



19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA

11 Número de publicación: **2 343 571**

51 Int. Cl.:

**C12N 9/10** (2006.01)

**C12N 9/90** (2006.01)

**C12N 15/54** (2006.01)

**C07K 1/14** (2006.01)

**C12Q 1/48** (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Número de solicitud europea: **04767632 .5**

96 Fecha de presentación : **08.07.2004**

97 Número de publicación de la solicitud: **1641920**

97 Fecha de publicación de la solicitud: **05.04.2006**

54

Título: **Glutamina:fructosa-6-fosfato amidotransferasa (GFAT) que comprende una etiqueta de purificación interna, y su utilización para el cribado de compuestos.**

30

Prioridad: **08.07.2003 FR 03 08350**

45

Fecha de publicación de la mención BOPI:  
**04.08.2010**

45

Fecha de la publicación del folleto de la patente:  
**04.08.2010**

73

Titular/es: **CENTRE NATIONAL DE LA  
RECHERCHE SCIENTIFIQUE (CNRS)  
3, rue Michel-Ange  
75794 Paris Cédex 16, FR**

72

Inventor/es:  
**Badet-Denisot, Marie-Ange, Juliette, Etienne y  
Badet, Bernard, François**

74

Agente: **Curell Suñol, Marcelino**

ES 2 343 571 T3

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

## DESCRIPCIÓN

Glutamina:fructosa-6-fosfato amidotransferasa (GFAT) que comprende una etiqueta de purificación interna, y su utilización para el cribado de compuestos.

La presente invención se refiere a una glutamina:fructosa-6-fosfato aminotransferasa modificada, purificable rápidamente y en cantidades suficientes para el cribado de compuestos que modifican su actividad.

Las glutamina:fructosa-6-fosfato aminotransferasas (GFAT), EC 2.6.1.16, denominadas asimismo glucosamina-6-fosfato sintasas o 2-desoxi-glucosa-6-fosfato cetol isomerasas, están implicadas en la vía de biosíntesis de las hexosaminas. La GFAT cataliza la primera etapa, limitante, de esta vía de biosíntesis según la reacción:

$$\text{L-glutamina} + \text{fructosa-6-fosfato} \rightarrow \text{L-glutamato} + \text{glucosamina-6-fosfato}$$
 mediante transferencia del nitrógeno amídico de la L-glutamina sobre la función cetona del fructosa-6-fosfato. Las GFAT controlan por lo tanto el flujo de glucosa en la vía de las hexosaminas, a través del fructosa-6-fosfato, y por consiguiente la formación de las hexosaminas producidas.

Una forma bacteriana recombinante de la GFAT, la glucosamina-6-fosfato sintasa de *Escherichia coli*, ha sido purificada hasta homogeneidad y estudiada de manera exhaustiva. Las propiedades y el mecanismo enzimático de la transferencia de la amida han sido en particular ampliamente descritas (para revisión Teplyakov *et al.*, Nat. Prod. Rep. (2002) 19:60). En particular, esta enzima, cuya estructura cristalina ha sido resuelta (Teplyakov *et al.*, J. Mol. Biol. (2001) 313:1093), está formada por dos dominios, uno que tiene una actividad hidrolasa (glutaminasa) y el otro una actividad isomerasa.

Por otro lado, se han secuenciado unas GFAT de eucariotas, incluyendo en particular la de hígado de rata (Huynh *et al.*, Arch. Biochem. Biophys. (2000) 379:307) y la de la levadura *Candida albicans* (Milewsky *et al.*, J. Biol. Chem. (1999) 274:4000).

En el ser humano, unos estudios preliminares han mostrado la presencia de una actividad GFAT en el hígado (Ghosh *et al.*, J. Biol. Chem. (1960) 235:1265). A partir de entonces se conocen varias GFAT. GFAT1, la forma principal, GFAT2, que se expresa preferentemente en el sistema nervioso central, y GFAT1Alt, una isoforma de GFAT1, expresada esencialmente en los músculos estriados. Las secuencias peptídicas de GFAT1 y GFAT2 poseen 75% de identidad de secuencias entre sí, y las de GFAT1 y GFAT1Alt son idénticas salvo para una inserción de 18 aminoácidos en la secuencia de GFAT1Alt. Las secuencias de GFAT están por lo tanto muy conservadas en el ser humano, pero igualmente entre las especies, puesto que las secuencias peptídicas de la GFAT1 humana y de la GFAT de *E. coli* o de la GFAT1 de ratón presentan respectivamente 35% y 99% de identidad.

El gen de la GFAT1 humana ha sido clonado en 1992 (McKnight *et al.*, J. Biol. Chem. (1992) 267:25208). Codifica una proteína de 77 kDa formada por dos dominios distintos (Teplyakov *et al.*, Nat. Prod. Rep. (2002) 19:60).

El aumento de la producción de UDP-NAc-GlcNH<sub>2</sub>, el producto final de la vía de biosíntesis de las hexosaminas, y su acumulación en los tejidos han estado recientemente implicados en el desarrollo de la resistencia a la insulina (Marshall *et al.*, FASEB J. (1991) 5:3031, Yki-Jarvinen *et al.*, Diabetes (1996) 45:302, Thompson *et al.*, J. Biol. Chem. (1997) 272: 7759, Hawkins *et al.*, J. Clin. Invest. (1997) 99:2173, Robinson *et al.*, Diabetes (1993) 42:1333, Daniels *et al.*, J. Clin. Invest. (1995) 96:1235, Baron *et al.*, J. Clin. Invest. (1995) 96:2792).

Así, se ha mostrado que un aumento del nivel celular de UDP-NAc-GlcNH<sub>2</sub> mediante una modesta sobreexpresión de GFAT1, o una aportación de glucosamina exógena, puede inducir una resistencia a la insulina al mismo tiempo *in vivo* y en unos adipocitos en cultivo (Robinson *et al.*, Diabetes (1993) 42:1333, Daniels *et al.*, J. Clin. Invest. (1995) 96:1235, Baron *et al.*, J. Clin. Invest. (1995) 96:2792, Hebert *et al.*, J. Clin. Invest. (1996) 98:930).

En efecto, la insulina activa su vía de transducción fijándose a su receptor, lo que induce la translocación de los vehículos de glucosa, tales como el receptor GLUT4, almacenados en la célula, hacia la membrana, y aumenta el flujo de glucosa. La glucosa entra así en la vía de la glicolisis y se convierte en glucosa-6-fosfato y después en fructosa-6-fosfato. Cuando el flujo de glucosa es excesivo, el fructosa-6-fosfato entra en la vía de biosíntesis de las hexosaminas y se convierte en glucosamina-6-fosfato por la GFAT. Varias observaciones indican que los metabolitos del glucosamina-6-fosfato impiden la translocación de los receptores a la glucosa hacia la membrana celular, disminuyendo así el flujo de la glucosa celular (Marshall *et al.*, FASEB J. (1991) 5:3031, Giacarrí *et al.*, Diabetologia (1995) 38:518, Marshall *et al.*, J. Biol. Chem. (1991) 266:4706, Paterson *et al.*, Endocrinology (1995) 136:2809).

El mecanismo mediante el cual los metabolitos del glucosamina-6-fosfato ejercen sus efectos fisiológicos no está claro. Se ha propuesto sin embargo una hipótesis: una concentración citosólica elevada de UDP-NAc-GlcNH<sub>2</sub> provocaría la hiperglicosilación de los sitios de fosforilación Ser o Thr, conduciendo de esta manera a la detención de la vía de señalización de la insulina (Comer *et al.*, J Biol. Chem (2000) 275:29179).

La actividad GFAT se considera por lo tanto como una de las causas de los altos niveles de glucosa en sangre; por otro lado, se conoce por ser elevada en los pacientes que padecen diabetes azucarada no-insulino-dependiente o diabetes de tipo II (Yki-Jarvinen *et al.*, Diabetes (1996) 45:302).

## ES 2 343 571 T3

La obtención de inhibidores de la GFAT permitiría reducir la glicemia en particular en los individuos que padecen patologías relacionadas con una hiperglicemia, tales como la diabetes de tipo II, la acidosis y/o la cetosis diabética, por ejemplo.

5 Unos inhibidores de la GFAT de planta o de hongo podrían asimismo permitir obtener respectivamente unos fungicidas y unos herbicidas.

Sin embargo, a pesar de la obtención de formas recombinantes de GFAT, la inestabilidad de las preparaciones enzimáticas obtenidas, su baja cantidad, y su nivel de purificación insuficiente, no han permitido obtener unos inhibidores eficaces de GFAT.

Un objetivo de la invención es por lo tanto proporcionar una GFAT modificada cuya actividad es estable y que se puede obtener en gran cantidad, con un alto nivel de pureza y de actividad.

15 La presente invención se refiere a una proteína enzimáticamente activa que comprende:

- una secuencia de GFAT y por lo menos una secuencia de una etiqueta de purificación, siendo la secuencia de la etiqueta de purificación insertada entre dos aminoácidos consecutivos de la secuencia de GFAT,

20 estando dichos aminoácidos comprendidos en:

- una secuencia de GFAT que corresponde a la secuencia que se extiende entre los aminoácidos 30 a 80 de la GFAT de *Escherichia coli*, o

- una secuencia de GFAT que corresponde a la secuencia que se extiende entre los aminoácidos 220 a 230 de la GFAT de *Escherichia coli*, o

- una secuencia que se deriva de la secuencia anterior por supresión, inserción o mutación de por lo menos un aminoácido, con la condición de que dicha proteína presente una actividad enzimática, o

30 - una secuencia que presenta por lo menos 35%, en particular por lo menos 90%, de identidad de secuencia y/o por lo menos 44%, en particular por lo menos 95%, de similitud de las secuencias con una de las secuencias anteriores, con la condición de que dicha proteína presente una actividad enzimática.

35 Se designa por el término GFAT una enzima de la clase E.C. 2.6.1.16 que cataliza la reacción siguiente:

L-glutamina + fructosa-6-fosfato → L-glutamato + glucosamina-6-fosfato en particular en las condiciones experimentales tales como las descritas en el ejemplo siguiente o en Broschat *et al.*, J. Biol. Chem. (2002) 277:14764.

40 GFAT se designa con el nombre de glutamina:fructosa-6-fosfato aminotransferasa, o también glucosamina-6-fosfato sintasa o 2-desoxi-glucosa-6-fosfato cetol isomerasa.

Mediante la expresión “proteína enzimáticamente activa” se designa una proteína que tiene una acción catalítica.

45 Ventajosamente, la proteína enzimáticamente activa posee una actividad GFAT.

Mediante la expresión “etiqueta de purificación” se designa una secuencia peptídica susceptible de unirse específicamente a un ligando determinado. Ventajosamente, la unión de dicho ligando a la etiqueta de purificación permite formar un complejo entre la proteína que contiene la etiqueta de purificación y dicho ligando, pudiendo dicho complejo ser aislado específicamente.

Ventajosamente, las etiquetas de purificación según la invención no están dispuestas en final de cadena peptídica, en el extremo N-terminal o C-terminal, sino en el interior de la cadena peptídica.

55 Mediante la expresión “identidad de secuencia” se designa el porcentaje de aminoácidos idénticos entre dos secuencias alineadas, en particular con la ayuda de algoritmos tales como el definido por Altschul *et al.*, Nucleic Acids Res. (1997) 25:3389, por ejemplo.

60 Mediante la expresión “similitud de secuencia” se designa el porcentaje de aminoácidos similares, es decir, de aminoácidos cuyas cadenas laterales poseen unas propiedades fisicoquímicas parecidas, entre dos secuencias alineadas, en particular con la ayuda de algoritmos tales como el definido por Altschul *et al.*, Nucleic Acids Res. (1997) 25:3389, por ejemplo.

65 La presente invención se refiere en particular a una proteína tal como se ha definido anteriormente, en la que la secuencia de GFAT corresponde a una secuencia de GFAT de bacteria o de eucariota, en particular de planta, de hongo o de animal, en particular de insecto o de mamífero, más particularmente de ser humano.

## ES 2 343 571 T3

La invención se refiere más particularmente a una proteína tal como se ha definido anteriormente, en la que la secuencia de la etiqueta de purificación se inserta entre dos aminoácidos consecutivos de la secuencia de GFAT, estando dichos aminoácidos comprendidos en:

- 5 - una parte de la secuencia de la GFAT que corresponde y/o que es homóloga a la secuencia que se extiende entre la hoja  $\beta 2$  y la hoja  $\beta 3$  de la GFAT de *Escherichia coli*, o
- una parte de la secuencia de la GFAT que corresponde y/o que es homóloga a la secuencia que se extiende entre la hoja  $\beta 13$  y la hoja  $\beta 14$  de la GFAT de *Escherichia coli*.

10 La estructura de la GFAT de *Escherichia coli* está descrita en particular por Teplyakov *et al.*, J. Mol. Biol. (2001) 313:1093 (proteína entera), por Isupov *et al.*, Structure (1996) 4:801 (dominio glutaminasa) y por Teplyakov *et al.* Structure (1998) 6:1047 (dominio isomerasa). La estructura de la proteína completa se puede consultar en particular con la ayuda del fichero de coordenadas atómicas 1JXA depositado en Protein Data Bank (<http://www.pdb.org>).

15 La secuencia peptídica de la GFAT de *E. coli* se define mediante la SEC ID nº 13.

La secuencia que se extiende entre la hoja  $\beta 2$  y la hoja  $\beta 3$  corresponde a la secuencia que se extiende aproximadamente entre los aminoácidos 30 a 80 de la GFAT de *E. coli*, situados en el dominio glutaminasa.

20 La secuencia que se extiende entre la hoja  $\beta 13$  y la hoja  $\beta 14$  corresponde a la secuencia que se extiende aproximadamente entre los aminoácidos 220 a 230 de la GFAT de *E. coli*, situados en el dominio glutaminasa.

Según un modo de realización particular, la invención se refiere por lo tanto a una proteína tal como se ha definido anteriormente, en la que la secuencia de la etiqueta de purificación se inserta entre dos aminoácidos consecutivos de la secuencia de GFAT, estando dichos aminoácidos comprendidos en:

- 25 - una parte de la secuencia de la GFAT que corresponde y/o que es homóloga a la secuencia que se extiende aproximadamente entre los aminoácidos 30 a 80 de la GFAT de *Escherichia coli*, o
- 30 - una parte de la secuencia de la GFAT que corresponde y/o que es homóloga a la secuencia que se extiende aproximadamente entre los aminoácidos 220 a 230 de la GFAT de *Escherichia coli*,

35 La identificación de las partes de secuencias de una GFAT que corresponde y/o que es homóloga a unas estructuras secundarias de la GFAT de *E. coli* se puede obtener alineando la secuencia de dicha GFAT con la de la GFAT de *E. coli*, en particular con la ayuda de un algoritmo tal como el definido por Altschul *et al.*, Nucleic Acids Res. (1997) 25:3389 o con la ayuda del programa Clustal W, bien conocido por el experto en la técnica y descrito por Thompson *et al.*, Nucleic Acids Res. (1994) 22 4673-4680, por ejemplo.

40 En particular, dos secuencias o partes de secuencias se denominan homólogas si el porcentaje de identidad entre estas dos secuencias o partes de secuencias es superior a aproximadamente 35%, y/o si el porcentaje de identidad entre estas dos secuencias o partes de secuencias es superior a aproximadamente 44%.

45 Más particularmente, dos secuencias o partes de secuencias se denominan homólogas si son susceptibles de hibridarse en unas condiciones astringentes, tales como las condiciones siguientes: formamida 50%, NaCl 0,75 mol/l, citrato de sodio 0,75 mmol/l, dodecilsulfato de sodio 1%, pH 7, 42°C.

50 Según otro modo de realización preferido, la invención se refiere a una proteína tal como se ha definido anteriormente, en la que la secuencia de la etiqueta de purificación se inserta entre dos aminoácidos consecutivos de la secuencia de una GFAT humana, estando dichos aminoácidos comprendidos entre los aminoácidos 40 a 50 y/o 290 a 330, de dicha secuencia de GFAT humana.

55 Los aminoácidos 40 a 50 de dicha secuencia de GFAT humana corresponden y/o son homólogos a la parte de la secuencia de la GFAT de *E. coli* que se extiende entre la hoja  $\beta 2$  y la hoja  $\beta 3$ , es decir, la secuencia que se extiende aproximadamente entre los aminoácidos 30 a 80 de la GFAT de *E. coli*.

Los aminoácidos 290 a 330 de dicha secuencia de GFAT humana corresponden y/o son homólogos a la parte de la secuencia de la GFAT de *E. coli* que se extiende entre la hoja  $\beta 13$  y la hoja  $\beta 14$ , es decir, la secuencia que se extiende aproximadamente entre los aminoácidos 220 a 230 de la GFAT de *E. coli*.

60 La invención se refiere en particular a una de las proteínas anteriores, en la que la secuencia de GFAT corresponde a:

- 65 - SEC ID nº 2, que corresponde a la secuencia de la GFAT1 humana,
- SEC ID nº 4, que corresponde a la secuencia de la GFAT2 humana,
- SEC ID nº 6, que corresponde a la secuencia de la GFAT1Alt humana.

## ES 2 343 571 T3

La secuencia de la GFAT1 humana se describe en particular en McKnight *et al.*, J. Biol. Chem. (1992) 267: 25208, y corresponde a la secuencia nucleotídica SEC ID n° 1.

5 La secuencia de la GFAT2 humana se describe en particular en Oki *et al.*, Genomics (1999) 57:227, y corresponde a la secuencia nucleotídica SEC ID n° 3.

La secuencia de la GFAT1Alt humana se describe en particular en DeHaven *et al.*, Diabetes (2001) 50:2419, y corresponde a la secuencia nucleotídica SEC ID n° 5.

10 La invención se refiere en particular a una de las proteínas anteriores, en la que la secuencia de la etiqueta de purificación se inserta entre dos aminoácidos consecutivos, estando dichos aminoácidos comprendidos entre los aminoácidos:

15 - 43 a 47 y/o 298 a 306, de SEC ID n° 2,

- 42 a 45 y/o 299 a 307, de SEC ID n° 4,

- 42 a 45 y/o 299 a 307, de SEC ID n° 4,

20 - 43 a 47 y/o 316 a 324, de SEC ID n° 6,

Los aminoácidos 43 a 47 de SEC ID n° 2, 42 a 45 de SEC ID n° 4 y 43 a 47 de SEC ID n° 6 corresponden, a saber son homólogos, a la parte de la secuencia de la GFAT de *E. coli* que se extiende entre la hoja  $\beta$ 2 y la hoja  $\beta$ 3, es decir, 25 a la secuencia que se extiende aproximadamente entre los aminoácidos 30 a 80 de la GFAT de *E. coli*.

Los aminoácidos 298 a 306 de SEC ID n° 2, 299 a 307 de SEC ID n° 4 y 325 a 330 de SEC ID n° 6 corresponden, a saber son homólogos, a la parte de la secuencia de la GFAT de *E. coli* que se extiende entre la hoja  $\beta$ 13 y la hoja  $\beta$ 14, es decir, a la secuencia que se extiende aproximadamente entre los aminoácidos 220 a 230 de la GFAT de *E. coli*.

