

19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 380 147**

51 Int. Cl.:

**C12N 9/02** (2006.01)

**C12N 15/52** (2006.01)

**C12Q 1/68** (2006.01)

**A61K 38/44** (2006.01)

**A61K 31/7088** (2006.01)

**G01N 33/68** (2006.01)

12

## TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Número de solicitud europea: **04709824 .9**

96 Fecha de presentación: **10.02.2004**

97 Número de publicación de la solicitud: **1592786**

97 Fecha de publicación de la solicitud: **09.11.2005**

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

30 Prioridad:  
**11.02.2003 US 447747 P**

45 Fecha de publicación de la mención BOPI:  
**09.05.2012**

45 Fecha de la publicación del folleto de la patente:  
**09.05.2012**

73 Titular/es:  
**Shire Human Genetic Therapies, Inc.  
300 Shire Way  
Lexington MA 02421, US**

72 Inventor/es:  
**VON FIGURA, Kurt;  
SCHMIDT, Bernhard;  
DIERKS, Thomas;  
HEARTLEIN, Michael, W.;  
BALLABIO, Andrea y  
COSMA, Maria, Pía**

74 Agente/Representante:  
**Linage González, Rafael**

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).

ES 2 380 147 T3

## DESCRIPCIÓN

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

Esta invención se refiere, entre otras cosas, a métodos y composiciones para el diagnóstico y tratamiento de deficiencia múltiple de sulfatasas (MSD) así como a otras deficiencias de sulfatasa. Más específicamente, la invención usa moléculas que modulan modificaciones postraduccionales en sulfatasas. Tales modificaciones son esenciales para una función sulfatasa apropiada.

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 determinadas procariontes), en Cp, proporcionando L-C<sub>α</sub>-formilglicina (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 endoplasmático 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 (SEQ ID NO: 32), presente en la región N-terminal 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).

Hasta la fecha se han identificado trece genes de sulfatasa en seres humanos. Codifican para enzimas con diferente especificidad de sustrato y ubicación subcelular tales como lisosomas, aparato de golgi y RE. Cuatro de esos genes, ARSC, ARSD, ARSE y ARSF, que codifican para arilsulfatasa C, D, E y F, respectivamente, se encuentran ubicados dentro de la misma región cromosómica (Xp22.3). Comparten una similitud de secuencia significativa y una organización genómica casi idéntica, lo 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 de seres humanos se ve destacada por la identificación de al menos ocho enfermedades monogénicas humanas provocadas por la deficiencia de actividades sulfatasa individuales. La mayoría de estos estados son trastornos de almacenamiento lisosomal 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 II, 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: Mc 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 lisosomales. Incluyen ictiosis ligada al cromosoma X, un trastorno cutáneo debido a deficiencia de sulfatasa esteroidea (STS/ARSC), y condrodistrofia punctata, un trastorno que afecta a los huesos y cartílagos debido a deficiencia de arilsulfatasa E (ARSE). Las sulfatasas también están implicadas 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.

En un trastorno monogénico humano intrigante, la deficiencia múltiple de sulfatasas (MSD), todas las actividades 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 cuanto a actividades sulfatasa incluso tras la transfección con ADNc que codifican para sulfatasas humanas, lo que sugiere 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 postraduccional de sulfatasas era defectuosa en un paciente con MSD, lo que sugiere que este trastorno está provocado por una mutación en un gen, o genes, implicado 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 poca frecuencia de pacientes con MSD y la consiguiente falta de casos familiares adecuados para realizar un mapeo genético.

Esta invención proporciona métodos y composiciones para el diagnóstico y el tratamiento de deficiencia múltiple de sulfatasas (MIM 272200), y el tratamiento de otras deficiencias de sulfatasa. Más específicamente, se ha identificado

un gen que codifica para la enzima generadora de formilglucina (FGE), una enzima responsable de la modificación postraducciona única que se produce en sulfatasas que es esencial para la función sulfatasa (formación de L-C<sub>α</sub>-formilglucina; 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, 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. A la 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.

La invención proporciona un método para determinar el nivel de expresión de enzima generadora de C<sub>α</sub>-formilglucina (FGE) en un sujeto, comprendiendo el método medir *in vitro* en una muestra obtenida de un sujeto la actividad de generación de C<sub>α</sub>-formilglucina de un péptido o polipéptido que tiene una secuencia seleccionada del grupo que consiste en aminoácidos 34-374 de SEQ ID NO: 2, SEQ ID NO: 2,5, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76 y 78.

Proporciona además un método para identificar un agente útil en la modulación de la actividad de generación de C<sub>α</sub>-formilglucina, que comprende:

- (a) poner en contacto una molécula que tiene actividad de generación de C<sub>α</sub>-formilglucina con un agente candidato,
- (b) medir la actividad de generación de C<sub>α</sub>-formilglucina de la molécula, y
- (c) comparar la actividad medida de generación de C<sub>α</sub>-formilglucina de la molécula con un control para determinar si el agente candidato modula la actividad de generación de C<sub>α</sub>-formilglucina de la molécula;

en el que la molécula es una molécula de ácido nucleico que tiene la secuencia de nucleótidos expuesta como SEQ ID NO: 1, o es un producto de expresión de la misma.

También proporciona un método para determinar una deficiencia de sulfatasa en un sujeto caracterizada por una expresión aberrante de una molécula de ácido nucleico o un producto de expresión de la misma, que comprende:

monitorizar una muestra de un paciente para detectar un parámetro seleccionado del grupo que consiste en

- (i) una molécula de ácido nucleico que tiene la secuencia de nucleótidos expuesta como SEQ ID NO: 1,
- (ii) un polipéptido codificado por la molécula de ácido nucleico,
- (iii) un péptido derivado del polipéptido, y
- (iv) un anticuerpo que se une selectivamente al polipéptido o péptido,

como determinación de una deficiencia de sulfatasa en el sujeto.

La muestra puede ser, por ejemplo, un líquido biológico o un tejido.

La etapa de monitorizar puede incluir poner en contacto la muestra con un agente detectable seleccionado del grupo que consiste en:

- (a) una molécula de ácido nucleico que se hibrida selectivamente en condiciones rigurosas con la molécula de ácido nucleico de (i),
- (b) un anticuerpo que se une selectivamente al polipéptido de (ii), o el péptido de (iii), y
- (c) un polipéptido o péptido que se une al anticuerpo de (iv).

El anticuerpo, el polipéptido, el péptido o el ácido nucleico pueden marcarse con un marcador radioactivo o una enzima.

El método puede incluir someter a ensayo la muestra para detectar el péptido

La invención proporciona además un kit, que comprende un envase que contiene un agente y un control; en el que el agente es un anticuerpo o fragmento del mismo que se une selectivamente a un péptido o polipéptido tal como se describió anteriormente, o es un ácido nucleico que se une selectivamente a un ácido nucleico que codifica para

dicho péptido o polipéptido; y en el que el control es para compararlo con un valor medido de unión de dicho agente a dicho péptido, polipéptido o ácido nucleico.

El control puede ser un valor predeterminado para compararlo con dicho valor medido.

5

Puede comprender un epítipo del péptido o polipéptido.

La invención también incluye el uso de una molécula de ácido nucleico, péptido o polipéptido que aumenta la actividad de generación de  $C_{\alpha}$ -formilglucina en la preparación de un medicamento para tratar una deficiencia de sulfatasa en un sujeto; en el que:

10

(i) la molécula de ácido nucleico se selecciona del grupo que consiste en:

(a) el complemento de una molécula de ácido nucleico que se hibrida en condiciones rigurosas con una molécula que consiste en una secuencia de nucleótidos expuesta como SEQ ID NO: 1;

15

(b) una molécula de ácido nucleico que se diferencia de una molécula de ácido nucleico de (a) debido a la degeneración del código genético; y

20

(c) un ácido nucleico que tiene una secuencia seleccionada del grupo que consiste en: SEQ ID NO: 1, 3, 4, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77 y 80-87; y

(ii) el péptido o el polipéptido se selecciona del grupo que consiste en:

25

(a) un péptido o un polipéptido codificado por una molécula de ácido nucleico tal como se describió anteriormente; y

(b) un péptido o un polipéptido que tiene una secuencia seleccionada del grupo que consiste en: aminoácidos 34-374 de SEQ ID NO: 2, SEQ ID NO: 2, 5, 46, 48, 50, 52, 54, 56; 58, 60, 62, 64, 66, 68, 70, 72, 74, 76 y 78.

30

El medicamento descrito con respecto al uso mencionado anteriormente puede comprender además un agente seleccionado del grupo que consiste en: una molécula de ácido nucleico que codifica para 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; un producto de expresión de la molécula de ácido nucleico; y un fragmento del producto de expresión de la molécula de ácido nucleico.

35

El agente que aumenta la actividad de generación de  $C_{\alpha}$ -formilglucina puede producirse por una célula que expresa dicha molécula de ácido nucleico. La célula puede expresar una molécula de ácido nucleico de FGE exógena o endógena.

40

La invención incluye además una molécula de ácido nucleico, péptido o polipéptido que modula la actividad de generación de  $C_{\alpha}$ -formilglucina; para su uso en el tratamiento de una deficiencia de sulfatasa en un sujeto; siendo la molécula de ácido nucleico, péptido o polipéptido tal como se describió anteriormente.

45

Una composición farmacéutica que comprende la molécula de ácido nucleico, péptido, o polipéptido, junto con un portador farmacéuticamente aceptable también está dentro del alcance de la presente invención.

La invención también incluye un método para identificar un agente candidato útil en el tratamiento de una deficiencia de sulfatasa, que comprende:

50

determinar la expresión de un conjunto de moléculas de ácido nucleico *in vitro* en una célula o tejido en condiciones que, en ausencia de un agente candidato, permiten una primera cantidad de expresión del conjunto de moléculas de ácido nucleico; en el que el conjunto de moléculas de ácido nucleico comprende al menos una molécula de ácido nucleico que es una molécula de ácido nucleico tal como se describió anteriormente;

55

poner en contacto la célula o tejido con el agente candidato *in vitro*, y

detectar una cantidad de prueba de expresión del conjunto de moléculas de ácido nucleico;

60

en el que un aumento en la cantidad de prueba de expresión en presencia del agente candidato con respecto a la primera cantidad de expresión indica que el agente candidato es útil en el tratamiento de la deficiencia de sulfatasa.

Incluye además una matriz de moléculas de ácido nucleico en fase sólida que consiste en un conjunto de moléculas de ácido nucleico, productos de expresión de las mismas, o fragmentos de los mismos, cuando se fija a un sustrato sólido; codificando cada molécula de ácido nucleico para un polipéptido seleccionado del grupo que consiste en aminoácidos 34-374 de SEQ ID NO: 2, SEQ ID NO: 2, 5, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76

65

y 78, 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.

- 5 La matriz de moléculas de ácido nucleico en fase sólida puede comprender al menos una molécula de ácido nucleico de control.

El conjunto de moléculas de ácido nucleico puede comprender al menos una o al menos dos moléculas de ácido nucleico que codifican para un polipéptido seleccionado del grupo que consiste en: aminoácidos 34-374 de SEQ ID NO: 2, SEQ ID NO: 2, 5, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76 y 78, 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.

- 15 También se incluye dentro del alcance de la presente invención un alelo variante aislado de un gen de FGE humana que codifica para un polipéptido de FGE variante, comprendiendo el polipéptido:

una secuencia de aminoácidos que comprende al menos una variación en la SEQ ID NO: 2, en el que la al menos una variación comprende: Met1Arg; Met1Val; Ser155Pro; Cys218Tyr; Ala279Val; Arg327Stop; Cys336Arg; Arg345Cys; Arg349Trp; Arg349Gln; Ser359Stop; o una combinación de las mismas.

La invención incluye un polipéptido de FGE humana variante aislado que comprende una secuencia de aminoácidos con al menos una variación en la SEQ ID NO: 2, en el que la al menos una variación comprende: Met1Arg; Met1Val; Ser155Pro; Cys218Tyr; Ala279Val; Arg327Stop; Cys336Arg; Arg345Cys; Arg349Trp; Arg349Gln; Ser359Stop; o una combinación de las mismas.

Un anticuerpo que tiene tal polipéptido de FGE humana variante como inmunógeno también está dentro del alcance de la presente invención.

- 30 El anticuerpo puede ser, por ejemplo, un anticuerpo policlonal, un anticuerpo monoclonal, un anticuerpo quimérico, o un anticuerpo marcado de manera detectable.

Si está marcado de manera detectable, el marcador detectable puede comprender, por ejemplo, un elemento radioactivo, un compuesto químico que fluoresce, o una enzima.

La invención incluye además un método de producción de una sulfatasa activada, comprendiendo el método:

(a) poner en contacto la sulfatasa con una enzima generadora de  $C_{\alpha}$ -formilglicina (FGE) *in vitro*; en el que la enzima generadora de  $C_{\alpha}$ -formilglicina es un polipéptido o péptido tal como se describe en la reivindicación 1; o

(b) proporcionar una célula que coexpresa una sulfatasa y una enzima generadora de  $C_{\alpha}$ -formilglicina (FGE) de modo que se produce una sulfatasa activada dentro de la célula; en el que la célula comprende ARN o ADN heterólogo que da como resultado una expresión aumentada de la sulfatasa activada, con respecto a lo que se produciría en ausencia del ARN o ADN heterólogo; y en el que la enzima generadora de  $C_{\alpha}$ -formilglicina es un polipéptido o péptido tal como se describió anteriormente.

La sulfatasa puede seleccionarse, por ejemplo, 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.

La deficiencia de sulfatasa a la que se hace referencia con respecto a los métodos o usos descritos anteriormente puede seleccionarse del grupo que consiste en: deficiencia múltiple de sulfatasas, mucopolisacaridosis II, mucopolisacaridosis IIIA, mucopolisacaridosis IVA, mucopolisacaridosis VI, mucopolisacaridosis VIII, leucodistrofia metacromática, condrodistrofia punctata I recesiva ligada al cromosoma X y ictiosis ligada al cromosoma X.

#### Breve descripción de las secuencias

SEQ ID NO: 1 es la secuencia de nucleótidos del ADNc de FGE humana.

SEQ ID NO: 2 es la secuencia de aminoácidos predicha del producto de traducción de ADNc de FGE humana (SEQ ID NO: 1).

SEQ ID NO: 3 es la secuencia de nucleótidos del ADNc de FGE humana que codifica para el polipéptido de SEQ ID NO: 2 (es decir, nucleótidos 20-1141 de SEQ ID NO: 1).

SEQ ID NO: 4 es la secuencia de nucleótidos con n.º de registro de GenBank AK075459.

SEQ ID NO: 5 es la secuencia de aminoácidos predicha del producto de traducción de SEQ ID NO: 4, un producto de proteína sin nombre que tiene el n.º de registro de GenBank BAC11634.

- 5 SEQ ID NO: 6 es la secuencia de nucleótidos del ADNc de iduronato 2-sulfatasa humana (n.º de registro de GenBank M58342).

SEQ ID NO: 7 es la secuencia de aminoácidos predicha del producto de traducción de ADNc de iduronato 2-sulfatasa humana (SEQ ID NO: 6).

- 10 SEQ ID NO: 8 es la secuencia de nucleótidos del ADNc de sulfamidasa humana (n.º de registro de GenBank U30894).

- 15 SEQ ID NO: 9 es la secuencia de aminoácidos predicha del producto de traducción de ADNc de sulfamidasa humana (SEQ ID NO: 8).

SEQ ID NO: 10 es la secuencia de nucleótidos del ADNc de N-acetilgalactosamina 6-sulfatasa humana (n.º de registro de GenBank U06088).

- 20 SEQ ID NO: 11 es la secuencia de aminoácidos predicha del producto de traducción de ADNc de N-acetilgalactosamina 6-sulfatasa humana (SEQ ID NO: 10).

SEQ ID NO: 12 es la secuencia de nucleótidos del ADNc de N-acetilglucosamina 6-sulfatasa humana (n.º de registro de GenBank Z12173).

- 25 SEQ ID NO: 13 es la secuencia de aminoácidos predicha del producto de traducción de ADNc de N-acetilglucosamina 6-sulfatasa humana (SEQ ID NO: 12).

- 30 SEQ ID NO: 14 es la secuencia de nucleótidos del ADNc de arilsulfatasa A humana (n.º de registro de GenBank X52151).

SEQ ID NO: 15 es la secuencia de aminoácidos predicha del producto de traducción de ADNc de arilsulfatasa A humana (SEQ ID NO: 14).

- 35 SEQ ID NO: 16 es la secuencia de nucleótidos del ADNc de arilsulfatasa B humana (n.º de registro de GenBank J05225).

SEQ ID NO: 17 es la secuencia de aminoácidos predicha del producto de traducción de ADNc de arilsulfatasa B humana (SEQ ID NO: 16).

- 40 SEQ ID NO: 18 es la secuencia de nucleótidos del ADNc de arilsulfatasa C humana (n.º de registro de GenBank J04964).

- 45 SEQ ID NO: 19 es la secuencia de aminoácidos predicha del producto de traducción de ADNc de arilsulfatasa C humana (SEQ ID NO: 18).

SEQ ID NO: 20 es la secuencia de nucleótidos del ADNc de arilsulfatasa D humana (n.º de registro de GenBank X83572).

- 50 SEQ ID NO: 21 es la secuencia de aminoácidos predicha del producto de traducción de ADNc de arilsulfatasa D humana (SEQ ID NO: 20).

SEQ ID NO: 22 es la secuencia de nucleótidos del ADNc de arilsulfatasa E humana (n.º de registro de GenBank X83573).

- 55 SEQ ID NO: 23 es la secuencia de aminoácidos predicha del producto de traducción de ADNc de arilsulfatasa E humana (SEQ ID NO: 22).

- 60 SEQ ID NO: 24 es la secuencia de nucleótidos del ADNc de arilsulfatasa F humana (n.º de registro de GenBank X97868).

SEQ ID NO: 25 es la secuencia de aminoácidos predicha del producto de traducción de ADNc de arilsulfatasa F humana (SEQ ID NO: 24).

- 65 SEQ ID NO: 26 es la secuencia de nucleótidos del ADNc de arilsulfatasa G humana (n.º de registro de GenBank BC012375).

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

5 SEQ ID NO: 28 es la secuencia de nucleótidos del ADNc de HSulf-1 (n.º de registro de GenBank AY101175).

SEQ ID NO: 29 es la secuencia de aminoácidos predicha del producto de traducción de ADNc de HSulf-1 (SEQ ID NO: 28).

10 SEQ ID NO: 30 es la secuencia de nucleótidos del ADNc de HSulf-2 (n.º de registro de GenBank AY101176).

SEQ ID NO: 31 es la secuencia de aminoácidos predicha del producto de traducción de ADNc de HSulf-2 (SEQ ID NO: 30).

15 SEQ ID NO: 32 es el hexapéptido altamente conservado L/V-FGly-X-P-S-R presente en sulfatasas.

SEQ ID NO: 33 es un sustrato de formación de FGly sintético; su secuencia primaria se deriva de arilsulfatasa A humana.

20 SEQ ID NO: 34 es el oligopéptido reorganizado al azar PVSLPTRSCAALLTGR.

SEQ ID NO: 35 es el oligopéptido Ser69 PVSLSTPSRAALLTGR.

SEQ ID NO: 36 es un cebador específico para FGE humana 1199nc.

25 SEQ ID NO: 37 es un cebador directo específico para FGE humana 1c.

SEQ ID NO: 38 es un cebador inverso específico para FGE humana 1182c.

30 SEQ ID NO: 39 es un cebador específico para 5'-FGE humana que contiene un sitio de EcoRI.

SEQ ID NO: 40 es un cebador específico para HA.

SEQ ID NO: 41 es un cebador específico para c-myc.

35 SEQ ID NO: 42 es un cebador específico para RGS-His6.

SEQ ID NO: 43 es un oligopéptido tríptico SQNTPDSSASNLGFR de una preparación de FGE humana.

40 SEQ ID NO: 44 es un oligopéptido tríptico MVPIPAVGFTMGTDPPQIK de una preparación de FGE humana.

SEQ ID NO: 45 es la secuencia de nucleótidos del parálogo de FGE2 humana (GenBank GI: 24308053).

45 SEQ ID NO: 46 es la secuencia de aminoácidos predicha del producto de traducción del parálogo de FGE2 humana (SEQ ID NO: 45).

SEQ ID NO: 47 es la secuencia de nucleótidos del parálogo de FGE de ratón (GenBank GI: 26344956).

50 SEQ ID NO: 48 es la secuencia de aminoácidos predicha del producto de traducción del parálogo de FGE de ratón (SEQ ID NO: 47).

SEQ ID NO: 49 es la secuencia de nucleótidos del ortólogo de FGE de ratón (GenBank GI: 22122361).

55 SEQ ID NO: 50 es la secuencia de aminoácidos predicha del producto de traducción del ortólogo de FGE de ratón (SEQ ID NO: 49).

SEQ ID NO: 51 es la secuencia de nucleótidos del ortólogo de FGE de mosca de la fruta (GenBank GI: 20130397).

60 SEQ ID NO: 52 es la secuencia de aminoácidos predicha del producto de traducción del ortólogo de FGE de mosca de la fruta (SEQ ID NO: 51).

SEQ ID NO: 53 es la secuencia de nucleótidos del ortólogo de FGE de mosquito (GenBank GI: 21289310).

65 SEQ ID NO: 54 es la secuencia de aminoácidos predicha del producto de traducción del ortólogo de FGE de mosquito (SEQ ID NO: 53).

SEQ ID NO: 55 es la secuencia de nucleótidos del ortólogo de FGE de *S. coelicolor* estrechamente relacionado (GenBank GI: 21225812).

5 SEQ ID NO: 56 es la secuencia de aminoácidos predicha del producto de traducción del ortólogo de FGE de *S. coelicolor* (SEQ ID NO: 55).

SEQ ID NO: 57 es la secuencia de nucleótidos del ortólogo de FGE de *C. efficiens* estrechamente relacionado (GenBank GI: 25028125).

10 SEQ ID NO: 58 es la secuencia de aminoácidos predicha del producto de traducción del ortólogo de FGE de *C. efficiens* (SEQ ID NO: 57).

SEQ ID NO: 59 es la secuencia de nucleótidos del ortólogo de FGE de *N. aromaticivorans* (GenBank GI: 23108562).

15 SEQ ID NO: 60 es la secuencia de aminoácidos predicha del producto de traducción del ortólogo de FGE de *N. aromaticivorans* (SEQ ID NO: 59).

SEQ ID NO: 61 es la secuencia de nucleótidos del ortólogo de FGE de *M. loti* (GenBank GI: 13474559).

20 SEQ ID NO: 62 es la secuencia de aminoácidos predicha del producto de traducción del ortólogo de FGE de *M. loti* (SEQ ID NO: 61).

SEQ ID NO: 63 es la secuencia de nucleótidos del ortólogo de FGE de *B. fungorum* (GenBank GI: 22988809).

25 SEQ ID NO: 64 es la secuencia de aminoácidos predicha del producto de traducción del ortólogo de FGE de *B. fungorum* (SEQ ID NO: 63).

SEQ ID NO: 65 es la secuencia de nucleótidos del ortólogo de FGE de *S. meliloti* (GenBank GI: 16264068).

30 SEQ ID NO: 66 es la secuencia de aminoácidos predicha del producto de traducción del ortólogo de FGE de *S. meliloti* (SEQ ID NO: 65).

SEQ ID NO: 67 es la secuencia de nucleótidos del ortólogo de FGE de *Microscilla sp.* (GenBank GI: 14518334).

35 SEQ ID NO: 68 es la secuencia de aminoácidos predicha del producto de traducción del ortólogo de FGE de *Microscilla sp.* (SEQ ID NO: 67).

SEQ ID NO: 69 es la secuencia de nucleótidos del ortólogo de FGE de *P. putida* KT2440 (GenBank GI: 26990068).

40 SEQ ID NO: 70 es la secuencia de aminoácidos predicha del producto de traducción del ortólogo de FGE de *P. putida* KT2440 (SEQ ID NO: 69).

SEQ ID NO: 71 es la secuencia de nucleótidos del ortólogo de FGE de *R. metallidurans* (GenBank GI: 22975289).

45 SEQ ID NO: 72 es la secuencia de aminoácidos predicha del producto de traducción del ortólogo de FGE de *R. metallidurans* (SEQ ID NO: 71).

SEQ ID NO: 73 es la secuencia de nucleótidos del ortólogo de FGE de *P. marinus* (GenBank GI: 23132010).

50 SEQ ID NO: 74 es la secuencia de aminoácidos predicha del producto de traducción del ortólogo de FGE de *P. marinus* (SEQ ID NO: 73).

SEQ ID NO: 75 es la secuencia de nucleótidos del ortólogo de FGE de *C. crescentus* CB15 (GenBank GI: 16125425).

55 SEQ ID NO: 76 es la secuencia de aminoácidos predicha del producto de traducción del ortólogo de FGE de *C. crescentus* CB 15 (SEQ ID NO: 75).

60 SEQ ID NO: 77 es la secuencia de nucleótidos del ortólogo de FGE de *M. tuberculosis* Ht37Rv (GenBank GI: 15607852).

SEQ ID NO: 78 es la secuencia de aminoácidos predicha del producto de traducción del ortólogo de FGE de *M. tuberculosis* Ht37Rv (SEQ ID NO: 77).

65 SEQ ID NO: 79 es el heptapéptido altamente conservado presente en el subdominio 3 de parálogos y ortólogos de FGE.



SEQ ID NO: 80 es la secuencia de nucleótidos del fragmento EST ortólogo de FGE que tiene n.º de registro de GenBank: CA379852.

5 SEQ ID NO: 81 es la secuencia de nucleótidos del fragmento EST ortólogo de FGE que tiene n.º de registro de GenBank: AI721440.

SEQ ID NO: 82 es la secuencia de nucleótidos del fragmento EST ortólogo de FGE que tiene n.º de registro de GenBank: BJ505402.

10 SEQ ID NO: 83 es la secuencia de nucleótidos del fragmento EST ortólogo de FGE que tiene n.º de registro de GenBank: BJ054666.

15 SEQ ID NO: 84 es la secuencia de nucleótidos del fragmento EST ortólogo de FGE que tiene n.º de registro de GenBank: AL892419.

SEQ ID NO: 85 es la secuencia de nucleótidos del fragmento EST ortólogo de FGE que tiene n.º de registro de GenBank: CA064079.

20 SEQ ID NO: 86 es la secuencia de nucleótidos del fragmento EST ortólogo de FGE que tiene n.º de registro de GenBank: BF189614.

SEQ ID NO: 87 es la secuencia de nucleótidos del fragmento EST ortólogo de FGE que tiene n.º de registro de GenBank: AV609121.

25 SEQ ID NO: 88 es la secuencia de nucleótidos del ADNc de HSulf-3.

SEQ ID NO: 89 es la secuencia de aminoácidos predicha del producto de traducción de ADNc de HSulf-3 (SEQ ID NO: 88).

30 SEQ ID NO: 90 es la secuencia de nucleótidos del ADNc de HSulf-4.

SEQ ID NO: 91 es la secuencia de aminoácidos predicha del producto de traducción de ADNc de HSulf-4 (SEQ ID NO: 90).

35 SEQ ID NO: 92 es la secuencia de nucleótidos del ADNc de HSulf-5.

SEQ ID NO: 93 es la secuencia de aminoácidos predicha del producto de traducción de ADNc de HSulf-5 (SEQ ID NO: 92).

40 SEQ ID NO: 94 es la secuencia de nucleótidos del ADNc de HSulf-6.

SEQ ID NO: 95 es la secuencia de aminoácidos predicha del producto de traducción de ADNc de HSulf-6 (SEQ ID NO: 94).

#### 45 **Breve descripción de los dibujos**

Figura 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.

50 Figura 2: un árbol filogenético derivado de una alineación de FGE humana y 21 proteínas de la semilla DUF323 en PFAM.

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

Figura 4: diagrama que muestra un mapa de plásmido pXMG.1.3 de expresión de FGE.

60 Figura 5: gráfico de barras que representa actividad N-acetilgalactosamina 6-sulfatasa en células 36F transfectadas de manera transitoria con plásmido de expresión de FGE.

Figura 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.

65 Figura 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.

Figura 8: gráfico que representa actividad iduronato 2-sulfatasa en células 30C6 transfectadas de manera transitoria con plásmido de expresión de FGE.

Figura 9: representa un kit que incluye características de la presente invención.

Los inventores han descubierto el gen que codifica para enzima generadora de formilglicina (FGE), una enzima responsable de la modificación postraduccional única que se produce en sulfatasas que es esencial para la función sulfatasa: la formación de L-C<sub>α</sub>-formilglicina (también conocida como FGly y/o ácido 2-amino-3-oxopropanoico). Se ha descubierto, de manera inesperada, que mutaciones en el gen de FGE conducen al desarrollo de deficiencia múltiple de sulfatasas (MSD) en sujetos. También se ha descubierto, de manera inesperada, 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 estadounidenses con números de publicación 20030073118, 20030147875, 20030148920, 20030162279 y 20030166283. A la vista de estos descubrimientos, las moléculas descritas en el presente documento pueden usarse en el diagnóstico y/o tratamiento de deficiencia múltiple de sulfatasas, así como el tratamiento de otras deficiencias de sulfatasa.

Se proporcionan métodos para usar moléculas en el diagnóstico de deficiencia múltiple de sulfatasas.

Adicionalmente, también se proporcionan métodos para usar esas moléculas *in vivo* o *in vitro* con el fin de modular la formación de FGly en sulfatasas, métodos para tratar estados asociados con tal modificación, y composiciones útiles en la preparación de preparaciones terapéuticas para el tratamiento de deficiencia múltiple de sulfatasas así como otras deficiencias de sulfatasa.

También se proporcionan polipéptidos que modulan la formación de FGly en sulfatasas, ácidos nucleicos aislados que codifican para esos polipéptidos, modificaciones y variantes funcionales de los anteriores, fragmentos útiles de los anteriores, así como productos terapéuticos y productos de diagnóstico, métodos de investigación, composiciones y herramientas referentes a los mismos.

“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 tal como se describe en otra parte en el presente documento, o un oligopéptido sintético (véase, por ejemplo, SEQ ID NO: 33, y los ejemplos). El sustrato contiene preferiblemente el hexapéptido conservado de SEQ ID NO: 32 [L/V-C(S)-X-P-S-R]. Métodos para someter a ensayo la formación de FGly son tal como se describe 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 (véase, por ejemplo, los ejemplos). Una “molécula”, tal como se usa en el presente documento, abarca tanto “ácidos nucleicos” como “polipéptidos”. Las moléculas de FGE pueden 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<sub>α</sub>-formilglicina, tal 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 tal como se describe en otra parte en el presente documento), dado que la regulación por incremento de cualquiera de estos procesos da como resultado 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<sub>α</sub>-formilglicina también se refiere a prevenir o inhibir la degradación de FGE (por ejemplo, mediante aumento de la ubiquitinización), regulación por disminución, etc., dando como resultado, por ejemplo, una t<sub>1/2</sub> (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<sub>α</sub>-formilglicina.

“Expresión”, tal 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<sub>α</sub>-formilglicina de la molécula).

Un método implica la clonación de un ADNc que codifica para FGE. FGE es una molécula de ácido nucleico aislada

que comprende una molécula de ácido nucleico de SEQ ID NO: 1, y codifica para un polipéptido con actividad de generación de C<sub>α</sub>-formilglicina. La secuencia del ADNc de FGE humana se presenta como SEQ ID NO: 1, y la secuencia de aminoácidos predicha de este producto de proteína codificado por ADNc se presenta como SEQ ID NO: 2.

5 Tal 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.

10 Se dan a conocer un polipéptido de FGE aislado, el ADNc que codifica para este polipéptido, modificaciones y variantes funcionales de los anteriores, fragmentos útiles de los anteriores, así como productos de diagnóstico y productos terapéuticos referentes a los mismos.

15 Tal 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, tal 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 dado a conocer 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 sólo un diminuto porcentaje del material en la célula en la que reside. Sin embargo, un ácido nucleico de este tipo está aislado tal como se usa el término en el presente documento porque se manipula fácilmente mediante técnicas convencionales conocidas por los expertos en la técnica.

20 Tal 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 propósitos descritos 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.

30 Moléculas de ácido nucleico aisladas que codifican para un polipéptido de FGE que tiene actividad de generación de C<sub>α</sub>-formilglicina incluyen: (a) moléculas de ácido nucleico que se hibridan en condiciones rigurosas con una molécula que consiste en un ácido nucleico de SEQ ID NO: 1 y que codifican para un polipéptido de FGE que tiene actividad de generación de C<sub>α</sub>-formilglicina, (b) delecciones, adiciones y sustituciones de (a) que codifican para un polipéptido de FGE respectivo que tiene actividad de generación de C<sub>α</sub>-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", tal como se usa en el presente documento, incluye "cadenas complementarias de longitud completa o cadenas complementarias al 100% de (a), (b) o (c)".

45 Se dan a conocer homólogos y alelos de los ácidos nucleicos de FGE que también tienen actividad de generación de C<sub>α</sub>-formilglicina. Los homólogos, tal como se describen en el presente documento, incluyen las moléculas identificadas en otra parte en el presente documento (véanse por ejemplo, SEQ ID NO: 4, 5, 45-78, y 80-87) es decir ortólogos y parálogos. Pueden identificarse homólogos adicionales siguiendo las enseñanzas en el presente documento así como mediante técnicas convencionales. Dado que los homólogos de FGE descritos en el presente documento comparten todos actividad de generación de C<sub>α</sub>-formilglicina, pueden usarse de manera intercambiable con la molécula de FGE humana en todos los aspectos descritos en el presente documento.

50 Las secuencias de ácido nucleico que codifican para polipéptidos de FGE pueden hibridarse con el complemento de una molécula de ácido nucleico que consiste en la región codificante de SEQ ID NO: 1, en condiciones rigurosas. En una realización importante, la expresión "condiciones rigurosas", tal como se usa en el presente documento, se refiere a parámetros con los que está familiarizada la técnica. Con ácidos nucleicos, se dice que las condiciones de hibridación son rigurosas normalmente en condiciones de baja fuerza iónica y una temperatura justo por debajo de la temperatura de fusión (T<sub>f</sub>) del complejo de híbrido de ADN (normalmente, aproximadamente 3°C por debajo de la T<sub>f</sub> del híbrido). Una rigurosidad superior proporciona una correlación más específica entre la secuencia sonda y la diana. Las condiciones rigurosas usadas en la hibridación de ácidos nucleicos se conocen bien en la técnica y pueden encontrarse 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. Un ejemplo de "condiciones rigurosas" es la hibridación a 65°C en 6 x SSC. Otro ejemplo de condiciones rigurosas es la hibridación a 65°C en tampón de hibridación que consiste en 3,5 x SSC, Ficoll al 0,02%, polivinilpirrolidona al 0,02%, albúmina sérica bovina al 0,02%, H<sub>2</sub>PO<sub>4</sub> [pH 7] 2,5 mM, SDS al 0,5%, EDTA 2 mM. (SSC es cloruro de sodio 0,15 M/citrato de sodio 0,15 M, pH 7; SDS es dodecilsulfato de sodio; y EDTA es ácido etilendiaminatetracético). Tras la hibridación, se lava la membrana sobre la que se transfiere el ADN a 2 x SSC a

temperatura ambiente y después a 0,1 x SSC/0,1 x SDS a temperaturas de hasta 68°C. En un ejemplo adicional, una alternativa al uso de una disolución de hibridación acuosa es el uso de una disolución de hibridación en formamida. Por tanto, pueden lograrse condiciones de hibridación rigurosas usando, por ejemplo, una disolución de formamida al 50% y 42°C. Hay otras condiciones, reactivos, etcétera, que pueden usarse, y darán como resultado un grado similar de rigurosidad. El experto en la técnica estará familiarizado con tales condiciones, y por tanto no se proporcionan en el presente documento. Sin embargo, se entenderá que el experto en la técnica podrá manipular las condiciones de una manera que permitan la clara identificación de homólogos y alelos de ácidos nucleicos de FGE descritos en el presente documento. El experto en la técnica también está familiarizado con la metodología para examinar células y bibliotecas para detectar la expresión de tales moléculas que entonces se aíslan de manera rutinaria, seguido por aislamiento de la molécula de ácido nucleico pertinente y secuenciación.

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 SEQ ID NO: 1 y SEQ ID NO: 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 SEQ ID NO: 1 y SEQ ID NO: 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 alineaciones 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 se abarcan los complementos de Watson-Crick de los ácidos nucleicos anteriores.

En la selección para detectar genes relacionados con FGE, tales como homólogos y alelos de FGE, puede realizarse una transferencia de tipo Southern usando las condiciones anteriores, junto con una sonda radioactiva. Tras 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 radioactividad para detectar la señal radioactiva.

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.

Ácidos nucleicos degenerados 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, resultará evidente para un experto en la técnica que puede emplearse cualquiera de los tripletes de nucleótidos que codifican para 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. De manera similar, tripletes de secuencias de nucleótidos que codifican para otros residuos de aminoácido incluyen, pero no se limitan a: CCA, CCC, CCG y CCT (codones de prolina); CGA, CGC, 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 de manera similar por múltiples secuencias de nucleótidos. Por tanto, pueden usarse ácidos nucleicos degenerados que se diferencian de los ácidos nucleicos aislados biológicamente en la secuencia de codones debido a la degeneración del código genético.

Se proporcionan fragmentos únicos aislados de SEQ ID NO: 1 o SEQ ID NO: 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 técnica 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 por las secuencias de nucleótidos seleccionadas del grupo que consiste en la SEQ ID NO: 4, y/u otras secuencias previamente publicadas en la fecha de presentación de esta solicitud.

Un fragmento que está completamente compuesto por la secuencia descrita en los depósitos de GenBank anteriores es una que no incluye ninguno de los nucleótidos únicos para las secuencias descritas en el presente documento. Por tanto, un fragmento único debe contener una secuencia de nucleótidos distinta de la secuencia exacta de esos 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 de tipo Southern y Northern para identificar tales ácidos nucleicos, o pueden usarse en ensayos de amplificación tales como los que emplean PCR.

Tal como saben los expertos en la técnica, se prefieren sondas grandes tales como 200, 250, 300 o más nucleótidos para determinados usos tales como transferencias de tipo 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, tal 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.

Tal como reconocerán los expertos en la técnica, el tamaño del fragmento único dependerá de su conservación en el código genético. Por tanto, algunas regiones de SEQ ID NO: 1 o SEQ ID NO: 3 y complementos requerirán que segmentos más largos sean únicos mientras que otras sólo 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 dada a conocer. Tal como se mencionó anteriormente, esta descripción pretende abarcar todos y cada uno de los fragmentos de cada secuencia, comenzando en el primer 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 tal como se describió anteriormente). Prácticamente cualquier segmento de la región de SEQ ID NO: 1 comenzando en el nucleótido 1 y terminando en el nucleótido 1180, o SEQ ID NO: 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 técnica están 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 aquéllas en bases de datos conocidas es todo lo que se necesita, aunque pueden realizarse hibridación confirmatoria *in vitro* y análisis de secuenciación.

Tal como se mencionó anteriormente, pueden usarse oligonucleótidos antisentido que se unen selectivamente a una molécula de ácido nucleico que codifica para un polipéptido de FGE para disminuir la actividad FGE.

Tal como se usa en el presente documento, la expresión "oligonucleótido antisentido" o "antisentido" describe un oligonucleótido que es un oligorribonucleótido, oligodesoxirribonucleótido, oligorribonucleótido modificado, u oligodesoxirribonucleótido modificado que se hibrida en condiciones fisiológicas con ADN que comprende un gen particular o con un transcrito de ARNm de ese gene y, de ese modo, inhibe la transcripción de ese gen y/o la traducción de ese ARNm. Las moléculas antisentido están diseñadas para interferir con la transcripción o traducción de un gen diana tras la hibridación con ese gen diana o transcrito. Los expertos en la técnica reconocerán que la longitud exacta del oligonucleótido antisentido y su grado de complementariedad con su diana dependerán de la diana específica seleccionada, incluyendo la secuencia de la diana y las bases particulares que comprende esa secuencia. Se prefiere que el oligonucleótido antisentido se construya y disponga para unirse selectivamente con la diana en condiciones fisiológicas, es decir, para hibridarse sustancialmente más a la secuencia diana que a cualquier otra secuencia en la célula diana en condiciones fisiológicas. Basándose en SEQ ID NO: 1 o en secuencias genómicas y/o de ADNc alélicas u homólogas, un experto en la técnica puede elegir fácilmente y sintetizar cualquiera de varias moléculas antisentido apropiadas. Con el fin de ser suficientemente selectivos y potentes para la inhibición, tales oligonucleótidos antisentido deben comprender al menos 10 y, más preferiblemente, al menos 15 bases consecutivas que son complementarias a la diana, aunque en determinados casos se han usado satisfactoriamente oligonucleótidos modificados de tan sólo 7 bases de longitud como oligonucleótidos antisentido (Wagner *et al.*, Nat. Med, 1995, 1(11):1116-1118; Nat. Biotech., 1996, 14:840-844). Lo más preferiblemente, los oligonucleótidos antisentido comprenden una secuencia complementaria de 20-30 bases. Aunque pueden elegirse oligonucleótidos que son antisentido para cualquier región del gen o transcritos de ARNm, en realizaciones preferidas los oligonucleótidos antisentido corresponden a sitios N-terminales o en sentido 5' tales como sitios de iniciación de la traducción, de iniciación de la transcripción o de promotor. Además, oligonucleótidos antisentido pueden dirigirse a regiones no traducidas en 3'. También se ha usado el direccionamiento a sitios de corte y empalme de ARNm en la técnica pero puede preferirse menos si se produce corte y empalme de ARNm alternativo. Además, el antisentido se dirige, preferiblemente, a sitios en los que no se espera una estructura secundaria de ARNm (véase, por ejemplo, Sainio *et al.*, Cell Mol. Neurobiol. 14(5):439-457, 1994) y a la que no se espera que se unan proteínas. Finalmente, aunque SEQ ID NO: 1 da a conocer una secuencia de ADNc, un experto habitual en la técnica puede derivar fácilmente el ADN genómico correspondiente para esta secuencia. Por tanto, pueden proporcionarse oligonucleótidos antisentido que son complementarios al ADN genómico correspondiente a SEQ ID NO: 1. De manera similar, se permiten oligonucleótidos antisentido para los ADN genómicos y ADNc de FGE homólogos o alélicos sin excesiva experimentación.

En un conjunto de realizaciones, los oligonucleótidos antisentido pueden estar compuestos por desoxirribonucleótidos "naturales", ribonucleótidos, o cualquier combinación de los mismos. Es decir, el extremo 5' de un nucleótido nativo y el extremo 3' de otro nucleótido nativo pueden estar unidos covalentemente, como en sistemas naturales, mediante un enlace internucleósido de fosfodiéster. Estos oligonucleótidos pueden prepararse mediante métodos reconocidos en la técnica que pueden llevarse a cabo de manera manual o mediante un

sintetizador automatizado. También pueden producirse de manera recombinante mediante vectores.

Sin embargo, en realizaciones preferidas los oligonucleótidos antisentido también pueden incluir oligonucleótidos "modificados". Es decir, los oligonucleótidos pueden modificarse de varias maneras que no impiden que se hibriden a su diana pero que potencian su estabilidad o direccionamiento o que de otro modo potencian su eficacia terapéutica.

La expresión "oligonucleótido modificado" tal como se usa en el presente documento describe un oligonucleótido en el que (1) al menos dos de sus nucleótidos están unidos covalentemente mediante un enlace internucleósido sintético (es decir, un enlace distinto de un enlace de fosfodiéster entre el extremo 5' de un nucleótido y el extremo 3' de otro nucleótido) y/o (2) se ha unido covalentemente al oligonucleótido un grupo químico que no está normalmente asociado con ácidos nucleicos. Enlaces internucleósido sintéticos preferidos son fosforotioatos, alquifosfonatos, fosforoditioatos, ésteres de fosfato, alquifosfonotioatos, fosforamidatos, carbamatos, carbonatos, triésteres de fosfato, acetamidatos, ésteres carboxilmetílicos y péptidos.

