcca 1500  
 ctltgtgacg cctgtgttcc agccagaggg agccgggtgc tgctatggaa gaaaggtctg 1560  
 cccgtgcttt ggggaaaaag tagtccacca cgatccacct ttgctctttg acctctcaag 1620  
 60 agacctctct gagaccaca tctcaccacc agcctcagag cccgtgttct atcaggtgat 1680  
 ggaacgagtc cagcaggcgg tgtgggaaca ccagcggaca ctcagcccag ttctctgca 1740  
 65 gctggacagg ctgggcaaca tctggagacc gtggctgcag ccctgctgtg gcccgttccc 1800  
 cctctgctgg tgccctaggg aagatgaccc acaataaatg totgcagtga aaagctgg 1858

ES 2 555 056 T3

5 <210> 23  
<211> 589  
<212> PRT  
<213> Homo sapiens

10 <400> 23

Met Leu His Leu His His Ser Cys Leu Cys Phe Arg Ser Trp Leu Pro  
1 5 10 15

15 Ala Met Leu Ala Val Leu Leu Ser Leu Ala Pro Ser Ala Ser Ser Asp  
20 25 30

20 Ile Ser Ala Ser Arg Pro Asn Ile Leu Leu Leu Met Ala Asp Asp Leu  
35 40 45

25

30

35

40

45

50

55

60

65

5 Gly Ile Gly Asp Ile Gly Cys Tyr Gly Asn Asn Thr Met Arg Thr Pro  
50 55 60

10 Asn Ile Asp Arg Leu Ala Glu Asp Gly Val Lys Leu Thr Gln His Ile  
65 70 75 80

15 Ser Ala Ala Ser Leu Cys Thr Pro Ser Arg Ala Ala Phe Leu Thr Gly  
85 90 95

20 Arg Tyr Pro Val Arg Ser Gly Met Val Ser Ser Ile Gly Tyr Arg Val  
100 105 110

25 Leu Gln Trp Thr Gly Ala Ser Gly Gly Leu Pro Thr Asn Glu Thr Thr  
115 120 125

30 Phe Ala Lys Ile Leu Lys Glu Lys Gly Tyr Ala Thr Gly Leu Ile Gly  
130 135 140

35 Lys Trp His Leu Gly Leu Asn Cys Glu Ser Ala Ser Asp His Cys His  
145 150 155 160

40 His Pro Leu His His Gly Phe Glu His Phe Tyr Gly Met Pro Phe Ser  
165 170 175

45 Leu Met Gly Asp Cys Ala Arg Trp Glu Leu Ser Glu Lys Arg Val Asn  
180 185 190

50 Leu Glu Gln Lys Leu Asn Phe Leu Phe Gln Val Leu Ala Leu Val Ala  
195 200 205

55 Leu Thr Leu Val Ala Gly Lys Leu Thr His Leu Ile Pro Val Ser Trp  
210 215 220

60 Met Pro Val Ile Trp Ser Ala Leu Ser Ala Val Leu Leu Leu Ala Ser  
225 230 235 240

65 Ser Tyr Phe Val Gly Ala Leu Ile Val His Ala Asp Cys Phe Leu Met  
245 250 255

70 Arg Asn His Thr Ile Thr Glu Gln Pro Met Cys Phe Gln Arg Thr Thr  
260 265 270

75 Pro Leu Ile Leu Gln Glu Val Ala Ser Phe Leu Lys Arg Asn Lys His  
275 280 285

80 Gly Pro Phe Leu Leu Phe Val Ser Phe Leu His Val His Ile Pro Leu  
290 295 300

85 Ile Thr Met Glu Asn Phe Leu Gly Lys Ser Leu His Gly Leu Tyr Gly  
305 310 315 320

90 Asp Asn Val Glu Glu Met Asp Trp Met Val Gly Arg Ile Leu Asp Thr  
325 330 335

95 Leu Asp Val Glu Gly Leu Ser Asn Ser Thr Leu Ile Tyr Phe Thr Ser  
340 345 350

100 Asp His Gly Gly Ser Leu Glu Asn Gln Leu Gly Asn Thr Gln Tyr Gly  
355 360 365

5  
 Gly Trp Asn Gly Ile Tyr Lys Gly Gly Lys Gly Met Gly Gly Trp Glu  
 370 375 380

10  
 Gly Gly Ile Arg Val Pro Gly Ile Phe Arg Trp Pro Gly Val Leu Pro  
 385 390 395 400

15  
 Ala Gly Arg Val Ile Gly Glu Pro Thr Ser Leu Met Asp Val Phe Pro  
 405 410 415

20  
 Thr Val Val Arg Leu Ala Gly Gly Glu Val Pro Gln Asp Arg Val Ile  
 420 425 430

25  
 Asp Gly Gln Asp Leu Leu Pro Leu Leu Leu Gly Thr Ala Gln His Ser  
 435 440 445

30  
 Asp His Glu Phe Leu Met His Tyr Cys Glu Arg Phe Leu His Ala Ala  
 450 455 460

35  
 Arg Trp His Gln Arg Asp Arg Gly Thr Met Trp Lys Val His Phe Val  
 465 470 475 480

40  
 Thr Pro Val Phe Gln Pro Glu Gly Ala Gly Ala Cys Tyr Gly Arg Lys  
 485 490 495

45  
 Val Cys Pro Cys Phe Gly Glu Lys Val Val His His Asp Pro Pro Leu  
 500 505 510

50  
 Leu Phe Asp Leu Ser Arg Asp Pro Ser Glu Thr His Ile Leu Thr Pro  
 515 520 525

55  
 Ala Ser Glu Pro Val Phe Tyr Gln Val Met Glu Arg Val Gln Gln Ala  
 530 535 540

60  
 Val Trp Glu His Gln Arg Thr Leu Ser Pro Val Pro Leu Gln Leu Asp  
 545 550 555 560

65  
 Arg Leu Gly Asn Ile Trp Arg Pro Trp Leu Gln Pro Cys Cys Gly Pro  
 565 570 575

Phe Pro Leu Cys Trp Cys Leu Arg Glu Asp Asp Pro Gln  
 580 585

<210> 24  
 <211> 1996  
 <212> ADN  
 <213> Homo sapiens

<400> 24

5  
 gggttctgct cctagacatt agagagataa tacggctgat agacaacaag aaggatttcc 60  
 aagctgcaca atgaggccca ggagaccgtt ggtcttcatg tctttgggtg gtgcactott 120  
 10 gaacacatgg ccagggcaca cagggtgcat gacgacaagg cctaatttg tctaatacat 180  
 ggttgatgac ctgggtattg gagatctggg ctgctacggc aatgacacca tgaggacgcc 240  
 tcacatogac cgccttgcca ggaaggogt gcgactgact cagcacatct ctgccgcctc 300  
 15 cctctgcage ccaagccggt ccgcttctt gacgggaaga taccocatcc gatcagggtat 360  
 ggtttctagt ggtaatagac gtgtcatoca aaatcttgca gtccccgag gcctccctct 420  
 taatgagaca acacttgag ccttgctaaa gaagcaagga tacagcacgg ggcttatagg 480  
 20 caaatggcac caaggcttga actgcgactc ccgaagtgac cagtgccacc atccatataa 540  
 ttatgggttt gactactact atggcatgcc gtctactctc gttgacagct gctggccgga 600  
 25 cccctctcgt aacacggaat tagccttga gactcagctc tggctctgtg tgcagctagt 660  
 tgcattgcc atcctcacc taaccttgg gaagctgagc ggctgggtct ctgttccctg 720  
 gctcctgac ttctccatga ttctgtttat ttctctcttg ggctatgctt ggttctccag 780  
 30 ccacacgtcc cctttatact gggactgcct cctcatgagg gggcacgaga tcacggagca 840  
 gccatgaag gctgaacgag ctggatccat tatggtgaag gaagcgattt cctttttaga 900  
 aaggcacagt aaggaaactt tccttctctt ttctctctt ctccacgtgc acacacctct 960  
 35 ccccaccacg gacgatttca ctggcaccag caagcatggc ttgtatgggg ataatgtgga 1020  
 agagatggac tccatggtgg gcaagattct tgatgctatc gatgattttg gcctaaggaa 1080  
 caacaccctt gtctacttta catcagatca cggagggcat ttggaagcta ggcgagggca 1140  
 40 tgcceaactt ggtggatgga atggaatata caaagggtgga aaaggcatgg ggggctggga 1200  
 aggtggaatc cgcgtcccag gaattgtccg atggcctgga aaggtagcag ctggacgggt 1260  
 45 gattaaggaa cctacaagtt taatggatat ttaccaact gtccatcag tgtcaggagg 1320  
 aagtctccct caggacaggg tcattgagg ccgagacctc atgcccttgc tgcagggcaa 1380  
 cgtcaggcac tcggagcatg aatttctttt ccactactgt ggctcctacc tgcacgccgt 1440  
 50 gcggtggatc cccaaggacg acagtgggtc agtttggag gctcactatg tgaccccggt 1500  
 attccagcca ccagcttctg gtggctgcta tgtcaacctc ttatgcagat gtttcggaga 1560  
 acaggttacc taccacaacc cccctctgct ctccgatctc tccagggacc cctcagagtc 1620  
 55 cacaccctg acacctgcca cagagcccct ctatgatttt gtgattaaaa aggtggccaa 1680  
 cgcctgaag gaaccaccag aaaccatcgt gcctgtgacc taccaactct cagaactgaa 1740  
 tcagggcagg acgtggctga agccttgctg tgggtgttc ccattttgtc tgtgtgacaa 1800  
 60 ggaagaggaa gtctctcagc ctccgggtcc taacgagaag agataattac aatcaggcta 1860  
 ccagaggaag cctttggtcc taacgagaag agataattac aatcaggcta ccaaggaag 1920  
 65 cactaacttt ggtgcttca agttggcaag gactgcattt aatagtcaat aaatcactct 1980  
 accattccag attatt 1996

ES 2 555 056 T3

5

<210> 25  
<211> 591  
<212> PRT  
<213> Homo sapiens

10

<400> 25

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

30

35

40

45

50

55

60

65

ES 2 555 056 T3

5  
 Glu Gly Val Arg Leu Thr Gln His Ile Ser Ala Ala Ser Leu Cys Ser  
 65 70 75 80

10  
 Pro Ser Arg Ser Ala Phe Leu Thr Gly Arg Tyr Pro Ile Arg Ser Gly  
 85 90 95

15  
 Met Val Ser Ser Gly Asn Arg Arg Val Ile Gln Asn Leu Ala Val Pro  
 100 105 110

20  
 Ala Gly Leu Pro Leu Asn Glu Thr Thr Leu Ala Ala Leu Leu Lys Lys  
 115 120 125

25  
 Gln Gly Tyr Ser Thr Gly Leu Ile Gly Lys Trp His Gln Gly Leu Asn  
 130 135 140

30  
 Cys Asp Ser Arg Ser Asp Gln Cys His His Pro Tyr Asn Tyr Gly Phe  
 145 150 155 160

35  
 Asp Tyr Tyr Tyr Gly Met Pro Phe Thr Leu Val Asp Ser Cys Trp Pro  
 165 170 175

40  
 Asp Pro Ser Arg Asn Thr Glu Leu Ala Phe Glu Ser Gln Leu Trp Leu  
 180 185 190

45  
 Cys Val Gln Leu Val Ala Ile Ala Ile Leu Thr Leu Thr Phe Gly Lys  
 195 200 205

50  
 Leu Ser Gly Trp Val Ser Val Pro Trp Leu Leu Ile Phe Ser Met Ile  
 210 215 220

55  
 Leu Phe Ile Phe Leu Leu Gly Tyr Ala Trp Phe Ser Ser His Thr Ser  
 225 230 235 240

60  
 Pro Leu Tyr Trp Asp Cys Leu Leu Met Arg Gly His Glu Ile Thr Glu  
 245 250 255

65  
 Gln Pro Met Lys Ala Glu Arg Ala Gly Ser Ile Met Val Lys Glu Ala  
 260 265 270

70  
 Ile Ser Phe Leu Glu Arg His Ser Lys Glu Thr Phe Leu Leu Phe Phe  
 275 280 285

75  
 Ser Phe Leu His Val His Thr Pro Leu Pro Thr Thr Asp Asp Phe Thr  
 290 295 300

80  
 Gly Thr Ser Lys His Gly Leu Tyr Gly Asp Asn Val Glu Glu Met Asp  
 305 310 315 320

85  
 Ser Met Val Gly Lys Ile Leu Asp Ala Ile Asp Asp Phe Gly Leu Arg  
 325 330 335

90  
 Asn Asn Thr Leu Val Tyr Phe Thr Ser Asp His Gly Gly His Leu Glu  
 340 345 350

95  
 Ala Arg Arg Gly His Ala Gln Leu Gly Gly Trp Asn Gly Ile Tyr Lys  
 355 360 365

100  
 Gly Gly Lys Gly Met Gly Gly Trp Glu Gly Gly Ile Arg Val Pro Gly  
 370 375 380

5  
 Ile Val Arg Trp Pro Gly Lys Val Pro Ala Gly Arg Leu Ile Lys Glu  
 385 390 395 400

10  
 Pro Thr Ser Leu Met Asp Ile Leu Pro Thr Val Ala Ser Val Ser Gly  
 405 410 415

15  
 Gly Ser Leu Pro Gln Asp Arg Val Ile Asp Gly Arg Asp Leu Met Pro  
 420 425 430

20  
 Leu Leu Gln Gly Asn Val Arg His Ser Glu His Glu Phe Leu Phe His  
 435 440 445

25  
 Tyr Cys Gly Ser Tyr Leu His Ala Val Arg Trp Ile Pro Lys Asp Asp  
 450 455 460

30  
 Ser Gly Ser Val Trp Lys Ala His Tyr Val Thr Pro Val Phe Gln Pro  
 465 470 475 480

35  
 Pro Ala Ser Gly Gly Cys Tyr Val Thr Ser Leu Cys Arg Cys Phe Gly  
 485 490 495

40  
 Glu Gln Val Thr Tyr His Asn Pro Pro Leu Leu Phe Asp Leu Ser Arg  
 500 505 510

45  
 Asp Pro Ser Glu Ser Thr Pro Leu Thr Pro Ala Thr Glu Pro Leu Tyr  
 515 520 525

50  
 Asp Phe Val Ile Lys Lys Val Ala Asn Ala Leu Lys Glu His Gln Glu  
 530 535 540

55  
 Thr Ile Val Pro Val Thr Tyr Gln Leu Ser Glu Leu Asn Gln Gly Arg  
 545 550 555

60  
 Thr Trp Leu Lys Pro Cys Cys Gly Val Phe Pro Phe Cys Leu Cys Asp  
 565 570 575

65  
 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



ES 2 555 056 T3

5 atgggctggc tttttctaaa ggttttggtg gcgggagtga gtttctcagg atttctttat 60  
cctcttgtgg atttttgcat cagtgggaaa acaagaggac agaagccaaa ctttgtgatt 120  
athttggccg atgacatggg gtgggggtgac ctggggagcaa actgggcaga aacaaaggac 180  
10 actgccaacc ttgataagat ggcttcggag ggaatgaggt ttgtggattt ccatgcagct 240  
gcctccacct gctcaccctc ccgggcttcc ttgctcaccg gccggcttgg ccttcgcaat 300  
15 ggagtccacac gcaactttgc agtcacttct gtgggaggcc ttccgctcaa cgagaccacc 360  
ttggcagagg tgctgcagca ggcggttac gtcactggga taataggcaa atggcatctt 420  
ggacaccacg gctcttatca cccaacttc cgtgggtttg attactactt tggaatccca 480  
20 tatagccatg atatgggctg tactgatact ccaggctaca accaccctcc ttgtccagcg 540  
tgtccacagg gtgatggacc atcaaggaac cttcaaagag actgttacac tgacgtggcc 600  
ctccctcttt atgaaaacct caacattgtg gaggagccgg tgaacttgag cagccttgcc 660  
25 cagaagtatg ctgagaaagc aaccagttc atccagcgtg caagcaccag cgggaggccc 720  
ttcctgctct atgtggctct ggcccacatg cacgtgcctt tacctgtgac tcagctacca 780  
gcagcgcac ggggcagaag cctgtatggt gcagggctct gggagatgga cagtctggtg 840  
30 ggccagatca aggacaaagt tgaccacaca gtgaaggaaa acacattcct ctggtttaca 900  
ggagacaatg gcccggtggc tcagaagtgt gagctagcgg gcagtgtggg tcccttact 960  
ggathttggc aaactcgtca agggggaagt ccagccaagc agacgacctg ggaaggaggg 1020  
35 caccgggtcc cagcactggc ttactggcct ggcagagttc cagttaatgt caccagcact 1080  
gccttgttaa gcgtgctgga catttttcca actgtggtag ccctggcca gccagctta 1140  
cctcaaggac ggcgctttga tgggtgtggac gtctccgagg tgccttttg ccggtcacag 1200  
40 cctgggcaca ggggtgctgtt ccaccccaac agcggggcag ctggagagtt tggagccctg 1260  
cagactgtcc gcctggagcg ttacaaggcc ttctacatta ccggtggagc cagggcgtgt 1320  
45 gatgggagca cggggcctga gctgcagcat aagtttctc tgathttcaa cctggaagac 1380  
gataccgcag aagctgtgcc cctagaaaga ggtggtgagg agtaccaggc tgtgctgccc 1440  
gaggtcagaa aggttcttgc agacgtctc caagacattg ccaacgaca catctccagc 1500  
50 gcagattaca ctcaggacct ttcagtaact ccctgctgta atccctacca aattgcctgc 1560  
cgctgtcaag ccgcataa 1578

55 <210> 27  
<211> 525  
<212> PRT  
<213> Homo sapiens

60 <400> 27

65

5  
 Met Gly Trp Leu Phe Leu Lys Val Leu Leu Ala Gly Val Ser Phe Ser  
 1 5 10 15

10  
 Gly Phe Leu Tyr Pro Leu Val Asp Phe Cys Ile Ser Gly Lys Thr Arg  
 20 25 30

15  
 Gly Gln Lys Pro Asn Phe Val Ile Ile Leu Ala Asp Asp Met Gly Trp  
 35 40 45

20  
 Gly Asp Leu Gly Ala Asn Trp Ala Glu Thr Lys Asp Thr Ala Asn Leu  
 50 55 60

25  
 Asp Lys Met Ala Ser Glu Gly Met Arg Phe Val Asp Phe His Ala Ala  
 65 70 75 80

30  
 Ala Ser Thr Cys Ser Pro Ser Arg Ala Ser Leu Leu Thr Gly Arg Leu  
 85 90 95

35  
 Gly Leu Arg Asn Gly Val Thr Arg Asn Phe Ala Val Thr Ser Val Gly  
 100 105 110

40  
 Gly Leu Pro Leu Asn Glu Thr Thr Leu Ala Glu Val Leu Gln Gln Ala  
 115 120 125

45  
 Gly Tyr Val Thr Gly Ile Ile Gly Lys Trp His Leu Gly His His Gly  
 130 135 140

50

55

60

65

ES 2 555 056 T3

5 Ser Tyr His Pro Asn Phe Arg Gly Phe Asp Tyr Tyr Phe Gly Ile Pro  
145 150 155 160

10 Tyr Ser His Asp Met Gly Cys Thr Asp Thr Pro Gly Tyr Asn His Pro  
165 170 175

15 Pro Cys Pro Ala Cys Pro Gln Gly Asp Gly Pro Ser Arg Asn Leu Gln  
180 185 190

20 Arg Asp Cys Tyr Thr Asp Val Ala Leu Pro Leu Tyr Glu Asn Leu Asn  
195 200 205

25 Ile Val Glu Gln Pro Val Asn Leu Ser Ser Leu Ala Gln Lys Tyr Ala  
210 215 220

30 Glu Lys Ala Thr Gln Phe Ile Gln Arg Ala Ser Thr Ser Gly Arg Pro  
225 230 235 240

35 Phe Leu Leu Tyr Val Ala Leu Ala His Met His Val Pro Leu Pro Val  
245 250 255

40 Thr Gln Leu Pro Ala Ala Pro Arg Gly Arg Ser Leu Tyr Gly Ala Gly  
260 265 270

45 Leu Trp Glu Met Asp Ser Leu Val Gly Gln Ile Lys Asp Lys Val Asp  
275 280 285

50 His Thr Val Lys Glu Asn Thr Phe Leu Trp Phe Thr Gly Asp Asn Gly  
290 295 300

55 Pro Trp Ala Gln Lys Cys Glu Leu Ala Gly Ser Val Gly Pro Phe Thr  
305 310 315 320

60 Gly Phe Trp Gln Thr Arg Gln Gly Gly Ser Pro Ala Lys Gln Thr Thr  
325 330 335

65 Trp Glu Gly Gly His Arg Val Pro Ala Leu Ala Tyr Trp Pro Gly Arg  
340 345 350

70 Val Pro Val Asn Val Thr Ser Thr Ala Leu Leu Ser Val Leu Asp Ile  
355 360 365

75 Phe Pro Thr Val Val Ala Leu Ala Gln Ala Ser Leu Pro Gln Gly Arg  
370 375 380

80 Arg Phe Asp Gly Val Asp Val Ser Glu Val Leu Phe Gly Arg Ser Gln  
385 390 395 400

85 Pro Gly His Arg Val Leu Phe His Pro Asn Ser Gly Ala Ala Gly Glu  
405 410 415

90 Phe Gly Ala Leu Gln Thr Val Arg Leu Glu Arg Tyr Lys Ala Phe Tyr  
420 425 430

95 Ile Thr Gly Gly Ala Arg Ala Cys Asp Gly Ser Thr Gly Pro Glu Leu  
435 440 445

ES 2 555 056 T3

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

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

25

<210> 28  
<211> 4669  
<212> ADN  
<213> Homo sapiens

30

<400> 28

cgcagaccgt cgctaataaa tcttggggcc ggtgtcgggc cggggcggct tgatcggcaa 60  
ctaggaaacc ccaggcgcag aggccaggag cgagggcagc gaggatcaga ggccaggcct 120  
35 tcccggctgc cggcgctcct cggaggtcag ggcagatgag gaacatgact ctcccccttc 180  
ggaggaggaa ggaagtcccg ctgccacctt atctctgctc ctctgcctcc tccctgttcc 240  
cagagctttt tctctagaga agattttgaa ggccgctttt gtgctgacgg ccaccacca 300  
40 tcatctaaag aagataaact tggcaaatga catgcagggt cttcaaggca gaataattgc 360  
agaaaatctt caaaggacct tatctgcaga tgtttgaat acctctgaga atagagattg 420  
attattcaac caggatacct aattcaagaa ctccagaaat caggagacgg agacattttg 480  
tcagttttgc aacattggac caaatacaat gaagtattct tgctgtgctc tggttttggc 540  
45 tgtcctgggc acagaattgc tgggaagcct ctgttogact gtcagatccc cgaggttcag 600  
aggacggata cagcaggaac gaaaaaacat ccgaccaac attattcttg tgcttaccga 660  
tgatcaagat gtggagctgg ggtccctgca agtcatgaac aaaacgagaa agattatgga 720  
50 acatgggggg gccaccttca tcaatgcctt tgtgactaca cccatgtgct gcccgtcacg 780  
gtctccatg ctccaccgga agtatgtgca caatcacaat gtctacacca acaacgagaa 840  
ctgctcttcc cctcctgtggc aggccatgca tgagcctcgg acttttgctg tatactctaa 900  
55 caacactggc tacagaacag ccttttttgg aaaatacctc aatgaatata atggcagcta 960  
cctccccctt ggggtggcag aatggccttg attaatcaag aattctcgtc tctataatta 1020  
cactgtttgt cgcaatggca tcaaagaaaa gcatggattt gattatgcaa aggactactt 1080  
60 cacagactta atcactaacg agagcattaa ttacttcaaa atgtctaaga gaatgtatcc 1140  
ccataggccc gttatgatgg tgatcagcca cgctgcgccc cacggccccg aggactcagc 1200  
cccacagttt tctaaactgt accccaatgc ttccaacac ataactccta gttataacta 1260  
tgcaccaa atggataaac actggattat gcagtacaca ggaccaatgc tgcccatcca 1320  
65 catggaattt acaaacatct tacagcgcaa aaggctccag actttgatgt cagtggatga 1380  
ttctgtggag aggctgtata acatgctcgt ggagacgggg gagctggaga atacttacat 1440  
catttacacc gccgaccatg gttaccatat tgggcagttt ggactgggtca aggggaaatc 1500  
catgccatat gactttgata ttcgtgtgcc tttttttatt cgtgggtccaa gtgtagaacc 1560