30 Según otro modo de realización particular, la invención se refiere más particularmente a una proteína tal como se ha definido anteriormente, en la que la secuencia de la etiqueta de purificación se inserta entre dos aminoácidos consecutivos de la secuencia de GFAT, estando dichos aminoácidos comprendidos en:

35 - una parte de la secuencia de la GFAT que corresponde y/o que es homóloga a la secuencia que se extiende aproximadamente entre los aminoácidos 43 a 47 de la GFAT1 humana,

40 - una parte de la secuencia de la GFAT que corresponde y/o que es homóloga a la secuencia que se extiende aproximadamente entre los aminoácidos 298 a 306, en particular 299 a 300, de la GFAT1 humana,

La invención se refiere más particularmente a una de las proteínas anteriores, en la que la secuencia de la etiqueta de purificación se inserta entre los aminoácidos:

45 - 299 y 300 de SEC ID n° 2,

- 300 y 301 de SEC ID n° 4,

- 317 y 318 de SEC ID n° 6.

50 La invención se refiere en particular a una de las proteínas anteriores, en la que la etiqueta de purificación corresponde a una secuencia de aproximadamente 2 a aproximadamente 10 aminoácidos, en particular de aproximadamente 4 a aproximadamente 8 aminoácidos.

55 Unas etiquetas de purificación preferidas según la invención se refieren en particular a unas etiquetas denominadas FLAG (Sigma-Aldrich, Francia). Estas etiquetas se unen específicamente a un paratope determinado, pudiendo dicho paratope pertenecer a un anticuerpo o a un fragmento de anticuerpo por ejemplo. Un ejemplo particular de etiqueta FLAG está constituido por la secuencia peptídica Asp-Tyr-Lys-Asp-Asp-Asp-Asp-Lys (SEC ID n° 18), por ejemplo.

60 Otras etiquetas preferidas según la invención son unas etiquetas formadas por varias histidinas. Estas etiquetas pueden formar unos complejos con unos cationes metálicos divalentes tales como Ni<sup>2+</sup> o Co<sup>2+</sup> por ejemplo.

La invención se refiere en particular a una proteína tal como se ha definido anteriormente, en la que la etiqueta de purificación es una hexa-histidina.

Se designa por el término "hexa-histidina" la secuencia His-His-His-His-His-His (SEC ID n° 19).

## ES 2 343 571 T3

La invención se refiere más particularmente a una proteína tal como se ha definido anteriormente, que corresponde a las secuencias:

- 5 - SEC ID nº 8, que corresponde a la secuencia SEC ID nº 2 en la que una hexa-histidina se inserta entre los aminoácidos 299 y 300,
- SEQ ID nº 10, que corresponde a la secuencia SEC ID nº 4 en la que una hexa-histidina se inserta entre los aminoácidos 300 y 301, y
- 10 - SEQ ID nº 12, que corresponde a la secuencia SEC ID nº 6 en la que una hexa-histidina se inserta entre los aminoácidos 317 y 318.

La presente invención se refiere asimismo a un ácido nucleico que comprende o que está constituido por una secuencia que codifica para una proteína tal como se ha definido anteriormente.

15 La invención se refiere más particularmente a un ácido nucleico que comprende o que está constituido por la secuencia nucleotídica:

- 20 - SEC ID nº 7 que codifica para la proteína SEC ID nº 8, o
- SEC ID nº 9 que codifica para la proteína SEC ID nº 10, o
- SEC ID nº 11 que codifica para la proteína SEC ID nº 12,

25 o por su complementario, o que se deriva de dicha secuencia por mutación, inserción o delección de por lo menos un nucleótido, con la condición de que dicha secuencia nucleotídica codifique para una proteína enzimáticamente activa.

Según otro modo de realización, la presente invención se refiere asimismo a un vector eucariota o procariota que comprende un ácido nucleico tal como se ha definido anteriormente.

30 Estos vectores permiten en particular sintetizar las proteínas según la invención en un organismo eucariota o procariota.

35 Ventajosamente, la invención se refiere a un vector de expresión de tipo baculovirus que permite la síntesis de las proteínas según la invención en unas células de insecto.

La presente invención se refiere asimismo a un procedimiento de purificación de una proteína tal como se ha definido anteriormente, a partir de una disolución que comprende dicha proteína, que comprende una etapa de puesta en presencia de dicha disolución con un compuesto que se une específicamente a la etiqueta de purificación de dicha proteína, y una etapa de separación del complejo formado por la unión de dicha proteína a dicho compuesto de los demás constituyentes de la disolución.

45 El compuesto puede ser fijado sobre un soporte sólido de manera que el complejo formado entre dicho compuesto y dicha proteína pueda ser recuperado mediante centrifugación o mediante filtración. Opcionalmente, dicho compuesto fijado sobre su soporte puede ser dispuesto en una columna a través de la cual se eluye dicha disolución.

Ventajosamente, el procedimiento anterior puede comprender asimismo una etapa de disociación del complejo formado por la unión de dicha proteína a dicho compuesto con el fin de recuperar la proteína purificada.

50 La invención se refiere más particularmente a un procedimiento de purificación tal como se ha definido anteriormente, que comprende una etapa de puesta en presencia de una disolución que comprende una proteína tal como se ha definido anteriormente con un compuesto que comprende un catión metálico divalente seleccionado de entre el Ni<sup>2+</sup> o el Co<sup>2+</sup>, en particular el Ni<sup>2+</sup>, y una etapa de separación del complejo formado por la unión de la proteína a dicho compuesto de los demás constituyentes de la disolución.

55 Ventajosamente, el procedimiento anterior puede comprender asimismo una etapa de disociación del complejo formado por la unión de dicha proteína a dicho compuesto que comprende un catión metálico divalente, en particular con la ayuda de imidazol, con el fin de recuperar la proteína purificada.

60 Según otro modo de realización, la presente invención se refiere a un procedimiento de conservación de una proteína tal como se ha definido anteriormente, en una forma enzimáticamente activa, en particular a -80°C o a 4°C, que comprende la adición de dicha proteína a una disolución que comprende:

- 65 - aproximadamente 1 mM a aproximadamente 10 mM de fructosa-6-fosfato, en particular aproximadamente 1 mM,
- aproximadamente 1 mM a aproximadamente 5 mM de Tris(2-carboxietil)fosfina, en particular aproximadamente 1 mM,

## ES 2 343 571 T3

- aproximadamente 5% a aproximadamente 20% de glicerol, en particular aproximadamente 10%.

El fructosa-6-fosfato es un sustrato de dicha proteína.

- 5 La tris(2-carboxietil)fosfina es un compuesto reductor que permite ventajosamente mantener las propiedades de resinas que contienen unos iones  $\text{Ni}^{2+}$  o  $\text{Co}^{2+}$ .

Ventajosamente, el glicerol es un agente crioprotector.

- 10 A este respecto, la presente invención se refiere asimismo a una composición que comprende una proteína GFAT activa unida a una etiqueta de purificación tal como una proteína tal como se ha definido anteriormente, siendo dicha proteína susceptible de poder ser conservada en una forma enzimáticamente activa, durante por lo menos 8 días a una temperatura de 2°C a 10°C, en particular aproximadamente 4°C, y durante por lo menos 12 meses a una temperatura de -100°C a -20°C, en particular aproximadamente -80°C, estando dicha proteína en asociación con:

- 15 - aproximadamente 1 mM a aproximadamente 10 mM de fructosa-6-fosfato, en particular aproximadamente 1 mM,
- 20 - aproximadamente 1 mM a aproximadamente 5 mM de tris(2-carboxietil)fosfina, en particular aproximadamente 1 mM,
- aproximadamente 5% a aproximadamente 20% de glicerol, en particular de aproximadamente 10%.

- 25 La presente invención se refiere asimismo a la utilización de una proteína tal como se ha definido anteriormente, para el cribado de compuestos que modifican la actividad de dicha proteína, en particular para el cribado de inhibidor de dicha proteína.

La actividad de las proteínas según la invención se puede medir en particular con la ayuda de los métodos siguientes:

- 30 - el método radiométrico descrito por Broschat *et al.*, Analytical Biochem. (2002) 305:10-15,
- el método denominado con Nitro Bleu Tétrazolium descrito por Nakata *et al.*, J. Antibio. (2001) 54:737-743.
- 35 - el método Morgan-Elson descrito por Ghosh *et al.*, Method. Enzymol. (1960) 5:414 y detallado en el ejemplo siguiente.
- el método APAD descrito por Badet *et al.*, Biochemistry (1987) 26:1940 y detallado en el ejemplo siguiente.

- 40 Ventajosamente, estos métodos se pueden utilizar para el cribado, en particular de alto caudal, de compuestos que modifican la actividad de las proteínas según la invención.

- La invención se refiere en particular a la utilización tal como se ha definido anteriormente, para el cribado de compuestos útiles en el marco del tratamiento o de la prevención de la diabetes, en particular de la diabetes de tipo II, de la obesidad, de la acidosis, de la cetosis, de la artritis, de los cánceres o de la osteoporosis.

### Descripción de la figura 1

- 50 La figura 1 representa el plásmido pFastBac-gfat-His6 de peso molecular 6,89 kb. El casete "Ampr" representa un gen de resistencia a la ampicilina, el casete "ori" representa un origen de replicación bacteriana, el casete "Gmr" representa un gen de resistencia a la gentamicina, el casete "Polh Pr" representa el promotor de la polihedrina, el casete gfat-his6 representa el gen gfat1 modificado mediante la inserción de una secuencia que codifica para una etiqueta hexahistidina. Están representados asimismo los sitios de restricción *Xba*I en posición 4,11 kb, y *Eco*RI en posiciones 4,56 kb y 6,60 kb.

### 55 Ejemplo 1

#### 1. Síntesis y clonación del gen *gfat1-His6*

- 60 El fragmento *Eco*RI de un ADNc que corresponde al gen *gfat1* humano ha sido clonado en el sitio *Eco*RI I del vector pCRII (Invitrogen) para formar el plásmido pCRII-gfat1. La secuencia nucleotídica de una etiqueta de purificación interna compuesta por 6 restos de histidina ha sido introducida en la posición 898 de la secuencia del gen *gfat1* clonado en pCRII mediante PCR con la *platinum pfx polymerase* (Roche) y el par de cebadores apropiados:

- 65 - Start Aat II-His6:

5'TGGACGTCTTTCTATCCATCGAATTAACGAACTGCAGGACATCACCATCACCATCACGATCACC  
CCGGACG3' (SEC ID n° 14)

## ES 2 343 571 T3

- End Hinc II:

5'CAAAGTTGACTCTTCCTCTCATTGTGTTACGACAGACTCTGGC3' (SEC ID nº 15)

5 según el protocolo siguiente: 94°C, durante 2 min. y después 30 ciclos (94°C durante 45 s, 55°C durante 1 min., 72°C durante 5 min.) seguidos de una polimerización de 5 min. a 72°C y vuelta a 4°C.

Después de la digestión por *AatII* y *HincII* y después purificación sobre gel de agarosa Seaplaque 1,5% (Tebu), el amplicón (170 pb) ha sido insertado a nivel de los sitios de restricción correspondientes en la construcción pCRII-gfat1. El inserto de 170 pb ha sido introducido mediante ligación en la construcción con una relación 3:1 a 16°C durante la noche en presencia de T4 DNA ligase (Nebs). La mezcla de ligación (20 µl) así obtenida ha permitido transformar una cepa de *E. coli* JM109. A continuación, el fragmento *XbaI-HindIII* del plásmido recombinante pCRII-gfat1-His6 ha sido clonado en el plásmido donante pFastBac1 (Life Technologies Ltd). El plásmido pFastBac-gfat-His6 así generado (figura 1) ha sido verificado mediante múltiples digestiones: *SmaI*, *AccI/DraI*, *PstEI/XbaI*, y mediante secuenciación. Con vistas a mejorar la construcción, la secuencia corriente arriba del codón de partida ha sido mutada sobre dos posiciones mediante PCR, con el par de cebadores siguiente, con el fin de retirar dos fases de lectura abiertas corriente arriba del gen *gfat1*:

- Start XbaI

5'AATCTAGATTCATGCTCGAGCGGCCCGCCAGTGTGATTGATATC3' (SEC ID nº 16)

- End AfeI

5'ATTTTTATCAGAGCGCTGGGGGTGGCTATTGACAGG3' (SEC ID nº 17)

según el protocolo: 94°C durante 2 min., después 30 ciclos (94°C durante 15 s, 55°C durante 30 s, 68°C durante 1 min.) seguidos de una polimerización de 1 min. a 68°C y vuelta a 4°C.

El fragmento de PCR obtenido, que contiene las dos mutaciones, ha sido purificado sobre gel Seaplaque (Tebu) a 0,7%, y después digerido por *XbaI* y *AfeI* para sustituir a su homólogo en pFastBac-gfat-His6 con el fin de proporcionar el plásmido donante pFastBac-gfat-His6-orf que sirve para la transposición en las células DH10Bac (Life Technologies Ltd). La construcción ha sido verificada mediante digestiones *SmaI*, *XbaI/PstEI*, *XbaI/HindIII*, y mediante secuenciación.

Un báculo recombinante ha sido aislado después de la transposición en las células DH19Bac y utilizado para transfectar unas células de insecto Sf9 en presencia de Lipofectin (Life Technologies Ltd). Los baculovirus obtenidos han sido amplificados en las células Sf9 y el título vírico ha sido medido a 5.10<sup>7</sup> pfu/ml.

### 40 2. Producción de la proteína GFAT1-His6

Unas células de insectos Sf9 han sido cultivadas a 28°C en presencia de medio SF900II (Life Technologies Ltd) en frascos de 5 l bajo agitación a 100 rpm. Las células con una densidad de 2.10<sup>9</sup> células/l han sido infectadas por el baculovirus recombinante obtenido anteriormente con una multiplicidad de infección de 2 (pfu/célula), y después cultivadas durante 72 h.

Las células y el sobrenadante han sido separados por centrifugación (2.500 g, 10 min a 4°C). Los residuos celulares se han lavado en presencia de tampón Tris-HCl 20 mM, pH 7, centrifugado (4.000 g, 45 min a 4°C) y congelado a -80°C.

### 50 3. Purificación de la proteína GFAT1-His6

El residuo celular (20 g) ha sido recogido en 50 ml de tampón de lisis (NaPO<sub>4</sub> 50 mM pH 7,5, NaCl 300 mM, imidazol 10 mM, fructosa-6-fosfato (fructosa-6P) 1 mM, TCEP (Tris(2-carboxietil)fosfina) 1 mM, PMSF 1 mM (fluoruro de fenilmetilsulfonilo), 10% de glicerol y 1 tableta de mezcla de inhibidores de proteasas sin EDTA (Roche Applied Sciences) y sometido a una trituración con DynoMill a 4.500 rpm. (4 ciclos durante 30 s) en presencia de 40 g de microbolas de 0,2 mm de diámetro. El conjunto ha sido enfriado mediante circulación de etilenglicol/agua regulado a -15°C. El extracto bruto obtenido (100 ml, 445 mg de proteínas totales) ha sido centrifugado a 4°C durante 20 min. a 12.000 rpm. El sobrenadante ha sido sometido a una ultracentrifugación a 4°C (350.000 rpm, 1 h). El sobrenadante así obtenido ha sido mezclado con 5 ml de matriz 50% Ni-NTA (Qiagen) durante 2 h a 4°C. La mezcla ha sido vertida en una columna vacía y después aclarada con 40 ml de tampón de lavado (NaPO<sub>4</sub> 50 mM pH 7,5, NaCl 300 mM, imidazol 40 mM, fructosa-6P 1 mM, TCEP 1 mM, PMSF 1 mM y 1 tableta de mezcla de inhibidores de proteasas sin EDTA (Roche Applied Sciences). La elución ha sido realizada por etapas sucesivas a 125 y 500 mM de imidazol en el mismo tampón que anteriormente. Se han obtenido así 12 mg de GFAT1-His6 funcional (dosificación proteica según el método de Bradford).

## ES 2 343 571 T3

### 4. Conservación de la enzima GFAT1-His6

La enzima se almacenó entonces en fracciones de 100  $\mu$ l en presencia de 1 mM de fructosa-6P, 1 mM de TCEP y 10% de gliceril a -80°C. La estabilidad de la enzima es de varios meses a -80°C y superior a 8 días a 4°C.

### 5. Dosificación de la actividad de la enzima GFAT1-His6

Se han usado diferentes ensayos de dosificación de la actividad enzimática de la GFAT1-His6. Estos ensayos pueden ser utilizados asimismo con el fin de cribar unos compuestos que modifican, y en particular que inhiben, la actividad de la GFAT1-His6. Es posible adaptarlos fácilmente a un cribado de alto caudal.

#### Dosificación de Morgan-Elson

En este caso, la actividad enzimática se sigue mediante un ensayo colorimétrico cuyo principio es el siguiente: la D-glucosamina-6P liberada por la enzima es N-acetilada por el anhídrido acético en medio alcalino (Ghosh *et al.*, Method Enzymol. (1962) 5:414), y después la disolución se trata mediante el agente reactivo de Ehrlich (para-dimetil-amino-benzaldehído, PDAB) en medio ácido concentrado; el compuesto rosa formado absorbe a 585 nm.

La reacción enzimática se desarrolla durante 30 minutos a 37°C en presencia de:

- 0,2 ml de fructosa-6P a 100 mM
- 0,25 mM de L-glutamina a 60 mM
- 0,25 ml de tampón  $KPO_4$  a 150 mM pH 7
- 0,1 ml de EDTA (tetra-acetato de etilendiamina) a 25 mM, pH 7
- hasta 200  $\mu$ l de muestra (a completar con  $H_2O$  si es necesario)

La reacción se detiene mediante inmersión durante 4 minutos en un baño de agua a 100°C y después se centrifuga. Se extraen 0,8 ml del sobrenadante para la dosificación de la glucosamina-6P según el protocolo siguiente:

- adición de 0,1 ml de  $NaHCO_3$  saturado,
- adición de 0,1 ml de una disolución de anhídrido acético al 5% en agua preparada extemporáneamente,
- agitación e incubación durante 5 minutos a temperatura ambiente,
- incubación durante 5 minutos en un baño a 100°C,
- adición de 0,2 ml de borato de potasio 0,8 M pH 9,1 (a ajustar con KOH 10 N),
- agitación e incubación durante 7 minutos en un baño a 100°C,
- adición de 3 ml de agente reactivo de Ehrlich diluido 10 veces en ácido acético, preparado extemporáneamente, sobre la disolución enfriada en hielo,
- incubación durante 20 minutos a 37°C.

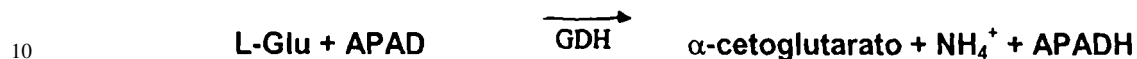
La actividad de la GFAT ha sido determinada en comparación con una curva patrón establecida utilizando la D-glucosamina como estándar en una gama de concentración de 0 a 200 nmoles. La actividad específica de la GFAT1-His6 obtenida ha sido así medida a 1,7 U/mg. Lo que es superior al valor de 0,4 U/mg obtenido por Broschat *et al.*, J. Biol. Chem. (2002) 277:14764, para la purificación de una GFAT1 humana recombinante. Esto traduce una actividad superior de la GFAT1-His6 y/o una pureza más importante de la preparación enzimática según la invención.

Los parámetros cinéticos de la GFAT1-His6 han sido caracterizados frente a la glutamina ( $K_m^{Gln} = 0,2$  M) y a la fructosa-6P (F6P) ( $K_m^{F6P} = 0,006$  mM) mediante una dosificación espectrofotométrica acoplada con la glutamato deshidrogenasa según el ensayo APAD. Lo que está de acuerdo con los valores citados en la técnica anterior ( $K_m^{Gln} = 0,26$  mM y  $K_m^{F6P} = 0,007$  mM por Broschat *et al.*, J. Biol. Chem. (2002) 277:14764).