La expresión "oligonucleótido modificado" también abarca oligonucleótidos con una base y/o azúcar modificado covalentemente. Por ejemplo, oligonucleótidos modificados incluyen oligonucleótidos que tienen azúcares de estructura principal que están covalentemente unidos a grupos orgánicos de bajo peso molecular distintos de un grupo hidroxilo en la posición 3' y distintos de un grupo fosfato en la posición 5'. Por tanto, oligonucleótidos modificados pueden incluir un grupo ribosa 2'-O-alquilado. Además, oligonucleótidos modificados pueden incluir azúcares tales como arabinosa en vez de ribosa. Por tanto, se contemplan preparaciones farmacéuticas que contienen moléculas antisentido modificadas que son complementarias a, y pueden hibridarse en condiciones fisiológicas con, ácidos nucleicos que codifican para polipéptidos de FGE, junto con portadores farmacéuticamente aceptables. Pueden administrarse oligonucleótidos antisentido como parte de una composición farmacéutica. Una composición farmacéutica de este tipo puede incluir el oligonucleótido antisentido en combinación con cualquier portador convencional fisiológico y/o farmacéuticamente aceptable que se conoce en la técnica. Las composiciones deben ser estériles y contener una cantidad terapéuticamente eficaz de los oligonucleótidos antisentido en una unidad de peso o volumen adecuada para su administración a un paciente. La expresión "farmacéuticamente aceptable" significa un material no tóxico que no interfiere con la eficacia de la actividad biológica de los principios activos. La expresión "fisiológicamente aceptable" se refiere a un material no tóxico que es compatible con un sistema biológico tal como una célula, cultivo celular, tejido u organismo. Las características del portador dependerán de la vía de administración. Portadores fisiológicos y farmacéuticamente aceptables incluyen diluyentes, cargas, sales, tampones, estabilizadores, solubilizadores y otros materiales que se conocen bien en la técnica.

También se proporcionan métodos para aumentar la actividad de generación de C<sub>α</sub>-formilglicina en una célula. En realizaciones importantes, esto se logra mediante el uso de vectores ("vectores de expresión" y/o "vectores de direccionamiento").

"Vectores", tal 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 por 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 puede 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 sólo una única vez por huésped antes de que el huésped se reproduzca mediante 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 SEQ ID NO: 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 para proteínas que aumentan o disminuyen o bien su resistencia o bien su sensibilidad a antibióticos u otros compuestos, genes que codifican para enzimas cuyas actividades pueden detectarse mediante ensayos convencionales conocidos en la técnica (por ejemplo, β-galactosidasa o fosfatasa alcalina), y genes que afectan visiblemente al fenotipo de células, huéspedes, colonias o placas transformadas o transfectadas (por ejemplo, proteína verde fluorescente).

Un "vector de direccionamiento" es uno que contiene normalmente constructos/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 sentido 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 SEQ ID NO: 1) y otras secuencias necesarias para el direccionamiento del gen a una ubicación preferida en el genoma (por ejemplo, una ubicación transcripcionalmente

activa, por ejemplo en el sentido 3' de un promotor endógeno de un gen no relacionado). La construcción de vectores y constructos de direccionamiento se describe en detalle en las patentes estadounidenses 5.641.670 y 6.270.989.

5 Puede usarse prácticamente cualquier célula, procariota o eucariota, que puede transformarse 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 médula ósea), células del músculo y precursores de esos tipos de células somáticas incluyendo células madre embrionarias. Cuando las células deben usarse en terapia génica, se obtienen preferiblemente células primarias del individuo a quien se le 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 los constructos de ADN y métodos descritos en el presente documento 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. *et 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 mediante 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 describe una evaluación adicional de los tipos de células que pueden usarse en las patentes estadounidenses 5.641.670 y 6.270.989. También pueden usarse sistemas de transcripción libres de células en lugar de células.

Las células se mantienen en condiciones, tal como se conoce en la técnica, que dan como resultado 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 mediante vías farmacéuticas convencionales tal como se conoce en la técnica (por ejemplo, oral, intravenosa, intramuscular, intranasal, intratraqueal o subcutánea). Tal 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 son ventajosas para fines de producción de proteína en los que la FGE no humana producida es terapéuticamente útil.

Tal como se usa en el presente documento, se dice que una secuencia codificante y secuencias reguladoras están "operativamente" unidas cuando están unidas covalentemente 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' da como resultado la transcripción de la secuencia codificante y si la naturaleza del enlace entre las dos secuencias de ADN no (1) da como resultado la introducción de una mutación de cambio de marco, (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 estará operativamente unida a una secuencia codificante si la región promotora puede realizar a la transcripción de esa secuencia de ADN de tal manera que el transcrito resultante puede traducirse en la proteína o polipéptido deseado.

La naturaleza precisa de las secuencias reguladoras necesarias para la expresión génica puede variar entre tipos de especie o célula, pero en general incluirán, como 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 sentido 5' según se desee. Los vectores 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 técnica.

Vectores de expresión que contienen todos los elementos necesarios para la expresión están comercialmente disponibles y los conocen los expertos en la técnica. 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 para polipéptido de FGE o fragmento o variante 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.

5 Sistemas preferidos para la expresión de ARNm en células de mamífero son aquéllos tales como pRc/CMV (disponible de Invitrogen, Carlsbad, CA) que contiene un marcador seleccionable tal como un gen que confiere resistencia a G418 (lo que facilita la selección de líneas celulares transfectadas de manera estable) y las secuencias de potenciador-promotor del citomegalovirus (CMV) humano. Adicionalmente, es adecuado para la expresión en  
10 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 la transcripción eficaz *in vitro*. El plásmido se describe por Mishizuma y Nagata (Nuc. Acids Res. 18:5322, 1990), y su uso en experimentos de transfección se da a conocer, por ejemplo, por Demoulin (Mol. Cell. Biol. 16:4710-4716, 1996). Todavía otro vector de expresión preferido es un  
15 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.P1A se da a conocer por Warnier *et al.*, en la inyección intradérmica en ratones para su inmunización frente a P1A (Int. J. Cancer, 67:303-310, 1996).

20 Pueden proporcionarse kits de expresión, que permiten al experto en la técnica preparar un vector o vectores de expresión deseados. Tales kits de expresión incluyen al menos partes separadas de cada una de las secuencias codificantes anteriormente comentadas. Pueden añadirse otros componentes, según se desee, siempre que se incluyan las secuencias anteriormente mencionadas, que se requieren.

25 También se reconocerá que pueden usarse vectores de expresión que contienen la secuencia de ADNc de FGE descrita anteriormente para transfectar líneas celulares y células huésped, ya sean procariotas (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,  
30 e incluyen células primarias y líneas celulares inmortalizadas tal 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. La construcción de "deficientes" en gen de FGE en células y en animales es útil para proporcionar materiales para estudiar determinados aspectos de la actividad FGE.

35 Pueden proporcionarse polipéptidos aislados (incluyendo proteínas completas y proteínas parciales), codificados por los ácidos nucleicos de FGE anteriores, e incluyen el polipéptido de SEQ ID NO: 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  
40 incluyendo 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 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 únicos tal como se comentó anteriormente en relación con ácidos nucleicos. Tal como reconocerán los expertos en la técnica, el tamaño del fragmento único dependerá de factores tales como si el fragmento constituye una parte de  
50 un dominio de proteína conservado. Por tanto, algunas regiones de SEQ ID NO: 2 requerirán segmentos más largos para ser únicas mientras que otras sólo 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).

55 Fragmentos únicos de un polipéptido son preferiblemente aquellos fragmentos que conservan una capacidad funcional diferenciada del polipéptido. Las capacidades funcionales que pueden conservarse 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 identificar el polipéptido. Los expertos en la técnica conocerán bien métodos para seleccionar secuencias de  
60 aminoácidos únicas, normalmente basándose en la capacidad del fragmento único para distinguir selectivamente la secuencia de interés de otras que no son miembros de la familia. Una comparación de la secuencia del fragmento con las de bases de datos conocidas es normalmente todo lo que se necesita.

Las variantes de los polipéptidos de FGE descritos anteriormente son útiles. Tal como se usa en el presente documento, una "variante" de un polipéptido de FGE es un polipéptido que contiene una o más modificaciones de la  
65 secuencia de aminoácidos primaria de un polipéptido de FGE. Las modificaciones que crean una variante de



polipéptido de FGE se realizan normalmente en el ácido nucleico que codifica para 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) 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 propiedad o actividad 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 realizarse modificaciones directamente en el polipéptido, tal como mediante escisión, adición de una molécula ligadora, adición de un resto detectable, tal como biotina, adición de un ácido graso, y similares. Las modificaciones también abarcan proteínas de fusión que comprenden la totalidad o parte de la secuencia de aminoácidos de FGE. Un experto en la técnica estará familiarizado con métodos para predecir el 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 variante según métodos conocidos. Un ejemplo de un método de este tipo se describe por Dahiyat y Mayo en *Science* 278:82-87, 1997, mediante lo cual pueden diseñarse proteínas *de novo*. El método puede aplicarse a una proteína conocida para variar sólo una parte de la secuencia del polipéptido. Aplicando los métodos computacionales de Dahiyat y Mayo, pueden proponerse variantes específicas del polipéptido de FGE y someterse a prueba para determinar si la variante conserva una conformación deseada.

Las variantes 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 eliminarse residuos de cisteína para prevenir enlaces disulfuro no deseados. De manera similar, pueden cambiarse determinados aminoácidos para potenciar la expresión de un polipéptido de FGE eliminando la proteólisis mediante proteasas en un sistema de expresión (por ejemplo, residuos de aminoácido dibásico en sistemas de expresión de levadura en los que está presente actividad proteasa KEX2).

Las mutaciones de un ácido nucleico que codifica para 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 perjudiciales para la expresión del polipéptido variante.

Pueden realizarse mutaciones seleccionando una sustitución de aminoácidos, o mediante mutagénesis al azar de un sitio seleccionado en un ácido nucleico que codifica para el polipéptido. Entonces se expresan polipéptidos variantes y se someten a prueba para detectar una o más actividades para determinar qué mutación proporciona un polipéptido variante con las propiedades deseadas. Pueden realizarse mutaciones adicionales en variantes (o en polipéptidos de FGE no variantes) 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. los conocen bien los expertos en la técnica. Pueden realizarse todavía otras mutaciones en las secuencias no codificantes de un clon de ADNc o gen de FGE para potenciar la expresión del polipéptido.

El experto en la técnica constatará que pueden realizarse sustituciones de aminoácidos conservativas en polipéptidos de FGE para proporcionar variantes funcionalmente equivalentes de los polipéptidos anteriores, es decir, las variantes conservan las capacidades funcionales de los polipéptidos de FGE. Tal 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 variantes según métodos para alterar la secuencia del polipéptido conocidos para un experto habitual en la técnica, 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. Variantes funcionalmente equivalentes a modo de ejemplo de los polipéptidos de FGE incluyen sustituciones de aminoácidos conservativas de SEQ ID NO: 2. Las sustituciones conservativas de aminoácidos incluyen sustituciones realizadas 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 variantes funcionalmente equivalentes de polipéptidos de FGE, es decir, variantes de polipéptidos de FGE que conservan la función de los polipéptidos de FGE naturales. Se realizan normalmente sustituciones de aminoácidos conservativas en la secuencia de aminoácidos de polipéptidos de FGE para producir variantes funcionalmente equivalentes de polipéptidos de FGE mediante alteración de un ácido nucleico que codifica para polipéptidos de FGE (SEQ ID NO: 1, 3). Tales sustituciones pueden realizarse mediante una variedad de métodos conocidos por un experto habitual en la técnica. Por ejemplo, pueden realizarse sustituciones de aminoácidos mediante mutación dirigida por PCR, mutagénesis dirigida al sitio según el método de Kunkel (Kunkel, *Proc. Nat. Acad. Sci. U.S.A.* 82: 488-492, 1985), o mediante síntesis química de un gen que codifica para un polipéptido de FGE. La actividad de fragmentos funcionalmente equivalentes de polipéptidos de FGE puede someterse a prueba clonando el gen que codifica para 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 sometiendo a prueba para detectar una capacidad funcional de los polipéptidos de FGE tal como se da a conocer en el presente documento (por ejemplo, actividad de generación de C<sub>α</sub>-formilglicina, etc.).

La información proporcionada en el presente documento tiene varios usos, algunos de los cuales se describen en otra parte en el presente documento. En primer lugar, permite el aislamiento de polipéptidos de FGE. Puede utilizarse una variedad de metodologías bien conocidas por el experto en la técnica 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. Los expertos en la técnica también pueden seguir fácilmente métodos conocidos para aislar polipéptidos de FGE. Incluyen, pero no se limitan a, inmunocromatografía, HPLC, cromatografía de exclusión molecular, cromatografía de intercambio iónico y cromatografía de inmunoafinidad.

En determinadas realizaciones, pueden proporcionarse polipéptidos "negativos dominantes", derivados de polipéptidos de FGE. Un polipéptido negativo dominante es una variante inactiva de una proteína, que, interaccionando con la maquinaria celular, desplaza a una proteína activa de su interacción con la maquinaria celular o compite con la proteína activa, reduciendo así el efecto de la proteína activa. Por ejemplo, un receptor negativo dominante que se une a un ligando pero que no transmite una señal en respuesta a la unión del ligando puede reducir el efecto biológico de la expresión del ligando. Asimismo, una cinasa catalíticamente inactiva negativa dominante que interacciona normalmente con proteínas diana pero que no fosforila a las proteínas diana puede reducir la fosforilación de las proteínas diana en respuesta a una señal celular. De manera similar, un factor de transcripción negativo dominante que se une a un sitio de promotor en la región de control de un gen pero que no aumenta la transcripción del gen puede reducir el efecto de un factor de transcripción normal ocupando sitios de unión a promotor sin aumentar la transcripción.

El resultado final de la expresión de un polipéptido negativo dominante en una célula es una reducción de la función de proteínas activas. Un experto habitual en la técnica puede evaluar la posibilidad de una variante negativa dominante de una proteína, y usar técnicas de mutagénesis convencionales para crear uno o más polipéptidos variantes negativos dominantes. Véase, por ejemplo, la patente estadounidense n.º 5.580.723 y Sambrook *et al.*, Molecular Cloning: A Laboratory Manual, segunda edición, Cold Spring Harbor Laboratory Press, 1989. El experto en la técnica puede someter entonces a prueba la población de polipéptidos sometidos a mutagénesis para detectar una disminución de una actividad seleccionada y/o para detectar la conservación de una actividad de este tipo. Otros métodos similares para crear y someter a prueba variantes negativas dominantes de una proteína resultarán evidentes para un experto en la técnica.

El aislamiento del ADNc de FGE también hace posible que el experto en la técnica diagnostique un trastorno caracterizado por una expresión aberrante de FGE. Estos métodos suponen determinar la expresión del gen de FGE, y/o polipéptidos de FGE derivados del mismo. En la primera 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 sometiendo a ensayo con sondas de hibridación marcadas tal como se muestra a modo de ejemplo 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 es deficiencia múltiple de sulfatasas.

Son útiles 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. Anticuerpos incluyen anticuerpos policlonales y monoclonales, preparados según metodología convencional. En determinadas realizaciones, se excluyen agentes de unión (por ejemplo, anticuerpos) que se unen a los polipéptidos codificados por los ácidos nucleicos de SEQ ID NO: 4.

De manera significativa, tal como se conoce en la técnica, sólo una pequeña parte de una molécula de anticuerpo, el paratopo, participa en la unión del anticuerpo a su epítipo (véase, en general, Clark, W.R. (1986) The Experimental Foundations of Modern Immunology, Wiley & Sons, Inc., Nueva York; Roitt, L (1991) Essential Immunology, 7ª 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 fragmento F(ab')<sub>2</sub>, conserva ambos sitios de unión a antígeno de un anticuerpo intacto. De manera similar, un anticuerpo del que se ha escindido enzimáticamente la región Fc, o que se ha producido sin la región Fc, denominado fragmento Fab, conserva 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 parte de la cadena pesada de anticuerpo denominada Fd. Los fragmentos Fd son los determinantes principales de la especificidad de anticuerpo (un fragmento Fd individual puede estar asociado con hasta diez cadenas ligeras diferentes si alterar la especificidad de anticuerpo) y los fragmentos Fd conservan capacidad de unión a epítipo en aislamiento.

Dentro de la parte de unión a antígeno de un anticuerpo, tal como se conoce bien en la técnica, hay regiones determinantes de complementariedad (CDR), que interaccionan directamente con el epítipo del antígeno, y regiones de entramado (FR), que mantienen la estructura terciaria del paratopo (véase, en general, Clark, 1986; Roitt, 1991). Tanto en el fragmento Fd de cadena pesada como en la cadena ligera de inmunoglobulinas IgG, hay cuatro regiones de entramado (de FR1 a FR4) separadas respectivamente por tres regiones determinantes de 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 conservan la especificidad epítipo 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. Véase, por ejemplo, las patentes estadounidenses 4.816.567, 5.225.539, 5.585.089, 5.693.762 y 5.859.205. Por tanto, por ejemplo, la publicación internacional PCT n.º WO 92/04381 enseña la producción y el uso de anticuerpos frente a VSR murinos humanizados en los que al menos una parte 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 con frecuencia anticuerpos "quiméricos".

Por tanto, tal como resultará evidente a un experto habitual en la técnica, 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 parejas 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 resumen, se prepara una biblioteca de fagos (usando por ejemplo fago m13, fd, o lambda), que presentan insertos de desde 4 hasta 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 una pareja de unión. Este proceso puede repetirse a lo largo de varios ciclos de reselección de fago que se une 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 parte 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 parte lineal mínima más uno o más residuos degenerados adicionales en sentido 5' o en sentido 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, o un fragmento de los mismos, o complejos de FGE y una pareja de unión, pueden usarse para examinar bibliotecas de péptidos, incluyendo bibliotecas de presentación en fagos, para identificar y seleccionar parejas de unión a péptidos de los polipéptidos de FGE. Tales moléculas pueden usarse, tal como se describe, para ensayos de selección, para protocolos de purificación, para interferir directamente con el funcionamiento de FGE y para otros fines que resultarán evidentes para los expertos habituales en la técnica.

También puede usarse un polipéptido de FGE, o un fragmento del mismo, para aislar sus parejas de unión nativas. El aislamiento de parejas 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 una pareja 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 la pareja de unión. Otras proteínas que son parejas de unión para FGE pueden aislarse mediante métodos similares sin experimentación excesiva. Una pareja de unión preferida es una sulfatasa.

Pueden proporcionarse métodos para medir el nivel de expresión de FGE en un sujeto. Esto puede realizarse

obteniendo en primer lugar una muestra de prueba del sujeto. La muestra de prueba puede ser tejido o líquido biológico. Los tejidos incluyen cerebro, corazón, suero, mama, colon, vejiga, útero, próstata, estómago, testículos, ovario, páncreas, glándula pituitaria, glándula suprarrenal, glándula tiroidea, glándula salivar, glándula mamaria, riñón, hígado, intestino, bazo, timo, vasos sanguíneos, médula ósea, tráquea y pulmón. En determinadas realizaciones, las muestras de prueba se originan de tejidos del corazón y vasos sanguíneos, y los líquidos biológicos incluyen sangre, saliva y orina. Pueden usarse técnica tanto invasivas como no invasivas para obtener tales muestras y están bien documentadas en la técnica. A nivel molecular pueden usarse tanto PCR como transferencia de tipo Northern para determinar el nivel de ARNm de FGE usando productos descritos en el presente documento, y protocolos bien conocidos en la técnica que se encuentran en referencias que compilan tales métodos.

A nivel de proteína, puede determinarse la expresión de FGE usando sueros anti-FGE o bien policlonales o bien monoclonales en combinación con ensayos inmunológicos convencionales. Los métodos preferidos compararán el nivel medido de expresión de FGE de la muestra de prueba con un control. Un control puede incluir una cantidad conocida de una sonda de ácido nucleico, un epítipo de FGE (tal como un producto de expresión de FGE), o una muestra de prueba similar de un sujeto con un nivel de control o "normal" de expresión de FGE.

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 químicas que comprenden una fusión de una proteína FGE con otro polipéptido, por ejemplo, un polipéptido que puede 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 fragmento o polipéptido de FGE también puede proporcionar medios para detectar fácilmente la proteína de fusión, por ejemplo, mediante reconocimiento inmunológico o mediante marcado fluorescente.

La información proporcionada en el presente documento también es útil en la generación de animales transgénicos no humanos. Tal como se usa en el presente documento, "animales transgénicos no humanos" incluyen animales no humanos que tienen una o más moléculas de ácido nucleico exógenas incorporadas en células de línea germinal y/o células somáticas. Por tanto, los animales transgénicos incluyen animales "deficientes" que tienen una alteración de genes homocigóticos o heterocigóticos mediante recombinación homóloga, animales que tienen vectores de expresión incorporados de manera episomal o cromosómica, etc. Pueden prepararse animales deficientes mediante recombinación homóloga usando células madre embrionarias tal como se conoce bien en la técnica. La recombinación puede facilitarse usando, por ejemplo, el sistema cre/lox u otros sistemas de recombinasa conocidos por un experto habitual en la técnica. En determinadas realizaciones, el propio sistema de recombinasa se expresa de manera condicionada, por ejemplo, en determinados tejidos o tipos de célula, en determinados estadios de desarrollo embrionario o posembionario, se induce mediante la adición de un compuesto que aumenta o disminuye la expresión, y similares. En general, los vectores de expresión condicionados usados en tales sistemas usan una variedad de promotores que confieren el patrón de expresión génica deseado (por ejemplo, temporal o espacial). También pueden unirse operativamente promotores condicionados a moléculas de ácido nucleico de FGE para aumentar la expresión de FGE de una manera regulada o condicionada. También pueden unirse operativamente reguladores negativos que actúan en trans de la actividad o expresión de FGE a un promotor condicionado tal como se describió anteriormente. Tales reguladores que actúan en trans incluyen moléculas de ácidos nucleicos de FGE antisentido, moléculas de ácido nucleico que codifican para moléculas de FGE negativas dominantes, moléculas de ribozima específicas para ácidos nucleicos de FGE, y similares. Los animales transgénicos no humanos son útiles en experimentos dirigidos a someter a prueba los efectos bioquímicos o fisiológicos de productos de diagnóstico o productos terapéuticos para detectar estados caracterizados por una expresión de FGE aumentada o disminuida. Otros usos resultarán evidentes para un experto habitual en la técnica.

También se contempla la terapia génica. El procedimiento para realizar terapia génica *ex vivo* se explica en la patente estadounidense 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(s) célula(s) de un sujeto que contiene(n) 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. Los 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 WO95/00654. También se contempla la terapia génica *in vivo* usando vectores tales como adenovirus, retrovirus, herpesvirus, y liposomas seleccionados como diana.

Pueden proporcionarse métodos eficaces de identificación de agentes o compuestos de partida para agentes activos a nivel de una función celular dependiente de FGE o fragmento de FGE. En particular, tales funciones incluyen interacción con otros polipéptidos o fragmentos. Generalmente, los métodos de selección suponen someter a ensayo para detectar compuestos que interfieren con la actividad FGE (tal como actividad de generación de  $C_{\alpha}$ -formilglicina), aunque también pueden someterse a ensayo compuestos que potencian la actividad FGE de generación de  $C_{\alpha}$ -formilglicina usando los métodos de selección. Tales métodos pueden adaptarse para la selección automatizada de alto rendimiento de compuestos. Indicaciones diana incluyen procesos celulares modulados por FGE tales como actividad de generación de  $C_{\alpha}$ -formilglicina.

Se proporciona una amplia variedad de ensayos para agentes candidatos (farmacológicos), incluyendo, ensayos de unión proteína-ácido *in vitro* marcados, ensayos de desplazamiento de movilidad electroforética, inmunoensayos, ensayos basados en células tales como selecciones de dos o tres híbridos, ensayos de expresión, etc. Los ácidos nucleicos transfectados pueden codificar, por ejemplo, para bibliotecas de péptidos combinatorias o bibliotecas de ADNc. En la técnica se conocen reactivos convenientes para tales ensayos, por ejemplo, proteínas de fusión de GAL4. Un ensayo basado en células a modo de ejemplo implica transferir a una célula un ácido nucleico que codifica para un polipéptido de FGE fusionado a un dominio de unión a ADN de GAL4 y un ácido nucleico que codifica para un gen indicador operativamente unido a una región reguladora de expresión génica, tal como uno o más sitios de unión a GAL4. La activación de transcripción de gen indicador se produce cuando el polipéptido de fusión de FGE e indicador se une de tal manera que permite la transcripción del gen indicador. Entonces se detectan agentes que modulan una función celular mediada por polipéptido de FGE mediante un cambio en la expresión del gen indicador. En la técnica se conocen métodos para determinar cambios en la expresión de un gen indicador.

Los fragmentos de FGE usados en los métodos, cuando no se producen mediante un ácido nucleico transfectado, se añaden a una mezcla de ensayo como un polipéptido aislado. Preferiblemente se producen polipéptidos de FGE de manera recombinante, aunque tales polipéptidos pueden aislarse de extractos biológicos. Los polipéptidos de FGE producidos de manera recombinante incluyen proteínas quiméricas que comprende una fusión de una proteína FGE con otro polipéptido, por ejemplo, un polipéptido que puede proporcionar o potenciar la unión proteína-proteína, unión a 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 o epítipo Flag.

La mezcla de ensayo comprende una diana de unión a FGE intracelular natural que puede interactuar con FGE. Aunque pueden usarse dianas de unión a FGE naturales, con frecuencia se prefiere usar partes (por ejemplo, péptidos (véase por ejemplo, el péptido de SEQ ID NO: 33) o fragmentos de ácido nucleico) o análogos (es decir, agentes que imitan las propiedades de unión a FGE de la diana de unión natural para fines del ensayo) de la diana de unión a FGE siempre que la parte o el análogo proporcione avidez y afinidad de unión al fragmento de FGE medible en el ensayo.

La mezcla de ensayo también comprende un agente candidato. Normalmente, se aplica una pluralidad de mezclas de ensayo en paralelo con diferentes concentraciones de agente para obtener una respuesta diferente a las diversas concentraciones. Normalmente, una de esas concentraciones sirve como control negativo, es decir, a una concentración nula de agente o a una concentración de agente por debajo de los límites de detección del ensayo. Agentes candidatos abarcan numerosas clases químicas, aunque normalmente son compuestos orgánicos. Preferiblemente, los agentes candidatos son compuestos orgánicos pequeños, es decir, aquellos que tienen un peso molecular de más de 50 pero menos de aproximadamente 2500, preferiblemente menos de aproximadamente 1000 y, más preferiblemente, menos de aproximadamente 500. Agentes candidatos comprenden grupos químicos funcionales necesarios para interacciones estructurales con polipéptidos y/o ácidos nucleicos, y normalmente incluyen al menos un grupo amina, carbonilo, hidroxilo o carboxilo, preferiblemente al menos dos de los grupos químicos funcionales y más preferiblemente al menos tres de los grupos químicos funcionales. Los agentes candidatos pueden comprender una estructura heterocíclica o de carbono cíclica y/o estructuras aromáticas o poliaromáticas sustituidas con uno o más de los grupos funcionales identificados anteriormente. Los agentes candidatos también pueden ser biomoléculas tales como péptidos, sacáridos, ácidos grasos, esteroides, isoprenoides, purinas, pirimidinas, derivados o análogos estructurales de los anteriores, o combinaciones de los mismos y similares. Cuando el agente es un ácido nucleico, el agente es normalmente una molécula de ADN o ARN, aunque también se contemplan ácidos nucleicos modificados tal como se define en el presente documento.

Los agentes candidatos se obtienen de una amplia variedad de fuentes incluyendo bibliotecas de compuestos sintéticos o naturales. Por ejemplo, están disponibles numerosos medios para la síntesis al azar y dirigida de una amplia variedad de compuestos orgánicos y biomoléculas, incluyendo la expresión de oligonucleótidos aleatorizados, bibliotecas combinatorias orgánicas sintéticas, bibliotecas de presentación en fagos de péptidos al azar y similares. Alternativamente, están disponibles o pueden producirse fácilmente bibliotecas de compuestos naturales en forma de extractos bacterianos, fúngicos, vegetales y animales. Adicionalmente, pueden modificarse compuestos y bibliotecas naturales y producidas sintéticamente mediante medios químicos, físicos y bioquímicos convencionales. Además, pueden someterse agentes conocidos (farmacológicos) a modificaciones químicas dirigidas o al azar tales como acilación, alquilación, esterificación, amidificación, etc. para producir análogos estructurales de los agentes.

También puede incluirse una variedad de otros reactivos en la mezcla. Incluyen reactivos tales como sales, tampones, proteínas neutras (por ejemplo, albúmina), detergentes, etc. que pueden usarse para facilitar la unión óptima proteína-proteína y/o proteína-ácido nucleico. Un reactivo de este tipo también puede reducir las interacciones no específicas o de fondo de los componentes de reacción. También pueden usarse otros reactivos que mejoran la eficacia del ensayo tales como proteasa, inhibidores, inhibidores de nucleasa, agentes antimicrobianos y similares.

La mezcla de los materiales de ensayo anteriores se incubaba en condiciones mediante las cuales, salvo por la

presencia del agente candidato, el polipéptido de FGE se une específicamente a una diana de unión celular, una parte de la misma o análogo de la misma. El orden de adición de componentes, la temperatura de incubación, tiempo de incubación, y otros parámetros del ensayo pueden determinarse fácilmente. Tal experimentación implica simplemente la optimización de los parámetros del ensayo, no la composición fundamental del ensayo. Las temperaturas de incubación son normalmente de entre 4°C y 40°C. Los tiempos de incubación se minimizan preferiblemente para facilitar una selección rápida de alto rendimiento, y son normalmente de entre 0,1 y 10 horas.

Tras la incubación, se detecta la presencia o ausencia de unión específica entre el polipéptido de FGE y una o más dianas de unión mediante cualquier método conveniente disponible para el usuario. Para ensayos de tipo de unión libre de células, con frecuencia se usa una etapa de separación para separar componentes unidos de no unidos. La etapa de separación puede lograrse de una variedad de maneras. Convenientemente, al menos uno de los componentes se inmoviliza sobre un sustrato sólido, del que pueden separarse fácilmente los componentes no unidos. El sustrato sólido puede prepararse de una amplia variedad de materiales y en una amplia variedad de formas, por ejemplo, placa de microtitulación, microperla, tira reactiva, partícula de resina, etc. El sustrato se elige preferiblemente para relaciones señal-ruido máximas, principalmente para minimizar la unión de fondo, así como para la facilidad de separación y coste.

La separación puede realizarse, por ejemplo, retirando una perla o tira reactiva de un depósito, vaciando o diluyendo un depósito tal como un pocillo de placa de microtitulación aclarando una perla, partícula, columna cromatográfica o filtro con una disolución de lavado o disolvente. La etapa de separación incluye preferiblemente múltiples aclarados o lavados. Por ejemplo, cuando el sustrato sólido es una placa de microtitulación, los pocillos pueden lavarse varias veces con una disolución de lavado, que incluye normalmente los componentes de la mezcla de incubación que no participan en uniones específicas tales como sales, tampón, detergente, proteína no específica, etc. Cuando el sustrato sólido es una perla magnética, las perlas pueden lavarse una o más veces con una disolución de lavado y aislarse usando un imán.

Puede realizarse la detección de cualquier manera conveniente para ensayos basados en células tales como selecciones de dos o tres híbridos. El transcrito resultante de un ensayo de transcripción de gen indicador de polipéptido de FGE que interacciona con una molécula diana codifica normalmente para un producto directa o indirectamente detectable, por ejemplo, actividad  $\beta$ -galactosidasa, actividad luciferasa y similares. Para ensayos de unión libres de células, uno de los componentes comprende habitualmente, o está acoplado a, un marcador detectable. Puede usarse una amplia variedad de marcadores, tales como los que proporcionan detección directa (por ejemplo, radioactividad, luminiscencia, densidad óptica o electrónica, etc.), o detección indirecta (por ejemplo, etiqueta de epítipo tal como el epítipo FLAG, etiqueta enzimática tal como peroxidasa del rábano, etc.). El marcador puede unirse a una pareja de unión a FGE, o incorporarse en la estructura de la pareja de unión.

Puede usarse una variedad de métodos para detectar el marcador, dependiendo de la naturaleza del marcador y otros componentes del ensayo. Por ejemplo, el marcador puede detectarse mientras está unido al sustrato sólido o posteriormente a la separación del sustrato sólido. Pueden detectarse directamente marcadores mediante densidad óptica o electrónica, emisiones radioactivas, transferencias de energía no radiada, etc. o detectarse indirectamente con conjugados de anticuerpo, conjugados de estreptavidina-biotina, etc. En la técnica se conocen bien métodos para detectar los marcadores.

Pueden proporcionarse 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 selecciones de dos híbridos, y agentes de unión no naturales intracelulares y extracelulares identificados en exploraciones 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 pueden unirse selectivamente a un polipéptido de FGE tienen preferiblemente constantes de equilibrio de unión de al menos aproximadamente  $10^7 \text{ M}^{-1}$ , más preferiblemente al menos aproximadamente  $10^8 \text{ M}^{-1}$ , y lo más preferiblemente al menos aproximadamente  $10^9 \text{ M}^{-1}$ . Puede usarse una amplia variedad de ensayos basados en células y libres de células para demostrar la unión específica para FGE. Ensayos basados en células incluyen selecciones de uno, dos y tres híbridos, ensayos en los que se inhibe o se incrementa la transcripción mediada por FGE, etc. Ensayos libres de células incluyen ensayos de unión FGE-proteína, inmunoensayos, etc. Otros ensayos útiles para seleccionar 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 para identificar un agente útil en la modulación de la actividad de generación de  $\text{C}_\alpha$ -formilglutamina. El método implica (a) poner en contacto una molécula que tiene actividad de generación de  $\text{C}_\alpha$ -formilglutamina con un agente candidato, (b) medir la actividad de generación de  $\text{C}_\alpha$ -formilglutamina de la molécula, y (c)

comparar la actividad de generación de  $C_{\alpha}$ -formilglicina medida de la molécula con un control para determinar si el agente candidato modula la actividad de generación de  $C_{\alpha}$ -formilglicina de la molécula, en el que la molécula es una molécula de ácido nucleico de FGE tal como se describe en el presente documento, o un producto de expresión de la misma. "Poner en contacto" se refiere a poner en contacto tanto directo como indirecto una molécula que tiene actividad de generación de  $C_{\alpha}$ -formilglicina con el agente candidato. Poner en contacto "indirecto" significa que el agente candidato ejerce su efecto sobre la actividad de generación de  $C_{\alpha}$ -formilglicina de la molécula mediante un tercer agente (por ejemplo, una molécula mensajera, un receptor, etc.). En determinadas realizaciones, el control es la actividad de generación de  $C_{\alpha}$ -formilglicina de la molécula medida en ausencia del agente candidato. Los métodos de ensayo y agentes candidatos son tal como se describieron anteriormente en las realizaciones anteriores con respecto a FGE.

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 tal como se describe en el presente documento. El trastorno es deficiencia múltiple de sulfatasas. Las mutaciones en el gen de FGE que provocan la expresión aberrante de moléculas de FGE da como resultado los siguientes cambios de aminoácido en SEQ ID NO: 2: Met1Arg; Met1Val; Leu20Phe; Ser155Pro; Ala177Pro; Cys218Tyr; Arg224Trp; Asn259De; Pro266Leu; Ala279Val; Arg327Stop; Cys336Arg; Arg345Cys; Ala348Pro; Arg349Gln; Arg349Trp; Arg349Trp; Ser359Stop; o una combinación de los mismos.

En el caso en el que la molécula es una molécula de ácido nucleico, tales determinaciones pueden llevarse a cabo mediante cualquier ensayo de determinación de ácido nucleico convencional, incluyendo la reacción en cadena de la polimerasa, o sometiendo a ensayo con sondas de hibridación marcadas tal como se muestra a modo de ejemplo en el presente documento. En el caso en el que la molécula es 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 convencional usando, por ejemplo, anticuerpos que se unen a cualquiera de los productos de expresión de polipéptido.

"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", tal como se usa en el presente documento, se refiere a un sujeto que, según patrones médicos convencionales, 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 sujetos, si se examinan por un profesional médico, se caracterizarán como sanos y libres de síntomas de una deficiencia múltiple de sulfatasas. 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.

Pueden usarse kits novedosos para medir los niveles de los ácidos nucleicos descritos en el presente documento, o productos de expresión de los mismos.

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 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.

En el caso de detección de ácido nucleico, pueden incluirse pares de cebadores para amplificar una molécula de ácido nucleico. Los kits preferidos incluirán controles tales como cantidades conocidas de sondas de ácido nucleico, epítipos (tales como 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, 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 un estado de deficiencia de sulfatasa basándose en el desenlace del ensayo. Los reactivos pueden envasarse en recipientes y/o recubrirse sobre pocillos en cantidades predeterminadas, y los kits pueden

incluir materiales convencionales 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 generalmente designado con el número 11. El kit 11 comprende los siguientes elementos principales: un envase 15, un agente 17 tal 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 el agente 17, un vial (o varios viales) que contiene un agente 19 de control, e instrucciones 21. Los expertos en la técnica pueden modificar fácilmente el envase 15 para adaptarse a necesidades individuales.

Pueden proporcionarse métodos para tratar deficiencia múltiple de sulfatasas en un sujeto. Un método implica administrar a un sujeto que necesita 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 coadministrar un agente seleccionado del grupo que consiste en una molécula de ácido nucleico que codifica para 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, un producto de expresión de la molécula de ácido nucleico, y/o un fragmento del producto de expresión de la molécula de ácido nucleico.

“Agentes que modulan la expresión” de un ácido nucleico o un polipéptido, tal 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 frente a los polipéptidos, y similares. Cualquier agente que modula la expresión de una molécula (y tal como se describe en el presente documento, modula su actividad) es útil. 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 SEQ ID NO: 3). En realizaciones importantes, el agente que modula la actividad de generación de  $C_{\alpha}$ -formilglicina es un péptido (por ejemplo, un péptido de SEQ ID NO: 2). En algunas realizaciones, el agente que modula la actividad de generación de  $C_{\alpha}$ -formilglicina es un ácido nucleico sentido.

Puede proporcionarse un método para aumentar la actividad de generación de  $C_{\alpha}$ -formilglicina en un sujeto. El método implica administrar una molécula de ácido nucleico de FGE aislada descrita en el presente documento, y/o un producto de expresión de la misma, a un sujeto, en una cantidad eficaz para aumentar la actividad de generación de  $C_{\alpha}$ -formilglicina en el sujeto.

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 SEQ ID NO: 1), o un producto de expresión de la misma (por ejemplo, un péptido de SEQ ID NO: 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 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 tal como una célula HT-1080. La “secuencia de expresión génica” es cualquier secuencia de nucleótidos reguladora, tal como una secuencia promotora o combinación promotor-potenciador, lo 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 de virus, 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 (HFT), adenosina desaminasa, piruvato cinasa, promotor de  $\alpha$ -actina y otros promotores constitutivos. Promotores de virus a modo de ejemplo que funcionan constitutivamente en células eucariotas incluyen, por ejemplo, promotores del virus del simio, virus del papiloma, adenovirus, virus de inmunodeficiencia humana (VIH), virus del sarcoma de Rous, citomegalovirus, las repeticiones terminales largas (LTR) de virus de leucemia de Moloney y otros retrovirus, y el promotor de timidina cinasa del virus del herpes simple. Otros promotores constitutivos los conocen los expertos habituales en la técnica. Los promotores útiles como secuencias de expresión génica 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 los conocen los expertos habituales en la técnica.

En general, la secuencia de expresión génica incluirá, 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 de ácido nucleico operativamente unido. Las secuencias de expresión génica incluyen opcionalmente secuencias potenciadoras o secuencias activadoras en sentido 5' según se desee.



Preferiblemente, cualquiera de las moléculas de ácido nucleico de FGE descritas en el presente documento está unida 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 celular de este tipo, 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. Los expertos habituales en la técnica podrán identificar fácilmente promotores alternativos que pueden 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 unidas” 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, SEQ 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 unidas si la inducción de un promotor en la secuencia de expresión génica en 5' da como resultado la transcripción de la secuencia de ácido nucleico y si la naturaleza del enlace entre las dos secuencias de ADN no (1) da como resultado la introducción de una mutación de cambio de marco, (2) interfiere con la capacidad de la región promotora para dirigir la transcripción de la secuencia de ácido nucleico, y/o (3) interfiere con la capacidad del transcrito de ARN correspondiente para traducirse en una proteína. Por tanto, una secuencia de expresión génica estará operativamente unida a una secuencia de ácido nucleico si la secuencia de expresión génica puede realizar la transcripción de esa secuencia de ácido nucleico de tal manera que el transcrito resultante pueda traducirse en la proteína o polipéptido deseado.

Las moléculas descritas en el presente documento pueden suministrarse a los tipos de célula preferidos solas o en asociación con un vector (véase también la explicación 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 que puede facilitar: (1) el suministro de una molécula a una célula diana y/o (2) la captación de la molécula por una célula diana. Preferiblemente, los vectores de suministro transportan la molécula al interior de la célula diana con degradación reducida con respecto al grado de degradación que resultaría en ausencia del vector. Opcionalmente, puede fijarse un “ligando de direccionamiento” al vector para suministrar 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 suministrarse selectivamente a una neurona. Metodologías para el direccionamiento incluyen conjugados, tales como los descritos en la patente estadounidense 5.391.723 concedida a Priest. Otro ejemplo de un vehículo de direccionamiento bien conocido es un liposoma. Hay liposomas comercialmente disponibles de Gibco BRL. Están publicados numerosos métodos para preparar liposomas de direccionamiento.

En general, vectores útiles 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 fijarse a las secuencias de ácido nucleico. 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 adenoasociado; retrovirus, tales como virus de leucemia murina de Moloney; virus del sarcoma murino de Harvey; virus de tumor de mama murino; virus del sarcoma de Rouse; virus de tipo SV40; virus del poliovirus; virus de Epstein-Barr; virus del papiloma; virus del herpes; virus vaccinia; 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 adenoasociado, un virus de ADN bicatenario. El virus adenoasociado puede 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 disolventes lipídicos, altas frecuencias de transducción en células de diversos linajes, incluyendo células hematopoyéticas, y falta de inhibición de superinfección permitiendo por tanto múltiples series de transducciones. Se notifica que el virus adenoasociado puede integrarse en el ADN celular humano de una manera específica del sitio, minimizando así la posibilidad de mutagénesis por inserción y variabilidad de la expresión del gen insertado. Además, se ha realizado un seguimiento de infecciones por virus adenoasociado de tipo natural en cultivo tisular durante más de 100 pasos en ausencia de presión selectiva, lo que supone que la integración genómica del virus adenoasociado es un acontecimiento relativamente estable. El virus adenoasociado 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 para dar ADN con la posterior integración proviral en el ADN celular huésped. Se han aprobado adenovirus y retrovirus para ensayos de terapia génica en seres humanos. En general, los retrovirus son deficientes para la replicación (es decir, pueden dirigir la síntesis de las proteínas deseadas, pero

son 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 convencionales 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 a partir 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 de virus de leucemia murina de Moloney, tal como se describe Nabel, E.G., *et al.*, Science, 1990, 249:1285-1288. Se notifica que estos vectores fueron eficaces para el suministro de genes a las tres capas de la pared arterial, incluyendo la media. Otros vectores preferidos se dan a conocer en Flugelman, *et al.*, Circulation, 1992, 85:1110-1117. Vectores adicionales que son útiles para suministrar moléculas se describen en la patente estadounidense n.º 5.674.722 de Mulligan, *et al.*

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

Un método de suministro de este tipo preferido 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 suministro *in vivo* o *in vitro*. Se ha mostrado que vasos unilamelares grandes (LUV), cuyo tamaño oscila entre 0,2 y 4,0  $\mu\text{m}$  pueden encapsular macromoléculas grandes. Pueden encapsularse ARN, ADN y viriones intactos dentro del interior acuoso y suministrarse a células en una forma biológicamente activa (Fraley, *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) suministro 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 vírica 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.

Hay liposomas 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-dioleiloxy)-propil]-N,N,N-trimetilamonio (DOTMA) y bromuro de dimetildioctadecilamonio (DDAB). Métodos para preparar 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 el suministro intracelular de macromoléculas, incluyendo ácidos nucleicos, en la solicitud internacional PCT n.º PCT/US96/07572 (publicación n.º WO 96/40060, titulada "Intracellular Delivery of Macromolecules").