ES 2 555 056 T3

5	aggatcaata	gtccccacaga	tcgttctcaa	cattgacttg	gccccacga	tcctggatat	1620
	tgctgggctc	gacacacctc	ctgatgtgga	eggcaagtct	gtcctcaaac	ttctggaccc	1680
	agaaaagcca	ggtaacaggt	ttcgaacaaa	caagaaggcc	aaaatttggc	gtgatacatt	1740
10	cctagtggaa	agaggcaaat	ttctacgtaa	gaaggaagaa	tccagcaaga	atatccaaca	1800
	gtcaaatcac	ttgccccaat	atgaacgggt	caaagaacta	tgccagcagg	ccaggtacca	1860
	gacagcctgt	gaacaaccgg	ggcagaagtg	gcaatgcatt	gaggatacat	ctggcaagct	1920
15	togaattcac	aagtgtaaag	gacccagtga	cctgctcaca	gtccggcaga	gcacgcggaa	1980
	cctctacgct	cgcggttcc	atgacaaa	caaagagtgc	agttgtaggg	agtctggtta	2040
	ccgtgccagc	agaagccaaa	gaaagagtca	acggcaattc	ttgagaaacc	aggggactcc	2100
20	aaagtacaag	cccagatttg	tccatactcg	gcagacacgt	tccttgtccg	tcgaatttga	2160
	aggtgaaata	tatgacataa	atctggaaga	agaagaagaa	ttgcaagtgt	tgcaaccaag	2220
	aaacattgct	aagcgtcatg	atgaaggcca	caaggggcca	agagatctcc	aggcttccag	2280
25	tggtggcaac	aggggcagga	tgctggcaga	tagcagcaac	gccgtgggcc	cacctaccac	2340
	tgtccgagtg	acacacaagt	gttttattct	tccaatgac	totatccatt	gtgagagaga	2400
	actgtacca	tcggccagag	cgtggaagga	ccataaggca	tacattgaca	aagagattga	2460
30	agctctgcaa	gataaaatta	agaatttaag	agaagtgaga	ggacatctga	agagaaggaa	2520
	gcctgaggaa	tgtagctgca	gtaaacaaag	ctattacaat	aaagagaaag	gtgtaaaaaa	2580
	gcaagagaaa	ttaaagagcc	atcttcaccc	attcaaggag	gctgctcagg	aagtagatag	2640
35	caaactgcaa	cttttcaagg	agaacaaccg	taggaggaag	aaggagagga	aggagaagag	2700
	acggcagagg	aagggggaag	agtgcagcct	gcctggcctc	acttgcttca	cgcatgacaa	2760
	caaccactgg	cagacagccc	cgttctggaa	cctgggatct	ttctgtgctt	gcacgagttc	2820
40	taacaataac	acctactggg	gtttgcgtac	agttaatgag	acgcataatt	ttcttttctg	2880
	tgagtttgct	actggctttt	tggagtattt	tgatatgaat	acagatcctt	atcagctcac	2940
	aaatacagtg	cacacggtag	aacgaggcat	tttgaatcag	ctacacgtac	aactaatgga	3000
45	gctcagaagc	tgtcaaggat	ataagcagtg	caacccaaga	cctaagaatc	ttgatgttgg	3060
	aaataaagat	ggaggaagct	atgacctaca	cagaggacag	ttatgggatg	gatgggaagg	3120
	ttaatcagcc	ccgtctcact	gcagacatca	actggcaagg	cctagaggag	ctacacagtg	3180
50	tgaatgaaaa	catctatgag	tacagacaaa	actacagact	tagtctgggtg	gactggacta	3240
	attaactgaa	ggatttagat	agagtatttg	cactgctgaa	gagtcactat	gagcaaaata	3300
	aaacaaataa	gactcaaact	gctcaaagtg	acgggttctt	ggttgtctct	gctgagcacg	3360
55	ctgtgtcaat	ggagatggcc	tctgctgact	cagatgaaga	cccaaggcat	aaggttggga	3420
	aaacacctca	tttgaccttg	ccagctgacc	ttcaaaccct	gcatttgaac	cgaccaacat	3480
60	taagtccaga	gagtaaactt	gaatggaata	acgacattcc	agaagttaat	catttgaatt	3540
	ctgaacactg	gagaaaaacc	gaaaaatgga	cggggcatga	agagactaat	catctggaaa	3600
	ccgatttcag	tggcgatggc	atgacagagc	tagagctcgg	gcccagcccc	aggctgcagc	3660
65	ccattcacag	gcacccgaaa	gaacttcccc	agtatggtgg	tcctggaaag	gacatttttg	3720
	aagatcaact	atatcttcc	gtgcattccg	atggaatttc	agttcatcag	atgttcacca	3780
	tggccaccgc	agaacaccga	agtaattcca	gcatagcggg	gaagatggtg	accaaggtgg	3840
	agaagaatca	cgaaaaggag	aagtcacagc	acctagaagg	cagcacctcc	tcttactct	3900

ES 2 555 056 T3

5 cctctgatta gatgaaactg ttaccttacc ctaaacacag tatttcctttt taactttttt 3960  
 atttgtaaac taataaaggt aatcacagcc accaacattc caagctaccc tgggtacctt 4020  
 10 tgtgcagtag aagctagtga gcatgtgagc aagcgggtgtg cacacggaga ctcatcgta 4080  
 taatttacta tctgccaaga gtagaaagaa aggcctgggga tatttggggtt ggcttgggtt 4140  
 tgattttttg cttgtttgtt tgttttgtac taaaacagta ttatcttttg aatctcgtag 4200  
 15 ggacataagt atatacatgt tatccaatca agatggctag aatgggtgcct ttctgagtg 4260  
 ctaaaacttg acacccttg taaatctttc aacacacttc cactgcctgc gtaatgaagt 4320  
 tttgattcat ttttaaccac tgggaattttt caatgcctgc attttcagtt agatgatttt 4380  
 20 gcactttgag attaaaatgc catgtctatt tgattagtct tattttttta tttttacagg 4440  
 cttatcagtc tcaactgttg ctgtcattgt gacaaagtca aataaacccc caaggacgac 4500  
 acacagtatg gatcacatat tgtttgacat taagcttttg ccagaaaatg ttgcatgtgt 4560  
 25 tttacctcga cttgctaaaa tggattagca gaaaggcatg gctaataatg ttgggtggtga 4620  
 aaataaataa ataagtaaat gaaaaaaaa aaaaaaaaa aaaaaaaaa 4669

<210> 29  
 <211> 871  
 30 <212> PRT  
 <213> Homo sapiens  
 <400> 29

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

ES 2 555 056 T3

5 Gly Ile Lys Glu Lys His Gly Phe Asp Tyr Ala Lys Asp Tyr Phe Thr  
180 185 190

10 Asp Leu Ile Thr Asn Glu Ser Ile Asn Tyr Phe Lys Met Ser Lys Arg  
195 200 205

15 Met Tyr Pro His Arg Pro Val Met Met Val Ile Ser His Ala Ala Pro  
210 215 220

20 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

25 Lys His Trp Ile Met Gln Tyr Thr Gly Pro Met Leu Pro Ile His Met  
260 265 270

30 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

35 Glu Leu Glu Asn Thr Tyr Ile Ile Tyr Thr Ala Asp His Gly Tyr His  
305 310 315 320

40 Ile Gly Gln Phe Gly Leu Val Lys Gly Lys Ser Met Pro Tyr Asp Phe  
325 330 335

45 Asp Ile Arg Val Pro Phe Phe Ile Arg Gly Pro Ser Val Glu Pro Gly  
340 345 350

Ser Ile Val Pro Gln Ile Val Leu Asn Ile Asp Leu Ala Pro Thr Ile  
355 360 365

50 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

55 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

60 Asn His Leu Pro Lys Tyr Glu Arg Val Lys Glu Leu Cys Gln Gln Ala  
435 440 445

Arg Tyr Gln Thr Ala Cys Glu Gln Pro Gly Gln Lys Trp Gln Cys Ile  
450 455 460

65 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

ES 2 555 056 T3

5 Phe His Asp Lys Asp Lys Glu Cys Ser Cys Arg Glu Ser Gly Tyr Arg  
500 505 510

10 Ala Ser Arg Ser Gln Arg Lys Ser Gln Arg Gln Phe Leu Arg Asn Gln  
515 520 525

15 Gly Thr Pro Lys Tyr Lys Pro Arg Phe Val His Thr Arg Gln Thr Arg  
530 535 540

20 Ser Leu Ser Val Glu Phe Glu Gly Glu Ile Tyr Asp Ile Asn Leu Glu  
545 550 555 560

25 Glu Glu Glu Glu Leu Gln Val Leu Gln Pro Arg Asn Ile Ala Lys Arg  
565 570 575

30 His Asp Glu Gly His Lys Gly Pro Arg Asp Leu Gln Ala Ser Ser Gly  
580 585 590

35 Gly Asn Arg Gly Arg Met Leu Ala Asp Ser Ser Asn Ala Val Gly Pro  
595 600 605

40 Pro Thr Thr Val Arg Val Thr His Lys Cys Phe Ile Leu Pro Asn Asp  
610 615 620

45 Ser Ile His Cys Glu Arg Glu Leu Tyr Gln Ser Ala Arg Ala Trp Lys  
625 630 635 640

50 Asp His Lys Ala Tyr Ile Asp Lys Glu Ile Glu Ala Leu Gln Asp Lys  
645 650 655

55 Ile Lys Asn Leu Arg Glu Val Arg Gly His Leu Lys Arg Arg Lys Pro  
660 665 670

60 Glu Glu Cys Ser Cys Ser Lys Gln Ser Tyr Tyr Asn Lys Glu Lys Gly  
675 680 685

65 Val Lys Lys Gln Glu Lys Leu Lys Ser His Leu His Pro Phe Lys Glu  
690 695 700

70 Ala Ala Gln Glu Val Asp Ser Lys Leu Gln Leu Phe Lys Glu Asn Asn  
705 710 715 720

75 Arg Arg Arg Lys Lys Glu Arg Lys Glu Lys Arg Arg Gln Arg Lys Gly  
725 730 735

80 Glu Glu Cys Ser Leu Pro Gly Leu Thr Cys Phe Thr His Asp Asn Asn  
740 745 750

85 His Trp Gln Thr Ala Pro Phe Trp Asn Leu Gly Ser Phe Cys Ala Cys  
755 760 765

90 Thr Ser Ser Asn Asn Asn Thr Tyr Trp Cys Leu Arg Thr Val Asn Glu  
770 775 780

95 Thr His Asn Phe Leu Phe Cys Glu Phe Ala Thr Gly Phe Leu Glu Tyr  
785 790 795 800



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

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

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

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

25 Leu Trp Asp Gly Trp Glu Gly  
865 870

25 <210> 30  
<211> 4279  
<212> ADN  
<213> Homo sapiens

30 <400> 30

35

40

45

50

55

60

65

ES 2 555 056 T3

5  
 gggccatttc tggacaacag ctgctatfff cacttgagcc caagttaatt tctcggggag 60  
 ttctcgggag cgcacaggca gctcggtttg ccctgcgatt gagctgcggg tcgcgggcgg 120  
 10 cgccggcctc tccaatggca aatgtgtgtg gctggaggcg agcgcgaggc tttcggcaaa 180  
 ggcagtcgag tgtttgacaga cgggggagag tctgtgaaa gcagataaaa gaaaacattt 240  
 attaacgtgt cattacgagg ggagcgcggc gccggggctg tcgcaactcc cgcggaacat 300  
 15 ttggctccct ccagctccta gagaggagaa gaagaaagcg gaaaagaggc agattcacgt 360  
 cgtttccagc caagtggacc tgatcgatgg ccctcctgaa tttatcacga tatttgattt 420  
 attagcgatg cccctgggtt tgtgtgttac gcacacacac gtgcacacaa ggctctggct 480  
 20 cgcttcctc cctcgtttcc agctcctggg cgaatcccac atctgtttca actctccgcc 540  
 gagggcgagc aggagcgaga gtgtgtcgaa tctgcgagtg aagagggagc agggaaaaga 600  
 25 aacaaagcca cagacgcaac ttgagactcc cgcacccaa aagaagcacc agatcagcaa 660  
 aaaaagaaga tgggcccccc gagcctcgtg ctgtgcttgc tgtccgcaac tgtgtttctc 720  
 ctgctgggtg gaagctcggc cttcctgtcg caccaccgcc tgaaaggcag gtttcagagg 780  
 30 gaccgcagga acatccgcc caacatcctc ctgggtctga cggacgacca ggatgtggag 840  
 ctgggttcca tgcagtgat gaacaagacc cggcgcacatca tggagcaggg cggggcgcac 900  
 ttcatcaacg ccttcgtgac cacacccatg tgctgcccct cacgctcctc catcctcacc 960  
 35 ggcaagtacg tccacaacca caacacctac accaacaatg agaactgctc ctccgccctc 1020  
 tggcaggcac agcacgagag ccgcacctt gccgtgtacc tcaatagcac tggctaccgg 1080  
 acagotttct tcgggaagta tcttaatgaa tacaacggct cctacgtgcc acccggtgg 1140  
 40 aaggagtggg tcggactcct taaaaactcc cgcttttata actacacgct gtgtcggaac 1200  
 ggggtgaaag agaagcacgg ctccgactac tccaaggatt acctcacaga cctcatcacc 1260  
 45 aatgacagcg tgagcttctt ccgcacgtcc aagaagatgt acccgcacag gccagtcctc 1320  
 atggctatca gccatgcagc cccccacggc cctgaggatt cagccccaca atattcacgc 1380  
 ctcttcccaa acgcatctca gcacatcacg ccgagctaca actacgggcc caaccggac 1440  
 50 aaacactgga tcatgcgcta cacggggccc atgaagccca tccacatgga attcaccaac 1500  
 atgctccagc ggaagcgtct gcagacctc atgtcgggtg acgactccat ggagacgatt 1560

55

60

65

ES 2 555 056 T3

5 tacaacatgc tggttgagac gggcgagctg gacaacacgt acatcgtata caccgccgac 1620  
cacggttacc acatcggcca gtttggcctg gtgaaagga aatccatgcc atatgagttt 1680  
gacatcaggg tcccgttcta cgtgaggggc cccaacgtgg aagccggctg tctgaatccc 1740  
10 cacatcgtcc tcaacattga cctggccccc accatcctgg acattgcagg cctggacata 1800  
cctgcgggata tggacgggaa atccatcctc aagctgctgg acacggagcg gccggtgaat 1860  
cggtttcact tgaaaaagaa gatgagggtc tggcgggact ccttcttggg ggagagagggc 1920  
15 aagctgctac acaagagaga caatgacaag gtggacgccc aggaggagaa ctttctgccc 1980  
aagtaccagc gtgtgaagga cctgtgtcag cgtgctgagt accagacggc gtgtgagcag 2040  
ctgggacaga agtggcagtg tgtggaggac gccacgggga agctgaagct gcataagtgc 2100  
20 aagggcccca tgcggtctgg cggcagcaga gccctctcca acctcgtgcc caagtactac 2160  
gggcagggca gcgaggcctg cacctgtgac agcggggact acaagctcag cctggccgga 2220  
cgccggaaaa aactcttcaa gaagaagtac aaggccagct atgtccgag tcgctccatc 2280  
25 cgctcagtg ccacgcaggt ggacggcagg gtgtaccacg taggcctggg tgatgccgcc 2340  
cagccccgaa acctcaccaa gcggcactgg ccaggggccc ctgaggacca agatgacaag 2400  
gatggtgggg acttcagtg cactggaggc cttcccact actcagccgc caaccccatt 2460  
30 aaagtgacac atcgggtgcta catcctagag aacgacacag tccagtgtga cctggacctg 2520  
tacaagtccc tgcaggcctg gaaagaccac aagctgcaca tcgaccacga gattgaaacc 2580  
ctgcagaaca aaattaagaa cctgagggaa gtccgaggtc acctgaagaa aaagcggcca 2640  
35 gaagaatgtg actgtcacia aatcagctac cacaccacg acaaaggccg cctcaagcac 2700  
agaggctcca gtctgcatcc tttcaggaag ggcctgcaag agaaggacaa ggtgtggctg 2760  
ttgcccggagc agaagcgcaa gaagaaactc cgcaagctgc tcaagcgcct gcagaacaac 2820  
40 gacacgtgca gcatgccagg cctcacgtgc ttcaccacg acaaccagca ctggcagacg 2880  
gggcctttct ggacactggg gcctttctgt gcctgcacca ggcacaaca taacacgtac 2940  
tggtgcatga ggaccatcaa tgagactcac aatttcctct tctgtgaatt tgcaactggc 3000  
45 ttctagagat actttgatct caacacagac ccctaccagc tgatgaatgc agtgaacaca 3060  
ctggacaggg atgtcctcaa ccagctacac gtacagctca tggagctgag gagctgcaag 3120  
ggttacaagc agtgtaaccc cggactcga aacatggacc tgggacttaa agatggagga 3180  
50 agctatgagc aatacaggca gtttcagcgt cgaaagtggc cagaaatgaa gagacctct 3240  
tccaaatcac tgggacaact gtgggaaggc tgggaaggtt aagaaacaac agaggtggac 3300  
ctccaaaaac atagaggcat cacctgactg cacaggcaat gaaaaacat gtgggtgatt 3360  
55 tccagcagac ctgtgctatt ggccaggagg cctgagaaag caagcacgca ctctcagtca 3420  
acatgacaga ttctggagga taaccagcag gagcagagat aacttcagga agtccatctt 3480  
tgcccctgct tttgctttgg attatacctc accagctgca caaaatgcat tttttcgtat 3540  
60 caaaaagtca cactaacc cccccagaa gctcacaag gaaaacggag agagcgagcg 3600  
agagagattt ccttggaaat ttctccaag ggcgaaagtc attggaattt ttaaatcata 3660  
ggggaaaagc agtctgttc taaatcctct tattcttttg gtttgtcaca aagaaggaac 3720  
65 taagaagcag gacagaggca acgtggagag gctgaaaaca gtgcagagac gtttgacaat 3780  
gagtcagtag cacaaaagag atgacattta cctagcatat aaaccctggt tgcctctgaa 3840  
gaaactgcct toattgtata tatgtgacta tttacatgta atcaacatgg gaacttttag 3900

ES 2 555 056 T3

5 gggaacctaa taagaaatcc caattttcag gagtgggtgggt gtcaataaac gctctgtggc 3960  
 cagtgtaaaa gaaaaaaaaa aaaaattgtg gacattttctg ttctgtcca gataccattt 4020  
 10 ctctagtagt ttctttgtta tgtcccagaa ctgatgtttt ttttttaagg tactgaaaag 4080  
 aaatgaagtt gatgtatgtc ccaagttttg atgaaactgt atttgtaaaa aaaattttgt 4140  
 agtttaagta ttgtcataca gtgttcaaaa cccagccaa tgaccagcag ttggtatgaa 4200  
 15 gaacctttga ctttttgtaa aaggccattt cttggggaaa aaaaaaaaaa aaaaaaaaaa 4260  
 aaaaaaaaaa aaaaaaaaaa 4279

20 <210> 31  
 <211> 870  
 <212> PRT  
 <213> Homo sapiens

25 <400> 31

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

ES 2 555 056 T3

5 Pro His Gly Pro Glu Asp Ser Ala Pro Gln Tyr Ser Arg Leu Phe Pro  
225 230 235 240

10 Asn Ala Ser Gln His Ile Thr Pro Ser Tyr Asn Tyr Ala Pro Asn Pro  
245 250 255

15 Asp Lys His Trp Ile Met Arg Tyr Thr Gly Pro Met Lys Pro Ile His  
260 265 270

20 Met Glu Phe Thr Asn Met Leu Gln Arg Lys Arg Leu Gln Thr Leu Met  
275 280 285

25 Ser Val Asp Asp Ser Met Glu Thr Ile Tyr Asn Met Leu Val Glu Thr  
290 295 300

30 Gly Glu Leu Asp Asn Thr Tyr Ile Val Tyr Thr Ala Asp His Gly Tyr  
305 310 315 320

35 His Ile Gly Gln Phe Gly Leu Val Lys Gly Lys Ser Met Pro Tyr Glu  
325 330 335

40 Phe Asp Ile Arg Val Pro Phe Tyr Val Arg Gly Pro Asn Val Glu Ala  
340 345 350

45 Gly Cys Leu Asn Pro His Ile Val Leu Asn Ile Asp Leu Ala Pro Thr  
355 360 365

50 Ile Leu Asp Ile Ala Gly Leu Asp Ile Pro Ala Asp Met Asp Gly Lys  
370 375 380

55 Ser Ile Leu Lys Leu Leu Asp Thr Glu Arg Pro Val Asn Arg Phe His  
385 390 395 400

60 Leu Lys Lys Lys Met Arg Val Trp Arg Asp Ser Phe Leu Val Glu Arg  
405 410 415

65 Gly Lys Leu Leu His Lys Arg Asp Asn Asp Lys Val Asp Ala Gln Glu  
420 425 430

70 Glu Asn Phe Leu Pro Lys Tyr Gln Arg Val Lys Asp Leu Cys Gln Arg  
435 440 445

75 Ala Glu Tyr Gln Thr Ala Cys Glu Gln Leu Gly Gln Lys Trp Gln Cys  
450 455 460

80 Val Glu Asp Ala Thr Gly Lys Leu Lys Leu His Lys Cys Lys Gly Pro  
465 470 475 480

85 Met Arg Leu Gly Gly Ser Arg Ala Leu Ser Asn Leu Val Pro Lys Tyr  
485 490 495

90 Tyr Gly Gln Gly Ser Glu Ala Cys Thr Cys Asp Ser Gly Asp Tyr Lys  
500 505 510

95 Leu Ser Leu Ala Gly Arg Arg Lys Lys Leu Phe Lys Lys Lys Tyr Lys  
515 520 525

100 Ala Ser Tyr Val Arg Ser Arg Ser Ile Arg Ser Val Ala Ile Glu Val  
530 535 540

ES 2 555 056 T3

5 Asp Gly Arg Val Tyr His Val Gly Leu Gly Asp Ala Ala Gln Pro Arg  
545 550 555 560

10 Asn Leu Thr Lys Arg His Trp Pro Gly Ala Pro Glu Asp Gln Asp Asp  
565 570 575

15 Lys Asp Gly Gly Asp Phe Ser Gly Thr Gly Gly Leu Pro Asp Tyr Ser  
580 585 590

Ala Ala Asn Pro Ile Lys Val Thr His Arg Cys Tyr Ile Leu Glu Asn  
595 600 605

20 Asp Thr Val Gln Cys Asp Leu Asp Leu Tyr Lys Ser Leu Gln Ala Trp  
610 615 620

25 Lys Asp His Lys Leu His Ile Asp His Glu Ile Glu Thr Leu Gln Asn  
625 630 635 640

Lys Ile Lys Asn Leu Arg Glu Val Arg Gly His Leu Lys Lys Lys Arg  
645 650 655

30 Pro Glu Glu Cys Asp Cys His Lys Ile Ser Tyr His Thr Gln His Lys  
660 665 670

35 Gly Arg Leu Lys His Arg Gly Ser Ser Leu His Pro Phe Arg Lys Gly  
675 680 685

Leu Gln Glu Lys Asp Lys Val Trp Leu Leu Arg Glu Gln Lys Arg Lys  
690 695 700

40 Lys Lys Leu Arg Lys Leu Leu Lys Arg Leu Gln Asn Asn Asp Thr Cys  
705 710 715 720

45 Ser Met Pro Gly Leu Thr Cys Phe Thr His Asp Asn Gln His Trp Gln  
725 730 735

Thr Ala Pro Phe Trp Thr Leu Gly Pro Phe Cys Ala Cys Thr Ser Ala  
740 745 750

50 Asn Asn Asn Thr Tyr Trp Cys Met Arg Thr Ile Asn Glu Thr His Asn  
755 760 765

55 Phe Leu Phe Cys Glu Phe Ala Thr Gly Phe Leu Glu Tyr Phe Asp Leu  
770 775 780

60 Asn Thr Asp Pro Tyr Gln Leu Met Asn Ala Val Asn Thr Leu Asp Arg  
785 790 795 800

Asp Val Leu Asn Gln Leu His Val Gln Leu Met Glu Leu Arg Ser Cys  
805 810 815

65 Lys Gly Tyr Lys Gln Cys Asn Pro Arg Thr Arg Asn Met Asp Leu Gly  
820 825 830

Leu Lys Asp Gly Gly Ser Tyr Glu Gln Tyr Arg Gln Phe Gln Arg Arg  
835 840 845

5

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

10

Trp Glu Gly Trp Glu Gly  
865 870

15

<210> 32  
<211> 6  
<212> PRT  
<213> Homo sapiens

20

<220>  
<221> VARIANTE  
<222> (1)..(1)  
<223> Leu OR Val

25

<220>  
<221> misc\_característica  
<222> (1)..(3)  
<223> Xaa puede ser cualquier aminoácido de origen natural

