



19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA

11 Número de publicación: **2 319 828**

51 Int. Cl.:  
**C12N 15/56** (2006.01)  
**C12N 9/24** (2006.01)  
**C12Q 1/68** (2006.01)  
**C07K 16/40** (2006.01)  
**A61K 38/47** (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Número de solicitud europea: **96938846 .1**  
96 Fecha de presentación : **22.11.1996**  
97 Número de publicación de la solicitud: **0870036**  
97 Fecha de publicación de la solicitud: **14.10.1998**

54 Título:  **$\alpha$ -N-acetilglucosaminidasa sintética de mamíferos y secuencias genéticas que la codifican.**

30 Prioridad: **23.11.1995 AU PN6748**

45 Fecha de publicación de la mención BOPI:  
**12.05.2009**

45 Fecha de la publicación del folleto de la patente:  
**12.05.2009**

73 Titular/es: **WOMEN'S AND CHILDREN'S HOSPITAL  
72 King William Road  
North Adelaide, S.A. 5006, AU**

72 Inventor/es: **Hopwood, John, Joseph;  
Scott, Hamish, Steele;  
Weber, Birgit;  
Blanch, Lianne y  
Anson, Donald, Stewart**

74 Agente: **Elzaburu Márquez, Alberto**

ES 2 319 828 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

$\alpha$ -N-acetilglucosaminidasa sintética de mamíferos y secuencias genéticas que la codifican.

5 La presente invención se refiere en general a la  $\alpha$ -N-acetilglucosaminidasa de mamíferos y a secuencias genéticas que la codifican y al uso de éstas en la investigación, diagnóstico y tratamiento de individuos que padecen, o se sospecha que padecen una deficiencia de  $\alpha$ -N-acetilglucosaminidasa.

10 Los detalles bibliográficos de las publicaciones mencionadas por el autor en esta memoria descriptiva se reúnen al final de la descripción. Los números de identidad de secuencias (SEQ ID NOS.) para las secuencias de nucleótidos y aminoácidos mencionados en la memoria descriptiva se definen a continuación de la bibliografía.

15 A lo largo de esta memoria descriptiva, a menos que el contexto lo requiera de otro modo, se considerará que la palabra “comprender”, o variaciones tales como “comprende” o “que comprende”, implican la inclusión de un elemento indicado o número entero o grupo de elementos o números enteros pero no la exclusión de ningún otro elemento o número entero o grupo de elementos o números enteros.

20 La creciente sofisticación de la tecnología de ADN recombinante está facilitando en gran medida la eficacia de muchas industrias de importancia comercial que incluyen las áreas de investigación y desarrollo médico y farmacéutico. La capacidad para purificar proteínas nativas y posteriormente clonar secuencias genéticas que codifican estas proteínas es un primer paso importante en el desarrollo de una variedad de procedimientos terapéuticos y diagnósticos. Sin embargo, los profesionales se han topado con muchas dificultades para purificar moléculas diana en un grado suficiente para determinar secuencias de aminoácidos que permitan el desarrollo de sondas de oligonucleótidos para asistir en la clonación de secuencias genéticas que codifican las moléculas diana.

25 Se han topado con tales dificultades particularmente en la investigación y desarrollo de enzimas lisosómicas. Una enzima lisosómica importante es la  $\alpha$ -N-acetilglucosaminidasa (EC 2.1.50). Esta enzima actúa como una exoglicosidasa en los lisosomas para hidrolizar los restos de  $\alpha$ -N-acetilglucosamina terminales presentes en el extremo terminal no reductor de los fragmentos de heparán sulfato y heparina (Hopwood, 1989). Una deficiencia de esta hidrolasa lisosómica es responsable de la patogénesis del síndrome de Sanfilippo B (mucopolisacaridosis tipo IIIB (MPS-IIIB) (von-Figura y Kresse, 1972; O'Brien, 1972). Este es un trastorno recesivo autosómico del catabolismo de los glicosaminoglicanos que lleva al almacenamiento y excreción de cantidades excesivas de heparán sulfato y una variedad de fenotipos clínicos, pero que presenta clásicamente retardo mental progresivo en combinación con deformidades esqueléticas (McKusick y Neufeld, 1983).

30 En consecuencia, se necesita purificar la  $\alpha$ -N-acetilglucosaminidasa y clonar secuencias genéticas que la codifiquen para permitir el desarrollo de una gama de procedimientos terapéuticos y diagnósticos para asistir en el diagnóstico y tratamiento de estados patológicos que se originan en la deficiencia de  $\alpha$ -N-acetilglucosaminidasa. En un primer aspecto de la presente invención, se proporciona un método para la producción de  $\alpha$ -N-acetilglucosaminidasa o un fragmento o derivado de la misma en donde la  $\alpha$ -N-acetilglucosaminidasa o fragmento o derivado de la misma hidroliza restos de  $\alpha$ -N-acetilglucosamina a partir del extremo terminal no reductor del heparán sulfato, que comprende:

35 (i) transfectar células de mamífero con un vector de expresión que comprende una molécula de ADN que codifica  $\alpha$ -N-acetilglucosaminidasa o un fragmento o derivado de la misma al menos 40% idéntico a la SEQ ID NO: 1 o SEQ ID NO: 3;

(ii) cultivar las células transfectadas, por lo que se produce la proteína deseada y se secreta al medio; y

(iii) recuperar la proteína deseada del medio.

50 Las realizaciones preferidas de la invención en cualquiera de sus diversos aspectos son como se describen a continuación o como se definen en las sub-reivindicaciones.

55 La Figura 1 es una representación fotográfica de  $\alpha$ -N-acetilglucosaminidasa purificada de placenta humana después de SDS/PAGE. Calle 1: M= estándares (kDa); Calles 2 y 3:  $\alpha$ -N-acetilglucosaminidasa purificada de placenta humana. Calles 4 y 5, albúmina sérica bovina.

60 La Figura 2 es una representación fotográfica de un gel de SDS/poliacrilamida que muestra los pesos moleculares de los polipéptidos de  $\alpha$ -N-acetilglucosaminidasa producidos en células CHO antes (-) y después (+) de la digestión con PNGasa F. Se indican las fracciones de NaCl 50 mM y NaCl 75 mM. Los pesos moleculares de los polipéptidos de  $\alpha$ -N-acetilglucosaminidasa están indicados a la izquierda de la figura. Los pesos moleculares de las proteínas marcadoras están indicados a la derecha de la figura (calle 5).

65 Las abreviaturas de una y tres letras de los restos de aminoácidos convencionales usadas en la presente memoria se definen en la Tabla 1.

Las sustituciones de aminoácidos adecuadas mencionadas en la presente memoria se definen en la Tabla 2.

## ES 2 319 828 T3

Los códigos para los restos de aminoácidos no convencionales usados en la presente se definen en la Tabla 3.

5

TABLA 1

10

Aminoácido	Abreviatura de tres letras	Símbolo de una letra
Alanina	Ala	A
Arginina	Arg	R
15 Asparagina	Asn	N
Ácido aspártico	Asp	D
Cisteína	Cys	C
20 Glutamina	Gln	Q
Ácido glutámico	Glu	E
25 Glicina	Gly	G
Histidina	His	H
Isoleucina	Ile	I
30 Leucina	Leu	L
Lisina	Lys	K
Metionina	Met	M
35 Fenilalanina	Phe	F
Prolina	Pro	P
40 Serina	Ser	S
Treonina	Thr	T
45 Triptofano	Trp	W
Tirosina	Tyr	Y
Valina	Val	V
50 Cualquier resto	Xaa	X

55

60

65

# ES 2 319 828 T3

TABLA 2

*Restos adecuados para sustituciones de aminoácidos*

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

Resto original	Ejemplos de sustituciones
Ala	Ser
Arg	Lys
Asn	Gln; His
Asp	Glu
Cys	Ser
Gln	Asn
Glu	Asp
Gly	Pro
His	Asn; Gln
Ile	Leu; Val
Leu	Ile; Val
Lys	Arg; Gln; Glu
Met	Leu; Ile
Phe	Met; Leu; Tyr
Ser	Thr
Thr	Ser
Trp	Tyr
Tyr	Trp; Phe
Val	Ile; Leu