## ES 2 343 571 T3

### Dosificación APAD

Se trata de una dosificación espectrofotométrica en el ultravioleta de la actividad GFAT. Se basa en la determinación, en continuo, de la cantidad de L-glutamato formado con la ayuda de la GFAT y de un análogo del NAD (nicotinamida adenina dinucleótido), la APAD (3-acetilpiridina adenina dinucleótido), según la reacción siguiente (catalizada por la glutamato deshidrogenasa (GDH)):



La medición se efectúa a 365 nm, a 37°C. En estas condiciones, una unidad de absorbancia corresponde a 0,11  $\mu$ moles de ADADH formado.

El ensayo comprende:

- 100  $\mu$ l de ADAP 3 mM (2 mg/ml)
- 25  $\mu$ l de KCl 2M
- 100  $\mu$ l de tampón KPO<sub>4</sub> 1 M pH 7,2
- 100  $\mu$ l de fructosa-6P 100 mM (30,41 mg/ml)
- 100  $\mu$ l de L-glutamina 60 mM purificada (8,77 mg/ml)
- H<sub>2</sub>O csp 1 ml (teniendo en cuenta los volúmenes a añadir a continuación)
- 50  $\mu$ l de GDH
- muestra a dosificar: 0,5  $\mu$ g

Es posible utilizar asimismo otros procedimientos de dosificación, tales como la dosificación radiométrica descrita por Broschat *et al.*, Analytical Biochem. (2002) 305:10-15 o la dosificación denominada Nitro Bleu Tétrazolium descrita por Nakata *et al.*, J. Antibio. (2001) 54:737-743.

REIVINDICACIONES

1. Proteína que posee una actividad GFAT que comprende:

5 una secuencia de GFAT y por lo menos una secuencia de una etiqueta de purificación, siendo la secuencia de la etiqueta de purificación insertada entre dos aminoácidos consecutivos de la secuencia de GFAT, estando dichos aminoácidos comprendidos en:

- 10
- una secuencia de GFAT que corresponde a la secuencia que se extiende entre los aminoácidos 30 a 80 de la GFAT de *Escherichia coli*, o
  - una secuencia de GFAT que corresponde a la secuencia que se extiende entre los aminoácidos 220 a 230 de la GFAT de *Escherichia coli*.

15 2. Proteína según la reivindicación 1, en la que la secuencia de GFAT corresponde a una secuencia de GFAT de bacteria o de eucariota, en particular de planta, de hongo o de animal, en particular de insecto o de mamífero, más particularmente del ser humano.

20 3. Proteína según la reivindicación 1 ó 2, en la que la secuencia de la etiqueta de purificación se inserta entre dos aminoácidos consecutivos de la secuencia de una GFAT humana, estando dichos aminoácidos comprendidos entre los aminoácidos 40 a 50 y/o 290 a 330 de dicha secuencia de GFAT humana.

4. Proteína según una de las reivindicaciones 1 a 3, en la que la secuencia de GFAT corresponde a:

- 25
- SEC ID nº 2, que corresponde a la secuencia de la GFAT1 humana, o
  - SEC ID nº 4, que corresponde a la secuencia de la GFAT2 humana, o
  - SEC ID nº 6, que corresponde a la secuencia de la GFAT1Alt humana.
- 30

5. Proteína según una de las reivindicaciones 1 a 4, en la que la secuencia de la etiqueta de purificación se inserta entre dos aminoácidos consecutivos, estando dichos aminoácidos comprendidos entre los aminoácidos:

- 35
- 43 a 47 y/o 298 a 306, de SEC ID nº 2, o
  - 42 a 45 y/o 299 a 307, de SEC ID nº 4, o
  - 43 a 47 y/o 316 a 324, de SEC ID nº 6,

40 6. Proteína según la reivindicación 5, en la que la secuencia de la etiqueta de purificación se inserta entre los aminoácidos:

- 45
- 299 y 300 de SEC ID nº 2, o
  - 300 y 301 de SEC ID nº 4, o
  - 317 y 318 de SEC ID nº 6.

50 7. Proteína según una de las reivindicaciones 1 a 6, en la que la etiqueta de purificación corresponde a una secuencia de 2 a 10 aminoácidos, en particular de 4 a 8 aminoácidos.

8. Proteína según una de las reivindicaciones 1 a 7, en la que la etiqueta de purificación es una hexa-histidina.

55 9. Proteína según las reivindicaciones 6 y 8, que corresponde a las secuencias:

- 60
- SEC ID nº 8, que corresponde a la secuencia SEC ID nº 2 en la que una hexa-histidina se inserta entre los aminoácidos 299 y 300,
  - SEQ ID nº 10, que corresponde a la secuencia SEC ID nº 4 en la que una hexa-histidina se inserta entre los aminoácidos 300 y 301, y
  - SEQ ID nº 12, que corresponde a la secuencia SEC ID nº 6 en la que una hexa-histidina se inserta entre los aminoácidos 317 y 318.

65 10. Ácido nucleico que comprende o que está constituido por una secuencia que codifica para una proteína según una de las reivindicaciones 1 a 9.

## ES 2 343 571 T3

11. Ácido nucleico que comprende o que está constituido por la secuencia nucleotídica:

- SEC ID nº 7 que codifica para la proteína SEC ID nº 8, o
- SEC ID nº 9 que codifica para la proteína SEC ID nº 10, o
- SEC ID nº 11 que codifica para la proteína SEC ID nº 12.

12. Vector eucariota o procariota que comprende un ácido nucleico según la reivindicación 10 u 11.

13. Procedimiento de purificación de una proteína según una de las reivindicaciones 1 a 9, a partir de una disolución que comprende dicha proteína, que comprende una etapa de puesta en presencia de dicha disolución con un compuesto que se une específicamente a la etiqueta de purificación de dicha proteína, y una etapa de separación del complejo formado por la unión de dicha proteína a dicho compuesto de los demás constituyentes de la disolución.

14. Procedimiento de purificación según la reivindicación 13, que comprende una etapa de puesta en presencia de una disolución que comprende una proteína según la reivindicación 9, con un compuesto que comprende un catión metálico divalente seleccionado de entre el  $\text{Ni}^{2+}$  o el  $\text{Co}^{2+}$ , en particular el  $\text{Ni}^{2+}$ , y una etapa de separación del complejo formado por la unión de la proteína a dicho compuesto de los demás constituyentes de la disolución.

15. Procedimiento de conservación de una proteína según una de las reivindicaciones 1 a 9 en forma enzimáticamente activa, en particular a  $-80^{\circ}\text{C}$  o a  $4^{\circ}\text{C}$ , que comprende la adición de dicha proteína a una disolución que comprende:

- 1 mM a 10 mM de fructosa-6-fosfato, en particular aproximadamente 1 mM,
- 1 mM a 5 mM de Tris(2-carboxietil)fosfina, en particular aproximadamente 1 mM,
- 5% a 20% de glicerol, en particular aproximadamente 10%.

16. Composición que comprende una proteína GFAT activa unida a una etiqueta de purificación según una de las reivindicaciones 1 a 9, siendo dicha proteína susceptible de poder ser conservada en una forma enzimáticamente activa, durante por lo menos 8 días a una temperatura de  $2^{\circ}\text{C}$  a  $10^{\circ}\text{C}$ , en particular  $4^{\circ}\text{C}$ , y durante por lo menos 12 meses a una temperatura de  $-100^{\circ}\text{C}$  a  $-20^{\circ}\text{C}$ , en particular  $-80^{\circ}\text{C}$ , estando dicha proteína en asociación con:

- 1 mM a 10 mM de fructosa-6-fosfato, en particular 1 mM,
- 1 mM a 5 mM de Tris(2-carboxietil)fosfina, en particular 1 mM,
- 5% a 20% de glicerol, en particular 10%.

17. Utilización de una proteína según una de las reivindicaciones 1 a 9, para el cribado de compuestos que modifican la actividad de dicha proteína, en particular para el cribado de un inhibidor de dicha proteína.

18. Utilización según la reivindicación 17, para el cribado de compuestos útiles en el marco del tratamiento o de la prevención de la diabetes, en particular de la diabetes de tipo II, de la obesidad, de la acidosis, de la cetosis, de la artritis, de los cánceres, o de la osteoporosis.

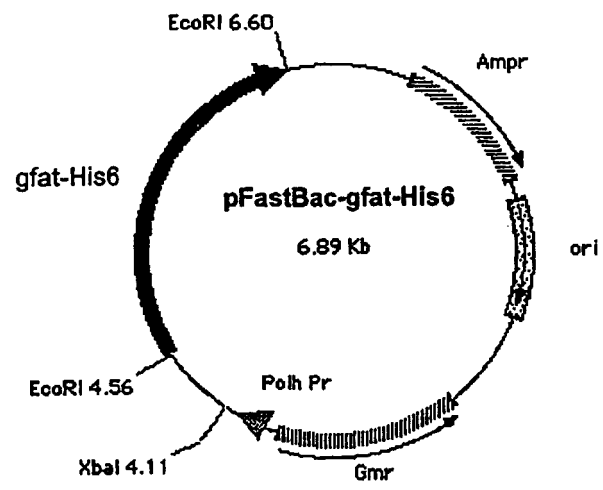


Figura 1

# ES 2 343 571 T3

## LISTA DE SECUENCIAS

- <110> CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE
- 5 <120> GLUTAMINA:FRUCTOSA-6-FOSFATO AMIDOTRANSFERASA (GFAT) QUE COMPRENDE UNA ETIQUETA DE PURIFICACIÓN INTERNA, Y SU UTILIZACIÓN PARA EL CRIBADO DE COMPUESTOS
- 10 <130> WOB 03 BP CNR GFAT
- <160>19
- <170> PatentIn versión 3.1
- 15 <210>1
- <211> 2046
- <212> ADN
- 20 <213> *Homo sapiens*
- <220>
- <221> CDS
- 25 <222> (1)..(2046)
- <223>
- <220>
- 30 <221> misc\_feature
- <222> (170)..(170)
- <223> t o c
- 35 <400> 1
- |    |   |     |
|----|---|-----|
|    | atg tgt ggt ata ttt gct tac tta aac tac cat gtt cct cga acg aga | 48  |
|    | Met Cys Gly Ile Phe Ala Tyr Leu Asn Tyr His Val Pro Arg Thr Arg |     |
| 40 | 1 5 10 15   |     |
|    | cga gaa atc ctg gag acc cta atc aaa ggc ctt cag aga ctg gag tac | 96  |
|    | Arg Glu Ile Leu Glu Thr Leu Ile Lys Gly Leu Gln Arg Leu Glu Tyr |     |
|    | 20 25 30  |     |
| 45 | aga gga tat gat tct gct ggt gtg gga ttt gat gga ggc aat gat aaa | 144 |
|    | Arg Gly Tyr Asp Ser Ala Gly Val Gly Phe Asp Gly Gly Asn Asp Lys |     |
|    | 35 40 45  |     |
|    | gat tgg gaa gcc aat gcc tgc aaa anc cag ctt att aag aag aaa gga | 192 |
|    | Asp Trp Glu Ala Asn Ala Cys Lys Xaa Gln Leu Ile Lys Lys Lys Gly |     |
| 50 | 50 55 60  |     |
|    | aaa gtt aag gca ctg gat gaa gaa gtt cac aag caa caa gat atg gat | 240 |
|    | Lys Val Lys Ala Leu Asp Glu Glu Val His Lys Gln Gln Asp Met Asp |     |
|    | 65 70 75 80   |     |
| 55 | ttg gat ata gaa ttt gat gta cac ctt gga ata gct cat acc cgt tgg | 288 |
|    | Leu Asp Ile Glu Phe Asp Val His Leu Gly Ile Ala His Thr Arg Trp |     |
|    | 85 90 95  |     |
|    | gca aca cat gga gaa ccc agt cct gtc aat agc cac ccc cag cgc tct | 336 |
|    | Ala Thr His Gly Glu Pro Ser Pro Val Asn Ser His Pro Gln Arg Ser |     |
| 60 | 100 105 110   |     |
|    | gat aaa aat aat gaa ttt atc gtt att cac aat gga atc atc acc aac | 384 |
|    | Asp Lys Asn Asn Glu Phe Ile Val Ile His Asn Gly Ile Ile Thr Asn |     |
|    | 115 120 125   |     |
| 65 | tac aaa gac ttg aaa aag ttt ttg gaa agc aaa ggc tat gac ttc gaa | 432 |
|    | Tyr Lys Asp Leu Lys Lys Phe Leu Glu Ser Lys Gly Tyr Asp Phe Glu |     |

ES 2 343 571 T3

	130					135						140					
5	tct	gaa	aca	gac	aca	gag	aca	att	gcc	aag	ctc	ggt	aag	tat	atg	tat	480
	Ser	Glu	Thr	Asp	Thr	Glu	Thr	Ile	Ala	Lys	Leu	Val	Lys	Tyr	Met	Tyr	
	145					150					155					160	
10	gac	aat	cgg	gaa	agt	caa	gat	acc	agc	ttt	act	acc	ttg	gtg	gag	aga	528
	Asp	Asn	Arg	Glu	Ser	Gln	Asp	Thr	Ser	Phe	Thr	Thr	Leu	Val	Glu	Arg	
				165						170					175		
15	gtt	atc	caa	caa	ttg	gaa	ggt	gct	ttt	gca	ctt	gtg	ttt	aaa	agt	gtt	576
	Val	Ile	Gln	Gln	Leu	Glu	Gly	Ala	Phe	Ala	Leu	Val	Phe	Lys	Ser	Val	
			180					185					190				
20	cat	ttt	ccc	ggg	caa	gca	ggt	ggc	aca	agg	cga	ggt	agc	cct	ctg	ttg	624
	His	Phe	Pro	Gly	Gln	Ala	Val	Gly	Thr	Arg	Arg	Gly	Ser	Pro	Leu	Leu	
			195				200						205				
25	att	ggt	gta	cgg	agt	gaa	cat	aaa	ctt	tct	act	gat	cac	att	cct	ata	672
	Ile	Gly	Val	Arg	Ser	Glu	His	Lys	Leu	Ser	Thr	Asp	His	Ile	Pro	Ile	
	210					215					220						
30	ctc	tac	aga	aca	ggc	aaa	gac	aag	aaa	gga	agc	tgc	aat	ctc	tct	cgt	720
	Leu	Tyr	Arg	Thr	Gly	Lys	Asp	Lys	Lys	Gly	Ser	Cys	Asn	Leu	Ser	Arg	
	225				230					235						240	
35	gtg	gac	agc	aca	acc	tgc	ctt	ttc	ccg	gtg	gaa	gaa	aaa	gca	gtg	gag	768
	Val	Asp	Ser	Thr	Thr	Cys	Leu	Phe	Pro	Val	Glu	Glu	Lys	Ala	Val	Glu	
				245					250						255		
40	tat	tac	ttt	gct	tct	gat	gca	agt	gct	gtc	ata	gaa	cac	acc	aat	cgc	816
	Tyr	Tyr	Phe	Ala	Ser	Asp	Ala	Ser	Ala	Val	Ile	Glu	His	Thr	Asn	Arg	
			260				265						270				
45	gtc	atc	ttt	ctg	gaa	gat	gat	gat	ggt	gca	gca	gta	gtg	gat	gga	cgt	864
	Val	Ile	Phe	Leu	Glu	Asp	Asp	Asp	Val	Ala	Ala	Val	Val	Asp	Gly	Arg	
		275				280							285				
50	ctt	tct	atc	cat	cga	att	aaa	cga	act	gca	gga	gat	cac	ccc	gga	cga	912
	Leu	Ser	Ile	His	Arg	Ile	Lys	Arg	Thr	Ala	Gly	Asp	His	Pro	Gly	Arg	
	290					295						300					
55	gct	gtg	caa	aca	ctc	cag	atg	gaa	ctc	cag	cag	atc	atg	aag	ggc	aac	960
	Ala	Val	Gln	Thr	Leu	Gln	Met	Glu	Leu	Gln	Gln	Ile	Met	Lys	Gly	Asn	
	305				310					315					320		
60	ttc	agt	tca	ttt	atg	cag	aag	gaa	ata	ttt	gag	cag	cca	gag	tct	gtc	1008
	Phe	Ser	Ser	Phe	Met	Gln	Lys	Glu	Ile	Phe	Glu	Gln	Pro	Glu	Ser	Val	
				325					330					335			
65	gtg	aac	aca	atg	aga	gga	aga	gtc	aac	ttt	gat	gac	tat	act	gtg	aat	1056
	Val	Asn	Thr	Met	Arg	Gly	Arg	Val	Asn	Phe	Asp	Asp	Tyr	Thr	Val	Asn	
			340					345					350				
70	ttg	ggt	ggt	ttg	aag	gat	cac	ata	aag	gag	atc	cag	aga	tgc	cgg	cgt	1104
	Leu	Gly	Gly	Leu	Lys	Asp	His	Ile	Lys	Glu	Ile	Gln	Arg	Cys	Arg	Arg	
		355					360					365					
75	ttg	att	ctt	att	gct	tgt	gga	aca	agt	tac	cat	gct	ggt	gta	gca	aca	1152
	Leu	Ile	Leu	Ile	Ala	Cys	Gly	Thr	Ser	Tyr	His	Ala	Gly	Val	Ala	Thr	
	370					375					380						
80	cgt	caa	gtt	ctt	gag	gag	ctg	act	gag	ttg	cct	gtg	atg	gtg	gaa	cta	1200
	Arg	Gln	Val	Leu	Glu	Glu	Leu	Thr	Glu	Leu	Pro	Val	Met	Val	Glu	Leu	
	385				390						395				400		

ES 2 343 571 T3

5	gca agt gac ttc ctg gac aga aac aca cca gtc ttt cga gat gat gtt Ala Ser Asp Phe Leu Asp Arg Asn Thr Pro Val Phe Arg Asp Asp Val	1248
	405 410 415	
10	tgc ttt ttc ctt agt caa tca ggt gag aca gca gat act ttg atg ggt Cys Phe Phe Leu Ser Gln Ser Gly Glu Thr Ala Asp Thr Leu Met Gly	1296
	420 425 430	
15	ctt cgt tac tgt aag gag aga gga gct tta act gtg ggg atc aca aac Leu Arg Tyr Cys Lys Glu Arg Gly Ala Leu Thr Val Gly Ile Thr Asn	1344
	435 440 445	
20	aca gtt ggc agt tcc ata tca cgg gag aca gat tgt gga gtt cat att Thr Val Gly Ser Ser Ile Ser Arg Glu Thr Asp Cys Gly Val His Ile	1392
	450 455 460	
25	aat gct ggt cct gag att ggt gtg gcc agt aca aag gct tat acc agc Asn Ala Gly Pro Glu Ile Gly Val Ala Ser Thr Lys Ala Tyr Thr Ser	1440
	465 470 475 480	
30	cag ttt gta tcc ctt gtg atg ttt gcc ctt atg atg tgt gat gat cgg Gln Phe Val Ser Leu Val Met Phe Ala Leu Met Met Cys Asp Asp Arg	1488
	485 490 495	
35	atc tcc atg caa gaa aga cgc aaa gag atc atg ctt gga ttg aaa cgg Ile Ser Met Gln Glu Arg Arg Lys Glu Ile Met Leu Gly Leu Lys Arg	1536
	500 505 510	
40	ctg cct gat ttg att aag gaa gta ctg agc atg gat gac gaa att cag Leu Pro Asp Leu Ile Lys Glu Val Leu Ser Met Asp Asp Glu Ile Gln	1584
	515 520 525	
45	aaa cta gca aca gaa ctt tat cat cag aag tca gtt ctg ata atg gga Lys Leu Ala Thr Glu Leu Tyr His Gln Lys Ser Val Leu Ile Met Gly	1632
	530 535 540	
50	cga ggc tat cat tat gct act tgt ctt gaa ggg gca ctg aaa atc aaa Arg Gly Tyr His Tyr Ala Thr Cys Leu Glu Gly Ala Leu Lys Ile Lys	1680
	545 550 555 560	
55	gaa att act tat atg cac tct gaa ggc atc ctt gct ggt gaa ttg aaa Glu Ile Thr Tyr Met His Ser Glu Gly Ile Leu Ala Gly Glu Leu Lys	1728
	565 570 575	
60	cat ggc cct ctg gct ttg gtg gat aaa ttg atg cct gtg atc atg atc His Gly Pro Leu Ala Leu Val Asp Lys Leu Met Pro Val Ile Met Ile	1776
	580 585 590	
65	atc atg aga gat cac act tat gcc aag tgt cag aat gct ctt cag caa Ile Met Arg Asp His Thr Tyr Ala Lys Cys Gln Asn Ala Leu Gln Gln	1824
	595 600 605	
70	gtg gtt gct cgg cag ggg cgg cct gtg gta att tgt gat aag gag gat Val Val Ala Arg Gln Gly Arg Pro Val Val Ile Cys Asp Lys Glu Asp	1872
	610 615 620	
75	act gag acc att aag aac aca aaa aga acg atc aag gtg ccc cac tca Thr Glu Thr Ile Lys Asn Thr Lys Arg Thr Ile Lys Val Pro His Ser	1920
	625 630 635 640	
80	gtg gac tgc ttg cag ggc att ctc agc gtg atc cct tta cag ttg ctg Val Asp Cys Leu Gln Gly Ile Leu Ser Val Ile Pro Leu Gln Leu Leu	1968
	645 650 655	

ES 2 343 571 T3

gct ttc cac ctt gct gtg ctg aga ggc tat gat gtt gat ttc cca egg 2016  
 Ala Phe His Leu Ala Val Leu Arg Gly Tyr Asp Val Asp Phe Pro Arg  
                   660                  665                  670

5 aat ctt gcc aaa tct gtg act gta gag tga 2046  
 Asn Leu Ala Lys Ser Val Thr Val Glu  
                   675                  680

10 <210>2  
 <211> 681  
 <212> PRT  
 <213> *Homo sapiens*

15 <220>  
 <221> misc\_feature  
 <222> (57)..(57)

20 <223> El "Xaa" en posición 57 representa Thr o Ile.