En una realización particular, el vehículo preferido es un implante o micropartícula biocompatible 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 n.º PCT/LUS/03307 (publicación n.º WO 95/24929, titulada "Polymeric Gene Delivery System", que reivindica prioridad de la solicitud de patente estadounidense con n.º 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 un pliegue sostenido 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 dada a conocer 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 matriz polimérica para contener los ácidos nucleicos incluyen películas, recubrimientos, geles, implantes y endoprótesis. El tamaño y la composición del dispositivo de matriz polimérica se seleccionan para dar como resultado 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 suministro que va a usarse, normalmente inyección en un tejido o administración de una suspensión mediante aerosol en las zonas nasal y/o pulmonar. La composición de matriz polimérica puede seleccionarse para que tenga tanto tasas de degradación favorables como también para estar formada por un material que es bioadhesivo, para aumentar adicionalmente la eficacia de transferencia cuando se administra el dispositivo a una superficie vascular. La composición de matriz también puede seleccionarse para no degradarse, sino más bien liberar mediante difusión a lo largo de un periodo de tiempo extendido.

Pueden usarse matrices poliméricas tanto biodegradables como no biodegradables para suministrar los ácidos nucleicos 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 de múltiples valencias u otros polímeros.

En general, los ácidos nucleicos descritos en el presente documento pueden suministrarse 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 suministro biodegradable incluyen: poliamidas, policarbonatos, polialquilenos, polialquilenglicoles, poli(óxidos de alquileo), poli(tereftalatos de alquileo), poli(alcoholes vinílicos), poliviniléteres, poli(ésteres vinílicos), poli(haluros de vinilo), polivinilpirrolidona, poliglicolidos, polisiloxanos, poliuretanos y copolímeros de los mismos, alquilcelulosa, hidroxialquilcelulosas, éteres de celulosa, ésteres de celulosa, nitrocelulosas, polímeros de ésteres acrílicos y metacrílicos, metilcelulosa, etilcelulosa, hidroxipropilcelulosa, hipromelosa, hidroxibutylmetilcelulosa, acetato de celulosa, propionato de celulosa, acetato-butirato 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 de manera rutinaria por los expertos en la técnica), 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, cuyas enseñanzas se incorporan en el presente documento, 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 lo siguiente: una composición de las moléculas descritas anteriormente para su uso como medicamento, métodos para preparar el medicamento y métodos para la liberación sostenida del medicamento *in vivo*.

También pueden usarse agentes de compactación en combinación con un vector. Un "agente de compactación", tal 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 por tanto 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 suministrar un ácido nucleico aislado tal como se describe 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 incluyen fosfato de calcio y otros mediadores químicos de transporte intracelular, composiciones de microinyección y electroporación.

Pueden proporcionarse métodos para aumentar la actividad sulfatasa en una célula. Tales métodos implican poner en contacto una célula que expresa una sulfatasa con una molécula de ácido nucleico aislada tal como se describe en el presente documento (por ejemplo, una molécula de ácido nucleico aislada tal como se reivindica en una cualquiera de las reivindicaciones 1-8, una molécula de ácido nucleico de FGE que tiene una secuencia seleccionada del grupo que consiste en la SEQ ID NO: 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 tal como se reivindica en las reivindicaciones 11-15, 19, 20, o un péptido que tiene una secuencia seleccionada del grupo que consiste en la SEQ ID NO: 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 sulfatasa en la célula. “Aumentar” la actividad sulfatasa, tal 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 de tipo natural. Por “aumentar la actividad 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 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 portador farmacéuticamente aceptable. En el presente documento dicha célula se ha puesto en contacto con un agente que comprende una molécula de ácido nucleico aislada tal como se describe en el presente documento (por ejemplo, tal como se reivindica en las reivindicaciones 1-8, o un ácido nucleico tal como se reivindica en las reivindicaciones 1-8, o una molécula de ácido nucleico que tiene una secuencia seleccionada del grupo que consiste en la SEQ ID NO: 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 SEQ ID NO: 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, sólo una fracción de las sulfatasas recombinantes es activa. Se ha descubierto, inesperadamente, que la expresión/actividad aumentada de FGE en una célula con expresión/actividad aumentada de una sulfatasa da como resultado la producción de una sulfatasa que es más activa. Dado que la presencia de *FGly* en una molécula de sulfatasa está asociada con actividad sulfatasa, puede cuantificarse la “sulfatasa activa” determinando la presencia de *FGly* en el producto celular de sulfatasa usando espectrometría de masas MALDI-TOF, tal como se describe en otra parte en el presente documento. Entonces puede determinarse fácilmente la razón con respecto a 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), condrodistrofia punctata I recesiva ligada al cromosoma X e ictiosis ligada al cromosoma X (deficiencia de sulfatasa esteroidea).

Los métodos son útiles en el tratamiento tanto agudo como profiláctico de cualquiera de los estados anteriores. Tal como se usa en el presente documento, un tratamiento agudo se refiere al tratamiento de sujetos que tienen un estado particular. El tratamiento profiláctico se refiere al tratamiento de sujetos con riesgo de tener el estado, pero que actualmente no tienen o experimentan los síntomas del estado.

En su sentido más amplio, los términos “tratamiento” o “tratar” se refieren a tratamientos tanto agudos como profilácticos. Si el sujeto que necesita tratamiento está experimentando un estado (o tiene o está teniendo un estado particular), entonces tratar el estado se refiere a mejorar, reducir o eliminar el estado o uno o más síntomas que surgen del estado. En algunas realizaciones preferidas, tratar el estado se refiere a mejorar, reducir o eliminar un síntoma específico o un subconjunto específico de síntomas asociados con el estado. Si el sujeto que necesita tratamiento es uno que está en riesgo de desarrollar un estado, entonces tratar al sujeto se refiere a reducir el riesgo del sujeto de desarrollar el estado.

El modo de administración y la dosificación de un agente terapéutico variarán con el estadio particular del estado 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.

Tal como se describe en el presente documento, pueden administrarse agentes terapéuticos 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 un estado particular tal como una reducción, alivio o eliminación de un síntoma o de un estado 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 suponer sólo ralentizar la progresión del estado temporalmente, aunque más preferiblemente, implica detener la progresión del estado permanentemente o retrasar la aparición de, o prevenir que se produzca, el estado. Esto puede monitorizarse mediante métodos rutinarios. Generalmente, las dosis de compuestos activos serán de desde aproximadamente 0,01 mg/kg al día hasta 1000 mg/kg al día. Se espera que dosis que oscilan entre 50 □g y 500 mg/kg serán adecuadas, preferiblemente por

Tales cantidades dependerán, por supuesto, del estado particular que está tratándose, la gravedad del estado, 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 al día para lograr niveles sistemáticos apropiados de compuestos. Se prefiere generalmente que se use una dosis máxima, es decir, la dosis segura más alta según criterio médico sensato. Sin embargo, los expertos habituales en la técnica 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 portador farmacéuticamente aceptable para formar una preparación farmacéutica. La expresión "portador farmacéuticamente aceptable" tal como se usa en el presente documento significa una o más cargas, diluyentes o sustancias encapsulantes sólidas o líquidas compatibles, que son adecuadas para su administración en un ser humano. El término "portador" representa 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 descritas en el presente documento, y entre sí, de una manera tal que no hay interacción que afectaría sustancialmente a la eficacia farmacéutica deseada. En algunos aspectos, las preparaciones farmacéuticas comprenden un agente tal como se describe 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.

Hay una variedad de vías de administración disponibles. El modo particular seleccionado dependerá, por supuesto, del fármaco particular seleccionado, la gravedad del estado 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 es médicamente aceptable, lo que significa cualquier modo que produce 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, intraomental, intramuscular o infusión. Las vías intravenosa o intramuscular no son particularmente adecuadas para profilaxis y terapia a largo plazo. Como 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 pulverizaciones 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 portador 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 portador líquido, un portador sólido finamente dividido o ambos, y después, si es necesario, conformando el producto.

Pueden presentarse composiciones adecuadas para su administración oral como unidades diferenciadas, tales como cápsulas, comprimidos, pastillas para chupar, conteniendo cada 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.

Las composiciones adecuadas para su administración parenteral comprenden convenientemente una preparación acuosa estéril de un agente descrito en el presente documento, 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 suspensión o disolución inyectable estéril en un disolvente o diluyente 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 de cloruro de sodio isotónica. Además, se emplean convencionalmente aceites estériles fijos como medio disolvente o de suspensión. Con este fin puede emplearse cualquier aceite fijo suave incluyendo mono o di-glicéridos sintéticos. Además, pueden usarse ácidos grasos tales como ácido oleico en la preparación de productos 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 SEQ ID NO: 1), o un producto de expresión de la misma (por ejemplo, un péptido de SEQ ID NO: 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 en la célula.

La expresión "permitir la entrada" de una molécula en una a 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 usar 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, el uso 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 ácidos nucleicos 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 suministrar un ácido nucleico 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 suministro de ácido nucleico. Por ejemplo, cuando se emplean liposomas para suministrar los ácidos nucleicos, 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 ubicación intracelular y potencian la semivida intracelular, y similares. También se han usado satisfactoriamente sistemas de suministro poliméricos para suministrar ácidos nucleicos al interior de células, tal como conocen los expertos en la técnica. Tales sistemas permiten incluso la administración oral de ácidos nucleicos.

Otros sistemas de suministro pueden incluir sistemas de suministro de liberación en el tiempo, liberación retrasada o liberación sostenida. Tales sistemas pueden evitar administraciones repetidas de un agente tal como se describe en el presente documento, aumentando la conveniencia para el sujeto y el médico. Hay muchos tipos de sistemas de suministro de liberación disponibles y los conocen los expertos habituales en la técnica. Incluyen sistemas basados en polímero tales como poli(lactida-glicolida), copolioxalatos, policaprolactonas, poliésteramidas, poliortoésteres, poli(ácido hidroxibutírico) y polianhídridos. Se describen microcápsulas de los polímeros anteriores que contienen fármacos, por ejemplo, en la patente estadounidense 5.075.109. Los sistemas de suministro 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 sometidos a compresión 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 tal como se describe en el presente documento está contenido en una forma dentro de una matriz tal como se describe en las patentes estadounidenses n.ºs 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 estadounidenses n.ºs 3.854.480, 5.133.974 y 5.407.686. Además, pueden usarse sistemas de suministro de hardware basados en bombas, algunos de los cuales están adaptados para su implantación.

El uso de un implante de liberación sostenida a largo plazo puede ser deseable. La liberación a largo plazo, tal como se usa en el presente documento, significa que el implante se construye y dispone para suministrar niveles terapéuticos del principio activo durante al menos 30 días, y preferiblemente 60 días. Los 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 estadounidense n.º 4.748.024, y la patente canadiense n.º 1330939.

- 5 Puede realizarse la 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 tal como se describe en el presente documento para: (i) modular la actividad de generación de  $C_{\alpha}$ -formilglicina, y (ii) tratar cualquiera de los estados en los que participa la actividad de generación de  $C_{\alpha}$ -formilglicina de una molécula de la invención (por ejemplo, una deficiencia de sulfatasa incluyendo MSD). Agentes distintos de las moléculas de FGE incluyen iduronato 2-sulfatasa, 10 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.
- 15 "Coadministrar", tal como se usa en el presente documento se refiere a administrar simultáneamente dos o más compuestos tal como se describe en el presente documento (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 20 incluso sinérgico.

- Pueden proporcionarse matrices de moléculas de ácido nucleico en fase sólida. La matriz puede consistir 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 25 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 30 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 35 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.

- Puede usarse una técnica de hibridación convencional de tecnología de micromatrices para evaluar patrones de expresión de ácido nucleico e identificar la expresión de ácido nucleico. La tecnología de micromatriz, que también 40 se conoce por otros nombres incluyendo: tecnología de chip de ADN, tecnología de chip génico y tecnología de red de ácido nucleico en fase sólida, la conocen bien los 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 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, 45 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, etiquetas radioactivas, quimioluminiscentes o fluorescentes tales como fluoresceína, Cy3-dLTP o Cy5-dUTP), hibridar ácidos nucleicos diana con las sondas, y evaluar la hibridación diana-sonda. En general, una sonda con una secuencia de ácido nucleico que corresponde perfectamente con la secuencia diana dará como resultado la detección de una señal de 50 molécula indicadora más fuerte que sondas con correspondencias menos perfectas. Se presentan muchos componentes y técnicas usadas en la tecnología de micromatriz de ácido nucleico en The Chipping Forecast, Nature Genetics, Vol. 21, enero de 1999.

- 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 55 se prefiere un sustrato de vidrio. Pueden seleccionarse 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 de 20 a 25 meros y las sondas de ADN/ADNc tienen preferiblemente de 500 a 5000 bases de longitud, aunque pueden usarse otras longitudes. Un experto habitual en la 60 técnica 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 SEQ ID NO: 1, 3, 4, 6, 8, 10 y/o 12. Pueden purificarse sondas para eliminar contaminantes usando métodos convencionales conocidos por los expertos habituales en la técnica tales como filtración en gel o precipitación.
- 65 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 unir covalentemente el primer 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 fijan 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 fijándose la restante a los grupos reactivos. Estos grupos hidrocarbilo pueden contener hasta aproximadamente diez átomos de carbono, preferiblemente hasta aproximadamente seis átomos de carbono. Habitualmente se prefieren radicales alquileo que contienen de dos a cuatro átomos de carbono en la cadena principal. Estos y otros detalles del procedimiento se dan a conocer, por ejemplo, en la patente estadounidense 4.458.066.
- 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 suministro de precursores de nucleótidos al sustrato y posterior producción de sonda.
- 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, usando un robot controlado por ordenador para aplicar una 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 unirse covalentemente sondas al sustrato con métodos que incluyen, pero no se limitan a, irradiación UV. En otra realización se unen sondas al sustrato con calor.
- 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, puede unirse una o más moléculas de ácido nucleico de control al sustrato. Preferiblemente, 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.
- 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 micromatriz 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 software informático Permax, aunque puede usarse cualquier paquete estadístico convencional que puede 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 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 una enfermedad de expresión de deficiencia de sulfatasa relacionada con moléculas de ácido nucleico usando métodos de medición de proteínas para determinar la expresión de SEQ ID NO: 2, por ejemplo, determinando la expresión de polipéptidos codificados por SEQ ID NO: 1 y/o 3. Métodos preferidos para medir específica y cuantitativamente proteínas incluyen, pero no se limitan a: métodos basados en espectroscopia de masas tales como desorción/ionización por láser potenciada en superficie (SELDI; por ejemplo, sistema ProteinChip de Ciphergen), métodos no basados en espectroscopia de masas, y métodos basados en inmunohistoquímica tales como electroforesis en gel en dos dimensiones.
- La metodología SELDI puede usarse, mediante procedimientos conocidos por los 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 usarse ensayos basados en SELDI para caracterizar deficiencia múltiple de sulfatasas así como estadios de tales estados. 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 selección de proteínas (por ejemplo, con geles en 2-D), mediante "SELDI de proteína total" optimizada para visualizar los marcadores particulares de interés de entre SEQ ID NO: 1, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26 y/o 28. Pueden usarse modelos predictivos de una deficiencia de sulfatasa específica de medición de SELDI de múltiples marcadores de SEQ ID NO: 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 realizarse con métodos rutinarios conocidos por



los 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.

5 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) de 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  
10 célula en ausencia de la enzima generadora de formilglicina.

Una "sulfatasa con expresión aumentada", tal 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 tal como se describe en otra parte en el presente documento), ya que la regulación por incremento de cualquiera de esos procesos da como resultado 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 abarca el locus de sulfatasa, activación del gen de sulfatasa endógeno colocando, por ejemplo, un elemento promotor fuerte en sentido 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 estadounidenses n.ºs 5.733.761, 6.270.989 y 6.565.844), etc. Un control típico será una célula idéntica transfectada con un(os) plásmido(s) vector(es). Potenciar (o aumentar) la actividad 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 expresión de genes de sulfatasa también puede determinarse indirectamente detectando un cambio en la actividad sulfatasa.

De manera similar, una "enzima generadora de formilglicina con una expresión aumentada", tal como se usa en el presente documento, se refiere normalmente a expresión aumentada de un ácido nucleico de FGE 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 (tal como se describe en otra parte en el presente documento), ya que la regulación por incremento de cualquiera de esos procesos da como resultado 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 que necesita 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. Tal 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 tal como se describe en otra parte en el presente documento. Resultará evidente para un experto habitual en la técnica que una FGE puede entrar en contacto con una sulfatasa y modificarla si los ácidos nucleicos que codifican para FGE y una sulfatasa se coexpresan 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 coadministrarse 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 se necesita administrar cantidades inferiores 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. Sin embargo, se pretende que estos ejemplos ilustren meramente las realizaciones de la invención y no deben interpretarse como

limitativos del alcance de la invención.

## Ejemplos

### 5 Ejemplo 1:

*La deficiencia múltiple de sulfatasas está producida por mutaciones en el gen que codifica para la enzima generadora de C<sub>α</sub>-formilglutamina (FGE) humana*

### 10 Procedimientos experimentales

#### Materiales y métodos

#### Ensayo *in vitro* para FGE

15 Para monitorizar la actividad FGE, se usó el péptido P23 de 23-meros N-acetilado y C-amidado (MTDFYVPVSLCTPSRAALLTGRS) (SEQ ID NO: 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 μM de P23 en acetonitrilo al 30% y ácido trifluoroacético al 0,1% (TFA). En condiciones normales, se incubaron 6 pmol de P23 a 37°C con hasta 10 μl de enzima en un volumen final de 30 μl de Tris/HCl 50 mM, pH 9,0, que contenía NaCl 67 mM, CaCl<sub>2</sub> 15 μM, DTT 2 mM, y albúmina sérica bovina 0,33 mg/ml. Para detener la reacción enzimática, se añadieron 1,5 μl de TFA al 10%. Entonces P23 se unió a ZipTip C18 (Millipore), se lavó con TFA al 0,1% y se eluyó en 3 μl de acetonitrilo al 50%, TFA al 0,1%. Se mezclaron 0,5 μl del eluato con 0,5 μl de disolución de matriz (ácido α-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 espectrómetro 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<sup>+</sup> 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

35 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 licuadora Waring Blendor y mediante tres rondas de licuado a motor. Se realizó la preparación de microsomas rugosos (RM) mediante fraccionamiento celular del homogenizado obtenido tal 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 Ti45, 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 tras la centrifugación durante 210 minutos a 45000 rpm en un rotor Ti45 a 4°C. Habitualmente, se obtuvieron 100000-150000 equivalentes de RM, tal 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, tal 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 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 habían 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. Tras 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 (PVSLPTRSRAALLTGR) (SEQ ID NO: 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). Tras 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) (SEQ ID NO: 35) y se lavaron con tampón A. Tras 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 substituyó una alícuota del eluato con albúmina sérica bovina 1 mg/ml, se dializó frente al tampón A y se analizó para determinar 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 tras las etapas cromatográficas mediante SDS-PAGE (15% de acrilamida, 0,16% de bisacrilamida) y tinción con SYPRO Ruby (Bio-Rad Laboratories).

## 10 Identificación de FGE mediante espectrometría de masas

Para el análisis de huella de masa peptídica, se digirieron los polipéptidos purificados colocados 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 obtenidos a partir 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 convencionales 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 Human Mouse Homology Map (mapa de homología entre ser humano y ratón) también de NCBI, (Bethesda, MD).

## 35 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™ (Qiagen, Inc., Valencia, CA) y o bien un cebador de oligo(dT) o bien el cebador específico de FGE, 1199nc (CCAATGTAGGTCAGACACG) (SEQ ID NO: 36). Se amplificó el ADNc de primera hebra mediante PCR usando el cebador directo 1c (ACATGGCCCCGCGGGAC) (SEQ ID NO: 37) y, como cebador inverso, o bien 1199nc o bien 1182nc (CGACTGCTCCTTGGACTGG) (SEQ ID NO: 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 (SEQ ID NO: 1 y 3).

Detección de mutaciones, secuenciación genómica, mutagénesis dirigida al sitio y análisis de transferencia de tipo Northern

Los protocolos convencionales utilizados en este estudio fueron esencialmente tal 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 de tipo 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.

## 55 Líneas celulares y cultivo celular

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 las células CHO, BHK21, HT-1080 y fibroblastos de piel humana a 37°C en 5% de CO<sub>2</sub> en medio Eagle modificado por Dulbecco que contenía un 10% de suero bovino fetal.

Transfección, inmunofluorescencia indirecta, análisis de inmunotransferencia de tipo Western y detección de actividad FEG

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) (SEQ ID NO: 39), CCCAAGCTTATGCGTAGTCAGGCACATCATACGGATAGTCCATGGTGGGCAGGC (HA) (SEQ ID NO: 40), CCCAAGCTTACAGGTCTTCTTCAGAAATCAGCTTTTGTTCGTCCATGGTGGGCAG GC (c-Myc) (SEQ ID NO: 41), CCCAAGCTTAGTGATGGTGATGGTGATGCGATC CTCTGTCCATGGTGGGCAGGC (RGS-His<sub>6</sub>) (SEQ ID NO: 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 tras la transfección, se analizaron las células mediante inmunofluorescencia indirecta tal como se describió anteriormente (Lübke *et al.*, Nat. Gen., 2001, 28:73-76; Hansske *et al.*, J. Clin. Invest., 2002, 109:725-733), usando anticuerpos monoclonales IgG1 frente a 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 endoplasmático, 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 inmunotransferencia tipo Western se usaron los mismos anticuerpos monoclonales y una IgG anti-ratón conjugada con HRP como anticuerpo secundario. Para la determinación de la actividad 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 µM, aprotinina 0,16 µM, leupeptina 4,2 µM, bestatina 7,2 µM, pepstatina A 3 µM, E-64 2.8 µM E-64), 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 una décima parte del volumen reunido original antes de la determinación de la actividad FEG con el péptido P23.

#### Transducción retroviral

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 tal 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 PT63 transducidas en pLPCX con puomicina antes de la determinación de las actividades sulfatasa.

#### Ensayos de sulfatasa

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

#### Resultados

Un ensayo basado en péptidos rápido para determinar la actividad FEG

Se ha 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 triptica, separación de los péptidos mediante RP-HPLC e identificación y cuantificación de péptido triptico 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 de 16-meros que correspondía a los residuos 65-80 de ASA y que contenía el motivo de secuencia requerido para la formación de FGly inhibió la actividad FEG en el ensayo *in vitro*. Esto sugirió que péptidos tales como ASA65-80 pueden servir como sustratos para FGE. Se sintetizó el péptido de 23-meros P23 (SEQ ID NO: 33), que corresponde a 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 terminales, respectivamente. Las formas que contienen FGly y la cisteína 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<sub>m</sub> para P23 fue de 13 nM. Los efectos del glutatión reducido y oxidado, Ca<sup>2+</sup> 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

5 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 se mezcló esta fracción con concanavalina A-Sepharose, se unió FGE. Pudo eluirse el 30-40% de la actividad de partida con  $\alpha$ -metilmanósido 0,5 M. Las dos etapas de purificación finales fueron cromatografía en matrices de afinidad derivatizadas con péptidos de 16-meros. La primera matriz de afinidad fue Affigel 10 sustituida con una variante 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 -SEQ ID NO: 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 conservó la actividad FEG. La cromatografía en la matriz de afinidad de péptido reorganizado eliminó las proteínas de unión al péptido incluyendo las chaperonas en el retículo endoplasmático. La segunda matriz de afinidad fue Affigel 10 sustituida con una variante del péptido ASA65-80, en el que se sustituyó la Cys69 por una serina (péptido Ser69 PVSLSTPSRAALLTGR-SEQ ID NO: 35). La matriz de afinidad del péptido Ser69 unió de manera eficaz FGE. La actividad 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 de tipo natural 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 figura 2, se muestra el patrón de polipéptido del material de partida y las fracciones activas obtenidas tras 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).

30 Los polipéptidos purificados de 39,5 y 41,5 kDa se codifican 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 tripticos de los dos polipéptidos obtenidos mediante espectrometría de masas MALDI-TOF fueron en su mayor parte solapantes, lo que sugiere que las dos proteínas se originan a partir del mismo gen. Entre los péptidos tripticos de ambos polipéptidos se encontraron dos péptidos abundantes  $MH^+$  1580,73, SQNTPDSSASNLGFR (SEQ ID NO: 43) y  $MH^+$  2049,91, MVPIPAGVFTMGTDPPQIK - SEQ ID NO: 44 más dos oxidaciones de metionina), lo que correspondía a la proteína codificada por un ADNc con n.º de registro de GenBank AK075459 (SEQ ID NO: 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 ortóloga bovina del ADNc humano que cubre la parte C-terminal de la FGE y que corresponde a las secuencias de ambos péptidos proporcionó la información de secuencia adicional para la FGE bovina.

#### Conservación evolutiva y estructura de dominio de FGE

45 El gen para la FGE humana está codificado por el ADNc de (SEQ ID NO: 1 y/o 3) y está ubicado 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 ubicado 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 procariotas se encontraron 12 homólogos de FGE humana. Se predijo que el ADNc para la FGE humana codificaba para una proteína de 374 residuos (figura 3 y SEQ ID NO: 2). La proteína contiene una secuencia señal escindible de 33 residuos, lo que indica translocación de FGE al interior del retículo endoplasmático, 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.

65 El subdominio N-terminal (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 C-terminal (residuos 327-366 en FGE humana) es la secuencia más altamente conservada dentro de la familia de FGE. La identidad de secuencia del subdominio C-terminal 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 procariotas. Los 12 miembros procariotas 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 procariota de FGE debido a secuencias no conservadas de longitudes variables que separan los subdominios entre sí. El genoma humano y de ratón codifican para dos homólogos estrechamente relacionados de FGE (SEQ ID NO: 43 y 44, n.º de registro de GenBank NM\_015411, en el hombre, y SEQ ID NO: 45 y 46, n.º de registro de GenBank AK076022, en ratón). Los dos parálogos son idénticos en un 86%. Sus genes están ubicados en regiones sinténicas 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, ubicación subcelular y formas moleculares

Puede detectarse un único transcrito de 2,1 kb mediante análisis de transferencia de tipo 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 sometieron a ensayo 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-terminal por un marcador de HA, Myc o His<sub>6</sub> para determinar la actividad FEG y la ubicación 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 coubicación de las formas de FGE marcadas de manera diversa con la proteína disulfuro isomerasa, una proteína luminal del retículo endoplasmático. El análisis de inmunotransferencia tipo Western de los extractos de células BHK21 transfectadas de manera transitoria con ADNc que codifica para las formas marcadas de FGE mostró una única banda inmunoreactiva con un tamaño aparente de entre 42 y 44 kDa.

#### El gen de FGE porta mutaciones en MSD

La MSD está producida por una deficiencia para generar residuos de *FGly* en sulfatasas (Schmidt, B., *et al.*, Cell, 1995, 82: 271-278). El gen de FGE, por tanto, es un gen candidato para la MSD. Se amplificó y se secuenció el ADNc que codifica para FGE de siete pacientes con MSD y se encontraron diez mutaciones diferentes que se confirmaron mediante secuenciación del ADN genómico (tabla 1).

Tabla 1: Mutaciones en pacientes con MSD

Mutación	Efecto sobre la proteína	Observaciones	Paciente
1076C>A	S359X	Truncamiento de los 16 residuos del extremo C-terminal	1*
IVS3+5-8 del	Deleción de residuos 149-173	Deleción en marco del exón 3	1, 2
979C>T	R327X	Pérdida del subdominio 3	2
1045C>T	R349W	Substitución de un residuo conservado en el subdominio 3	3, 7
1046G>A	R349Q	Substitución de un residuo conservado en el subdominio 3	4
1006T>C	C336R	Substitución de un residuo conservado en el subdominio 3	4
836C>T	A279V	Substitución de un residuo conservado en el subdominio 2	5
243delC	cambio de marco y truncamiento	Pérdida de los tres subdominios	5
661delG	cambio de marco y truncamiento	Pérdida del tercero del C-terminal de FGE que incluye el subdominio 3	6**
IVS6-1G>A	Deleción de residuos 281-318	Deleción en marco del exón 7	5

\*El paciente 1 es el paciente con MSD Mo. en Schmidt, B., *et al.*, Cell, 1995, 82:271-278 y Rommerskirch y von Figura, Proc. Natl. Acad. Sci., EE.UU., 1992, 89:2561-2565.

\*\*El paciente 6 es el paciente con MSD notificado por Burk *et al.*, J. Pediatr., 1984, 104:574-578.

Los otros pacientes representan casos no publicados.

El primer 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 deleción de los 25 residuos 149-173 que están codificados por el exón 3 y separan los dominios primero y segundo de la proteína. La secuenciación genómica

reveló una delección de nucleótidos +5-8 del tercer intrón (IVS3+5-8 del) destruyéndose de ese modo 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 trunca codificada por el alelo 979C>T carece de la mayor parte del subdominio 3. El tercer paciente era homocigoto para una sustitución 1045C>T 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 tras la prolina 81 y que produce una detención de la traducción tras el 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 para 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 pacientes control y 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 de tipo natural 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, se transdujo el vector retroviral sin inserto de ADNc. Para monitorizar la complementación del defecto metabólico con la actividad de ASA, se midió la esteroide sulfatasa (STS) y la N-acetilgalactosamina 6-sulfatasa (GalNAc6S) en los fibroblastos transducidos antes o tras la selección. La transducción de la FGE de tipo natural restableció 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, el restablecimiento fue sólo parcial tras la selección de los fibroblastos que alcanzaban del 20 al 50% de la actividad normal. Para la STS, se encontró que la actividad se restablecía a la de los fibroblastos control tras la selección. La selección aumentó la actividad de ASA y STS en del 50 al 80%, lo 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 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 restableció las actividades sulfatasa (tabla 2).

Tabla 2: Complementación de fibroblastos con MSD por transducción de ADNc de FGE de tipo natural o mutante

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

<sup>1</sup>Los valores proporcionan la razón entre ASA (mU/mg de proteína celular), STS (μU/mg de proteína celular), GalNAc6S (μU/mg de proteína celular) y la de β-hexosaminidasa (U/mg de proteína celular). Para los fibroblastos control se proporciona la media y la variación de 6-11 líneas celulares. Cuando está indicado, se proporciona el intervalo de dos cultivos transducidos en paralelo para fibroblastos con MSD.

<sup>o</sup> El número de fibroblastos con MSD se refiere al del paciente en la tabla 1.

<sup>+</sup> Determinación de la actividad antes de la selección.

<sup>++</sup> Determinación de la actividad tras la selección.

n.d.: no determinado

#### Discusión

FGE es una glicoproteína altamente conservada del retículo endoplasmático.

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 substitución 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 la translocación a través de la membrana del retículo endoplasmático. 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 extrae electrones del grupo tiol de la cisteína y los transfiere a un aceptor. El tioaldehído 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 donador de electrones tal como FADH<sub>2</sub>. El derivado de hidrato de tioaldehído 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 comprobar si los elementos estructurales que median el reconocimiento del motivo *FGly* y la unión de un aceptor de electrones o un donador de electrones se correlacionan con la estructura de dominio de FGE.

FGE recombinante se ubica en el retículo endoplasmático, lo 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 endoplasmático (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 endoplasmático 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.

#### Las mutaciones en FGE producen MSD

Se ha demostrado que las mutaciones en el gen que codifica para FGE producen MSD. La FGE también puede interaccionar con otros componentes, y los defectos en genes que codifican para estos últimos podrían producir MSD de igual forma. En siete pacientes con MSD se 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-terminal de diversas longitudes (cuatro mutaciones) o grandes deleciones en macro (dos mutaciones). Para dos líneas celulares con MSD y una de las mutaciones en MSD se demostró que la transducción del ADNc de FEG de tipo natural, pero no del mutante, restablece parcialmente las actividades 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 una forma común de hidrocefalia 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. Desde el punto de vista bioquímico, 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 notificado que la actividad residual se somete a variaciones dependiendo de las condiciones del cultivo celular y de factores desconocidos. Desde el punto de vista bioquímico, 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 notificados en el presente documento se encuentran dentro del grupo I, excepto el paciente 5, que pertenece al 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 deleción 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 una mejor comprensión de la correlación



genotipo-fenotipo en MSD.

## Ejemplo 2:

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

### Bioinformática

- 10 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).

- 15 Se detectaron los dominios funcionales buscando mediante la búsqueda de modelos ocultos de Harkov 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 alineaciones 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 (*tree bootstrapping*) para obtener resultados significativos. Se visualizaron los árboles usando Njplot (Perriere, G. y Gouy, M., Biochimie, 1996, 78:364-369). Las alineaciones se representaron gráficamente usando el comando prettyplot de EMBOSS.

- 25 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 de  $10^{-40}$ , y parámetros por defecto. Se alcanzó la convergencia tras 5 iteraciones. Para las secuencias de nucleótidos, se realizó la búsqueda con Psi-TBlastn: usando NR y la secuencia proteica 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  $10^{-20}$ .

- 35 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).

- 40 Se calcularon puntuaciones de similitud de los subdominios a partir de las alineaciones usando el algoritmo cons. de EMBOSS con parámetros por defecto. Se generaron las meta-alineaciones mediante la alineación de 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.

## 45 Resultados y discusión

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

- 50 El gen de FGE humana (SEQ ID NO: 1, 3) codifica para la proteína FGE (SEQ ID NO: 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-terminal se escinden tras 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 tras el residuo 34. La proteína tiene un único sitio posible de N-glicosilación en Asn 141.

- 55 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  $7,9 \cdot 10^{-114}$ . 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 una alineación múltiple de FGE con las secuencias de la semilla DUF323. Basándose en esto, se construyó y remuestreó 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 la FGE humana y los genes restantes, menos relacionados.

Las 7 proteínas superiores tienen una distancia filogenético de entre 0,41 y 0,73 con respecto a la FGE humana. Sólo 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 filogenético de entre 1,14 y 1,93). Su dominio DUF323 diverge considerablemente del dominio DUF323 altamente conservado del primer 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:O84147) 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 para 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. Tal como se comenta en otra parte en el presente documento, la FGE también podría ejercer su función de modificación de cisteína como una oxigenasa, lo que sugiere 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 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 procariotas y dos parálogos en el hombre y el ratón (tabla 3).

Tabla 3: La familia de genes de FGE en eucariotas y procariotas

SEQ ID NO: NA, AA [GI]	ESPECIE	LONGITUD [AA]	SUBGRUPO
1/3, 2	<i>Homo sapiens</i>	374	E1
49, 50 [22122361]	<i>Mus musculus</i>	372f	E1
51, 52 [20130397]	<i>Drosophila melanogaster</i>	336	E1
53, 54 [21289310]	<i>Anopheles gambiae</i>	290	E1
47, 48 [26344956]	<i>Mus musculus</i>	308	E2
45, 46 [24308053]	<i>Homo sapiens</i>	301	E2
55, 56 [21225812]	<i>Streptomyces coelicolor</i> A3(2)	314	P1
57, 58 [25028125]	<i>Corynebacterium efficiens</i> YS-314	334	P1
59, 60 [23108562]	<i>Novosphingobium aromaticivorans</i>	338	P2
61, 62 [13474559]	<i>Mesorhizobium loti</i>	372	P2
63, 64 [22988809]	<i>Burkholderia fungorum</i>	416	P2
65, 66 [16264068]	<i>Sinorhizobium meliloti</i>	303	P2

67, 68 [14518334]	<i>Microscilla sp.</i>	354	P2
69, 70 [26990068]	<i>Pseudomonas putida</i> KT2440	291	P2
71, 72 [22975289]	<i>Ralstonia metallidurans</i>	259	P2
73, 74 [23132010]	<i>Prochlorococcus marinus</i>	291	P2
75, 76 [16125425]	<i>Caulobacter crescentus</i> CB15	338	P2
77, 78 [15607852]	<i>Mycobacterium tuberculosis</i> Ht37Rv	299	P2

GI - Identificador de proteínas de GenBank

NA - ácido nucleico AA - aminoácidos,

E1 - ortólogos eucariotas E2 - parálogos eucariotas

P1 - ortólogos procariotas estrechamente relacionados P2 - otros ortólogos procariotas

f- secuencia de proteínas mal predicha en GenBank

Obsérvese que en GenBank se predice que la secuencia GI 22122361 de ratón codifica para una proteína de 284 aa, aunque la secuencia NM 145937 de ADNc codifica para una proteína de 372 residuos. Esta mala predicción se basa en la omisión del primer exón del gen murino de FGE. Todas las secuencias encontradas en la base de datos NR son de procariotas 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' terminal un alto grado de conservación a nivel de proteína que no se enumeran en la base de datos NR. Estas secuencias no abarcan la parte codificante completa de los ARNm y proceden todos ellos a eucariotas superiores (tabla 4).

Tabla 4: Fragmentos EST ortólogos de FGE en eucariotas

SEQ ID NO: NA [GB]	ESPECIE
80 [CA379852]	<i>Oncorhynchus mykiss</i>
81 [AI721440]	<i>Danio rerio</i>
82 [BJ505402]	<i>Oryzias latipes</i>
83 [BJ054666]	<i>Xenopus laevis</i>
84 [AL892419]	<i>Silurana tropicalis</i>
85 [CA064079]	<i>Salmo salar</i>
86 [BF189614]	<i>Sus scrofa</i>
87 [AV609121]	<i>Bos taurus</i>

GB - n.º de registro de GenBank; NA - ácido nucleico

La alineación 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 (parólogo 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 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.

- 5 Tal como se comenta 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* y contribuir a las actividades residuales de las sulfatasas encontradas en pacientes con MSD.

#### Subdominios de FGE

- 10 Los miembros de la familia de genes de FGE tienen tres partes/dominios altamente conservados (tal como se describe en otra parte en el presente documento). Además de las dos secuencias no conservadas que separan a la primera, tienen extensiones no conservadas en los extremos N y C terminales. Se considera que las tres partes representan subdominios del dominio DUF323 porque 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 terminales muestran una variación incluso más fuerte en la longitud (extremo N-terminal: 0-90 AA, extremo C-terminal: 0-28 AA). La secuencia para el gen de FGE de *Ralstonia metallidurans* está probablemente incompleta, puesto que carece del primer subdominio.
- 15
- 20 Para verificar la plausibilidad de definir subdominios de DUF323, se realizó 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 lo que confirma la predicción de señal del algoritmo de von-Heijne. La región no conservada N-terminal (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
- 25 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 láminas  $\beta$ . El primer subdominio está constituido por varias láminas  $\beta$  y una hélice  $\alpha$ , el segundo subdominio contiene dos láminas  $\beta$  y cuatro hélices  $\alpha$ . El tercer subdominio tiene una región de hélice  $\alpha$  flanqueada por una lámina al comienzo 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 láminas  $\beta$ .
- 30

- 35 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.

Tabla 5: Homología (% de similitud) de los subdominios de la familia de FGE

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

- 40 P1 - ortólogos procariotas estrechamente relacionados; P2 - otros ortólogos procariotas

- El primer subdominio de la familia de FGE muestra la homología más débil en todos los subgrupos. En los ortólogos eucariotas porta el sitio de N-glicosilación: en el residuo Asn 141 en ser humano, en Asn 139 en el ratón y en 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 en Asn 130 en *Anopheles*. Por tanto, es necesario volver a secuenciar la secuencia que abarca 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.
- 45

- 50 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 que dan lugar a la enfermedad en pacientes con MSD. Siete de las nueve mutaciones identificadas en seis pacientes con MSD descritos en el ejemplo 1 están ubicadas en secuencias que codifican para los 40 residuos del subdominio 3. Los residuos contienen cuatro
- 55 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 (SEQ ID NO: 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 primero 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 (SEQ ID NO: 79) y el motivo GFR terminal.

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

El gen de FGE humana está ubicado 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 ubicado 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 de ser humano y el 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 para LMCD1, KIAA0212, ITPR1, AXCAM, y IL5RA. En la dirección centromérica, ambos loci de FGE están flanqueados por los loci de CAV3 y OXTR.

#### 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 para una proteína citosólica que contiene motivos de agrupamiento (*cluster*) 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).

Por tanto, fue de interés examinar si los genes de FGE procariotas estaban ubicados 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, lo 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 han sido ORF solapantes, lo 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 tienen sulfatasas de tipo cisteína (*S. coelicolor*: PID 24413927, *M. luti*: 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 microorganismos, 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 lisosomal 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 actúan 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 sulfatasa en líneas celulares que sobreexpresan una sulfatasa

Se desea examinar los efectos de FGE sobre células que expresan/sobreexpresan 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 un constructo de expresión de FGE, pXMG.1.3 (tabla 7 y figura 4) o un plásmido 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 sometieron a prueba las muestras de medio para determinar la actividad sulfatasa respectiva mediante ensayo de actividad y el nivel de proteína sulfatasa total estimado mediante ELISA específico para o bien iduronato 2-sulfatasa o N-acetilgalactosamina 6-sulfatasa.

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

Cepa celular	Plásmido	Sulfatasa expresada
36F	XFM4A.1	N-acetilgalactosamina 6-sulfatasa
30C6	pXI2S6	Iduronato 2-sulfatasa

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

Plásmido	Configuración de elementos de secuencia de ADN principales*
pXMG.1.3 (expresión de FGE)	> Potenciador/promotor de CMV de 1,6 kb > ADNc de FGE de 1,1 kb > secuencia no traducida de hGH3' < amp < casete de DHFR < casete de Cneo (neomicina fosfotransferasa)
pXMG.1.2 (control, orientación inversa de FGE)	> Potenciador/promotor de CMV de 1,6 kb < ADNc de FGE de 1,1 kb < secuencia no traducida de hGH3' < amp < casete de DHFR < casete de Cneo (neomicina fosfotransferasa)

\* > indica orientación de 5' a 3'

#### Procedimientos experimentales

#### Materiales y métodos

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

Se recogieron células HT-1080 hasta obtener  $9-12 \times 10^6$  células para cada electroporación. Se transfectaron dos plásmidos por duplicado: uno para someterse a prueba (FGE) y un control; en este caso, el plásmido control contenía el ADNc de FGE clonado en 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.

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. Tras 24 horas, se volvió a alimentar el matraz con 20 ml de DMEM/10 para eliminar las células muertas. 48-72 horas tras la transfección, se recogieron muestras de los medios y se recogieron las células de los matraces T75 por duplicado.

#### Preparación del medio

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

Se transfectaron las células en medios sin metotrexato (MTX). 24 horas después, se volvió a alimentar a las células con medios que contenían las cantidades apropiadas de MTX (36F = MTX 10 µM, 30C6 = MTX 0,1 M). Se recogió el medio y se recogieron las células 24, 48, y 72 horas tras volver a alimentar.

#### Ensayos de actividad

Iduronato 2-sulfatasa (I2S). Se equilibraron columnas de desalación NAP5 (Amersham Pharmacia Biotech AB, Uppsala, Suecia) con tampón de diálisis (acetato de sodio 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 de tampón de diálisis. Se diluyeron adicionalmente las muestras desaladas hasta aproximadamente I2S 100 ng/ml tampón de reacción (acetato de sodio 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-umbeliferilo (Fluka, Buchs, Suiza) en tampón de sustrato (acetato de sodio 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. Tras la incubación, se añadieron 190 µl de tampón de parada (glicina 332,5 mM, carbonato de sodio 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 µL 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 µL de tampón de parada a cada pocillo restante en la placa de 96 pocillos. Se realizaron diluciones en serie dos veces 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 extrapola la fluorescencia de las muestras desconocidas a partir de esta curva. Los resultados se notifican como unidades/ml en las que una unidad de actividad era igual a 1 µmol de 4-MUF producido por minuto a 37°C.

N-acetilgalactosamina 6-sulfatasa (GALNS). El ensayo de actividad GALNS hace uso del sustrato fluorescente 4-metilumbeliferil-β-D-galactopiranosido-6-sulfato (Toronto Research Chemicals Inc., n.º de catálogo M33448). El ensayo comprendió dos etapas. En la primera etapa, se incubaron 75 µl del sustrato 1,3 mM preparado en tampón de reacción (acetato de sodio 0,1 M, cloruro de sodio 0,1 M, pH 4,3) durante 4 horas a 37°C con 10 µl de muestra de proteína/medios 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 GALNS. Tras la adición de aproximadamente 500 U de β-galactosidasa de *Aspergillus oryzae* (Sigma, n.º 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 mediante el uso de 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, n.º de catálogo M1508) que sirve como patrón de referencia. Una unidad de la actividad corresponde a nmoles de 4-metilumbeliferona liberados por hora.