## ES 2 319 828 T3

TABLA 3

	Aminoácido no convencional	Código	Aminoácido no convencional	Código
5	ácido $\alpha$ -aminobutírico	Abu	L-N-metilalanina	Nmala
	$\alpha$ -amino- $\alpha$ -metilbutirato	Mgabu	L-N-metilarginina	Nmarg
10	aminociclopropan- carboxilato	Cpro	L-N-metilasparagina	Nmasn
	ácido aminoisobutírico	Aib	ácido L-N-metilaspártico	Nmasp
15	aminonorbornil- carboxilato	Norb	L-N-metilcisteína	Nmcys
	ciclohexilalanina	Chexa	L-N-metilglutamina	Nmgln
20	ciclopentilalanina	Cpen	ácido L-N-metilglutámico	Nmglu
	D-alanina	Dal	L-N-metilhistidina	Nmhis
25	D-arginina	Darg	L-N-metilisoleucina	Nmile
	ácido D-aspártico	Dasp	L-N-metilleucina	Nmleu
30	D-cisteína	Dcys	L-N-metillisina	Nmlis
	D-glutamina	Dgln	L-N-metilmetionina	Nmmet
	ácido D-glutámico	Dglu	L-N-metilnorleucina	Nmnle
35	D-histidina	Dhis	L-N-metilnorvalina	Nmnva
	D-isoleucina	Dile	L-N-metilornitina	Nmorn
40	D-leucina	Dleu	L-N-metilfenilalanina	Nmfe
	D-lisina	Dlis	L-N-metilprolina	Nmpro
	D-metionina	Dmet	L-N-metilserina	Nmser
45	D-ornitina	Dorn	L-N-metiltreonina	Nmtr
	D-fenilalanina	Dfe	L-N-metiltryptofano	Nmtrp
	D-prolina	Dpro	L-N-metiltirosina	Nmtir
50	D-serina	Dser	L-N-metilvalina	Nmval
	D-treonina	Dtr	L-N-metiletilglicina	Nmetg
	D-tryptofano	Dtrp	L-N-metil-t-butilglicina	Nmtbug
55	D-tirosina	Dtyr	L-norleucina	Nle
	D-valina	Dval	L-norvalina	Nva
60	D- $\alpha$ -metilalanina	Dmala	$\alpha$ -metil-aminoisobutirato	Maib
	D- $\alpha$ -metilarginina	Dmarg	$\alpha$ -metil- $\gamma$ -aminoisobutirato	Mgabu
	D- $\alpha$ -metilasparagina	Dmasn	$\alpha$ -metilciclohexilalanina	Mchexa
65	D- $\alpha$ -metilaspartato	Dmasp	$\alpha$ -metilciclopentilalanina	Mcpen
			$\alpha$ -metil- $\alpha$ -naftilalanina	Manap
			$\alpha$ -metilpenicilamina	Mpen

ES 2 319 828 T3

	D- $\alpha$ -metilcisteína	Dmcys	N-(4-aminobutil)glicina	Nglu
5	D- $\alpha$ -metilglutamina	Dmgln	N-(2-aminoetil)glicina	Naeg
	D- $\alpha$ -metilhistidina	Dmhis	N-(3-aminopropil)glicina	Norn
	D- $\alpha$ -metilisoleucina	Dmile	N-amino- $\alpha$ -metilbutirato	Nmaabu
10	D- $\alpha$ -metilleucina	Dmleu	$\alpha$ -naftilalanina	Anap
	D- $\alpha$ - metillisina	Dmlys	N-bencilglicina	Nphe
	D- $\alpha$ -metilmetionina	Dmmet	N-(2-carbamiletil)glicina	Ngln
15	D- $\alpha$ -metilornitina	Dmorn	N-(carbamilmetil)glicina	Nasn
	D- $\alpha$ -metilfenilalanina	Dmphe	N-(2-carboxietil)glicina	Nglu
20	D- $\alpha$ -metilprolina	Dmpro	N-(carboximetil)glicina	Nasp
	D- $\alpha$ -metilserina	Dmser	N-ciclobutilglicina	Ncbut
	D- $\alpha$ -metiltreonina	Dmtr	N-cicloheptilglicina	Nchep
25	D- $\alpha$ -metiltriptofano	Dmtrp	N-ciclohexilglicina	Nchex
	D- $\alpha$ -metiltirosina	Dmty	N-ciclododecilglicina	Ncdec
30	D- $\alpha$ -metilvalina	Dmval	N-cilcododecilglicina	Nedod
	D-N-metilalanina	Dnmala	N-ciclooctilglicina	Ncoct
	D-N-metilarginina	Dnmarg	N-ciclopropilglicina	Ncpro
35	D-N-metilasparagina	Dnmasn	N-cicloundecilglicina	Ncund
	D-N-metilaspartato	Dnmasp	N-(2,2-difeniletil)glicina	Nbhm
	D-N-metilcisteína	Dnmcys	N-(3,3-difenilpropil)-glicina	Nbhe
40	D-N-metilglutamina	Dnmgln	N-(3-guanidinopropil)- glicina	Narg
	D-N-metilglutamato	Dnmglu	N-(1-hidroxietil)glicina	Nthr.
45	D-N-metilhistidina	Dnmhis	N-(hidroxietil)glicina	Nser
	D-N-metilisoleucina	Dnmile	N-(imidazoliletil)glicina	Nhis
50	D-N-metilleucina	Dnmleu	N-(3-indoliletil)glicina	Nhtrp
	D-N-metillisina	Dnmlys	N-metil- $\gamma$ -aminobutirato	Nmgabu
	N-metilciclohexilalanina	Nmchexa	D-N-metilmetionina	Dnmmet
55	D-N-metilornitina	Dnmorn	N-metilciclopentilalanina	Nmcpen
	N-metilglicina	Nala	D-N-metilfenilalanina	Dnmphe
	N-metilaminoisobutirato	Nmaib	D-N-metilprolina	Dnmpro
60	N-(1-metilpropil)glicina	Nile	D-N-metilserina	Dnmser
	N-(2-metilpropil)glicina	Nleu	D-N-metiltreonina	Dnmtyr
65	D-N-metiltriptofano	Dnmtrp	N-(1-metiletil)glicina	Nval
	D-N-metiltirosina	Dnmtyr	N-metil-naftilalanina	Nmanap

5	D-N-metilvalina	Dnmval	N-metilpenicilamina	Nmpen
	ácido $\gamma$ -aminobutírico	Gabu	N-( <i>p</i> -hidroxifenil)glicina	Nhtyr
	L- <i>t</i> -butilglicina	Tug	N-(tiometil)glicina	Ncis
	L-etilglicina	Etg	penicilamina	Pen
10	L-homofenilalanina	Hfe	L- $\alpha$ -metilalanina	Mala
	L- $\alpha$ -metilarginina	Marg	L- $\alpha$ -metilasparagina	Masn
	L- $\alpha$ -metilaspartato	Masp	L- $\alpha$ -metil- <i>t</i> -butilglicina	Mtbug
15	L- $\alpha$ -metilcisteína	Mcys	L-metiletilglicina	Metg
	L- $\alpha$ -metilglutamina	Mgln	L- $\alpha$ -metilglutamato	Mglu
20	L- $\alpha$ -metilhistidina	Mhis	L- $\alpha$ -metilhomofenilalanina	Mhfe
	L- $\alpha$ -metilisoleucina	Mile	N-(2-metiltioetil)glicina	Nmet
	L- $\alpha$ -metilleucina	Mleu	L- $\alpha$ -metillisina	Mlys
25	L- $\alpha$ -metilmetionina	Mmet	L- $\alpha$ -metilnorleucina	Mnle
	L- $\alpha$ -metilnorvalina	Mnva	L- $\alpha$ -metilornitina	Morn
	L- $\alpha$ -metilfenilalanina	Mfe	L- $\alpha$ -metilprolina	Mpro
30	L- $\alpha$ -metilserina	Mser	L- $\alpha$ -metiltreonina	Mtr
	L- $\alpha$ -metiltriptofano	Mtrp	L- $\alpha$ -metiltirosina	Mtyr
35	L- $\alpha$ -metilvalina	Mval	L-N-metilhomofenilalanina	Nmhfe
	N-(N-(2,2-difeniletíl)- carbamilmetil)glicina	Nnbhm	N-(N-(3,3-difenilpropil)- carbamilmetil)glicina	Nnbhe
40	1-carboxi-1-(2,2-difenil- etilamino)ciclopropano	Nmbc		

45

La presente invención proporciona una molécula de ácido nucleico aislado que comprende una secuencia de nucleótidos que codifica, o son complementarios de una secuencia que codifica, una  $\alpha$ -N-acetilglucosaminidasa de mamífero o fragmento o derivado de la misma o su molécula semejante.

50

55

Preferentemente, el mamífero es un ser humano, animal de cría, animal de compañía, animal salvaje o animal empleado en ensayos de laboratorio (por ej., conejo, rata, ratón o cobaya). Con máxima preferencia, el mamífero es un ser humano. En forma conveniente, la  $\alpha$ -N-acetilglucosaminidasa se puede aislar del hígado, riñón o placenta. Sin embargo, la presente invención se extiende a todas las enzimas  $\alpha$ -N-acetilglucosaminidasas de mamíferos y de cualquier fuente anatómica o celular y/o de cualquier fuente de fluido biológico, tal como, sin limitación, plasma, suero, extracto celular o fluido linfático.