30

<220>  
<221> VARIANTE  
<222> (2)..(2)  
<223> Cys OR Ser

35

<220>  
<221> VARIANTE  
<222> (3)..(3)  
<223> Cualquier Aminoácido

40

<400> 32

Xaa Xaa Xaa Pro Ser Arg  
1 5

45

<210> 33  
<211> 23  
<212> PRT  
<213> Artificial

50

<220>  
<223> Secuencia derivada de humano Arilsulfatasa A

55

<220>  
<221> PEPTIDO  
<222> (1)..(23)  
<223> sustrato sintético formación FGly; secuencia primaria de humano Arilsulfatasa A

<400> 33

60

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

65

Ala Leu Leu Thr Gly Arg Ser  
20

5 <210> 34  
 <211> 16  
 <212> PRT  
 <213> Artificial

10 <220>  
 <223> una VARIANTEe del péptido ASA65-80, en las que los residuos Cys69, Pro71 y Arg73, crítica para la formación de FGly,fueron revueltos  
 <220>  
 <221> MISC\_CHARACTERÍSTICA

15 <222> (1)..(16)  
 <223> oligopéptido revueltos

20 <400> 34

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

25 <210> 35  
 <211> 16  
 <212> PRT  
 <213> Artificial

30 <220>  
 <223> una VARIANTEe del péptido ASA65-80, en el que la Cys69 fue sustituida por una serina  
 <220>  
 <221> MISC\_CHARACTERÍSTICA

35 <222> (1)..(16)  
 <223> Ser69 oligopeptido

40 <400> 35

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

45 <210> 36  
 <211> 19  
 <212> ADN  
 <213> Artificial

50 <220>  
 <223> humano FGE-especifico PCR imprimación

55 <220>  
 <221> misc\_característica  
 <222> (1)..(19)  
 <223> humano FGE-especifico PCR imprimación 1199nc

60 <400> 36  
 ccaatgtagg tcagacacg 19

65 <210> 37  
 <211> 16  
 <212> ADN  
 <213> Artificial

<220>



5 <223> humano FGE-especifico PCR imprimación  
 <220>  
 <221> misc\_característica  
 <222> (1)..(16)  
 10 <223> humano FGE-especifico avazadoPCR imprimacióno 1c  
 <400> 37  
 acatggcccg cgggac 16  
 15 <210> 38  
 <211> 19  
 <212> ADN  
 <213> Artificial  
 20 <220>  
 <223> humano FGE-especifico PCR imprimacióno  
 <220>  
 <221> misc\_característica  
 25 <222> (1)..(19)  
 <223> humano FGE-especifico reverso PCR imprimacióno 1182c  
 <400> 38  
 cgactgctcc ttggactgg 19  
 30 <210> 39  
 <211> 24  
 <212> ADN  
 <213> Artificial  
 35 <220>  
 <223> humano FGE-especifico PCR imprimacióno  
 <220>  
 40 <221> misc\_característica  
 <222> (1)..(24)  
 <223> humano 5' -FGE-especifico PCR imprimacióno conteniendo EcoRI  
 <400> 39  
 45 ggaattcggg acaacatggc tgcg 24  
 <210> 40  
 <211> 54  
 <212> ADN  
 50 <213> Artificial  
 <220>  
 <223> HA-especifico imprimacióno  
 55 <220>  
 <221> misc\_característica  
 <222> (1)..(54)  
 <223> HA-especifico imprimación  
 60 <400> 40  
 cccaagctta tgcgtagtca ggcacatcat acggatagtc catggtgggc aggc 54

65

5 <210> 41  
 <211> 57  
 <212> ADN  
 <213> Artificial

10 <220>  
 <223> imprimación específico de c-myc

<220>  
 <221> misc\_característica

15 <222> (1)..(57)  
 <223> imprimación específico de c-myc

<400> 41  
 cccaagctta caggctctt tcagaaatca gcttttgctc gtccatggcg ggcaggc 57

20 <210> 42  
 <211> 54  
 <212> ADN  
 <213> Artificial

25 <220>  
 <223> RGS-His6 - imprimación específico

<220>  
 <221> misc\_característica

30 <222> (1)..(54)  
 <223> RGS-His6 - imprimación específico

<400> 42  
 cccaagctta gtgatggga tggatgatgc atcctctgct catggtgggc aggc 54

35 <210> 43  
 <211> 15  
 <212> PRT

40 <213> Artificial

<220>  
 <223> oligopéptido tríptico partir de una preparación FGE humana

45 <220>  
 <221> MISC\_CHARACTERÍSTICA  
 <222> (1)..(15)  
 <223> oligopéptido tríptico partir de una preparación FGE humana

50 <400> 43

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

55 <210> 44  
 <211> 19  
 <212> PRT  
 <213> Artificial

60 <220>  
 <223> oligopéptido tríptico partir de una preparación FGE humana

<220>  
 <221> MISC\_CHARACTERÍSTICA

65 <222> (1)..(19)



ES 2 555 056 T3

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

ES 2 555 056 T3

5 <210> 47  
 <211> 927  
 <212> ADN  
 <213> Mus musculus

10 <400> 47

```

atgCGctctg agttctgggt cccCagcatg ggTtccttgc tccctccggT gttgctgctg      60
aggctcctgt cctgccccag gcttcagcta ggacatgcc aggatcctgc catggtgcat      120
ctgccaggTg gccggtttct gatggggaca gacgctccag atggcagaga cggTgaaggg      180
cctgcccggg aagtGacagt aaaacccttt gccatcgaca tatttccagt caccaataaa      240
gacttcaggg agtttGtcag ggagaagaag taccagactg aagccgaggc attcgggtgg      300
agcttcgtct ttgaggattt tgtctcccct gagctcagaa agcaagaaaa tctgatgccg      360
gctgttCact ggtggcagcc agtgccaaag gcattttgga ggcagcctgc aggtcccggc      420
tctggcatcc gagagaaact ggagcttccc gtggtacacg tgagctggaa cgacgctggT      480
gcttactgCg catggcgggg gagacgcttg cccacagaag aggagtggga gtttgcagcc      540
cgagggggct tgaagggtca ggtttatcca tgggggaacc ggttccagcc aaaccgcacc      600
aacttatggc agggaaagtt ccccaaaggt gacaaagctg aagatggttt tcatggactg      660
tcaccagtga acgctttccc cccacagaac aactacggac tgtatgacct catgggcaat      720
gtgtgggagt ggacagcgtc cacataccaa cctgctggcc aggacatgcg tgtcctccgg      780
ggggcatcat ggatcgacac cgcagacggc tctgctaatc acagggctcg ggtcaccacc      840
aggatgggaa aactccaga ctCagcctca gacaacctgg gcttccgctg cgctccagt      900
gcaggccgac cgaaggagga cctgtga                                          927
  
```

45 <210> 48  
 <211> 308  
 <212> PRT  
 <213> Mus musculus

50 <400> 48

```

Met Arg Ser Glu Phe Trp Phe Pro Ser Met Gly Ser Leu Leu Pro Pro
 1          5          10
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
  
```

ES 2 555 056 T3

5 Ala Phe Gly Trp Ser Phe Val Phe Glu Asp Phe Val Ser Pro Glu Leu  
100 105 110

10 Arg Lys Gln Glu Asn Leu Met Pro Ala Val His Trp Trp Gln Pro Val  
115 120 125

15 Pro Lys Ala Phe Trp Arg Gln Pro Ala Gly Pro Gly Ser Gly Ile Arg  
130 135 140

20 Glu Lys Leu Glu Leu Pro Val Val His Val Ser Trp Asn Asp Ala Gly  
145 150 155 160

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

30 Glu Phe Ala Ala Arg Gly Gly Leu Lys Gly Gln Val Tyr Pro Trp Gly  
180 185 190

35 Asn Arg Phe Gln Pro Asn Arg Thr Asn Leu Trp Gln Gly Lys Phe Pro  
195 200 205

40 Lys Gly Asp Lys Ala Glu Asp Gly Phe His Gly Leu Ser Pro Val Asn  
210 215 220

45 Ala Phe Pro Pro Gln Asn Asn Tyr Gly Leu Tyr Asp Leu Met Gly Asn  
225 230 235 240

50 Val Trp Glu Trp Thr Ala Ser Thr Tyr Gln Pro Ala Gly Gln Asp Met  
245 250 255

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

60 Asn His Arg Ala Arg Val Thr Thr Arg Met Gly Asn Thr Pro Asp Ser  
275 280 285

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

Lys Glu Asp Leu  
305

<210> 49  
<211> 855  
<212> ADN  
<213> Mus musculus

<400> 49

ES 2 555 056 T3

5 atggtcccca ttcoctgctgg agtattcaca atgggcactg atgatoctca gatcaggcag 60  
gatggagaag cccctgccag gagagtcaact gttgatggct ttacatgga cgcctatgaa 120  
10 gtcagcaatg cggatthtga gaagthtgtg aactcgactg gctatthtga agaggctgag 180  
aagthtggag actctthtct cthtgaaggc atgthtggcg agcaagtga aacgcataac 240  
caccaggcag ttgcagctgc tccatggtgg ttgcctgtca agggagctaa ttggagacac 300  
15 ccagaggggc cggactccag tathctgcac aggtcaaatc atccggthct ccagthtcc 360  
tggaacgatg ctgthtgcctc ctgcacatgg gcgggcaaga ggtthtctac tgaggcagag 420  
20 tgggaataca gctgtagagg aggcctgcag aacaggctth tcccctgggg caacaaactg 480  
cagcccaaag gacagcatta tgccaacac ttggcaggga agthtctctg gagcaaacact 540  
25 ggcgaggatg gctthcaagg aactgcccc gttgatgcct ttcoctccaa tggctatggc 600  
ttatacaaca tagtggggaa tgtgtgggag tggacctcag actggtggac tgttcaccat 660  
tctgthtggg aaacgttcaa ccaaagggg cccactthct ggaaagaccg agtgaagaag 720  
30 ggtggatcct acatgtgcca taagtcctat tgcctatagg accgctgtgc agctcgaagc 780  
cagaacacac cagatagctc tgcacccaac ctgggathct gatgtgcagc cgaccacctg 840  
35 cccaccgcag actga 855

<210> 50  
<211> 284  
40 <212> PRT  
<213> Mus musculus

<400> 50

45

50

55

60

65

ES 2 555 056 T3

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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



ES 2 555 056 T3

5 <210> 51  
 <211> 1011  
 <212> ADN  
 <213> Drosophila melanogaster

10 <400> 51

```

atgacaacaa ttatattagt cctctttatt tggatagttt tattcaatga cgtatccagc      60
gactgtggct gccaaaagct cgaccggaag gccccggata tgccgtccat ttccggacaa      120
gtgtgccagc aacgagcaca ggggtgcacac agccactacc gggattacta tggcgaactg      180
gagccaaata ttgcggacat gtcactgctt cggggaggca cggtttacat gggtaactgac      240
aaaccgcact ttccggccga ccgcgaggct ccggaacggc aggtgaagct gaatgacttc      300
tacatcgaca agtatgaggt ttccaacgaa gcctttgcca agtttgttct gcacactaac      360
tacaccaggg aggctgagcg atatggcgac agttttctgt ttaagagcct tttgagccca      420
ttggagcaga agaacctaga ggacttccga gtggcgagcg ctgtctggtg gtacaaagtg      480
gccggcgtga actggcgaca tccaaatggc gtggacagcg atatagacca cttaggccga      540
cacccggtag tgcacgtatc gtggcgcgac gctgtggagt actgtaagtg ggccggcaag      600
cggttgccca gcgaggcgga gtgggaggcg gcttgcaggg gcggcaagga gcgcaaactg      660
tttccctggg gcaacaagct gatgccaaag aatgaacatt ggctgaacat ctggcagggga      720
gactttcccg atggcaacct ggctgaagat gggtttgagt acaccagccc cgtggatgcc      780
ttccgacaga atatttacga cctgcacaac atgggtgggca acgtctggga gtggacggca      840
gatctgtggg acgtaaata gaattagcag aatccaaatc gggtaagaa gggcggttct      900
tatctgtgtc acaagtccta ctgctacagg tacaggtgcg cggcacgctc gcagaacaca      960
gaagacagtt cagccggtaa cctgggtttt cgggtgcgca agaatgctg a              1011
  
```

45 <210> 52  
 <211> 336  
 <212> PRT  
 <213> Drosophila melanogaster

50 <400> 52

```

Met Thr Thr Ile Ile Leu Val Leu Phe Ile Trp Ile Val Leu Phe Asn
55 1           5           10           15

Asp Val Ser Ser Asp Cys Gly Cys Gln Lys Leu Asp Arg Lys Ala Pro
60           20           25           30

Asp Met Pro Ser Ile Ser Gly Gln Val Cys Gln Gln Arg Ala Gln Gly
65           35           40           45

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

ES 2 555 056 T3

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

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

ES 2 555 056 T3

5 <210> 53  
 <211> 870  
 <212> ADN  
 <213> Anopheles gambiae

10 <400> 53

ccggagagct tgctcgatct ggtggaacat tccaagcggc tcgaagacat gaggccttatc 60

15 ccaggagggtg aatatgtaat cggcacaaat gaacctatct tcgtcaagga tcgcgaatca 120

ccggcccggc ccgcgacgat ccgcgacttt tacctcgacc agtacgaagt ctccaacgca 180

20 cagttcaagg cattcgtcga ccagacgggc tacgtcacgg aggcggaaaa gtttggcgac 240

agcttctgtct tccagcagct gctcagcgaa ccgggtgcgc agcagtacga agatttccgc 300

gtggcggcgg cgcctctggtg gtacaaggta cgtggagcct cctggcagca tccggaaggc 360

25 gatgtgtcac gtgatataag cgaccgattg gaccatccgg tgggtgcacgt gtcctggaac 420

gatgcggctc cgtactgcgc ctggaaaggc aagcgcctgc cgacggaagc ggaatgggaa 480

gcccctgcc gggcggtcgc caagcagaag ctgttcccct ggggtaacaa gctgatgccg 540

30 aaggagcagc acatgatgaa catatggcag ggcgagttcc cggacagcaa tctgaaggag 600

gatggctacg agaccacctg cccggtgacg tccttccgcc agaaccctt cgagctgtac 660

35 aacatcgttg gcaacgtgtg ggagtggacg gcggatcttt gggacgcgaa ggatgocggc 720

atcgagcgca agccgggcag cgatccaccg aatcgggtga aaaagggtgg ctcatacctg 780

tgtcacgaat cgtactgcta tcgctatcgc tgtcgggctc gatcgcagaa caccgaggac 840

40 agttcggcgg gcaatctggg ctcccggtgc 870

45 <210> 54  
 <211> 290  
 <212> PRT  
 <213> Anopheles gambiae

50 <400> 54

55

60

65

ES 2 555 056 T3

5 Pro Glu Ser Leu Leu Asp Leu Val Glu His Ser Lys Arg Phe Glu Asp  
1 5 10 15

10 Met Ser Leu Ile Pro Gly Gly Glu Tyr Val Ile Gly Thr Asn Glu Pro  
20 25 30

Ile Phe Val Lys Asp Arg Glu Ser Pro Ala Arg Pro Ala Thr Ile Arg  
35 40 45

15 Asp Phe Tyr Leu Asp Gln Tyr Glu Val Ser Asn Ala Gln Phe Lys Ala  
50 55 60

20 Phe Val Asp Gln Thr Gly Tyr Val Thr Glu Ala Glu Lys Phe Gly Asp  
65 70 75 80

Ser Phe Val Phe Gln Gln Leu Leu Ser Glu Pro Val Arg Gln Gln Tyr  
85 90 95

25 Glu Asp Phe Arg Val Ala Ala Ala Pro Trp Trp Tyr Lys Val Arg Gly  
100 105 110

30 Ala Ser Trp Gln His Pro Glu Gly Asp Val Ser Arg Asp Ile Ser Asp  
115 120 125

Arg Leu Asp His Pro Val Val His Val Ser Trp Asn Asp Ala Val Ala  
130 135 140

35 Tyr Cys Ala Trp Lys Gly Lys Arg Leu Pro Thr Glu Ala Glu Trp Glu  
145 150 155 160

40 Ala Ala Cys Arg Gly Gly Arg Lys Gln Lys Leu Phe Pro Trp Gly Asn  
165 170 175

45 Lys Leu Met Pro Lys Glu Gln His Met Met Asn Ile Trp Gln Gly Glu  
180 185 190

50 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

55 Asn Val Trp Glu Trp Thr Ala Asp Leu Trp Asp Ala Lys Asp Ala Ala  
225 230 235 240

60 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

65 Ala Arg Ser Gln Asn Thr Glu Asp Ser Ser Ala Gly Asn Leu Gly Phe  
275 280 285

Arg Cys  
290

ES 2 555 056 T3

5 <210> 55  
 <211> 945  
 <212> ADN  
 <213> Streptomyces coelicolor

10 <400> 55

	gtggcctgg cgcgccgctc ccccgccggc gccgcggagc cggggcccg cgcctgctcg	60
15	cgctcgaccc ggcggacaggt ggcctgccc ggcggtgagt tcgcatggg ggacgccttc	120
	ggggagggat atccggccga cggcgagaca cccgtgcaca cggtgccct gcggcccttc	180
	cacatcgacg agaccgccc caccacgcc cggctcggc ccttcgtcaa ggcgaccggc	240
20	catgtgaccg acgcccgaac ctccggctcc tcggccgtct tccacctggc cgtcgcggcc	300
	cgggacggc acgtcctcgg cagcgcggcc ggcgccccct ggtggatcaa cgtcgggggc	360
	gcccactggc gccgccccga gggcgcccgc tccgacatca ccggccggcc gaacctccg	420
25	gtcgtccacg tctcctggaa cgatgccacc gcctacggc ggtgggcccg caagcgctg	480
	cccaccgagg ccgaatggga gtacggccc cgcgggggac tggccggccg ccgctacgcc	540
30	tggggcgacg agctgacccc gggcgcccg gggcgctgca acatctggca gggccgcttc	600
	ccgcacgtca acaaggccga ggacgggac ctgagcaccg cacgggtcaa gtcctaccgg	660
	cccaacggcc acggcctgtg gaacaccggc ggcaacgtgt gggaatggtg ctccgactgg	720
35	ttctcgccc cctactacgc cgaatcacc accgtcgacc cgcacggccc cgggaccggg	780
	gggcacggg tgctgcggc cggctcctac ctgtgccacg actcctactg caaccgctac	840
	cgggtcggc cccgctcctc caacaccgg gactcctgt ccggcaacct cggattccgc	900
40	tgcgccaacg acgaggacct cacgtccgga tcagccgctg agtga	945

45 <210> 56  
 <211> 314  
 <212> PRT  
 <213> Streptomyces coelicolor

50 <400> 56 ,

55

60

65

ES 2 555 056 T3

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

Ala Ala Arg Pro Arg Ser Thr Arg Gly Gln Val Arg Leu Pro Gly Gly  
20 25 30 10

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 15

Thr Ala Val Thr Asn Ala Arg Phe Ala Ala Phe Val Lys Ala Thr Gly  
65 70 75 80 20

His Val Thr Asp Ala Glu Arg Phe Gly Ser Ser Ala Val Phe His Leu  
85 90 95 20

Val Val Ala Ala Pro Asp Ala Asp Val Leu Gly Ser Ala Ala Gly Ala  
100 105 110 25

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 30

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 40

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 50

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 55

Pro Gly Thr Gly Ala Ala Arg Val Leu Arg Gly Gly Ser Tyr Leu Cys  
260 265 270 60

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 65

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

ES 2 555 056 T3

5 <210> 57  
 <211> 1005  
 <212> ADN  
 <213> Corynebacterium efficiens

10 <400> 57