#### 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 hasta 10 µg/ml en bicarbonato de sodio 50 mM 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 convencional. 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. Tras 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, hasta 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 convencional. Se usó el primer anticuerpo para la captura de GALNS para recubrir una placa F96 MaxiSorp Nunc-Immuno (Nalge Nunc, n.º de catálogo 442404) en un tampón de recubrimiento (bicarbonato de sodio 50 mM, pH 9,6). Tras la incubación durante una hora a 37°C y el lavado con un tampón de lavado, se bloqueó la placa con un 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. Tras el lavado 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. Tras 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 se obtuvieron los resultados espectrofotométricamente usando un lector de placas de Molecular Device a una longitud de onda de 450 nm.

## Discusión

### Efecto de FGE sobre la actividad sulfatasa

5 GALNS. Se observó un aumento de aproximadamente 50 veces en la actividad GALNS total con respecto a los niveles control (figura 5). Se observó este nivel de actividad aumentada con los tres puntos de tiempo de toma de muestras de medio. Además, la actividad 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 puntos de tiempo de 48 horas y 72 horas.

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

20 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 puntos de tiempo en el estudio y pareció aumentar durante los tres días de acumulación tras la transfección.

25 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 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 afecta a las rutas de traducción y secreción implicadas en la producción de sulfatasa.

30 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 sobreexpresan GALNS e I2S.

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

35 Para someter a prueba el efecto de FGE (SUMF1) sobre actividades sulfatasa adicionales en células normales se sobreexpresó ADNc de ARSA (SEQ ID NO: 14), ARSC (SEQ ID NO: 18) y ARSE (SEQ ID NO: 22) en diversas líneas celulares con y sin cotransfección del ADNc de FGE (SUMF1) y se midieron las actividades sulfatasa. La sobreexpresión de ADNc de sulfatasa en células Cos-7 dio como resultado un aumento moderado de la actividad sulfatasa, mientras que se observó un sorprendente aumento sinérgico (de 20 a 50 veces) cuando se coexpresaron 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 sobreexpresión simultánea de múltiples ADNc de sulfatasa dio como resultado un aumento inferior de cada actividad sulfatasa específica en comparación con la sobreexpresión de una única sulfatasa, lo que indica la presencia de competición de las diferentes sulfatasas por la maquinaria de modificación.

50 Para someter a prueba la conservación funcional del gen de FGE (SUMF1) durante la evolución se sobreexpresó ADNc de ARSA, ARSC y ARSE en diversas líneas celulares con y sin cotransfección del ADNc de MSD y se midieron las actividades 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, lo 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), lo que sugiere 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 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:

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

65 En un experimento separado usando transferencia de cromosomas mediada por microcélulas por medio de complementación funcional se confirmó que el gen mutado en MSD es FGE. Los hallazgos 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 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.

#### 5 El gen para MSD mapea en el cromosoma 3p26

Para identificar la ubicación cromosómica del gen mutado en MSD, se intentó 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 seleccionable 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. Aproximadamente se recogieron 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 sometieron a prueba las actividades arilsulfatasa A (ARSA) (SEQ ID NO: 15), arilsulfatasa B (ARSB) (SEQ ID NO: 17), y arilsulfatasa C (ARSC) (SEQ ID NO: 19) para cada uno de los aproximadamente 440 clones (20 x 22). Este análisis indicó claramente que las actividades 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 donadora, se usó 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 partiendo de la base de tener diferentes alelos entre las líneas celulares donadoras y del paciente. Esto permitió examinar la presencia del cromosoma donador e identificar la posible pérdida de regiones específicas debido a la rotura cromosómica imprevista. Cada clon que tenía alta actividad enzimática conservaba el cromosoma 3 completo procedente de la línea celular donadora, mientras que los clones con bajas actividades parecían haber perdido el cromosoma completo partiendo de la base de la ausencia de alelos del cromosoma 3 procedentes de la línea celular donadora. Estos últimos clones conservaban probablemente una región pequeña del cromosoma donador que contenía el gen marcador seleccionable que les permitía sobrevivir en medio que contenía higromicina. Estos datos indican que un cromosoma 3 humano normal podía complementar el defecto observado en la línea celular del paciente con MSD.

Para determinar la región cromosómica específica que contenía el gen responsable para la actividad de complementación, se usaron híbridos del cromosoma 3 marcados con Neo que se encontró que habían perdido varias partes del cromosoma. Además, se realizó transferencia de cromosomas mediada por microcélulas irradiadas de cromosomas 3 humanos marcados con HyTK. Se sometieron a prueba ciento quince híbridos irradiados de cromosoma 3 para determinar las actividades sulfatasa y se obtuvo el genotipo usando un panel de 31 marcadores microsatélites polimórficos que abarcaban el cromosoma completo. Todos los clones que presentaban altas actividades enzimáticas parecían tener el cromosoma 3p26 conservado. Un análisis de resolución superior usando marcadores adicionales de esta región mapeó la posible ubicación para el gen de complementación entre los marcadores D3S3630 y D3S2397.

#### Identificación del gen mutado en MSD

Se investigaron los genes de la región genómica 3p26 para determinar 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 anteriormente y siete casos no publicados. Se identificaron varias mutaciones de la cohorte de MSD en el marcador de secuencia expresada (EST) AK075459 (SEQ ID NO: 4,5), que correspondía a un gen de función desconocida, lo que sugiere fuertemente que éste era el gen implicado en MSD. Se encontró que cada mutación estaba ausente en 100 individuos 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 al corte y empalme. También se identificaron mutaciones de cambio de marco, terminadora, de corte y empalme y de aminoácido, lo que sugiere que la enfermedad está producida por un mecanismo de pérdida de función, tal 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).

Tabla 8: Mutaciones en MSD adicionales identificadas

Caso	referencia	fenotipo	exón	cambio de nucleótido	cambio de aminoácido
1. BA426	Conary <i>et al</i> , 1988	moderado	3	463T>C	S 155P
			3	463T>C	S155P
2. BA428	Burch <i>et al</i> , 1986	neonatal grave	5	661delG	cambio de marco

3. BA431	Zenger <i>et al</i> , 1989	moderado	1	2T>G	MIR
			2	276delC	cambio de marco
4. BA799	Burk <i>et al</i> , 1981	leve-moderado	3	463T>C	S155P
			3	463T>C	S155P
5. BA806	no publicado	neonatal grave	9	1045T>C	R349W
6. BA807	Schmidt <i>et al</i> , 1995	desconocido	3	c519+4delGTAA	salto de ex 3
			9	1076C>A	S359X
7. BA809	Couchot <i>et al</i> , 1974	leve-moderado	1	1A>G	M1V
			9	1042G>C	A348P
8. BA810	no publicado	grave	8	1006T>C	C336R
			9	1046G>A	R349Q
9. BA811	no publicado	neonatal grave	3	c519+4delGTAA	salto de ex 3
			8	979C>T	R327X
10. BA815	no publicado	moderado	5	c.603-6delC	salto de ex 6
			6	836C>T	A279V
11. BA919	no publicado	leve-moderado	9	1033C> T	R345C
			9	1033C>T	R345C
12. BA920	no publicado	moderado	5	653G>A	C218Y
			9	1033C>T	R345C

Se identificaron mutaciones en cada paciente con MSD sometido a prueba, excluyendo así la heterogeneidad de locus. No se observó correlación obvia entre los tipos de mutaciones identificados y la gravedad del fenotipo notificado en los pacientes, lo que sugiere que la variabilidad clínica no está producida por heterogeneidad alélica.

- 5 En tres casos, se encontró que diferentes pacientes (casos 1 y 4, casos 6 y 9, y casos 11 y 12 en la tabla 6) portaban la misma mutación. Dos de estos pacientes (casos 11 y 12) proceden de la misma ciudad en Sicilia, lo que sugiere 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 sólo algunos eran homocigotos. Aunque compatible con la ausencia de consanguinidad notificada 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

- 15 Se reunió la secuencia de ADNc consenso del ADNc de la FGE humana (también usada de manera intercambiable en el presente documento como SUMF1) (SEQ ID NO: 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 ubicado en el cromosoma 7 humano que se designó FGE2 (también usado de manera intercambiable en el presente documento como SUMF2) (SEQ ID NO: 45, 46).

Complementación funcional de deficiencias de sulfatasa

- 25 Se infectaron fibroblastos de dos pacientes (casos 1 y 12 en la tabla 8) con MSD en los que se identificaron mutaciones del gen de FGE (SUMF1) (líneas celulares BA426 y BA920) con virus VHS que contenían el tipo natural y dos formas mutada del ADNc de FGE (SUMF1) (R327X y  $\Delta$ ex3). Se sometieron a prueba las actividades ARSA, ARSB y ARSC 72 horas tras la infección. La expresión del ADNc de FGE de tipo natural (SUMF1) dio como resultado 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 dan como resultado deficiencia de sulfatasa, demostrando por tanto que FGE (SUMF1) es un factor esencial para la actividad sulfatasa.

Tabla 9: Complementación funcional de deficiencias de sulfatasa

Línea celular receptora con MSD	Constructo	ARSA <sup>(1)</sup>	ARSB <sup>(1)</sup>	ARSC <sup>(1)</sup>
BA426	amplicón de VHS	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 (4,9 x)	37,8 (1,7 x)	0,62 (4,1 x)
BA920	amplicón de VHS	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 (4,0 x)	21,6 (1,9 x)	0,42 (2,8 x)
Intervalo de control		123,7 – 394,6	50,6 – 60,7	1,80 – 1,58

<sup>(1)</sup>Todas las actividades enzimáticas se expresan como nmoles de 4-metilumbeliferona liberados por mg de proteína<sup>-1</sup>. 3 h. Se infectaron las líneas celulares BA426 y BA920 de MSD con el amplicón de VHS solo, y con constructos que portaban o bien el ADNc de SUMF1 mutante o bien de tipo natural. El aumento de las actividades arilsulfatasas individuales en fibroblastos infectados con el gen de SUMF1 de tipo natural, en comparación con el de las células infectadas con el vector solo, se indica entre paréntesis. Se indican las actividades medidas en los fibroblastos control no infectados.

#### Base molecular de MSD

- 10 Basándose en la hipótesis de que el gen de la enfermedad debe poder complementar la deficiencia enzimática en una línea celular del paciente, se realizó 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
- 15 enzimática se 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 donadoras y receptoras,
- 20 como a nivel funcional sometiendo a prueba las actividades sulfatasa. El análisis de 130 de tales híbridos dio como resultado 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 sometidos a prueba, lo que
- 25 demuestra que un único gen está implicado en la MSD. Las mutaciones encontradas eran de diferentes tipos, dando como resultado 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, tal como se espera para una enfermedad recesiva. La mayoría de las mutaciones de aminoácido afectan a codones correspondientes a aminoácidos que están altamente conservados durante la evolución, lo que sugiere que estas mutaciones también producen una pérdida de función.
- 30 No pudieron establecerse correlaciones entre el tipo de mutación y la gravedad del fenotipo, lo que indica que este último se debe a factores no relacionados. De manera inesperada 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.

#### 35 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 un
- 40 restablecimiento constante, aunque parcial, de todas las actividades sometidas a prueba, en comparación con las líneas celulares de pacientes control transfectadas con vectores vacíos. Como 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 tal como se somete a prueba mediante un análisis con proteína marcadora (GFP). En los mismos experimentos, se usaron vectores que contenían ADNc de FGE (SUMF1) que
- 45 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.

Tal como se menciona en otra parte en el presente documento, Schmidt *et al.* descubrieron en primer lugar que las sulfatasas experimentan una modificación tras la traducción 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 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. Se proporciona evidencia de que el homólogo de *Drosophila* del gen de FGE humana (SUMF1) puede activar sulfatasas humanas sobreexpresadas, 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 para sulfatasas, lo que indica que este microorganismo no requiere la función sulfatasa y sugiere la presencia de una influencia recíproca en la evolución de los genes de FGE (SUMF1) y sulfatasas.

Resulta interesante que hay dos genes homólogos, FGE (SUMF1) y FGE2 (SUMF2), en los genomas de todos los vertebrados analizados, incluyendo los seres humanos. Como resulta evidente a partir del árbol filogenético, el gen de FGE2 (SUMF2) parece haber evolucionado independientemente del gen de FGE (SUMF1). En los ensayos, el gen de FGE2 (SUMF2) también puede 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 sulfatasa residual encontrada en pacientes con MSD y sugiere que una deficiencia de sulfatasa completa sería letal. En este momento no se puede descartar la posibilidad de que el gen de FGE2 (SUMF2) tenga una función adicional, todavía desconocida.

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

Se observó un fuerte aumento, de hasta 50 veces, de las actividades sulfatasa en células que sobreexpresan ADNc de FGE (SUMF1) junto con o bien ADNc de ARSA, ARSC, o bien ARSE, en comparación con células que sobreexpresan sulfatasas individuales solas. En todas las líneas celulares se encontró un efecto sinérgico significativo, lo que indica que FGE (SUMF1) es un factor limitante para la actividad 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. Compatible con estas observaciones, se encontró 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 suministradas 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 notificado 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:

##### *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 epifisario (placa de crecimiento) debido a la deficiencia de la sulfatasa lisosomal, GALNS. El objetivo primario de estudios de investigación *in vivo* fue determinar si GALNS activada por FGE administrada por vía intravenosa (i.v.) podía 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 del esqueleto, 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 demuestran la actividad enzimática de GALNS funcional dirigida hacia la eliminación de GAG acumulados en los tejidos afectados (estudio farmacodinámico).

Los resultados de estos estudios demuestran 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 lisosomal en todos los tejidos blandos examinados. En resumen, los experimentos demuestran satisfactoriamente el suministro de GALNS a los condrocitos de la placa de crecimiento y demuestran la actividad bioquímica en lo que se refiere al aclaramiento de GAG en múltiples tejidos.

#### Estudio de biodistribución

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 el suministro de GALNS a los condrocitos de la placa de crecimiento ósea en todos los cortes de placa de crecimiento observados en comparación con los controles con vehículo.

#### Estudio farmacodinámico en ratones MKI

Se administraron a ratones MKI o de tipo natural 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 con vehículo (sin inyección para los ratones de tipo natural). Se sacrificaron todos los animales para realizar el examen histológico a las 20 semanas de edad y se tiñeron usando los siguientes métodos: hematoxilina y eosina para determinar la morfología celular, azul alciano para la detección de GAG.

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 sólo en los ratones a los que se inyectó GALNS. Aunque la placa de crecimiento en los ratones MKI funcionaba normalmente tal 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 en las placas de crecimiento no resultaron afectados por el suministro 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 a los que se inyectó 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 en ratones. Estas observaciones particulares reflejan probablemente el hecho de que, en ratones *knock-in*, están presentes altos niveles de GALNS mutante (en contraposición a los ratones *knock-out* 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 del suministro de GALNS activada por FGE a múltiples tejidos y de la actividad enzimática *in vivo*.

#### Descripción detallada de los dibujos

Figura 1: Espectros de masas MALDI-TOF de P23 tras la incubación en ausencia (A) o presencia (B) de un extracto soluble de microsomas de testículos bovinos. Se incubaron 6 pmol de P23 en condiciones normales durante 10 min. a 37°C en ausencia o presencia de 1 µl de extracto microsomal. Se prepararon las muestras para espectrometría de masas MALDI-TOF tal como se describe en Procedimientos experimentales. Se indican las masas monoisotópicas MH<sup>+</sup> de P23 (2526,28) y su derivado de FGly (2508,29).

Figura 2: Árbol filogenético derivado de una alineación de FGE humana y 21 proteínas de la semilla DUF323 en PFAM. Los números de 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.

Figura 3: Organización del locus del gen de FGE humana y murina. Los exones se muestran a escala como cuadros oscuros (locus humano) y cuadros 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.

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

Figura 5: Gráfico de barras que representa 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 metotrexato (MTX). Se volvió a alimentar a las células 24 horas más tarde con medios que contenían MTX 1,0  $\mu$ M. Se recogió el medio y se recogieron las células 24, 48, y 72 horas tras 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.

Figura 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. 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 metotrexato (MTX). Se volvió a alimentar a las células 24 horas más tarde con medios que contenían MTX 1,0  $\mu$ M. Se recogió el medio y se recogieron las células 24, 48, y 72 horas tras 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.

Figura 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 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 metotrexato (MTX). Se volvió a alimentar a las células 24 horas más tarde con medios que contenían MTX 1,0  $\mu$ M. Se recogió el medio y se recogieron las células 24, 48, y 72 horas tras 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.

Figura 8: Gráfico que representa actividad iduronato 2-sulfatasa en células 30C6 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 metotrexato (MTX). Se volvió a alimentar a las células 24 horas más tarde con medios que contenían MTX 1,0  $\mu$ M. 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.

Figura 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.

#### Lista de secuencias

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

<120> DIAGNÓSTICO Y TRATAMIENTO DE DEFICIENCIA MÚLTIPLE DE SULFATASAS Y OTRAS DEFICIENCIAS DE SULFATASAS

<130> 0403WO

<150> US 60/447.747

<151> 11-02-2003

<160> 95

<170> PatentIn versión 3.2

<210> 1

<211> 1180

5 <212> ADN

<213> *Homo sapiens*

<220>

10 <221> CDS

<222> (20)..(1141)

15 <223> ADNc de FGE

<400> 1

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

20

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

<210> 2

5 <211> 374

<212> PRT

<213> *Homo sapiens*

10

<400> 2



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

# ES 2 380 147 T3

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

<210> 3

<211> 1122

5 <212> ADN

<213> *Homo sapiens*

<400> 3

10  
 atggctgcgc ccgcactagg gctgggtgtgt ggacgttgcc ctgagctggg tctcgtcctc 60  
 ttgctgctgc tgctctcgct gctgtgtgga ggggcaggga gccaggaggc cgggaccggt 120  
 gcgggcgcgg ggtcccttgc gggttcttgc ggctgcggca cgcgccagcg gcctggcgcc 180  
 catggcagtt cggcagccgc tcaccgatac tcgcgggagg ctaacgctcc gggccccgta 240  
 cccggagagc ggcaactcgc gcactcaaag atgggtcccca tccctgctgg agtatttaca 300  
 atgggcacag atgatcctca gataaagcag gatggggaag cacctgcgag gagagttact 360  
 attgatgcct ttacatgga tgcttatgaa gtcagtaata ctgaatttga gaagtttgtg 420  
 aactcaactg gctatttgac agaggctgag aagtttggcg actcctttgt ctttgaaggc 480  
 atgttgagtg agcaagtga gaccaatatt caacaggcag ttgcagctgc tccctgggtg 540  
 ttacctgtga aaggcgctaa ctggagacac ccagaagggc ctgactctac tattctgcac 600  
 aggcgggac atccagttct ccatgtgtcc tggaatgatg cggttgccca ctgcacttgg 660  
 gcagggaagc ggctgcccac ggaagctgag tgggaataca gctgtcgagg aggcctgcat 720  
 aatagacttt tcccctgggg caacaaactg cagcccaaag gccagcatta tgccaacatt 780  
 tggcagggcg agtttccggt gaccaacact ggtgaggatg gcttccaagg aactgcgctt 840  
 gttgatgcct tccctcccaa tggttatggc ttatacaaca tagtggggaa cgcaggggaa 900  
 tggacttcag actggtggac tgttcacat tctgttgaag aaacgcttaa cccaaaaggt 960  
 ccccttctg ggaaagaccg agtgaagaaa ggtggatcct acatgtgcca taggtcttat 1020  
 tgttacaggt atcgctgtgc tgctcggagc cagaacacac ctgatagctc tgcttcgaat 1080  
 ctgggattcc gctgtgcagc cgaccgcctg cccaccatgg ac 1122

<210> 4

15 <211> 2130

<212> ADN

<213> *Homo sapiens*

20

<400> 4

acatggcccg cgggacaaca tggctgcgcc cgcactaggg ctggtgtgtg gacgttgccc	60
tgagctgggt ctcgteectt tgetgctgct gctctcgctg ctgtgtggag cggcagggag	120
ccaggaggcc gggaccggtg cgggcgcggg gtcccttgcg ggttcttgcg gctgcggcac	180
gccccagcgg cctggcgccc atggcagttc ggcagccgct caccgatact cgcgggaggg	240
taacgctccg ggccccgtac ccggagagcg gcaactcgcg cactcaaaga tggccccat	300
ccctgctgga gtatttacia tgggcacaga tgatcctcag ataaagcagg atggggaagc	360
acctgcgagg agagttacta ttgatgccct ttacatggat gcctatgaag tcagtaatac	420
tgaatttgag aagtttgta actcaactgg ctatttgaca gaggctgaga agtttggcga	480
ctcctttgtc tttgaaggca tgttgagtga gcaagtgaag accaatattc aacaggcagt	540
tgcagctgct ccctggtggt tacctgtgaa aggcgctaac tggagacacc cagaagggcc	600
tgactctact attctgcaca ggcgggatca tccagttctc catgtgtcct ggaatgatgc	660
ggttgectac tgcacttggg cagggaagcg gctgcccacg gaagctgagt ggggaatacag	720
ctgtcgagga ggctgcata atagactttt cccctggggc aacaaactgc agcccaaagg	780

ccagcattat	gccaacattt	ggcagggcga	ttttccggtg	accaacactg	gtgaggatgg	840
cttccaagga	actgcgcctg	ttgatgcctt	ccctcccaat	ggttatggct	tataacaacat	900
agtggggaac	gcatgggaat	ggacttcaga	ctggtggact	gttcatcatt	ctgttgaaga	960
aacgcttaac	ccaaaaggte	ccccttctgg	gaaagaccga	gtgaagaaag	gtggatccta	1020
catgtgccat	aggtcttatt	gttacaggta	tcgctgtgct	gctcggagcc	agaacacacc	1080
tgatagctct	gcttcgaatc	tgggatcccg	ctgtgcagcc	gaccgcctgc	ccaccatgga	1140
ctgacaacca	agggtagtct	tccccagtc	aaggagcagt	cgtgtctgac	ctacattggg	1200
ctttcctcag	aactttgaac	gatcccatgc	aaagaattcc	caccctgagg	tgggttacat	1260
acctgcccac	tggccaaagg	aaccgccttg	tgagacccaa	ttgctgacct	gggtcagtgc	1320
atgtgcttta	tgggtgtggtg	catctttgga	gatcatcacc	atattttact	tttgagagtc	1380
tttaaagagg	aaggggagtg	gagggaaacc	tgagctaggc	ttcaggaggc	ccgcaccta	1440
cgcaggctct	gccacagggg	ttagacccca	ggtccgacgc	ttgaccttcc	tgggcctcaa	1500
gtgccctccc	ctatcaaagt	aaggaatgga	cagcatgacc	tctgggtgtc	tctccaactc	1560
accagttcta	aaaaggggat	cagattctat	tgtgacttca	tagaatttat	gatagattat	1620
tttttagcta	ttttttccat	gtgtgaacct	tgagtgatac	taatcatgta	aagtaagagt	1680
tctcttatgt	attatgttcg	gaagaggggt	gtggtgactc	ctttatatte	gtactgcact	1740
ttgtttttcc	aaggaaatca	gtgtctttta	cgttggttatg	atgaatccca	catggggccg	1800
gtgatgggat	gctgaagttc	agccgttgaa	cacataggaa	tgtctgtggg	gtgactctac	1860
tgtgctttat	cttttaacat	taagtgcctt	tgggttcagag	gggcagtcac	aagctctggt	1920
tccccctctc	cccaaagcct	tcagcgaacg	tgaaatgtgc	gctaaacggg	gaaacctggt	1980
taattctaga	tatagggaac	aaggaacgag	gaccttgaat	gagctatatt	cagggtatcc	2040
ggtattttgt	aataggggat	aggaaacctt	gttggtctgtg	gaatatccga	tgctttgaat	2100
catgcactgt	gttgaataaa	cgtatctgct				2130

<210> 5

5 <211> 374

<212> PRT

<213> *Homo sapiens*

10

<400> 5

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Arg Leu Pro Thr Met Asp  
370

<210> 6

<211> 2297

5 <212> ADN

<213> *Homo sapiens*

<400> 6

10



cggtctgtgtt ggcagctctt catgggttcc cgacgaggag gtctctgtgg ctgcggcggc	60
tgctaactgc gccacctgct gcagcctgtc cccgccgctc tgaagcggcc gcgtcgaagc	120
cgaaatgccg ccaccccgga ccggccgagg ccttctctgg ctgggtctgg ttctgagetc	180
cgtctgcgtc gccctcggat ccgaaacgca ggccaactcg accacagatg ctctgaacgt	240
tcttctcctc atcgtggatg acctgcgccc ctccctgggc tgttatgggg ataagctggt	300
gagggtcccca aatattgacc aactggcatc ccacagcctc ctcttccaga atgcctttgc	360
gcagcaagca gtgtgcgccc cgagccgegt ttctttcctc actggcagga gacctgacac	420
caccgcgctg tacgaattca actcctactg gaggggtgac gctggaaaact tctccaccat	480
ccccagttac ttcaaggaga atggctatgt gaccatgtcg gtgggaaaag tctttcacc	540
tgggatattc tctaaccata ccgatgatc tccgtatagc tggctctttc caccctatca	600
tccttctctc gagaagtatg aaaacactaa gacatgtcga gggccagatg gagaactcca	660
tgccaacctg ctttgccctg tggatgtgct ggatgttccc gagggcacct tgcctgacaa	720
acagagcact gagcaagcca tacagtgtgt ggaaaagatg aaaacgtcag ccagtccttt	780
cttccctggc gttgggtatc ataagccaca catccccttc agatacccca aggaatttca	840
gaagtgtgat cccttggaaga acatcacccct ggccccgat cccgaggtcc ctgatggcct	900
acccctgtg gcctacaacc cctggatgga catcaggcaa cgggaagacg tccaagcctt	960
aaacatcagt gtgccgtatg gtccaattcc tgtggacttt cagcggaaaa tccgccagag	1020
ctactttgcc tctgtgtcat atttggtatc acaggtcggc cgctcttga gtgctttgga	1080
cgatcttcag ctggccaaca gcaccatcat tgcatttacc tcggatcatg ggtgggctct	1140
aggtgaacat ggagaatggg ccaaatacag caattttgat gttgctaccc atgttcccc	1200
gatattctat gtctctggaa ggacggcttc acttccggag gcaggcgaga agcttttccc	1260
ttacctcgac ctttttgatt ccgcctcaca gttgatggag ccaggcaggc aatccatgga	1320
ccttgtggaa cttgtgtctc tttttccac gctggctgga cttgcaggac tgcaggttcc	1380
acctcgctgc ccggttcctt catttcacgt tgagctgtgc agagaaggca agaaccttct	1440
gaagcatttt cgattccgtg acttggaaga ggatccgtac ctccctggta atccccgtga	1500
actgattgcc tatagccagt atccccggcc ttcagacatc cctcagtgga attctgacaa	1560
gccgagttta aaagatataa agatcatggg ctattccata cgcaccatag actataggta	1620
tactgtgtgg gttggcttca atcctgatga atttctagct aacttttctg acatccatgc	1680
aggggaactg tattttgtgg attctgaccc attgcaggat cacaatatgt ataattgatc	1740
ccaaggtgga gatcttttcc agttgttgat gccttgagtt ttgccaacca tggatggcaa	1800
atgtgatgtg ctcccttcca gctgggtgaga ggaggagtta gagctggctg ttttgtgatt	1860
accataata ttggaagcag cctgagggct agttaatcca aacatgcac aacaatttgg	1920
cctgagaata tgtaacagcc aaaccttttc gtttagtctt tattaaaatt tataattggt	1980
aattggacca gttttttttt taatttccct ctttttaaaa cagttacggc ttatttactg	2040
aataaataca aagcaaacaa actcaagtta tgtcatacct ttggatacga agaccataca	2100
taataaccaa acataacatt atacacaaag aatactttca ttatttgtgg aatttagtgc	2160
atttcaaaaa gtaatcatat atcaaactag gcaccacact aagttcctga ttattttggt	2220
tataatttaa taatatatct tatgagccct atatatcaa aatattatgt taacatgtaa	2280
tccatgttcc tttttcc	2297

<210> 7

<211> 550

5 <212> PRT

<213> *Homo sapiens*

<400> 7

10

Met Pro Pro Pro Arg Thr Gly Arg Gly Leu Leu Trp Leu Gly Leu Val  
1 5 10 15

Leu Ser Ser Val Cys Val Ala Leu Gly Ser Glu Thr Gln Ala Asn Ser  
20 25 30

Thr Thr Asp Ala Leu Asn Val Leu Leu Ile Ile Val Asp Asp Leu Arg  
35 40 45

Pro Ser Leu Gly Cys Tyr Gly Asp Lys Leu Val Arg Ser Pro Asn Ile  
50 55 60

Asp Gln Leu Ala Ser His Ser Leu Leu Phe Gln Asn Ala Phe Ala Gln  
65 70 75 80

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

Pro Asp Thr Thr Arg Leu Tyr Asp Phe Asn Ser Tyr Trp Arg Val His  
100 105 110

Ala Gly Asn Phe Ser Thr Ile Pro Gln Tyr Phe Lys Glu Asn Gly Tyr  
115 120 125

Val Thr Met Ser Val Gly Lys Val Phe His Pro Gly Ile Ser Ser Asn  
130 135 140

His Thr Asp Asp Ser Pro Tyr Ser Trp Ser Phe Pro Pro Tyr His Pro  
145 150 155 160

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

Glu Leu His Ala Asn Leu Leu Cys Pro Val Asp Val Leu Asp Val Pro  
180 185 190

Glu Gly Thr Leu Pro Asp Lys Gln Ser Thr Glu Gln Ala Ile Gln Leu  
195 200 205

Leu Glu Lys Met Lys Thr Ser Ala Ser Pro Phe Phe Leu Ala Val Gly  
210 215 220

Tyr His Lys Pro His Ile Pro Phe Arg Tyr Pro Lys Glu Phe Gln Lys  
225 230 235 240

Leu Tyr Pro Leu Glu Asn Ile Thr Leu Ala Pro Asp Pro Glu Val Pro  
245 250 255

Asp Gly Leu Pro Pro Val Ala Tyr Asn Pro Trp Met Asp Ile Arg Gln  
260 265 270

Arg Glu Asp Val Gln Ala Leu Asn Ile Ser Val Pro Tyr Gly Pro Ile  
275 280 285

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

## ES 2 380 147 T3

<210> 8

<211> 2657

5 <212> ADN

<213> *Homo sapiens*

<400> 8

10 **gaattccggg ccatgagctg ccccgtagcc gcctgctgcg cgctgctgct agtcctgggg 60**

ctctgccggg	cgcgccccg	gaacgcactg	ctgctcctcg	cggatgacgg	aggctttgag	120
agtggcgcgt	acaacaacag	cgccatcgcc	accccgaccc	tggacgcctt	ggcccgcgcg	180
agcctcctct	ttcgcaatgc	cttcacctcg	gtcagcagct	gctctcccag	ccgcgccagc	240
ctcctcactg	gcctgcccc	gcatacagaat	gggatgtacg	ggctgcacca	ggacgtgcac	300
cacttcaact	ccttcgacaa	gggtgcggagc	ctgccgctgc	tgctcagcca	agctgggtgtg	360
cgcacaggca	tcatacggaa	gaagcacgtg	gggccggaga	ccgtgtaccc	gtttgacttt	420
gcgtacacgg	aggagaatgg	ctccgtcttc	caggtggggc	ggaacatcac	tagaattaag	480
ctgctcgtcc	ggaaattcct	gcagactcag	gatgaccggc	ctttcttcct	ctacgtcgcc	540
ttccacgacc	cccaccgtg	tgggcactcc	cagccccagt	acggaacctt	ctgtgagaag	600
tttggaacg	gagagagcgg	catgggtcgt	atcccagact	ggacccccca	ggcctacgac	660
ccactggacg	tgctggtgcc	ttacttcgtc	cccaacaccc	cggcagcccc	agccgacctg	720
gccgctcagt	acaccaccgt	cggccgcatg	gaccaaggag	ttggactggg	gctccaggag	780
ctgcgtgacg	ccggtgtcct	gaacgacaca	ctggtgatct	tcacgtccga	caacgggatc	840
cccttcccca	gcggcaggac	caacctgtac	tggccgggca	ctgctgaacc	cttactgggtg	900
tcataccccc	agcacccaaa	acgctggggc	caagtacagc	aggcctacgt	gagcctccta	960
gacctcaacg	ccaccatctt	ggattgggtt	tcgatcccg	acccagcta	cgccatcttt	1020
ggctcgaaga	ccatccacct	cactggccgg	tcctcctctg	cggcgctgga	ggccgagccc	1080
ctctgggcca	ccgtcttttg	cagccagagc	caccacgagg	tcaccatgtc	ctaccccatg	1140
cgtcccgctg	agcaccggca	cttcgccttc	gtgcacaacc	tcaacttcaa	gatgcccttt	1200
cccatcgacc	aggactttta	cgtctcacc	accttcagg	acctcctgaa	ccgcaccaca	1260
gctggtcagc	ccacgggctg	gtacaaggac	ctccgtcatt	actactaccg	ggcgcgctgg	1320
gagctctacg	accggagccg	ggacccccac	gagaccaga	acctggccac	cgacccgcgc	1380
tttgctcagc	ttctggagat	gcttcgggac	cagctggcca	agtggcagtg	ggagaccac	1440
gacccctggg	tgtgcgcccc	cgacggcgtc	ctggaggaga	agctctctcc	ccagtgccag	1500
cccctccaca	atgagctgtg	accatcccag	gaggcctgtg	cacacatccc	aggcatgtcc	1560
cagacacatc	ccacacgtgt	ccgtgtggcc	ggccagcctg	gggagtagtg	gcaacagccc	1620
ttccgtccac	actcccatcc	aaggagggtt	cttccttcct	gtggggtcac	tcttgccatt	1680
gcctggaggg	ggaccagagc	atgtgaccag	agcatgtgcc	cagccccctc	accaccaggg	1740
gcactgccgt	catggcaggg	gacacagttg	tccttgtgtc	tgaacctagt	cccagcacgg	1800
gaattctaga	catacgtggg	ctgcggacag	ggcagcgccc	ccagcccatg	acaagggagt	1860
cttggttttct	ggcttggttt	ggggacctgc	aaatgggagg	cctgaggccc	tcttcaggct	1920
ttggcagcca	cagatacttc	tgaacccttc	acagagagca	ggcaggggct	tcggtgccgc	1980
gtgggcagta	cgcagggtccc	accgacactc	acctgggagc	acggcgcttg	gctcttacca	2040
gcgtctggcc	tagaggaagc	ctttgagcga	cctttgggca	ggtttctgct	tcttctgttt	2100
tgcccatggg	caagtccctg	ttccccaggc	aggtttcagc	tgattggcag	caggctccct	2160
gagtgatgag	cttgaacctg	tgggtgtttct	gggcagaagc	ttatcttttt	tgagagtgtc	2220
cgaagatgaa	ggcatggcga	tgcccgctct	ctggcttggg	ttaattcttc	ggtgacactg	2280
gcattgctgg	gtggtgatgc	ccgtcctctg	gcttgggtta	attcttcggg	gacactggcg	2340
ttgctgggtg	gcaatgcccc	tcctctgggt	tgggttaatt	cttcgggtgac	actggcggtg	2400

# ES 2 380 147 T3

```

ctgggtggcg atgcccgccc tctggccttg gttaattctt ggatgacgtc ggcgttgctg 2460
ggagaatgtg ccgttctctgc cctgcctcca cccacctcgg gagcagaagc ccggcctgga 2520
caccctcggg cctggacacc cctcgaagga gagggcgctt ccttgagtag gtgggctccc 2580
cttgcccttc cctccctatc actccatact ggggtgggct ggaggaggcc acaggccagc 2640
tattgtaaaa gctttttt 2657

```

<210> 9

5 <211> 502

<212> PRT

<213> *Homo sapiens*

10 <400> 9

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



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

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

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

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

Trp Gly Gln Val Ser Glu Ala Tyr Val Ser Leu Leu Asp Leu Thr Pro  
305 310 315 320

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

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

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

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

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

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

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

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

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

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

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

Pro Leu His Asn Glu Leu  
500

<210> 10

<211> 1014

5 <212> ADN

<213> *Homo sapiens*

<400> 10

10  
 cgtgcctgta atcccagcag ctactcactc aggaggctga ggcaggagaa tctcttgaac 60  
 ccggaaggca gaggttgcag tgagccaaga tcgcgccact gaactccagc ctgggtgaca 120  
 gagtgagact gtctcagaac agcaacaaca aaatgcccgc tgctgctggg tccagaagag 180  
 cttgaataac tgcattgtct tttctctcaat tttcatttcc cagaactggg cacctccggg 240  
 ctgtgaaaag ttagggaagt gtctgacacc tccagaatcc attcccaaga agtgcctctg 300  
 gtcccactag cacctgcgca gactcaggcc aggcctagaa tctccagttg gccctgcaag 360  
 tgcctggagg aaggatggct ctggcctcgg tcttccccca accctgcca agccagacag 420  
 acagcacctg cagacgcagg gggactgcac aattccacct gccaggacc tgaccctggc 480  
 gtgtgcttgg cctcctcct cgcccacggc gcctcagatt tcaggacct cctcctcgcc 540  
 cagggcgcct cagacctcag gacctgccc tctcagcct ttgtgaacct caaatatctg 600  
 agaccagtct cagtttatct tgccaaggct aaggatgcac ctgtgacagc ctccaggagg 660  
 cctgacaaca ggtgcccag gtggctgggg atacagtctg cctttataca tcttagggag 720  
 acacaagatc agtatgtgta tggcgctacat tgggttcagtc agccttccac tgaatacacg 780  
 attgagtctg gccagtgaa tccgcatttt tatgtaaaca gtaaggggaa ggggcaatca 840  
 tataagcgtt tgtctcaggg gagccccaga gggatgactt ccagttccgt ctgtcctttg 900  
 tccacaagga atttccctgg gcgctaatta tgaggagggc gtgtagcttc ttatcattgt 960  
 agctatgtta tttagaaata aaacgggagg cagggtttgcc taattcccag gttg 1014

15 <210> 11

<211> 522

<212> PRT

20 <213> *Homo sapiens*

<400> 11

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

# ES 2 380 147 T3

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

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

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

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

<210> 12

5 <211> 2379

<212> ADN

<213> *Homo sapiens*

10 <400> 12

ggaattccgg	tcggcctctc	gcccttcage	tacctgtgcg	tccctccgtc	ccgtcccgtc	60
ccggggtcac	cccggagcct	gtccgctatg	cggctcctgc	ctctagcccc	aggtcggctc	120
cggcggggca	gccccgcga	cctgccctcc	tgcagcccag	cgctgtact	gctgggtgctg	180
ggcggtgcc	tgggggtctt	cggggtgget	gcggaaccc	ggaggcccaa	cgtgggtgctg	240
ctcctcaagg	acgaccagga	cgaagtgtc	ggcggcatga	caccactaaa	gaaaaccaa	300
gctctcatcg	gagagatggg	gatgactttt	tccagtgtct	atgtgccaa	tgctctctgc	360
tgccccagca	gagccagtat	cctgacagga	aagtaccac	ataatcatca	cgttgtgaac	420
aacactctgg	aggggaactg	cagtagtaag	tcctggcaga	agatccaaga	accaaatact	480
ttcccagcaa	ttctcagatc	aatgtgtggt	tatcagacct	tttttgcagg	gaaatattta	540
aatgagtacg	gagccccaga	tgcagggtga	ctagaacacg	ttcctctggg	ttggagttac	600
tggtatgcct	tggaaaagaa	ttctaagtat	tataattaca	ccctgtctat	caatgggaag	660
gcacggaagc	atggtgaaaa	ctatagtgtg	gactacctga	cagatgtttt	ggctaattgtc	720
tccttggaact	ttctggacta	caagtccaac	tttgagccct	tcttcatgat	gatcgccact	780
ccagcgctc	attcgcttg	gacagctgca	cctcagtacc	agaaggcttt	ccagaatgtc	840
tttgaccaa	gaaacaagaa	cttcaacatc	catggaacga	acaagcactg	gttaattagg	900
caagccaaga	ctccaatgac	taattcttca	atacagtttt	tagataatgc	atttaggaaa	960
aggtggcaaa	ctctcctctc	agttgatgac	cttgtggaga	aactgggtcaa	gaggctggag	1020
ttcactgggg	agctcaacaa	cacttacatc	ttctatacct	cagacaatgg	ctatcacaca	1080
ggacagtttt	ccttgccaat	agacaagaga	cagctgtatg	agtttgatat	caaagttcca	1140
ctgttggttc	gaggacctgg	gatcaaacca	aatcagacaa	gcaagatgct	ggttgccaac	1200
attgacttgg	gtcctactat	tttggacatt	gctggctacg	acctaaataa	gacacagatg	1260
gatgggatgt	ccttattgcc	catttttgaga	ggtgccagta	acttgacctg	gcgatcagat	1320
gtcctgggtg	aataccaagg	agaaggccgt	aacgtcactg	acccaacatg	cccttccttg	1380
agtcctggcg	tatctcaatg	cttcccagac	tgtgtatgtg	aagatgctta	taacaatacc	1440
tatgcctgtg	tgaggacaat	gtcagcattg	tgggaatttgc	agtattgcga	gtttgatgac	1500
caggaggtgt	ttgtagaagt	ctataatctg	actgcagacc	cagaccagat	cactaacatt	1560
gctaaaacca	tagaccacga	gcttttagga	aagatgaact	atcggttaat	gatgttacag	1620
tcctgttctg	ggccaacctg	tcgcactcca	ggggtttttg	accccgata	caggtttgac	1680
ccccgtctca	tgttcagcaa	tcgcggcagt	gtcaggactc	gaagattttc	caaacatctt	1740

```

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

```

<210> 13

5 <211> 552

<212> PRT

<213> *Homo sapiens*

10 <400> 13

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



ES 2 380 147 T3

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

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

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

Arg Arg Phe Ser Lys His Leu Leu  
545 550

<210> 14

5 <211> 2022

<212> ADN

<213> *Homo sapiens*

10 <400> 14

ccggtaccgg	ctcctcctgg	gtccctctta	gcgccttccc	cccggccccga	ctgcctggtc	60
agcgccaagt	gaattacgcc	cccgaacctg	agccccggacc	gctaggcgag	gaggatcaga	120
tctccgctcg	agaatctgaa	ggtgccctgg	tcctggagga	gttccgtccc	agccctgcgg	180
tctcccggta	ctgctcgccc	cggccctctg	gagcttcagg	aggcggccgt	cagggtcggg	240
gagtatttgg	gtccgggggc	tcagggaagg	gcggcgccctg	ggtctgcggt	atcggaaga	300
gcctgctgga	gccaaagtagc	cctccctctc	ttgggacaga	cccctcggtc	ccatgtccat	360
gggggcaccg	cggtcctccc	tcctggccct	ggctgctggc	ctggccggtg	cccgtccgcc	420
caacatcgctg	ctgatctttg	ccgacgacct	cggctatggg	gacctgggct	gctatgggca	480
ccccagctct	accactccca	acctggacca	gctggcgggc	ggagggctgc	ggttcacaga	540
cttctacgtg	cctgtgtctc	tgtgcacacc	ctctaggggc	gccctcctga	ccggccggct	600
cccggttcgg	atgggcatgt	accctggcgt	cctggtgccc	agctcccggg	ggggcctgcc	660
cctggaggag	gtgaccgtgg	ccgaagtcct	ggctgccccga	ggctacctca	caggaatggc	720
cggcaagtgg	caccttgggg	tggggcctga	gggggccttc	ctgccccccc	atcagggtct	780
ccatcgattt	ctaggcatcc	cgtactccca	cgaccagggc	ccctgccaga	acctgacctg	840
cttcccgcgg	gccactcctt	gcgacgggtg	ctgtgaccag	ggcctgggtc	ccatcccact	900
gttggccaac	ctgtccgtgg	aggcgagacc	cccctggctg	cccggactag	aggcccgtca	960
catggctttc	gcccatgacc	tcattggccga	cgcccagcgc	caggatcgcc	ccttcttctt	1020
gtactatgcc	tctcaccaca	cccactaccc	tcagttcagt	gggcagagct	ttgcagagcg	1080
ttcaggccgc	gggccatttg	gggactccct	gatggagctg	gatgcagctg	tggggaccct	1140
gatgacagcc	ataggggacc	tggggctgct	tgaagagacg	ctggtcatct	tcactgcaga	1200
caatggacct	gagaccatgc	gtatgtcccg	aggcggctgc	tccggtctct	tgcggtgtgg	1260
aaagggaaacg	acctacgagg	gcggtgtccg	agagcctgcc	ttggccttct	ggccaggtea	1320
tatcgctccc	ggcgtgacct	acgagctggc	cagctccctg	gacctgctgc	ctaccctggc	1380
agccctggct	ggggccccac	tgcccaatgt	caccttggat	ggctttgacc	tcagccccct	1440
gctgctgggc	acaggcaaga	gccctcggca	gtctctcttc	ttctaccctg	cctaccacaga	1500
cgaggtcctg	ggggtttttg	ctgtgcggac	tggaaagtac	aaggctcact	tcttcaccca	1560
gggctctgcc	cacagtgata	ccactgcaga	ccctgcctgc	cacgcctcca	gctctctgac	1620
tgctcatgag	cccccgctgc	tctatgacct	gtccaaggac	cctggtgaga	actacaacct	1680
gctgggggggt	gtggccgggg	ccaccccaga	ggtgctgcaa	gccctgaaac	agcttcagct	1740