60

Si bien la presente invención contempla el uso de  $\alpha$ -N-acetilglucosaminidasa humana o secuencias genéticas genómicas o recombinantes (por ej., ADNc) que codifican la misma en la investigación, diagnóstico y/o tratamiento de individuos humanos (es decir, un sistema homólogo), los expertos en la técnica apreciarán que la enzima o secuencias genéticas que la codifican provenientes de un animal no humano también pueden ser útiles. Tal sistema heterólogo está comprendido en la presente invención.

65

El término "molécula de ácido nucleico" como se usa en la presente memoria se considera que se refiere a cualquier molécula de ARN o ADN (por ej., ADNc), sea monocatenaria o bicatenaria o en forma lineal o cerrada covalentemente. La molécula de ácido nucleico también puede ser ADN correspondiente al gen genómico entero o una porción sustancial del mismo o un fragmento o derivado del mismo.

## ES 2 319 828 T3

La molécula de ácido nucleico de la presente invención puede constituir únicamente la secuencia de nucleótidos que codifica  $\alpha$ -N-acetilglucosaminidasa o una molécula semejante a  $\alpha$ -N-acetilglucosaminidasa o puede ser parte de una molécula de ácido nucleico más grande. Por consiguiente, la presente invención se extiende al gen de  $\alpha$ -N-acetilglucosaminidasa genómico aislado. Las secuencias no traducidas de una molécula de ácido nucleico más grande pueden incluir vector, secuencias reguladoras de transcripción y/o traducción, promotor, terminador, potenciador, secuencias de replicación o señal o regiones no codificadoras (por ej., secuencias de intrón) de un gen genómico aislado.

La referencia en la presente memoria a un “gen” se toma en su contexto más amplio e incluye:

(i) un gen genómico clásico que consiste en secuencias reguladoras de transcripción y/o traducción, y/o una región codificadora y/o secuencias no traducidas (es decir, intrones, secuencias no traducidas 5' y 3');

(ii) ARNm o ADNc correspondiente a las regiones codificadoras (es decir, exones) que comprenden opcionalmente secuencias no traducidas 5' o 3' del gen; o

(iii) fragmentos sintéticos de ADN amplificado, u otras moléculas de ácido nucleico recombinante producidas *in vitro* y que comprenden toda o parte de la región codificadora y/o secuencias no traducidas 5' o 3' del gen.

El término “gen” también se usa para describir las moléculas sintéticas o de fusión que codifican el total o parte de un producto funcional. Un producto funcional es uno que comprende una secuencia de nucleótidos o es complementario a una secuencia de nucleótidos que codifica un polipéptido funcional, en particular un polipéptido que tiene actividad catalítica de  $\alpha$ -N-acetilglucosaminidasa o un homólogo, análogo o derivado de este.

Para el presente propósito, se considerará que “homólogos” de una secuencia de nucleótidos se refiere a una molécula de ácido nucleico aislada que es sustancialmente igual a la molécula de ácido nucleico de la presente invención o su secuencia de nucleótidos complementaria, a pesar de la aparición dentro de dicha secuencia, de una o más sustituciones, inserciones, supresiones o reordenamientos de nucleótidos.

Se considerará que los “análogos” de una secuencia de nucleótidos expuestos en la presente memoria se refieren a una molécula de ácido nucleico aislada que es sustancialmente igual a una molécula de ácido nucleico de la presente invención o su secuencia de nucleótidos complementaria, a pesar de la aparición de cualquier constituyente no nucleotídico normalmente ausente en dicha molécula de ácido nucleico aislado, por ejemplo carbohidratos, compuestos radioquímicos que incluyen radionucleótidos, moléculas indicadoras tales como, sin limitación, DIG, fosfatasa alcalina o peroxidasa de rábano picante, entre otras.

Se considera que los “derivados” de una secuencia de nucleótidos expuestos en la presente memoria se refieren a cualquier molécula de ácido nucleico aislado que contiene similitud de secuencia significativa con dicha secuencia o una parte de la misma. Generalmente, la secuencia de nucleótidos de la presente invención se puede someter a mutagénesis para producir sustituciones, supresiones y/o inserciones de nucleótidos únicas o múltiples. Los derivados por inserción de nucleótidos de la secuencia de nucleótidos de la presente invención incluyen las fusiones terminales 5' y 3' además de las inserciones intra-secuencia de nucleótidos únicos o múltiples o análogos de nucleótidos. Las variantes de secuencias de nucleótidos por inserción son aquellas en las que uno o más nucleótidos o análogos de nucleótidos se introducen en un sitio predeterminado en la secuencia de nucleótidos de dicha secuencia, aunque la inserción aleatoria también es posible con la adecuada identificación del producto resultante preformado. Las variantes por supresión se caracterizan por la eliminación de uno o más nucleótidos de la secuencia de nucleótidos. Las variantes por sustitución de nucleótidos son aquellas en las que al menos un nucleótido de la secuencia ha sido eliminado y se ha colocado un nucleótido diferente o análogo de nucleótido en su lugar.

Preferentemente, un homólogo, análogo o derivado de un gen de  $\alpha$ -N-acetilglucosaminidasa de acuerdo con cualquiera de las realizaciones descritas en la presente memoria, comprende una secuencia de nucleótidos de al menos 10 nucleótidos contiguos derivados de la SEQ ID NO: 1 o la SEQ ID NO: 3 o una hebra complementaria de las mismas, en donde la secuencia de dicho homólogo, análogo o derivado es al menos 40% idéntica a la SEQ ID NO: 1 o la SEQ ID NO: 3 o una hebra complementaria de esta o donde dicho homólogo, análogo o derivado es capaz de hibridarse a dicha secuencia en condiciones de hibridación de al menos baja rigurosidad.

Para el fin de la nomenclatura, la secuencia de nucleótidos expuesta en la SEQ ID NO: 1 se refiere al ADNc que codifica la enzima  $\alpha$ -N-acetilglucosaminidasa humana.

La secuencia de nucleótidos expuesta en la SEQ ID NO: 3 se refiere al gen genómico equivalente del ADNc que codifica la enzima  $\alpha$ -N-acetilglucosaminidasa hepática humana. Los expertos en la técnica sabrán que las secuencias de exón específicas descritas en la SEQ ID NO: 3 corresponden a las regiones codificadoras del gen de  $\alpha$ -N-acetilglucosaminidasa, comprendiendo además dichas regiones de exón el marco de lectura abierto completo de la secuencia de ADNc expuesta en la SEQ ID NO: 1, cuando se alinea en la configuración cabeza-cola. Las secuencias de intrón de la SEQ ID NO: 3, que corresponden a las regiones no codificadoras del gen empalmadas del producto de transcripción primario de este, aunque no definidas explícitamente, pueden ser fácilmente deducidas por los expertos en la técnica, cuando se proveen con los datos de secuencia de exón provistos en el listado de secuencias de nucleótidos.

## ES 2 319 828 T3

La secuencia de nucleótidos de la presente invención puede corresponder a la secuencia del gen de  $\alpha$ -N-acetilglucosaminidasa natural o puede comprender un homólogo, análogo o derivado del mismo que contiene sustituciones, supresiones y/o adiciones de nucleótidos únicas o múltiples. Todos estos homólogos, análogos o derivados codifican  $\alpha$ -N-acetilglucosaminidasa o moléculas semejantes a  $\alpha$ -N-acetilglucosaminidasa o un homólogo, análogo o derivado de la misma contemplados por la presente invención. La longitud de la secuencia de nucleótidos puede variar desde unas pocas bases, tales como en sondas o cebadores de ácidos nucleicos, hasta una secuencia de longitud completa.

La presente invención se refiere al ácido nucleico en forma de ADNc y en particular cuando se inserta en un vector de expresión. El vector de expresión se puede replicar en una célula eucariótica o procariótica y puede producir ARNm o el ARNm se puede traducir posteriormente en  $\alpha$ -N-acetilglucosaminidasa o molécula semejante. Las células eucarióticas particularmente preferidas incluyen células CHO pero pueden estar en otras células o líneas celulares de mamíferos o células no mamíferas tales como células de levaduras o insectos.