```

gtggttcgcc atcgactggg ccaccggccc tgcacactga ggattacgtc catgagtaac      60
15  tgctgctccc cgtcaagegc acaatggcgt accactacce gggatttata agatcctgtc      120
    aatcccacca ctccatgcaa cccggaacaa tcccgcgatg ctgtgacact gccgggtgga      180
    gctttccaca tgggcgatca tcacggggag gggatcccgg cggacgggga ggggccagta      240
20  catgaggttc acctcgcccc cttcggcatt aatgtcacca cggtcacgaa tgccgagttc      300
    ggacgattta ttgaagccac agggatatac acgacagcgg aacgctacgg tgtctcggct      360
    gtattctacg cagcgttcca agggcaacgc gctgacatlc ttccgccagg tcccggcgtg      420
25  ccctgggtggc tggcgggtcaa ggggtgcgaac tggcagcgtc ccaacggccc cggatccacc      480
    ctggacgggc ttgaggacca ccccgctcgtt cacgtttctt gggatgatgc cgttgcctac      540
30  tgcacctggg ctggcgggtcg totgcccacc gaagccgagt ggggaatacgc cgcccgggggt      600
    ggactgcagg gcgcacgata tgccctggggg gataacctcg ccctagacgg gaggtggaac      660
    tgcaatatct ggcagggggg ctcccccatg gagaacaccg ccgcggatgg ttacctcacc      720
35  actgcaccgg tgaagacctc cacgcaccaat ggatacggtc tgtggcagat ggcaggggat      780
    gtatggggaat ggtgccagga ctggtttgat gcgagtagct actcccgtgc ttctctcctc      840
40  aaccgcgggg gaccggatac cgggtgcgcgc cgggtgatgc ggggaggctc gtatctctgc      900
    catgattcct actgcaacag ataccgggtg gccgcccgca attcgaacac cccggattcc      960
    acctcgggga ataccggttt cgggtgcggtt ttccgatagtc cttga                          1005

```

45 <210> 58  
 <211> 334  
 <212> PRT  
 <213> Corynebacterium efficiens

50 <400> 58

55

60

65

5 Met Val Arg His Arg Leu Gly His Arg Pro Cys Thr Leu Arg Ile Thr  
1 5 10

10 Ser Met Ser Asn Cys Cys Ser Pro Ser Ser Ala Gln Trp Arg Thr Thr  
20 25 30

15 Thr Arg Asp Leu Ser Asp Pro Val Asn Pro Thr Thr Pro Cys Asn Pro  
35 40 45

20 Glu Gln Ser Arg Asp Ala Val Thr Leu Pro Gly Gly Ala Phe His Met  
50 55 60

25 Gly Asp His His Gly Glu Gly Tyr Pro Ala Asp Gly Glu Gly Pro Val  
65 70 75 80

30 His Glu Val His Leu Ala Pro Phe Gly Ile Asn Val Thr Thr Val Thr  
85 90 95

35 Asn Ala Glu Phe Gly Arg Phe Ile Glu Ala Thr Gly Tyr Thr Thr Thr

40

45

50

55

60

65



5  
100 105 110

Ala Glu Arg Tyr Gly Val Ser Ala Val Phe Tyr Ala Ala Phe Gln Gly  
115 120 125

Gln Arg Ala Asp Ile Leu Arg Gln Val Pro Gly Val Pro Trp Trp Leu  
130 135 140

Ala Val Lys Gly Ala Asn Trp Gln Arg Pro Asn Gly Pro Gly Ser Thr  
145 150 155 160

Leu Asp Gly Leu Glu Asp His Pro Val Val His Val Ser Trp Asp Asp  
165 170 175

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

Glu Trp Glu Tyr Ala Ala Arg Gly Gly Leu Gln Gly Ala Arg Tyr Ala  
195 200 205

Trp Gly Asp Asn Leu Ala Leu Asp Gly Arg Trp Asn Cys Asn Ile Trp  
210 215 220

Gln Gly Gly Phe Pro Met Glu Asn Thr Ala Ala Asp Gly Tyr Leu Thr  
225 230 235 240

Thr Ala Pro Val Lys Thr Tyr Thr Pro Asn Gly Tyr Gly Leu Trp Gln  
245 250 255

Met Ala Gly Asn Val Trp Glu Trp Cys Gln Asp Trp Phe Asp Ala Glu  
260 265 270

Tyr Tyr Ser Arg Ala Ser Ser Ile Asn Pro Arg Gly Pro Asp Thr Gly  
275 280 285

Ala Arg Arg Val Met Arg Gly Gly Ser Tyr Leu Cys His Asp Ser Tyr  
290 295 300

Cys Asn Arg Tyr Arg Val Ala Ala Arg Asn Ser Asn Thr Pro Asp Ser  
305 310 315 320

Thr Ser Gly Asn Thr Gly Phe Arg Cys Val Phe Asp Ser Pro  
325 330

65 <210> 59  
<211> 1017  
<212> ADN

ES 2 555 056 T3

<213> Novosphingobium aromaticivorans

<400> 59

5 atgggogcaac cattccgatc gacggcggcc agtcgtacaa gtattgaacg ccatctcgaa 60  
 cccaattgca ggagcacgtc gcgaatggtc gaacgccccg gcatgcgcct gatcgaaggc 120  
 10 ggcactttca ccatgggctc ggaagccttc taccgggagg aagcgccgct tcgcccgggtg 180  
 aaggtagaca gcttctggat cgatgaagcg ccggtgacga acgcacagtt cgcgcgcatc 240  
 gtggaggcca cgggatacgt cactgtggcc gagatcgagc cggatcccaa ggactacccc 300  
 15 ggcattgctcc cgggcatgga ccgcgcggga tcgctggtgt tccagaaaac agcagggccg 360  
 gtcgacatgg cggatgcgtc caactggtgg cactttacct ttggcgctg ctggaagcat 420  
 20 ccacttggac cgggcagttc catcgatggg atcgaggacc atcccgtcgt tcacgtcgcc 480  
 tatgccgatg ccgaggccta tgccaaatgg gcgggcaagg atctgccgac cgaagccgag 540  
 25 ttcgaatatg ctgcgcgcgg cgggttggac ggttccgaat tttctgggg agacgaactc 600  
 gcacctgaag gccggatgat ggccaactac tggcaaggcc tgtttccctt cgccaaccag 660  
 tgcctcgatg gctgggaacg gacatcgccc gtccgcaact tcccgccaa cggctatggt 720  
 30 ctttacgaca tgatcgggaa cacgtgggag tggacctgcg attggtgggc cgacaagccg 780  
 ctgactccgc aaaggaaatc ggcattgctgc gcgatcagca atccgcgcgg cggcaagctc 840  
 aaggacagct tcgaccgctc gcaaccgca atgcgcatcg gccggaaggc cataaagggc 900  
 35 ggttcgcacc tgtgtgcggc caattactgc cagecctatc gccccgcagc acgccatcct 960  
 gaaatggttg ataccgcgac gacgcacatc ggcttcaggc gtgtggtgcg gccctga 1017

40 <210> 60  
 <211> 338  
 <212> PRT  
 <213> Novosphingobium aromaticivorans  
 45 <400> 60

50

55

60

65

5 Met Ala Gln Pro Phe Arg Ser Thr Ala Ala Ser Arg Thr Ser Ile Glu  
1 5 10

10 Arg His Leu Glu Pro Asn Cys Arg Ser Thr Ser Arg Met Val Glu Arg  
20 25 30

15 Pro Gly Met Arg Leu Ile Glu Gly Gly Thr Phe Thr Met Gly Ser Glu  
35 40 45

20 Ala Phe Tyr Pro Glu Glu Ala Pro Leu Arg Arg Val Lys Val Asp Ser  
50 55 60

25 Phe Trp Ile Asp Glu Ala Pro Val Thr Asn Ala Gln Phe Ala Ala Phe  
65 70 75 80

30 Val Glu Ala Thr Gly Tyr Val Thr Val Ala Glu Ile Glu Pro Asp Pro  
85 90 95

35 Lys Asp Tyr Pro Gly Met Leu Pro Gly Met Asp Arg Ala Gly Ser Leu  
100 105 110

40 Val Phe Gln Lys Thr Ala Gly Pro Val Asp Met Ala Asp Ala Ser Asn  
115 120 125

45 Trp Trp His Phe Thr Phe Gly Ala Cys Trp Lys His Pro Leu Gly Pro  
130 135 140

50 Gly Ser Ser Ile Asp Gly Ile Glu Asp His Pro Val Val His Val Ala  
145 150 155 160

55 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  
 10 Asn Tyr Trp Gln Gly Leu Phe Pro Phe Ala Asn Gln Cys Leu Asp Gly  
 210 215 220  
 15 Trp Glu Arg Thr Ser Pro Val Arg Asn Phe Pro Pro Asn Gly Tyr Gly  
 225 230 235 240  
 20 Leu Tyr Asp Met Ile Gly Asn Thr Trp Glu Trp Thr Cys Asp Trp Trp  
 245 250 255  
 25 Ala Asp Lys Pro Leu Thr Pro Gln Arg Lys Ser Ala Cys Cys Ala Ile  
 260 265 270  
 30 Ser Asn Pro Arg Gly Gly Lys Leu Lys Asp Ser Phe Asp Pro Ser Gln  
 275 280 285  
 35 Pro Ala Met Arg Ile Gly Arg Lys Val Ile Lys Gly Gly Ser His Leu  
 290 295 300  
 40 Cys Ala Ala Asn Tyr Cys Gln Arg Tyr Arg Pro Ala Ala Arg His Pro  
 305 310 315 320  
 45 Glu Met Val Asp Thr Ala Thr Thr His Ile Gly Phe Arg Cys Val Val  
 325 330 335  
 50 Arg Pro

45 <210> 61  
 <211> 1119  
 <212> ADN  
 <213> Mesorhizobium loti  
 50 <400> 61

55  
 60  
 65

ES 2 555 056 T3

5 atgggcccac gaggtcgagg tcaaaaaccg catgaaaggc gacgcggtca tgttcgacat 60  
 tgccgggaag ttctagccga tagcgggtgg gggctgatg gagatgagca cgccgtgtca 120  
 10 ttccgggatc ttctgatgaa cgcccctgcc gaagtcttgc agcgcgctgc agccgaacgg 180  
 tcgtaccccg gaatggtctg gatccccggc ggtaccttc tgatgggctc agacaaccac 240  
 tatccggagg aggcaccggc ccaccgggtc agggctcgacg gcttctggat ggacaaatc 300  
 15 accgtctcca accgcgactt cgaacgcttc gttgcggcga caggacatgt cactcttgcc 360  
 gagaaacccg ccaatcccga cgactatccc ggtgccttac ccgatctgct ggctccgctc 420  
 tcgatgatgt tcaggaagcc ggcgggcoct gtcgaccttg gcaatcacta caattggtgg 480  
 20 gtctatgtcc gcggcgcca ctggcgccat ccacgcgggc cggcaagtac aatcaagaag 540  
 gttgcagatc atccggctgt gcatgtggcc tacgaggatg tcgtggccta tgccaactgg 600  
 gcaggcaagg aacttcccac cgaggccgag tgggaattcg cggcgcgagg cggcctcgat 660  
 25 gccgccgaat acgtctgggg caacgagctt acgcccggccg ggaagcacat ggccaacatc 720  
 tggcaaggag actttcccta ccggaatact gtcgacgacg gttacgaata tacggcccca 780  
 30 gtaggctcgt tcccggccaa cgactacggt ctctacgaca tggccggcaa tgtctggcaa 840  
 tggacgaccg actggtacca ggaccacaag gcgatcgaca gcccgctgct caccgctgtc 900  
 aatccgcgtg ggggccatcg cgaagcgagc tatgacaccc ggctacctga cgttaagatc 960  
 35 cctcgcaagg tcaccaaggg tggctcccat ctgtgcgcgc cgaactactg tcggogctac 1020  
 cgcccgcggg cgcgaatggc gcaaccgctc gacactgcaa tctcccatct cggctttcgc 1080  
 40 tgcacgtgct gaaggaaaat ggaattgaac gcgcagtaa 1119

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

<400> 62

50

55

60

65

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

10  
 His Val Arg His Cys Arg Glu Val Leu Ala Asp Ser Gly Trp Ala Ala  
 20 25 30

15  
 Asp Gly Asp Glu His Ala Val Ser Phe Arg Asp Leu Ser Met Asn Ala  
 35 40 45

20  
 Pro Ala Glu Val Phe Glu Arg Ala Ala Ala Glu Arg Ser Tyr Pro Gly  
 50 55 60

25  
 Met Val Trp Ile Pro Gly Gly Thr Phe Leu Met Gly Ser Asp Asn His  
 65 70 75 80

30  
 Tyr Pro Glu Glu Ala Pro Ala His Arg Val Arg Val Asp Gly Phe Trp  
 85 90 95

35  
 Met Asp Lys Phe Thr Val Ser Asn Arg Asp Phe Glu Arg Phe Val Ala  
 100 105 110

40  
 Ala Thr Gly His Val Thr Leu Ala Glu Lys Pro Ala Asn Pro Asp Asp  
 115 120 125

45  
 Tyr Pro Gly Ala Leu Pro Asp Leu Leu Ala Pro Ser Ser Met Met Phe  
 130 135 140

50  
 Arg Lys Pro Ala Gly Pro Val Asp Leu Gly Asn His Tyr Asn Trp Trp  
 145 150 155 160

55  
 Val Tyr Val Arg Gly Ala Asn Trp Arg His Pro Arg Gly Pro Ala Ser  
 165 170 175

60  
 Thr Ile Lys Lys Val Ala Asp His Pro Val Val His Val Ala Tyr Glu  
 180 185 190

65  
 Asp Val Val Ala Tyr Ala Asn Trp Ala Gly Lys Glu Leu Pro Thr Glu  
 195 200 205

70  
 Ala Glu Trp Glu Phe Ala Ala Arg Gly Gly Leu Asp Ala Ala Glu Tyr  
 210 215 220

75  
 Val Trp Gly Asn Glu Leu Thr Pro Ala Gly Lys His Met Ala Asn Ile  
 225 230 235 240

80  
 Trp Gln Gly Asp Phe Pro Tyr Arg Asn Thr Val Asp Asp Gly Tyr Glu  
 245 250 255

85  
 Tyr Thr Ala Pro Val Gly Ser Phe Pro Ala Asn Asp Tyr Gly Leu Tyr

5  
 260 265 270  
 10 Asp Met Ala Gly Asn Val Trp Gln Trp Thr Thr Asp Trp Tyr Gln Asp  
 275 280 285  
 15 His Lys Ala Ile Asp Ser Pro Cys Cys Thr Ala Val Asn Pro Arg Gly  
 290 300  
 20 Gly His Arg Glu Ala Ser Tyr Asp Thr Arg Leu Pro Asp Val Lys Ile  
 305 310 315 320  
 25 Pro Arg Lys Val Thr Lys Gly Gly Ser His Leu Cys Ala Pro Asn Tyr  
 325 330 335  
 30 Cys Arg Arg Tyr Arg Pro Ala Ala Arg Met Ala Gln Pro Val Asp Thr  
 340 345 350  
 35 Ala Ile Ser His Leu Gly Phe Arg Cys Ile Val Arg Arg Lys Met Glu  
 355 360 365  
 40 Leu Asn Ala Gln  
 370

35 <210> 63  
 <211> 1251  
 <212> ADN  
 <213> Burkholderia fungorum

40 <400> 63

45

50

55

60

65

ES 2 555 056 T3

5 atgaagagtg aaagagatcg agagcccgca aagtcgtccc gctcgaacgg gtcgggtcgca 60  
gcaaccctaaa cgcgcgcccg tgcggtgccc aaactaatgt tgtggggcgc cctgctcgtc 120  
10 atactgcccg cctgtgtcgg cgcgcgggtc agttgggcct tcacgcccga cgcacccgct 180  
caccgcgaaa tggttttcgg cgacggcagc catgggtccc tcggcatggc gtgggtgccc 240  
ggcggccagt tcctcatggg cagcgcagcc aaacaggcgc aaccgaacga acgcccgcg 300  
15 cacaaggtca aggtgcacgg cttctggatg gaccgcccac acgtgaccaa cgccgaattc 360  
cgccgcttcg togaagcgac cggctacgtc accacggccg agaagaaacc cgactgggag 420  
accctgaaag tccagttgcc gcccggcagc ccgcgcccgc ccgagagcgc gatggtggcg 480  
20 ggtgcaatgg tgttcgtcgg caccagccgt cccgtgcccgc tagacgacta ttcgcagtgg 540  
tggcgctatg tgcctggcgc taactggcgt catccagccg ggccctgagag caacatcacc 600  
25 ggtaaagatg atcaccocgt ggttcaagtg tctacgaag atgcgagggc ttatgcaaaa 660  
tgggcccggca agcgtctgcc gaccgaagcc gaatgggaat tcgcccgcgc cggcggcctc 720  
gaacaggcca cgtatgcgtg gggcgatcag ttctctccca acggcaaaca gatggccaac 780  
30 gtctggcagg gccagcagcc gcagtctttc cccgttgtca acccgaaagc ggggtggcgcg 840  
ctcgggtacaa gtcgggtggg tactttcccg gccaacggct acggcctttc cgacatgacc 900  
ggcaacgcct ggcagtgggg tgccgactgg tatcgcgcgg atcagttcag gcgtgaggcg 960  
35 gtaagcacca gcgcgatcga caatccggtg ggcccagagc agtcgtggga ccccgcagac 1020  
cagggcgtgc ccgtcaacgc gcccaagcgt gtcacacgcg gcggttcggt cctctgcaac 1080  
40 gaaatctatt gcctgagcta ccggcccagc gcgagacgcg gcaccgatcc ctacaacagc 1140  
atgtcgcacc tgggcttccg gctgggtgat gacgaagaca cctggaaaga agccggtgct 1200  
cgccaggctt cggcgaaagc tgccggcgcg cctggaacce ctggcggcta g 1251

45  
<210> 64  
<211> 416  
<212> PRT  
50 <213> Burkholderia fungorum  
<400> 64

55  
60  
65



ES 2 555 056 T3

5 Met Lys Ser Glu Arg Asp Arg Glu Pro Ala Lys Ser Ser Arg Ser Asn  
1 5 10

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

15 Ala Val Ser Trp Ala Phe Thr Pro His Ala Pro Ala His Pro Gln Ile  
50 55 60

20 Val Phe Gly Asp Gly Thr His Gly Pro Leu Gly Met Ala Trp Val Pro  
65 70 75 80

25 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

30 His His Val Thr Asn Ala Glu Phe Arg Arg Phe Val Glu Ala Thr Gly  
115 120 125

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

40 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

45 Tyr Ser Gln Trp Trp Arg Tyr Val Pro Gly Ala Asn Trp Arg His Pro  
180 185 190

50 Ala Gly Pro Glu Ser Asn Ile Ile Gly Lys Asp Asp His Pro Val Val  
195 200 205

55 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

60 Glu Gln Ala Thr Tyr Ala Trp Gly Asp Gln Phe Ser Pro Asn Gly Lys  
245 250 255

65 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



ES 2 555 056 T3

5 atgggtctggg ttccccggagc gacottcatg atgggggtcga acgaccatta cccggaggaa 60  
 ggcgccgtgc atccggtaac cgtcgacgga ttctggatcg atgtgacacc ggtaacgaac 120  
 10 cgccagtttc tcgaattcgt aaatgcgacg gggcatgtga ccttcgcgga aagaaagccg 180  
 cgcgccgaag actatccggg cgtcccgcca tccaatctaa gggccggttc gctcgtcttc 240  
 acacccccga agcgcaccgt gcaggggaacg gatatatcgc agtgggtggat attcacgctg 300  
 15 ggtgccaaact ggcggcacc cctcggggcg aagagcagca tcggagcgat tctggatcat 360  
 ccggtcgtcc atgtcgctta cagcgcgca aaggcctatg ccgaatgggc cggcaaggac 420  
 ctccccgaccg agaccgagtg ggagctggcg gcccgcgcg gctcgcgatgg ggctgaattt 480  
 20 tctggggcg gcgagcttgc gccggggcgga aatcacatgg ccaatacttg gcaggggaagt 540  
 tttccggctg agaattctat ggacgatggt ttgcgcgaa catcgccggt cagattttac 600  
 ccgccgaacg gctacggcct ctacgacatg atcggcaatg tgtgggagtg gaccacggat 660  
 tactggctcg tgcgccacc ggaagcggcc gccaaagcctt gctgcattcc gagcaatccc 720  
 cgcaatgccg atgccgatgc gagtatcgat ccggcggcga gcgtgaaagt tccgcgcggg 780  
 30 gtgctcaagg gtggatcgca tctctgcgcg ccgaactact gccggcggtta ccgccctgcg 840  
 gcgagggcacg cccaggaaat cgacacgacg accagccatg tcggtttccg atgtgtcagg 900  
 35 cgcgttcgat aa 912

<210> 66

<211> 303

<212> PRT

40 <213> Sinorhizobium meliloti

<400> 66

45

50

55

60

65

ES 2 555 056 T3

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

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

15 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

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

25 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

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

35 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

40 Ser Trp Gly Gly Glu Leu Ala Pro Gly Gly Asn His Met Ala Asn Thr  
165 170 175

45 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

50 Asp Met Ile Gly Asn Val Trp Glu Trp Thr Thr Asp Tyr Trp Ser Val  
210 215 220

55 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

60 Val Pro Arg Arg Val Leu Lys Gly Gly Ser His Leu Cys Ala Pro Asn  
260 265 270

65 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

ES 2 555 056 T3

5

<210> 67  
 <211> 1065  
 <212> ADN  
 <213> Microscilla sp.

10

<400> 67

	atgaaataca	tttttttagt	tcttttctta	tgggccttga	ccgatgtac	cggaaagtat	60
15	gaggacaaga	gagtggaaac	tgatacttcc	agaccaaag	ccgaagcgtc	agatataaaa	120
	gttcccgaag	gaatggctta	tattcccgcg	ggccagtaca	tgatgggagg	taaatacagac	180
	caggcttata	aggatgaata	tccccgccat	aacgtgaagg	tttcggcttt	ttatatggac	240
20	cttacagaag	tgaccaatgc	ggagtttaag	cggttttag	acgaaacggg	ctacgtgacc	300
	attgctgaga	aagatattga	ctgggaagag	ttaaagtctc	aggtgccaca	gggtaccccg	360
25	aagcctcctg	attctgtgct	tcaggcaggc	tcactggttt	tcaagcagac	agatgaacce	420
	gtttctctcc	aggattatc	acagtggctg	gaatggacta	tcggagccaa	ctggcgaaat	480
	ccggagggc	caggtagtac	gattgaggat	cgtatggatc	atccggtggt	acacgtttcc	540
30	tttgaagatg	tccaagcgta	tgccgattgg	gccggtaagc	gcctgcctac	tgaggcagaa	600
	tgggaatggg	ccgccatggg	aggccaaaat	gacgtgaaat	atccatgggg	aaatgaatcg	660
	gtcgaacaag	catccgataa	agcaaacttt	tggcagggga	attttccaca	tcaaaactat	720
35	gcctcgcgatg	gattogaacg	caccgcccct	gtacgctcct	tcccagcga	tgggtaacggc	780
	ctatatgata	tggctggcaa	tgtgtgggaa	tggtgccagg	ataagtatga	tgtcaatgct	840
40	tatgaaagct	ataagcaaaa	aggactgaca	gaagacccca	ggggttctga	gcactacaac	900
	gaccctaggg	aaccgtatac	tcctaagcat	gtgatcagag	ggggttcttt	cctatgcaat	960
	gacagctact	gtagtgggta	tcgtgtttca	cgctgatga	gttccagtag	agattcaggt	1020
45	tttaatcata	cgggattcag	gtgtgtgaaa	gatgtaaatg	gatag		1065

50

<210> 68  
 <211> 354  
 <212> PRT  
 <213> Microscilla sp.

<400> 68

55

60

65

5

Met Lys Tyr Ile Phe Leu Val Leu Phe Leu Trp Ala Leu Thr Arg Cys  
 1 5 10 15

10

Thr Gly Lys Tyr Glu Asp Lys Arg Val Glu Thr Asp Thr Ser Arg Pro  
 20 25 30

15

Lys Ala Glu Ala Ser Asp Ile Lys Val Pro Glu Gly Met Ala Tyr Ile  
 35 40 45

20

Pro Ala Gly Gln Tyr Met Met Gly Gly Lys Ser Asp Gln Ala Tyr Lys  
 50 55 60

25

Asp Glu Tyr Pro Arg His Asn Val Lys Val Ser Ala Phe Tyr Met Asp  
 65 70 75 80

30

Leu Thr Glu Val Thr Asn Ala Glu Phe Lys Arg Phe Val Asp Glu Thr  
 85 90 95

35

Gly Tyr Val Thr Ile Ala Glu Lys Asp Ile Asp Trp Glu Glu Leu Lys  
 100 105 110

40

45

50

55

60

65

5

Ser Gln Val Pro Gln Gly Thr Pro Lys Pro Pro Asp Ser Val Leu Gln  
 115 120 125

10

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

15

Asp Tyr Ser Gln Trp Trp Glu Trp Thr Ile Gly Ala Asn Trp Arg Asn  
 145 150 155 160

20

Pro Glu Gly Pro Gly Ser Thr Ile Glu Asp Arg Met Asp His Pro Val  
 165 170 175

25

Val His Val Ser Phe Glu Asp Val Gln Ala Tyr Ala Asp Trp Ala Gly  
 180 185 190

30

Lys Arg Leu Pro Thr Glu Ala Glu Trp Glu Trp Ala Ala Met Gly Gly  
 195 200 205

35

Gln Asn Asp Val Lys Tyr Pro Trp Gly Asn Glu Ser Val Glu Gln Ala  
 210 215 220

40

Ser Asp Lys Ala Asn Phe Trp Gln Gly Asn Phe Pro His Gln Asn Tyr  
 225 230 235 240

45

Ala Leu Asp Gly Phe Glu Arg Thr Ala Pro Val Arg Ser Phe Pro Ala  
 245 250 255

50

Asn Gly Tyr Gly Leu Tyr Asp Met Ala Gly Asn Val Trp Glu Trp Cys  
 260 265 270

55

Gln Asp Lys Tyr Asp Val Asn Ala Tyr Glu Ser Tyr Lys Gln Lys Gly  
 275 280 285

60

Leu Thr Glu Asp Pro Thr Gly Ser Glu His Tyr Asn Asp Pro Arg Glu  
 290 295 300

65

Pro Tyr Thr Pro Lys His Val Ile Arg Gly Gly Ser Phe Leu Cys Asn  
 305 310 315 320

Asp Ser Tyr Cys Ser Gly Tyr Arg Val Ser Arg Arg Met Ser Ser Ser  
 325 330 335

Arg Asp Ser Gly Phe Asn His Thr Gly Phe Arg Cys Val Lys Asp Val  
 340 345 350

Asn Gly

ES 2 555 056 T3

5  
 <210> 69  
 <211> 876  
 <212> ADN  
 <213> Pseudomonas putida KT2440

10  
 <400> 69

atggtgcacg tgccgggagg cgagttcage tttggttcaa gccgctttaa cgacgaagaa 60  
 15 ggcccgcctc accccgcca ggtgtccggc ttctggattg acgtgcatcc ggtcaccaac 120  
 gccagttcg cgcgcttcgt caaggccaag gggatgtca cccatgccga gcgcggtacc 180  
 20 cgtgtcagagg acgaccctgc cctgcccagc gcgctgcgga taccgggtgc gatgggtgtt 240

catcagggtg cggacgtgct cggccccggc tggcagttcg tgcccggcgc caactggcga 300  
 25 caccgcgaag ggccgggag cagcctggcc gggctggaca accatccggt ggtgcagatc 360  
 gccctggaag atgccaggc ctatgcccgc tgggcaggcc gcgaactgcc cagcgaggcg 420  
 cagctggaat acgccatgcg cggcggcctg accgatgccg acttcagctg gggtagacc 480  
 30 g'agcagccca agggcaagct catggccaat acctggcagg gtcagttccc ttatcgcaat 540  
 gcggcgaag atggttttac cggtagatcg cccgtgggtt gcttcccggc caacggcttt 600  
 35 ggcctgttcg atgccggcgg caatgtctgg gagctgactc gcacgggcta tggccaggc 660  
 catgacgcac agcgcgacgc caagctcagc cctcaggcc cggccctgag tgacagcttc 720  
 gaccggcag accccggcgt gccggtggcg gtaatcaag gcggctcgca cctgtgttcg 780  
 40 ggggaccgct gcatgcgcta ccgcccctcg gcacgccagc cgcagccggt gttcatgacg 840  
 acctcgcacg tgggtttcag aacgattcgg caatga 876

45  
 <210> 70  
 <211> 291  
 <212> PRT  
 <213> Pseudomonas putida KT2440

50  
 <400> 70

55

60

65



ES 2 555 056 T3

5 Met Val His Val Pro Gly Gly Glu Phe Ser Phe Gly Ser Ser Arg Phe  
1 5 10 15

10 Tyr Asp Glu Glu Gly Pro Pro His Pro Ala Lys Val Ser Gly Phe Trp  
20 25 30

15 Ile Asp Val His Pro Val Thr Asn Ala Gln Phe Ala Arg Phe Val Lys  
35 40 45

Ala Thr Gly Tyr Val Thr His Ala Glu Arg Gly Thr Arg Val Glu Asp  
50 55 60

20 Asp Pro Ala Leu Pro Asp Ala Leu Arg Ile Pro Gly Ala Met Val Phe  
65 70 75 80

His Gln Gly Ala Asp Val Leu Gly Pro Gly Trp Gln Phe Val Pro Gly  
85 90 95

25 Ala Asn Trp Arg His Pro Gln Gly Pro Gly Ser Ser Leu Ala Gly Leu  
100 105 110

30 Asp Asn His Pro Val Val Gln Ile Ala Leu Glu Asp Ala Gln Ala Tyr  
115 120 125

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

35 Ala Met Arg Gly Gly Leu Thr Asp Ala Asp Phe Ser Trp Gly Thr Thr  
145 150 155 160

40 Glu Gln Pro Lys Gly Lys Leu Met Ala Asn Thr Trp Gln Gly Gln Phe  
165 170 175

Pro Tyr Arg Asn Ala Ala Lys Asp Gly Phe Thr Gly Thr Ser Pro Val  
180 185 190

45 Gly Cys Phe Pro Ala Asn Gly Phe Gly Leu Phe Asp Ala Gly Gly Asn  
195 200 205

50 Val Trp Glu Leu Thr Arg Thr Gly Tyr Arg Pro Gly His Asp Ala Gln  
210 215 220

55 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

60 His Leu Cys Ser Ala Asp Arg Cys Met Arg Tyr Arg Pro Ser Ala Arg  
260 265 270

65 Gln Pro Gln Pro Val Phe Met Thr Thr Ser His Val Gly Phe Arg Thr  
275 280 285

Ile Arg Gln  
290

ES 2 555 056 T3

5 <210> 71  
 <211> 780  
 <212> ADN  
 <213> Ralstonia metallidurans

10 <400> 71