# ES 2 380 147 T3

```

gctcaaggcc cagttagacg cagctgtgac cttcggcccc agccagggtgg cccggggcgga 1800
ggacccccgc ctgcagatct gctgtcatcc tggctgcacc ccccgcccag cttgctgcca 1860
ttgcccagat ccccatgcct gagggccctt cggttggcct gggcatgtga tggctcctca 1920
ctgggagcct gtgggggagg ctcagggtgc tggagggggt ttgtgectga taacgtaata 1980
acaccagtgg agacttgac atctgaaaaa aaaaaaaaaa aa 2022

```

<210> 15

5 <211> 507

<212> PRT

<213> *Homo sapiens*

10 <400> 15

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

## ES 2 380 147 T3

<210> 16

<211> 2228

5 <212> ADN

<213> *Homo sapiens*

<400> 16

10    **acaaggatgg gtccgcgcgg cgcggcgagc ttgccccgag gcccgggacc tcggcggctg        60**  
      **ctcctccccg tcgtcctccc gctgctgctg ctgctgttgt tggcgcgcgc gggctcgggc        120**

```

gccggggcca gccggccgcc ccacctgggc ttcttgctgg cagacgacct aggctggaac 180
gacgtcggct tccacggctc ccgcatccgc acgccgcacc tggacgcgct ggccggccggc 240
ggggtgctcc tggacaacta ctacacgcag ccgctgtgca cgccgtcgcg gagccagctg 300
ctcactggcc gctaccagat ccgtacaggt ttacagcacc aaataatctg gccctgtcag 360
cccagctgtg ttcctctgga tgaaaaactc ctgccccagc tcctaaaaga agcaggttat 420
actaccata tggtcggaaa atggcacctg ggaatgtacc ggaaagaatg ccttccaacc 480
cgccgaggat ttgataccta ctttggatat ctctgggta gtgaagatta ttattcccat 540
gaacgctgta cattaattga cgctctgaat gtcacacgat gtgctcttga ttttcgagat 600
ggcgaagaag ttgcaacagg atataaaaaat atgtattcaa caaacatatt caccaaaagg 660
gctatagccc tcataactaa ccatccacca gagaagcctc tgtttctcta ccttgctctc 720
cagtctgtgc atgagccccct tcaggctccct gaggaatact tgaagccata tgactttatc 780
caagacaaga acaggcatca ctatgcagga atgggtgtccc ttatggatga agcagtagga 840
aatgtcactg cagctttaa aagcagtggg ctctggaaca acacggtgtt catcttttct 900
acagataacg gagggcagac tttggcaggg ggttaataact ggccccctcg aggaagaaaa 960
tggagcctgt ggggaaggagg cgtccgaggg gtgggctttg tggcaagccc cttgctgaag 1020
cagaagggcg tgaagaaccg ggagctcacc cacatctctg actggctgcc aacactcgtg 1080
aagctggcca ggggacacac caatggcaca aagcctctgg atggcttcga cgtgtggaaa 1140
accatcagtg aaggaagccc atccccaga attgagctgc tgcataatat tgacccaaac 1200
ttcgtggact cttcaccgtg tcccaggaac agcatggctc cagcaaagga tgactcttct 1260
cttcagaat attcagcctt taacacatct gtccatgctg caattagaca tggaaattgg 1320
aaactcctca cgggctaccc aggctgtggg tactgggtcc ctccaccgtc tcaatacaat 1380
gtttctgaga taccctcatc agaccacca accaagaccc tctggctctt tgatattgat 1440
cgggaccctg aagaaagaca tgacctgtcc agagaatatc ctacatcgt cacaaagctc 1500
ctgtcccgcc tacagttcta ccataaacac tcagtccccg tgtacttccc tgcacaggac 1560
ccccgctgtg atcccaaggc cactgggggtg tggggccctt ggatgtagga tttcagggag 1620
gctagaaaac ctttcaattg gaagtggac ctcaggcctt ttctcagac tcttgtctca 1680
tttgttatcc caacctgggt tcacttggcc cttctcttgc tcttaaacca caccgaggtg 1740
tctaatttca acccctaatt catttaagaa gctgataaaa tctgcaacac tctgtctgtt 1800
ggctggagca tgtgtctaga ggtgggggtg gctgggttta tcccccttc ctaagccttg 1860
ggacagctgg gaacttaact tgaaatagga agttctcact gaatcctgga ggctggaaca 1920
gctggctctt ttagactcac aagtcagacg ttcgattccc ctctgccaat agccagtttt 1980
attggagtga atcacatttc ttacgcaa atgaaggagca gacagtgatt aatggttctg 2040
ttggccaagg cttctccctg tcggtgaagg atcatgttca ggcaactcaa gtgaaccacc 2100
cctcttgggt cacccttac tcacttatct catcacagag cataaggccc attttgttgt 2160
tcagggtcaac agcaaatgg cctgcaccat gactgtggct tttaaaataa agaaatgtgt 2220
ttttatcg 2228

```



<210> 17

<211> 533

5 <212> PRT

<213> *Homo sapiens*

<400> 17

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Thr Leu Ala Gly Gly Asn Asn Trp Pro Leu Arg Gly Arg Lys Trp Ser  
305 310 315 320

Leu Trp Glu Gly Gly Val Arg Gly Val Gly Phe Val Ala Ser Pro Leu  
325 330 335

Leu Lys Gln Lys Gly Val Lys Asn Arg Glu Leu Ile His Ile Ser Asp  
340 345 350

Trp Leu Pro Thr Leu Val Lys Leu Ala Arg Gly His Thr Asn Gly Thr  
355 360 365

Lys Pro Leu Asp Gly Phe Asp Val Trp Lys Thr Ile Ser Glu Gly Ser  
370 375 380

Pro Ser Pro Arg Ile Glu Leu Leu His Asn Ile Asp Pro Asn Phe Val  
385 390 395 400

Asp Ser Ser Pro Cys Pro Arg Asn Ser Met Ala Pro Ala Lys Asp Asp  
405 410 415

Ser Ser Leu Pro Glu Tyr Ser Ala Phe Asn Thr Ser Val His Ala Ala  
420 425 430

Ile Arg His Gly Asn Trp Lys Leu Leu Thr Gly Tyr Pro Gly Cys Gly  
435 440 445

Tyr Trp Phe Pro Pro Pro Ser Gln Tyr Asn Val Ser Glu Ile Pro Ser  
450 455 460

Ser Asp Pro Pro Thr Lys Thr Leu Trp Leu Phe Asp Ile Asp Arg Asp  
465 470 475 480

Pro Glu Glu Arg His Asp Leu Ser Arg Glu Tyr Pro His Ile Val Thr  
485 490 495

Lys Leu Leu Ser Arg Leu Gln Phe Tyr His Lys His Ser Val Pro Val  
500 505 510

Tyr Phe Pro Ala Gln Asp Pro Arg Cys Asp Pro Lys Ala Thr Gly Val  
515 520 525

Trp Gly Pro Trp Met  
530

<210> 18

5 <211> 2401

<212> ADN

<213> *Homo sapiens*

# ES 2 380 147 T3

<400> 18

gcctccagca gctgacggga cccagctgta gtgagggtgc agtgattgag taggattggc	60
ctgcttcaaa gcagagggtt ctcattggga tatgcttatt aaactccac tgggtgcagaa	120
accatgaaca gaggatgaac aagtgaagtt gcaatctcct ccatcacagc tcagttcccc	180
aacaacagga tcacaagctg gagatgcctt taaggaagat gaagatccct ttcctcctac	240
tgttctttct gtgggaagcc gagagccacg cagcatcaag gccgaacatc atcctggtga	300
tggctgacga cctcggcatt ggagatcctg ggtgctatgg gaacaaaact atcaggactc	360

ccaatatega	ccggttggcc	agtgggggag	tgaactcac	tcagcacctg	gcagcatcac	420
cgctgtgcac	accaagcagg	gcagccttca	tgactggccg	gtaccctgtc	cgatcaggaa	480
tggcatcttg	gtcccgcact	ggagttttcc	tcttcacagc	ctcttcggga	ggacttccca	540
ccgatgagat	tacctttgct	aagcttctga	aggatcaagg	ttattcaaca	gcactgatag	600
ggaaatggca	ccttgggatg	agctgtcaca	gcaagactga	cttctgtcac	caccttttac	660
atcacggctt	caattatttc	tatgggatct	ctttgaccaa	tctgagagac	tgcaagcccg	720
gagagggcag	tgtcttcacc	acgggcttca	agaggctggt	cttcctcccc	ctgcagatcg	780
tcggggtcac	cctccttacc	cttgcctgcac	tcaattgtct	ggggctactc	cacgtgcctc	840
taggcgtttt	tttcagcctt	ctcttcctag	cagccctaata	cctgaccctt	ttcttgggct	900
tccttcatta	cttcggcccc	ctgaactgct	tcatgatgag	gaactacgag	atcattcagc	960
agcccatgtc	ctatgacaat	ctcaccacga	ggctaacggt	ggaggcggcc	cagttcatac	1020
agcggaacac	tgagactccg	ttcctgcttg	tcttgtccta	cctccacgtg	cacacagccc	1080
tgttctccag	caaagacttt	gctggcaaaa	gtcaacacgg	agtctacggg	gatgctgttg	1140
aggaaatgga	ctggagtgtg	gggcagatct	tgaaccttct	ggatgagctg	agattggcta	1200
atgataccct	catctacttc	acatcggacc	aggagacaca	tgtagaggag	gtgtcttcca	1260
aaggagaaat	tcatggcgga	agtaatggga	tctataaagg	aggaaaagca	aacaactggg	1320
aaggaggtat	ccgggttcca	ggcatccttc	gttggcccag	ggtgatacag	gctggccaga	1380
agattgatga	gccactagc	aacatggaca	tatttcctac	agtagccaag	ctggctggag	1440
ctcccttgcc	tgaggacagg	atcattgatg	gacgtgatct	gatgccctg	cttgaaggaa	1500
aaagccaacg	ctccgatcat	gagtttctct	tccattactg	caacgcctac	ttaaatgctg	1560
tgcgctggca	ccctcagaac	agcacatcca	tctggaaggc	ctttttcttc	acccccaact	1620
tcaaccccg	gggttccaac	ggatgctttg	ccacacacgt	gtgcttctgt	ttcgggagtt	1680
atgtcaccca	tcacgacca	cctttactct	ttgatatttc	caaagatccc	agagagagaa	1740
accactaac	tccagcatcc	gagccccggt	tttatgaaat	cctcaaagtc	atgcaggaag	1800
ctgcggacag	acacaccag	accctgccag	agggtgcccga	tcagttttca	tggaacaact	1860
ttcttttgaa	gccctggctt	cagctgtgct	gtccttccac	cggcctgtct	tgccagtgtg	1920
atagagaaaa	acaggataag	agactgagcc	gctagcagcg	cctggggacc	agacagacgc	1980
atgtggcaaa	gctcaccatc	ttcactacaa	acacgcctga	gagtggcact	ggggaaacat	2040
aactccatct	acaccttggg	tttggactga	ttctccattt	tatcacctga	aggcttgggc	2100
cagagctcaa	cagctactca	actggagggg	tgagggggat	aaggtctgta	gtatacagac	2160
aggaagatgg	taggtttatg	ccttctgtgg	ccagagtctt	ggactcatgg	aaatagaatg	2220
aatagagggg	cattcacaag	gcacaccagt	gcaagcagat	gacaaaaagg	tgcaagaaggc	2280
aatcttaaaa	cagaaagggtg	caggaggtac	cttaactcac	ccctcagcaa	atacctatgt	2340
caacagtata	agttaccatt	tactctataa	tctgcagtga	tgcaataacc	agcataataa	2400
a						2401

<210> 19

<211> 583

5

<212> PRT

<213> *Homo sapiens*

10

<400> 19

# ES 2 380 147 T3

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Gln Asp Lys Arg Leu Ser Arg  
580



## ES 2 380 147 T3

<210> 20

<211> 1945

5 <212> ADN

<213> *Homo sapiens*

<400> 20

10 **ggaagccttg gcactagcgg cgcccgggcg cggagtgcgc agggcaaggt cctgcgctct** 60

gggccagcgc tccggccatgc gatccgcgcgc gcggagggga cgcgccgcgc ccgccgccag	120
ggactctttg ccggtgctac tgtttttatg cttgcttctg aagacgtgtg aacctaaaac	180
tgcaaatgcc tttaaaccaa atatcctact gatcatggcg gatgatctag gcaactgggga	240
tctcggttgc tacgggaaca atacactgag aacgccgaat attgaccagc ttgcagagga	300
aggtgtgagg ctcactcagc acctggcggc cgccccgcgc tgcaccccaa gccgagctgc	360
attcctcaca gggagacatt ccttcagatc aggcattggac gccagcaatg gataccgggc	420
ccttcagtgg aacgcaggct caggtggact ccctgagaac gaaaccactt ttgcaagaat	480
cttgccagcag catggctatg caaccggcct cataggaaaa tggcaccagg gtgtgaattg	540
tgcattccgc ggggatcact gccaccaccc cctgaaccac ggatttgact atttctacgg	600
catgcccttc acgctcacia acgactgtga cccaggcagg cccccgaag tggacgccgc	660
cctgagggcg cagctctggg gttacacca gttcctggcg ctggggattc tcacctggc	720
tgcgggccag acctgcggtt tcttctctgt ctccgcgaga gcagtcaccg gcatggccgg	780
cgtgggctgc ctgtttttca tctcttggtt ctctctcttc gggtttgtgc gacgctggaa	840
ctgtatcctg atgagaaacc atgacgtcac ggagcaacc cctgttcttg agaaaacagc	900
gagtcttatg ctaaaggaag ctgttttcta tattgaaaga cacaagcatg ggccatttct	960
cctcttctct tctttgctgc atgtgcacat tccccttggt accacgagtg cattcctggg	1020
gaaaagtcag catggcttat atggtgataa tgtggaggag atggactggc tcataggtaa	1080
ggttcttaaat gccatcgaag acaatggttt aaagaactca acattcacgt atttcacctc	1140
tgaccatgga ggacatttag aggcaagaga tggacacagc cagttagggg gatggaacgg	1200
aatttacaaa ggtgggaagg gcatgggagg atgggaaggt gggatccgag tgcccgggat	1260
cttccactgg ccgggggtgc tcccggccgg ccgagtgatt ggagagccca cgagcctgat	1320
ggacgtgttc cctactgtgg tccagctggt ggggtggcag gtgccccagg acagggtgat	1380
tgatggccac agcctggtac ccttgctgca gggagctgag gcacgctcgg cacatgagtt	1440
cctgtttcat tactgtgggc agcatcttca cgcagcacgc tggcaccaga aggacagtgg	1500
aagcgtctgg aaggttcatt acacgacccc gcagttccac cccgaggagc ggggcctgct	1560
aacggccgag gcgtctgccc atgctgaatg gggaggcgtg acccatcaca gacccccctt	1620
gctctttgac ctctccaggg acccctccga ggcacggccc ctgacccccg actccgagcc	1680
cctgtaccac gccgtgatag caagggtagg tgccgcggtg tcggagcatt ggcagaccct	1740
gagtcctgtg cccagcagt tttccatgag caacatcctg tggaagccgt ggctgcagcc	1800
gtgctgcgga catttccgt tctgttcatg ccacgaggat ggggatggca cccctgaat	1860
gccaggactg tgagagagga tccaggagag cctgactgcg ttgcaaacia aattctccaa	1920
gcttggttct atcttcagtc cggaa	1945

<211> 593

<212> PRT

5 <213> *Homo sapiens*

<400> 21

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

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

10

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Pro

<210> 22

<211> 1858

5 <212> ADN

<213> *Homo sapiens*

<400> 22

10    **ccttcctctt cttgatcggg gattcaggaa ggagcccagg agcagaggaa gtagagagag    60**  
      **agacaacatg ttacatctgc accattcttg tttgtgtttc aggagctggc tgccagcgat    120**

gctcgtctgta ctgctaagtt tggcaccatc agcttccagc gacatttccg cctccccgacc 180  
 gaacatcctt cttctgatgg cggacgacct tggcattggg gacattggct gctatggcaa 240  
 caacaccatg aggactccga atattgaccg ccttgacagag gacggcgtga agctgaccca 300  
 acacatctct gccgcatctt tgtgcacccc aagcagagcc gccttctca cgggcagata 360  
 ccctgtgcga tcagggatgg tttccagcat tggttaccgt gttcttcagt ggaccggagc 420  
 atctggaggt cttccaacaa atgagacaac ttttgcaaaa atactgaaag agaaaggcta 480  
 tgccactgga ctcatggaa aatggcatct ggggtctcaac tgtgagtcag ccagtgatca 540  
 ttgccaccac cctctccatc atggctttga gcatttctac ggaatgcctt tctccttgat 600  
 gggtgattgc gcccgctggg aactctcaga gaagcgtgtc aacctggaac aaaaactcaa 660  
 ctctctcttc caagtcctgg ccttggttgc cctcacactg gtagcagga agctcacaca 720  
 cctgataccc gtctcgtgga tgccggtcac ctggtcagcc ctttcggccg tcctcctcct 780  
 cgcaagctcc tattttgtgg gtgctctgat tgtccatgcc gattgcttct tgatgagaaa 840  
 ccacaccatc acggagcagc ccatgtgctt ccaaagaacg acaccctta ttctgcagga 900  
 ggttgcgctc tttctcaaaa ggaataagca tgggcctttc ctctctttg tttcctttct 960  
 acacgttcac atccctctta tcaactatgga gaacttctc gggaagagtc tccacgggct 1020  
 gtatggggac aacgtagagg agatggactg gatggtagga cggatccttg acactttgga 1080  
 cgtggagggg ttgagcaaca gcaccctcat ttattttacg tcggatcacg gcgggttcct 1140  
 agagaatcaa cttggaaaca ccagtatgg tggctggaat ggaatttata aagggtgggaa 1200  
 gggcatggga ggatgggaag gtgggatccg cgtgcccggg atcttccgct ggcccggggg 1260  
 gctcccggcc ggccgagtga ttggcgagcc cacgagtctg atggacgtgt tccccaccgt 1320  
 ggtccggctg gcgggcggcg aggtgcccc aagacagagt attgacggcc aagaccttct 1380  
 gcccttgctc ctggggacag cccaacactc agaccacgag ttctgatgc attattgtga 1440  
 gaggtttctg cagcgagcca ggtggcatca acgggacaga ggaacaatgt ggaaagtcca 1500  
 ctttgtgacg cctgtgttcc agccagaggg agccggtgcc tgctatggaa gaaaggctctg 1560  
 cccgtgcttt ggggaaaaag tagtccacca cgatccacct ttgctctttg acctctcaag 1620  
 agaccttct gagacccaca tcctcacacc agcctcagag cccgtgttct atcagggtgat 1680  
 ggaacgagtc cagcaggcgg tgtgggaaca ccagcggaca ctacagccag ttctctgca 1740  
 gctggacagg ctgggcaaca tctggagacc gtggctgcag cctgctgtg gcccgttccc 1800  
 cctctgctgg tgccttaggg aagatgaccc acaataaatg tctgcagtga aaagctgg 1858

<210> 23

5 <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

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

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



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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

<210> 24

5 <211> 1996

<212> ADN

<213> *Homo sapiens*

5 <400> 24

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

taatgagaca acacttgcag ccttgctaaa gaagcaagga tacagcacgg ggcttatagg	480
caaatggcac caaggcttga actgcgactc cgaagtgac cagtgccacc atccatataa	540
ttatggggtt gactactact atggcatgcc gttcactctc gttgacagct gctggccgga	600
cccctctcgt aacacggaat tagcctttga gagtcagctc tggctctgtg tgcagctagt	660
tgccattgcc atcctcaccc taacctttgg gaagctgagc ggctgggtct ctgttcctg	720
gctcctgata ttctccatga ttctgtttat ttctctcttg ggctatgctt ggttctccag	780
ccacacgtcc cctttataact gggactgcct cctcatgcgg gggcacgaga tcacggagca	840
gcccataaag gctgaacgag ctggatccat tatggtgaag gaagcgattt cctttttaga	900
aaggcacagt aaggaaactt tccttctctt ttctctctt cttcacgtgc acacacctct	960
ccccaccacg gacgatttca ctggcaccag caagcatggc ttgtatgggg ataatgtgga	1020
agagatggac tccatggtgg gcaagattct tgatgctatc gatgattttg gcctaaggaa	1080
caacaccctt gtctacttta catcagatca cggagggcat ttggaagcta ggcgagggca	1140
tgcccaactt ggtggatgga atggaatata caaaggtgga aaaggcatgg ggggctggga	1200
aggtggaatc cgcgccccag gaattgtccg atggcctgga aaggtaccag ctggacgggt	1260
gattaaggaa cctacaagtt taatggatat ttaccaact gtcgcatcag tgtcaggagg	1320
aagtctccct caggacaggg tcattgacgg ccgagacctc atgcccttgc tgcagggcaa	1380
cgtcaggcac tcggagcatg aatttctttt ccactactgt ggctcctacc tgcacgccgt	1440
gcggtggatc cccaaggacg acagtgggtc agtttggaag gctcactatg tgaccccggt	1500
attccagcca ccagcttctg gtggctgcta tgtcacctca ttatgcagat gtttcggaga	1560
acaggttacc taccacaacc cccctctgct cttcgatctc tccagggacc cctcagagtc	1620
cacacccttg acacctgcca cagagcccct ctatgatttt gtgattaaaa aggtggccaa	1680
cgccctgaag gaacaacagg aaaccatcgt gcctgtgacc taccaactct cagaactgaa	1740
tcagggcagg acgtggctga agccttgctg tggggtgttc ccattttgtc tgtgtgacaa	1800
ggaagaggaa gtctctcagc ctcggggtcc taacgagaag agataattac aatcaggcta	1860
ccagaggaag cctttggtcc taacgagaag agataattac aatcaggcta ccaaaggaag	1920
cactaacttt ggtgctttca agttggcaag gagtgcattt aatagtcaat aaattcatct	1980
accattccag attatt	1996

<210> 25

5 <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				

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

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

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

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

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

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

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

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

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

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

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

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

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

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

<210> 26

5 <211> 1578

<212> ADN

10 <213> *Homo sapiens*

<400> 26

```

atgggctggc tttttctaaa ggttttgttg gcgggagtga gtttctcagg atttctttat    60
cctcttgttg atttttgcat cagtgggaaa acaagaggac agaagccaaa ctttgtgatt    120
attttggccg atgacatggg gtgggggtgac ctgggagcaa actgggcaga aacaaaggac    180
actgccaacc ttgataagat ggcttcggag ggaatgaggt ttgtggattt ccatgcagct    240
gcctccacct gctcaccctc ccgggcttcc ttgctcaccg gccggcttgg ccttcgcaat    300
ggagtcacac gcaactttgc agtcacttct gtgggaggcc ttccgctcaa cgagaccacc    360
ttggcagagg tgctgcagca ggcggttac gtcactggga taataggcaa atggcatctt    420
ggacaccacg gctcttatca ccccaacttc cgtgggttttg attactactt tggaatccca    480
tatagccatg atatgggctg tactgatact ccaggctaca accaccctcc ttgtccagcg    540

tgtccacagg gtgatggacc atcaaggaac cttcaaagag actgttacac tgacgtggcc    600
ctccctcttt atgaaaacct caacattgtg gagcagccgg tgaacttgag cagccttgcc    660
cagaagtatg ctgagaaagc aaccagttc atccagcgtg caagcaccag cgggaggccc    720
ttctgctct atgtggctct ggcccacatg cacgtgccct tacctgtgac tcagctacca    780
gcagcgccac ggggcagaag cctgtatggt gcagggtctt gggagatgga cagtctggtg    840
ggccagatca aggacaaagt tgaccacaca gtgaaggaaa acacattcct ctggtttaca    900
ggagacaatg gcccggtggc tcagaagtgt gagctagcgg gcagtgtggg tcccttcact    960
ggatttttgg aaactcgtca agggggaagt ccagccaagc agacgacctg ggaaggaggg   1020
caccgggtcc cagcactggc ttactggcct ggcagagttc cagttaatgt caccagcact   1080
gccttgttaa gcgtgctgga cttttttcca actgtggtag ccctggccca ggccagctta   1140
cctcaaggac ggcgctttga tgggtgtggac gtctccgagg tgctcttttg ccggtcacag   1200
cctgggcaca ggggtgctgt ccaccccaac agcggggcag ctggagagtt tggagccctg   1260
cagactgtcc gcctggagcg ttacaaggcc ttctacatta ccggtggagc cagggcgtgt   1320
gatgggagca cggggcctga gctgcagcat aagtttcctc tgattttcaa cctggaagac   1380
gataccgcag aagctgtgcc cctagaaaga ggtgggtgcg agtaccaggc tgtgctgccc   1440
gaggtcagaa aggttcttgc agacgtcctc caagacattg ccaacgacaa catctccagc   1500
gcagattaca ctcaggaccc ttcagtaact ccctgctgta atccctacca aattgcctgc   1560
cgctgtcaag ccgcataa                                1578

```

5 <210> 27

<211> 525

<212> PRT

10

<213> *Homo sapiens*

<400> 27



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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Gln His Lys phe Pro Leu Ile Phe Asn Leu Glu Asp Asp Thr Ala Glu  
450 455 460

Ala Val Pro Leu Glu Arg Gly Gly Ala Glu Tyr Gln Ala Val Leu Pro  
465 470 475 480

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

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

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

<210> 28

5 <211> 4669

<212> ADN

<213> *Homo sapiens*

10 <400> 28

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

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

```

cctctgatta gatgaaactg ttaccttacc ctaaacacag tatttctttt taactttttt 3960
atgtgtaaac taataaaggt aatcacagcc accaacattc caagetaccc tgggtacctt 4020
tgtgcagtag aagctagtga gcatgtgagc aagcgggtgtg cacacggaga ctcatcggtta 4080
taatttacta tctgccaaga gtagaaagaa aggctgggga tatttgggtt ggcttggttt 4140
tgattttttg cttgtttgtt tgttttgtac taaaacagta ttatcttttg aatatcgtag 4200
ggacataagt atatacatgt tatccaatca agatggctag aatggtgcct ttctgagtgt 4260
ctaaaacttg acacccttg taaatctttc aacacacttc cactgcctgc gtaatgaagt 4320
tttgattcat ttttaaccac tggaattttt caatgccgtc attttcagtt agatgatttt 4380
gcactttgag attaaaatgc catgtctatt tgattagtct tattttttta tttttacagg 4440
cttatcagtc tcaactgttg ctgtcattgt gacaaagtca aataaacccc caaggacgac 4500
acacagtatg gatcacatat tgtttgacat taagcttttg ccagaaaatg ttgcatgtgt 4560
tttacctcga cttgctaaaa tcgattagca gaaaggcatg gctaataatg ttggtggtga 4620
aaataataa ataagtaaat gaaaaaaaaa aaaaaaaaaa aaaaaaaaaa 4669

```

<210> 29

5 <211> 871

<212> PRT

<213> *Homo sapiens*

10 <400> 29

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

# ES 2 380 147 T3

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



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

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

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

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

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

Leu Trp Asp Gly Trp Glu Gly  
865 870

<210> 30

5 <211> 4279

<212> ADN

<213> *Homo sapiens*

10 <400> 30

gggccatttc	tggaacaacag	ctgctatttt	cacttgagcc	caagttaatt	tctcggggag	60
ttctcgggcg	cgcacaggca	gctcggtttg	ccctgcgatt	gagctgcggg	tcgcggccgg	120
cgccggcctc	tccaatggca	aatgtgtgtg	gctggaggcg	agcgcgaggc	tttcggcaaa	180
ggcagtcgag	tgtttgcaga	ccggggcgag	tcctgtgaaa	gcagataaaa	gaaaacattt	240
attaacgtgt	cattacgagg	ggagcgcccg	gccggggctg	tcgcactccc	cgcggaacat	300
ttgggtccct	ccagctccta	gagaggagaa	gaagaaagcg	gaaaagaggc	agattcacgt	360
cgtttccagc	caagtggacc	tgatcgatgg	ccctcctgaa	tttatcacga	tatttgattt	420
attagcgatg	ccccctgggt	tgtgtgttac	gcacacacac	gtgcacacaa	ggctctgggt	480
cgtttccctc	cctcgtttcc	agctcctggg	cgaatcccac	atctgtttca	actctccgcc	540
gagggcgagc	aggagcgaga	gtgtgtcgaa	tctgcgagtg	aagagggacg	agggaaaaga	600
aacaaagcca	cagacgcaac	ttgagactcc	cgcattccaa	aagaagcacc	agatcagcaa	660
aaaaagaaga	tgggcccccc	gagcctcgtg	ctgtgcttgc	tgcccgcaac	tgtgttctcc	720
ctgctgggtg	gaagctcggc	cttcctgtcg	caccaccgcc	tgaaaggcag	gtttcagagg	780
gaccgcagga	acatccgccc	caacatcate	ctggtgctga	cggacgacca	ggatgtggag	840
ctgggttcca	tgcaggtgat	gaacaagacc	cggcgcacat	tggagcaggg	cggggcgcac	900
ttcatcaacg	ccttcgtgac	cacacccatg	tgtgtccctc	cacgtcctc	catcctcacc	960
ggcaagtacg	tccacaacca	caacacctac	accaacaatg	agaactgctc	ctcgccctcc	1020
tggcaggcac	agcacgagag	cgcaccttt	gccgtgtacc	tcaatagcac	tggctaccgg	1080
acagctttct	tcgggaagta	tcttaatgaa	tacaacggct	cctacgtgcc	acccggctgg	1140
aaggagtggg	tcggactcct	taaaaactcc	cgtttttata	actacacgct	gtgtcggaac	1200
ggggtgaaag	agaagcacgg	ctccgactac	tccaaggatt	acctcacaga	cctcatcacc	1260
aatgacagcg	tgagcttctt	cgcacgtcc	aagaagatgt	acccgcacag	gccagtctc	1320
atggtcacat	gccatgcagc	ccccacggc	cctgaggatt	cagcccaca	atattcacgc	1380
ctcttcccaa	acgcattctc	gcacatcacg	cgcagctaca	actacgcgcc	caaccgggac	1440
aaacactgga	tcatgcgcta	cacggggccc	atgaagccca	tccacatgga	attcaccaac	1500
atgctccagc	ggaagcgctt	gcagaccctc	atgtcgggtg	acgactccat	ggagacgatt	1560

tacaacatgc	tggttgagac	gggCGagctg	gacaacacgt	acatcgata	caccgccgac	1620
cacggttacc	acatcgcca	gtttggcctg	gtgaaagga	aatccatgcc	atatgagttt	1680
gacatcaggg	tcccgttcta	cgtgaggggc	cccaacgtgg	aagccggctg	tctgaatccc	1740
cacatcgctc	tcaacattga	cctggccccc	accatcctgg	acattgcagg	cctggacata	1800
cctgcggata	tggaacggaa	atccatcctc	aagctgctgg	acacggagcg	gccggtgaat	1860
cggtttcact	tgaaaaagaa	gatgagggtc	tggggggact	ccttcttggg	ggagagaggg	1920
aagctgctac	acaagagaga	caatgacaag	gtggacgccc	aggaggagaa	ctttctgccc	1980
aagtaccagc	gtgtgaagga	cctgtgtcag	cgtgctgagt	accagacggc	gtgtgagcag	2040
ctgggacaga	agtggcagtg	tgtggaggac	gccacgggga	agctgaagct	gcataagtgc	2100
aagggcccca	tgcggctggg	cggcagcaga	gccctctcca	acctcgtgcc	caagtactac	2160
gggcagggca	gcgaggcctg	cacctgtgac	agcggggact	acaagctcag	cctggccgga	2220
cgccggaaaa	aactcttcaa	gaagaagtac	aaggccagct	atgtccgag	tcgctccatc	2280
cgctcagtg	ccatcgaggt	ggacggcagg	gtgtaccacg	taggcctggg	tgatgccgcc	2340
cagccccgaa	acctcaccaa	gcggcactgg	ccaggggccc	ctgaggacca	agatgacaag	2400
gatggtgggg	acttcagtg	cactggaggc	cttcccagct	actcagccgc	caacccattt	2460
aaagtgcac	atcgggtgta	catcctagag	aacgacacag	tccagtgtga	cctggacctg	2520
tacaagtccc	tgcaggcctg	gaaagaccac	aagctgcaca	tcgaccacga	gattgaaacc	2580
ctgcagaaca	aaattaagaa	cctgagggaa	gtccgaggtc	acctgaagaa	aaagcggcca	2640
gaagaatgtg	actgtcacaa	aatcagctac	cacaccagc	acaaaggccg	cctcaagcac	2700
agaggctcca	gtctgcatcc	tttcaggaag	ggcctgcaag	agaaggacaa	ggtgtggctg	2760
ttgccccgag	agaagcgcaa	gaagaaactc	cgcaagctgc	tcaagcgctc	gcagaacaac	2820
gacacgtgca	gcatgccagg	cctcacgtgc	ttcacccacg	acaaccagca	ctggcagacg	2880
gcgcctttct	ggacactggg	gcctttctgt	gcctgcacca	gcgccaacaa	taacacgtac	2940
tggtgcatga	ggaccatcaa	tgagactcac	aatttcctct	tctgtgaatt	tgcaactggc	3000
ttcctagagt	actttgatct	caacacagac	ccctaccagc	tgatgaatgc	agtgaacaca	3060
ctggacaggg	atgtcctcaa	ccagctacac	gtacagctca	tggaagctgag	gagctgcaag	3120
ggttacaagc	agtgtaaccc	ccggactcga	aacatggacc	tggaacttaa	agatggagga	3180
agctatgagc	aatacaggca	gtttcagcgt	cgaaagtggc	cagaaatgaa	gagaccttct	3240
tccaaatcac	tggaacaact	gtgggaaggc	tggaagggtt	aagaaacaac	agaggtggac	3300
ctccaaaaac	atagaggcat	cacctgactg	cacaggcaat	gaaaaaccat	gtgggtgatt	3360
tccagcagac	ctgtgctatt	ggccaggagg	cctgagaaag	caagcacgca	ctctcagtca	3420
acatgcagag	ttctggagga	taaccagcag	gagcagagat	aacttcagga	agtccatttt	3480
tgcccttgct	tttgctttgg	attatacctc	accagctgca	caaatgcat	tttttcgtat	3540
caaaaagtca	ccactaaccc	tccccagaa	gtcacaaaag	gaaaacggag	agagcgagcg	3600
agagagattt	ccttggaagt	ttctcccaag	ggcgaaagtc	attggaattt	ttaaatcata	3660
ggggaaaagc	agtcctgttc	taaatcctct	tattcttttg	gtttgtcaca	aagaaggaaac	3720
taagaagcag	gacagaggca	acgtggagag	gctgaaaaca	gtgcagagac	gtttgacaat	3780
gagtcagtag	cacaaaagag	atgacattta	cctagcatat	aaaccctggg	tgccctctgaa	3840
gaaactgcct	tcattgtata	tatgtgacta	tttacatgta	atcaacatgg	gaacttttag	3900

# ES 2 380 147 T3

```

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

```

<210> 31

5 <211> 870

<212> PRT

<213> *Homo sapiens*

10 <400> 31

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

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

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



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

**Trp Glu Gly Trp Glu Gly**  
**865 870**

<210> 32

5 <211> 6

<212> PRT

<213> *Homo sapiens*

10 <220>

<221> VARIANTE

15 <222> (1)..(1)

<223> Leu o Val

<220>

20 <221> misc\_feature

<222> (1)..(3)

25 <223> Xaa puede ser un aminoácido que se produce de manera natural

<220>

<221> VARIANTE

30 <222> (2)..(2)

<223> Cys o Ser

35 <220>

<221> VARIANTE

<222> (3)..(3)

40 <223> Cualquier aminoácido

<400> 32

**Xaa Xaa Xaa Pro Ser Arg**  
**1 5**

45 <210> 33

<211> 23

50 <212> PRT

<213> Artificial

55 <220>

<223> Secuencia derivada de arilsulfatasa A humana

<220>

<221> PÉPTIDO

5 <222> (1)..(23)

<223> Sustrato de formación de *FGly* sintético; secuencia primaria de arilsulfatasa A humana

<400> 33

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

**Ala Leu Leu Thr Gly Arg Ser**  
**20**

<210> 34

15 <211> 16

<212> PRT

<213> Artificial

20 <220>

<223> Una variante del péptido ASA65-80, en la que los residuos Cys69, Pro71 y Arg73, críticos para la formación de *FGly*, estaban reorganizados al azar

25 <220>

<221> MISC\_FEATURE

30 <222> (1)..(16)

<223> Oligopéptido reorganizado al azar

<400> 34

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

<210> 35

40 <211> 16

<212> PRT

<213> Artificial

45 <220>

<223> Una variante del péptido ASA65-80, en la que la Cys69 se sustituyó por una serina

50 <220>

<221> MISC\_FEATURE

<222> (1)..(16)

55 <223> Oligopéptido de Ser69

<400> 35

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

<210> 36

5 <211> 19

<212> ADN

<213> Artificial

10 <220>

<223> Cebador de PCR específico de FGE humana

15 <220>

<221> misc\_feature

<222> (1)..(19)

20 <223> Cebador 1199nc de PCR específico de FGE humana

<400> 36

25 ccaatgtagg tcagacacg 19

<210> 37

<211> 16

30 <212> ADN

<213> Artificial

35 <220>

<223> Cebador de PCR específico de FGE humana

<220>

40 <221> misc\_feature

<222> (1)..(16)

45 <223> Cebador 1c directo específico de FGE humana

<400> 37

acatggcccg cgggac 16

50 <210> 38

<211> 19

55 <212> ADN

<213> Artificial

<220>

60 <223> Cebador de PCR específico de FGE humana

<220>

<221> misc\_feature  
 <222> (1)..(19)  
 5 <223> Cebador 1182c inverso específico de FGE humana  
 <400> 38  
 cgactgctcc ttggactgg 19  
 10 <210> 39  
 <211> 24  
 15 <212> ADN  
 <213> Artificial  
 <220>  
 20 <223> Cebador de PCR específico de FGE humana  
 <220>  
 25 <221> misc\_feature  
 <222> (1)..(24)  
 <223> Cebador 5' de PCR específico de FGE humana que contiene EcoRI  
 30 <400> 39  
 ggaattcggg acaacatggc tgcg 24  
 35 <210> 40  
 <211> 54  
 <212> ADN  
 40 <213> Artificial  
 <220>  
 45 <223> Cebador específico de HA  
 <220>  
 <221> misc\_feature  
 50 <222> (1)..(54)  
 <223> Cebador específico de HA  
 55 <400> 40  
 cccaagctta tgcgtagtca ggcacatcat acggatagtc catggtgggc aggc 54  
 60 <210> 41  
 <211> 57  
 <212> ADN  
 65 <213> Artificial

<220>  
 <223> Cebador específico de c-myc  
 5 <220>  
 <221> misc\_feature  
 <222> (1)..(57)  
 10 <223> Cebador específico de c-myc  
 <400> 41  
 15 cccaagctta caggctctct tcagaaatca gctttgttc gtccatgggt ggcaggc 57  
 <210> 42  
 <211> 54  
 20 <212> ADN  
 <213> Artificial  
 25 <220>  
 <223> Cebador específico de RGS-His6  
 <220>  
 30 <221> misc\_feature  
 <222> (1)..(54)  
 35 <223> Cebador específico de RGS-His6  
 <400> 42  
 cccaagctta gtgatgtga tggatgatgcg atcctctgtc catggtgggc aggc 54  
 40 <210> 43  
 <211> 15  
 45 <212> PRT  
 <213> Artificial  
 <220>  
 50 <223> Oligopéptido tríptico de una preparación de FGE humana  
 <220>  
 55 <221> MISC\_FEATURE  
 <222> (1)..(15)  
 <223> Oligopéptido tríptico de una preparación de FGE humana  
 60 <400> 43

<b>Ser</b>	<b>Gln</b>	<b>Asn</b>	<b>Thr</b>	<b>Pro</b>	<b>Asp</b>	<b>Ser</b>	<b>Ser</b>	<b>Ala</b>	<b>Ser</b>	<b>Asn</b>	<b>Leu</b>	<b>Gly</b>	<b>Phe</b>	<b>Arg</b>
<b>1</b>				<b>5</b>					<b>10</b>					<b>15</b>

<210> 44

<211> 19

5 <212> PRT

<213> Artificial

<220>

10 <223> Oligopéptido tríptico de una preparación de FGE humana

<220>

15 <221> MISC\_FEATURE

<222> (1)..(19)

20 <223> Oligopéptido tríptico de una preparación de FGE humana

<400> 44

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

**Gln Ile Lys**

25 <210> 45

<211> 906

<212> ADN

30 <213> *Homo sapiens*

<400> 45

atggcccggc atgggttacc gctgctgccc ctgctgtcgc tcctggtcgg cgcgaggctc	60
aagctaggaa atggacaggc tactagcatg gtccaactgc aggggtgggag attcctgatg	120
ggaacaaatt ctccagacag cagagatggt gaagggcctg tgcgggaggc gacagtgaaa	180
ccctttgcca tcgacatatt tcctgtcacc aacaaagatt tcagggattt tgtcaggag	240
aaaaagtatc ggacagaagc tgagatgttt ggatggagct ttgtctttga ggactttgtc	300
tctgatgagc tgagaaacaa agccacccag ccaatgaagt ctgtactctg gtggcttcca	360
gtggaaaagg cattttggag gcagcctgca ggtcctggct ctggcatccg agagagactg	420
gagcaccagc tgttacacgt gagctggaat gacgcccgtg cctactgtgc ttggcgggga	480
aaacgactgc ccacggagga agagtgggag tttgccgccc gagggggctt gaagggtaa	540
gtttacccat ggggggaactg gttccagcca aaccgcacca acctgtggca gggaaagttc	600
ccaagggag acaaagctga ggatggcttc catggagtct cccagtgaa tgctttcccc	660
gccagaaca actacgggct ctatgacctc ctggggaacg tgtgggagtg gacagcatca	720
ccgtaccagg ctgctgagca ggacatgcgc gtcctccggg gggcatcctg gatcgacaca	780
gctgatggct ctgccaatca ccgggcccgg gtcaccacca ggatgggcaa cactccagat	840
tcagcctcag acaacctcgg tttccgctgt gctgcagacg caggccggcc gccaggggag	900

ctgtaa 906

5 <210> 46

<211> 301

<212> PRT

10 <213> *Homo sapiens*

<400> 46

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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