Alternativamente, la presente invención proporciona una molécula de ácido nucleico que comprende una secuencia de nucleótidos que codifica o es complementaria a una secuencia que codifica un polipéptido capaz de hidrolizar los restos de  $\alpha$ -N-acetilglucosamina de los extremos terminales no reductores del heparán sulfato y fragmentos de heparina y en donde dicha secuencia de nucleótidos es capaz de hibridarse en condiciones de al menos baja rigurosidad a una secuencia de nucleótidos expuesta en la SEQ ID NO: 1 o SEQ ID NO: 3 o un homólogo, análogo o derivado de la misma.

La invención también proporciona una molécula de ácido nucleico aislado que comprende una secuencia de nucleótidos que es capaz de hibridarse en condiciones de al menos baja rigurosidad a una secuencia de nucleótidos expuesta en la SEQ ID NO: 1 o SEQ ID NO: 3 o una hebra complementaria de un homólogo, análogo o derivado de la misma.

Preferentemente, la hibridación es posible en al menos condiciones de rigurosidad media. Con más preferencia, la hibridación es posible en condiciones de alta rigurosidad.

Para el fin de definir el nivel de rigurosidad, se puede hacer referencia en forma conveniente a Sambrook *et al* (1989) o Ausubel *et al* (1987).

Una baja rigurosidad se define en la presente memoria como una hibridación y/o lavado que se lleva a cabo en 4-6X SSC/0,1-0,5% p/v SDS a 37-45°C durante 2-3 horas. Una hibridación y/o lavado de rigurosidad media se lleva a cabo en 1-4X SSC/0,25-0,5% p/v SDS a 45°C durante 2-3 horas y una hibridación y/o lavado de rigurosidad alta se lleva a cabo en 0,1-1X SSC/0,1% p/v SDS a 60°C durante 1-3 horas.

Se pueden emplear condiciones de rigurosidad alternativas a las mencionadas específicamente en la presente memoria. En general, la rigurosidad aumenta por reducción de la concentración del tampón SSC y/o aumento de la concentración de SDS y/o aumento de la temperatura de hibridación y/o lavado. Los expertos en la técnica sabrán que las condiciones de hibridación y/o lavado pueden variar de acuerdo con la naturaleza de la membrana de hibridación o el tipo de sonda de hibridación usada. Las condiciones para la hibridación y lavados son bien conocidas por los expertos en la técnica. Para el fin de clarificar los parámetros que afectan la hibridación entre las moléculas de ácido nucleico, la referencia se halla en las páginas 2.10.8 a 2.10.16. de Ausubel *et al.* (1987).

Los expertos en la técnica sabrán que las secuencias de nucleótidos expuestas en la SEQ ID NO: 1 y SEQ ID NO: 3 se pueden usar para aislar los genes correspondientes de otros tejidos humanos o alternativamente, de los tejidos o células de otras especies, sin experimentación indebida. Los medios para el aislamiento de dichas secuencias relacionadas también serán conocidos por los expertos en la técnica, por ejemplo hibridación de ácidos nucleicos, reacción en cadena de la polimerasa, selección de anticuerpos a partir de genotecas de expresión, selección funcional a partir de genotecas de expresión o complementación de mutantes, entre otros. La presente invención no está limitada por la fuente de la que se han aislado las secuencias génicas específicas descritas en la presente memoria o por los medios usados para aislar dichas secuencias.

La invención también se refiere a la situación en donde una secuencia genética relacionada que comprende ADN genómico o ARNm o ADNc se pone en contacto con una cantidad efectiva de hibridación de una secuencia genética que codifica  $\alpha$ -N-acetilglucosaminidasa, o su secuencia de nucleótidos complementaria o un homólogo, análogo, derivado o parte funcional de la misma y posteriormente dicha hibridación se detecta usando un medio de detección adecuado.

La secuencia genética relacionada puede estar en forma recombinante, en una partícula viral, partícula de bacteriófago, célula de levadura; célula de animal o una célula de planta. Preferentemente, la secuencia genética relacionada proviene de una especie animal o un ser humano. Con mayor preferencia, la secuencia genética relacionada proviene de un ser humano.

Preferentemente, la secuencia genética que codifica  $\alpha$ -N-acetilglucosaminidasa (es decir, sonda o última secuencia genética) comprende una secuencia de nucleótidos de al menos 10 nucleótidos, con más preferencia al menos 20 nucleótidos, aún con más preferencia al menos 50 nucleótidos y aún con más preferencia al menos 100 nucleótidos de-

## ES 2 319 828 T3

rivados de la secuencia de nucleótidos expuesta en la SEC ID. NO: 1 o SEQ ID NO: 3 o una secuencia complementaria o un homólogo, análogo o derivado de la misma.

Preferentemente, el medio de detección es una molécula indicadora capaz de dar una señal identificable (por ej., radioisótopos tales como  $^{32}\text{P}$  o  $^{35}\text{S}$  o una molécula biotinilada) unida en forma covalente a la sonda de  $\alpha$ -N-acetilglucosaminidasa.

Alternativamente, el medio de detección es una reacción en cadena de la polimerasa.

En particular, dos “moléculas cebadoras” de ácidos nucleicos no complementarias opuestas de al menos 10 nucleótidos de longitud, con más preferencia al menos 20 nucleótidos de longitud, derivadas de la secuencia de nucleótidos expuesta en la SEQ ID NO: 1 o SEQ ID NO: 3, pueden ponerse en contacto con una “molécula molde” de ácido nucleico y copias de moléculas de ácido nucleico específicas de la molécula del molde amplificarse en una reacción en cadena de la polimerasa.

Las moléculas cebadoras opuestas se seleccionan de modo que cada una sea capaz de hibridarse a las hebras complementarias de la misma molécula molde, en donde la síntesis de ADN dependiente de ADN polimerasa que se produce a partir de una primera molécula cebadora opuesta será en una dirección hacia la segunda molécula cebadora opuesta.

Por consiguiente, ambos cebadores hibridizan a dicha molécula molde de modo que, en presencia de una enzima ADN polimerasa, un cofactor y sustrato apropiado, la síntesis de ADN se produce en la dirección 5' a 3' a partir de cada molécula cebadora hacia la posición del ADN donde la otra molécula cebadora se hibrida, amplificándose de este modo el ADN interviniente.

Los expertos en la técnica saben de los requerimientos técnicos de la reacción en cadena de la polimerasa y son capaces de cualquiera de las modificaciones que se pueden realizar a las condiciones de reacción. Por ejemplo, la reacción en cadena de la polimerasa se puede usar en cualquier formato adecuado, tal como un polimorfismo de longitud de fragmento amplificado (AFLP), polimorfismo de cadena monocatenaria (SSCP), detección de amplificación y errores de apareamiento (AMD), reacción en cadena de la polimerasa de secuencia repetida interpuesta (IRS-PCR), reacción en cadena de la polimerasa inversa (iPCR) y reacción en cadena de la polimerasa de transcripción inversa (RT-PCR), entre otros, para aislar una secuencia relacionada del gen de  $\alpha$ -N-acetilglucosaminidasa o identificar una mutación de una secuencia genética de  $\alpha$ -N-acetilglucosaminidasa. Tales variaciones de la reacción en cadena de la polimerasa son tratadas en detalle por McPherson *et al* (1991), que se incorpora en la presente memoria como referencia. La presente invención abarca todas estas variaciones, siendo el único requerimiento que el producto final de la reacción sea un ácido nucleico aislado que sea capaz de codificar  $\alpha$ -N-acetilglucosaminidasa o un homólogo, análogo o derivado de la misma.

Preferentemente, la primera molécula cebadora con preferencia deriva de la hebra homsentido de un gen que codifica  $\alpha$ -N-acetilglucosaminidasa, en particular de la secuencia de nucleótidos expuesta en la SEQ ID NO: 1 o SEQ ID NO: 3 o un homólogo, derivado o análogo de la misma, y la segunda molécula cebadora preferentemente deriva de la hebra antisentido de dicho gen.

Los expertos en la técnica sabrán que no es esencial para la realización de la invención que las moléculas cebadoras deriven del mismo gen.

La molécula cebadora de ácido nucleico además puede consistir en una combinación de cualquiera de los nucleótidos adenina, citidina, guanina, timidina o inosina, o análogos funcionales o derivados de los mismos, capaces de incorporarse en una molécula de polinucleótido con la condición de que sea capaz de hibridarse en condiciones de al menos baja rigurosidad a la molécula de ácido nucleico expuesta en la SEQ ID NO: 1 o SEQ ID NO: 3 o un homólogo, análogo o derivado de la misma.

Las moléculas cebadoras de ácido nucleico además pueden estar cada una contenidas en una mezcla acuosa que comprende otras moléculas cebadoras de ácido nucleico. Con mayor preferencia, la molécula cebadora de ácido nucleico está en forma sustancialmente pura.