<400> 2

25 Met Cys Gly Ile Phe Ala Tyr Leu Asn Tyr His Val Pro Arg Thr Arg  
   1                  5                  10                  15

30 Arg Glu Ile Leu Glu Thr Leu Ile Lys Gly Leu Gln Arg Leu Glu Tyr  
                   20                  25                  30

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

40 Asp Trp Glu Ala Asn Ala Cys Lys Xaa Gln Leu Ile Lys Lys Lys Gly  
   50                  55                  60

45 Lys Val Lys Ala Leu Asp Glu Glu Val His Lys Gln Gln Asp Met Asp  
   65                  70                  75                  80

50 Leu Asp Ile Glu Phe Asp Val His Leu Gly Ile Ala His Thr Arg Trp  
                   85                  90                  95

55 Ala Thr His Gly Glu Pro Ser Pro Val Asn Ser His Pro Gln Arg Ser  
                   100                  105                  110

60 Asp Lys Asn Asn Glu Phe Ile Val Ile His Asn Gly Ile Ile Thr Asn  
                   115                  120                  125

65 Tyr Lys Asp Leu Lys Lys Phe Leu Glu Ser Lys Gly Tyr Asp Phe Glu  
   130                  135                  140

70 Ser Glu Thr Asp Thr Glu Thr Ile Ala Lys Leu Val Lys Tyr Met Tyr  
   145                  150                  155                  160

75 Asp Asn Arg Glu Ser Gln Asp Thr Ser Phe Thr Thr Leu Val Glu Arg  
                   165                  170                  175

ES 2 343 571 T3

Val Ile Gln Gln Leu Glu Gly Ala Phe Ala Leu Val Phe Lys Ser Val  
 180 185 190  
 5  
 His Phe Pro Gly Gln Ala Val Gly Thr Arg Arg Gly Ser Pro Leu Leu  
 195 200 205  
 10  
 Ile Gly Val Arg Ser Glu His Lys Leu Ser Thr Asp His Ile Pro Ile  
 210 215 220  
 15  
 Leu Tyr Arg Thr Gly Lys Asp Lys Lys Gly Ser Cys Asn Leu Ser Arg  
 225 230 235 240  
 20  
 Val Asp Ser Thr Thr Cys Leu Phe Pro Val Glu Glu Lys Ala Val Glu  
 245 250 255  
 25  
 Tyr Tyr Phe Ala Ser Asp Ala Ser Ala Val Ile Glu His Thr Asn Arg  
 260 265 270  
 30  
 Val Ile Phe Leu Glu Asp Asp Asp Val Ala Ala Val Val Asp Gly Arg  
 275 280 285  
 35  
 Leu Ser Ile His Arg Ile Lys Arg Thr Ala Gly Asp His Pro Gly Arg  
 290 295 300  
 40  
 Ala Val Gln Thr Leu Gln Met Glu Leu Gln Gln Ile Met Lys Gly Asn  
 305 310 315 320  
 45  
 Phe Ser Ser Phe Met Gln Lys Glu Ile Phe Glu Gln Pro Glu Ser Val  
 325 330 335  
 50  
 Val Asn Thr Met Arg Gly Arg Val Asn Phe Asp Asp Tyr Thr Val Asn  
 340 345 350  
 55  
 Leu Gly Gly Leu Lys Asp His Ile Lys Glu Ile Gln Arg Cys Arg Arg  
 355 360 365  
 60  
 Leu Ile Leu Ile Ala Cys Gly Thr Ser Tyr His Ala Gly Val Ala Thr  
 370 375 380  
 65  
 Arg Gln Val Leu Glu Glu Leu Thr Glu Leu Pro Val Met Val Glu Leu  
 385 390 395 400  
 Ala Ser Asp Phe Leu Asp Arg Asn Thr Pro Val Phe Arg Asp Asp Val  
 405 410 415  
 Cys Phe Phe Leu Ser Gln Ser Gly Glu Thr Ala Asp Thr Leu Met Gly  
 420 425 430  
 Leu Arg Tyr Cys Lys Glu Arg Gly Ala Leu Thr Val Gly Ile Thr Asn

ES 2 343 571 T3

		435				440						445				
5	Thr	Val	Gly	Ser	Ser	Ile	Ser	Arg	Glu	Thr	Asp	Cys	Gly	Val	His	Ile
		450					455					460				
10	Asn	Ala	Gly	Pro	Glu	Ile	Gly	Val	Ala	Ser	Thr	Lys	Ala	Tyr	Thr	Ser
	465					470					475					480
15	Gln	Phe	Val	Ser	Leu	Val	Met	Phe	Ala	Leu	Met	Met	Cys	Asp	Asp	Arg
					485					490					495	
20	Ile	Ser	Met	Gln	Glu	Arg	Arg	Lys	Glu	Ile	Met	Leu	Gly	Leu	Lys	Arg
				500					505					510		
25	Leu	Pro	Asp	Leu	Ile	Lys	Glu	Val	Leu	Ser	Met	Asp	Asp	Glu	Ile	Gln
			515					520					525			
30	Lys	Leu	Ala	Thr	Glu	Leu	Tyr	His	Gln	Lys	Ser	Val	Leu	Ile	Met	Gly
	530						535					540				
35	Arg	Gly	Tyr	His	Tyr	Ala	Thr	Cys	Leu	Glu	Gly	Ala	Leu	Lys	Ile	Lys
	545					550					555					560
40	Glu	Ile	Thr	Tyr	Met	His	Ser	Glu	Gly	Ile	Leu	Ala	Gly	Glu	Leu	Lys
					565					570					575	
45	His	Gly	Pro	Leu	Ala	Leu	Val	Asp	Lys	Leu	Met	Pro	Val	Ile	Met	Ile
				580					585					590		
50	Ile	Met	Arg	Asp	His	Thr	Tyr	Ala	Lys	Cys	Gln	Asn	Ala	Leu	Gln	Gln
			595					600					605			
55	Val	Val	Ala	Arg	Gln	Gly	Arg	Pro	Val	Val	Ile	Cys	Asp	Lys	Glu	Asp
		610					615					620				
60	Thr	Glu	Thr	Ile	Lys	Asn	Thr	Lys	Arg	Thr	Ile	Lys	Val	Pro	His	Ser
	625					630					635					640
65	Val	Asp	Cys	Leu	Gln	Gly	Ile	Leu	Ser	Val	Ile	Pro	Leu	Gln	Leu	Leu
					645					650					655	
70	Ala	Phe	His	Leu	Ala	Val	Leu	Arg	Gly	Tyr	Asp	Val	Asp	Phe	Pro	Arg
				660					665					670		
75	Asn	Leu	Ala	Lys	Ser	Val	Thr	Val	Glu							
			675					680								

60 <210> 3  
 <211> 2049  
 <212> ADN  
 <213> *Homo sapiens*  
 65  
 <220>  
 <221> CDS

ES 2 343 571 T3

<222> (1)..(2049)

<223>

5 <400> 3

	atg tgc gga atc ttt gcc tac atg aac tac aga gtc ccc cgg acg agg	48
	Met Cys Gly Ile Phe Ala Tyr Met Asn Tyr Arg Val Pro Arg Thr Arg	
	1 5 10 15	
10	aag gag atc ttc gaa acc ctc atc aag ggc ctg cag cgg ctg gag tac	96
	Lys Glu Ile Phe Glu Thr Leu Ile Lys Gly Leu Gln Arg Leu Glu Tyr	
	20 25 30	
15	aga ggc tac gac tgc gca ggt gtg gcg atc gat ggg aat aat cac gaa	144
	Arg Gly Tyr Asp Ser Ala Gly Val Ala Ile Asp Gly Asn Asn His Glu	
	35 40 45	
20	gtc aaa gaa aga cac att cag ctg gtc aag aaa agg ggg aaa gtc aag	192
	Val Lys Glu Arg His Ile Gln Leu Val Lys Lys Arg Gly Lys Val Lys	
	50 55 60	
25	gct ctc gat gaa gaa ctt tac aaa caa gac agc atg gac tta aaa gtg	240
	Ala Leu Asp Glu Glu Leu Tyr Lys Gln Asp Ser Met Asp Leu Lys Val	
	65 70 75 80	
30	gag ttt gag aca cac ttc ggc att gcc cac acg cgc tgg gcc acc cac	288
	Glu Phe Glu Thr His Phe Gly Ile Ala His Thr Arg Trp Ala Thr His	
	85 90 95	
35	ggg gtc ccc agt gct gtc aac agc cac cct cag cgc tca gac aaa ggc	336
	Gly Val Pro Ser Ala Val Asn Ser His Pro Gln Arg Ser Asp Lys Gly	
	100 105 110	
40	aac gaa ttt gtt gtc atc cac aat ggg atc atc aca aat tac aaa gat	384
	Asn Glu Phe Val Val Ile His Asn Gly Ile Ile Thr Asn Tyr Lys Asp	
	115 120 125	
45	ctg agg aaa ttt ctg gaa agc aaa ggc tac gag ttt gag tca gaa aca	432
	Leu Arg Lys Phe Leu Glu Ser Lys Gly Tyr Glu Phe Glu Ser Glu Thr	
	130 135 140	
50	gat aca gag acc atc gcc aag ctg att aaa tat gtg ttc gac aac aga	480
	Asp Thr Glu Thr Ile Ala Lys Leu Ile Lys Tyr Val Phe Asp Asn Arg	
	145 150 155 160	
55	gaa act gag gac att acg ttt tca acg ttg gtc gag aga gtc att cag	528
	Glu Thr Glu Asp Ile Thr Phe Ser Thr Leu Val Glu Arg Val Ile Gln	
	165 170 175	
60	cag ttg gaa ggt gca ttc gcg ctg gtt ttc aag agt gtc cac tac cca	576
	Gln Leu Glu Gly Ala Phe Ala Leu Val Phe Lys Ser Val His Tyr Pro	
	180 185 190	
65	gga gaa gcc gtt gcc aca cgg aga ggc agc ccc ctg ctc atc gga gtc	624
	Gly Glu Ala Val Ala Thr Arg Arg Gly Ser Pro Leu Leu Ile Gly Val	
	195 200 205	
70	cgg agc aaa tac aag ctc tcc aca gaa cag atc cct atc tta tac agg	672
	Arg Ser Lys Tyr Lys Leu Ser Thr Glu Gln Ile Pro Ile Leu Tyr Arg	
	210 215 220	

ES 2 343 571 T3

5 acg tgc act ctg gag aat gtg aag aat atc tgt aag aca cgg atg aag 720  
 Thr Cys Thr Leu Glu Asn Val Lys Asn Ile Cys Lys Thr Arg Met Lys  
 225 230 235 240

agg ctg gac agc tcc gcc tgc ctg cat gct gtg ggc gac aag gcc gtg 768  
 Arg Leu Asp Ser Ser Ala Cys Leu His Ala Val Gly Asp Lys Ala Val  
 245 250 255

10 gaa ttc ttc ttt gct tct gat gca agc gct atc ata gag cac acc aac 816  
 Glu Phe Phe Phe Ala Ser Asp Ala Ser Ala Ile Ile Glu His Thr Asn  
 260 265 270

15 cgg gtc atc ttc ctg gag gac gat gac atc gcc gca gtg gct gat ggg 864  
 Arg Val Ile Phe Leu Glu Asp Asp Ile Ala Ala Val Ala Asp Gly  
 275 280 285

20 aaa ctc tcc att cac cgg gtc aag cgc tcg gcc agt gat gac cca tct 912  
 Lys Leu Ser Ile His Arg Val Lys Arg Ser Ala Ser Asp Asp Pro Ser  
 290 295 300

25 cga gcc atc cag acc ttg cag atg gaa ctg cag caa atc atg aaa ggt 960  
 Arg Ala Ile Gln Thr Leu Gln Met Glu Leu Gln Gln Ile Met Lys Gly  
 305 310 315 320

aac ttc agt gcg ttt atg cag aag gag atc ttc gaa cag cca gaa tca 1008  
 Asn Phe Ser Ala Phe Met Gln Lys Glu Ile Phe Glu Gln Pro Glu Ser  
 325 330 335

30 gtt ttc aat act atg aga ggt cgg gtg aat ttt gaa acc aac aca gtg 1056  
 Val Phe Asn Thr Met Arg Gly Arg Val Asn Phe Glu Thr Asn Thr Val  
 340 345 350

35 ctc ctg ggt ggc ttg aag gac cac ttg aag gag att cga cga tgc cga 1104  
 Leu Leu Gly Gly Leu Lys Asp His Leu Lys Glu Ile Arg Arg Cys Arg  
 355 360 365

40 cgg ctc atc<sup>1</sup> gtg att ggc tgt gga acc agc tac cac gct gcc gtg gct 1152  
 Arg Leu Ile Val Ile Gly Cys Gly Thr Ser Tyr His Ala Ala Val Ala  
 370 375 380

45 acg cgg caa gtt ttg gag gaa ctg act gag ctt cct gtg atg gtt gaa 1200  
 Thr Arg Gln Val Leu Glu Glu Leu Thr Glu Leu Pro Val Met Val Glu  
 385 390 395 400

ctt gct agt gat ttt ctg gac agg aac aca cct gtg ttc agg gat gac 1248  
 Leu Ala Ser Asp Phe Leu Asp Arg Asn Thr Pro Val Phe Arg Asp Asp  
 405 410 415

50 gtt tgc ttt ttc atc agc cag tca ggc gag acc gcg gac acc ctc ctg 1296  
 Val Cys Phe Phe Ile Ser Gln Ser Gly Glu Thr Ala Asp Thr Leu Leu  
 420 425 430

55 gcg ctg cgc tac tgt aag gac cgc ggc gct ctc acc gtg ggc gtc acc 1344  
 Ala Leu Arg Tyr Cys Lys Asp Arg Gly Ala Leu Thr Val Gly Val Thr  
 435 440 445

aac acc gtg ggc agc tcc atc tct cgc gag acc gac tgc ggc gtc cac 1392  
 Asn Thr Val Gly Ser Ser Ile Ser Arg Glu Thr Asp Cys Gly Val His  
 450 455 460

60 atc aac gca ggg ccg gag gtc ggc gtg gcc agc acc aag gct tat acc 1440  
 Ile Asn Ala Gly Pro Glu Val Gly Val Ala Ser Thr Lys Ala Tyr Thr  
 465 470 475 480

65 agt cag ttc atc tct ctg gtg atg ttt ggt ttg atg atg tct gaa gac 1488

ES 2 343 571 T3

Ser Gln Phe Ile Ser Leu Val Met Phe Gly Leu Met Met Ser Glu Asp  
 485 490 495  
 5 cga att tca cta caa aac agg agg caa gag atc atc cgt ggc ttg aga 1536  
 Arg Ile Ser Leu Gln Asn Arg Arg Gln Glu Ile Ile Arg Gly Leu Arg  
 500 505 510  
 10 tct tta cct gag ctg atc aag gaa gtg ctg tct ctg gag gag aag atc 1584  
 Ser Leu Pro Glu Leu Ile Lys Glu Val Leu Ser Leu Glu Glu Lys Ile  
 515 520 525  
 15 cac gac ttg gcc ctg gag ctc tac acg cag aga tcg ctg ctg gtg atg 1632  
 His Asp Leu Ala Leu Glu Leu Tyr Thr Gln Arg Ser Leu Leu Val Met  
 530 535 540  
 20 ggg cgg ggc tac aac tat gcc acc tgc ctg gaa gga gcc ctg aaa att 1680  
 Gly Arg Gly Tyr Asn Tyr Ala Thr Cys Leu Glu Gly Ala Leu Lys Ile  
 545 550 555 560  
 25 aaa gag ata acc tac atg cac tca gaa ggc atc ctg gct ggg gag ctg 1728  
 Lys Glu Ile Thr Tyr Met His Ser Glu Gly Ile Leu Ala Gly Glu Leu  
 565 570 575  
 30 aag cac ggg ccc ctg gca ctg att gac aag cag atg ccc gtc atc atg 1776  
 Lys His Gly Pro Leu Ala Leu Ile Asp Lys Gln Met Pro Val Ile Met  
 580 585 590  
 35 gtc att atg aag gat cct tgc ttc gcc aaa tgc cag aac gcc ctg cag 1824  
 Val Ile Met Lys Asp Pro Cys Phe Ala Lys Cys Gln Asn Ala Leu Gln  
 595 600 605  
 40 caa gtc acg gcc cgc cag ggt cgc ccc att ata ctg tgc tcc aag gac 1872  
 Gln Val Thr Ala Arg Gln Gly Arg Pro Ile Ile Leu Cys Ser Lys Asp  
 610 615 620  
 45 gat act gaa agt tcc aag ttt gcg tat aag aca atc gag ctg ccc cac 1920  
 Asp Thr Glu Ser Ser Lys Phe Ala Tyr Lys Thr Ile Glu Leu Pro His  
 625 630 635 640  
 50 act gtg gac tgc ctc cag ggc atc ctg agc gtg att ccg ctg cag ctg 1968  
 Thr Val Asp Cys Leu Gln Gly Ile Leu Ser Val Ile Pro Leu Gln Leu  
 645 650 655  
 55 ctg tcc ttc cac ctg gct gtt ctc cga gga tat gac gtt gac ttc ccc 2016  
 Leu Ser Phe His Leu Ala Val Leu Arg Gly Tyr Asp Val Asp Phe Pro  
 660 665 670  
 60 aga aat ctg gcc aag tct gta act gtg gaa tga 2049  
 Arg Asn Leu Ala Lys Ser Val Thr Val Glu  
 675 680