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

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

<210> 47

5 <211> 927

<212> ADN

<213> *Mus musculus*

10

<400> 47

```

atgcgctctg agttctgggt cccagcatg gggtccttgc tccctccggt gttgctgctg      60
aggtcctctg cctgccccag gcttcagcta ggacatgcc aggatcctgc catggtgcat      120
ctgccaggtg gccggtttct gatggggaca gacgctccag atggcagaga cggatgaaggg      180
cctgccccggg aagtgcagct aaaacccttt gccatcgaca tatttccagt caccaataaa      240
gacttcaggg agtttgtcag ggagaagaag taccagactg aagccgaggc attcgggtgg      300
agcttcgtct ttgaggattt tgtctccctt gagctcagaa agcaagaaaa tctgatgccg      360
gctgttcact ggtggcagcc agtgccaaag gcatttttga ggcagcctgc aggtcccggc      420
tctggcatcc gagagaaact ggagcttccc gtggtacacg tgagctggaa cgacgtgggt      480
gcttactgcg catggcgggg gagacgcttg cccacagaag aggagtggga gtttgcagcc      540
cgaggggggt tgaagggtca ggtttatcca tgggggaacc ggttcagcc aaaccgcacc      600
aacttatggc agggaaagtt ccccaaaggt gacaaagctg aagatggttt tcatggactg      660
tcaccagtga acgctttccc cccacagaac aactacggac tgtatgacct catgggcaat      720
gtgtgggagt ggacagcgtc cacataccaa cctgctggcc aggacatgcg tgcctccgg      780
ggggcatcat ggatcgacac cgcagacggc tctgctaate acagggctcg ggtcaccacc      840
aggatgggaa aactccaga ctcagcctca gacaacctgg gcttccgctg cgcctccagt      900
gcaggccgac cgaaggagga cctgtga                                     927

```

15 <210> 48

<211> 308

<212> PRT

20

<213> *Mus musculus*

<400> 48

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

Ala Phe Gly Trp Ser Phe Val Phe Glu Asp Phe Val Ser Pro Glu Leu  
100 105 110

Arg Lys Gln Glu Asn Leu Met Pro Ala Val His Trp Trp Gln Pro Val  
115 120 125

Pro Lys Ala Phe Trp Arg Gln Pro Ala Gly Pro Gly Ser Gly Ile Arg  
130 135 140

Glu Lys Leu Glu Leu Pro Val Val His Val Ser Trp Asn Asp Ala Gly  
145 150 155 160

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

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

Asn Arg Phe Gln Pro Asn Arg Thr Asn Leu Trp Gln Gly Lys Phe Pro  
195 200 205

Lys Gly Asp Lys Ala Glu Asp Gly Phe His Gly Leu Ser Pro Val Asn  
210 215 220

Ala Phe Pro Pro Gln Asn Asn Tyr Gly Leu Tyr Asp Leu Met Gly Asn  
225 230 235 240

Val Trp Glu Trp Thr Ala Ser Thr Tyr Gln Pro Ala Gly Gln Asp Met  
245 250 255

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

Asn His Arg Ala Arg Val Thr Thr Arg Met Gly Asn Thr Pro Asp Ser  
275 280 285

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

Lys Glu Asp Leu  
305

<210> 49

5 <211> 855

<212> ADN

<213> *Mus musculus*

10

<400> 49

atgggtcccca	ttcctgctgg	agtattcaca	atggggcactg	atgatacctca	gatcaggcag	60
gatggagaag	cccctgccag	gagagtcact	gttgatggct	tttacaatgga	cgccatatgaa	120
gtcagcaatg	cggattttga	gaagtttgtg	aactcgactg	gctatattgac	agaggctgag	180
aagtttggag	actctttcgt	ctttgaaggc	atgttgagcg	agcaagtga	aacgcataatc	240
caccaggcag	ttgcagctgc	tccatggtgg	ttgcctgtca	agggagctaa	ttggagacac	300
ccagaggggtc	cggactccag	tattctgcac	aggtcaaate	atccgggttct	ccatgtttcc	360
tggaacgatg	ctgttgctta	ctgcacatgg	gcgggcaaga	ggttgcctac	tgaggcagag	420
tgggaataca	gctgtagagg	aggcctgcag	aacaggcttt	tcccctgggg	caacaaactg	480
cagcccaaag	gacagcatta	tgccaacatc	tggcagggca	agtttcctgt	gagcaacact	540
ggcgaggatg	gcttccaagg	aactgcccc	gttgatgcct	ttcctcccaa	tggctatggc	600
ttatacaaca	tagtggggaa	tgtgtgggag	tggacctcag	actggtggac	tgttcaccat	660
tctgttgagg	aaacgttcaa	cccaaagggt	cccacttctg	ggaaagaccg	agtgaagaag	720
ggtggatcct	acatgtgcca	taagtcctat	tgtataggt	accgctgtgc	agctcgaagc	780
cagaacacac	cagatagctc	tgcataccaac	ctgggattcc	gatgtgcagc	cgaccacctg	840
cccaccgcag	actga					855

5 <210> 50

<211> 284

<212> PRT

10

<213> *Mus musculus*

<400> 50

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

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

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

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

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

<210> 51

5 <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 ccgggaggca cggtttacat ggggtactgac      240
aaaccgcact ttccggccga ccgcgaggct ccggaacggc aggtgaagct gaatgacttc      300
tacatcgaca agtatgaggt ttccaacgaa gcctttgcga agtttggtct gcacactaac      360
tacaccacgg aggtgtagcg atatggcgac agttttctgt ttaagagcct tttagagcca      420
ttggagcaga agaacctaga ggacttccga gtggcgagcg ctgtctggtg gtacaaagtg      480
gccggcgtga actggcgaca tccaaatggc gtggacagcg atatagacca cttaggccga      540
cacccggtag tgcacgtatc gtggcgcgac gctgtggagt actgtaagtg ggccggcaag      600
cggttgcccc gcgaggcgga gtgggaggcg gcttgaggcg gcggcaagga gcgcaaactg      660
tttccttggg gcaacaagct gatgccaagg aatgaacatt ggctgaacat ctggcaggga      720
gactttcccc atggcaacct ggctgaagat gggtttgagt acaccagccc cgtggatgcc      780
ttccgacaga atatttacga cctgcacaac atggtgggca acgtctggga gtggacggca      840
gatctgtggg acgtaaatga cgtttagcgat aatccaaatc gggtaagaa gggcggttct      900
tatctgtgtc acaagtccta ctgtacagg tacaggtgcg cggcacgctc gcagaacaca      960
gaagacagtt cagccggtaa cctgggtttt cgggtgcgcca agaattgcgtg a          1011
```

15 <210> 52

<211> 336

<212> PRT

<213> *Drosophila melanogaster*

5 <400> 52

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

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

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

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

ES 2 380 147 T3

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



<210> 53

<211> 870

5 <212> ADN

<213> *Anopheles gambiae*

<400> 53

10 **ccggagagct tgctcgatct ggtggaacat tccaagcggc tcgaagacat gaggccttattc 60**

**ccaggagggtg aatatgtaat cggcacaaat gaacctatct tcgtcaagga tcgcgaatca 120**

**ccggccccggc ccgcgacgat ccgcgacttt taectcgacc agtacgaagt ctccaacgca 180**

**cagttcaagg cattcgtcga ccagacgggc tacgtcacgg aggcggaaaa gtttggcgac 240**

**agcttcgtct tccagcagct gctcagcgaa ccggtgcgcc agcagtacga agatttccgc 300**

**gtggcggcgg cgccctggtg gtacaaggta cgtggagcct cctggcagca tccggaaggc 360**

**gatgtgtcac gtgatataag cgaccgattg gaccatccgg tgggtgcacgt gtcctggaac 420**

**gatgcggtcg cgtactgcgc ctggaaaggg aagcgctgc cgacggaagc ggaatgggaa 480**

**gcggcctgcc ggggcggtcg caagcagaag ctgttccctt ggggtaacaa gctgatgccg 540**

**aaggagcagc acatgatgaa catatggcag ggcgagttcc cggacagcaa tctgaaggag 600**

**gatggctacg agaccacctg cccggtgacg tccttccgcc agaaccggtt cgagctgtac 660**

**aacatcgttg gcaacgtgtg ggagtggacg gcggatcttt gggacgcgaa ggatgcggcc 720**

**atcgagcgca agccgggcag cgatccaccg aatcgggtga aaaaggggtg ctcataacctg 780**

**tgtcacgaat cgtactgcta tcgctatcgc tgtgcggctc gatcgcagaa caccgaggac 840**

**agttcggcgg gcaatctggg cttccgggtgc 870**

15 <210> 54

<211> 290

<212> PRT

20 <213> *Anopheles gambiae*

<400> 54

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

[illegible]

<210> 55

5 <211> 945

<212> ADN

<213> *Streptomyces coelicolor*

10

<400> 55

```

gtggcgtgg ccgccccgtc ccccgcgggc gccgcggagc cggggccccgc cgcccgtcgc 60
cgctcgaccc gcggacaggt gcgcctgccg ggcggtgagt tcgcgatggg ggacgccttc 120
ggggagggat atccggccga cggcgagaca cccgtgcaca cggtgcgccct gcggcccttc 180
cacatcgacg agaccgccgt caccaacgcc cggttcgccg ccttcgtcaa ggcgaccggc 240
catgtgaccg acgccgaacg cttegggtcc tcggcgtctt tccacctggg cgtcgccgcc 300
ccggacgccg acgtcctcgg cagcgccgcc ggcgccccct ggtggatcaa cgtgcggggc 360
gcccactggc gccgccccga gggcgccccg tccgacatca ccggccggcc gaaccatccg 420
gtcgtccacg tctcctggaa cgatgccacc gcctacgcgc ggtggggccg caagcgctg 480
cccaccgagg ccgaatggga gtacgccgcc cgcgggggac tggccggccg ccgctacgcc 540
tggggcgacg agctgacccc gggcgggccg tggcgctgca acatctggca gggccgcttc 600
ccgcacgtca acacggccga ggacgggcac ctgagcaccg caccggtcaa gtcctaccgg 660
cccaacggcc acggcctgtg gaacaccgcg ggcaacgtgt gggaatggtg ctccgactgg 720
ttctcgccca cctactacgc cgaatcacc accgtcgacc cgcacggccc cgggaccggg 780
gcggcacggg tgctgcgcgg cggctcctac ctgtgccacg actcctactg caaccgctac 840
cgggtcgccg cccgctcctc caacaccccg gactcctcgt ccggcaacct cggtattccg 900
tgcgccaacg acgcggacct cacgtccgga tcagccgctg agtga 945

```

<210> 56

5 <211> 314

<212> PRT

<213> *Streptomyces coelicolor*

10

<400> 56

```

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

```

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

<210> 57

5 <211> 1005

<212> ADN

<213> *Corynebacterium efficiens*

10

<400> 57

```
gtggttcgcc atcgactggg ccaccggccc tgcacactga ggattacgtc catgagtaac      60
tgctgctccc cgtcaagcgc acaatggcgt accactaccc gggatttatc agatcctgtc      120
aatcccacca ctccatgcaa cccggaacaa tcccgcgatg ctgtgacact gccgggtgga      180
gctttccaca tgggcgatca tcacggggag gggtagccgg cggacgggga ggggccagta      240
catgagggttc acctcgcccc cttcggcatt aatgtcacca cggtcacgaa tgccgagttc      300
ggacgattta ttgaagccac agggatatcg acgacagcgg aacgctacgg tgtctcggct      360
gtattctacg cagcgttcca agggcaacgc gctgacattc ttcgccaggt tcccggcgtg      420
ccctgggtggc tggcgggtcaa ggggtgcgaac tggcagcgtc ccaacggccc cggatccacc      480
ctggacggggc ttgaggacca ccccgctcgtt cacgtttcct gggatgatgc cgttgcctac      540
tgcacctggg ctggcggtcg tctgcccacc gaagccgagt ggaataacgc cggccgggggt      600
ggactgcagg gcgcacgata tgccctggggg gataacctcg ccctagacgg gaggtggaac      660
tgcaatatct ggcagggggg ctcccccatg gagaacaccg ccgcggatgg ttacctcacc      720
actgcaccgg tgaagacctt cacgcccatt ggatacggtc tgtggcagat ggcagggaat      780
gtatgggaat ggtgccagga ctggtttgat gcggagtact actcccgtgc ttcctccatc      840
aaccgcgggg gaccggatac cgggtgcgcgc cgggtgatgc gcggaggctc gtatctctgc      900
catgattcct actgcaacag ataccgggtg gccgcccgcg attcgaacac cccggattcc      960
acctcgggga ataccggttt cgggtgcggt ttcgatagtc cttga                      1005
```

15 <210> 58

<211> 334

<212> PRT

20

<213> *Corynebacterium efficiens*

<400> 58

# ES 2 380 147 T3

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

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											

<210> 59

5 <211> 1017

<212> ADN

<213> *Novosphingobium aromaticivorans*



<400> 59

atggcgcaac cattccgatac gacggcgggc agtcgtacaa gtattgaacg ccatctcgaa	60
cccaattgca ggagcacgtc gcgaatggtc gaacgccccg gcatgcgcct gatcgaaggc	120
ggcactttca ccatgggctc ggaagccttc taccgggagg aagcgccgct tcgcccgggtg	180
aaggtagaca gcttctggat cgatgaagcg ccggtgacga acgcacagtt cgccgcattc	240
gtggaggcca cgggatacgt cactgtggcc gagatcgagc cggatcccaa ggactacccc	300
ggcatgctcc cgggcatgga ccgcgcggga tcgctgggtg tccagaaaac agcagggccg	360
gtcgacatgg cggatgcgtc caactgggtg cactttacct ttggcgctg ctggaagcat	420
ccacttggac cgggcagttc catcgatggg atcgaggacc atcccgtcgt tcacgtcgcc	480
tatgccgatg ccgaggccta tgccaaatgg gcgggcaagg atctgccgac cgaagccgag	540
ttcgaatatg ctgcgcgcgg cgggttggac ggttccgaat tttcctgggg agacgaactc	600
gcacctgaag gccggatgat ggccaactac tggcaaggcc tgtttccctt cgccaaccag	660
tgcctcgatg gctgggaacg gacatcgccc gtccgcaact tcccgcccaa cggctatggt	720
ctttacgaca tgatcgggaa cacgtgggag tggacctgcg attggtgggc cgacaagccg	780
ctgactccgc aaaggaaatc ggcatgctgc gcgatcagca atccgcgcgg cggcaagctc	840
aaggacagct tcgaccgctc gcaaccgcga atgcgcatcg gccggaaggc cataaagggc	900
ggttcgcacc tgtgtgcggc caattactgc cagcgctatc gcccgcagc acgccatcct	960
5 gaaatggttg ataccgcgac gacgcacatc ggcttcaggt gtgtggtgcg gccctga	1017

<210> 60

<211> 338

10

<212> PRT

<213> *Novosphingobium aromaticivorans*

15

<400> 60

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

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

Arg Pro

<210> 61

5 <211> 1119

<212> ADN

<213> *Mesorhizobium loti*

10 <400> 61

atgggcccac gaggtcgagg tcaaaaaccg catgaaaggc gacgcggtca tgttcgacat	60
tgccgggaag ttctagccga tagcgggtgg gcggctgatg gagatgagca cgccgtgtca	120
tttcgggatc ttctgatgaa cgcacctgcc gaagtcttcg agcgcgctgc agccgaacgg	180
tcgtaccccc gaatggtctg gatccccggc ggtaccttcc tgatgggctc agacaaccac	240
tatccggagg aggcaccggc ccaccgggtc agggtcgacg gcttctggat ggacaaattc	300
accgtctcca accgcgaactt cgaacgcttc gttgcggcga caggacatgt cactcttgcc	360
gagaaacccg ccaatcccga cgactatccc ggtgccttac ccgatctgct ggctccgtcc	420
tcgatgatgt tcaggaagcc ggccggccct gtcgaccttg gcaatcacta caattggtgg	480
gtctatgtcc gcggcgccaa ctggcgccat ccacgcgggc cggcaagtac aatcaagaag	540
gttcgagatc atccggtcgt gcatgtggcc tacgaggatg tcgtggccta tgccaactgg	600
gcaggcaagg aacttcccac cgaggccgag tgggaattcg cggcgcgagg cggcctcgat	660
gccgccgaat acgtctgggg caacgagctt acgcgggccg ggaagcacat ggccaacatc	720
tggcaaggag actttcccta ccggaatact gtcgacgacg gttacgaata tacggcccca	780
gtaggctcgt tcccggccaa cgactacggt ctctacgaca tggccggcaa tgtctggcaa	840
tggacgaccg actggtacca ggaccacaag gcgatcgaca gcccgtgctg caccgctgtc	900
aatccgcgtg gcggccatcg cgaagcgagc tatgacaccc ggctacctga cgttaagatc	960
cctcgcaagg tcaccaaggg tggctcccat ctgtgcgcgc cgaactactg tcggcgctac	1020
 cggccccgcg cgcgaaatggc gcaaccgctc gacactgcaa tctcccatct cggctttcgc	 1080
tgcategtgc gaaggaaaat ggaattgaac gcgcagtaa	1119

5 <210> 62

<211> 372

<212> PRT

10

<213> *Mesorhizobium loti*

<400> 62

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

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

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

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

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

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

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

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

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

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

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

Thr Ile Lys Lys Val Ala Asp His Pro Val Val His Val Ala Tyr Glu  
180 185 190

Asp Val Val Ala Tyr Ala Asn Trp Ala Gly Lys Glu Leu Pro Thr Glu  
195 200 205

Ala Glu Trp Glu Phe Ala Ala Arg Gly Gly Leu Asp Ala Ala Glu Tyr  
210 215 220

Val Trp Gly Asn Glu Leu Thr Pro Ala Gly Lys His Met Ala Asn Ile  
225 230 235 240

Trp Gln Gly Asp Phe Pro Tyr Arg Asn Thr Val Asp Asp Gly Tyr Glu  
245 250 255

Tyr Thr Ala Pro Val Gly Ser Phe Pro Ala Asn Asp Tyr Gly Leu Tyr

260					265					270					
Asp	Met	Ala	Gly	Asn	Val	Trp	Gln	Trp	Thr	Thr	Asp	Trp	Tyr	Gln	Asp
		275					280					285			
His	Lys	Ala	Ile	Asp	Ser	Pro	Cys	Cys	Thr	Ala	Val	Asn	Pro	Arg	Gly
	290					295					300				
Gly	His	Arg	Glu	Ala	Ser	Tyr	Asp	Thr	Arg	Leu	Pro	Asp	Val	Lys	Ile
305					310					315					320
Pro	Arg	Lys	Val	Thr	Lys	Gly	Gly	Ser	His	Leu	Cys	Ala	Pro	Asn	Tyr
				325					330					335	
Cys	Arg	Arg	Tyr	Arg	Pro	Ala	Ala	Arg	Met	Ala	Gln	Pro	Val	Asp	Thr
			340					345					350		
Ala	Ile	Ser	His	Leu	Gly	Phe	Arg	Cys	Ile	Val	Arg	Arg	Lys	Met	Glu
		355					360					365			
Leu	Asn	Ala	Gln												
		370													

<210> 63

5 <211> 1251

<212> ADN

<213> *Burkholderia fungorum*

10

<400> 63

```

atgaagagtg aaagagatcg agagcccgca aagtcgtccc gctcgaacgg gtcggtcgca      60
gcaacccaaa cgcgcgccgg tcgctgtcgc aaactaatgt tgtggggcgc cctgctcgtc      120
atactgcccc cctgtgtcgg cgccgcggtc agttgggcct tcacgccgca cgcacccgct      180
cacccgcaaa tcgttttcgg cgacggcacg catgggtccg tcggcatggc gtgggtgccc      240
ggcggccagt tcctcatggg cagcgacgcc aacaggcgc aaccgaacga acgccccgcg      300
cacaaggtca aggtgcacgg cttctggatg gaccgccatc acgtgaccaa cgccgaattc      360
cgccgcttcg tcgaagcgac cggctacgtc accacggccg agaagaaacc cgactgggag      420
accctgaaag tccagttgcc gcccggcacg ccgcgcccg ccgagagcgc gatggtggcg      480
ggtgcaatgg tgttcgtcgg caccagccgt cccgtgccgc tagacgacta ttgcagtg      540
tgccgctatg tgcctggcgc taactggcgt catccagccg ggcctgagag caacatcatc      600
ggtaaagatg atcaccccg ggttcaagtg tcctacgaag atgcgcaggc ttatgcgaaa      660
tgggccggca agcgtctgcc gaccgaagcc gaatgggaat tcgccgcgcg cggcggcctc      720
gaacaggcca cgtatgcgtg gggcgatcag ttctctccca acggcaaaca gatggccaac      780
gtctggcagg gccagcagcc gcagtctttc cccgttgtca acccgaaagc ggggtggcgcg      840
ctcggtaaaa gtccggtggg tactttcccg gccaacggct acggcctttc cgacatgacc      900
ggcaacgcct ggcagtgggt tgccgactgg tatcgcgcgg atcagttcag gcgtgaggcg      960
gtaagcacca gcgcgatcga caatccggtg ggcccgagcg agtcgtggga ccccgagac     1020
cagggcggtgc ccgtcaacgc gcccaagcgt gtcacacgcg gcggttcgtt cctctgcaac     1080
gaaatctatt gcctgagcta ccggcccagc gcgagacgcg gcaccgatcc ctacaacagc     1140
atgtcgcate tgggcttccg gctgggtgatg gacgaagaca cctggaaaga agccggtgct     1200
cgccaggctt cggcgaaagc tgccggcgcg cctggaacct ctggcggtta g               1251

```

<210> 64

5 <211> 416

<212> PRT

<213> *Burkholderia fungorum*

10

<400> 64

ES 2 380 147 T3

Met Lys Ser Glu Arg Asp Arg Glu Pro Ala Lys Ser Ser Arg Ser Asn  
1 5 10 15

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

Met Leu Trp Gly Ala Leu Leu Val Ile Leu Pro Ala Cys Val Gly Ala  
35 40 45

Ala Val Ser Trp Ala Phe Thr Pro His Ala Pro Ala His Pro Gln Ile  
50 55 60

Val Phe Gly Asp Gly Thr His Gly Pro Leu Gly Met Ala Trp Val Pro  
65 70 75 80

Gly Gly Gln Phe Leu Met Gly Ser Asp Ala Lys Gln Ala Gln Pro Asn  
85 90 95

Glu Arg Pro Ala His Lys Val Lys Val His Gly Phe Trp Met Asp Arg  
100 105 110

His His Val Thr Asn Ala Glu Phe Arg Arg Phe Val Glu Ala Thr Gly  
115 120 125

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

Gln Leu Pro Pro Gly Thr Pro Arg Pro Pro Glu Ser Ala Met Val Ala  
145 150 155 160

Gly Ala Met Val Phe Val Gly Thr Ser Arg Pro Val Pro Leu Asp Asp  
165 170 175

Tyr Ser Gln Trp Trp Arg Tyr Val Pro Gly Ala Asn Trp Arg His Pro  
180 185 190

Ala Gly Pro Glu Ser Asn Ile Ile Gly Lys Asp Asp His Pro Val Val  
195 200 205

Gln Val Ser Tyr Glu Asp Ala Gln Ala Tyr Ala Lys Trp Ala Gly Lys  
210 215 220

Arg Leu Pro Thr Glu Ala Glu Trp Glu Phe Ala Ala Arg Gly Gly Leu  
225 230 235 240

Glu Gln Ala Thr Tyr Ala Trp Gly Asp Gln Phe Ser Pro Asn Gly Lys  
245 250 255

Gln Met Ala Asn Val Trp Gln Gly Gln Gln Pro Gln Ser Phe Pro Val  
260 265 270

Val Asn Pro Lys Ala Gly Gly Ala Leu Gly Thr Ser Pro Val Gly Thr



# ES 2 380 147 T3

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

<210> 65

5 <211> 912

<212> ADN

<213> *Sinorhizobium meliloti*

10

<400> 65

```

atggtctggg tccccggagc gaccttcatg atgggggtcga acgaccatta cccggaggaa 60
gcgcccgtgc atccggtaac cgtcgacgga ttctggatcg atgtgacacc ggtaacgaac 120
cgccagtttc tcgaattcgt aaatgcgacg gggcatgtga ccttcgcgga aagaaagccg 180
cgcgccgaag actatccggg cgctccgcca tccaatctaa gggccgggttc gctcgtcttc 240
acacccccga agcgaccgct gcaggggaacg gatatatcgc agtgggtggat attcacgctg 300
ggtgccaaact ggcggcaccc gctcggggcgc aagagcagca tcggagcgat tctggatcat 360
ccggtcgtcc atgtcgctta cagcgacgca aaggcctatg ccgaatgggc cggcaaggac 420
ctcccgaacc agaccgagtg ggagctggcg gcccgcggcg gcctcgatgg ggctgaattt 480
tcctggggcg gcgagcttgc gccgggcgga aatcacatgg ccaatacttg gcaggggaagt 540
tttccggtcg agaattctat ggacgatggt ttccgcgcga catcgccggt cagattttac 600
ccgccgaacg gctacggcct ctacgacatg atcggcaatg tgtgggagtg gaccacggat 660
tactggtccg tgcgccaccc ggaagcggcc gccaaagcctt gctgcattcc gagcaatccc 720
cgcaatgccg atgccgatgc gagtatcgat ccggcggcga gcgtgaaagt tccgcgccgg 780
gtgctcaagg gtggatcgca tctctgcgcg ccgaactact gccggcggtta ccgccctgcg 840
gcgaggcacg cccaggaaat cgacacgacg accagccatg tcggtttccg atgtgtcagg 900
cgcgttcgat aa 912

```

<210> 66

5 <211> 303

<212> PRT

<213> *Sinorhizobium meliloti*

10

<400> 66

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

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

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

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

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

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

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

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

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

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

Ser Trp Gly Gly Glu Leu Ala Pro Gly Gly Asn His Met Ala Asn Thr  
165 170 175

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

Arg Thr Ser Pro Val Arg Phe Tyr Pro Pro Asn Gly Tyr Gly Leu Tyr  
195 200 205

Asp Met Ile Gly Asn Val Trp Glu Trp Thr Thr Asp Tyr Trp Ser Val  
210 215 220

Arg His Pro Glu Ala Ala Ala Lys Pro Cys Cys Ile Pro Ser Asn Pro  
225 230 235 240

Arg Asn Ala Asp Ala Asp Ala Ser Ile Asp Pro Ala Ala Ser Val Lys  
245 250 255

Val Pro Arg Arg Val Leu Lys Gly Gly Ser His Leu Cys Ala Pro Asn  
260 265 270

Tyr Cys Arg Arg Tyr Arg Pro Ala Ala Arg His Ala Gln Glu Ile Asp  
275 280 285

Thr Thr Thr Ser His Val Gly Phe Arg Cys Val Arg Arg Val Arg  
290 295 300

<210> 67

<211> 1065

5 <212> ADN

<213> *Microscilla* sp.

<400> 67

10  
atgaaataca tttttttagt tcttttctta tgggccttga cccgatgtac cggaaagtat 60  
gaggacaaga gagtggaaac tgatacttcc agaccaaaag ccgaagcgtc agatataaaa 120  
gttcccgaag gaatggctta tattcccgcg ggccagtaca tgatgggagg taaatcagac 180  
caggcttata aggatgaata tccccgcat aacgtgaagg ttccggcttt ttatatggac 240  
cttacagaag tgaccaatgc ggagtttaag cgggtttgtag acgaaacggg ctacgtgacc 300  
attgctgaga aagatattga ctgggaagag ttaaagtctc aggtgccaca gggtagcccg 360  
aagcctcctg attctgtgct tcaggcaggc tcaactgggtt tcaagcagac agatgaaccc 420  
gtttctctcc aggattatcc acagtgggtg gaatggacta tcggagccaa ctggcgaaat 480  
ccggaggggc caggtagtac gattgaggat cgtatggatc atccgggtgt acacgtttcc 540  
tttgaagatg tccaagcgta tgcggattgg gccggtaagc gcctgcctac tgaggcagaa 600  
tgggaatggg ccgccatggg aggccaaaat gacgtgaaat atccatgggg aaatgaatcg 660  
gtcgaacaag catccgataa agcaaaacttt tggcagggga attttccaca tcaaaactat 720  
gccctcgatg gattcgaacg caccgcccct gtacgctcct tcccagcgaa tgggtacggc 780  
ctatatgata tggctggcaa tgtgtgggaa tgggtgccagg ataagtatga tgtcaatgct 840  
tatgaaagct ataagcaaaa aggactgaca gaagacccca cgggttctga gcactacaac 900  
gaccctaggg aaccgtatac tcctaagcat gtgatcagag ggggttcttt cctatgcaat 960  
gacagctact gtagtgggta tcgtgtttca cgtcgtatga gttccagtag agattcaggt 1020  
tttaatcata cgggattcag gtgtgtgaaa gatgtaaatg gatag 1065

<210> 68

15 <211> 354

<212> PRT

<213> *Microscilla* sp.

20

<400> 68

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

Ser Gln Val Pro Gln Gly Thr Pro Lys Pro Pro Asp Ser Val Leu Gln  
115 120 125

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

Asp Tyr Ser Gln Trp Trp Glu Trp Thr Ile Gly Ala Asn Trp Arg Asn  
145 150 155 160

Pro Glu Gly Pro Gly Ser Thr Ile Glu Asp Arg Met Asp His Pro Val  
165 170 175

Val His Val Ser Phe Glu Asp Val Gln Ala Tyr Ala Asp Trp Ala Gly  
180 185 190

Lys Arg Leu Pro Thr Glu Ala Glu Trp Glu Trp Ala Ala Met Gly Gly  
195 200 205

Gln Asn Asp Val Lys Tyr Pro Trp Gly Asn Glu Ser Val Glu Gln Ala  
210 215 220

Ser Asp Lys Ala Asn Phe Trp Gln Gly Asn Phe Pro His Gln Asn Tyr  
225 230 235 240

Ala Leu Asp Gly Phe Glu Arg Thr Ala Pro Val Arg Ser Phe Pro Ala  
245 250 255

Asn Gly Tyr Gly Leu Tyr Asp Met Ala Gly Asn Val Trp Glu Trp Cys  
260 265 270

Gln Asp Lys Tyr Asp Val Asn Ala Tyr Glu Ser Tyr Lys Gln Lys Gly  
275 280 285

Leu Thr Glu Asp Pro Thr Gly Ser Glu His Tyr Asn Asp Pro Arg Glu  
290 295 300

Pro Tyr Thr Pro Lys His Val Ile Arg Gly Gly Ser Phe Leu Cys Asn  
305 310 315 320

Asp Ser Tyr Cys Ser Gly Tyr Arg Val Ser Arg Arg Met Ser Ser Ser  
325 330 335

Arg Asp Ser Gly Phe Asn His Thr Gly Phe Arg Cys Val Lys Asp Val  
340 345 350

Asn Gly

<210> 69

5 <211> 876

<212> ADN

<213> *Pseudomonas putida* KT2440

<400> 69

5 atggtgcacg tgccgggcg cgagttcagc tttggttcaa gccgctttta cgacgaagaa 60  
ggcccgctc accccgcaa ggtgtccggc ttctggattg acgtgcatcc ggtcaccaac 120  
gccagttcg cgcgcttcgt caaggccacg gggatatgtca cccatgccga gcgcggtacc 180  
cgtgtcgagg acgacctgc cctgcccgc gcgctgcgga taccgggtgc gatggtgttt 240  
catcaggggtg cggacgtgct cggccccggc tggcagttcg tgcccggcgc caactggcga 300  
caccgcgaag ggccgggcag cagcctggcc gggctggaca accatccggt ggtgcagatc 360  
gccctggaag atgcccaggc ctatgccgc tgggcaggcc gcgaactgcc cagcgaggcg 420  
cagctggaat acgccatgcg cggcggcctg accgatgccg acttcagctg ggggtaccacc 480  
gagcagccca agggcaagct catggccaat acctggcagg gtcagttccc ttatcgcaat 540  
gcggcgaagg atggttttac cgggtacatcg ccctggggtt gcttcccggc caacggcttt 600  
ggcctgttcg atgccggcg caatgtctgg gagctgactc gcacgggcta tcggccaggc 660  
catgacgcac agcgcgacgc caagctcgac ccctcaggcc cggccctgag tgacagcttc 720  
gaccggcgag accccggcgt gccgggtggcg gtaatcaaag gcggctcgca cctgtgttcg 780  
gcggaccgct gcatgcgcta ccgcccctcg gcacgccagc cgcagccggt gttcatgacg 840  
acctcgcacg tgggtttcag aacgattcgg caatga 876

<210> 70

10

<211> 291

<212> PRT

15 <213> *Pseudomonas putida* KT2440

<400> 70

# ES 2 380 147 T3

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



Gly Cys Phe Pro Ala Asn Gly Phe Gly Leu Phe Asp Ala Gly Gly Asn  
195 200 205

Val Trp Glu Leu Thr Arg Thr Gly Tyr Arg Pro Gly His Asp Ala Gln  
210 215 220

Arg Asp Ala Lys Leu Asp Pro Ser Gly Pro Ala Leu Ser Asp Ser Phe  
225 230 235 240

Asp Pro Ala Asp Pro Gly Val Pro Val Ala Val Ile Lys Gly Gly Ser  
245 250 255

His Leu Cys Ser Ala Asp Arg Cys Met Arg Tyr Arg Pro Ser Ala Arg  
260 265 270

Gln Pro Gln Pro Val Phe Met Thr Thr Ser His Val Gly Phe Arg Thr  
275 280 285

Ile Arg Gln  
290

<210> 71

5 <211> 780

<212> ADN

<213> *Ralstonia metallidurans*

10

<400> 71

atgggtcgcg	gcgggatggt	gttcgtcggc	accaacagcc	cggtgccgct	gcgcgaatac	60
tggcgctggt	ggcgcttcgt	acctggcgcg	gactggcgtc	acccgaccgg	cccgggcagt	120
tccatcgaag	gcaaggacaa	tcattccgct	gtgcaggtct	cgtatgaaga	cgcgcaggcg	180
tacgccaagt	gggccggcaa	gcgtctgccc	accgaggccg	agtgggagtt	tgccgcccgt	240
ggcgccctgg	agcaggccac	ctacgcctgg	ggtgacaagt	tcgcgccgga	tggccggcag	300
atggcgcaatg	tctggcaggg	ccagcaggtg	cagccgttcc	cggtgggtcag	cgccaaggcg	360
ggcgccgcgg	ctggcaccag	tgctgtcggc	acgttcccgg	gcaatggcta	tgggctctat	420
gacatgaccg	gcaacgcctg	gcagtgggtg	gccgactggt	atcgegcgga	ccagttccgc	480
cggaagcca	cggtggcggc	agtgtgtcag	aatccgaccg	gcccgccgga	ttcgtgggac	540
ccgaccgaac	ctggcgtgcc	ggtgtcggcg	cccaagcggg	tcacgcgcgg	tggctcgttc	600
ctctgcaacg	aggacttctg	cctcagctac	cgcccagagt	cccggcgcg	taccgaccgc	660
tacaccagca	tgctgcacct	aggcttccgg	ctcgtgatgg	atgacgccc	ttgggcagaa	720
gttcgcaagc	agccagccgt	ggcaatggcc	gcgggcgggc	agcagaacgt	gcagaaataa	780

15 <210> 72

<211> 259

<212> PRT

<213> *Ralstonia metallidurans*

5 <400> 72

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

Leu Arg Glu Tyr Trp Arg Trp Trp Arg Phe Val Pro Gly Ala Asp Trp  
20 25 30

Arg His Pro Thr Gly Pro Gly Ser Ser Ile Glu Gly Lys Asp Asn His

# ES 2 380 147 T3

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

Val Gln Lys

<210> 73

5 <211> 876

<212> ADN

<213> *Prochlorococcus marinus*

10

<400> 73

gtgaccacat ctttgccagt agagatggta accatccccg cagggctcta tcgagttggc	60
tgtgatcgct gctatccgga tggttcagtt cgctgctatc cggaggaaac acccgcgca	120
gaagtgcagc ttgactcatt ccagatcgac gtagggccag tcaccaatgc ccagttccga	180
gctttcgtta gcgccacgca gcatctcaca gtctcggagc taccacctga tccaacgctc	240
tatccccgatc tagcgccga ggaacgcac cctgaatcag ttgtctttca accgcctcca	300
gcaacgggtgg atcgagcaa acccttgagc tgggtggacc tcatggctgg ggctgattgg	360
cgatcatcccc aaggaccga aagcacgac gatggccttg atgatcacc tgcgtgcat	420
gtcgctatg cgcagccat cgcctatgcc cattgggctg gcaagcgtct cccctctgct	480
gaagagtggg aagtagccgc ccgcgggggt cttgtcgatg cccaatagc ctgggggaat	540
gaactcactc ccaataaccg ctggatggcg aacatctggc aaggtccttt cccttggcac	600
aacgaggagc tagacggctg gttctggacc tcgcccgttg gcagctttcc tgccaacggc	660
tatggactct tggatgtttg cggcaatgtg tgggaatgga ccaactctgt ttatcccggtg	720
gcgtcaggcc accaggaacg gcgaactatc aaaggcggat cgtttctctg cgcagataat	780
tactgcgtac gttatcgacc ctctgcaact caaggccaga cagtagacac tgccacctgt	840
cacatgggct ttcgctgtgc aaaaggagg ccttga	876

5 <210> 74

<211> 291

<212> PRT

10

<213> *Prochlorococcus marinus*

<400> 74

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

210

215

220

Asp Val Cys Gly Asn Val Trp Glu Trp Thr Asn Ser Val Tyr Pro Val  
225 230 235 240

Ala Ser Gly His Gln Glu Arg Arg Thr Ile Lys Gly Gly Ser Phe Leu  
245 250 255

Cys Ala Asp Asn Tyr Cys Val Arg Tyr Arg Pro Ser Ala Leu Gln Gly  
260 265 270

Gln Thr Val Asp Thr Ala Thr Cys His Met Gly Phe Arg Cys Ala Lys  
275 280 285

Gly Gly Pro  
290

<210> 75

5 <211> 1017

<212> ADN

<213> *Caulobacter crescentus* CB15

10

<400> 75

```

ttgggaaaac tgacggcgct tcccgctctg atgcttctgg cgctggccgg ctgcgggccag    60
ccggcgccca aggcttgect ggcggaacct cgggttccag atccccagaa ccgcacggcg    120
ggtatggttc ggctggcggg cggcgacttc cagatgggcg ctgcgccgct gcgtccggag    180
gagggaccgc cccagacggt cacggtcccc ccgttctgga tcgatcagac agaggtcacc    240
aacgccgect tcgcgcggtt cgtcgagggc acgggttate gcaccgtggc cgagcgaccg    300
ctcgaccccc cgcgctacgc ccacgtaccg gcggcgagc ggcgctccggc ctcgctcgtc    360
ttcgtggggg cgaagggggc gaggtcggac gatccttccc aatggtggca ggtgatcccc    420
ggcgccgact ggcgggcatcc cgaaggtccc ggctcgaaca tccggggcag ggacgcctgg    480
ccggtggtgc atatecgctg ggaggacgcc atggcctacg cccgctggct gggccgtgac    540
ctgcccacag aggccgaatg ggagtaagcc gcgcgcggcg ggctggttgg caagcgctac    600
acctggggcg accaggetca ggatcctgca aagccgcgcg ccaatacttg gcaaggcgtg    660
ttcccgggcc aggaccttgg caatgacggc ttcaaggcca agcccgcgcc ggtcggctgc    720
ttcccgccca acggctatgg cctgcgcgac atggccggca atgtctggga gtggaccgcg    780
gactggttca agccgggcct ggatccggtc agcgtcctcg aaaccggcgg gccgcccag    840
gcccgcgcgc tggatcccgga ggacccgaac acgcccgaag acgtcgtgaa gggcggttcg    900
ttcctgtgcg ccgacgacta ctgcttccgc tategacctg cggcgcggaac gccggggccg    960
ccggacagcg gcgcatcgca tgtcggtttc cgcaccgtgc tccgcgccga gcgctga    1017

```

<210> 76

5 <211> 338

<212> PRT

<213> *Caulobacter crescentus* CB15

10

<400> 76

```

Met Gly Lys Leu Thr Ala Leu Pro Val Leu Met Leu Leu Ala Leu Ala
1           5           10           15