La molécula de ácido nucleico molde puede estar en forma recombinante, en una partícula viral, partícula de bacteriófago, célula de levadura, célula animal o célula de planta. Preferentemente, la secuencia genética relacionada proviene de una célula, tejido u órgano derivado de una especie animal o un ser humano. Con mayor preferencia, la secuencia genética relacionada proviene de una célula, tejido u órgano derivado de un ser humano.

Por consiguiente, la presente invención se extiende a una molécula de ácido nucleico aislado que es al menos 40% idéntica a la secuencia de nucleótidos expuesta en la SEQ ID NO: 1 o SEQ ID NO: 3 o a una hebra complementaria de las mismas o un homólogo, análogo o derivado de las mismas.

Preferentemente, el porcentaje de identidad con la SEQ ID NO: 1 o la SEQ ID NO: 3 es al menos aproximadamente 55%, aún con mayor preferencia al menos aproximadamente 65%, aún con mayor preferencia al menos aproximadamente 75-80% y aún con mayor preferencia al menos aproximadamente 85-95%.

## ES 2 319 828 T3

Aún con mayor preferencia, la presente invención proporciona una molécula de ácido nucleico aislado que es al menos 40% idéntica a la secuencia de nucleótidos expuesta en la SEQ ID NO: 1 o SEQ ID NO: 3 o a una hebra complementaria de las mismas o un homólogo, análogo o derivado de las mismas y es capaz de hibridarse en condiciones de al menos baja rigurosidad a una secuencia de nucleótidos expuesta en la SEQ ID NO: 1 o SEQ ID NO: 3.

En particular, la molécula de ácido nucleico aislada descrita en la presente memoria es además capaz de codificar una secuencia de aminoácidos que es capaz de llevar a cabo la reacción enzimática catalizada por una enzima  $\alpha$ -N-acetilglucosaminidasa.

La molécula de ácido nucleico aislada de la presente invención también es útil para desarrollar una construcción genética que comprende una molécula homosentido, para la expresión o sobreexpresión de  $\alpha$ -N-acetilglucosaminidasa en células procarióticas o eucarióticas. Las células eucarióticas particularmente preferidas incluyen células CHO pero pueden estar en cualquier otra célula o líneas celulares de mamífero adecuadas o células no mamíferas tales como células de levaduras o insectos.

Se considera que el término “molécula homosentido” como se usa en la presente memoria se refiere a una molécula de ácido nucleico aislada de la invención descrita en la presente memoria, que se provee en un formato adecuado para su expresión para producir un polipéptido recombinante, cuando dicha molécula homosentido se introduce en una célula hospedante.

En particular, una molécula homosentido que codifica la  $\alpha$ -N-acetilglucosaminidasa comprende una secuencia de nucleótidos expuesta en la SEQ ID NO: 1 o SEQ ID NO: 3 o una hebra complementaria, homólogo, análogo o derivado de las mismas.

Más particularmente, la molécula homosentido de la invención comprende la secuencia de nucleótidos expuesta en la SEQ ID NO: 1 o una hebra complementaria, homólogo, análogo o derivado de la misma.

Los expertos en la técnica sabrán que la expresión de una molécula homosentido puede requerir que la molécula de ácido nucleico de la invención sea colocada en una conexión operable con una secuencia promotora para producir una “construcción homosentido”. La elección del promotor para el presente fin puede variar de acuerdo con el nivel de expresión requerido de la molécula homosentido y/o la especificidad de tejido o especificidad de desarrollo de la expresión de la molécula homosentido que se requiere. La construcción homosentido además puede comprender una secuencia terminadora e introducirse en una célula hospedante adecuada donde sea capaz de expresarse para producir un polipéptido recombinante como producto génico.

En el contexto de la presente invención, una molécula homosentido que corresponde a una secuencia genética o molécula de ácido nucleico aislado que codifica el polipéptido  $\alpha$ -N-acetilglucosaminidasa o un homólogo, análogo o derivado del mismo, colocado en forma operativa bajo el control de una secuencia promotora adecuada, se introduce en una célula mediante cualquier método para la transformación de dicha célula y dicha secuencia genética o molécula de ácido nucleico aislado se expresa allí para producir dicho polipéptido.

La presente invención se extiende claramente a construcciones genéticas diseñadas para facilitar la expresión de cualquier molécula de ácido nucleico descrita en la presente memoria.

Una construcción genética de la presente invención comprende la anterior molécula homosentido, colocada operativamente bajo el control de una secuencia promotora capaz de regular la expresión de dicha molécula de ácido nucleico en una célula procariota o eucariota, preferentemente una célula de mamífero tal como una célula CHO, una célula de levadura, una célula de insecto o una célula bacteriana. Dicha construcción genética opcionalmente comprende, además de un promotor y una molécula homosentido, una secuencia terminadora.

El término “terminadora” se refiere a una secuencia de ADN en el extremo de una unidad de transcripción que señala la terminación de la transcripción. Las terminadoras son secuencias de ADN no traducidas 3' que contienen una señal de poliadenilación, que facilita la adición de secuencias de poliadenilato al extremo 3' de un transcrito primario. Las terminadoras activas en células de plantas son conocidas y se describen en la bibliografía. Pueden aislarse de bacterias, hongos, virus, animales y/o plantas.

La referencia en la presente memoria a un “promotor” debe tomarse en su contexto más amplio e incluye las secuencias reguladoras de transcripción de un gen genómico clásico, que incluye la caja TATA que se requiere para la iniciación de la transcripción exacta, con o sin una secuencia de la caja CCAAT y elementos reguladores adicionales (es decir, secuencias activadoras corriente arriba, potenciadores y silenciadores) que alteran la expresión génica en respuesta a estímulos del desarrollo y/o externos, o en forma específica de tejido. Un promotor usualmente, pero no necesariamente, se ubica corriente arriba o 5', de un gen estructural, cuya expresión regula. Además, los elementos reguladores que comprenden un promotor se ubican usualmente dentro de 2 kb del sitio de inicio de la transcripción del gen.

En el presente contexto, el término “promotor” también se usa para describir una molécula sintética o de fusión o derivado que confiere, activa o potencia la expresión de dicha molécula homosentido en una célula.

## ES 2 319 828 T3

Los promotores preferidos pueden contener copias adicionales de uno o más elementos reguladores específicos, para aumentar adicionalmente la expresión de la molécula homosentido y/o para alterar la expresión espacial y/o temporal de dicha molécula homosentido. Por ejemplo, los elementos reguladores que confieren inducibilidad por cobre se pueden colocar adyacentes a una secuencia promotora heteróloga que dirige la expresión de una molécula homosentido, confiriendo de este modo inducibilidad por cobre sobre la expresión de dicha molécula.

Colocar una molécula homosentido bajo el control regulador de una secuencia promotora significa ubicar dicha molécula de modo que la expresión sea controlada por la secuencia promotora. Los promotores generalmente se ubican 5' (corriente arriba) de los genes que controlan. En la construcción de combinaciones génicas heterólogas de promotor/estructurales se prefiere generalmente ubicar el promotor a una distancia del sitio de comienzo de transcripción génica que sea aproximadamente igual a la distancia que existe entre el promotor y el gen que controla en su marco natural, es decir, el gen del cual deriva el promotor. Como se sabe en la técnica, esta distancia puede contener alguna variación sin pérdida de la función promotora. De modo similar, la ubicación preferida de un elemento regulador de la secuencia con respecto a un gen heterólogo que se ha de colocar bajo su control está definida por la ubicación del elemento en su marco natural, es decir, los genes de los que deriva. Nuevamente, como se sabe en la técnica, también puede producirse alguna variación de esta distancia.

Los ejemplos de promotores adecuados para usar en las construcciones genéticas de la presente invención incluyen promotores virales, fúngicos, bacterianos, derivados de animales y plantas capaces de funcionar en células animales, humanas, de levadura, de insectos o bacterianas. El promotor puede regular la expresión de dicha molécula en forma constitutiva o diferencial con respecto al tejido en que se produce la expresión o, con respecto a la etapa de desarrollo en que se produce la expresión o en respuesta a estímulos externos tales como estrés fisiológicos, o agentes patógenos de plantas, o iones metálicos, entre otros. Preferentemente, el promotor es capaz de regular la expresión de una molécula homosentido en una célula derivada de una especie animal o un ser humano.

En particular, el promotor deriva del gen genómico que codifica  $\alpha$ -N-acetilglucosaminidasa, preferentemente el gen de  $\alpha$ -N-acetilglucosaminidasa humana. En una realización más preferida, sin embargo, el promotor deriva de la secuencia de nucleótidos expuesta en la SEQ ID NO: 3 o es al menos capaz de hibridarse a los restos de nucleótidos 1 a 989 de la SEQ ID NO: 3 o al menos 20 nucleótidos contiguos derivados de los mismos.