```

    atgggtcgcgg gcgggatggg gttcgtcggc accaacagcc cggtgccgct gcgcaatac    60
    tggcgctggg ggcgcttcgt acctggcgcg gactggcgtc acccgaccgg cccgggcagt    120
15  tccatcgaag gcaaggacaa tcatcccgtc gtgcaggtct cgtatgaaga cgcgcaggcg    180
    tacgccaagt gggccggcaa gcgtctgcc accgaggccg agtgggagtt tgccgcccgt    240
20  ggcggcctgg agcaggccac ctacgcctgg ggtgacaagt tcgcgccgga tggccggcag    300
    atggcgcaatg tctggcaggg ccagcaggtg cagccgttcc cggtggtcag cgccaaggcg    360
    ggcggcgcgg ctggcaccag tgctgtcggc acgttcccgg gcaatggcta tgggctctat    420
25  gacatgaccg gcaacgcctg gcagtgggtg gccgactggt atcgcgcgga ccagttccgc    480
    cgggaagcca cggtgggcggc agtgctgcag aatccgaccg gcccggccga ttcgtgggac    540
30  ccgaccgaac ctggcgtgcc ggtgtcggcg cccaagcggg tcacgcgcgg tggctcgttc    600
    ctctgcaacg aggacttctg cctcagctac cgcccagatg cccggcgcgg taccgaccgg    660
    tacaccagca tgtcgcacct aggcctccgg ctctgtgatg atgacgccg ttgggcagaa    720
35  gttcgcgaagc agccagccgt ggcaatggcc gcgggcgggc agcagaacgt gcagaaataa    780
  
```

40 <210> 72  
 <211> 259  
 <212> PRT  
 <213> Ralstonia metallidurans

<400> 72

```

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

60

65

5

35

40

45

10

Pro Val Val Gln Val Ser Tyr Glu Asp Ala Gln Ala Tyr Ala Lys Trp  
50 55 60

15

Ala Gly Lys Arg Leu Pro Thr Glu Ala Glu Trp Glu Phe Ala Ala Arg  
65 70 75 80

20

Gly Gly Leu Glu Gln Ala Thr Tyr Ala Trp Gly Asp Lys Phe Ala Pro  
85 90 95

25

Asp Gly Arg Gln Met Ala Asn Val Trp Gln Gly Gln Gln Val Gln Pro  
100 105 110

30

Phe Pro Val Val Ser Ala Lys Ala Gly Gly Ala Ala Gly Thr Ser Ala  
115 120 125

35

Val Gly Thr Phe Pro Gly Asn Gly Tyr Gly Leu Tyr Asp Met Thr Gly  
130 135 140

40

Asn Ala Trp Gln Trp Val Ala Asp Trp Tyr Arg Ala Asp Gln Phe Arg  
145 150 155 160

45

Arg Glu Ala Thr Val Ala Ala Val Leu Gln Asn Pro Thr Gly Pro Ala  
165 170 175

50

Asp Ser Trp Asp Pro Thr Glu Pro Gly Val Pro Val Ser Ala Pro Lys  
180 185 190

55

Arg Val Thr Arg Gly Gly Ser Phe Leu Cys Asn Glu Asp Phe Cys Leu  
195 200 205

60

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

65

<210> 73  
<211> 876  
<212> ADN

ES 2 555 056 T3

<213> Prochlorococcus marinus

<400> 73

5  
 gtgaccacat ctttgccagt agagatggta accatccccg cagggctcta togagttggc 60  
 tgtgatcgct gctatccgga tggttcagtt cgctgctatc cggaggaaac acccgcgcga 120  
 10 gaagtgcagc ttgactcatt ccagatcgac gtagggccag tcaccaatgc ccagttccga 180  
 gctttcgtta gcgccacgca gcattctaca gtctcggagc taccacctga tccaacgctc 240  
 tatcccgatc tagcgcccga ggaacgcata cctgaatcag ttgtctttca accgcctcca 300  
 15 gcaacgggtgg atcgcagcaa acccttgagc tgggtggacc tcatggctgg ggcctgattgg 360  
 cgtcatcccc aaggaccgga aagcacgac gatggccttg atgatcacc tgtcgtgcat 420  
 20 gtcgcctatg ccgacgccat cgcctatgcc cattgggctg gcaagcgtct cccctctgct 480  
 gaagagtggg aagtagccgc ccgcggggggt cttgtcgatg cccaatacgc ctgggggaat 540  
 25 gaactcactc ccaataaccg ctggatggcg aacatctggc aaggtecttt cccttggcac 600  
 aacgaggagc tagacggctg gttctggacc tcgcccgttg gcagctttcc tgccaacggc 660  
 tatggactct tggatgttg cggcaatgtg tgggaatgga ccaactctgt ttatcccgctg 720  
 30 gcgtcaggcc accaggaacg gcgaactatc aaaggcggat cgtttctctg cgcagataat 780  
 tactgcgtac gttatcgacc ctctgcacta caaggccaga cagtagacac tgccacctgt 840  
 35 cacatgggct ttcgctgtgc aaaaggaggg ccttga 876

<210> 74

<211> 291

<212> PRT

40 <213> Prochlorococcus marinus

<400> 74

45

50

55

60

65

ES 2 555 056 T3

5 Met Thr Thr Ser Leu Pro Val Glu Met Val Thr Ile Pro Ala Gly Leu  
1 5 10 15

10 Tyr Arg Val Gly Cys Asp Arg Cys Tyr Pro Asp Gly Ser Val Arg Cys  
20 25 30

15 Tyr Pro Glu Glu Thr Pro Ala Arg Glu Val Gln Leu Asp Ser Phe Gln  
35 40 45

20 Ile Asp Val Gly Pro Val Thr Asn Ala Gln Phe Arg Ala Phe Val Ser  
50 55 60

25 Ala Thr Gln His Leu Thr Val Ser Glu Leu Pro Pro Asp Pro Thr Leu  
65 70 75 80

30 Tyr Pro Asp Leu Ala Pro Glu Glu Arg Ile Pro Glu Ser Val Val Phe  
85 90 95

35 Gln Pro Pro Pro Ala Thr Val Asp Arg Ser Lys Pro Leu Ser Trp Trp  
100 105 110

40 Thr Leu Met Ala Gly Ala Asp Trp Arg His Pro Gln Gly Pro Glu Ser  
115 120 125

45 Thr Ile Asp Gly Leu Asp Asp His Pro Val Val His Val Ala Tyr Ala  
130 135 140

50 Asp Ala Ile Ala Tyr Ala His Trp Ala Gly Lys Arg Leu Pro Ser Ala  
145 150 155 160

55 Glu Glu Trp Glu Val Ala Ala Arg Gly Gly Leu Val Asp Ala Gln Tyr  
165 170 175

60 Ala Trp Gly Asn Glu Leu Thr Pro Asn Asn Arg Trp Met Ala Asn Ile  
180 185 190

65 Trp Gln Gly Pro Phe Pro Trp His Asn Glu Glu Leu Asp Gly Trp Phe  
195 200 205

210 215 220

70 Asp Val Cys Gly Asn Val Trp Glu Trp Thr Asn Ser Val Tyr Pro Val  
225 230 235 240

75 Ala Ser Gly His Gln Glu Arg Arg Thr Ile Lys Gly Gly Ser Phe Leu  
245 250 255

80 Cys Ala Asp Asn Tyr Cys Val Arg Tyr Arg Pro Ser Ala Leu Gln Gly  
260 265 270

85 Gln Thr Val Asp Thr Ala Thr Cys His Met Gly Phe Arg Cys Ala Lys  
275 280 285

90 Gly Gly Pro  
290

<210> 75

ES 2 555 056 T3

5 <211> 1017  
 <212> ADN  
 <213> Caulobacter crescentus CB15

<400> 75

10 ttgggaaaac tgacggcgct tcccgctcctg atgcttctgg cgctggccgg ctgceggccag 60  
 ccggcgccca aggcttgccct ggcggacctg ccggttccag atccccagaa ccgcacggcg 120  
 15 ggtatggttc ggctggcggg cggcgacttc cagatggggc ctgcccgcct gcgtccggag 180  
 gagggaccgc cccagacggg caccgtcccc ccgttctgga tcgatcagac agaggtcacc 240  
 aacgcgcctc tcgcgcgggt cgtcagaggc acgggttata gcaccgtggc cgagcggaccg 300  
 20 ctcgaccccc cgcgctaacc ccacgtaccg gcggcgacgc ggcgtccggc ctcgctcgtc 360  
 ttcgtggggg cgaagggggc gaggtcggac gatccttccc aatggtggca ggtgatcccc 420  
 25 ggcgcccact ggcggcatcc cgaaggctcc ggctcgaaca tccggggcag ggacgcctgg 480  
 ccggtgggtgc atatcgctg ggaggacgcc atggcctacg cccgctggct gggccgtgac 540  
 ctgcccacag aggcgcaatg ggagtacgcc gcgcgcggcg ggctgggttg caagcgctac 600  
 30 acctggggcg accaggctca ggatcctgca aagccgcgcg ccaatacttg gcaaggcgtg 660  
 ttcccggccc aggaccttgg caatgacggc ttcaaggcca agcccgcgcc ggtcggctgc 720  
 ttcccgccca accgctatgg cctgcgcgac atggccggca atgtctggga gtggaccgcg 780  
 35 gactggttca agccgggcct ggatccggtc agcgtcctcg aaaccggcgg gccgcccag 840  
 gcccgcgcgc tggatcccga ggacccgaac acgcccgaag acgtcgtgaa gggcggttcg 900  
 40 ttctgtgctg ccgacgacta ctgcttccgc tategacctg cggcgcgaac gccggggccg 960  
 ccggacagcg gcgcatcgca tgtcgggttc cgcaccgtgc tccgcgccga gcgctga 1017

<210> 76  
 <211> 338  
 <212> PRT  
 <213> Caulobacter crescentus CB15

<400> 76

50  
 Met Gly Lys Leu Thr Ala Leu Pro Val Leu Met Leu Leu Ala Leu Ala  
 1 5 10 15  
 55 Gly Cys Gly Gln Pro Ala Pro Lys Ala Cys Leu Ala Asp Leu Pro Val  
 20 25 30  
 60  
 65

ES 2 555 056 T3

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

Pro Asp Pro Gln Asn Arg Thr Ala Gly Met Val Arg Leu Ala Gly Gly  
 35 40 45

Asp Phe Gln Met Gly Ala Ala Pro Leu Arg Pro Glu Glu Gly Pro Pro  
 50 55 60

Gln Thr Val Thr Val Pro Pro Phe Trp Ile Asp Gln Thr Glu Val Thr  
 65 70 75 80

Asn Ala Ala Phe Ala Arg Phe Val Glu Ala Thr Gly Tyr Arg Thr Val  
 85 90 95

Ala Glu Arg Pro Leu Asp Pro Ala Arg Tyr Ala His Val Pro Ala Ala  
 100 105 110

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

Ser Asp Asp Pro Ser Gln Trp Trp Gln Val Ile Pro Gly Ala Asp Trp  
 130 135 140

Arg His Pro Glu Gly Pro Gly Ser Asn Ile Arg Gly Arg Asp Ala Trp  
 145 150 155 160

Pro Val Val His Ile Ala Trp Glu Asp Ala Met Ala Tyr Ala Arg Trp  
 165 170 175

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

Gly Gly Leu Val Gly Lys Arg Tyr Thr Trp Gly Asp Gln Ala Gln Asp  
 195 200 205

Pro Ala Lys Pro Arg Ala Asn Thr Trp Gln Gly Val Phe Pro Ala Gln  
 210 215 220

Asp Leu Gly Asn Asp Gly Phe Lys Ala Lys Pro Ala Pro Val Gly Cys  
 225 230 235 240

Phe Pro Pro Asn Gly Tyr Gly Leu Arg Asp Met Ala Gly Asn Val Trp  
 245 250 255

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

Leu Glu Thr Gly Gly Pro Pro Glu Ala Arg Ala Leu Asp Pro Glu Asp  
 275 280 285

Pro Asn Thr Pro Lys His Val Val Lys Gly Gly Ser Phe Leu Cys Ala  
 290 295 300

Asp Asp Tyr Cys Phe Arg Tyr Arg Pro Ala Ala Arg Thr Pro Gly Pro  
 305 310 315 320

Pro Asp Ser Gly Ala Ser His Val Gly Phe Arg Thr Val Leu Arg Ala  
 325 330 335

Glu Arg

<210> 77

ES 2 555 056 T3

5

<211> 900  
 <212> ADN  
 <213> Mycobacterium tuberculosis H37Rv

10 <400> 77

	gtgctgaccg agttgggtga cctgcccggc ggatcgttcc gcatgggctc gacgcgcttc	60
15	taccccgaag aagcgcgat tcataccgtg accgtgcgcg cctttgcggt agagcgacac	120
	ccggtgacca acgcgcaatt tgccgaattc gtctccgca caggctatgt gacggttgca	180
	gaacaacccc ttgaccccgg gctctacca ggagtggacg cagcagacct gtgtcccggg	240
20	gcgatggtgt tttgtccgac ggccggggccg gtcgacctgc gtgactggcg gcaatgggtg	300
	gactgggtac ctggcgectg ctggcgccat ccgtttggcc gggacagcga tctcgccgac	360
	cgagcccggc acccggctgt acaggtggcc tatccggacg ccgtggccta cgcacgatgg	420
25	gctgggtcgac gcctaccgac cgaggccgag tgggagtacg cggcccgtgg cggaaccacg	480
	gcaacctatg cgtggggcga ccaggagaag ccggggggca tgctcatggc gaacacctgg	540
30	cagggccggt ttccttaccg caacgacggt gcattgggct ggggtgggaac ctccccggtg	600
	ggcaggtttc cggccaacgg gtttggcttg ctcgacatga tcggaaacgt ttgggagtgg	660
	accaccaacc agttctatcc acaccatcgc atcgatccac cctcgacggc ctgctgcgca	720
35	ccggtcaagc tcgctacagc cgccgacccg acgatcagcc agaccctcaa gggcggctcg	780
	cacctgtgcg cgccggagta ctgccaccgc taccgccggc cggcgcgctc gccgcagtcg	840
40	caggacaccg cgaccacca tctcgggttc cgggtcgtgg ccgacccggt gtcccggtag	900

<210> 78  
 <211> 299  
 <212> PRT  
 45 <213> Mycobacterium tuberculosis H37Rv  
 <400> 78

50

55

60

65



5 Met Leu Thr Glu Leu Val Asp Leu Pro Gly Gly Ser Phe Arg Met Gly  
 1 5 10 15

10 Ser Thr Arg Phe Tyr Pro Glu Glu Ala Pro Ile His Thr Val Thr Val  
 20 25 30

15 Arg Ala Phe Ala Val Glu Arg His Pro Val Thr Asn Ala Gln Phe Ala  
 35 40 45

20 Glu Phe Val Ser Ala Thr Gly Tyr Val Thr Val Ala Glu Gln Pro Leu  
 50 55 60

25 Asp Pro Gly Leu Tyr Pro Gly Val Asp Ala Ala Asp Leu Cys Pro Gly  
 65 70 75 80

30 Ala Met Val Phe Cys Pro Thr Ala Gly Pro Val Asp Leu Arg Asp Trp  
 85 90 95

35 Arg Gln Trp Trp Asp Trp Val Pro Gly Ala Cys Trp Arg His Pro Phe  
 100 105 110

40 Gly Arg Asp Ser Asp Ile Ala Asp Arg Ala Gly His Pro Val Val Gln  
 115 120 125

45

50

55

60

65

5 Val Ala Tyr Pro Asp Ala Val Ala Tyr Ala Arg Trp Ala Gly Arg Arg  
130 135 140

10 Leu Pro Thr Glu Ala Glu Trp Glu Tyr Ala Ala Arg Gly Gly Thr Thr  
145 150 155 160

15 Ala Thr Tyr Ala Trp Gly Asp Gln Glu Lys Pro Gly Gly Met Leu Met  
165 170 175

20 Ala Asn Thr Trp Gln Gly Arg Phe Pro Tyr Arg Asn Asp Gly Ala Leu  
180 185 190

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

30 Gly Leu Leu Asp Met Ile Gly Asn Val Trp Glu Trp Thr Thr Thr Glu  
210 215 220

35 Phe Tyr Pro His His Arg Ile Asp Pro Pro Ser Thr Ala Cys Cys Ala  
225 230 235 240

40 Pro Val Lys Leu Ala Thr Ala Ala Asp Pro Thr Ile Ser Gln Thr Leu  
245 250 255

45 Lys Gly Gly Ser His Leu Cys Ala Pro Glu Tyr Cys His Arg Tyr Arg  
260 265 270

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

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

50 <210> 79  
<211> 7  
<212> PRT  
<213> Artificial

55 <220>  
<223> dominio conservado en procariotas y procariotas

60 <220>  
<221> DOMINIO  
<222> (1)..(7)  
<223> dominio conservado

65 <220>  
<221> MISC\_CARACTERÍSTICA  
<222> (3)..(4)  
<223> Cualquier Aminoácido

<220>  
 <221> MISC\_CHARACTERÍSTICA  
 <222> (6)..(6)  
 <223> Cualquier Aminoácido  
 5 <220>  
 <221> MISC\_CHARACTERÍSTICA  
 <222> (6)..(6)  
 <223> Gly o Ala  
 10 <400> 79

Arg Val Xaa Xaa Gly Xaa Ser  
 1 5

<210> 80  
 <211> 630  
 <212> ADN  
 20 <213> Oncorhynchus mykiss

<400> 80

25	tcaggtggct gctgccccct ggtgggttgc tgtcagagga gcagactgga ggcaccctga	60
	gggccccgac tccagcatca cagacaggct ggaccaccct gtgctgcatg tgtcatggca	120
	ggacgctgtg gcctactgct cctgggccta caagagacta cccacagagg ctgagtggga	180
30	gtacgcctgc agagggggcc tacaggagag actttaccocg tgggggaaca aactgaaacc	240
	taaaggacag cactacgcca acctctggca gggaaagtcc cccacacaca actcagaaga	300
	ggacgggtac actaaaacct caccagtga gtcatttccct gcaaattggct atggcctgta	360
35	caacatggta gggaaatgcat gggagtggac atctgactgg tggactgtac accacaccac	420
	agatgaacag cacaaccocg caggtccacc atcaggcaca gaccgagtga agaaaggagg	480
	ctcctacatg tgcataagt catactgtta caggtacagg tgtgcagcac ggagtcagaa	540
40	caccocctgac agctctgect ctaacctagg gttccgctgt gtctcccagg agcagccgta	600
	acctttcacc ctcgaccctg acatgggtag	630

45 <210> 81  
 <211> 655  
 <212> ADN  
 <213> Danio rerio

50 <220>  
 <221> misc\_característica  
 <222> (590)..(590)  
 <223> n is a, c, g, or t

55 <220>  
 <221> misc\_característica  
 <222> (626)..(626)  
 <223> n is a, c, g, or t

60 <400> 81

65

5  
 caaatgggtt tatttacata aaaaaatcct cttagtttga agtgaagac agtgagatta 60  
 gtgatgtttg aggttatgga tcaacatcag aggcgcagcg gaagccaag ttcgaggctg 120  
 10 aactgtccgg tgtgttctga ctgcgagcgg cacacctgta tctgtagcag taagacttgt 180  
 ggcacatgta ggatcctcct ttcttgactc tgtctgtccc tgattctggg ccctttgggt 240  
 taaacttgtc ttctgcagtg tgatgcacag tccaccagtc tgccgtccac tcccacgcat 300  
 15 ttcccaccat gtcatacagg ccaaagccat tgggaggaaa agacatcacc ggggatgtgt 360  
 tggcatagcc gtcctctgca gtgttgtgat tagggaaate tccctgccac aggttagcat 420  
 20 agtgctgccc tcttggcatt aatttatttc cccatgggta catcctgtcc tgtagtcctc 480  
 ctctacaggg caactcccat tcagctcttg taggaagtct gcgtttggcc cattgacagt 540  
 acgcccgtgc atcatcccat gaaacatgca gagcagggtg attcattctn gtgtgtatgg 600  
 25 ttgaatctgg tctttctgg tgtctncagt ctgcacctt cactggtgac cacca 655

30 <210> 82  
 <211> 773  
 <212> ADN  
 <213> Oryzias latipes

35 <220>  
 <221> misc\_característica  
 <222> (690)..(690)  
 <223> n is a, c, g, o t

40 <220>  
 <221> misc\_característica  
 <222> (755)..(755)  
 <223> n is a, c, g, o t

45 <400> 82  
 tctccttttt tcataaata acattagagt ccttacatc tgccctttaca tacattgtca 60  
 gagacagtac aaaaaatctg cctttgtaaa attagagtta caaaaatata ttttagattt 120  
 50 gacttcttca gaattgtcgg tggcagcaaa agaatcggat tgatctcatg acaagagcgt 180  
 gagccagaag ttcttggatc aaactgattt ggttctgtca tcgtttctgt tcagcagcac 240  
 agcgaaaacc aagattggaa gcggagctgt ctggagtgtt ttggcttcga gcagcacatc 300  
 55 tgtacotgta acaataagac ttgtggcaca tgtacgagcc tctttctctc acctatctg 360  
 tgccctgacgg aggaccggtt gggttgtgct gatggctctgt tgtgtggtgc acgctccacc 420  
 agtctgaggt cactcccat gcgttcccca ccatgtcata cagacaaaa gcattgcctg 480  
 60 ggaaggacat caccggggag gtttttagtgt agccatcctc tgcagagtgt tgtgtggtgga 540  
 attccccctg ccagaggttg gcgtaatgct gtccttttgg gtttagcttg tttccccagg 600  
 65 ggtagagtct gtccttcagg ccgcccctgc aggcaacctc cactctgccc tcagtgggaa 660  
 gtctcttgtt gaccaggag cagtaagccn aggcacatc cccagaaacc tgaacgacgg 720  
 atgatccatc ctgtctgtga tgttggagtc tggandtca ggggtgctcc agt 773

ES 2 555 056 T3

5 <210> 83  
 <211> 566  
 <212> ADN  
 <213> *Xenopus laevis*

10 <220>  
 <221> misc\_característica  
 <222> (6)..(6)  
 <223> n is a, c, g, o t

15 <220>  
 <221> misc\_característica  
 <222> (47)..(47)  
 <223> n is a, c, g, o t

20 <220>  
 <221> misc\_característica  
 <222> (81)..(81)  
 <223> n is a, c, g, o t