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

```

Pro Asp Pro Gln Asn Arg Thr Ala Gly Met Val Arg Leu Ala Gly Gly  
35 40 45

Asp Phe Gln Met Gly Ala Ala Pro Leu Arg Pro Glu Glu Gly Pro Pro  
50 55 60

Gln Thr Val Thr Val Pro Pro Phe Trp Ile Asp Gln Thr Glu Val Thr  
65 70 75 80

Asn Ala Ala Phe Ala Arg Phe Val Glu Ala Thr Gly Tyr Arg Thr Val  
85 90 95

Ala Glu Arg Pro Leu Asp Pro Ala Arg Tyr Ala His Val Pro Ala Ala  
100 105 110

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

Ser Asp Asp Pro Ser Gln Trp Trp Gln Val Ile Pro Gly Ala Asp Trp  
130 135 140

Arg His Pro Glu Gly Pro Gly Ser Asn Ile Arg Gly Arg Asp Ala Trp  
145 150 155 160

Pro Val Val His Ile Ala Trp Glu Asp Ala Met Ala Tyr Ala Arg Trp  
165 170 175

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

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

Pro Ala Lys Pro Arg Ala Asn Thr Trp Gln Gly Val Phe Pro Ala Gln  
210 215 220

Asp Leu Gly Asn Asp Gly Phe Lys Ala Lys Pro Ala Pro Val Gly Cys  
225 230 235 240

Phe Pro Pro Asn Gly Tyr Gly Leu Arg Asp Met Ala Gly Asn Val Trp  
245 250 255

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

Leu Glu Thr Gly Gly Pro Pro Glu Ala Arg Ala Leu Asp Pro Glu Asp  
275 280 285

Pro Asn Thr Pro Lys His Val Val Lys Gly Gly Ser Phe Leu Cys Ala  
290 295 300

Asp Asp Tyr Cys Phe Arg Tyr Arg Pro Ala Ala Arg Thr Pro Gly Pro  
305 310 315 320

Pro Asp Ser Gly Ala Ser His Val Gly Phe Arg Thr Val Leu Arg Ala  
325 330 335

Glu Arg



<210> 77

<211> 900

5 <212> ADN

<213> *Mycobacterium tuberculosis* H37Rv

<400> 77

10  
 gtgctgaccg agttggttga cctgcccggc ggatcggttc gcatgggctc gacgcgcttc 60  
 tacccegaag aagcgccgat tcataccgtg accgtgcgcg cctttgcggt agagcgacac 120  
 ccggtgacca acgcgcaatt tgccgaattc gtctccgcga caggctatgt gacggttgca 180  
 gaacaacccc ttgaccccg gctctaccca ggagtggacg cagcagacct gtgtcccggt 240  
 gcgatgggtgt tttgtccgac ggccggggccg gtcgacctgc gtgactggcg gcaatgggtgg 300  
 gactgggtac ctggcgccctg ctggcgccat ccgtttggcc gggacagcga tatcgccgac 360  
 cgagccggcc acccggtcgt acagggtggcc tatccggacg ccgtggccta cgcacgatgg 420  
 gctgggtcgac gcctaccgac cgaggccgag tgggagtacg cggcccggtg cggaaccacg 480  
 gcaacctatg cgtggggcga ccaggagaag ccggggggcga tgctcatggc gaacacctgg 540  
 cagggccggt ttccttaccg caacgacggt gcattgggct ggggtgggaac ctccccggtg 600  
 ggcaggtttc cggccaacgg gtttggttg ctcgacatga tcggaaacgt ttgggagtgg 660  
 accaccaccg agttctatcc acaccatcgc atcgatccac cctcgacggc ctgctgcgca 720  
 ccggtcaagc tcgctacagc cgccgaccg acgatcagcc agaccctcaa gggcggctcg 780  
 cacctgtgcg cgccggagta ctgccaccgc taccgcccgg cggcgcgctc gccgcagtcg 840  
 caggacaccg cgaccacca tatcggttcc cgggtgcgtgg ccgaccgggt gtccgggtag 900

<210> 78

15 <211> 299

<212> PRT

<213> *Mycobacterium tuberculosis* H37Rv

20

<400> 78

ES 2 380 147 T3

Met Leu Thr Glu Leu Val Asp Leu Pro Gly Gly Ser Phe Arg Met Gly  
1 5 10 15

Ser Thr Arg Phe Tyr Pro Glu Glu Ala Pro Ile His Thr Val Thr Val  
20 25 30

Arg Ala Phe Ala Val Glu Arg His Pro Val Thr Asn Ala Gln Phe Ala  
35 40 45

Glu Phe Val Ser Ala Thr Gly Tyr Val Thr Val Ala Glu Gln Pro Leu  
50 55 60

Asp Pro Gly Leu Tyr Pro Gly Val Asp Ala Ala Asp Leu Cys Pro Gly  
65 70 75 80

Ala Met Val Phe Cys Pro Thr Ala Gly Pro Val Asp Leu Arg Asp Trp  
85 90 95

Arg Gln Trp Trp Asp Trp Val Pro Gly Ala Cys Trp Arg His Pro Phe  
100 105 110

Gly Arg Asp Ser Asp Ile Ala Asp Arg Ala Gly His Pro Val Val Gln  
115 120 125

Val Ala Tyr Pro Asp Ala Val Ala Tyr Ala Arg Trp Ala Gly Arg Arg  
130 135 140

Leu Pro Thr Glu Ala Glu Trp Glu Tyr Ala Ala Arg Gly Gly Thr Thr  
145 150 155 160

Ala Thr Tyr Ala Trp Gly Asp Gln Glu Lys Pro Gly Gly Met Leu Met  
165 170 175

Ala Asn Thr Trp Gln Gly Arg Phe Pro Tyr Arg Asn Asp Gly Ala Leu  
180 185 190

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

Gly Leu Leu Asp Met Ile Gly Asn Val Trp Glu Trp Thr Thr Thr Glu  
210 215 220

Phe Tyr Pro His His Arg Ile Asp Pro Pro Ser Thr Ala Cys Cys Ala  
225 230 235 240

Pro Val Lys Leu Ala Thr Ala Ala Asp Pro Thr Ile Ser Gln Thr Leu  
245 250 255

Lys Gly Gly Ser His Leu Cys Ala Pro Glu Tyr Cys His Arg Tyr Arg  
260 265 270

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

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

<210> 79  
 <211> 7  
 5 <212> PRT  
 <213> Artificial  
 <220>  
 10 <223> dominio conservado en procariotas y procariotas  
 <220>  
 15 <221> DOMINIO  
 <222> (1)..(7)  
 <223> dominio conservado  
 20 <220>  
 <221> MISC\_FEATURE  
 25 <222> (3)..(4)  
 <223> cualquier aminoácido  
 <220>  
 30 <221> MISC\_FEATURE  
 <222> (6)..(6)  
 35 <223> cualquier aminoácido  
 <220>  
 <221> MISC\_FEATURE  
 40 <222> (6)..(6)  
 <223> Gly o Ala  
 45 <400> 79  
**Arg Val Xaa Xaa Gly Xaa Ser**  
**1 5**  
 <210> 80  
 50 <211> 630  
 <212> ADN  
 55 <213> *Oncorhynchus mykiss*  
 <400> 80

tcagggtggct gctgccccct ggtggttgcc tgtcagagga gcagactgga ggcaccctga 60  
 gggcccccagac tccagcatca cagacaggct ggaccaccct gtgctgcatg tgtcatggca 120  
 ggacgctgtg gcctactgct cctgggccta caagagacta cccacagagg ctgagtggga 180  
 gtacgcctgc agagggggcc tacaggagag actttaccg tgggggaaca aactgaaacc 240  
 taaaggacag cactacgcca acctctggca gggaaagtcc cccacacaca actcagaaga 300  
 ggacgggtac actaaaacct caccagtga gtcatttcc gcaaattggct atggcctgta 360  
 caacatggta gggaaatgcat gggagtggac atctgactgg tggactgtac accacaccac 420  
 agatgaacag cacaacccgg cagggtccacc atcaggcaca gaccgagtga agaaaggagg 480  
 ctctacatg tgccataagt catactgtta caggtacagg tgtgcagcac ggagtcagaa 540  
 caccctgac agctctgct ctaacctagg gtcccgctgt gtctcccagg agcagccgta 600  
 acctttcacc ctcgacctg acatgggtag 630

<210> 81

5 <211> 655

<212> ADN

<213> *Danio rerio*

10 <220>

<221> misc\_feature

15 <222> (590)..(590)

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

<220>

20 <221> misc\_feature

<222> (626)..(626)

25 <223> n es a, c, g o t

<400> 81

caaatggttt tatttacata aaaaaatcct cttagtttga agtgtaagac agtgagatta 60  
 gtgatgtttg aggttatgga tcaacatcag aggcgcagcg gaagcccaag ttcgaggctg 120  
 aactgtccgg tgtgtttctga ctgcgagcgg cacacctgta tctgtagcag taagacttgt 180  
 ggcacatgta ggatcctcct ttcttgactc tgtctgtccc tgattctggg ccctttgggt 240  
 taaacttgtc ttctgcagtg tgatgcacag tccaccagtc tgccgtccac tcccacgcat 300  
 tccccacat gtcatacagg ccaaagccat tgggaggaaa agacatcacc ggggatgtgt 360  
 tggcatagcc gtcctctgca gtgttgat tagggaaatc tccctgccac aggttagcat 420  
 agtctgccc tcttggcatt aatttatttc cccatgggta catcctgtcc tgtagtcctc 480  
 ctctacaggg caactcccat tcagcttctg taggaagtct gcgtttggcc cattgacagt 540  
 acgcccgtgc atcatcccat gaaacatgca gagcagggtg attcattctn gtgtgtatgg 600  
 ttgaatctgg tcctttctgg tgtctncagt ctgcaccttt cactggtgac cacca 655

<210> 82

<211> 773

5 <212> ADN

<213> *Oryzias latipes*

<220>

10 <221> misc\_feature

<222> (690)..(690)

15 <223> n es a, c, g o t

<220>

<221> misc\_feature

20 <222> (755)..(755)

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

25 <400> 82

tctccttttt	tccataaata	acattagagt	ccttacattc	tgcctttaca	tacattgtca	60
gagacagtac	aaaaaatctg	cctttgtaaa	attagagtta	caaaaatata	ttttagattt	120
gacttcttca	gaattgtcgg	tggcagcaaa	agaatcggat	tgatctcatg	acaagagcgt	180
gagccagaag	ttcttggatc	aaactgattt	ggttctgtca	tcgtttctgt	tcagcagcac	240
agcgaaaacc	aagattggaa	gcggagctgt	ctggagtgtt	ttggcttcga	gcagcacatc	300
tgtacctgta	acaataagac	ttgtggcaca	tgtacgagcc	tcctttcttc	accttatctg	360
tgcctgacgg	aggacccggt	gggttgtgct	gatgggtctgt	tgtgtggtgc	acgctccacc	420
agtctgaggt	ccactcccat	gcgttcccca	ccatgtcata	cagaccaaaa	gcattgcctg	480
ggaaggacat	caccggggag	gttttagtgt	agccatcctc	tgcagagttg	tgtgctggga	540
attccccctg	ccagaggttg	gcgtaatgct	gtccctttgg	gtttagcttg	tttccccagg	600
ggtagagtct	gtccttcagg	cgcgccctgc	aggcaacctc	ccactctgcc	tcagtgggaa	660
gtctcttggt	gacccaggag	cagtaagccn	aggcatcatt	cccagaaacc	tgaacgacgg	720
atgatccatc	ctgtctgtga	tgttggagtc	tggancttca	gggtgcttcc	agt	773

<210> 83

30 <211> 566

<212> ADN

35 <213> *Xenopus laevis*

<220>

<221> misc\_feature

40 <222> (6)..(6)

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

<220>

5 <221> misc\_feature

<222> (47)..(47)

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

10 <220>

<221> misc\_feature

15 <222> (81)..(81)

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

<400> 83

20  
**atatgnaact aaaggtaatg taattggaat gatggatttc acaaggnetg agagttccct 60**  
**attgctcctg cttgtcgtgt nacaggtcac ggagccggcg ccacacagcg aaatcccagg 120**  
**ttggaggccg agctgtcggg tgtattctga cttogagcag cacagcgata cctgtagcaa 180**  
**taggactcat ggcacatgta ggagcctcct ttcttcactc tatcatttcc cgtagaaggt 240**  
**cctttcgggt tgtgaacctc atctgctgta tgatgagtgt cccaccaatc agatgtccac 300**  
**tcccaagcat ttcccaccat gttatataga ccataaccat tggctgggaa agcagttaca 360**  
**ggtgaagtct gcacataacc atcctctcca gtgttttggg ttggaaaatc cccctgccag 420**  
**acattcgcat aatgttgtcc ctttggttcc agcttggtcc cccatggaaa aatcctgttc 480**  
**tcaagtcccc cgcggcaggc gtattcccac tcagcttcag ttggaaggcg tttacctgcc 540**  
**cagggtgcaga aagcagaagc atcatt 566**

<210> 84

25 <211> 647

<212> ADN

<213> *Silurana tropicalis*

30 <400> 84

gccgctttttt tttttttttt tttttttttt catcacaaaa ataattttat taataaaaata 60  
 ggatttttgtg ttcattctta ttatgaagga caaggaatgt cattgaaatt tttgttttca 120  
 caaggctcttg ggagttcctt cctgctcagg tcatttttgca gtggtcacgg agccgacgcc 180  
 acgcagcgga atcccagggt agaggccgag ctgtcagggtg tattctgact tcgagcagca 240  
 cagcgataacc tgtagcagta ggactcatgg cacatgtatg agcctccttt tttcaccttg 300  
 tcttttcccg taaaaggacc tttcgggttg taagtctcat ctgctgtatg atgagtgtcc 360  
 caccaatcgg atgtccactc ccaagcattt cccaccatgt tatataggct ataaccattg 420  
 gctgggaaag cggttacagg tgaagtctgc acatagccgt cctctccagt gttttgggtt 480  
 ggaaattccc cctgccagac attcgcataa tgttctccct ttgggtccag cttgttcccc 540  
 cacggaaaaa gcctgttctc aagtccccca cgggagggcat attcccactc agcttctgtc 600  
 ggaaggcgct taccgcacca ggtgcagaag gcagaagcat cgttcca 647

<210> 85

5 <211> 636

<212> ADN

<213> *Salmo salar*

10 <400> 85

atagacattt tttaaatatt ttacaacaaa atatattcca taaatatcca catgtcatgc 60  
 ggtaatcctg catttcatga agaacactga catcactggc tgtatgaaga ggtgcacttg 120  
 atttgtttcg cctggcgggc aagataggca gagttagcac cctagactag agccaatggc 180  
 gaatggtaca aaaagggaaa agtcagacta cccatgtcag ggtcaagggt aaaaggttac 240  
 ggctgctcct gggagacaca gcggaacctt aggttagagg cagagctgtc aggggtgttc 300  
 tgactccttg ctgcacacct gtacctgtaa cagtatgact tatggcacat gtaggagcct 360  
 cttttcttca ctgggtctgt gcctgatggg ggacctgccg ggttgtgccg ttcactctgtg 420  
 gtgtggtgta cagtccacca gtcagatgtc cactcccatg cattccctac catgttgtac 480  
 aggccatagc catttgcagg aaatgacttc actgggtgagg ttttggtgta cccgtcctct 540  
 tctgagttgt gtgtggggaa ctttccctgc cagagggttg cgtagtgtg tcccttaggt 600  
 ttcagtttgt tccccacgg gtaaagtctg tcctgt 636

15 <210> 86

<211> 415

<212> ADN

20 <213> *Sus scrofa*

<400> 86

agtttctctgt gaccaacacc ggagaggatg gcttccgagg aactgcgcct gttgatgcct 60  
 ttctctccaa tggttatggc ctttacaata tagtagggaa cgcttgggaa tggacctcag 120  
 actggtggac cattcaccat gctgctgaag aaacaattaa cccatcaagt tcttctctgct 180  
 gcaccgaata acagagccgc cactacgtga tgaaagcaga gaaaggcccc ccttctggga 240  
 aagaccgggt gaagaaaggg ggatcctata tgtgccataa gtcctactgc tacaggtacc 300  
 gctgtgctgc tcgaagccag aacacgccgg acagctcggc ttcaaactctg gggttccgct 360  
 gtgcagctga ccaccagccc accacagget gagtcaggaa gagtcttccc gaatc 415

5 <210> 87

<211> 595

<212> ADN

10

<213> *Bos taurus*

<400> 87

ccacgcgtcc gggggcaaca aactgcagcc gaaaggccag cattatagcc aacatcttgg 60  
 caaggcgagt ttctgtgac caacaccggg gaggacggct tccgaggagc cgcgctgttt 120  
 gacgcctttc ctcccaatgg ttattggctt atacaatata gtagggaacg cctgggagtg 180  
 gacttcagac tgggtggactg ttcaccattc tgctgaagaa acgattaacc caaaaggccc 240  
 cccttctggg aaagaccggg tgaagaaagg tggatcctac atgtgccata aatcctattg 300  
 ctacaggtat cgctgtgctg ctgaagcca gaacacaccc gacagctctg cttcgaatct 360  
 gggattccgt tgtgcagctg accacctgcc caccacaggc taagagccaa aaagagcctt 420  
 cccgaaccgc agaagtcgtg tctactctgc acgcggttcc cctcagaagg ctgaacaacc 480  
 tgctgtgaag aattcccacc ccaaggtggg ttacatacct tgcccagtgg ccaaaggacc 540  
 tatggcaaga ccaaattgct gagctgatca gcatgtgcgc tttattgggg gatgg 595

15

<210> 88

<211> 1611

20

<212> ADN

<213> *Homo sapiens*

25

<220>

<221> CDS

<222> (1)..(1608)

30

<223> hSULF3

<400> 88



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

tgg aca aga gat gtt gct ttc tta ctc aga caa gaa ggc agg ccc atg Trp Thr Arg Asp Val Ala Phe Leu Leu Arg Gln Glu Gly Arg Pro Met 145 150 155 160	480
gtt aat ctt atc cgt aac agg act aaa gtc aga gtg atg gaa agg gat Val Asn Leu Ile Arg Asn Arg Thr Lys Val Arg Val Met Glu Arg Asp 165 170 175	528
tgg cag aat aca gac aaa gca gta aac tgg tta aga aag gaa gca att Trp Gln Asn Thr Asp Lys Ala Val Asn Trp Leu Arg Lys Glu Ala Ile 180 185 190	576
aat tac act gaa cca ttt gtt att tac ttg gga tta aat tta cca cac Asn Tyr Thr Glu Pro Phe Val Ile Tyr Leu Gly Leu Asn Leu Pro His 195 200 205	624
cct tac cct tca cca tct tct gga gaa aat ttt gga tct tca aca ttt Pro Tyr Pro Ser Pro Ser Ser Gly Glu Asn Phe Gly Ser Ser Thr Phe 210 215 220	672
cac aca tct ctt tat tgg ctt gaa aaa gtg tct cat gat gcc atc aaa His Thr Ser Leu Tyr Trp Leu Glu Lys Val Ser His Asp Ala Ile Lys 225 230 235 240	720
atc cca aag tgg tca cct ttg tca gaa atg cac cct gta gat tat tac Ile Pro Lys Trp Ser Pro Leu Ser Glu Met His Pro Val Asp Tyr Tyr 245 250 255	768
tct tct tat aca aaa aac tgc act gga aga ttt aca aaa gaa att Ser Ser Tyr Thr Lys Asn Cys Thr Gly Arg Phe Thr Lys Glu Ile 260 265 270	816
aag aat att aga gca ttt tat tat gct atg tgt gct gag aca gat gcc Lys Asn Ile Arg Ala Phe Tyr Tyr Ala Met Cys Ala Glu Thr Asp Ala 275 280 285	864
atg ctt ggt gaa att att ttg gcc ctt cat caa tta gat ctt ctt cag Met Leu Gly Glu Ile Ile Leu Ala Leu His Gln Leu Asp Leu Leu Gln 290 295 300	912
aaa act att gtc ata tac tcc tca gac cat gga gag ctg gcc atg gaa Lys Thr Ile Val Ile Tyr Ser Ser Asp His Gly Glu Leu Ala Met Glu 305 310 315 320	960
cat cga cag ttt tat aaa atg agc atg tac gag gct agt gca cat gtt His Arg Gln Phe Tyr Lys Met Ser Met Tyr Glu Ala Ser Ala His Val 325 330 335	1008
ccg ctt ttg atg atg gga cca gga att aaa gcc ggc cta caa gta tca Pro Leu Leu Met Met Gly Pro Gly Ile Lys Ala Gly Leu Gln Val Ser 340 345 350	1056
aat gtg gtt tct ctt gtg gat att tac cct acc atg ctt gat att gct Asn Val Val Ser Leu Val Asp Ile Tyr Pro Thr Met Leu Asp Ile Ala 355 360 365	1104
gga att cct ctg cct cag aac ctg agt gga tac tct ttg ttg ccg tta Gly Ile Pro Leu Pro Gln Asn Leu Ser Gly Tyr Ser Leu Leu Pro Leu 370 375 380	1152
tca tca gaa aca ttt aag aat gaa cat aaa gtc aaa aac ctg cat cca Ser Ser Glu Thr Phe Lys Asn Glu His Lys Val Lys Asn Leu His Pro 385 390 395 400	1200
ccc tgg att ctg agt gaa ttc cat gga tgt aat gtg aat gcc tcc acc Pro Trp Ile Leu Ser Glu Phe His Gly Cys Asn Val Asn Ala Ser Thr 405 410 415	1248
tac atg ctt cga act aac cac tgg aaa tat ata gcc tat tcg gat ggt Tyr Met Leu Arg Thr Asn His Trp Lys Tyr Ile Ala Tyr Ser Asp Gly 420 425 430	1296
gca tca ata ttg cct caa ctc ttt gat ctt tcc tcg gat cca gat gaa Ala Ser Ile Leu Pro Gln Leu Phe Asp Leu Ser Ser Asp Pro Asp Glu 435 440 445	1344
tta aca aat gtt gct gta aaa ttt cca gaa att act tat tct ttg gat Leu Thr Asn Val Ala Val Lys Phe Pro Glu Ile Thr Tyr Ser Leu Asp 450 455 460	1392

# ES 2 380 147 T3

450	455	460	
cag aag ctt cat tcc att ata aac tac cct aaa gtt tct gct tct gtc Gln Lys Leu His Ser Ile Ile Asn Tyr Pro Lys Val Ser Ala Ser Val 465 470 475 480			1440
cac cag tat aat aaa gag cag ttt atc aag tgg aaa caa agt ata gga His Gln Tyr Asn Lys Glu Gln Phe Ile Lys Trp Lys Gln Ser Ile Gly 485 490 495			1488
cag aat tat tca aac gtt ata gca aat ctt agg tgg cac caa gac tgg Gln Asn Tyr Ser Asn Val Ile Ala Asn Leu Arg Trp His Gln Asp Trp 500 505 510			1536
cag aag gaa cca agg aag tat gaa aat gca att gat cag tgg ctt aaa Gln Lys Glu Pro Arg Lys Tyr Glu Asn Ala Ile Asp Gln Trp Leu Lys 515 520 525			1584
acc cat atg aat cca aga gca gtt tga Thr His Met Asn Pro Arg Ala Val 530 535			1611

<210> 89

5 <211> 536

<212> PRT

<213> *Homo sapiens*

10

<400> 89

# ES 2 380 147 T3

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

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

Gln Lys Glu Pro Arg Lys Tyr Glu Asn Ala Ile Asp Gln Trp Leu Lys  
515 520 525

Thr His Met Asn Pro Arg Ala Val  
530 535

<210> 90

5 <211> 1722

<212> ADN

<213> *Homo sapiens*

10 <220>

<221> CDS

15 <222> (1)..(1719)

<223> hSULF4

<400> 90

20

ES 2 380 147 T3

atg ggg gcg ctg gca gga ttc tgg atc ctc tgc ctc ctc act tat ggt Met Gly Ala Leu Ala Gly Phe Trp Ile Leu Cys Leu Leu Thr Tyr Gly 1 5 10 15	48
tac ctg tcc tgg ggc cag gcc tta gaa gag gag gaa gaa ggg gcc tta Tyr Leu Ser Trp Gly Gln Ala Leu Glu Glu Glu Glu Gly Ala Leu 20 25 30	96
cta gct caa gct gga gag aaa cta gag ccc agc aca act tcc acc tcc Leu Ala Gln Ala Gly Glu Lys Leu Glu Pro Ser Thr Thr Ser Thr Ser 35 40 45	144
cag ccc cat ctc att ttc atc cta gcg gat gat cag gga ttt aga gat Gln Pro His Leu Ile Phe Ile Leu Ala Asp Asp Gln Gly Phe Arg Asp 50 55 60	192
gtg ggt tac cac gga tct gag att aaa aca cct act ctt gac aag ctc Val Gly Tyr His Gly Ser Glu Ile Lys Thr Pro Thr Leu Asp Lys Leu 65 70 75 80	240
gct gcc gaa gga gtt aaa ctg gag aac tac tat gtc cag cct att tgc Ala Ala Glu Gly Val Lys Leu Glu Asn Tyr Tyr Val Gln Pro Ile Cys 85 90 95	288
aca cca tcc agg agt cag ttt att act gga aag tat cag ata cac acc Thr Pro Ser Arg Ser Gln Phe Ile Thr Gly Lys Tyr Gln Ile His Thr 100 105 110	336
gga ctt caa cat tct atc ata aga cct acc caa ccc aac tgt tta cct Gly Leu Gln His Ser Ile Ile Arg Pro Thr Gln Pro Asn Cys Leu Pro 115 120 125	384
ctg gac aat gcc acc cta cct cag aaa ctg aag gag gtt gga tat tca Leu Asp Asn Ala Thr Leu Pro Gln Lys Leu Lys Glu Val Gly Tyr Ser 130 135 140	432
acg cat atg gtc gga aaa tgg cac ttg ggt ttt tac aga aaa gaa tgc Thr His Met Val Gly Lys Trp His Leu Gly Phe Tyr Arg Lys Glu Cys 145 150 155 160	480
atg ccc acc aga aga gga ttt gat acc ttt ttt ggt tcc ctt ttg gga Met Pro Thr Arg Arg Gly Phe Asp Thr Phe Phe Gly Ser Leu Leu Gly 165 170 175	528
agt ggg gat tac tat aca cac tac aaa tgt gac agt cct ggg atg tgt Ser Gly Asp Tyr Tyr Thr His Tyr Lys Cys Asp Ser Pro Gly Met Cys 180 185 190	576
ggc tat gac ttg tat gaa aac gac aat gct gcc tgg gac tat gac aat Gly Tyr Asp Leu Tyr Glu Asn Asp Asn Ala Ala Trp Asp Tyr Asp Asn 195 200 205	624
ggc ata tac tcc aca cag atg tac act cag aga gta cag caa atc tta Gly Ile Tyr Ser Thr Gln Met Tyr Thr Gln Arg Val Gln Gln Ile Leu 210 215 220	672

gct tcc cat aac ccc aca aag cct ata ttt tta tat att gcc tat caa Ala Ser His Asn Pro Thr Lys Pro Ile Phe Leu Tyr Ile Ala Tyr Gln 225 230 235 240	720
gct gtt cat tca cca ctg caa gct cct ggc agg tat ttc gaa cac tac Ala Val His Ser Pro Leu Gln Ala Pro Gly Arg Tyr Phe Glu His Tyr 245 250 255	768
cga tcc att atc aac ata aac agg agg aga tat gct gcc atg ctt tcc Arg Ser Ile Ile Asn Ile Asn Arg Arg Tyr Ala Ala Met Leu Ser 260 265 270	816
tgc tta gat gaa gca atc aac aac gtg aca ttg gct cta aag act tat Cys Leu Asp Glu Ala Ile Asn Asn Val Thr Leu Ala Leu Lys Thr Tyr 275 280 285	864
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
cag cct acg gca gga ggg agt aac tgg cct ctc aga ggt agc aaa gga Gln Pro Thr Ala Gly Gly Ser Asn Trp Pro Leu Arg Gly Ser Lys Gly 305 310 315 320	960
aca tat tgg gaa gga ggg atc cgg gct gta ggc ttt gtg cat agc cca Thr Tyr Trp Glu Gly Gly Ile Arg Ala Val Gly Phe Val His Ser Pro 325 330 335	1008
ctt ctg aaa aac aag gga aca gtg tgt aag gaa ctt gtg cac atc act Leu Leu Lys Asn Lys Gly Thr Val Cys Lys Glu Leu Val His Ile Thr 340 345 350	1056
gac tgg tac ccc act ctc att tca ctg gct gaa gga cag att gat gag Asp Trp Tyr Pro Thr Leu Ile Ser Leu Ala Glu Gly Gln Ile Asp Glu 355 360 365	1104
gac att caa cta gat ggc tat gat atc tgg gag acc ata agt gag ggt Asp Ile Gln Leu Asp Gly Tyr Asp Ile Trp Glu Thr Ile Ser Glu Gly 370 375 380	1152
ctt cgc tca ccc cga gta gat att ttg cat aac att gac ccc ata tac Leu Arg Ser Pro Arg Val Asp Ile Leu His Asn Ile Asp Pro Ile Tyr 385 390 395 400	1200
acc aag gca aaa aat ggc tcc tgg gca gca ggc tat ggg atc tgg aac Thr Lys Ala Lys Asn Gly Ser Trp Ala Ala Gly Tyr Gly Ile Trp Asn 405 410 415	1248
act gca atc cag tca gcc atc aga gtg cag cac tgg aaa ttg ctt aca Thr Ala Ile Gln Ser Ala Ile Arg Val Gln His Trp Lys Leu Leu Thr 420 425 430	1296
gga aat cct ggc tac agc gac tgg gtc ccc cct cag tct ttc agc aac Gly Asn Pro Gly Tyr Ser Asp Trp Val Pro Pro Gln Ser Phe Ser Asn 435 440 445	1344
ctg gga ccg aac cgg tgg cac aat gaa cgg atc acc ttg tca act ggc Leu Gly Pro Asn Arg Trp His Asn Glu Arg Ile Thr Leu Ser Thr Gly 450 455 460	1392
aaa agt gta tgg ctt ttc aac atc aca gcc gac cca tat gag agg gtg Lys Ser Val Trp Leu Phe Asn Ile Thr Ala Asp Pro Tyr Glu Arg Val 465 470 475 480	1440
gac cta tct aac agg tat cca gga atc gtg aag aag ctc cta cgg agg Asp Leu Ser Asn Arg Tyr Pro Gly Ile Val Lys Lys Leu Leu Arg Arg 485 490 495	1488
ctc tca cag ttc aac aaa act gca gtg ccg gtc agg tat ccc ccc aaa Leu Ser Gln Phe Asn Lys Thr Ala Val Pro Val Arg Tyr Pro Pro Lys 500 505 510	1536
gac ccc aga agt aac cct agg ctc aat gga ggg gtc tgg gga cca tgg Asp Pro Arg Ser Asn Pro Arg Leu Asn Gly Gly Val Trp Gly Pro Trp 515 520 525	1584
tat aaa gag gaa acc aag aaa aag aag cca agc aaa aat cag gct gag Tyr Lys Glu Glu Thr Lys Lys Lys Lys Pro Ser Lys Asn Gln Ala Glu 530 535 540 545	1632



# ES 2 380 147 T3

530	535	540	
aaa aag caa aag aaa agc	aaa aaa aag aag aag	aaa cag cag aaa gca	1680
Lys Lys Gln Lys Lys Ser	Lys Lys Lys Lys Lys	Lys Gln Gln Lys Ala	
545	550	555	560
gtc tca ggt tca act tgc cat tca ggt gtt act tgt gga taa			1722
Val Ser Gly Ser Thr Cys His Ser Gly Val Thr Cys Gly			
	565	570	

<210> 91

5 <211> 573

<212> PRT

<213> *Homo sapiens*

10 <400> 91

# ES 2 380 147 T3

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

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

Val Ser Gly Ser Thr Cys His Ser Gly Val Thr Cys Gly  
565 570

<210> 92

5 <211> 1710

<212> ADN

<213> *Homo sapiens*

10 <220>

<221> CDS

15 <222> (1)..(1707)

<223> hSULF5

20 <400> 92

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
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
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
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
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
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
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
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
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
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
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
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
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
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
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

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 245 250 255	768
aat gtg gcc cgg cgg aag tac gcg gcc atg gtg acc tgc atg gat gag Asn Val Ala Arg Arg Lys Tyr Ala Ala Met Val Thr Cys Met Asp Glu 260 265 270	816
gct gtg cgc aac atc acc tgg gcc ctc aag cgc tac ggt ttc tac aac Ala Val Arg Asn Ile Thr Trp Ala Leu Lys Arg Tyr Gly Phe Tyr Asn 275 280 285	864
aac agt gtc atc atc ttc tcc agt gac aat ggt ggc cag act ttc tcg Asn Ser Val Ile Ile Phe Ser Ser Asp Asn Gly Gly Gln Thr Phe Ser 290 295 300	912
ggg ggc agc aac tgg ccg ctc cga gga cgc aag ggc act tat tgg gaa Gly Gly Ser Asn Trp Pro Leu Arg Gly Arg Lys Gly Thr Tyr Trp Glu 305 310 315 320	960
ggt ggc gtg cgg ggc cta ggc ttt gtc cac agt ccc ctg ctc aag cga Gly Gly Val Arg Gly Leu Gly Phe Val His Ser Pro Leu Lys Arg 325 330 335	1008
aag caa cgg aca agc cgg gca ctg atg cac atc act gac tgg tac ccg Lys Gln Arg Thr Ser Arg Ala Leu Met His Ile Thr Asp Trp Tyr Pro 340 345 350	1056
acc ctg gtg ggt ctg gca ggt ggt acc acc tca gca gcc gat ggg cta Thr Leu Val Gly Leu Ala Gly Thr Thr Ser Ala Asp Gly Leu 355 360 365	1104
gat ggc tac gac gtg tgg ccg gcc atc agc gag ggc cgg gcc tca cca Asp Gly Tyr Asp Val Trp Pro Ala Ile Ser Glu Gly Arg Ala Ser Pro 370 375 380	1152
cgc acg gag atc ctg cac aac att gac cca ctc tac aac cat gcc cag Arg Thr Glu Ile Leu His Asn Ile Asp Pro Leu Tyr Asn His Ala Gln 385 390 395 400	1200
cat ggc tcc ctg gag ggc ggc ttt ggc atc tgg aac acc gcc gtg cag His Gly Ser Leu Glu Gly Gly Phe Gly Ile Trp Asn Thr Ala Val Gln 405 410 415	1248
gct gcc atc cgc gtg ggt gag tgg aag ctg ctg aca gga gac ccc ggc Ala Ala Ile Arg Val Gly Glu Trp Lys Leu Leu Thr Gly Asp Pro Gly 420 425 430	1296
tat ggc gat tgg atc cca ccg cag aca ctg gcc acc ttc ccg ggt agc Tyr Gly Asp Trp Ile Pro Pro Gln Thr Leu Ala Thr Phe Pro Gly Ser 435 440 445	1344