Aún más particularmente, el promotor es una secuencia promotora del CMV o una secuencia promotora derivada del mismo.

En forma alternativa, la invención se refiere a una construcción que comprende un promotor o derivado funcional, fragmento parcial, homólogo o análogo del mismo, derivado del gen genómico de  $\alpha$ -N-acetilglucosaminidasa definido por la SEQ ID NO: 3.

Preferentemente, dicha construcción genética además comprende la secuencia de  $\alpha$ -N-acetilglucosaminidasa definida por la SEQ ID NO: 1 ubicada en conexión operativa con dicho promotor.

La presente invención también se refiere a la  $\alpha$ -N-acetilglucosaminidasa sintética o molécula semejante.

El término "sintético" según se usa en la presente memoria incluirá tanto moléculas recombinantes como sintetizadas químicamente producidas por la adición sucesiva de restos de aminoácidos o grupos de restos de aminoácidos en orden definido.

La invención se refiere a la  $\alpha$ -N-acetilglucosaminidasa recombinante o molécula semejante codificada por o expresada a partir de las moléculas de ácido nucleico descritas con anterioridad en la presente memoria.

La  $\alpha$ -N-acetilglucosaminidasa sintética o molécula semejante comprende una secuencia de aminoácidos que es al menos 40% idéntica a la secuencia de aminoácidos expuesta en alguna de las SEQ ID NOs: 2, 4, 5 ó 6.

Con mayor preferencia, el porcentaje de identidad es al menos 60% y aún con mayor preferencia al menos 80% o 85-90%.

En particular, la presente invención proporciona una  $\alpha$ -N-acetilglucosaminidasa sintética definida con anterioridad en la presente memoria que comprende una secuencia de aminoácidos sustancialmente igual a la expuesta en cualquiera de las SEQ ID NOs: 2, 4, 5 ó 6 o un homólogo, análogo o derivado de las mismas.

Para el fin de la nomenclatura, la secuencia de aminoácidos expuesta en la SEQ ID NO: 2; comprende el producto de traducción de longitud completa del gen de la  $\alpha$ -N-acetilglucosaminidasa humana (es decir, de aquí en adelante denominado "polipéptido  $\alpha$ -N-acetilglucosaminidasa" o "SEQ ID NO: 2) producido por expresión de la secuencia de ADNc definida por SEQ ID NO: 1 o el gen genómico definido por SEQ ID NO: 3. El polipéptido de  $\alpha$ -N-acetilglucosaminidasa comprende al menos siete restos de Asn potencialmente glicosilados, en las posiciones 261, 272, 435, 503, 513, 526 y 532. Además, la secuencia de aminoácidos del polipéptido de  $\alpha$ -N-acetilglucosaminidasa puede comprender un péptido señal de aproximadamente 23 restos de aminoácidos de longitud, apareciendo un probable sitio para la ruptura del péptido señal por la peptidasa entre Gly<sub>23</sub> y Asp<sub>24</sub>.

## ES 2 319 828 T3

Las secuencias de aminoácidos expuestas en las SEQ ID NOs: 4-6 se refieren a secuencias de aminoácidos N-terminales e internas (es decir, CNBr) derivadas de  $\alpha$ -N-acetilglucosaminidasa humana, purificadas según se describe en el Ejemplo 1. Según se describe en el Ejemplo 2, la forma purificada de la enzima comprende dos polipéptidos que tienen pesos moleculares aproximados de 82 y 77 kDa. La secuencia expuesta en la SEQ ID NO: 4 se refiere a la secuencia N-terminal del polipéptido de 82 kDa, mientras que la SEC NO: 5 se refiere a la secuencia N-terminal del polipéptido de 77 kDa. Además, la SEQ ID NO: 4 comprende los restos de aminoácidos 24-43 de la SEQ ID NO: 2, mientras que la SEC NO: 5 comprende los restos de aminoácidos 59-76 de la SEQ ID NO: 2.

La secuencia de aminoácidos definida por la SEQ ID NO: 6 se refiere al péptido escindido por CNBr de la  $\alpha$ -N-acetilglucosaminidasa humana purificada. Esta secuencia de aminoácidos se alinea con los restos de aminoácidos 540-554 del polipéptido de  $\alpha$ -N-acetilglucosaminidasa (SEQ ID NO: 2).

En el presente contexto, los “homólogos” de un polipéptido se refieren a aquellos polipéptidos, enzimas o proteínas que tienen una actividad enzimática similar a la de  $\alpha$ -N-acetilglucosaminidasa, independientemente de cualquier sustitución, adición o supresión de aminoácidos a los mismos. Un homólogo se puede aislar o derivar de la misma especie animal o de otra.

Además, los aminoácidos de un polipéptido homólogo se pueden reemplazar con otros aminoácidos que tengan propiedades similares, por ejemplo hidrofobia, hidrofilia, momento hidrofóbico, carga o antigenicidad, y demás.

“Análogos” abarca a los polipéptidos de  $\alpha$ -N-acetilglucosaminidasa y sus derivados peptídicos, sin tener en cuenta la presencia de aminoácidos no naturales análogos en los mismos.

El término “derivado” con respecto a los polipéptidos de la invención se refieren a mutantes, partes o fragmentos de una molécula funcional. Los derivados incluyen péptidos modificados en los que los ligandos se unen a uno o más de los restos de aminoácidos allí contenidos, tales como carbohidratos, enzimas, proteínas, polipéptidos o moléculas indicadoras tales como radionucleidos o compuestos fluorescentes. Las formas glicosiladas, fluorescentes, aciladas o alquiladas de los presentes péptidos están particularmente contempladas en la presente invención. Además, los derivados de un polipéptido pueden comprender fragmentos o partes de una secuencia de aminoácidos revelada en la presente memoria y están dentro del alcance de la invención, como son los homopolímeros o heteropolímeros que comprenden dos o más copias de los presentes polipéptidos. Los procedimientos para derivatizar péptidos son bien conocidos en la técnica.

Por consiguiente, la presente invención se refiere a cualquier molécula proteínica que comprende una secuencia de aminoácidos que corresponde a la enzima  $\alpha$ -N-acetilglucosaminidasa de mamífero de longitud completa o a una molécula semejante. La molécula semejante, en consecuencia, comprende partes, derivados y/o porciones de la enzima  $\alpha$ -N-acetilglucosaminidasa, sea funcional o no.

Preferentemente, el mamífero es un ser humano pero puede ser de origen no humano como se contempló con anterioridad.

La  $\alpha$ -N-acetilglucosaminidasa sintética o recombinante de la presente invención puede comprender una secuencia de aminoácidos que corresponde a la secuencia de aminoácidos natural o puede contener sustituciones, supresiones y/o adiciones de aminoácidos únicas o múltiples. La longitud de la secuencia de aminoácidos puede variar de unos pocos restos a una molécula de longitud completa.

Las sustituciones de aminoácidos normalmente son de restos únicos. Las inserciones de aminoácidos usualmente estarán en el orden de aproximadamente 1-10 restos de aminoácidos y las supresiones variarán de aproximadamente 1-20 restos. Preferentemente, las supresiones o inserciones se realizan en pares adyacentes, es decir una supresión de dos restos o inserción de dos restos.

Los derivados por inserción de aminoácidos de la  $\alpha$ -N-acetilglucosaminidasa de la presente invención incluyen fusiones amino- y/o carboxilo-terminales además de inserciones intra-secuencia de aminoácidos únicos o múltiples. Las variantes de secuencias de aminoácidos por inserción son aquellas en las que se introducen uno o más restos de aminoácidos en un sitio predeterminado de la proteína, si bien la inserción aleatoria también es posible con adecuada identificación del producto resultante. Las variantes por supresión se caracterizan por la eliminación de uno o más restos de aminoácidos de la secuencia. Las variantes por sustitución de aminoácidos son aquellas en las que se ha eliminado al menos un resto de la secuencia y se ha insertado un resto diferente en su lugar. Las sustituciones típicas son las que se realizan de acuerdo con la siguiente Tabla 2:

Las variantes de aminoácidos mencionadas anteriormente se pueden obtener fácilmente mediante técnicas de síntesis de péptidos bien conocidas en la técnica, tales como síntesis de péptidos en fase sólida (síntesis de Merrifield) y similares, o por manipulaciones de ADN recombinante. Las técnicas para obtener mutaciones por sustitución en sitios predeterminados del ADN que tienen una secuencia de ADN conocida o parcialmente conocida son bien conocidas e incluyen, por ejemplo, la mutagénesis M13. La manipulación de la secuencia del ADN para producir variantes de proteínas que se manifiestan como variantes por sustitución, inserción o supresión se describen en forma conveniente en otro lugar tal como Sambrook *et al.*, 1989 *Molecular Cloning: A Laboratory Manual* Cold Spring Harbor Laboratories, Cold Spring Harbor, NY.