25 <400> 83

	<b>atatgnaact aaaggtaatg taattggaat gatggatttc acaaggncctg agagttccct</b>	<b>60</b>
	<b>attgctcctg cttgtcgtgt nacaggtcac ggagccggcg ccacacagcg aaatcccagg</b>	<b>120</b>
30	<b>ttggaggcgc agctgtcggg tgtattctga cttcgagcag cacagcgata cctgtagcaa</b>	<b>180</b>
	<b>taggactcat ggcacatgta ggagcctcct ttcttctactc tatcatttcc cgtagaaggt</b>	<b>240</b>
35	<b>cctttcgggt tgtgaacctc atctgctgta tgatgagtgt cccaccaatc agatgtccac</b>	<b>300</b>
	<b>tcccaagcat ttcccaccat gttatataga ccataaccat tggctgggaa agcagttaca</b>	<b>360</b>
	<b>ggatgaagtct gcacataacc atcctctcca gtgttttggg ttggaaaatc cccctgccag</b>	<b>420</b>
40	<b>acattcgcac aatgttgtcc ctttggttcc agcttgttcc cccatggaaa aatcctgttc</b>	<b>480</b>
	<b>tcaagtcccc cgcggcaggc gtattcccac tcagcttcag ttggaaggcg tttacctgcc</b>	<b>540</b>
45	<b>cagggtgcaga aagcagaagc atcatt</b>	<b>566</b>

50 <210> 84  
 <211> 647  
 <212> ADN  
 <213> *Silurana tropicalis*

<400> 84

55

60

65

ES 2 555 056 T3

5           gcgccttttt tttttttttt tttttttttt catcacaataa ataattttat taataaaaata       60  
             ggattttgtg ttcattotta ttatgaagga caaggaatgt cattgaaatt tttgttttca       120  
             caaggtcttg ggagttcctt cctgctcagg tcatttttga gtggtcacgg agccgacgcc       180  
 10           acgcagcggg atcccagggt agaggccgag ctgtcagggt tattctgact tcgagcagca       240  
             cagcgatacc tgtagcagta ggactcatgg cacatgtatg agcctccttt tttcaccttg       300  
             tcttttcccc taaaaggacc tttcgggttg taagtctcat ctgctgtatg atgagtgtcc       360  
 15           caccaatcgg atgtccactc ccaagcattt cccaccatgt tatataggct ataaccattg       420  
             gctgggaaag cggttacagg tgaagtctgc acatagccgt cctctccagt gttttgggtt       480  
             ggaaattccc cctgccagac attcgcataa tgtttctcct ttggttccag cttgttcccc       540  
 20           cacggaaaaa gcctgttctc aagtccccca cgggaggcat attcccactc agcttctgtc       600  
             ggaaggcgct taccgcacca ggtgcagaag gcagaagcat cgttcca               647

25           <210> 85  
             <211> 636  
             <212> ADN  
             <213> Salmo salar

30           <400> 85  
             atagacattt tttaaatatt ttacaacaaa atatattcca taaatatcca catgtcatgc       60  
 35           ggtaatcctg catttcatga agaacactga catcactggc tgtatgaaga ggtgcacttg       120  
             atltgtttcg cctggcgggc aagataggca gagtttagcac cctagactag agccaatggc       180  
             gaatggtaca aaaagggaaa agtcagacta cccatgtcag ggtcaagggt aaaagggttac       240  
 40           ggctgctcct gggagacaca gcggaacctt aggttagagg cagagctgtc aggggtgttc       300  
             tgactccgtg ctgcacacct gtacctgtaa cagtatgact tatggcacat gtaggagcct       360  
 45           cctttcttca ctcggtctgt gcctgatggg ggacctgccg ggttgtgccg ttcattctgtg       420  
             gtgtggtgta cagtccacca gtcagatgtc cactcccatg cattccctac catgtttgtac       480  
             aggccatagc catttgcagg aaatgacttc actggtgagg ttttgggtgta cccgtcctct       540  
 50           tctgagttgt gtgtggggaa ctttccctgc cagaggttgg cgtagtgtct tcccttaggt       600  
             ttcagtttgt tccccacggg gtaaagtctg tctctg                       636

55           <210> 86  
             <211> 415  
             <212> ADN  
             <213> Sus scrofa

60           <400> 86

65

5 agtttccctgt gaccaacacc ggagaggatg gcttccgagg aactgcgcct gttgatgcct 60  
 ttctcccaa tggttatggc ctttacaata tagtagggaa cgcctgggaa tggacctcag 120  
 actgggtggac cattcaccat gctgctgaag aaacaattaa cccatcaagt tcttccctgct 180  
 10 gcaccgaata acagagccgc cactacgtga tgaagcaga gaaaggcccc ccttctggga 240  
 aagaccgggt gaagaaaggg ggatcctata tgtgccataa gtcctactgc tacaggtacc 300  
 15 gctgtgctgc tgaagccag aacacgccgg acagctcggc ttcaaactcg gggttccgct 360

gtgcagctga ccaccagccc accacaggct gagtcaggaa gagtcttccc gaatc 415

20 <210> 87  
 <211> 595  
 <212> ADN  
 <213> Bos taurus

25 <400> 87

ccacgcgtcc gggggcaaca aactgcagcc gaaaggccag cattatagcc aacatcttgg 60  
 30 caaggcgagt ttctgtgac caacacoggg gaggacggct tccgaggac cgcgcctggt 120  
 gaogccttcc ctcccaatgg ttattggctt atacaatata gtagggaacg cctggggagtg 180  
 gacttcagac tgggtggactg ttcaccattc tgctgaagaa acgattaacc caaaaggccc 240  
 35 ccttctggg aaagaccggg tgaagaaagg tggatcctac atgtgccata aatcctattg 300  
 ctacaggtat cgctgtgctg ctogaagcca gaacacaccc gacagctctg cttogaatct 360  
 40 gggattccgt tgtgcagctg accacctgcc caccacaggg taagagccaa aaagagcctt 420  
 cccgaacccg agaagtcgtg tctactctgc acgcggttc cctcagaagg ctgaacaacc 480  
 tgctgtgaag aattcccacc ccaaggtggg ttacatacct tgcccagtgg ccaaaggacc 540  
 45 tatggcaaga ccaaattgct gagctgatca gcatgtgcgc tttattgggg gatgg 595

50 <210> 88  
 <211> 1611  
 <212> ADN  
 <213> Homo sapiens

55 <220>  
 <221> CDS  
 <222> (1)..(1608)  
 <223> hSULF3

<400> 88

60

65

5	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	
10	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	
15	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	
20	cat cca gga agt cag gta gtg aaa ctt 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	
25	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	
30	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	
35	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	
40	atg gat gtc atg gag agg cat ggc 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	
45	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	
50		
55		
60		
65		



ES 2 555 056 T3

5 tgg aca aga gat gtt gct 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 ccg 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 gct 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 oct cag aac ctg agt gga tac tct ttg ttg ccg 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 tcg 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 tcg 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

ES 2 555 056 T3

5	450		455		460												
	cag	aag	ctt	cat	tcc	att	ata	aac	tac	cct	aaa	gtt	tct	gct	tct	gtc	1440
	Gln	Lys	Leu	His	Ser	Ile	Ile	Asn	Tyr	Pro	Lys	Val	Ser	Ala	Ser	Val	
10	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		
15	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			
20	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				
25	acc	cat	atg	aat	cca	aga	gca	ggt	tga								1611
	Thr	His	Met	Asn	Pro	Arg	Ala	Val									
		530					535										

<210> 89  
 <211> 536  
 <212> PRT  
 <213> Homo sapiens  
 <400> 89

35  
 40  
 45  
 50  
 55  
 60  
 65

5 Met Leu Leu Leu Trp Val Ser Val Val Ala Ala Leu Ala Leu Ala Val  
1 5 10

10 Leu Ala Pro Gly Ala Gly Glu Gln Arg Arg Arg Ala Ala Lys Ala Pro  
20 25 30

15 Asn Val Val Leu Val Val Ser Asp Ser Phe Asp Gly Arg Leu Thr Phe  
35 40 45

20 His Pro Gly Ser Gln Val Val Lys Leu Pro Phe Ile Asn Phe Met Lys  
50 55 60

25 Thr Arg Gly Thr Ser Phe Leu Asn Ala Tyr Thr Asn Ser Pro Ile Cys  
65 70 75 80

30 Cys Pro Ser Arg Ala Ala Met Trp Ser Gly Leu Phe Thr His Leu Thr  
85 90 95

35 Glu Ser Trp Asn Asn Phe Lys Gly Leu Asp Pro Asn Tyr Thr Thr Trp  
100 105 110

40 Met Asp Val Met Glu Arg His Gly Tyr Arg Thr Gln Lys Phe Gly Lys  
115 120 125

45 Leu Asp Tyr Thr Ser Gly His His Ser Ile Ser Asn Arg Val Glu Ala  
130 135 140

50 Trp Thr Arg Asp Val Ala Phe Leu Leu Arg Gln Glu Gly Arg Pro Met  
145 150 155 160

55 Val Asn Leu Ile Arg Asn Arg Thr Lys Val Arg Val Met Glu Arg Asp  
165 170 175

60 Trp Gln Asn Thr Asp Lys Ala Val Asn Trp Leu Arg Lys Glu Ala Ile  
180 185 190

55  
60  
65

ES 2 555 056 T3

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

15 His Thr Ser Leu Tyr Trp Leu Glu Lys Val Ser His Asp Ala Ile Lys  
225 230 235 240

20 Ile Pro Lys Trp Ser Pro Leu Ser Glu Met His Pro Val Asp Tyr Tyr  
245 250 255

25 Ser Ser Tyr Thr Lys Asn Cys Thr Gly Arg Phe Thr Lys Lys Glu Ile  
260 265 270

30 Lys Asn Ile Arg Ala Phe Tyr Tyr Ala Met Cys Ala Glu Thr Asp Ala  
275 280 285

35 Met Leu Gly Glu Ile Ile Leu Ala Leu His Gln Leu Asp Leu Leu Gln  
290 295 300

40 Lys Thr Ile Val Ile Tyr Ser Ser Asp His Gly Glu Leu Ala Met Glu  
305 310 315 320

45 His Arg Gln Phe Tyr Lys Met Ser Met Tyr Glu Ala Ser Ala His Val  
325 330 335

50 Pro Leu Leu Met Met Gly Pro Gly Ile Lys Ala Gly Leu Gln Val Ser  
340 345 350

55 Asn Val Val Ser Leu Val Asp Ile Tyr Pro Thr Met Leu Asp Ile Ala  
355 360 365

60 Gly Ile Pro Leu Pro Gln Asn Leu Ser Gly Tyr Ser Leu Leu Pro Leu  
370 375 380

65 Ser Ser Glu Thr Phe Lys Asn Glu His Lys Val Lys Asn Leu His Pro  
385 390 395 400

70 Pro Trp Ile Leu Ser Glu Phe His Gly Cys Asn Val Asn Ala Ser Thr  
405 410 415

75 Tyr Met Leu Arg Thr Asn His Trp Lys Tyr Ile Ala Tyr Ser Asp Gly  
420 425 430

80 Ala Ser Ile Leu Pro Gln Leu Phe Asp Leu Ser Ser Asp Pro Asp Glu  
435 440 445

85 Leu Thr Asn Val Ala Val Lys Phe Pro Glu Ile Thr Tyr Ser Leu Asp  
450 455 460

90 Gln Lys Leu His Ser Ile Ile Asn Tyr Pro Lys Val Ser Ala Ser Val  
465 470 475 480

95 His Gln Tyr Asn Lys Glu Gln Phe Ile Lys Trp Lys Gln Ser Ile Gly  
485 490 495

100 Gln Asn Tyr Ser Asn Val Ile Ala Asn Leu Arg Trp His Gln Asp Trp  
500 505 510

5

Gln Lys Glu Pro Arg Lys Tyr Glu Asn Ala Ile Asp Gln Trp Leu Lys  
515 520 525

10

Thr His Met Asn Pro Arg Ala Val  
530 535

15

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

20

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

25

<400> 90

30

35

40

45

50

55

60

65

5	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	
10	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 Glu Gly Ala Leu	
	20 25 30	
15	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	
20	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	
25	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	
25	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	
30	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	
35	gga ctt caa cat tct atc ata aga cct acc caa 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	
35	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	
40	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	
45	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	
50	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	
50	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	
55	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	

60

65

ES 2 555 056 T3

5	gct tcc cat aac ccc aca aag cct ata ttt tta tat att gcc tat caa Ala Ser His Asn Pro Thr Lys Pro Ile Phe Leu Tyr Ile Ala Tyr Gln 225 230 235 240	720
10	gct gtt cat tca cca ctg caa gct cct ggc agg tat ttc gaa cac tac Ala Val His Ser Pro Leu Gln Ala Pro Gly Arg Tyr Phe Glu His Tyr 245 250 255	768
15	cga tcc att atc aac ata aac agg agg aga tat gct gcc atg ctt tcc Arg Ser Ile Ile Asn Ile Asn Arg Arg Arg Tyr Ala Ala Met Leu Ser 260 265 270	816
20	tgc tta gat gaa gca atc aac aac gtg aca ttg gct cta aag act tat Cys Leu Asp Glu Ala Ile Asn Asn Val Thr Leu Ala Leu Lys Thr Tyr 275 280 285	864
25	ggg ttc tat aac aac agc att atc att tac tct tca gat aat ggt ggc Gly Phe Tyr Asn Asn Ser Ile Ile Ile Tyr Ser Ser Asp Asn Gly Gly 290 295 300	912
30	cag cct acg gca gga ggg agt aac tgg cct ctc aga ggt agc aaa gga Gln Pro Thr Ala Gly Gly Ser Asn Trp Pro Leu Arg Gly Ser Lys Gly 305 310 315 320	960
35	aca tat tgg gaa gga ggg atc cgg gct gta ggc ttt gtg cat agc cca Thr Tyr Trp Glu Gly Gly Ile Arg Ala Val Gly Phe Val His Ser Pro 325 330 335	1008
40	ctt ctg aaa aac aag gga aca gtg tgt aag gaa ctt gtg cac atc act Leu Leu Lys Asn Lys Gly Thr Val Cys Lys Glu Leu Val His Ile Thr 340 345 350	1056
45	gac tgg tac ccc act ctc att tca ctg gct gaa gga cag att gat gag Asp Trp Tyr Pro Thr Leu Ile Ser Leu Ala Glu Gly Gln Ile Asp Glu 355 360 365	1104
50	gac att caa cta gat ggc tat gat atc tgg gag acc ata agt gag ggt Asp Ile Gln Leu Asp Gly Tyr Asp Ile Trp Glu Thr Ile Ser Glu Gly 370 375 380	1152
55	ctt cgc tca ccc cga gta gat att ttg cat aac att gac ccc ata tac Leu Arg Ser Pro Arg Val Asp Ile Leu His Asn Ile Asp Pro Ile Tyr 385 390 395 400	1200
60	acc aag gca aaa aat ggc tcc tgg gca gca ggc tat ggg atc tgg aac Thr Lys Ala Lys Asn Gly Ser Trp Ala Ala Gly Tyr Gly Ile Trp Asn 405 410 415	1248
65	act gca atc cag tca gcc atc aga gtg cag cac tgg aaa ttg ctt aca Thr Ala Ile Gln Ser Ala Ile Arg Val Gln His Trp Lys Leu Leu Thr 420 425 430	1296
70	gga aat cct ggc tac agc gac tgg gtc ccc cct cag tct ttc agc aac Gly Asn Pro Gly Tyr Ser Asp Trp Val Pro Pro Gln Ser Phe Ser Asn 435 440 445	1344
75	ctg gga ccg aac cgg tgg cac aat gaa cgg atc acc ttg tca act ggc Leu Gly Pro Asn Arg Trp His Asn Glu Arg Ile Thr Leu Ser Thr Gly 450 455 460	1392
80	aaa agt gta tgg ctt ttc aac atc aca gcc gac cca tat gag agg gtg Lys Ser Val Trp Leu Phe Asn Ile Thr Ala Asp Pro Tyr Glu Arg Val 465 470 475 480	1440
85	gac cta tct aac agg tat cca gga atc gtg aag aag ctc cta cgg agg Asp Leu Ser Asn Arg Tyr Pro Gly Ile Val Lys Lys Leu Leu Arg Arg 485 490 495	1488
90	ctc tca cag ttc aac aaa act gca gtg ccg gtc agg tat ccc ccc aaa Leu Ser Gln Phe Asn Lys Thr Ala Val Pro Val Arg Tyr Pro Pro Lys 500 505 510	1536
95	gac ccc aga agt aac cct agg ctc aat gga ggg gtc tgg gga cca tgg Asp Pro Arg Ser Asn Pro Arg Leu Asn Gly Gly Val Trp Gly Pro Trp 515 520 525	1584
100	tat aaa gag gaa acc aag aaa aag aag cca agc aaa aat cag gct gag Tyr Lys Glu Glu Thr Lys Lys Lys Lys Pro Ser Lys Asn Gln Ala Glu	1632

ES 2 555 056 T3

5																	
		530				535					540						
	aaa	aag	caa	aag	aaa	agc	aaa	aaa	aag	aag	aag	aaa	cag	cag	aaa	gca	1680
	Lys	Lys	Gln	Lys	Lys	Ser	Lys	Lys	Lys	Lys	Lys	Lys	Gln	Gln	Lys	Ala	
10	545					550					555				560		
	gtc	tca	ggt	tca	act	tgc	cat	tca	ggt	gtt	act	tgt	gga	taa			1722
	Val	Ser	Gly	Ser	Thr	Cys	His	Ser	Gly	Val	Thr	Cys	Gly				
					565					570							

15  
 <210> 91  
 <211> 573  
 <212> PRT  
 <213> Homo sapiens

20  
 <400> 91

25

30

35

40

45

50

55

60

65



5  
 Met Gly Ala Leu Ala Gly Phe Trp Ile Leu Cys Leu Leu Thr Tyr Gly  
 1 5 10 15

10  
 Tyr Leu Ser Trp Gly Gln Ala Leu Glu Glu Glu Glu Gly Ala Leu  
 20 25 30

15  
 Leu Ala Gln Ala Gly Glu Lys Leu Glu Pro Ser Thr Thr Ser Thr Ser  
 35 40 45

20  
 Gln Pro His Leu Ile Phe Ile Leu Ala Asp Asp Gln Gly Phe Arg Asp  
 50 55 60

25  
 Val Gly Tyr His Gly Ser Glu Ile Lys Thr Pro Thr Leu Asp Lys Leu  
 65 70 75 80

30  
 Ala Ala Glu Gly Val Lys Leu Glu Asn Tyr Tyr Val Gln Pro Ile Cys  
 85 90 95

35  
 Thr Pro Ser Arg Ser Gln Phe Ile Thr Gly Lys Tyr Gln Ile His Thr  
 100 105 110

40  
 Gly Leu Gln His Ser Ile Ile Arg Pro Thr Gln Pro Asn Cys Leu Pro  
 115 120 125

45  
 Leu Asp Asn Ala Thr Leu Pro Gln Lys Leu Lys Glu Val Gly Tyr Ser  
 130 135 140

50  
 Thr His Met Val Gly Lys Trp His Leu Gly Phe Tyr Arg Lys Glu Cys  
 145 150 155 160

55  
 Met Pro Thr Arg Arg Gly Phe Asp Thr Phe Phe Gly Ser Leu Leu Gly  
 165 170 175

60  
 Ser Gly Asp Tyr Tyr Thr His Tyr Lys Cys Asp Ser Pro Gly Met Cys  
 180 185 190

65  
 Gly Tyr Asp Leu Tyr Glu Asn Asp Asn Ala Ala Trp Asp Tyr Asp Asn  
 195 200 205

70  
 Gly Ile Tyr Ser Thr Gln Met Tyr Thr Gln Arg Val Gln Gln Ile Leu  
 210 215 220

75  
 Ala Ser His Asn Pro Thr Lys Pro Ile Phe Leu Tyr Ile Ala Tyr Gln  
 225 230 235 240

ES 2 555 056 T3

5 Ala Val His Ser Pro Leu Gln Ala Pro Gly Arg Tyr Phe Glu His Tyr  
245 250 255

10 Arg Ser Ile Ile Asn Ile Asn Arg Arg Arg Tyr Ala Ala Met Leu Ser  
260 265 270

15 Cys Leu Asp Glu Ala Ile Asn Asn Val Thr Leu Ala Leu Lys Thr Tyr  
275 280 285

20 Gly Phe Tyr Asn Asn Ser Ile Ile Ile Tyr Ser Ser Asp Asn Gly Gly  
290 295 300

25 Gln Pro Thr Ala Gly Gly Ser Asn Trp Pro Leu Arg Gly Ser Lys Gly  
305 310 315

30 Thr Tyr Trp Glu Gly Gly Ile Arg Ala Val Gly Phe Val His Ser Pro  
325 330 335

35 Leu Leu Lys Asn Lys Gly Thr Val Cys Lys Glu Leu Val His Ile Thr  
340 345 350

40 Asp Trp Tyr Pro Thr Leu Ile Ser Leu Ala Glu Gly Gln Ile Asp Glu  
355 360 365

45 Asp Ile Gln Leu Asp Gly Tyr Asp Ile Trp Glu Thr Ile Ser Glu Gly  
370 375 380

50 Leu Arg Ser Pro Arg Val Asp Ile Leu His Asn Ile Asp Pro Ile Tyr  
385 390 395 400

55 Thr Lys Ala Lys Asn Gly Ser Trp Ala Ala Gly Tyr Gly Ile Trp Asn  
405 410 415

60 Thr Ala Ile Gln Ser Ala Ile Arg Val Gln His Trp Lys Leu Leu Thr  
420 425 430

65 Gly Asn Pro Gly Tyr Ser Asp Trp Val Pro Pro Gln Ser Phe Ser Asn  
435 440 445

70 Leu Gly Pro Asn Arg Trp His Asn Glu Arg Ile Thr Leu Ser Thr Gly  
450 455 460

75 Lys Ser Val Trp Leu Phe Asn Ile Thr Ala Asp Pro Tyr Glu Arg Val  
465 470 475 480

80 Asp Leu Ser Asn Arg Tyr Pro Gly Ile Val Lys Lys Leu Leu Arg Arg  
485 490 495

85 Leu Ser Gln Phe Asn Lys Thr Ala Val Pro Val Arg Tyr Pro Pro Lys  
500 505 510

90 Asp Pro Arg Ser Asn Pro Arg Leu Asn Gly Gly Val Trp Gly Pro Trp  
515 520 525

95 Tyr Lys Glu Glu Thr Lys Lys Lys Lys Pro Ser Lys Asn Gln Ala Glu  
530 535 540

100 Lys Lys Gln Lys Lys Ser Lys Lys Lys Lys Lys Lys Gln Gln Lys Ala  
545 550 555 560

5 Val Ser Gly Ser Thr Cys His Ser Gly Val Thr Cys Gly  
565 570

10 <210> 92  
<211> 1710  
<212> ADN  
<213> Homo sapiens

15 <220>  
<221> CDS  
<222> (1)..(1707)  
<223> hSULF5

20 <400> 92

25

30

35

40

45

50

55

60

65

5	atg cac acc ctc act ggc ttc tcc ctg gtc agc ctg ctc agc ttc ggc Met His Thr Leu Thr Gly Phe Ser Leu Val Ser Leu Leu Ser Phe Gly 1 5 10 15	48
10	tac ctg tcc tgg gac tgg gcc aag ccg agc ttc gtg gcc gac ggg ccc Tyr Leu Ser Trp Asp Trp Ala Lys Pro Ser Phe Val Ala Asp Gly Pro 20 25 30	96
15	ggg gag gct ggc gag cag ccc tcg gcc gct ccg ccc cag cct ccc cac Gly Glu Ala Gly Glu Gln Pro Ser Ala Ala Pro Pro Gln Pro Pro His 35 40 45	144
20	atc atc ttc atc ctc acg gac gac caa ggc tac cac gac gtg ggc tac Ile Ile Phe Ile Leu Thr Asp Asp Gln Gly Tyr His Asp Val Gly Tyr 50 55 60	192
25	cat ggt tca gat atc gag acc cct acg ctg gac agg ctg gcg gcc aag His Gly Ser Asp Ile Glu Thr Pro Thr Leu Asp Arg Leu Ala Ala Lys 65 70 75 80	240
30	ggg gtc aag ttg gag aat tat tac atc cag ccc atc tgc acg cct tcg Gly Val Lys Leu Glu Asn Tyr Tyr Ile Gln Pro Ile Cys Thr Pro Ser 85 90 95	288
35	cgg agc cag ctc ctc act ggc agg tac cag atc cac aca gga ctc cag Arg Ser Gln Leu Leu Thr Gly Arg Tyr Gln Ile His Thr Gly Leu Gln 100 105 110	336
40	cat tcc atc atc cgc cca cag cag ccc aac tgc ctg ccc ctg gac cag His Ser Ile Ile Arg Pro Gln Gln Pro Asn Cys Leu Pro Leu Asp Gln 115 120 125	384
45	gtg aca ctg cca cag aag ctg cag gag gca ggt tat tcc acc cat atg Val Thr Leu Pro Gln Lys Leu Gln Glu Ala Gly Tyr Ser Thr His Met 130 135 140	432
50	gtg ggc aag tgg cac ctg ggc ttc tac cgg aag gag tgt ctg ccc acc Val Gly Lys Trp His Leu Gly Phe Tyr Arg Lys Glu Cys Leu Pro Thr 145 150 155 160	480
55	cgt cgg ggc ttc gac acc ttc ctg ggc tcg ctc acg ggc aat gtg gac Arg Arg Gly Phe Asp Thr Phe Leu Gly Ser Leu Thr Gly Asn Val Asp 165 170 175	528
60	tat tac acc tat gac aac tgt gat ggc cca ggc gtg tgc ggc ttc gac Tyr Tyr Thr Tyr Asp Asn Cys Asp Gly Pro Gly Val Cys Gly Phe Asp 180 185 190	576
65	ctg cac gag ggt gag aat gtg gcc tgg ggg ctc agc ggc cag tac tcc Leu His Glu Gly Glu Asn Val Ala Trp Gly Leu Ser Gly Gln Tyr Ser 195 200 205	624
70	act atg ctt tac gcc cag cgc gcc agc cat atc ctg gcc agc cac agc Thr Met Leu Tyr Ala Gln Arg Ala Ser His Ile Leu Ala Ser His Ser 210 215 220	672
75	cct cag cgt ccc ctc ttc ctc tat gtg gcc ttc cag gca gta cac aca Pro Gln Arg Pro Leu Phe Leu Tyr Val Ala Phe Gln Ala Val His Thr 225 230 235 240	720