<210> 4

<211> 682

<212> PRT

<213> *Homo sapiens*

55 <400> 4  
 Met Cys Gly Ile Phe Ala Tyr Met Asn Tyr Arg Val Pro Arg Thr Arg  
 1 5 10 15  
 60 Lys Glu Ile Phe Glu Thr Leu Ile Lys Gly Leu Gln Arg Leu Glu Tyr  
 20 25 30  
 65

ES 2 343 571 T3

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

ES 2 343 571 T3

5 Lys Leu Ser Ile His Arg Val Lys Arg Ser Ala Ser Asp Asp Pro Ser  
290 295 300

10 Arg Ala Ile Gln Thr Leu Gln Met Glu Leu Gln Gln Ile Met Lys Gly  
305 310 315 320

15 Asn Phe Ser Ala Phe Met Gln Lys Glu Ile Phe Glu Gln Pro Glu Ser  
325 330 335

20 Val Phe Asn Thr Met Arg Gly Arg Val Asn Phe Glu Thr Asn Thr Val  
340 345 350

25 Leu Leu Gly Gly Leu Lys Asp His Leu Lys Glu Ile Arg Arg Cys Arg  
355 360 365

30 Arg Leu Ile Val Ile Gly Cys Gly Thr Ser Tyr His Ala Ala Val Ala  
370 375 380

35 Thr Arg Gln Val Leu Glu Glu Leu Thr Glu Leu Pro Val Met Val Glu  
385 390 395 400

40 Leu Ala Ser Asp Phe Leu Asp Arg Asn Thr Pro Val Phe Arg Asp Asp  
405 410 415

45 Val Cys Phe Phe Ile Ser Gln Ser Gly Glu Thr Ala Asp Thr Leu Leu  
420 425 430

50 Ala Leu Arg Tyr Cys Lys Asp Arg Gly Ala Leu Thr Val Gly Val Thr  
435 440 445

55 Asn Thr Val Gly Ser Ser Ile Ser Arg Glu Thr Asp Cys Gly Val His  
450 455 460

60 Ile Asn Ala Gly Pro Glu Val Gly Val Ala Ser Thr Lys Ala Tyr Thr  
465 470 475 480

65 Ser Gln Phe Ile Ser Leu Val Met Phe Gly Leu Met Met Ser Glu Asp  
485 490 495

Arg Ile Ser Leu Gln Asn Arg Arg Gln Glu Ile Ile Arg Gly Leu Arg  
500 505 510

Ser Leu Pro Glu Leu Ile Lys Glu Val Leu Ser Leu Glu Glu Lys Ile  
515 520 525

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

ES 2 343 571 T3

5 Gly Arg Gly Tyr Asn Tyr Ala Thr Cys Leu Glu Gly Ala Leu Lys Ile  
545 550 555 560

Lys Glu Ile Thr Tyr Met His Ser Glu Gly Ile Leu Ala Gly Glu Leu  
565 570 575

10 Lys His Gly Pro Leu Ala Leu Ile Asp Lys Gln Met Pro Val Ile Met  
580 585 590

15 Val Ile Met Lys Asp Pro Cys Phe Ala Lys Cys Gln Asn Ala Leu Gln  
595 600 605

Gln Val Thr Ala Arg Gln Gly Arg Pro Ile Ile Leu Cys Ser Lys Asp  
610 615 620

20 Asp Thr Glu Ser Ser Lys Phe Ala Tyr Lys Thr Ile Glu Leu Pro His  
625 630 635 640

25 Thr Val Asp Cys Leu Gln Gly Ile Leu Ser Val Ile Pro Leu Gln Leu  
645 650 655

30 Leu Ser Phe His Leu Ala Val Leu Arg Gly Tyr Asp Val Asp Phe Pro  
660 665 670

Arg Asn Leu Ala Lys Ser Val Thr Val Glu  
675 680

35 <210> 5  
<211> 2100  
<212> ADN

40 <213> *Homo sapiens*  
<220>  
45 <221> CDS  
<222> (1)..(2100)  
<223>

50 <220>  
<221> misc\_feature  
<222> (170)..(170)  
<223> t ou c

55 <400> 5

60 atg tgt ggt ata ttt gct tac tta aac tac cat gtt cct cga acg aga 48  
Met Cys Gly Ile Phe Ala Tyr Leu Asn Tyr His Val Pro Arg Thr Arg  
1 5 10 15

cga gaa atc ctg gag acc cta atc aaa ggc ctt cag aga ctg gag tac 96  
Arg Glu Ile Leu Glu Thr Leu Ile Lys Gly Leu Gln Arg Leu Glu Tyr  
20 25 30

65 aga gga tat gat tct gct ggt gtg gga ttt gat gga ggc aat gat aaa 144  
Arg Gly Tyr Asp Ser Ala Gly Val Gly Phe Asp Gly Gly Asn Asp Lys  
35 40 45

# ES 2 343 571 T3

5	gat tgg gaa gcc aat gcc tgc aaa anc cag ctt att aag aag aaa gga Asp Trp Glu Ala Asn Ala Cys Lys Xaa Gln Leu Ile Lys Lys Lys Gly 50 55 60	192
10	aaa gtt aag gca ctg gat gaa gaa gtt cac aag caa caa gat atg gat Lys Val Lys Ala Leu Asp Glu Glu Val His Lys Gln Gln Asp Met Asp 65 70 75 80	240
15	ttg gat ata gaa ttt gat gta cac ctt gga ata gct cat acc cgt tgg Leu Asp Ile Glu Phe Asp Val His Leu Gly Ile Ala His Thr Arg Trp 85 90 95	288
20	gca aca cat gga gaa ccc agt cct gtc aat agc cac ccc cag cgc tct Ala Thr His Gly Glu Pro Ser Pro Val Asn Ser His Pro Gln Arg Ser 100 105 110	336
25	gat aaa aat aat gaa ttt atc gtt att cac aat gga atc atc acc aac Asp Lys Asn Asn Glu Phe Ile Val Ile His Asn Gly Ile Ile Thr Asn 115 120 125	384
30	tac aaa gac ttg aaa aag ttt ttg gaa agc aaa ggc tat gac ttc gaa Tyr Lys Asp Leu Lys Lys Phe Leu Glu Ser Lys Gly Tyr Asp Phe Glu 130 135 140	432
35	tct gaa aca gac aca gag aca att gcc aag ctc gtt aag tat atg tat Ser Glu Thr Asp Thr Glu Thr Ile Ala Lys Leu Val Lys Tyr Met Tyr 145 150 155 160	480
40	gac aat cgg gaa agt caa gat acc agc ttt act acc ttg gtg gag aga Asp Asn Arg Glu Ser Gln Asp Thr Ser Phe Thr Thr Leu Val Glu Arg 165 170 175	528
45	gtt atc caa caa ttg gaa ggt gct ttt gca ctt gtg ttt aaa agt gtt Val Ile Gln Gln Leu Glu Gly Ala Phe Ala Leu Val Phe Lys Ser Val 180 185 190	576
50	cat ttt ccc ggg caa gca gtt ggc aca agg cga ggt agc cct ctg ttg His Phe Pro Gly Gln Ala Val Gly Thr Arg Arg Gly Ser Pro Leu Leu 195 200 205	624
55	att ggt gta cgg agt gaa cat aaa ctt tct act gat cac att cct ata Ile Gly Val Arg Ser Glu His Lys Leu Ser Thr Asp His Ile Pro Ile 210 215 220	672
60	ctc tac aga aca gct agg act cag att gga tca aaa ttc aca cgg tgg Leu Tyr Arg Thr Ala Arg Thr Gln Ile Gly Ser Lys Phe Thr Arg Trp 225 230 235 240	720
65	gga tca cag gga gaa aga ggc aaa gac aag aaa gga agc tgc aat ctc Gly Ser Gln Gly Glu Arg Gly Lys Asp Lys Lys Gly Ser Cys Asn Leu 245 250 255	768
70	tct cgt gtg gac agc aca acc tgc ctt ttc cgg gtg gaa gaa aaa gca Ser Arg Val Asp Ser Thr Thr Cys Leu Phe Pro Val Glu Glu Lys Ala 260 265 270	816
75	gtg gag tat tac ttt gct tct gat gca agt gct gtc ata gaa cac acc Val Glu Tyr Tyr Phe Ala Ser Asp Ala Ser Ala Val Ile Glu His Thr 275 280 285	864
80	aat cgc gtc atc ttt ctg gaa gat gat gat gtt gca gca gta gtg gat Asn Arg Val Ile Phe Leu Glu Asp Asp Asp Val Ala Ala Val Val Asp 290 295 300	912
85	gga cgt ctt tct atc cat cga att aaa cga act gca gga gat cac ccc	960

ES 2 343 571 T3

	Gly	Arg	Leu	Ser	Ile	His	Arg	Ile	Lys	Arg	Thr	Ala	Gly	Asp	His	Pro	
	305					310					315					320	
5	gga	cga	gct	gtg	caa	aca	ctc	cag	atg	gaa	ctc	cag	cag	atc	atg	aag	1008
	Gly	Arg	Ala	Val	Gln	Thr	Leu	Gln	Met	Glu	Leu	Gln	Gln	Ile	Met	Lys	
					325					330					335		
10	ggc	aac	ttc	agt	tca	ttt	atg	cag	aag	gaa	ata	ttt	gag	cag	cca	gag	1056
	Gly	Asn	Phe	Ser	Ser	Phe	Met	Gln	Lys	Glu	Ile	Phe	Glu	Gln	Pro	Glu	
				340					345					350			
15	tct	gtc	gtg	aac	aca	atg	aga	gga	aga	gtc	aac	ttt	gat	gac	tat	act	1104
	Ser	Val	Val	Asn	Thr	Met	Arg	Gly	Arg	Val	Asn	Phe	Asp	Asp	Tyr	Thr	
			355					360					365				
20	gtg	aat	ttg	ggt	ggt	ttg	aag	gat	cac	ata	aag	gag	atc	cag	aga	tgc	1152
	Val	Asn	Leu	Gly	Gly	Leu	Lys	Asp	His	Ile	Lys	Glu	Ile	Gln	Arg	Cys	
		370					375					380					
25	cgg	cgt	ttg	att	ctt	att	gct	tgt	gga	aca	agt	tac	cat	gct	ggt	gta	1200
	Arg	Arg	Leu	Ile	Leu	Ile	Ala	Cys	Gly	Thr	Ser	Tyr	His	Ala	Gly	Val	
	385					390					395				400		
30	gca	aca	cgt	caa	gtt	ctt	gag	gag	ctg	act	gag	ttg	cct	gtg	atg	gtg	1248
	Ala	Thr	Arg	Gln	Val	Leu	Glu	Glu	Leu	Thr	Glu	Leu	Pro	Val	Met	Val	
				405					410						415		
35	gat	gtt	tgc	ttt	ttc	ctt	agt	caa	tca	ggt	gag	aca	gca	gat	act	ttg	1344
	Asp	Val	Cys	Phe	Phe	Leu	Ser	Gln	Ser	Gly	Glu	Thr	Ala	Asp	Thr	Leu	
			435					440					445				
40	atg	ggt	ctt	cgt	tac	tgt	aag	gag	aga	gga	gct	tta	act	gtg	ggg	atc	1392
	Met	Gly	Leu	Arg	Tyr	Cys	Lys	Glu	Arg	Gly	Ala	Leu	Thr	Val	Gly	Ile	
		450					455					460					
45	aca	aac	aca	gtt	ggc	agt	tcc	ata	tca	cgg	gag	aca	gat	tgt	gga	gtt	1440
	Thr	Asn	Thr	Val	Gly	Ser	Ser	Ile	Ser	Arg	Glu	Thr	Asp	Cys	Gly	Val	
	465					470				475					480		
50	cat	att	aat	gct	ggt	cct	gag	att	ggt	gtg	gcc	agt	aca	aag	gct	tat	1488
	His	Ile	Asn	Ala	Gly	Pro	Glu	Ile	Gly	Val	Ala	Ser	Thr	Lys	Ala	Tyr	
					485					490					495		
55	acc	agc	cag	ttt	gta	tcc	ctt	gtg	atg	ttt	gcc	ctt	atg	atg	tgt	gat	1536
	Thr	Ser	Gln	Phe	Val	Ser	Leu	Val	Met	Phe	Ala	Leu	Met	Met	Cys	Asp	
				500					505					510			
60	gat	cgg	atc	tcc	atg	caa	gaa	aga	cgc	aaa	gag	atc	atg	ctt	gga	ttg	1584
	Asp	Arg	Ile	Ser	Met	Gln	Glu	Arg	Arg	Lys	Glu	Ile	Met	Leu	Gly	Leu	
			515					520					525				
65	aaa	cgg	ctg	cct	gat	ttg	att	aag	gaa	gta	ctg	agc	atg	gat	gac	gaa	1632
	Lys	Arg	Leu	Pro	Asp	Leu	Ile	Lys	Glu	Val	Leu	Ser	Met	Asp	Asp	Glu	
			530					535				540					
70	att	cag	aaa	cta	gca	aca	gaa	ctt	tat	cat	cag	aag	tca	ggt	ctg	ata	1680
	Ile	Gln	Lys	Leu	Ala	Thr	Glu	Leu	Tyr	His	Gln	Lys	Ser	Val	Leu	Ile	
	545					550					555				560		
75	atg	gga	cga	ggc	tat	cat	tat	gct	act	tgt	ctt	gaa	ggg	gca	ctg	aaa	1728
	Met	Gly	Arg	Gly	Tyr	His	Tyr	Ala	Thr	Cys	Leu	Glu	Gly	Ala	Leu	Lys	

# ES 2 343 571 T3

	565	570	575	
5	atc aaa gaa att act tat atg cac tct gaa ggc atc ctt gct ggt gaa Ile Lys Glu Ile Thr Tyr Met His Ser Glu Gly Ile Leu Ala Gly Glu 580 585 590			1776
	ttg aaa cat ggc cct ctg gct ttg gtg gat aaa ttg atg cct gtg atc Leu Lys His Gly Pro Leu Ala Leu Val Asp Lys Leu Met Pro Val Ile 595 600 605			1824
10	atg atc atc atg aga gat cac act tat gcc aag tgt cag aat gct ctt Met Ile Ile Met Arg Asp His Thr Tyr Ala Lys Cys Gln Asn Ala Leu 610 615 620			1872
15	cag caa gtg gtt gct cgg cag ggg cgg cct gtg gta att tgt gat aag Gln Gln Val Val Ala Arg Gln Gly Arg Pro Val Val Ile Cys Asp Lys 625 630 635 640			1920
20	gag gat act gag acc att aag aac aca aaa aga acg atc aag gtg ccc Glu Asp Thr Glu Thr Ile Lys Asn Thr Lys Arg Thr Ile Lys Val Pro 645 650 655			1968
25	cac tca gtg gac tgc ttg cag ggc att ctc agr gtg atc cct tta cag His Ser Val Asp Cys Leu Gln Gly Ile Leu Ser Val Ile Pro Leu Gln 660 665 670			2016
30	ttg ctg gct ttc cac ctt gct gtg ctg aga ggc tat gat gtt gat ttc Leu Leu Ala Phe His Leu Ala Val Leu Arg Gly Tyr Asp Val Asp Phe 675 680 685			2064
	cca cgg aat ctt gcc aaa tct gtg act gta gag tga Pro Arg Asn Leu Ala Lys Ser Val Thr Val Glu 690 695			2100

<210> 6  
 <211> 699  
 <212> PRT  
 <213> *Homo sapiens*

<220>  
 <221> misc\_feature  
 <222> (57)..(57)  
 <223> El "Xaa" en posición 57 representa Thr o Ile.

<400> 6

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

ES 2 343 571 T3

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

ES 2 343 571 T3

5 Gly Asn Phe Ser Ser Phe Met Gln Lys Glu Ile Phe Glu Gln Pro Glu  
340 345 350

10 Ser Val Val Asn Thr Met Arg Gly Arg Val Asn Phe Asp Asp Tyr Thr  
355 360 365

15 Val Asn Leu Gly Gly Leu Lys Asp His Ile Lys Glu Ile Gln Arg Cys  
370 375 380

20 Arg Arg Leu Ile Leu Ile Ala Cys Gly Thr Ser Tyr His Ala Gly Val  
385 390 395 400

25 Ala Thr Arg Gln Val Leu Glu Glu Leu Thr Glu Leu Pro Val Met Val  
405 410 415

30 Glu Leu Ala Ser Asp Phe Leu Asp Arg Asn Thr Pro Val Phe Arg Asp  
420 425 430

35 Asp Val Cys Phe Phe Leu Ser Gln Ser Gly Glu Thr Ala Asp Thr Leu  
435 440 445

40 Met Gly Leu Arg Tyr Cys Lys Glu Arg Gly Ala Leu Thr Val Gly Ile  
450 455 460

45 Thr Asn Thr Val Gly Ser Ser Ile Ser Arg Glu Thr Asp Cys Gly Val  
465 470 475 480

50 His Ile Asn Ala Gly Pro Glu Ile Gly Val Ala Ser Thr Lys Ala Tyr  
485 490 495

55 Thr Ser Gln Phe Val Ser Leu Val Met Phe Ala Leu Met Met Cys Asp  
500 505 510

60 Asp Arg Ile Ser Met Gln Glu Arg Arg Lys Glu Ile Met Leu Gly<sup>1</sup>Leu  
515 520 525

65 Lys Arg Leu Pro Asp Leu Ile Lys Glu Val Leu Ser Met Asp Asp Glu  
530 535 540

70 Ile Gln Lys Leu Ala Thr Glu Leu Tyr His Gln Lys Ser Val Leu Ile  
545 550 555 560

75 Met Gly Arg Gly Tyr His Tyr Ala Thr Cys Leu Glu Gly Ala Leu Lys  
565 570 575

80 Ile Lys Glu Ile Thr Tyr Met His Ser Glu Gly Ile Leu Ala Gly Glu  
580 585 590

# ES 2 343 571 T3

```

Leu Lys His Gly Pro Leu Ala Leu Val Asp Lys Leu Met Pro Val Ile
      595                                600                                605

5  Met Ile Ile Met Arg Asp His Thr Tyr Ala Lys Cys Gln Asn Ala Leu
      610                                615                                620

Gln Gln Val Val Ala Arg Gln Gly Arg Pro Val Val Ile Cys Asp Lys
10  625                                630                                635                                640

Glu Asp Thr Glu Thr Ile Lys Asn Thr Lys Arg Thr Ile Lys Val Pro
      645                                650                                655

15  His Ser Val Asp Cys Leu Gln Gly Ile Leu Ser Val Ile Pro Leu Gln
      660                                665                                670

Leu Leu Ala Phe His Leu Ala Val Leu Arg Gly Tyr Asp Val Asp Phe
20  675                                680                                685

Pro Arg Asn Leu Ala Lys Ser Val Thr Val Glu
      690                                695

25  <210> 7
      <211> 2064
      <212> ADN
30  <213> Secuencia artificial

      <220>
      <223> GFAT1 modificada por una etiqueta de purificación interna
35  <220>
      <221> CDS
      <222> (1)..(2064)
40  <223>

      <220>
      <221> misc_feature
45  <222> (170)..(170)
      <223> t o c

50  <400> 7
      atg tgt ggt ata ttt gct tac tta aac tac cat gtt cct cga acg aga      48
      Met Cys Gly Ile Phe Ala Tyr Leu Asn Tyr His Val Pro Arg Thr Arg
      1      5      10      15

55  cga gaa atc ctg gag acc cta atc aaa ggc ctt cag aga ctg gag tac      96
      Arg Glu Ile Leu Glu Thr Leu Ile Lys Gly Leu Gln Arg Leu Glu Tyr
      20      25      30

      aga gga tat gat tct gct ggt gtg gga ttt gat gga ggc aat gat aaa      144
      Arg Gly Tyr Asp Ser Ala Gly Val Gly Phe Asp Gly Gly Asn Asp Lys
      35      40      45

      gat tgg gaa gcc aat gcc tgc aaa anc cag ctt att aag aag aaa gga      192
      Asp Trp Glu Ala Asn Ala Cys Lys Xaa Gln Leu Ile Lys Lys Lys Gly
      50      55      60

65  aaa gtt aag gca ctg gat gaa gaa gtt cac aag caa caa gat atg gat      240