tgg tgg aac ctg gaa cga atg gcc agt gtc cgc cag gcc gtg tgg ctc Trp Trp Asn Leu Glu Arg Met Ala Ser Val Arg Gln Ala Val Trp Leu 450 455 460	1392
ttc aac atc agt gct gac cct tat gaa cgg gag gac ctg gct ggc cag Phe Asn Ile Ser Ala Asp Pro Tyr Glu Arg Glu Asp Leu Ala Gly Gln 465 470 475 480	1440
cgg cct gat gtg gtc cgc acc ctg ctg gct cgc ctg gcc gaa tat aac Arg Pro Asp Val Val Arg Thr Leu Leu Ala Arg Leu Ala Glu Tyr Asn 485 490 495	1488
cgc aca gcc atc ccg gta cgc tac cca gct gag aac ccc cgg gct cat Arg Thr Ala Ile Pro Val Arg Tyr Pro Ala Glu Asn Pro Arg Ala His 500 505 510	1536
cct gac ttt aat ggg ggt gct tgg ggg ccc tgg gcc agt gat gag gaa Pro Asp Phe Asn Gly Gly Ala Trp Gly Pro Trp Ala Ser Asp Glu Glu 515 520 525	1584
gag gag gaa gag gaa ggg agg gct cga agc ttc tcc cgg ggt cgt cgc Glu Glu Glu Glu Glu Gly Arg Ala Arg Ser Phe Ser Arg Gly Arg Arg 530 535 540	1632
aag aaa aaa tgc aag att tgc aag ctt cga tcc ttt ttc cgt aaa ctc Lys Lys Lys Cys Lys Ile Cys Lys Leu Arg Ser Phe Phe Arg Lys Leu	1680

# ES 2 380 147 T3

545

550

555

560

aac acc agg cta atg tcc caa cgg atc tga  
 Asn Thr Arg Leu Met Ser Gln Arg Ile  
 565

1710

<210> 93

5 <211> 569

<212> PRT

<213> *Homo sapiens*

10  
 <400> 93

ES 2 380 147 T3

Met His Thr Leu Thr Gly Phe Ser Leu Val Ser Leu Leu Ser Phe Gly  
1 5 10 15

Tyr Leu Ser Trp Asp Trp Ala Lys Pro Ser Phe Val Ala Asp Gly Pro  
20 25 30

Gly Glu Ala Gly Glu Gln Pro Ser Ala Ala Pro Pro Gln Pro Pro His  
35 40 45

Ile Ile Phe Ile Leu Thr Asp Asp Gln Gly Tyr His Asp Val Gly Tyr  
50 55 60

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

Gly Val Lys Leu Glu Asn Tyr Tyr Ile Gln Pro Ile Cys Thr Pro Ser  
85 90 95

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

His Ser Ile Ile Arg Pro Gln Gln Pro Asn Cys Leu Pro Leu Asp Gln  
115 120 125

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

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

Arg Arg Gly Phe Asp Thr Phe Leu Gly Ser Leu Thr Gly Asn Val Asp  
165 170 175

Tyr Tyr Thr Tyr Asp Asn Cys Asp Gly Pro Gly Val Cys Gly Phe Asp  
180 185 190

Leu His Glu Gly Glu Asn Val Ala Trp Gly Leu Ser Gly Gln Tyr Ser  
195 200 205

Thr Met Leu Tyr Ala Gln Arg Ala Ser His Ile Leu Ala Ser His Ser  
210 215 220

Pro Gln Arg Pro Leu Phe Leu Tyr Val Ala Phe Gln Ala Val His Thr  
225 230 235 240

Pro Leu Gln Ser Pro Arg Glu Tyr Leu Tyr Arg Tyr Arg Thr Met Gly  
245 250 255



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

<210> 94

<211> 2067

5 <212> ADN

<213> *Homo sapiens*

<220>

10 <221> CDS

<222> (1)..(2064)

15 <223> hSULF6

<400> 94

atg cta att tca gga aga gaa gag aac caa ata gac ata tcc aag acc Met Leu Ile Ser Gly Arg Glu Glu Asn Gln Ile Asp Ile Ser Lys Thr 1 5 10 15	48
aca gag gta gat tgt ttt gtg gtt gaa tta gga agt cta cac aat cct Thr Glu Val Asp Cys Phe Val Val Glu Leu Gly Ser Leu His Asn Pro 20 25 30	96
aca cgg aac cca cag cga att ttc acc aag cac gtg gcc acc aag tca Thr Arg Asn Pro Gln Arg Ile Phe Thr Lys His Val Ala Thr Lys Ser 35 40 45	144
tcc agc tcc aaa tgt cag ctg gac caa ggt gga aaa agc ctg gtc cag Ser Ser Ser Lys Cys Gln Leu Asp Gln Gly Gly Lys Ser Leu Val Gln 50 55 60	192
tgc att tta ccc aga tct tca aag ctc ctc tca ccc ttg tgt ctc ccc Cys Ile Leu Pro Arg Ser Ser Lys Leu Leu Ser Pro Leu Cys Leu Pro 65 70 75 80	240
cat ccg tgt gga gct tta ctt ctg tat aga tcc tca gga atc gcc tct His Pro Cys Gly Ala Leu Leu Leu Tyr Arg Ser Ser Gly Ile Ala Ser 85 90 95	288
gct ctt gct gcc ttt aca gac tcc ctc tct agg agc tgc tgg ctg tca Ala Leu Ala Ala Phe Thr Asp Ser Leu Ser Arg Ser Cys Trp Leu Ser 100 105 110	336
gtg tcc ctg tgc tgt ttg ttt tgc ggt gtt gat ggc aca ttt atg aca Val Ser Leu Cys Cys Leu Phe Cys Gly Val Asp Gly Thr Phe Met Thr 115 120 125	384
aga aac gcc aga ccc aac att gtc ctg ctg atg gca gat gac ctt gga Arg Asn Ala Arg Pro Asn Ile Val Leu Leu Met Ala Asp Asp Leu Gly 130 135 140	432
gtg ggg gat ttg tgc tgc tac ggt aat aac tca gtg agc aca cct aat Val Gly Asp Leu Cys Cys Tyr Gly Asn Asn Ser Val Ser Thr Pro Asn 145 150 155 160	480
att gac cgc ctg gca agt gaa gga gtg agg ctt acc cag cat ctc gca Ile Asp Arg Leu Ala Ser Glu Gly Val Arg Leu Thr Gln His Leu Ala 165 170 175	528
gct gct tcc atg tgc acc cca agt cgg gct gcc ttc ctg acc ggc cgg Ala Ala Ser Met Cys Thr Pro Ser Arg Ala Ala Phe Leu Thr Gly Arg 180 185 190	576
tac ccc atc aga tca ggg atg gtg tct gcc tac aac ctg aac cgt gcc Tyr Pro Ile Arg Ser Gly Met Val Ser Ala Tyr Asn Leu Asn Arg Ala 195 200 205	624
ttc acg tgg ctt ggt ggg tca ggt ggt ctt ccc acc aat gaa acg act Phe Thr Trp Leu Gly Gly Ser Gly Gly Leu Pro Thr Asn Glu Thr Thr 210 215 220	672
ttt gcc aag ctg ctg cag cac cgt ggc tac cgc acg gga ctc ata ggc Phe Ala Lys Leu Leu Gln His Arg Gly Tyr Arg Thr Gly Leu Ile Gly 225 230 235 240	720
aaa tgg cac ctg ggt ttg agc tgc gcc tct cgg aat gat cac tgt tac Lys Trp His Leu Gly Leu Ser Cys Ala Ser Arg Asn Asp His Cys Tyr 245 250 255	768

cac ccg ctc aac cat ggt ttt cac tac ttt tac ggg gtg cct ttt gga His Pro Leu Asn His Gly Phe His Tyr Phe Tyr Gly Val Pro Phe Gly 260 265 270	816
ctt tta agc gac tgc cag gca tcc aag aca cca gaa ctg cac cgc tgg Leu Leu Ser Asp Cys Gln Ala Ser Lys Thr Pro Glu Leu His Arg Trp 275 280 285	864
ctc agg atc aaa ctg tgg atc tcc acg gta gcc ctt gcc ctg gtt cct Leu Arg Ile Lys Leu Trp Ile Ser Thr Val Ala Leu Ala Leu Val Pro 290 295 300	912
ttt ctg ctt ctc att ccc aag ttc gcc cgc tgg ttc tca gtg cca tgg Phe Leu Leu Leu Ile Pro Lys Phe Ala Arg Trp Phe Ser Val Pro Trp 305 310 315 320	960
aag gtc atc ttt gtc ttt gct ctc ctc gcc ttt ctg ttt ttc act tcc Lys Val Ile Phe Val Phe Ala Leu Leu Ala Phe Leu Phe Phe Thr Ser 325 330 335	1008
tgg tac tct agt tat gga ttt act cga cgt tgg aat tgc atc ctt atg Trp Tyr Ser Ser Tyr Gly Phe Thr Arg Arg Trp Asn Cys Ile Leu Met 340 345 350	1056
agg aac cat gaa att atc cag cag cca atg aaa gag gag aaa gta gct Arg Asn His Glu Ile Ile Gln Gln Pro Met Lys Glu Glu Lys Val Ala 355 360 365	1104
tcc ctc atg ctg aag gag gca ctt gct ttc att gaa agg tac aaa agg Ser Leu Met Leu Lys Glu Ala Leu Ala Phe Ile Glu Arg Tyr Lys Arg 370 375 380	1152
gaa cct ttt ctc ctc ttt ttt tcc ttc ctg cac gta cat act cca ctc Glu Pro Phe Leu Leu Phe Phe Ser Phe Leu His Val His Thr Pro Leu 385 390 395 400	1200
atc tcc aaa aag aag ttt gtt ggg cgc agt aaa tat ggc agg tat ggg Ile Ser Lys Lys Lys Phe Val Gly Arg Ser Lys Tyr Gly Arg Tyr Gly 405 410 415	1248
gac aat gta gaa gaa atg gat tgg atg gtg ggt aaa atc ctg gat gcc Asp Asn Val Glu Glu Met Asp Trp Met Val Gly Lys Ile Leu Asp Ala 420 425 430	1296
ctg gac cag gag cgc ctg gcc aac cac acc ttg gtg tac ttc acc tct Leu Asp Gln Glu Arg Leu Ala Asn His Thr Leu Val Tyr Phe Thr Ser 435 440 445	1344
gac aac ggg ggc cac ctg gag ccc ctg gac ggg gct gtt cag ctg ggt Asp Asn Gly Gly His Leu Glu Pro Leu Asp Gly Ala Val Gln Leu Gly 450 455 460	1392
ggc tgg aac ggg atc tac aaa ggt ggc aaa gga atg gga gga tgg gaa Gly Trp Asn Gly Ile Tyr Lys Gly Gly Lys Gly Met Gly Gly Trp Glu 465 470 475 480	1440
gga ggt atc cgt gtg cca ggg ata ttc cgg tgg ccg tca gtc ttg gag Gly Gly Ile Arg Val Pro Gly Ile Phe Arg Trp Pro Ser Val Leu Glu 485 490 495	1488
gct ggg aga gtg atc aat gag ccc acc agc tta atg gac atc tat ccg Ala Gly Arg Val Ile Asn Glu Pro Thr Ser Leu Met Asp Ile Tyr Pro 500 505 510	1536
acg ctg tct tat ata ggc gga ggg atc ttg tcc cag gac aga gtg att Thr Leu Ser Tyr Ile Gly Gly Gly Ile Leu Ser Gln Asp Arg Val Ile 515 520 525	1584
gac ggc cag aac cta atg ccc ctg ctg gaa gga agg gcg tcc cac tcc Asp Gly Gln Asn Leu Met Pro Leu Leu Glu Gly Arg Ala Ser His Ser 530 535 540	1632
gac cac gag ttc ctc ttc cac tac tgt ggg gtc tat ctg cac acg gtc Asp His Glu Phe Leu Phe His Tyr Cys Gly Val Tyr Leu His Thr Val 545 550 555 560	1680
agg tgg cat cag aag gac tgt gca act gtg tgg aaa gct cat tat gtg Arg Trp His Gln Lys Asp Cys Ala Thr Val Trp Lys Ala His Tyr Val	1728

	565	570	575	
	act cct aaa ttc tac cct gaa gga	aca ggt gcc tgc tat ggg agt gga	1776	
	Thr Pro Lys Phe Tyr Pro Glu Gly	Thr Gly Ala Cys Tyr Gly Ser Gly		
	580	585	590	
	ata tgt tca tgt tcg ggg gat gta	acc tac cac gac cca cca ctc ctc	1824	
	Ile Cys Ser Cys Ser Gly Asp Val	Thr Tyr His Asp Pro Pro Leu Leu		
	595	600	605	
	ttt gac atc tca aga gac cct tca	gaa gcc ctt cca ctg aac cct gac	1872	
	Phe Asp Ile Ser Arg Asp Pro Ser	Glu Ala Leu Pro Leu Asn Pro Asp		
	610	615	620	
	aat gag cca tta ttt gac tcc gtg	atc aaa aag atg gag gca gcc ata	1920	
	Asn Glu Pro Leu Phe Asp Ser Val	Ile Lys Lys Met Glu Ala Ala Ile		
	625	630	635	640
	aga gag cat cgt agg aca cta aca	cct gtc cca cag cag ttc tct gtg	1968	
	Arg Glu His Arg Arg Thr Leu Thr	Pro Val Pro Gln Gln Phe Ser Val		
	645	650	655	
	ttc aac aca att tgg aaa cca tgg	ctg cag cct tgc tgt ggg acc ttc	2016	
	Phe Asn Thr Ile Trp Lys Pro Trp	Leu Gln Pro Cys Cys Gly Thr Phe		
	660	665	670	
	ccc ttc tgt ggg tgt gac aag gaa	gat gac atc ctt ccc atg gct ccc	2064	
	Pro Phe Cys Gly Cys Asp Lys Glu	Asp Asp Ile Leu Pro Met Ala Pro		
	675	680	685	
	tga			2067

<210> 95

5 <211> 688

<212> PRT

<213> *Homo sapiens*

10 <400> 95

Met Leu Ile Ser Gly Arg Glu Glu Asn Gln Ile Asp Ile Ser Lys Thr  
1 5 10 15

Thr Glu Val Asp Cys Phe Val Val Glu Leu Gly Ser Leu His Asn Pro  
20 25 30

Thr Arg Asn Pro Gln Arg Ile Phe Thr Lys His Val Ala Thr Lys Ser  
35 40 45

Ser Ser Ser Lys Cys Gln Leu Asp Gln Gly Gly Lys Ser Leu Val Gln  
50 55 60

Cys Ile Leu Pro Arg Ser Ser Lys Leu Leu Ser Pro Leu Cys Leu Pro  
65 70 75 80

His Pro Cys Gly Ala Leu Leu Leu Tyr Arg Ser Ser Gly Ile Ala Ser  
85 90 95

Ala Leu Ala Ala Phe Thr Asp Ser Leu Ser Arg Ser Cys Trp Leu Ser  
100 105 110

Val Ser Leu Cys Cys Leu Phe Cys Gly Val Asp Gly Thr Phe Met Thr  
115 120 125

Arg Asn Ala Arg Pro Asn Ile Val Leu Leu Met Ala Asp Asp Leu Gly  
130 135 140

Val Gly Asp Leu Cys Cys Tyr Gly Asn Asn Ser Val Ser Thr Pro Asn  
145 150 155 160

Ile Asp Arg Leu Ala Ser Glu Gly Val Arg Leu Thr Gln His Leu Ala  
165 170 175

Ala Ala Ser Met Cys Thr Pro Ser Arg Ala Ala Phe Leu Thr Gly Arg  
180 185 190

Tyr Pro Ile Arg Ser Gly Met Val Ser Ala Tyr Asn Leu Asn Arg Ala  
195 200 205

Phe Thr Trp Leu Gly Gly Ser Gly Gly Leu Pro Thr Asn Glu Thr Thr  
210 215 220

Phe Ala Lys Leu Leu Gln His Arg Gly Tyr Arg Thr Gly Leu Ile Gly  
225 230 235 240

Lys Trp His Leu Gly Leu Ser Cys Ala Ser Arg Asn Asp His Cys Tyr  
245 250 255

His Pro Leu Asn His Gly Phe His Tyr Phe Tyr Gly Val Pro Phe Gly  
260 265 270

Leu Leu Ser Asp Cys Gln Ala Ser Lys Thr Pro Glu Leu His Arg Trp  
275 280 285

Leu Arg Ile Lys Leu Trp Ile Ser Thr Val Ala Leu Ala Leu Val Pro  
290 295 300

Phe Leu Leu Leu Ile Pro Lys Phe Ala Arg Trp Phe Ser Val Pro Trp  
305 310 315 320

Lys Val Ile Phe Val Phe Ala Leu Leu Ala Phe Leu Phe Phe Thr Ser  
325 330 335

Trp Tyr Ser Ser Tyr Gly Phe Thr Arg Arg Trp Asn Cys Ile Leu Met  
340 345 350

Arg Asn His Glu Ile Ile Gln Gln Pro Met Lys Glu Glu Lys Val Ala  
355 360 365

Ser Leu Met Leu Lys Glu Ala Leu Ala Phe Ile Glu Arg Tyr Lys Arg  
370 375 380

Glu Pro Phe Leu Leu Phe Phe Ser Phe Leu His Val His Thr Pro Leu  
385 390 395 400

Ile Ser Lys Lys Lys Phe Val Gly Arg Ser Lys Tyr Gly Arg Tyr Gly  
405 410 415

Asp Asn Val Glu Glu Met Asp Trp Met Val Gly Lys Ile Leu Asp Ala  
420 425 430

Leu Asp Gln Glu Arg Leu Ala Asn His Thr Leu Val Tyr Phe Thr Ser  
435 440 445

Asp Asn Gly Gly His Leu Glu Pro Leu Asp Gly Ala Val Gln Leu Gly  
450 455 460

Gly Trp Asn Gly Ile Tyr Lys Gly Gly Lys Gly Met Gly Gly Trp Glu  
 465 470 475 480  
 Gly Gly Ile Arg Val Pro Gly Ile Phe Arg Trp Pro Ser Val Leu Glu  
 485 490 495  
 Ala Gly Arg Val Ile Asn Glu Pro Thr Ser Leu Met Asp Ile Tyr Pro ,  
 500 505 510  
 Thr Leu Ser Tyr Ile Gly Gly Gly Ile Leu Ser Gln Asp Arg Val Ile  
 515 520 525  
 Asp Gly Gln Asn Leu Met Pro Leu Leu Glu Gly Arg Ala Ser His Ser  
 530 535 540  
 Asp His Glu Phe Leu Phe His Tyr Cys Gly Val Tyr Leu His Thr Val  
 545 550 555 560  
 Arg Trp His Gln Lys Asp Cys Ala Thr Val Trp Lys Ala His Tyr Val  
 565 570 575  
 Thr Pro Lys Phe Tyr Pro Glu Gly Thr Gly Ala Cys Tyr Gly Ser Gly  
 580 585 590  
 Ile Cys Ser Cys Ser Gly Asp Val Thr Tyr His Asp Pro Pro Leu Leu  
 595 600 605  
 Phe Asp Ile Ser Arg Asp Pro Ser Glu Ala Leu Pro Leu Asn Pro Asp  
 610 615 620  
 Asn Glu Pro Leu Phe Asp Ser Val Ile Lys Lys Met Glu Ala Ala Ile  
 625 630 635 640  
 Arg Glu His Arg Arg Thr Leu Thr Pro Val Pro Gln Gln Phe Ser Val  
 645 650 655  
 Phe Asn Thr Ile Trp Lys Pro Trp Leu Gln Pro Cys Cys Gly Thr Phe  
 660 665 670  
 Pro Phe Cys Gly Cys Asp Lys Glu Asp Asp Ile Leu Pro Met Ala Pro  
 675 680 685



## REIVINDICACIONES

1. Método para determinar el nivel de expresión de enzima generadora de C<sub>α</sub>-formilglicina (FGE) en un sujeto, comprendiendo el método medir *in vitro* en una muestra obtenida de un sujeto la actividad de generación de C<sub>α</sub>-formilglicina de un péptido o polipéptido que tiene una secuencia seleccionada del grupo que consiste en: aminoácidos 34-374 de SEQ ID NO: 2, SEQ ID NO: 2, 5, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76 y 78.
2. Método para identificar un agente útil en la modulación de la actividad de generación de C<sub>α</sub>-formilglicina, que comprende:
  - (a) poner en contacto una molécula que tiene actividad de generación de C<sub>α</sub>-formilglicina o una molécula de ácido nucleico que codifica para una molécula con dicha actividad con un agente candidato,
  - (b) medir la actividad de generación de C<sub>α</sub>-formilglicina de la molécula, y
  - (c) comparar la actividad de generación de C<sub>α</sub>-formilglicina medida de la molécula con un control para determinar si el agente candidato modula la actividad de generación de C<sub>α</sub>-formilglicina de la molécula;
- en el que la molécula es una molécula de ácido nucleico que tiene la secuencia de nucleótidos expuesta como SEQ ID NO: 1, o es un producto de expresión de la misma.
3. Método para determinar una deficiencia de sulfatasa en un sujeto caracterizada por una expresión aberrante de una molécula de ácido nucleico o un producto de expresión de la misma, que comprende:
  - monitorizar una muestra de un paciente para detectar un parámetro seleccionado del grupo que consiste en:
    - (i) una molécula de ácido nucleico que tiene la secuencia de nucleótidos expuesta como SEQ ID NO: 1,
    - (ii) un polipéptido codificado por la molécula de ácido nucleico,
    - (iii) un péptido derivado del polipéptido, y
    - (iv) un anticuerpo que se une selectivamente al polipéptido o péptido,
  - como determinación de una deficiencia de sulfatasa en el sujeto.
4. Método según la reivindicación 3, en el que la muestra es un líquido biológico o un tejido.
5. Método según la reivindicación 3, en el que la etapa de monitorizar comprende poner en contacto la muestra con un agente detectable seleccionado del grupo que consiste en
  - (a) una molécula de ácido nucleico que se hibrida selectivamente en condiciones rigurosas con la molécula de ácido nucleico de (i),
  - (b) un anticuerpo que se une selectivamente al polipéptido de (ii), o el péptido de (iii), y
  - (c) un polipéptido o péptido que se une al anticuerpo de (iv).
6. Método según la reivindicación 5, en el que el anticuerpo, el polipéptido, el péptido o el ácido nucleico se marca con un marcador radioactivo o una enzima.
7. Método según la reivindicación 3, que comprende someter a ensayo la muestra para detectar el péptido.
8. Kit, que comprende un envase que contiene un agente y un control; en el que el agente es un anticuerpo o fragmento del mismo que se une selectivamente a un péptido o polipéptido según la reivindicación 1, o es un ácido nucleico que se une selectivamente a un ácido nucleico que codifica para dicho péptido o polipéptido; y en el que el control es para comparar un valor medido de unión de dicho agente a dicho péptido, polipéptido o ácido nucleico.
9. Kit según la reivindicación 8; en el que el control es un valor predeterminado para compararlo con el valor medido.
10. Kit según la reivindicación 8; en el que el control comprende un epítipo del péptido o polipéptido.
11. Uso de una molécula de ácido nucleico, péptido o polipéptido que modula la actividad de generación de C<sub>α</sub>-formilglicina en la preparación de un medicamento para tratar una deficiencia de sulfatasa en un sujeto; en el que:

(i) la molécula de ácido nucleico se selecciona del grupo que consiste en:

(a) una molécula de ácido nucleico de hibridación que se hibrida en condiciones rigurosas con una molécula que consiste en una secuencia de nucleótidos expuesta como SEQ ID NO: 1, o un complemento de dicha molécula de hibridación que codifica para un polipéptido que tiene actividad FGE;

(b) una molécula de ácido nucleico que se diferencia de una molécula de ácido nucleico de (a) debido a la degeneración del código genético; y

(c) un ácido nucleico que tiene una secuencia seleccionada del grupo que consiste en: SEQ ID NO: 1, 3, 4, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77 y 80-87; y

(ii) el péptido o el polipéptido se selecciona del grupo que consiste en:

(a) un péptido o un polipéptido codificado por una molécula de ácido nucleico tal como se describió anteriormente; y

(b) un péptido o un polipéptido que tiene una secuencia seleccionada del grupo que consiste en: aminoácidos 34-374 de SEQ ID NO: 2, SEQ ID NO: 2, 5, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76 y 78.

12. Uso según la reivindicación 11; en el que el medicamento comprende además un agente seleccionado del grupo que consiste en: una molécula de ácido nucleico que codifica para 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; un producto de expresión de la molécula de ácido nucleico; y un fragmento del producto de expresión de la molécula de ácido nucleico.

13. Uso según la reivindicación 11; en el que el agente que modula la actividad de generación de  $C_{\alpha}$ -formilglicina se produce mediante una célula que expresa dicha molécula de ácido nucleico.

14. Uso según la reivindicación 13; en el que la célula expresa una molécula de ácido nucleico de FGE exógena.

15. Uso según la reivindicación 13, en el que la célula expresa una molécula de ácido nucleico de FGE endógena.

16. Molécula de ácido nucleico, péptido o polipéptido que modula la actividad de generación de  $C_{\alpha}$ -formilglicina; para su uso en el tratamiento de una deficiencia de sulfatasa en un sujeto; siendo la molécula de ácido nucleico, péptido o polipéptido según la reivindicación 11.

17. Composición farmacéutica que comprende una molécula de ácido nucleico, péptido, o polipéptido según la reivindicación 11, y un vehículo farmacéuticamente aceptable.

18. Método para identificar un agente candidato útil en el tratamiento de una deficiencia de sulfatasa, que comprende:

determinar la expresión de un conjunto de moléculas de ácido nucleico *in vitro* en una célula o tejido en condiciones que, en ausencia de un agente candidato, permiten una primera cantidad de expresión del conjunto de moléculas de ácido nucleico; en el que el conjunto de moléculas de ácido nucleico comprende al menos una molécula de ácido nucleico que es una molécula de ácido nucleico según la reivindicación 11;

poner en contacto la célula o tejido con el agente candidato *in vitro*, y

detectar una cantidad de prueba de expresión del conjunto de moléculas de ácido nucleico;

en el que un aumento en la cantidad de prueba de expresión en presencia del agente candidato con respecto a la primera cantidad de expresión indica que el agente candidato es útil en el tratamiento de la deficiencia de sulfatasa.

19. Matriz de moléculas de ácido nucleico en fase sólida que consiste en un conjunto de moléculas de ácido nucleico, productos de expresión de las mismas, o fragmentos de los mismos, cuando se fija a un sustrato sólido; codificando cada molécula de ácido nucleico para un polipéptido seleccionado del grupo que consiste en aminoácidos 34-374 de SEQ ID NO: 2, SEQ ID NO: 2, 5, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76 y 78, 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.

20. Matriz de moléculas de ácido nucleico en fase sólida según la reivindicación 19, que comprende además al menos una molécula de ácido nucleico de control.

21. Matriz de moléculas de ácido nucleico en fase sólida según la reivindicación 19; en la que el conjunto de moléculas de ácido nucleico comprende al menos una molécula de ácido nucleico que codifica para un polipéptido seleccionado del grupo que consiste en: aminoácidos 34-374 de SEQ ID NO: 2, SEQ ID NO: 2, 5, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76 y 78, 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.
22. Matriz de moléculas de ácido nucleico en fase sólida según la reivindicación 19; en la que el conjunto de moléculas de ácido nucleico comprende al menos dos moléculas de ácido nucleico, codificando cada molécula de ácido nucleico para un polipéptido seleccionado del grupo que consiste en: aminoácidos 34-374 de SEQ ID NO: 2, SEQ ID NO: 2, 5, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76 y 78, 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.
23. Alelo variante aislado de un gen de FGE humana que codifica para un polipéptido de FGE variante, comprendiendo el polipéptido: una secuencia de aminoácidos que comprende al menos una variación en la SEQ ID NO: 2, en el que la al menos una variación comprende: Met1Arg; Met1Val; Ser155Pro; Cys218Tyr; Ala279Val; Arg327Stop; Cys336Arg; Arg345Cys; Arg349Trp; Arg349Gln; Ser359Stop; o una combinación de las mismas.
24. Polipéptido de FGE humana variante aislado que comprende una secuencia de aminoácidos con al menos una variación en la SEQ ID NO: 2, en el que la al menos una variación comprende: Met1Arg; Met1Val; Ser155Pro; Cys218Tyr; Ala279Val; Arg327Stop; Cys336Arg; Arg345Cys; Arg349Trp; Arg349Gln; Ser359Stop; o a combinación de las mismas.
25. Anticuerpo que tiene el polipéptido de FGE humana variante según la reivindicación 23 como inmunógeno.
26. Anticuerpo según la reivindicación 25, que es un anticuerpo policlonal.
27. Anticuerpo según la reivindicación 25, que es un anticuerpo monoclonal.
28. Anticuerpo según la reivindicación 25, que es un anticuerpo quimérico.
29. Anticuerpo según la reivindicación 25, marcado de manera detectable.
30. Anticuerpo según la reivindicación 29, en el que dicho marcador detectable comprende un elemento radioactivo, un compuesto químico que fluoresce o una enzima.
31. Método de producción de una sulfatasa activada, comprendiendo el método:
- (a) poner en contacto la sulfatasa con una enzima generadora de  $C_{\alpha}$ -formilglicina (FGE) *in vitro*; en el que la enzima generadora de  $C_{\alpha}$ -formilglicina es un polipéptido o péptido según la reivindicación 1; o
- (b) proporcionar una célula que coexpresa una sulfatasa y una enzima generadora de  $C_{\alpha}$ -formilglicina (FGE) de modo que se produce sulfatasa activada dentro de la célula; en el que la célula comprende ARN o ADN heterólogo que da como resultado una expresión aumentada de la sulfatasa activada, con respecto a lo que se produciría en ausencia del ARN o ADN heterólogo; y en el que la enzima generadora de  $C_{\alpha}$ -formilglicina es un polipéptido o péptido según la reivindicación 1.
32. Método según la reivindicación 31, en el que la sulfatasa se selecciona 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.
33. Método según cualquiera de las reivindicaciones 3 a 7 ó 18, o uso según cualquiera de las reivindicaciones 11 a 16; en el que la deficiencia de sulfatasa a la que se hace referencia con respecto a dicho método o uso se selecciona del grupo que consiste en: deficiencia múltiple de sulfatasas, mucopolisacaridosis II, mucopolisacaridosis IIIA, mucopolisacaridosis IVA, mucopolisacaridosis VI, mucopolisacaridosis VIII, leucodistrofia metacromática, condrodisplasia punctata I recesiva ligada al cromosoma X e ictiosis ligada al cromosoma X.

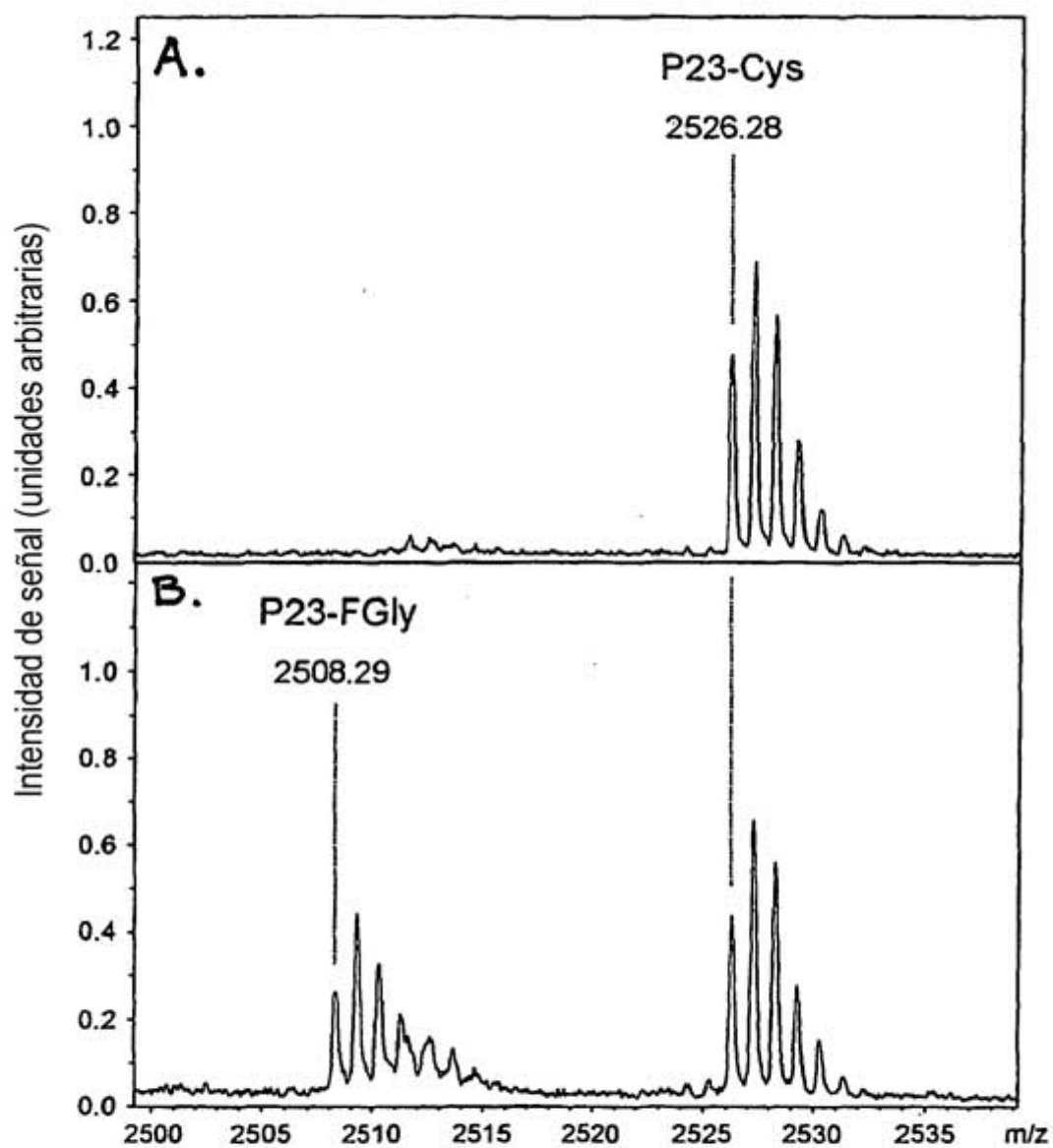


Fig. 1

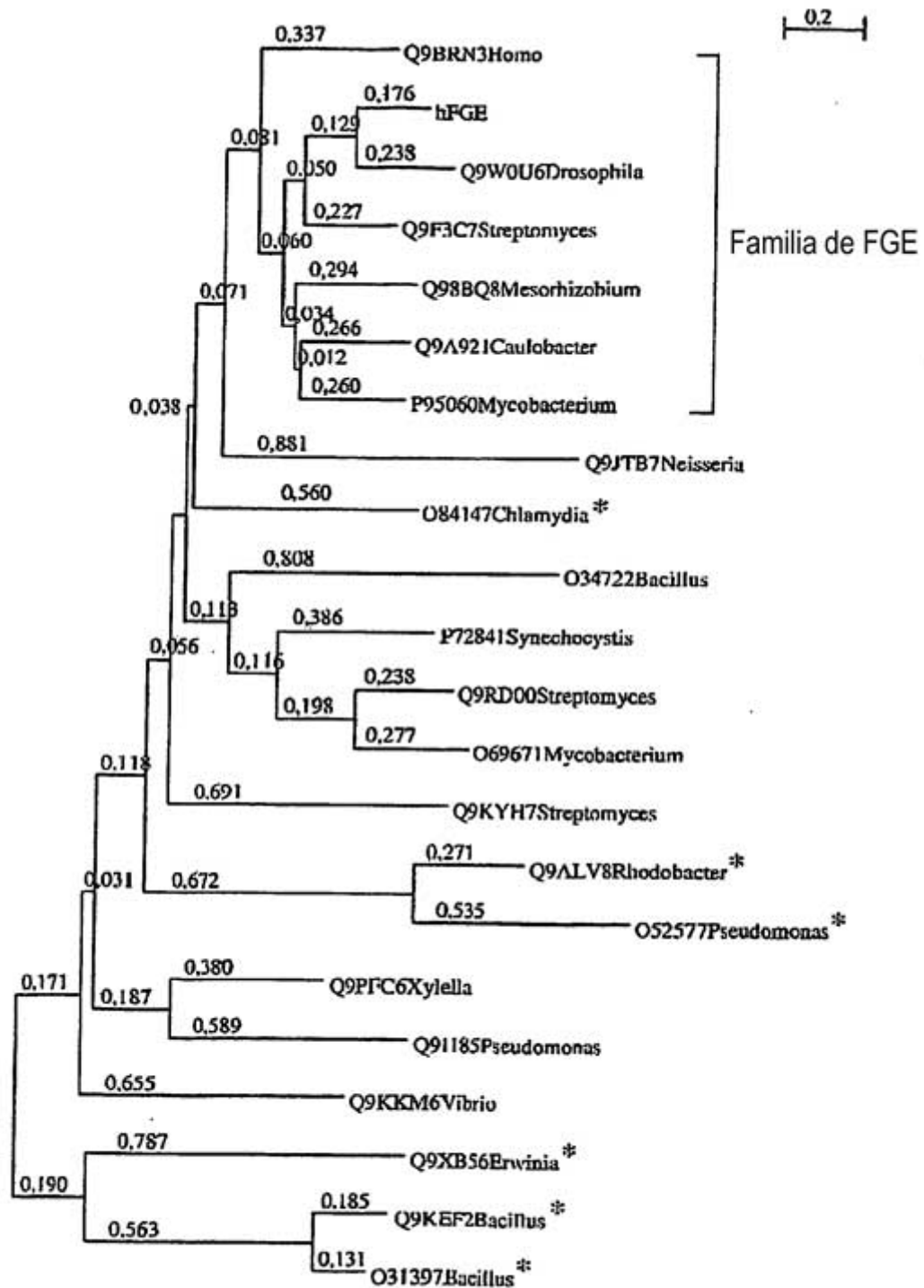


Fig. 2

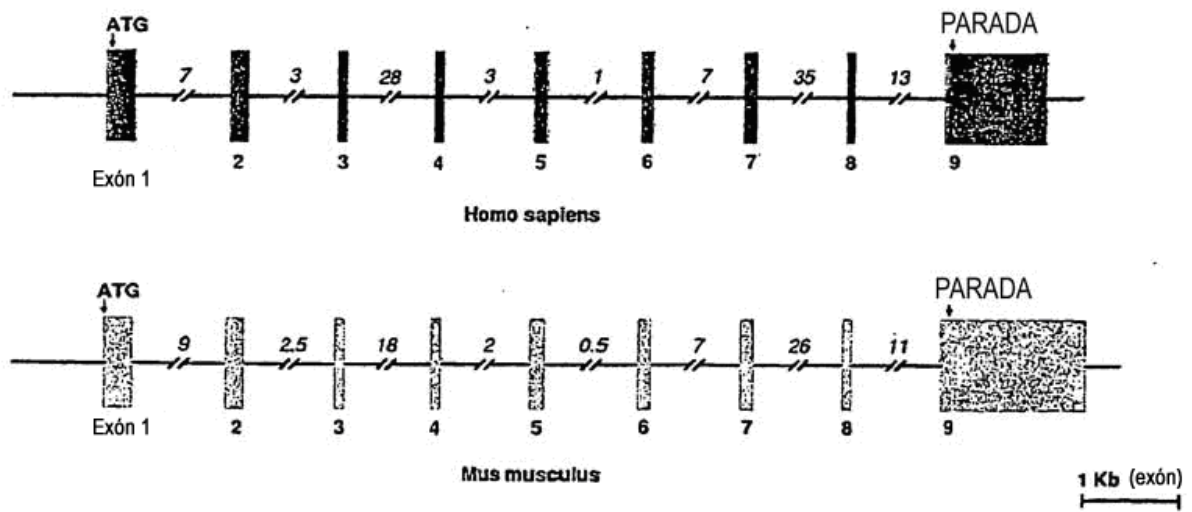


Fig. 3

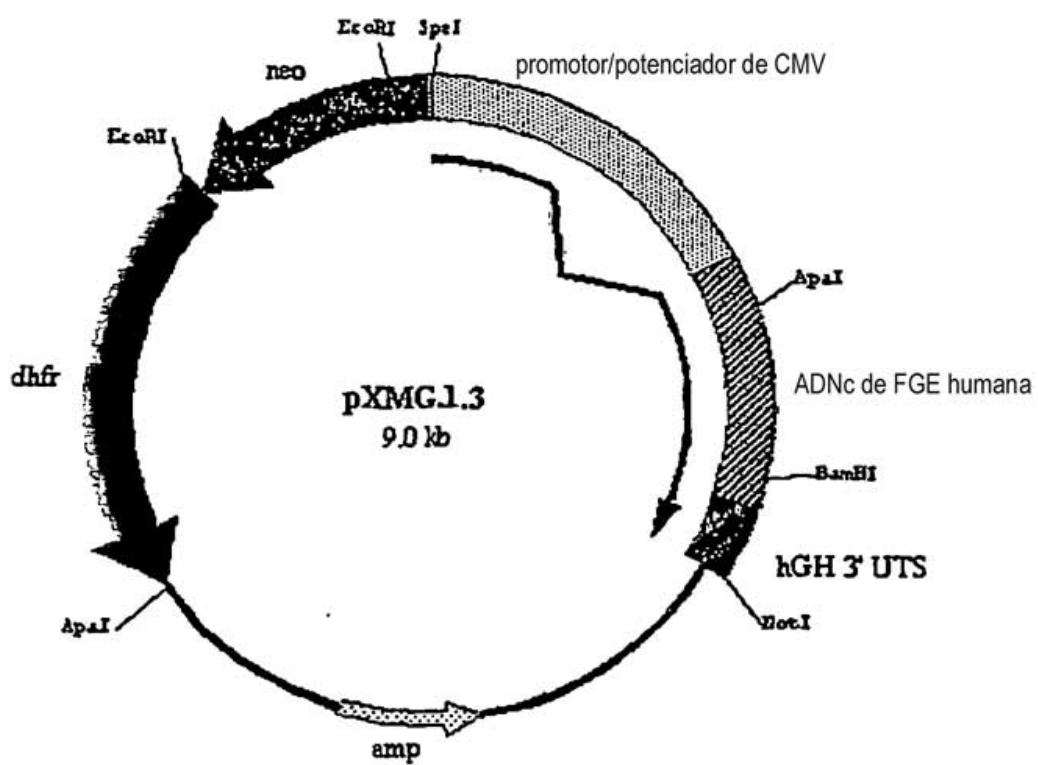


Fig. 4

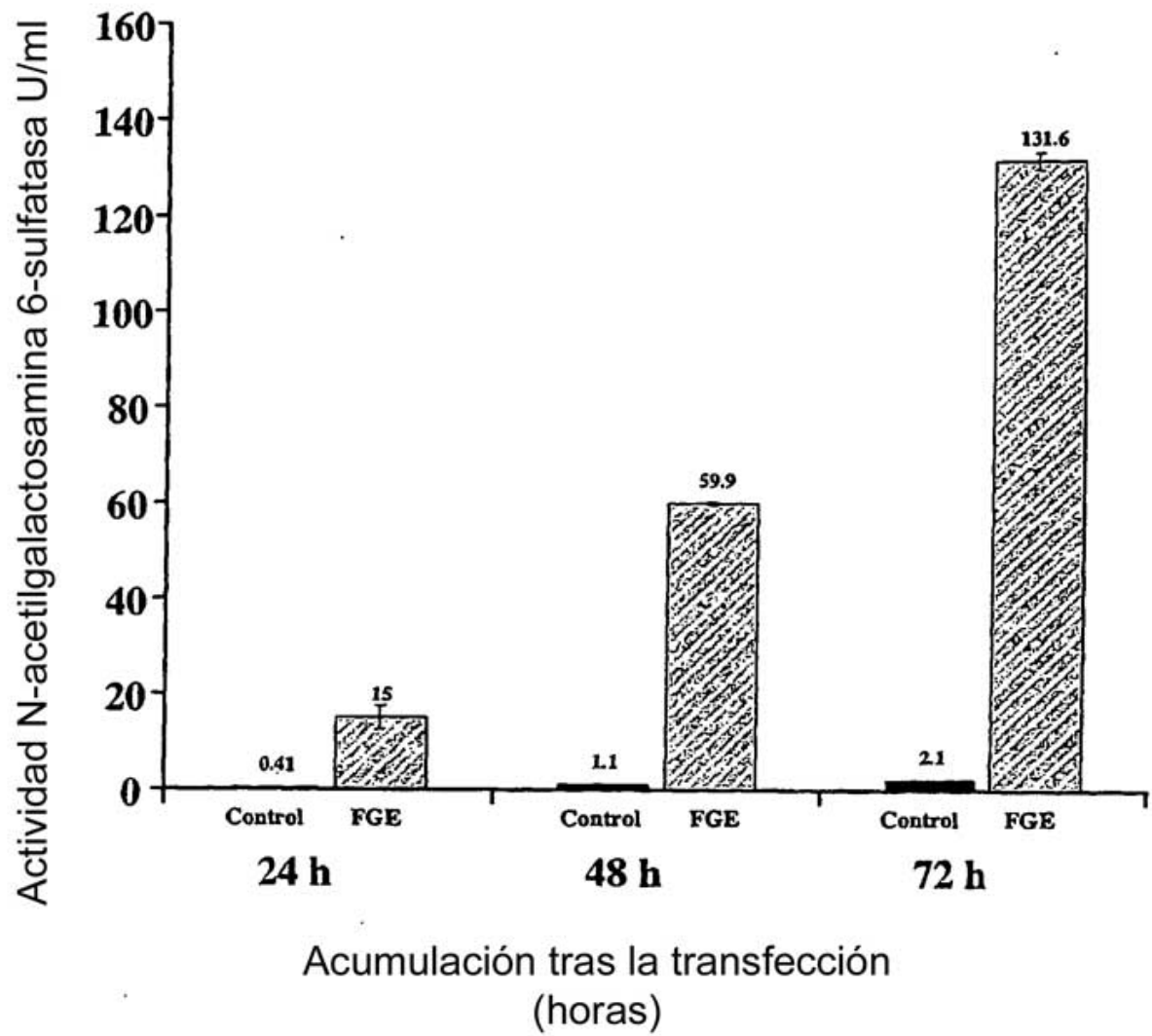


Fig. 5



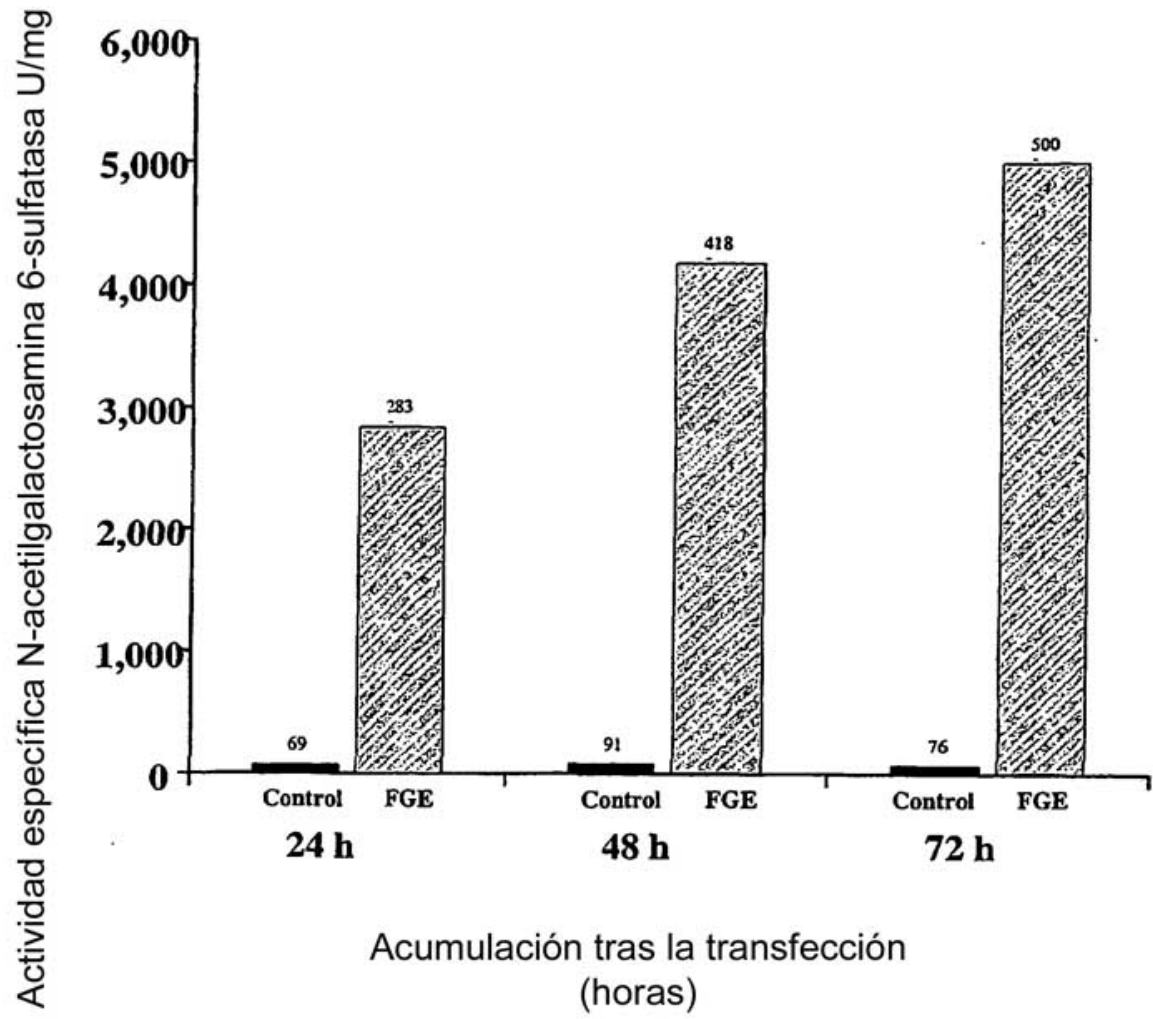


Fig. 6

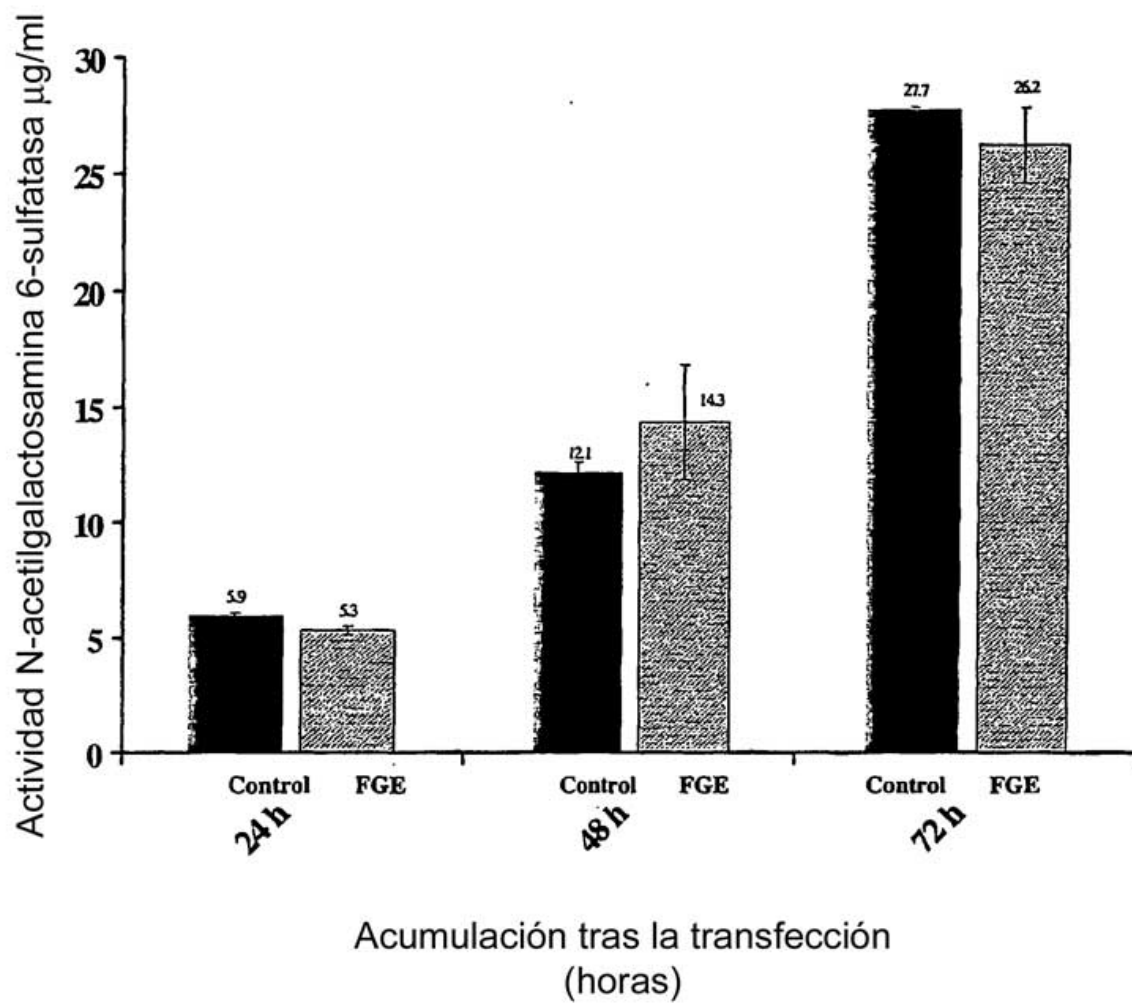


Fig. 7

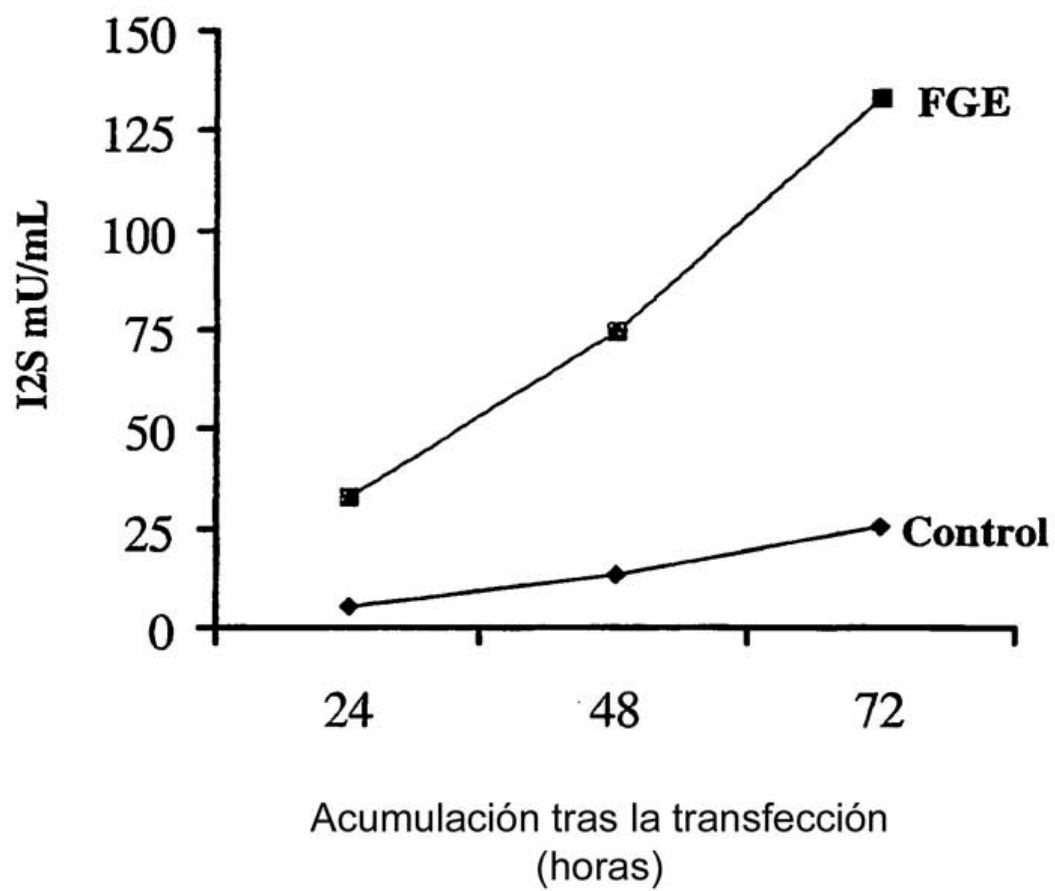


Fig. 8

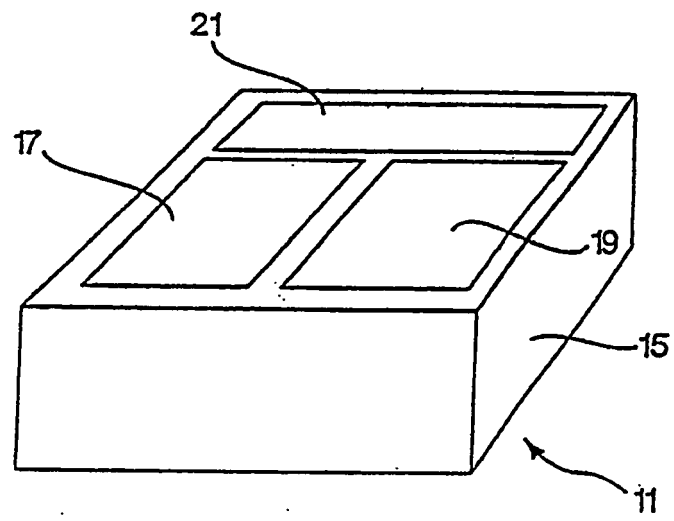


Fig. 9