Los derivados o moléculas semejantes incluyen sustituciones, supresiones y/o adiciones únicas o múltiples de cualesquiera componentes natural o artificialmente asociados con la enzima  $\alpha$ -N-acetilglucosaminidasa tales como un carbohidrato, lípido y/u otros restos proteínicos. Por ejemplo, la presente invención se extiende a las formas glicosiladas y no glicosiladas de la molécula. Todas estas moléculas están comprendidas en la expresión “mutantes”, “derivados”, “fragmentos”, “porciones” y moléculas “semejantes”. Estas moléculas pueden ser activas o no activas y pueden contener regiones específicas, tales como una región catalítica. En particular, las moléculas derivadas preferidas incluyen a aquellas con patrones de glicosilación alterados respecto de la molécula natural. Aún más particularmente, la molécula recombinante está más altamente glicosilada que la molécula natural. Tales derivados altamente glicosilados pueden presentar propiedades de captación mejoradas y vidas medias aumentadas.

Como se indica en los Ejemplos, el peso molecular de la  $\alpha$ -N-acetilglucosaminidasa purificada humana (es decir, 82kDa y 77kDa) y la  $\alpha$ -N-acetilglucosaminidasa recombinante de mamífero: producida en células CHO (es decir, 89 kDa y 79 kDa) son mayores que el peso molecular deducido del polipéptido de  $\alpha$ -N-acetilglucosaminidasa expuesto en la SEQ ID NO: 2 (es decir, 70 kDa), lo que sugiere que el polipéptido purificado y el recombinante se modifican en forma post-traduccional. Los datos presentados en el Ejemplo 8 indican además que la enzima  $\alpha$ -N-acetilglucosaminidasa recombinante producida en las células CHO, al menos, está glicosilada y que la diferencia de peso molecular determinada para los polipéptidos recombinantes y el polipéptido de la SEQ ID NO: 2 se debe casi totalmente a la glicosilación del polipéptido recombinante generado por las células CHO. Como se muestra en el Ejemplo 9, el polipéptido de  $\alpha$ -N-acetilglucosaminidasa recombinante glicosilado exhibe actividad enzimática.

La presente invención también se extiende a una  $\alpha$ -N-acetilglucosaminidasa sintética o moléculas semejantes cuando se fusionan con otras moléculas proteínicas. Estas últimas pueden incluir otra enzima, molécula indicadora, sitio de purificación o una secuencia de aminoácidos que facilite el transporte de la molécula fuera de una célula, tal como una secuencia señal.

La presente invención se extiende además a modificaciones post-traduccionales a la enzima  $\alpha$ -N-acetilglucosaminidasa. Las modificaciones se pueden realizar a la enzima natural o después de la síntesis por técnicas recombinantes. Las modificaciones pueden ser a nivel estructural o, por ejemplo, a nivel electroquímico, tal como la modificación de la carga neta o de la conformación estructural de la enzima. Tal modificación puede ser importante para facilitar la entrada o penetración de la enzima en tejidos seleccionados tales como cartílago o las barreras hematoencefálicas o para aumentar la vida media en circulación.

Los análogos de  $\alpha$ -N-acetilglucosaminidasa contemplados en la presente memoria incluyen, pero sin limitación, modificaciones de las cadenas laterales, incorporación de aminoácidos no naturales y/o sus derivados durante la síntesis de péptidos y el uso de agentes de entrecruzamiento y otros métodos que imponen restricciones conformacionales a la enzima.

Ejemplos de modificaciones de la cadena lateral contemplados por la presente invención incluyen modificaciones de grupos amino, tales como por alquilación reductiva por reacción con un aldehído seguida de reducción con  $\text{NaBH}_4$ ; amidación con metilacetimidato; acilación con anhídrido acético; carbamoilación de grupos amino con cianato; trinitrobenzoylación de grupos amino con ácido 2,4,6-trinitrobenzensulfónico (TNBS); acilación de grupos amino con anhídrido succínico y anhídrido tetrahydroftálico y piridoxilación de lisina con piridoxal-5'-fosfato seguida de reducción con  $\text{NaBH}_4$ .

El grupo guanidino de los restos de arginina se puede modificar por formación de productos de condensación heterocíclicos con reactivos tales como 2,3-butanodiona, fenilgloxal y glioxal.

El grupo carboxilo se puede modificar por activación con carbodiimida por formación de O-acilisourea seguida de ulterior derivatización, por ejemplo, a una amida correspondiente.

Los grupos sulfhidrilo se pueden modificar por métodos tales como carboximetilación con ácido yodoacético o yodoacetamida; oxidación con ácido perbromico a ácido cisteico; formación de disulfuros mixtos con otros compuestos con grupo tiol; reacción con maleimida, anhídrido maleico u otra maleimida sustituida; formación de derivados mercuriales usando 4-cloromercuribenzoato, ácido 4-cloromercurifenilsulfónico, cloruro de fenilmercurio, 2-cloromercurio-4-nitrofenol y otros mercuriales; carbamoilación con cianato a pH alcalino.

Los restos de triptofano se pueden modificar, por ejemplo, por oxidación con N-bromosuccinimida o alquilación del anillo indólico con bromuro de 2-hidroxi-5-nitrobenzilo o haluros de sulfenilo. Los restos de tirosina, por otra parte, se pueden modificar por nitración con tetranitrometano para formar un derivado de 3-nitrotirosina.

La modificación del anillo imidazólico de un resto de histidina se puede obtener por alquilación con derivados de ácido yodoacético o N-carboxilación con dietilpirocarbonato.

Los ejemplos de incorporación de aminoácidos no naturales y derivados durante la síntesis de péptidos incluyen, sin limitación, el uso de norleucina, ácido 4-aminobutírico, ácido 4-amino-3-hidroxi-5-fenilpentanoico, ácido 6-aminohexanoico, t-butilglicina, norvalina, fenilglicina, ornitina, sarcosina, ácido 4-amino-3-hidroxi-6-metilheptanoico, 2-tienilalanina y/o D-isómeros de aminoácidos. Los aminoácidos no naturales contemplados por la presente invención se incorporan en la presente memoria como Tabla 3.

## ES 2 319 828 T3

Se pueden usar agentes de entrecruzamiento, por ejemplo, para estabilizarlas conformaciones 3D, usando agentes de entrecruzamiento homo-bifuncionales tales como los imido-ésteres bifuncionales que tienen grupos espaciadores  $(CH_2)_n$  con  $n=1$  a  $n=6$ , glutaraldehído, ésteres de N-hidroxisuccinimida y reactivos hetero-bifuncionales que usualmente contienen un resto amino-reactivo tal como N-hidroxisuccinimida y otro resto reactivo con un grupo específico tal como el resto maleimido o ditio (SH) o carbodiimida (COOH). Además, la enzima se podría restringir desde un punto de vista conformacional, por ejemplo, por incorporación de  $C_\alpha$ - y  $N_\alpha$ -metilaminoácidos, introducción de dobles enlaces entre los átomos  $C_\alpha$  y  $C_\beta$  de los aminoácidos y la formación de péptidos cíclicos o análogos por introducción de enlaces covalentes tales como la formación un enlace amida entre los extremos N- y C-terminales entre dos cadenas laterales o entre una cadena lateral y el extremo N- o C-terminal.

Las modificaciones electroquímicas de la  $\alpha$ -N-acetilglucosaminidasa incluyen la interacción con polilisina o polietilenglicol u otro agente que ocasiona un cambio global en la carga neta de la enzima.

En forma ventajosa, la  $\alpha$ -N-acetilglucosaminidasa recombinante es una preparación biológicamente pura, lo que significa que ha experimentado alguna de purificación para eliminar otras proteínas y/o material no proteínico. La pureza de la preparación se puede representar como al menos 40% de la enzima, preferentemente al menos 60%, con mayor preferencia al menos 75%, aún con más preferencia al menos 85% y aún con mayor preferencia al menos 95% con respecto al material distinto de la  $\alpha$ -N-acetilglucosaminidasa determinado por peso, actividad, homología o similitud de aminoácidos, reactividad frente a anticuerpos u otros medios convenientes.

Se proveen métodos particularmente preferidos para la preparación y purificación de la  $\alpha$ -N-acetilglucosaminidasa recombinante en los Ejemplos 7 y 8.