```

ES 2 343 571 T3

	Lys	Val	Lys	Ala	Leu	Asp	Glu	Glu	Val	His	Lys	Gln	Gln	Asp	Met	Asp	
	65					70					75					80	
5	ttg	gat	ata	gaa	ttt	gat	gta	cac	ctt	gga	ata	gct	cat	acc	cgt	tgg	288
	Leu	Asp	Ile	Glu	Phe	Asp	Val	His	Leu	Gly	Ile	Ala	His	Thr	Arg	Trp	
				85						90					95		
10	gca	aca	cat	gga	gaa	ccc	agt	cct	gtc	aat	agc	cac	ccc	cag	cgc	tct	336
	Ala	Thr	His	Gly	Glu	Pro	Ser	Pro	Val	Asn	Ser	His	Pro	Gln	Arg	Ser	
				100					105					110			
15	gat	aaa	aat	aat	gaa	ttt	atc	gtt	att	cac	aat	gga	atc	atc	acc	aac	384
	Asp	Lys	Asn	Asn	Glu	Phe	Ile	Val	Ile	His	Asn	Gly	Ile	Ile	Thr	Asn	
			115					120					125				
20	tac	aaa	gac	ttg	aaa	aag	ttt	ttg	gaa	agc	aaa	ggc	tat	gac	ttc	gaa	432
	Tyr	Lys	Asp	Leu	Lys	Lys	Phe	Leu	Glu	Ser	Lys	Gly	Tyr	Asp	Phe	Glu	
		130					135					140					
25	tct	gaa	aca	gac	aca	gag	aca	att	gcc	aag	ctc	gtt	aag	tat	atg	tat	480
	Ser	Glu	Thr	Asp	Thr	Glu	Thr	Ile	Ala	Lys	Leu	Val	Lys	Tyr	Met	Tyr	
		145				150					155					160	
30	gac	aat	cgg	gaa	agt	caa	gat	acc	agc	ttt	act	acc	ttg	gtg	gag	aga	528
	Asp	Asn	Arg	Glu	Ser	Gln	Asp	Thr	Ser	Phe	Thr	Thr	Leu	Val	Glu	Arg	
					165					170					175		
35	gtt	atc	caa	caa	ttg	gaa	ggt	gct	ttt	gca	ctt	gtg	ttt	aaa	agt	gtt	576
	Val	Ile	Gln	Gln	Leu	Glu	Gly	Ala	Phe	Ala	Leu	Val	Phe	Lys	Ser	Val	
				180					185					190			
40	cat	ttt	ccc	ggg	caa	gca	gtt	ggc	aca	agg	cga	ggg	agc	cct	ctg	ttg	624
	His	Phe	Pro	Gly	Gln	Ala	Val	Gly	Thr	Arg	Arg	Gly	Ser	Pro	Leu	Leu	
			195					200					205				
45	att	ggt	gta	cgg	agt	gaa	cat	aaa	ctt	tct	act	gat	cac	att	cct	ata	672
	Ile	Gly	Val	Arg	Ser	Glu	His	Lys	Leu	Ser	Thr	Asp	His	Ile	Pro	Ile	
			210				215					220					
50	ctc	tac	aga	aca	ggc	aaa	gac	aag	aaa	gga	agc	tgc	aat	ctc	tct	cgt	720
	Leu	Tyr	Arg	Thr	Gly	Lys	Asp	Lys	Lys	Gly	Ser	Cys	Asn	Leu	Ser	Arg	
						230					235					240	
55	gtg	gac	agc	aca	acc	tgc	ctt	ttc	ccg	gtg	gaa	gaa	aaa	gca	gtg	gag	768
	Val	Asp	Ser	Thr	Thr	Cys	Leu	Phe	Pro	Val	Glu	Glu	Lys	Ala	Val	Glu	
					245					250					255		
60	tat	tac	ttt	gct	tct	gat	gca	agt	gct	gtc	ata	gaa	cac	acc	aat	cgc	816
	Tyr	Tyr	Phe	Ala	Ser	Asp	Ala	Ser	Ala	Val	Ile	Glu	His	Thr	Asn	Arg	
				260					265					270			
65	gtc	atc	ttt	ctg	gaa	gat	gat	gat	ggt	gca	gca	gta	gtg	gat	gga	cgt	864
	Val	Ile	Phe	Leu	Glu	Asp	Asp	Asp	Val	Ala	Ala	Val	Val	Val	Asp	Gly	
				275				280						285			
70	ctt	tct	atc	cat	cga	att	aaa	cga	act	gca	gga	cat	cac	cat	cac	cat	912
	Leu	Ser	Ile	His	Arg	Ile	Lys	Arg	Thr	Ala	Gly	His	His	His	His	His	
			290					295				300					
75	cac	gat	cac	ccc	gga	cga	gct	gtg	caa	aca	ctc	cag	atg	gaa	ctc	cag	960
	His	Asp	His	Pro	Gly	Arg	Ala	Val	Gln	Thr	Leu	Gln	Met	Glu	Leu	Gln	
						310					315					320	
80	cag	atc	atg	aag	ggc	aac	ttc	agt	tca	ttt	atg	cag	aag	gaa	ata	ttt	1008
	Gln	Ile	Met	Lys	Gly	Asn	Phe	Ser	Ser	Phe	Met	Gln	Lys	Glu	Ile	Phe	

ES 2 343 571 T3

	325					330					335						
5	gag Glu	cag Gln	cca Pro	gag Glu	tct Ser	gtc Val	gtg Val	aac Asn	aca Thr	atg Met	aga Arg	gga Gly	aga Arg	gtc Val	aac Asn	ttt Phe	1056
			340					345						350			
10	gat Asp	gac Asp	tat Tyr	act Thr	gtg Val	aat Asn	ttg Leu	ggt Gly	ggt Gly	ttg Leu	aag Lys	gat Asp	cac His	ata Ile	aag Lys	gag Glu	1104
			355					360						365			
15	atc Ile	cag Gln	aga Arg	tgc Cys	cgg Arg	cgt Arg	ttg Leu	att Ile	ctt Leu	att Ile	gct Ala	tgt Cys	gga Gly	aca Thr	agt Ser	tac Tyr	1152
			370				375					380					
20	cat His	gct Ala	ggt Gly	gta Val	gca Ala	aca Thr	cgt Arg	caa Gln	ggt Val	ctt Leu	gag Glu	gag Glu	ctg Leu	act Thr	gag Glu	ttg Leu	1200
			385			390					395					400	
25	cct Pro	gtg Val	atg Met	gtg Val	gaa Glu	cta Leu	gca Ala	agt Ser	gac Asp	ttc Phe	ctg Leu	gac Asp	aga Arg	aac Asn	aca Thr	cca Pro	1248
				405						410				415			
30	gtc Val	ttt Phe	cga Arg	gat Asp	gat Asp	ggt Val	tgc Cys	ttt Phe	ttc Phe	ctt Leu	agt Ser	caa Gln	tca Ser	ggt Gly	gag Glu	aca Thr	1296
				420				425						430			
35	gca Ala	gat Asp	act Thr	ttg Leu	atg Met	ggt Gly	ctt Leu	cgt Arg	tac Tyr	tgt Cys	aag Lys	gag Glu	aga Arg	gga Gly	gct Ala	tta Leu	1344
			435				440						445				
40	act Thr	gtg Val	ggg Gly	atc Ile	aca Thr	aac Asn	aca Thr	ggt Val	ggc Gly	agt Ser	tcc Ser	ata Ile	tca Ser	cgg Arg	gag Glu	aca Thr	1392
			450				455				460						
45	gat Asp	tgt Cys	gga Gly	gtt Val	cat His	att Ile	aat Asn	gct Ala	ggt Gly	cct Pro	gag Glu	att Ile	ggt Gly	gtg Val	gcc Ala	agt Ser	1440
			465			470				475						480	
50	aca Thr	aag Lys	gct Ala	tat Tyr	acc Thr	agc Ser	cag Gln	ttt Phe	gta Val	tcc Ser	ctt Leu	gtg Val	atg Met	ttt Phe	gcc Ala	ctt Leu	1488
				485					490					495			
55	atg Met	atg Met	tgt Cys	gat Asp	gat Asp	cgg Arg	atc Ile	tcc Ser	atg Met	caa Gln	gaa Glu	aga Arg	cgc Arg	aaa Lys	gag Glu	atc Ile	1536
			500				505						510				
60	atg Met	ctt Leu	gga Gly	ttg Leu	aaa Lys	cgg Arg	ctg Leu	cct Pro	gat Asp	ttg Leu	att Ile	aag Lys	gaa Glu	gta Val	ctg Leu	agc Ser	1584
			515				520						525				
65	atg Met	gat Asp	gac Asp	gaa Glu	att Ile	cag Gln	aaa Lys	cta Leu	gca Ala	aca Thr	gaa Glu	ctt Leu	tat Tyr	cat His	cag Gln	aag Lys	1632
			530				535				540						
70	tca Ser	ggt Val	ctg Leu	ata Ile	atg Met	gga Gly	cga Arg	ggc Gly	tat Tyr	cat His	tat Tyr	gct Ala	act Thr	tgt Cys	ctt Leu	gaa Glu	1680
			545			550					555					560	
75	ggg Gly	gca Ala	ctg Leu	aaa Lys	atc Ile	aaa Lys	gaa Glu	att Ile	act Thr	tat Tyr	atg Met	cac His	tct Ser	gaa Glu	ggc Gly	atc Ile	1728
				565					570					575			
80	ctt Leu	gct Ala	ggt Gly	gaa Glu	ttg Leu	aaa Lys	cat His	ggc Gly	cct Pro	ctg Leu	gct Ala	ttg Leu	gtg Val	gat Asp	aaa Lys	ttg Leu	1776
			580						585					590			

# ES 2 343 571 T3

```

5      atg cct gtg atc atg atc atc atg aga gat cac act tat gcc aag tgt      1824
      Met Pro Val Ile Met Ile Ile Met Arg Asp His Thr Tyr Ala Lys Cys
          595                      600                      605

10     cag aat gct ctt cag caa gtg gtt gct cgg cag ggg cgg cct gtg gta      1872
      Gln Asn Ala Leu Gln Gln Val Val Ala Arg Gln Gly Arg Pro Val Val
          610                      615                      620

15     att tgt gat aag gag gat act gag acc att aag aac aca aaa aga acg      1920
      Ile Cys Asp Lys Glu Asp Thr Glu Thr Ile Lys Asn Thr Lys Arg Thr
      625                      630                      635                      640

20     atc aag gtg ccc cac tca gtg gac tgc ttg cag ggc att ctc agc gtg      1968
      Ile Lys Val Pro His Ser Val Asp Cys Leu Gln Gly Ile Leu Ser Val
          645                      650                      655

25     atc cct tta cag ttg ctg gct ttc cac ctt gct gtg ctg aga ggc tat      2016
      Ile Pro Leu Gln Leu Leu Ala Phe His Leu Ala Val Leu Arg Gly Tyr
          660                      665                      670

30     gat gtt gat ttc cca cgg aat ctt gcc aaa tct gtg act gta gag tga      2064
      Asp Val Asp Phe Pro Arg Asn Leu Ala Lys Ser Val Thr Val Glu
          675                      680                      685

```

25 <210> 8

<211> 687

<212> PRT

30 <213> Secuencia artificial

<220>

<221> misc\_feature

35 <222> (57)..(57)

<223> El "Xaa" en posición 57 representa Thr o Ile.

<220>

40 <223> GFAT1 modificada por una etiqueta de purificación interna

<400> 8

```

45     Met Cys Gly Ile Phe Ala Tyr Leu Asn Tyr His Val Pro Arg Thr Arg
      1          5          10          15

50     Arg Glu Ile Leu Glu Thr Leu Ile Lys Gly Leu Gln Arg Leu Glu Tyr
      20          25          30

55     Arg Gly Tyr Asp Ser Ala Gly Val Gly Phe Asp Gly Gly Asn Asp Lys
      35          40          45

60     Asp Trp Glu Ala Asn Ala Cys Lys Xaa Gln Leu Ile Lys Lys Lys Gly
      50          55          60

65     Lys Val Lys Ala Leu Asp Glu Glu Val His Lys Gln Gln Asp Met Asp
      65          70          75          80

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

```

65

ES 2 343 571 T3

Ala Thr His Gly Glu Pro Ser Pro Val Asn Ser His Pro Gln Arg Ser  
 100 105 110

5 Asp Lys Asn Asn Glu Phe Ile Val Ile His Asn Gly Ile Ile Thr Asn  
 115 120 125

10 Tyr Lys Asp Leu Lys Lys Phe Leu Glu Ser Lys Gly Tyr Asp Phe Glu  
 130 135 140

15 Ser Glu Thr Asp Thr Glu Thr Ile Ala Lys Leu Val Lys Tyr Met Tyr  
 145 150 155 160

20 Asp Asn Arg Glu Ser Gln Asp Thr Ser Phe Thr Thr Leu Val Glu Arg  
 165 170 175

25 Val Ile Gln Gln Leu Glu Gly Ala Phe Ala Leu Val Phe Lys Ser Val  
 180 185 190

30 His Phe Pro Gly Gln Ala Val Gly Thr Arg Arg Gly Ser Pro Leu Leu  
 195 200 205

35 Ile Gly Val Arg Ser Glu His Lys Leu Ser Thr Asp His Ile Pro Ile  
 210 215 220

40 Leu Tyr Arg Thr Gly Lys Asp Lys Lys Gly Ser Cys Asn Leu Ser Arg  
 225 230 235 240

45 Val Asp Ser Thr Thr Cys Leu Phe Pro Val Glu Glu Lys Ala Val Glu  
 245 250 255

50 Tyr Tyr Phe Ala Ser Asp Ala Ser Ala Val Ile Glu His Thr Asn Arg  
 260 265 270

55 Val Ile Phe Leu Glu Asp Asp Asp Val Ala Ala Val Val Asp Gly Arg  
 275 280 285

60 Leu Ser Ile His Arg Ile Lys Arg Thr Ala Gly His His His His His  
 290 295 300

65 His Asp His Pro Gly Arg Ala Val Gln Thr Leu Gln Met Glu Leu Gln  
 305 310 315 320

Gln Ile Met Lys Gly Asn Phe Ser Ser Phe Met Gln Lys Glu Ile Phe  
 325 330 335

60 Glu Gln Pro Glu Ser Val Val Asn Thr Met Arg Gly Arg Val Asn Phe  
 340 345 350

65 Asp Asp Tyr Thr Val Asn Leu Gly Gly Leu Lys Asp His Ile Lys Glu

ES 2 343 571 T3

	355		360		365														
5	Ile	Gln	Arg	Cys	Arg	Arg	Leu	Ile	Leu	Ile	Ala	Cys	Gly	Thr	Ser	Tyr			
	370						375					380							
10	His	Ala	Gly	Val	Ala	Thr	Arg	Gln	Val	Leu	Glu	Glu	Leu	Thr	Glu	Leu			
	385					390					395					400			
15	Pro	Val	Met	Val	Glu	Leu	Ala	Ser	Asp	Phe	Leu	Asp	Arg	Asn	Thr	Pro			
				405						410				415					
20	Val	Phe	Arg	Asp	Asp	Val	Cys	Phe	Phe	Leu	Ser	Gln	Ser	Gly	Glu	Thr			
				420				425						430					
25	Ala	Asp	Thr	Leu	Met	Gly	Leu	Arg	Tyr	Cys	Lys	Glu	Arg	Gly	Ala	Leu			
		435						440					445						
30	Thr	Val	Gly	Ile	Thr	Asn	Thr	Val	Gly	Ser	Ser	Ile	Ser	Arg	Glu	Thr			
	450						455					460							
35	Asp	Cys	Gly	Val	His	Ile	Asn	Ala	Gly	Pro	Glu	Ile	Gly	Val	Ala	Ser			
	465					470					475					480			
40	Thr	Lys	Ala	Tyr	Thr	Ser	Gln	Phe	Val	Ser	Leu	Val	Met	Phe	Ala	Leu			
				485						490				495					
45	Met	Met	Cys	Asp	Asp	Arg	Ile	Ser	Met	Gln	Glu	Arg	Arg	Lys	Glu	Ile			
			500						505					510					
50	Met	Leu	Gly	Leu	Lys	Arg	Leu	Pro	Asp	Leu	Ile	Lys	Glu	Val	Leu	Ser			
		515						520					525						
55	Met	Asp	Asp	Glu	Ile	Gln	Lys	Leu	Ala	Thr	Glu	Leu	Tyr	His	Gln	Lys			
	530						535					540							
60	Ser	Val	Leu	Ile	Met	Gly	Arg	Gly	Tyr	His	Tyr	Ala	Thr	Cys	Leu	Glu			
	545					550					555					560			
65	Gly	Ala	Leu	Lys	Ile	Lys	Glu	Ile	Thr	Tyr	Met	His	Ser	Glu	Gly	Ile			
				565					570						575				
70	Leu	Ala	Gly	Glu	Leu	Lys	His	Gly	Pro	Leu	Ala	Leu	Val	Asp	Lys	Leu			
			580					585						590					
75	Met	Pro	Val	Ile	Met	Ile	Ile	Met	Arg	Asp	His	Thr	Tyr	Ala	Lys	Cys			
			595					600					605						
80	Gln	Asn	Ala	Leu	Gln	Gln	Val	Val	Ala	Arg	Gln	Gly	Arg	Pro	Val	Val			
	610						615					620							