ES 2 555 056 T3

5	ccc ctg cag tcc cct cgt gag tac ctg tac cgc tac cgc acc atg ggc Pro Leu Gln Ser Pro Arg Glu Tyr Leu Tyr Arg Tyr Arg Thr Met Gly	768
	245 250 255	
10	aat gtg gcc cgg cgg aag tac gcc gcc atg gtg acc tgc atg gat gag Asn Val Ala Arg Arg Lys Tyr Ala Ala Met Val Thr Cys Met Asp Glu	816
	260 265 270	
15	gct gtg cgc aac atc acc tgg gcc ctc aag cgc tac ggt ttc tac aac Ala Val Arg Asn Ile Thr Trp Ala Leu Lys Arg Tyr Gly Phe Tyr Asn	864
	275 280 285	
20	aac agt gtc atc atc ttc tcc agt gac aat ggt gcc cag act ttc tgc Asn Ser Val Ile Ile Phe Ser Ser Asp Asn Gly Gly Gln Thr Phe Ser	912
	290 295 300	
25	ggg gcc agc aac tgg ccg ctc cga gga cgc aag gcc act tat tgg gaa Gly Gly Ser Asn Trp Pro Leu Arg Gly Arg Lys Gly Thr Tyr Trp Glu	960
	305 310 315 320	
30	ggt gcc gtg cgg gcc cta gcc ttt gtc cac agt ccc ctg ctc aag cga Gly Gly Val Arg Gly Leu Gly Phe Val His Ser Pro Leu Leu Lys Arg	1008
	325 330 335	
35	aag caa cgg aca agc cgg gca ctg atg cac atc act gac tgg tac ccg Lys Gln Arg Thr Ser Arg Ala Leu Met His Ile Thr Asp Trp Tyr Pro	1056
	340 345 350	
40	acc ctg gtg ggt ctg gca ggt ggt acc acc tca gca gcc gat ggg cta Thr Leu Val Gly Leu Ala Gly Gly Thr Thr Ser Ala Ala Asp Gly Leu	1104
	355 360 365	
45	gat gcc tac gac gtg tgg ccg gcc atc agc gag gcc cgg gcc tca cca Arg Gly Tyr Asp Val Trp Pro Ala Ile Ser Glu Gly Arg Ala Ser Pro	1152
	370 375 380	
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 gcc atc tgg aac acc gcc gtg cag His Gly Ser Leu Glu Gly Gly Phe Gly Ile Trp Asn Thr Ala Val Gln	1248
	405 410 415	
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	1296
	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 Pro Ala Glu Asn Pro Arg Ala His	1536
	500 505 510	
90	cct gac ttt aat ggg ggt gct tgg ggg ccc tgg gcc agt gat gag gaa Pro Asp Phe Asn Gly Gly Ala Trp Gly Pro Trp Ala Ser Asp Glu Glu	1584
	515 520 525	
95	gag gag gaa gag gaa ggg agg gct cga agc ttc tcc cgg ggt cgt cgc Glu Glu Glu Glu Glu Gly Arg Ala Arg Ser Phe Ser Arg Gly Arg Arg	1632
	530 535 540	
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

ES 2 555 056 T3

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  
 25                      20                      25                      30  
 Gly Glu Ala Gly Glu Gln Pro Ser Ala Ala Pro Pro Gln Pro Pro His  
 30                      35                      40                      45  
 Ile Ile Phe Ile Leu Thr Asp Asp Gln Gly Tyr His Asp Val Gly Tyr  
 50                      55                      60  
 His Gly Ser Asp Ile Glu Thr Pro Thr Leu Asp Arg Leu Ala Ala Lys  
 65                      70                      75                      80  
 35 Gly Val Lys Leu Glu Asn Tyr Tyr Ile Gln Pro Ile Cys Thr Pro Ser  
 85                      90                      95  
 40 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  
 45 Val Thr Leu Pro Gln Lys Leu Gln Glu Ala Gly Tyr Ser Thr His Met  
 130                      135                      140  
 50 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  
 55 Tyr Tyr Thr Tyr Asp Asn Cys Asp Gly Pro Gly Val Cys Gly Phe Asp  
 180                      185                      190  
 60 Leu His Glu Gly Glu Asn Val Ala Trp Gly Leu Ser Gly Gln Tyr Ser  
 195                      200                      205  
 Thr Met Leu Tyr Ala Gln Arg Ala Ser His Ile Leu Ala Ser His Ser  
 210                      215                      220  
 65 Pro Gln Arg Pro Leu Phe Leu Tyr Val Ala Phe Gln Ala Val His Thr  
 225                      230                      235                      240  
 Pro Leu Gln Ser Pro Arg Glu Tyr Leu Tyr Arg Tyr Arg Thr Met Gly  
 245                      250                      255

ES 2 555 056 T3

5

Asn Val Ala Arg Arg Lys Tyr Ala Ala Met Val Thr Cys Met Asp Glu  
260 265 270

10

Ala Val Arg Asn Ile Thr Trp Ala Leu Lys Arg Tyr Gly Phe Tyr Asn  
275 280 285

15

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

20

Gly Gly Ser Asn Trp Pro Leu Arg Gly Arg Lys Gly Thr Tyr Trp Glu  
305 310 315 320

25

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

30

Thr Leu Val Gly Leu Ala Gly Gly Thr Thr Ser Ala Ala Asp Gly Leu  
355 360 365

Asp Gly Tyr Asp Val Trp Pro Ala Ile Ser Glu Gly Arg Ala Ser Pro  
370 375 380

35

Arg Thr Glu Ile Leu His Asn Ile Asp Pro Leu Tyr Asn His Ala Gln  
385 390 395 400

His Gly Ser Leu Glu Gly Gly Phe Gly Ile Trp Asn Thr Ala Val Gln  
405 410 415

40

Ala Ala Ile Arg Val Gly Glu Trp Lys Leu Leu Thr Gly Asp Pro Gly  
420 425 430

Tyr Gly Asp Trp Ile Pro Pro Gln Thr Leu Ala Thr Phe Pro Gly Ser  
435 440 445

45

Trp Trp Asn Leu Glu Arg Met Ala Ser Val Arg Gln Ala Val Trp Leu  
450 455 460

50

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

55

Arg Thr Ala Ile Pro Val Arg Tyr Pro Ala Glu Asn Pro Arg Ala His  
500 505 510

60

Pro Asp Phe Asn Gly Gly Ala Trp Gly Pro Trp Ala Ser Asp Glu Glu  
515 520 525

Glu Glu Glu Glu Gly Arg Ala Arg Ser Phe Ser Arg Gly Arg Arg  
530 535 540

65

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

ES 2 555 056 T3

5 <210> 94  
 <211> 2067  
 <212> ADN  
 <213> Homo sapiens

10 <220>  
 <221> CDS  
 <222> (1)..(2064)  
 <223> hSULF6

15 <400> 94

1	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 cgg 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 ggc 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 ggc 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	



ES 2 555 056 T3

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	ctt tta agc gac tgc cag gca tcc aag aca cca gaa ctg cac cgc tgg Leu Leu Ser Asp Cys Gln Ala Ser Lys Thr Pro Glu Leu His Arg Trp	864
	275 280 285	
	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	
15	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
	305 310 320	
20	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	
	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	
25	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	
30	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	
	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
	385 390 400	
35	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	
40	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	
	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	
45	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	
50	ggc tgg aac ggg atc tac aaa ggt ggc aaa gga atg gga gga tgg gaa Gly Trp Asn Gly Ile Tyr Lys Gly Gly Lys Gly Met Gly Gly Trp Glu	1440
	465 470 475 480	
	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	
55	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	
60	acg ctg tct tat ata ggc gga ggg atc ttg tcc cag gac aga gtg att Thr Leu Ser Tyr Ile Gly Gly Ile Leu Ser Gln Asp Arg Val Ile	1584
	515 520 525	
	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	
65	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	
	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

ES 2 555 056 T3

			565					570				575					
	act	cct	aaa	ttc	tac	cct	gaa	gga	aca	ggt	gcc	tgc	tat	ggg	agt	gga	1776
5	Thr	Pro	Lys	Phe	Tyr	Pro	Glu	Gly	Thr	Gly	Ala	Cys	Tyr	Gly	Ser	Gly	
			580					585					590				
	ata	tgt	tca	tgt	tcg	ggg	gat	gta	acc	tac	cac	gac	cca	cca	ctc	ctc	1824
	Ile	Cys	Ser	Cys	Ser	Gly	Asp	Val	Thr	Tyr	His	Asp	Pro	Pro	Leu	Leu	
			595					600					605				
10	ttt	gac	atc	tca	aga	gac	cct	tca	gaa	gcc	ctt	cca	ctg	aac	cct	gac	1872
	Phe	Asp	Ile	Ser	Arg	Asp	Pro	Ser	Glu	Ala	Leu	Pro	Leu	Asn	Pro	Asp	
		610				615						620					
15	aat	gag	cca	tta	ttt	gac	tcc	gtg	atc	aaa	aag	atg	gag	gca	gcc	ata	1920
	Asn	Glu	Pro	Leu	Phe	Asp	Ser	Val	Ile	Lys	Lys	Met	Glu	Ala	Ala	Ile	
		625				630					635				640		
20	aga	gag	cat	cgt	agg	aca	cta	aca	cct	gtc	cca	cag	cag	ttc	tct	gtg	1968
	Arg	Glu	His	Arg	Arg	Thr	Leu	Thr	Pro	Val	Pro	Gln	Gln	Phe	Ser	Val	
				645						650				655			
25	ttc	aac	aca	att	tgg	aaa	cca	tgg	ctg	cag	cct	tgc	tgt	ggg	acc	ttc	2016
	Phe	Asn	Thr	Ile	Trp	Lys	Pro	Trp	Leu	Gln	Pro	Cys	Cys	Gly	Thr	Phe	
				660					665					670			
30	ccc	ttc	tgt	ggg	tgt	gac	aag	gaa	gat	gac	atc	ctt	ccc	atg	gct	ccc	2064
	Pro	Phe	Cys	Gly	Cys	Asp	Lys	Glu	Asp	Asp	Ile	Leu	Pro	Met	Ala	Pro	
			675					680					685				
35	tga																2067
	<210>	95															
	<211>	688															
	<212>	PRT															
	<213>	Homo sapiens															
	<400>	95															
35	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				
50	Ser	Ser	Ser	Lys	Cys	Gln	Leu	Asp	Gln	Gly	Gly	Lys	Ser	Leu	Val	Gln	
		50					55					60					
55	Cys	Ile	Leu	Pro	Arg	Ser	Ser	Lys	Leu	Leu	Ser	Pro	Leu	Cys	Leu	Pro	
		65				70					75				80		
60	His	Pro	Cys	Gly	Ala	Leu	Leu	Leu	Tyr	Arg	Ser	Ser	Gly	Ile	Ala	Ser	
				85						90					95		
65	Ala	Leu	Ala	Ala	Phe	Thr	Asp	Ser	Leu	Ser	Arg	Ser	Cys	Trp	Leu	Ser	
				100					105					110			
70	Val	Ser	Leu	Cys	Cys	Leu	Phe	Cys	Gly	Val	Asp	Gly	Thr	Phe	Met	Thr	
			115					120					125				
75	Arg	Asn	Ala	Arg	Pro	Asn	Ile	Val	Leu	Leu	Met	Ala	Asp	Asp	Leu	Gly	
		130					135					140					
80	Val	Gly	Asp	Leu	Cys	Cys	Tyr	Gly	Asn	Asn	Ser	Val	Ser	Thr	Pro	Asn	
		145				150					155				160		

Ile Asp Arg Leu Ala Ser Glu Gly Val Arg Leu Thr Gln His Leu Ala  
 165 170 175

5 Ala Ala Ser Met Cys Thr Pro Ser Arg Ala Ala Phe Leu Thr Gly Arg  
 180 185 190

10 Tyr Pro Ile Arg Ser Gly Met Val Ser Ala Tyr Asn Leu Asn Arg Ala  
 195 200 205

Phe Thr Trp Leu Gly Gly Ser Gly Gly Leu Pro Thr Asn Glu Thr Thr  
 210 215 220

15 Phe Ala Lys Leu Leu Gln His Arg Gly Tyr Arg Thr Gly Leu Ile Gly  
 225 230 235 240

20 Lys Trp His Leu Gly Leu Ser Cys Ala Ser Arg Asn Asp His Cys Tyr  
 245 250 255

25 His Pro Leu Asn His Gly Phe His Tyr Phe Tyr Gly Val Pro Phe Gly  
 260 265 270

Leu Leu Ser Asp Cys Gln Ala Ser Lys Thr Pro Glu Leu His Arg Trp  
 275 280 285

30 Leu Arg Ile Lys Leu Trp Ile Ser Thr Val Ala Leu Ala Leu Val Pro  
 290 295 300

35 Phe Leu Leu Leu Ile Pro Lys Phe Ala Arg Trp Phe Ser Val Pro Trp  
 305 310 315 320

Lys Val Ile Phe Val Phe Ala Leu Leu Ala Phe Leu Phe Phe Thr Ser  
 325 330 335

40 Trp Tyr Ser Ser Tyr Gly Phe Thr Arg Arg Trp Asn Cys Ile Leu Met  
 340 345 350

45 Arg Asn His Glu Ile Ile Gln Gln Pro Met Lys Glu Glu Lys Val Ala  
 355 360 365

Ser Leu Met Leu Lys Glu Ala Leu Ala Phe Ile Glu Arg Tyr Lys Arg  
 370 375 380

50 Glu Pro Phe Leu Leu Phe Phe Ser Phe Leu His Val His Thr Pro Leu  
 385 390 395 400

55 Ile Ser Lys Lys Lys Phe Val Gly Arg Ser Lys Tyr Gly Arg Tyr Gly  
 405 410 415

60 Asp Asn Val Glu Glu Met Asp Trp Met Val Gly Lys Ile Leu Asp Ala  
 420 425 430

Leu Asp Gln Glu Arg Leu Ala Asn His Thr Leu Val Tyr Phe Thr Ser  
 435 440 445

65 Asp Asn Gly Gly His Leu Glu Pro Leu Asp Gly Ala Val Gln Leu Gly  
 450 455 460

Gly Trp Asn Gly Ile Tyr Lys Gly Gly Lys Gly Met Gly Gly Trp Glu  
 465 470 475 480  
 5  
 Gly Gly Ile Arg Val Pro Gly Ile Phe Arg Trp Pro Ser Val Leu Glu  
 485 490 495  
 10  
 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  
 30  
 Arg Trp His Gln Lys Asp Cys Ala Thr Val Trp Lys Ala His Tyr Val  
 565 570 575  
 35  
 Thr Pro Lys Phe Tyr Pro Glu Gly Thr Gly Ala Cys Tyr Gly Ser Gly  
 580 585 590  
 40  
 Ile Cys Ser Cys Ser Gly Asp Val Thr Tyr His Asp Pro Pro Leu Leu  
 595 600 605  
 45  
 Phe Asp Ile Ser Arg Asp Pro Ser Glu Ala Leu Pro Leu Asn Pro Asp  
 610 615 620  
 50  
 Asn Glu Pro Leu Phe Asp Ser Val Ile Lys Lys Met Glu Ala Ala Ile  
 625 630 635 640  
 55  
 Arg Glu His Arg Arg Thr Leu Thr Pro Val Pro Gln Gln Phe Ser Val  
 645 650 655  
 60  
 Phe Asn Thr Ile Trp Lys Pro Trp Leu Gln Pro Cys Cys Gly Thr Phe  
 660 665 670  
 65  
 Pro Phe Cys Gly Cys Asp Lys Glu Asp Asp Ile Leu Pro Met Ala Pro  
 675 680 685

## REIVINDICACIONES

1. Un polipéptido para su uso en el tratamiento de una deficiencia de sulfatasa, o el uso de un polipéptido para preparar un medicamento para su uso en el tratamiento de una deficiencia de sulfatasa; en el que dicho polipéptido tiene actividad de generación de C $\alpha$ -formilglicina (actividad de FGE) y:
- tiene una secuencia seleccionada de SEC ID N $^{\circ}$  2, 5, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, o aminoácidos 34-374 de SEC ID N $^{\circ}$ . 2; o
  - tiene al menos el 50 % de identidad de secuencia con SEC ID N $^{\circ}$  2; o
  - tiene una o más mutaciones de aminoácido conservativas con respecto a un polipéptido como se describe en a) anteriormente o con respecto a un polipéptido seleccionado del grupo que consiste en SEC ID N $^{\circ}$  2, 5, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, o aminoácidos 34-374 de SEC ID N $^{\circ}$ . 2; o
  - es un fragmento de un polipéptido como se describe en cualquiera de a) a c) anteriormente; o
  - es una proteína de fusión de cualquiera de a) a d) anteriormente.
2. Un polipéptido para su uso en el tratamiento de una deficiencia de sulfatasa según la reivindicación 1, o el uso del polipéptido para preparar un medicamento para su uso en el tratamiento de una deficiencia de sulfatasa según la reivindicación 1; en el que el grado de identidad de secuencia en la parte b) de la reivindicación 1 es de al menos el 75 %.
3. Un polipéptido para su uso en el tratamiento de una deficiencia de sulfatasa según la reivindicación 1, o el uso del polipéptido para preparar un medicamento para su uso en el tratamiento de una deficiencia de sulfatasa según la reivindicación 1; en el que el grado de identidad de secuencia en la parte b) de la reivindicación 1 es de al menos el 95 %.
4. Un polipéptido para su uso en el tratamiento de una deficiencia de sulfatasa según cualquiera de las reivindicaciones 1 a 3, o el uso del polipéptido para preparar un medicamento para su uso en el tratamiento de una deficiencia de sulfatasa según cualquiera de las reivindicaciones 1 a 3; en el que el polipéptido tiene actividad de generación de C $\alpha$ -formilglicina y tiene un subdominio 3 que comprende al menos uno de los siguientes:
- un motivo GFR
  - un motivo RVXXGG(A)S
  - un heptámero que contiene tres argininas
  - tres residuos de cisteína.
5. Un ácido nucleico o vector que codifica un polipéptido como se describe en cualquiera de las reivindicaciones 1 a 4 para su uso en el tratamiento de una deficiencia de sulfatasa, o el uso de un ácido nucleico o vector que codifica el polipéptido para preparar un medicamento para su uso en el tratamiento de una deficiencia de sulfatasa.
6. Un vector que codifica un polipéptido para su uso en el tratamiento de una deficiencia de sulfatasa según la reivindicación 5, o el uso de un vector que codifica el polipéptido para preparar un medicamento para su uso en el tratamiento de una deficiencia de sulfatasa según la reivindicación 5; en el que dicho vector es un vector viral.
7. Un vector viral que codifica un polipéptido para su uso en el tratamiento de una deficiencia de sulfatasa según la reivindicación 6, o el uso de un vector viral que codifica el polipéptido para preparar un medicamento para su uso en el tratamiento de una deficiencia de sulfatasa; en el que el vector viral es un adenovirus, un virus adeno-asociado o un retrovirus.
8. Un polipéptido para su uso en el tratamiento de una deficiencia de sulfatasa según cualquiera de las reivindicaciones 1 a 4, o el uso de un polipéptido para preparar un medicamento para su uso en el tratamiento de una deficiencia de sulfatasa según cualquiera de las reivindicaciones 5 a 7, o un ácido nucleico o vector que codifica un polipéptido para su uso en el tratamiento de una deficiencia de sulfatasa según cualquiera de las reivindicaciones 5 a 7, o el uso de un ácido nucleico o vector que codifica dicho polipéptido para preparar un medicamento que es para su uso en el tratamiento de una deficiencia de sulfatasa según cualquiera de las reivindicaciones 5 a 7; en el que el tratamiento es el tratamiento de un ser humano.
9. Un polipéptido para su uso en el tratamiento de una deficiencia de sulfatasa según cualquiera de las reivindicaciones 1 a 4 o 8, o el uso de un polipéptido para preparar un medicamento para su uso en el tratamiento de una deficiencia de sulfatasa según cualquiera de las reivindicaciones 1 a 4 o 8, o un ácido nucleico o vector que codifica un polipéptido para su uso en el tratamiento de una deficiencia de sulfatasa según cualquiera de las reivindicaciones 5 a 8, o el uso de un ácido nucleico o vector que codifica dicho polipéptido para preparar un medicamento que es para su uso en el tratamiento de una deficiencia de sulfatasa según cualquiera de las reivindicaciones 5 a 8;
- en el que dicho deficiencia de sulfatasa está seleccionada del grupo que consiste en: deficiencia múltiple de sulfatasas, mucopolisacaridosis II (MPS II; síndrome de Hunter), mucopolisacaridosis IIIA (MPS IIIA; síndrome de Sanfilippo A), mucopolisacaridosis VIII (MPS VIII), mucopolisacaridosis IVA (MPS IVA; síndrome de Morquio A),

mucopolisacaridosis VI (MPS VI; síndrome de Maroteaux-Lamy), leucodistrofia metacromática (MLD), condrod displasia punctata 1 recesiva ligada al cromosoma X o ictiosis ligada al cromosoma X (deficiencia de sulfatasa esteroidea).

5 10. Un polipéptido para su uso en el tratamiento de una deficiencia de sulfatasa según cualquiera de las reivindicaciones 1 a 4, 8 o 9, o el uso de un polipéptido para preparar un medicamento para su uso en el tratamiento de una deficiencia de sulfatasa según cualquiera de las reivindicaciones 1 a 4, 8 o 9, o un ácido nucleico o vector que codifica un polipéptido para su uso en el tratamiento de una deficiencia de sulfatasa según cualquiera de las reivindicaciones 5 a 9, o el uso de un ácido nucleico o vector que codifica dicho polipéptido para su uso en el  
10 tratamiento de una deficiencia de sulfatasa según cualquiera de las reivindicaciones 5 a 9; en el que el polipéptido es una FGE eucariota, de mamífero o humana.