# ES 2 343 571 T3

		Ile	Cys	Asp	Lys	Glu	Asp	Thr	Glu	Thr	Ile	Lys	Asn	Thr	Lys	Arg	Thr	
		625					630					635					640	
5																		
		Ile	Lys	Val	Pro	His	Ser	Val	Asp	Cys	Leu	Gln	Gly	Ile	Leu	Ser	Val	
						645					650					655		
10																		
		Ile	Pro	Leu	Gln	Leu	Leu	Ala	Phe	His	Leu	Ala	Val	Leu	Arg	Gly	Tyr	
				660					665						670			
15																		
		Asp	Val	Asp	Phe	Pro	Arg	Asn	Leu	Ala	Lys	Ser	Val	Thr	Val	Glu		
				675				680						685				
20																		
		<210> 9																
		<211> 2067																
		<212> ADN																
		<213> Secuencia artificial																
		<220>																
25		<223> GFAT2 modificada por una etiqueta de purificación interna																
		<220>																
		<221> CDS																
30		<222> (1)..(2067)																
		<223>																
		<400> 9																
35		atg	tgc	gga	atc	ttt	gcc	tac	atg	aac	tac	aga	gtc	ccc	cgg	acg	agg	48
		Met	Cys	Gly	Ile	Phe	Ala	Tyr	Met	Asn	Tyr	Arg	Val	Pro	Arg	Thr	Arg	
		1			5					10					15			
40		aag	gag	atc	ttc	gaa	acc	ctc	atc	aag	ggc	ctg	cag	cgg	ctg	gag	tac	96
		Lys	Glu	Ile	Phe	Glu	Thr	Leu	Ile	Lys	Gly	Leu	Gln	Arg	Leu	Glu	Tyr	
				20						25				30				
45		aga	ggc	tac	gac	tcg	gca	ggt	gtg	gcg	atc	gat	ggg	aat	aat	cac	gaa	144
		Arg	Gly	Tyr	Asp	Ser	Ala	Gly	Val	Ala	Ile	Asp	Gly	Asn	Asn	His	Glu	
				35				40					45					
50		gtc	aaa	gaa	aga	cac	att	cag	ctg	gtc	aag	aaa	agg	ggg	aaa	gtc	aag	192
		Val	Lys	Glu	Arg	His	Ile	Gln	Leu	Val	Lys	Lys	Arg	Gly	Lys	Val	Lys	
				50			55						60					
55		gct	ctc	gat	gaa	gaa	ctt	tac	aaa	caa	gac	agc	atg	gac	tta	aaa	gtg	240
		Ala	Leu	Asp	Glu	Glu	Leu	Tyr	Lys	Gln	Asp	Ser	Met	Asp	Leu	Lys	Val	
						70					75					80		
60		gag	ttt	gag	aca	cac	ttc	ggc	att	gcc	cac	acg	cgc	tgg	gcc	acc	cac	288
		Glu	Phe	Glu	Thr	His	Phe	Gly	Ile	Ala	His	Thr	Arg	Trp	Ala	Thr	His	
					85					90					95			
65		ggg	gtc	ccc	agt	gct	gtc	aac	agc	cac	cct	cag	cgc	tca	gac	aaa	ggc	336
		Gly	Val	Pro	Ser	Ala	Val	Asn	Ser	His	Pro	Gln	Arg	Ser	Asp	Lys	Gly	
					100					105					110			
70		aac	gaa	ttt	ggt	gtc	atc	cac	aat	ggg	atc	atc	aca	aat	tac	aaa	gat	384
		Asn	Glu	Phe	Val	Val	Ile	His	Asn	Gly	Ile	Ile	Thr	Asn	Tyr	Lys	Asp	
				115					120					125				
75		ctg	agg	aaa	ttt	ctg	gaa	agc	aaa	ggc	tac	gag	ttt	gag	tca	gaa	aca	432

ES 2 343 571 T3

	Leu	Arg	Lys	Phe	Leu	Glu	Ser	Lys	Gly	Tyr	Glu	Phe	Glu	Ser	Glu	Thr	
	130						135					140					
5	gat	aca	gag	acc	atc	gcc	aag	ctg	att	aaa	tat	gtg	ttc	gac	aac	aga	480
	Asp	Thr	Glu	Thr	Ile	Ala	Lys	Leu	Ile	Lys	Tyr	Val	Phe	Asp	Asn	Arg	
	145					150					155					160	
10	gaa	act	gag	gac	att	acg	ttt	tca	acg	ttg	gtc	gag	aga	gtc	att	cag	528
	Glu	Thr	Glu	Asp	Ile	Thr	Phe	Ser	Thr	Leu	Val	Glu	Arg	Val	Ile	Gln	
					165					170					175		
15	cag	ttg	gaa	ggt	gca	ttc	gcg	ctg	gtt	ttc	aag	agt	gtc	cac	tac	cca	576
	Gln	Leu	Glu	Ala	Phe	Ala	Leu	Val	Phe	Lys	Ser	Val	His	Tyr	Pro		
				180					185					190			
20	gga	gaa	gcc	ggt	gcc	aca	cgg	aga	ggc	agc	ccc	ctg	ctc	atc	gga	gtc	624
	Gly	Glu	Ala	Val	Ala	Thr	Arg	Arg	Gly	Ser	Pro	Leu	Leu	Ile	Gly	Val	
			195					200						205			
25	cgg	agc	aaa	tac	aag	ctc	tcc	aca	gaa	cag	atc	cct	atc	tta	tac	agg	672
	Arg	Ser	Lys	Tyr	Lys	Leu	Thr	Glu	Gln	Ile		Pro	Ile	Leu	Tyr	Arg	
		210					215					220					
30	acg	tgc	act	ctg	gag	aat	gtg	aag	aat	atc	tgt	aag	aca	cgg	atg	aag	720
	Thr	Cys	Thr	Leu	Glu	Asn	Val	Lys	Asn	Ile	Cys	Lys	Thr	Arg	Met	Lys	
		225				230					235				240		
35	agg	ctg	gac	agc	tcc	gcc	tgc	ctg	cat	gct	gtg	ggc	gac	aag	gcc	gtg	768
	Arg	Leu	Asp	Ser	Ser	Ala	Cys	Leu	His	Ala	Val	Gly	Asp	Lys	Ala	Val	
				245					250						255		
40	gaa	ttc	ttc	ttt	gct	tct	gat	gca	agc	gct	atc	ata	gag	cac	acc	aac	816
	Glu	Phe	Phe	Phe	Ala	Ser	Asp	Ala	Ser	Ala	Ile	Ile	Glu	His	Thr	Asn	
				260					265						270		
45	cgg	gtc	atc	ttc	ctg	gag	gac	gat	gac	atc	gcc	gca	gtg	gct	gat	ggg	864
	Arg	Val	Ile	Phe	Leu	Glu	Asp	Asp	Asp	Ile	Ala	Ala	Val	Ala	Asp	Gly	
			275					280						285			
50	aaa	ctc	tcc	att	cac	cgg	gtc	aag	cgc	tcg	gcc	agt	cat	cac	cat	cac	912
	Lys	Leu	Ser	Ile	His	Arg	Val	Lys	Arg	Ser	Ala	Ser	His	His	His	His	
		290					295					300					
55	cat	cac	gat	gac	cca	tct	cga	gcc	atc	cag	acc	ttg	cag	atg	gaa	ctg	960
	His	His	Asp	Asp	Pro	Ser	Arg	Ala	Ile	Gln	Thr	Leu	Gln	Met	Glu	Leu	
					310						315					320	
60	cag	caa	atc	atg	aaa	ggt	aac	ttc	agt	gcg	ttt	atg	cag	aag	gag	atc	1008
	Gln	Gln	Ile	Met	Lys	Gly	Asn	Phe	Ser	Ala	Phe	Met	Gln	Lys	Glu	Ile	
					325					330					335		
65	ttc	gaa	cag	cca	gaa	tca	ggt	ttc	aat	act	atg	aga	ggt	cgg	gtg	aat	1056
	Phe	Glu	Gln	Pro	Glu	Ser	Val	Phe	Asn	Thr	Met	Arg	Gly	Arg	Val	Asn	
				340					345					350			
70	ttt	gaa	acc	aac	aca	gtg	ctc	ctg	ggt	ggc	ttg	aag	gac	cac	ttg	aag	1104
	Phe	Glu	Thr	Asn	Thr	Val	Leu	Leu	Gly	Gly	Leu	Lys	Asp	His	Leu	Lys	
				355				360						365			
75	gag	att	cga	cga	tgc	cga	cgg	ctc	atc	gtg	att	ggc	tgt	gga	acc	agc	1152
	Glu	Ile	Arg	Arg	Cys	Arg	Arg	Leu	Ile	Val	Ile	Gly	Cys	Gly	Thr	Ser	
			370				375						380				
80	tac	cac	gct	gcc	gtg	gct	acg	cgg	caa	ggt	ttg	gag	gaa	ctg	act	gag	1200
	Tyr	His	Ala	Ala	Val	Ala	Thr	Arg	Gln	Val	Leu	Glu	Glu	Leu	Thr	Glu	

ES 2 343 571 T3

	385		390				395			400							
5	cct Leu	cct Pro	gtg Val	atg Met	gtt Val	gaa Glu	ctt Leu	gct Ala	agt Ser	gat Asp	ttt Phe	ctg Leu	gac Asp	agg Arg	aac Asn	aca Thr	1248
				405					410						415		
10	cct Pro	gtg Val	ttc Phe	agg Arg	gat Asp	gac Asp	gtt Val	tgc Cys	ttt Phe	ttc Phe	atc Ile	agc Ser	cag Gln	tca Ser	ggc Gly	gag Glu	1296
				420					425					430			
15	acc Thr	gcg Ala	gac Asp	acc Thr	ctc Leu	ctg Leu	gcg Ala	ctg Leu	cgc Arg	tac Tyr	tgt Cys	aag Lys	gac Asp	cgc Arg	ggc Gly	gct Ala	1344
				435				440					445				
20	ctc Leu	acc Thr	gtg Val	ggc Gly	gtc Val	acc Thr	aac Asn	acc Thr	gtg Val	ggc Gly	agc Ser	tcc Ser	atc Ile	tct Ser	cgc Arg	gag Glu	1392
				450			455					460					
25	acc Thr	gac Asp	tgc Cys	ggc Gly	gtc Val	cac His	atc Ile	aac Asn	gca Ala	ggg Gly	ccg Pro	gag Glu	gtc Val	ggc Gly	gtg Val	gcc Ala	1440
				465		470				475						480	
30	agc Ser	acc Thr	aag Lys	gct Ala	tat Tyr	acc Thr	agt Ser	cag Gln	ttc Phe	atc Ile	tct Ser	ctg Leu	gtg Val	atg Met	ttt Phe	ggt Gly	1488
					485					490					495		
35	ttg Leu	atg Met	atg Met	tct Ser	gaa Glu	gac Asp	cga Arg	att Ile	tca Ser	cta Leu	caa Gln	aac Asn	agg Arg	agg Arg	caa Gln	gag Glu	1536
				500				505						510			
40	atc Ile	atc Ile	cgt Arg	ggc Gly	ttg Leu	aga Arg	tct Ser	tta Leu	cct Pro	gag Glu	ctg Leu	atc Ile	aag Lys	gaa Glu	gtg Val	ctg Leu	1584
			515					520					525				
45	tct Ser	ctg Leu	gag Glu	gag Glu	aag Lys	atc Ile	cac His	gac Asp	ttg Leu	gcc Ala	ctg Leu	gag Glu	ctc Leu	tac Tyr	acg Thr	cag Gln	1632
			530				535					540					
50	aga Arg	tcg Ser	ctg Leu	ctg Leu	gtg Val	atg Met	ggg Gly	cgg Arg	ggc Gly	tac Tyr	aac Asn	tat Tyr	gcc Ala	acc Thr	tgc Cys	ctg Leu	1680
					550						555					560	
55	gaa Glu	gga Gly	gcc Ala	ctg Leu	aaa Lys	att Ile	aaa Lys	gag Glu	ata Ile	acc Thr	tac Tyr	atg Met	cac His	tca Ser	gaa Glu	ggc Gly	1728
				565					570					575			
60	atc Ile	ctg Leu	gct Ala	ggg Gly	gag Glu	ctg Leu	aag Lys	cac His	ggg Gly	ccc Pro	ctg Leu	gca Ala	ctg Leu	att Ile	gac Asp	aag Lys	1776
				580				585						590			
65	cag Gln	atg Met	ccc Pro	gtc Val	atc Ile	atg Met	gtc Val	att Ile	atg Met	aag Lys	gat Asp	cct Pro	tgc Cys	ttc Phe	gcc Ala	aaa Lys	1824
			595					600				605					
70	tgc Cys	cag Gln	aac Asn	gcc Ala	ctg Leu	cag Gln	caa Gln	gtc Val	acg Thr	gcc Ala	cgc Arg	cag Gln	ggt Gly	cgc Arg	ccc Pro	att Ile	1872
			610				615					620					
75	ata Ile	ctg Leu	tgc Cys	tcc Ser	aag Lys	gac Asp	gat Asp	act Thr	gaa Glu	agt Ser	tcc Ser	aag Lys	ttt Phe	gcg Ala	tat Tyr	aag Lys	1920
						630					635					640	
80	aca Thr	atc Ile	gag Glu	ctg Leu	ccc Pro	cac His	act Thr	gtg Val	gac Asp	tgc Cys	ctc Leu	cag Gln	ggc Gly	atc Ile	ctg Leu	agc Ser	1968
					645					650					655		



ES 2 343 571 T3

5 Gln Leu Glu Gly Ala Phe Ala Leu Val Phe Lys Ser Val His Tyr Pro  
 180 185 190  
 Gly Glu Ala Val Ala Thr Arg Arg Gly Ser Pro Leu Leu Ile Gly Val  
 195 200 205  
 10 Arg Ser Lys Tyr Lys Leu Ser Thr Glu Gln Ile Pro Ile Leu Tyr Arg  
 210 215 220  
 15 Thr Cys Thr Leu Glu Asn Val Lys Asn Ile Cys Lys Thr Arg Met Lys  
 225 230 235 240  
 Arg Leu Asp Ser Ser Ala Cys Leu His Ala Val Gly Asp Lys Ala Val  
 245 250 255  
 20 Glu Phe Phe Phe Ala Ser Asp Ala Ser Ala Ile Ile Glu His Thr Asn  
 260 265 270  
 25 Arg Val Ile Phe Leu Glu Asp Asp Asp Ile Ala Ala Val Ala Asp Gly  
 275 280 285  
 30 Lys Leu Ser Ile His Arg Val Lys Arg Ser Ala Ser His His His His  
 290 295 300  
 35 His His Asp Asp Pro Ser Arg Ala Ile Gln Thr Leu Gln Met Glu Leu  
 305 310 315 320  
 40 Gln Gln Ile Met Lys Gly Asn Phe Ser Ala Phe Met Gln Lys Glu Ile  
 325 330 335  
 Phe Glu Gln Pro Glu Ser Val Phe Asn Thr Met Arg Gly Arg Val Asn  
 340 345 350  
 45 Phe Glu Thr Asn Thr Val Leu Leu Gly Gly Leu Lys Asp His Leu Lys  
 355 360 365  
 50 Glu Ile Arg Arg Cys Arg Arg Leu Ile Val Ile Gly Cys Gly Thr Ser  
 370 375 380  
 55 Tyr His Ala Ala Val Ala Thr Arg Gln Val Leu Glu Glu Leu Thr Glu  
 385 390 395 400  
 60 Leu Pro Val Met Val Glu Leu Ala Ser Asp Phe Leu Asp Arg Asn Thr  
 405 410 415  
 65 Pro Val Phe Arg Asp Asp Val Cys Phe Phe Ile Ser Gln Ser Gly Glu  
 420 425 430

ES 2 343 571 T3

Thr Ala Asp Thr Leu Leu Ala Leu Arg Tyr Cys Lys Asp Arg Gly Ala  
 435 440 445  
 5  
 Leu Thr Val Gly Val Thr Asn Thr Val Gly Ser Ser Ile Ser Arg Glu  
 450 455 460  
 10  
 Thr Asp Cys Gly Val His Ile Asn Ala Gly Pro Glu Val Gly Val Ala  
 465 470 475 480  
 15  
 Ser Thr Lys Ala Tyr Thr Ser Gln Phe Ile Ser Leu Val Met Phe Gly  
 485 490 495  
 20  
 Leu Met Met Ser Glu Asp Arg Ile Ser Leu Gln Asn Arg Arg Gln Glu  
 500 505 510  
 25  
 Ile Ile Arg Gly Leu Arg Ser Leu Pro Glu Leu Ile Lys Glu Val Leu  
 515 520 525  
 30  
 Ser Leu Glu Glu Lys Ile His Asp Leu Ala Leu Glu Leu Tyr Thr Gln  
 530 535 540  
 35  
 Arg Ser Leu Leu Val Met Gly Arg Gly Tyr Asn Tyr Ala Thr Cys Leu  
 545 550 555 560  
 40  
 Glu Gly Ala Leu Lys Ile Lys Glu Ile Thr Tyr Met His Ser Glu Gly  
 565 570 575  
 45  
 Ile Leu Ala Gly Glu Leu Lys His Gly Pro Leu Ala Leu Ile Asp Lys  
 580 585 590  
 50  
 Gln Met Pro Val Ile Met Val Ile Met Lys Asp Pro Cys Phe Ala Lys  
 595 600 605  
 55  
 Cys Gln Asn Ala Leu Gln Gln Val Thr Ala Arg Gln Gly Arg Pro Ile  
 610 615 620  
 60  
 Ile Leu Cys Ser Lys Asp Asp Thr Glu Ser Ser Lys Phe Ala Tyr Lys  
 625 630 635 640  
 65  
 Thr Ile Glu Leu Pro His Thr Val Asp Cys Leu Gln Gly Ile Leu Ser  
 645 650 655  
 Val Ile Pro Leu Gln Leu Leu Ser Phe His Leu Ala Val Leu Arg Gly  
 660 665 670  
 Tyr Asp Val Asp Phe Pro Arg Asn Leu Ala Lys Ser Val Thr Val Glu  
 675 680 685

## ES 2 343 571 T3

<210> 11  
 <211> 2118  
 <212> ADN  
 5 <213> Secuencia artificial  
  
 <220>  
 <223> GPAT1Alt modificada por una etiqueta de purificación interna  
 10  
 <220>  
 <221> CDS  
 <222> (1)..(2118)  
 15 <223>  
 <220>  
 <221> misc\_feature  
 20 <222> (170)..(170)  
 <223> t o c  
  
 <400> 11  
 25           atg tgt ggt ata ttt gct tac tta aac tac cat gtt cct cga acg aga           48  
           Met Cys Gly Ile Phe Ala Tyr Leu Asn Tyr His Val Pro Arg Thr Arg  
           1                           5                           10                           15  
  
           cga gaa atc ctg gag acc cta atc aaa ggc ctt cag aga ctg gag tac           96  
           Arg Glu Ile Leu Glu Thr Leu Ile Lys Gly Leu Gln Arg Leu Glu Tyr  
                                   20                           25                           30  
  
           aga gga tat gat tct gct ggt gtg gga ttt gat gga ggc aat gat aaa           144  
           Arg Gly Tyr Asp Ser Ala Gly Val Gly Phe Asp Gly Gly Asn Asp Lys  
                                   35                           40                           45  
  
 35           gat tgg gaa gcc aat gcc tgc aaa anc cag ctt att aag aag aaa gga           192  
           Asp Trp Glu Ala Asn Ala Cys Lys Xaa Gln Leu Ile Lys Lys Lys Gly  
                                   50                           55                           60  
  
           aaa gtt aag gca ctg gat gaa gaa gtt cac aag caa caa gat atg gat           240  
           Lys Val Lys Ala Leu Asp Glu Glu Val His Lys Gln Gln Asp Met Asp  
                                   65                           70                           75                           80  
  
           ttg gat ata gaa ttt gat gta cac ctt gga ata gct cat acc cgt tgg           288  
           Leu Asp Ile Glu Phe Asp Val His Leu Gly Ile Ala His Thr Arg Trp  
                                   85                           90                           95  
  
 45           gca aca cat gga gaa ccc agt cct gtc aat agc cac ccc cag cgc tct           336  
           Ala Thr His Gly Glu Pro Ser Pro Val Asn Ser His Pro Gln Arg Ser  
                                   100                           105                           110  
  
           gat aaa aat aat gaa ttt atc gtt att cac aat gga atc atc acc aac           384  
           Asp Lys Asn Asn Glu Phe Ile Val Ile His Asn Gly Ile Ile Thr Asn  
                                   115                           120                           125  
  
           tac aaa gac ttg aaa aag ttt ttg gaa agc aaa ggc tat gac ttc gaa           432  
           Tyr Lys Asp Leu Lys Lys Phe Leu Glu Ser Lys Gly Tyr Asp Phe Glu  
                                   130                           135                           140  
  
 55           tct gaa aca gac aca gag aca att gcc aag ctc gtt aag tat atg tat           480  
           Ser Glu Thr Asp Thr Glu Thr Ile Ala Lys Leu Val Lys Tyr Met Tyr  
                                   145                           150                           155                           160  
  
 60           gac aat cgg gaa agt caa gat acc agc ttt act acc ttg gtg gag aga           528  
           Asp Asn Arg Glu Ser Gln Asp Thr Ser Phe Thr Thr Leu Val Glu Arg  
                                   165                           170                           175  
  
           gtt atc caa caa ttg gaa ggt gct ttt gca ctt gtg ttt aaa agt gtt           576  
           Val Ile Gln Gln Leu Glu Gly Ala Phe Ala Leu Val Phe Lys Ser Val

ES 2 343 571 T3

	180	185	190	
5	cat ttt ccc ggg caa gca gtt ggc His Phe Pro Gly Gln Ala Val Gly	aca agg cga ggt agc cct ctg ttg Thr Arg Arg Gly Ser Pro Leu Leu		624
	195	200	205	
10	att ggt gta cgg agt gaa cat aaa ctt tct act gat cac att cct ata Ile Gly Val Arg Ser Glu His Lys Leu Ser Thr Asp His Ile Pro Ile			672
	210	215	220	
15	ctc tac aga aca gct agg act cag att gga tca aaa ttc aca cgg tgg Leu Tyr Arg Thr Ala Arg Thr Gln Ile Gly Ser Lys Phe Thr Arg Trp			720
	225	230	235	240
20	gga tca cag gga gaa aga ggc aaa gac aag aaa gga agc tgc aat ctc Gly Ser Gln Gly Glu Arg Gly Lys Asp Lys Lys Gly Ser Cys Asn Leu			768
	245	250	255	
25	tct cgt gtg gac agc aca acc tgc ctt ttc ccg gtg gaa gaa aaa gca Ser Arg Val Asp Ser Thr Thr Cys Leu Phe Pro Val Glu Glu Lys Ala			816
	260	265	270	
30	gtg gag tat tac ttt gct tct gat gca agt gct gtc ata gaa cac acc Val Glu Tyr Tyr Phe Ala Ser Asp Ala Ser Ala Val Ile Glu His Thr			864
	275	280	285	
35	aat cgc gtc atc ttt ctg gaa gat gat gat gtt gca gca gta gtg gat Asn Arg Val Ile Phe Leu Glu Asp Asp Asp Val Ala Ala Val Val Asp			912
	290	295	300	
40	gga cgt ctt tct atc cat cga att aaa cga act gca gga cat cac cat Gly Arg Leu Ser Ile His Arg Ile Lys Arg Thr Ala Gly His His His			960
	305	310	315	320
45	cac cat cac gat cac ccc gga cga gct gtg caa aca ctc cag atg gaa His His His Asp His Pro Gly Arg Ala Val Gln Thr Leu Gln Met Glu			1008
	325	330	335	
50	ctc cag cag atc atg aag ggc aac ttc agt tca ttt atg cag aag gaa Leu Gln Gln Ile Met Lys Gly Asn Phe Ser Ser Phe Met Gln Lys Glu			1056
	340	345	350	
55	ata ttt gag cag cca gag tct gtc gtg aac aca atg aga gga aga gtc Ile Phe Glu Gln Pro Glu Ser Val Val Asn Thr Met Arg Gly Arg Val			1104
	355	360	365	
60	aac ttt gat gac tat act gtg aat ttg ggt ggt ttg aag gat cac ata Asn Phe Asp Asp Tyr Thr Val Asn Leu Gly Gly Leu Lys Asp His Ile			1152
	370	375	380	
65	aag gag atc cag aga tgc cgg cgt ttg att ctt att gct tgt gga aca Lys Glu Ile Gln Arg Cys Arg Arg Leu Ile Leu Ile Ala Cys Gly Thr			1200
	385	390	395	400
70	agt tac cat gct ggt gta gca aca cgt caa gtt ctt gag gag ctg act Ser Tyr His Ala Gly Val Ala Thr Arg Gln Val Leu Glu Glu Leu Thr			1248
	405	410	415	
75	gag ttg cct gtg atg gtg gaa cta gca agt gac ttc ctg gac aga aac Glu Leu Pro Val Met Val Glu Leu Ala Ser Asp Phe Leu Asp Arg Asn			1296
	420	425	430	
80	aca cca gtc ttt cga gat gat gtt tgc ttt ttc ctt agt caa tca ggt Thr Pro Val Phe Arg Asp Asp Val Cys Phe Phe Leu Ser Gln Ser Gly			1344
	435	440	445	

ES 2 343 571 T3

5 gag aca gca gat act ttg atg ggt ctt cgt tac tgt aag gag aga gga 1392  
 Glu Thr Ala Asp Thr Leu Met Gly Leu Arg Tyr Cys Lys Glu Arg Gly  
 450 455 460

10 gct tta act gtg ggg atc aca aac aca gtt ggc agt tcc ata tca cgg 1440  
 Ala Leu Thr Val Gly Ile Thr Asn Thr Val Gly Ser Ser Ile Ser Arg  
 465 470 475 480

15 gag aca gat tgt gga gtt cat att aat gct ggt cct gag att ggt gtg 1488  
 Glu Thr Asp Cys Gly Val His Ile Asn Ala Gly Pro Glu Ile Gly Val  
 485 490 495

20 gcc agt aca aag gct tat acc agc cag ttt gta tcc ctt gtg atg ttt 1536  
 Ala Ser Thr Lys Ala Tyr Thr Ser Gln Phe Val Ser Leu Val Met Phe  
 500 505 510

25 gcc ctt atg atg tgt gat gat cgg atc tcc atg caa gaa aga cgc aaa 1584  
 Ala Leu Met Met Cys Asp Asp Arg Ile Ser Met Gln Glu Arg Arg Lys  
 515 520 525

30 gag atc atg ctt gga ttg aaa cgg ctg cct gat ttg att aag gaa gta 1632  
 Glu Ile Met Leu Gly Leu Lys Arg Leu Pro Asp Leu Ile Lys Glu Val  
 530 535 540

35 ctg agc atg gat gac gaa att cag aaa cta gca aca gaa ctt tat cat 1680  
 Leu Ser Met Asp Asp Glu Ile Gln Lys Leu Ala Thr Glu Leu Tyr His  
 545 550 555 560

40 cag aag tca gtt ctg ata atg gga cga ggc tat cat tat gct act tgt 1728  
 Gln Lys Ser Val Leu Ile Met Gly Arg Gly Tyr His Tyr Ala Thr Cys  
 565 570 575

45 ctt gaa ggg gca ctg aaa atc aaa gaa att act tat atg cac tct gaa 1776  
 Leu Glu Gly Ala Leu Lys Ile Lys Glu Ile Thr Tyr Met His Ser Glu  
 580 585 590

50 ggc atc ctt gct ggt gaa ttg aaa cat ggc cct ctg gct ttg gtg gat 1824  
 Gly Ile Leu Ala Gly Glu Leu Lys His Gly Pro Leu Ala Leu Val Asp  
 595 600 605

55 aaa ttg atg cct gtg atc atg atc atc atg aga gat cac act tat gcc 1872  
 Lys Leu Met Pro Val Ile Met Ile Ile Met Arg Asp His Thr Tyr Ala  
 610 615 620

60 aag tgt cag aat gct ctt cag caa gtg gtt gct cgg cag ggg cgg cct 1920  
 Lys Cys Gln Asn Ala Leu Gln Gln Val Val Ala Arg Gln Gly Arg Pro  
 625 630 635 640

65 gtg gta att tgt gat aag gag gat act gag acc att aag aac aca aaa 1968  
 Val Val Ile Cys Asp Lys Glu Asp Thr Glu Thr Ile Lys Asn Thr Lys  
 645 650 655

70 aga acg atc aag gtg ccc cac tca gtg gac tgc ttg cag ggc att etc 2016  
 Arg Thr Ile Lys Val Pro His Ser Val Asp Cys Leu Gln Gly Ile Leu  
 660 665 670

75 agc gtg atc cct tta cag ttg ctg gct ttc cac ctt gct gtg ctg aga 2064  
 Ser Val Ile Pro Leu Gln Leu Leu Ala Phe His Leu Ala Val Leu Arg  
 675 680 685

80 ggc tat gat gtt gat ttc cca cgg aat ctt gcc aaa tct gtg act gta 2112  
 Gly Tyr Asp Val Asp Phe Pro Arg Asn Leu Ala Lys Ser Val Thr Val  
 690 695 700

ES 2 343 571 T3

**gag tga**  
**Glu**  
**705**

2118

5 <210> 12  
 <211> 705  
 <212> PRT  
 10 <213> Secuencia artificial

<220>  
 <221> misc\_feature  
 15 <222> (57)..(57)  
 <223> El "Xaa" en posición 57 representa Thr o Ile.

<220>  
 20 <223> GFAT1Alt modificada por una etiqueta de purificación interna

<400> 12

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

ES 2 343 571 T3

	180	185	190	
5	His Phe Pro Gly Gln Ala Val Gly Thr Arg Arg Gly Ser Pro Leu Leu	200	205	195
10	Ile Gly Val Arg Ser Glu His Lys Leu Ser Thr Asp His Ile Pro Ile	215	220	210
15	Leu Tyr Arg Thr Ala Arg Thr Gln Ile Gly Ser Lys Phe Thr Arg Trp	230	235	225
20	Gly Ser Gln Gly Glu Arg Gly Lys Asp Lys Lys Gly Ser Cys Asn Leu	245	250	255
25	Ser Arg Val Asp Ser Thr Thr Cys Leu Phe Pro Val Glu Glu Lys Ala	260	265	270
30	Val Glu Tyr Tyr Phe Ala Ser Asp Ala Ser Ala Val Ile Glu His Thr	280	285	275
35	Asn Arg Val Ile Phe Leu Glu Asp Asp Asp Val Ala Ala Val Val Asp	295	300	290
40	Gly Arg Leu Ser Ile His Arg Ile Lys Arg Thr Ala Gly His His His	310	315	305
45	His His His Asp His Pro Gly Arg Ala Val Gln Thr Leu Gln Met Glu	325	330	335
50	Leu Gln Gln Ile Met Lys Gly Asn Phe Ser Ser Phe Met Gln Lys Glu	340	345	350
55	Ile Phe Glu Gln Pro Glu Ser Val Val Asn Thr Met Arg Gly Arg Val	360	365	355
60	Asn Phe Asp Asp Tyr Thr Val Asn Leu Gly Gly Leu Lys Asp His Ile	375	380	370
65	Lys Glu Ile Gln Arg Cys Arg Arg Leu Ile Leu Ile Ala Cys Gly Thr	390	395	400
70	Ser Tyr His Ala Gly Val Ala Thr Arg Gln Val Leu Glu Glu Leu Thr	405	410	415
75	Glu Leu Pro Val Met Val Glu Leu Ala Ser Asp Phe Leu Asp Arg Asn	420	425	430
80	Thr Pro Val Phe Arg Asp Asp Val Cys Phe Phe Leu Ser Gln Ser Gly	435	440	445

ES 2 343 571 T3

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

Glu Thr Ala Asp Thr Leu Met Gly Leu Arg Tyr Cys Lys Glu Arg Gly  
 450 455 460

Ala Leu Thr Val Gly Ile Thr Asn Thr Val Gly Ser Ser Ile Ser Arg  
 465 470 475 480

Glu Thr Asp Cys Gly Val His Ile Asn Ala Gly Pro Glu Ile Gly Val  
 485 490 495

Ala Ser Thr Lys Ala Tyr Thr Ser Gln Phe Val Ser Leu Val Met Phe  
 500 505 510

Ala Leu Met Met Cys Asp Asp Arg Ile Ser Met Gln Glu Arg Arg Lys  
 515 520 525

Glu Ile Met Leu Gly Leu Lys Arg Leu Pro Asp Leu Ile Lys Glu Val  
 530 535 540

Leu Ser Met Asp Asp Glu Ile Gln Lys Leu Ala Thr Glu Leu Tyr His  
 545 550 555 560

Gln Lys Ser Val Leu Ile Met Gly Arg Gly Tyr His Tyr Ala Thr Cys  
 565 570 575

Leu Glu Gly Ala Leu Lys Ile Lys Glu Ile Thr Tyr Met His Ser Glu  
 580 585 590

Gly Ile Leu Ala Gly Glu Leu Lys His Gly Pro Leu Ala Leu Val Asp  
 595 600 605

Lys Leu Met Pro Val Ile Met Ile Ile Met Arg Asp His Thr Tyr Ala  
 610 615 620

Lys Cys Gln Asn Ala Leu Gln Gln Val Val Ala Arg Gln Gly Arg Pro  
 625 630 635 640

Val Val Ile Cys Asp Lys Glu Asp Thr Glu Thr Ile Lys Asn Thr Lys  
 645 650 655

Arg Thr Ile Lys Val Pro His Ser Val Asp Cys Leu Gln Gly Ile Leu  
 660 665 670

Ser Val Ile Pro Leu Gln Leu Leu Ala Phe His Leu Ala Val Leu Arg  
 675 680 685

Gly Tyr Asp Val Asp Phe Pro Arg Asn Leu Ala Lys Ser Val Thr Val  
 690 695 700

Glu  
 705

ES 2 343 571 T3

<210> 13

<211> 608

<212> PRT

5 <213> *Escherichia coli*

<400> 13

10 Cys Gly Ile Val Gly Ala Ile Ala Gln Arg Asp Val Ala Glu Ile Leu  
1 5 10 15

15 Leu Glu Gly Leu Arg Arg Leu Glu Tyr Arg Gly Tyr Asp Ser Ala Gly  
20 25 30

20 Leu Ala Val Val Asp Ala Glu Gly His Met Thr Arg Leu Arg Arg Leu  
35 40 45

25 Gly Lys Val Gln Met Leu Ala Gln Ala Ala Glu Glu His Pro Leu His  
50 55 60

30 Gly Gly Thr Gly Ile Ala His Thr Arg Trp Ala Thr His Gly Glu Pro  
65 70 75 80

35 Ser Glu Val Asn Ala His Pro His Val Ser Glu His Ile Val Val Val  
85 90 95

40 His Asn Gly Ile Ile Glu Asn His Glu Pro Leu Arg Glu Glu Leu Lys  
100 105 110

45 Ala Arg Gly Tyr Thr Phe Val Ser Glu Thr Asp Thr Glu Val Ile Ala  
115 120 125

50 His Leu Val Asn Trp Glu Leu Lys Gln Gly Gly Thr Leu Arg Glu Ala  
130 135 140

55 Val Leu Arg Ala Ile Pro Gln Leu Arg Gly Ala Tyr Gly Thr Val Ile  
145 150 155 160

60 Met Asp Ser Arg His Pro Asp Thr Leu Leu Ala Ala Arg Ser Gly Ser  
165 170 175

65 Pro Leu Val Ile Gly Leu Gly Met Gly Glu Asn Phe Ile Ala Ser Asp  
180 185 190

Gln Leu Ala Leu Leu Pro Val Thr Arg Arg Phe Ile Phe Leu Glu Glu  
195 200 205

Gly Asp Ile Ala Glu Ile Thr Arg Arg Ser Val Asn Ile Phe Asp Lys

ES 2 343 571 T3

	210		215		220														
5	Thr	Gly	Ala	Glu	Val	Lys	Arg	Gln	Asp	Ile	Glu	Ser	Asn	Leu	Gln	Tyr			
	225					230					235					240			
10	Asp	Ala	Gly	Asp	Lys	Gly	Ile	Tyr	Arg	His	Tyr	Met	Gln	Lys	Glu	Ile			
					245					250					255				
15	Tyr	Glu	Gln	Pro	Asn	Ala	Ile	Lys	Asn	Thr	Leu	Thr	Gly	Arg	Ile	Ser			
				260					265					270					
20	His	Gly	Gln	Val	Asp	Leu	Ser	Glu	Leu	Gly	Pro	Asn	Ala	Asp	Glu	Leu			
			275					280					285						
25	Leu	Ser	Lys	Val	Glu	His	Ile	Gln	Ile	Leu	Ala	Cys	Gly	Thr	Ser	Tyr			
	290						295					300							
30	Asn	Ser	Gly	Met	Val	Ser	Arg	Tyr	Trp	Phe	Glu	Ser	Leu	Ala	Gly	Ile			
	305					310					315					320			
35	Pro	Cys	Asp	Val	Glu	Ile	Ala	Ser	Glu	Phe	Arg	Tyr	Arg	Lys	Ser	Ala			
				325						330				335					
40	Val	Arg	Arg	Asn	Ser	Leu	Met	Ile	Thr	Leu	Ser	Gln	Ser	Gly	Glu	Thr			
				340					345					350					
45	Ala	Asp	Thr	Leu	Ala	Gly	Leu	Arg	Leu	Ser	Lys	Glu	Leu	Gly	Tyr	Leu			
			355					360					365						
50	Gly	Ser	Leu	Ala	Ile	Cys	Asn	Val	Pro	Gly	Ser	Ser	Leu	Val	Arg	Glu			
		370					375					380							
55	Ser	Asp	Leu	Ala	Leu	Met	Thr	Asn	Ala	Gly	Thr	Glu	Ile	Gly	Val	Ala			
	385					390					395					400			
60	Ser	Thr	Lys	Ala	Phe	Thr	Thr	Gln	Leu	Thr	Val	Leu	Leu	Met	Leu	Val			
				405						410					415				
65	Ala	Lys	Leu	Ser	Arg	Leu	Lys	Gly	Leu	Asp	Ala	Ser	Ile	Glu	His	Asp			
				420					425					430					
70	Ile	Val	His	Gly	Leu	Gln	Ala	Leu	Pro	Ser	Arg	Ile	Glu	Gln	Met	Leu			
			435					440					445						
75	Ser	Gln	Asp	Lys	Arg	Ile	Glu	Ala	Leu	Ala	Glu	Asp	Phe	Ser	Asp	Lys			
	450						455					460							
80	His	His	Ala	Leu	Phe	Leu	Gly	Arg	Gly	Asp	Gln	Tyr	Pro	Ile	Ala	Leu			
	465					470					475					480			



## ES 2 343 571 T3

<211> 43  
<212> ADN  
<213> Secuencia artificial  
5  
<220>  
<223> Cebador  
10 <400>16  
aatctagatt catgctcgag cggccgccag tgtgattgat atc 43  
15 <210> 17  
<211> 36  
<212> ADN  
<213> Secuencia artificial  
20  
<220>  
<223> Cebador  
25 <400>17  
atatttatca gagcgctggg ggtggctatt gacagg 36  
30 <210> 18  
<211> 8  
<212> PRT  
<213> Secuencia artificial  
35  
<220>  
<223> Etiqueta FLAG  
40 <400>18  
**Asp Thr Lys Asp Asp Asp Asp Lys**  
45 **1 5**  
<210> 19  
<211> 6  
50 <212> PRT  
<213> Secuencia artificial  
<220>  
55 <223> Etiqueta hexa-histidina  
<400>19  
60 **His His His His His His**  
**1 5**  
65