11. Un agente que se une selectivamente a un polipéptido como se describe en cualquiera de las reivindicaciones 1 a 4, para su uso en un método de diagnóstico o pronóstico de una deficiencia de sulfatasa que se realiza en el cuerpo humano o animal; en el que el agente es un anticuerpo, un fragmento de anticuerpo o un receptor de la superficie celular.  
15

12. El agente como se describe en la reivindicación 11 para el uso descrito en ella, en el que la deficiencia de sulfatasa está seleccionada del grupo que consiste en: deficiencia múltiple de sulfatasas, mucopolisacaridosis II (MPS II; síndrome de Hunter), mucopolisacaridosis IIIA (MPS IIIA; síndrome de Sanfilippo A), mucopolisacaridosis VIII (MPS VIII), mucopolisacaridosis IVA (MPS IVA; síndrome de Morquio A), mucopolisacaridosis VI (MPS VI; síndrome de Maroteaux-Lamy), leucodistrofia metacromática (MLD), condrod displasia punctata 1 recesiva ligada al cromosoma X o ictiosis ligada al cromosoma X (deficiencia de sulfatasa esteroidea); y en el que el agente es un anticuerpo, un fragmento de anticuerpo, o un receptor de la superficie celular.  
20

13. Un método de diagnóstico o pronóstico de una deficiencia de sulfatasa que comprende poner en contacto una muestra biológica obtenida de un sujeto con un agente que se une selectivamente a un polipéptido como se describe en cualquiera de las reivindicaciones 1 a 4; en el que el agente es un anticuerpo, un fragmento de anticuerpo o un receptor de la superficie celular.  
25

14. Un método de diagnóstico o pronóstico según la reivindicación 13; en el que la deficiencia de sulfatasa está seleccionada del grupo que consiste en: deficiencia múltiple de sulfatasas, mucopolisacaridosis II (MPS II; síndrome de Hunter), mucopolisacaridosis IIIA (MPS IIIA; síndrome de Sanfilippo A), mucopolisacaridosis VIII (MPS VIII), mucopolisacaridosis IVA (MPS IVA; síndrome de Morquio A), mucopolisacaridosis VI (MPS VI; síndrome de Maroteaux-Lamy), leucodistrofia metacromática (MLD), condrod displasia punctata 1 recesiva ligada al cromosoma X o ictiosis ligada al cromosoma X (deficiencia de sulfatasa esteroidea).  
30  
35

15. Un polipéptido para su uso en el tratamiento de una deficiencia de sulfatasa según cualquiera de las reivindicaciones 1 a 4, 8 o 9, o el uso de un polipéptido para preparar un medicamento para su uso en el tratamiento de una deficiencia de sulfatasa según cualquiera de las reivindicaciones 1 a 4, 8 o 9, o un ácido nucleico o vector que codifica un polipéptido para su uso en el tratamiento de una deficiencia de sulfatasa según cualquiera de las reivindicaciones 5 a 9, o el uso de un ácido nucleico o vector que codifica dicho polipéptido para su uso en el tratamiento de una deficiencia de sulfatasa según cualquiera de las reivindicaciones 5 a 9; en el que dicho polipéptido, ácido nucleico, vector o agente se proporciona en una composición farmacéutica junto con un vehículo farmacéuticamente aceptable.  
40  
45

50

55

60

65

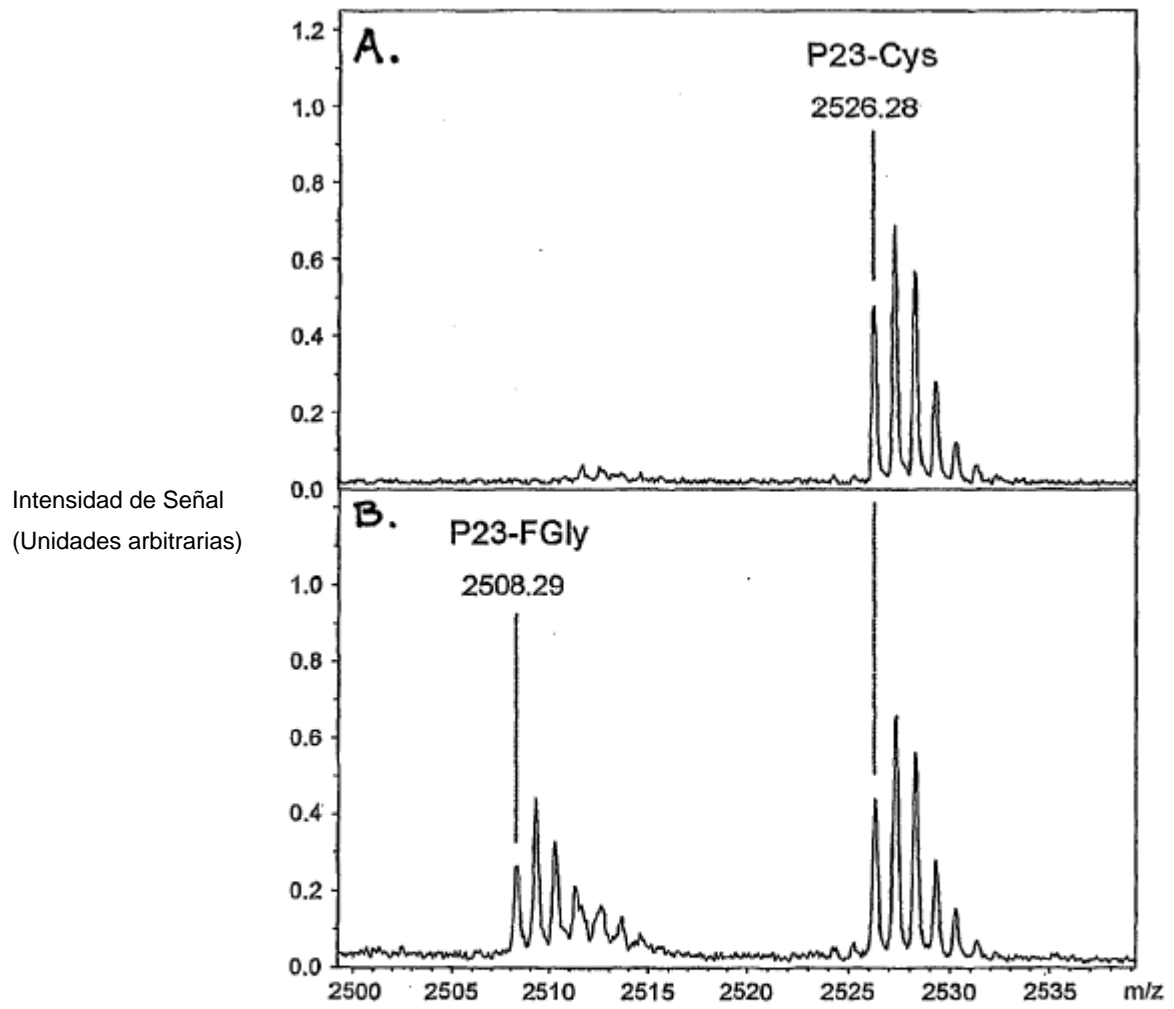


Fig. 1

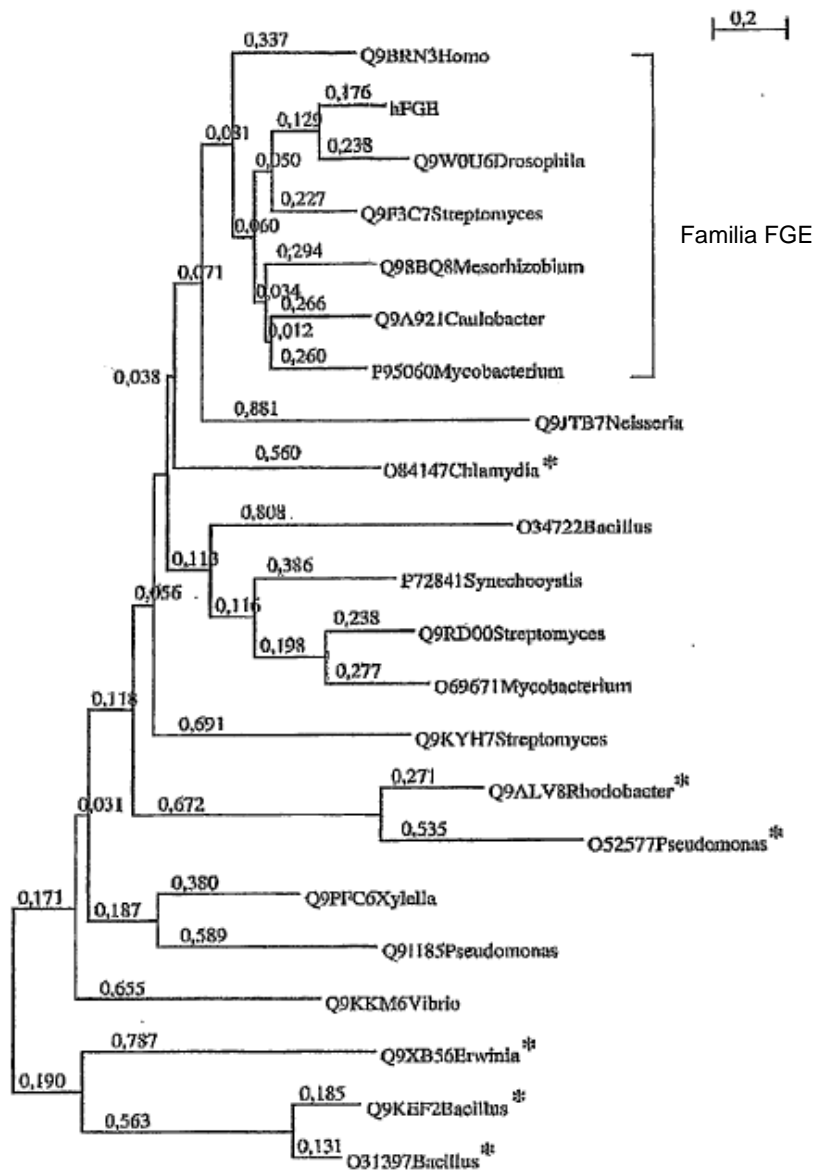


Fig. 2



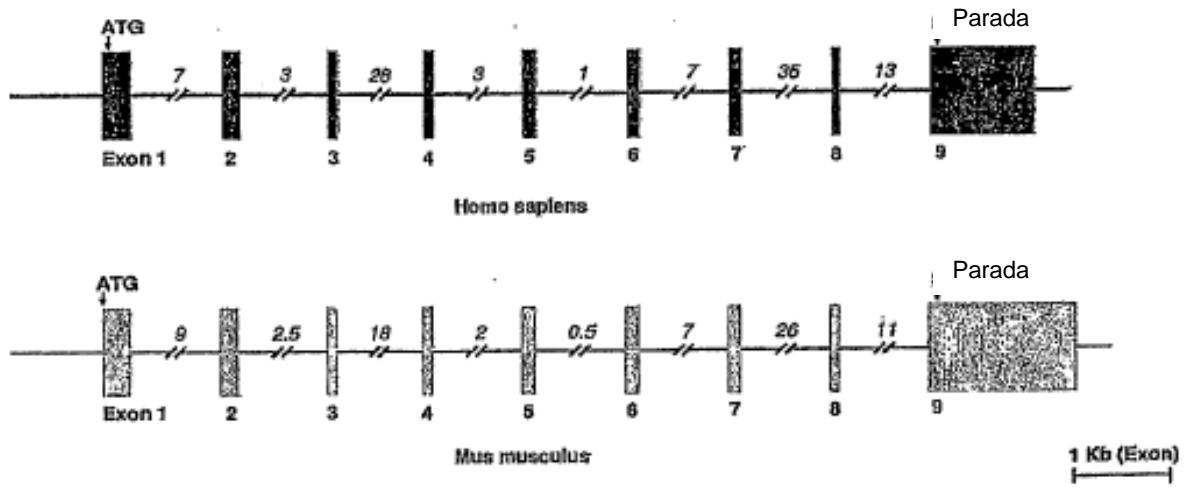


Fig. 3

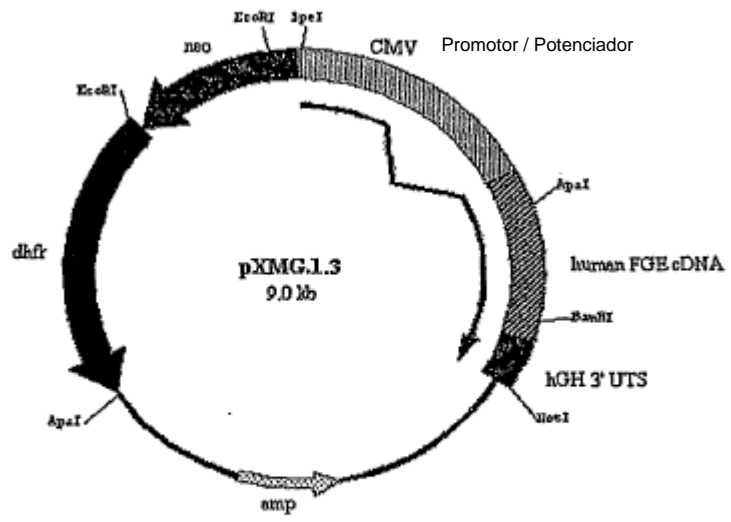


Fig. 4

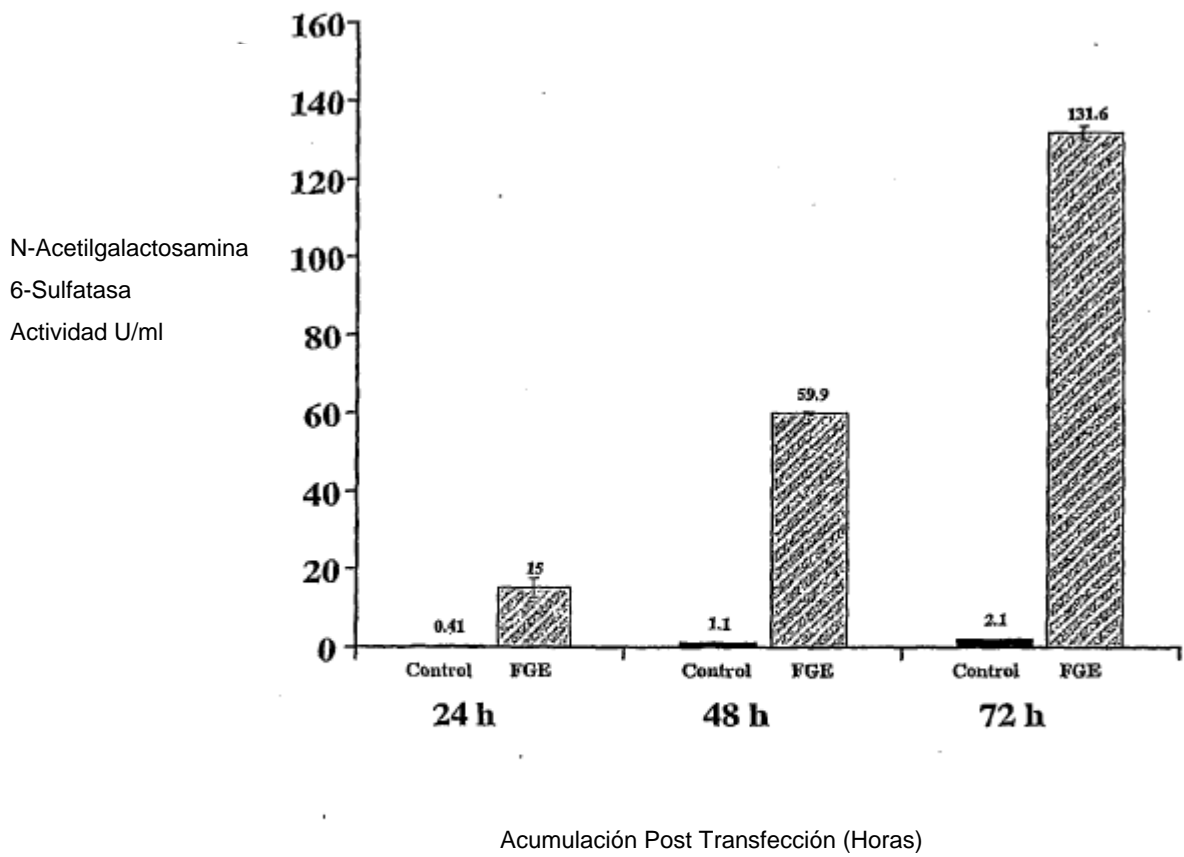
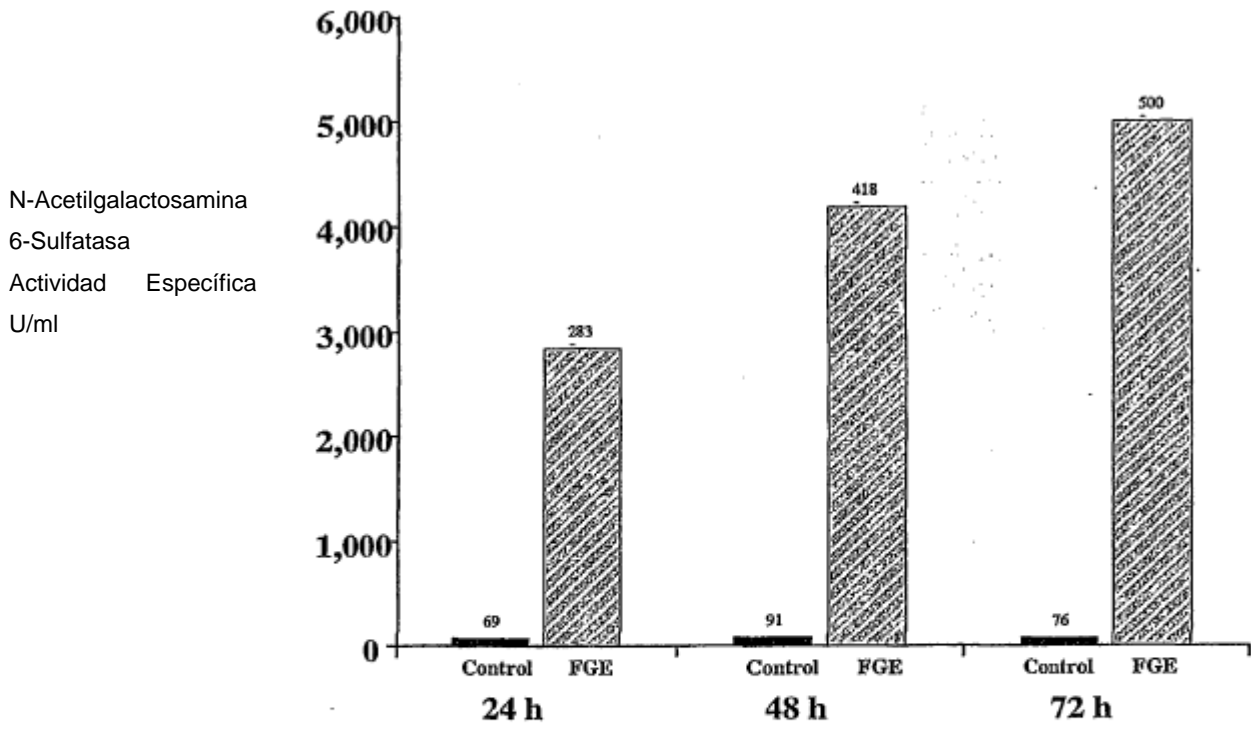


Fig. 5



Acumulación Post Transfección (Horas)

Fig. 6

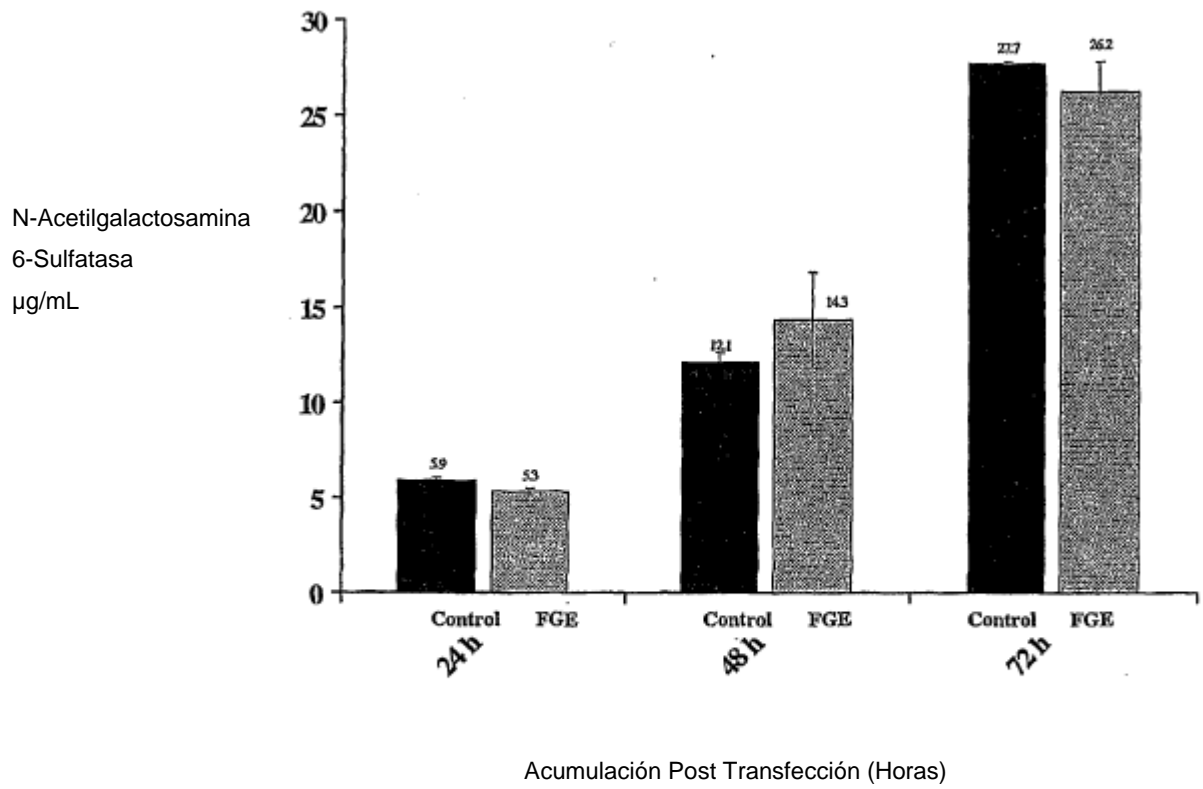


Fig. 7

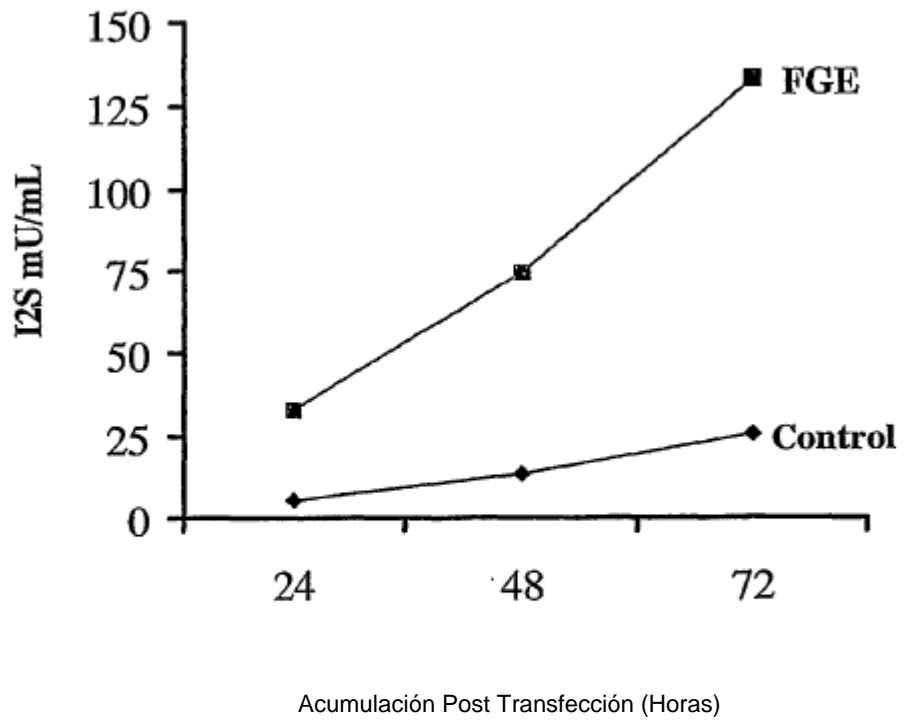


Fig. 8

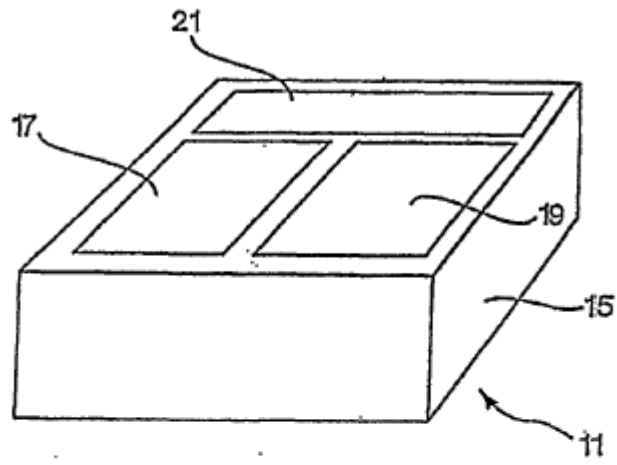


Fig. 9