Los expertos en la técnica conocerán los medios para purificar una  $\alpha$ -N-acetilglucosaminidasa sintética o recombinante proveniente de varias fuentes sin experimentación indebida y para expresar el grado de pureza de tal preparación purificada de la enzima.

La presente invención además contempla anticuerpos para  $\alpha$ -N-acetilglucosaminidasa y preferentemente  $\alpha$ -N-acetilglucosaminidasa sintética o molécula semejante. Los anticuerpos pueden ser policlonales o monoclonales, naturales o sintéticos (incluyendo formas recombinantes, fragmentarias o de fusión). Tales anticuerpos serán útiles para desarrollar inmunoensayos para  $\alpha$ -N-acetilglucosaminidasa y para identificar secuencias genéticas adicionales que son capaces de expresar polipéptidos de  $\alpha$ -N-acetilglucosaminidasa u homólogos, análogos o derivados de los mismos.

Los anticuerpos policlonales y monoclonales se pueden obtener por inmunización con un producto génico sintético o recombinante apropiado, o epítipo o fragmento peptídico de un producto génico, en particular un polipéptido de  $\alpha$ -N-acetilglucosaminidasa o un homólogo, análogo o derivado del mismo.

Alternativamente, se pueden usar fragmentos de anticuerpos, tales como fragmentos Fab. La presente invención se extiende además para incluir anticuerpos recombinantes y sintéticos y a híbridos de anticuerpos. Se considera que un "anticuerpo sintético" de la presente memoria incluye fragmentos e híbridos de anticuerpos.

La presente invención contempla un método de detección de mutaciones u otras aberraciones en el gen de  $\alpha$ -N-acetilglucosaminidasa en un paciente humano o animal. Tal método se puede llevar a cabo de varias maneras que incluyen el aislamiento de una fuente de ADN para ensayar o ARNm del mismo y la hibridación de éste a una molécula de ácido nucleico como se describió en la presente con anterioridad. Generalmente, el ácido nucleico es del tamaño de una sonda o cebador y la reacción en cadena de la polimerasa es un medio conveniente por el que se analiza el ARN o ADN. Otros ensayos adecuados incluyen la reacción en cadena de ligación y los métodos de amplificación de desplazamiento de hebra. La secuencia de la  $\alpha$ -N-acetilglucosaminidasa también se puede determinar y comparar con la secuencia natural. Tales métodos pueden ser útiles en adultos y niños y se pueden adaptar para un examen prenatal. El ADN a examinar incluye una muestra genómica que porta el gen de  $\alpha$ -N-acetilglucosaminidasa, un clon de ADNc y/o un producto de amplificación.

De acuerdo con la presente invención se provee un método de detección de aberraciones del gen de  $\alpha$ -N-acetilglucosaminidasa que incluye la ausencia de tal gen o una porción o una porción sustancial del mismo que comprende aislar una muestra de ADN o ARNm correspondiente a una región de dicho ADN y ponerla en contacto con una sonda de oligonucleótido capaz de hibridarse a una o más secuencias complementarias dentro del gen de  $\alpha$ -N-acetilglucosaminidasa y luego detectar la hibridación, el grado de hibridación o la ausencia de hibridación.

Alternativamente, la sonda es un cebador capaz de dirigir la amplificación de una o más regiones de dicho gen de  $\alpha$ -N-acetilglucosaminidasa y los productos de amplificación y/o el perfil de los productos de amplificación se compara con un individuo que porta el gen completo o con una base de datos de referencia.

En forma conveniente, los productos de amplificación se secuencian para determinar la presencia o ausencia del gen completo.

La presente invención se extiende al uso de algunos y todos los formatos de hibridación basados en ADN o basados en ácidos nucleicos y/o en la reacción en cadena de la polimerasa descritos en la presente memoria, para el diagnóstico de un trastorno que involucra el gen de  $\alpha$ -N-acetilglucosaminidasa en un paciente humano o animal.

La presente invención además se extiende a un método de tratamiento de pacientes que padecen una deficiencia de  $\alpha$ -N-acetilglucosaminidasa, tal como en MPS-IIIIB, comprendiendo dicho método administrar a dicho paciente una cantidad efectiva de  $\alpha$ -N-acetilglucosaminidasa o una forma activa semejante de la misma.

5 Preferentemente, la  $\alpha$ -N-acetilglucosaminidasa está en forma recombinante. Dicho método se denomina “terapia enzimática”. Alternativamente, se puede emplear la terapia génica, que incluye introducir un gen activo (es decir, una molécula de ácido nucleico como se describió en la presente memoria con anterioridad) o a partes del gen u otras secuencias que facilitan la expresión de un gen de  $\alpha$ -N-acetilglucosaminidasa natural.

10 La administración de  $\alpha$ -N-acetilglucosaminidasa para la terapia enzimática puede ser por administración oral, intravenosa, supositorio, intraperitoneal, intramuscular, intranasal, intradérmica o subcutánea o por infusión o implante. La  $\alpha$ -N-acetilglucosaminidasa preferentemente es como se describió en la presente memoria con anterioridad, incluyendo mutantes activas o sus derivados y sus variantes de glicosilación. La administración también puede ser por medio de una terapia génica que incluye la expresión del gen por inclusión del gen en vectores virales que se introducen en el huésped animal (por ej., ser humano) a tratar. Alternativamente, el gen se puede expresar en un hospedante bacteriano que luego se introduce y se convierte en parte de la flora bacteriana del animal a examinar.

15 La presente invención también se refiere a una composición farmacéutica que comprende  $\alpha$ -N-acetilglucosaminidasa sintética (por ej., recombinante) o molécula semejante, que incluye los derivados activos y sus fragmentos, solos o en combinación con otras moléculas activas. Tales otras moléculas pueden actuar en forma sinérgica con la enzima o facilitan su entrada a una célula diana. La composición también contendrá uno o más vehículos y/o diluyentes farmacéuticamente aceptables. En forma alternativa, la composición puede comprender un componente genético útil para la terapia génica.

25 Se contempla que los componentes útiles activos de la composición farmacéutica que comprenden la  $\alpha$ -N-acetilglucosaminidasa sintética o recombinante o sus mutantes o fragmentos exhiben excelente actividad para tratar pacientes con una deficiencia de la enzima cuando se administran en una cantidad que depende del caso particular. La variación depende, por ejemplo, del paciente y la  $\alpha$ -N-acetilglucosaminidasa utilizada. Por ejemplo, se puede administrar de aproximadamente 0,5  $\mu$ g a aproximadamente 20 mg de enzima por cuerpo del animal o, dependiendo del animal y otros factores, por kilogramo de peso corporal. El régimen de dosis se puede ajustar para proporcionar la respuesta terapéutica óptima. Por ejemplo, se pueden administrar varias dosis divididas en forma diaria, semanal, mensual o en otros intervalos de tiempo adecuados o la dosis se puede reducir proporcionalmente según lo indiquen las exigencias de la situación. Por consiguiente, se pueden administrar dosis alternativas en el orden de 1,0  $\mu$ g a 15 mg, 2,0  $\mu$ g a 10 mg o 10  $\mu$ g a 5 mg en una dosis única o como parte de dosis múltiples. El compuesto activo se puede administrar de manera conveniente tal como por las vías oral, intravenosa (cuando es hidrosoluble), intramuscular, subcutánea, intranasal, intradérmica o supositorio o implante (por ej., usando moléculas de liberación lenta). De acuerdo con la vía de administración, se puede requerir que los componentes activos que comprenden una  $\alpha$ -N-acetilglucosaminidasa sintética (por ej., recombinante) o sus fragmentos, derivados o mutantes estén recubiertos con un material para protegerlos de la acción de enzimas, ácidos y otras condiciones naturales que pueden inactivar dichos componentes. Por ejemplo, la baja lipofilicidad de la  $\alpha$ -N-acetilglucosaminidasa permitirá que ésta sea destruida en el aparato gastrointestinal por las enzimas capaces de escindir los enlaces peptídicos y en el estómago por hidrólisis ácida. Con el fin de administrar la vacuna por una administración diferente de la parenteral, la enzima se recubrirá o administrará con un material para evitar su inactivación. Por ejemplo, la enzima se puede administrar en un adyuvante, coadministrar con inhibidores enzimáticos o en liposomas. El adyuvante se usa en su sentido más amplio e incluye cualquier compuesto inmunoestimulante tal como interferón. Los adyuvantes contemplados en la presente memoria incluyen resorcinoles, tensioactivos no iónicos tales como polioxietileno-olefín éter y n-hexadecil-polietileno éter. En forma conveniente, el adyuvante es adyuvante de Freund Completo o Incompleto. Los inhibidores enzimáticos incluyen al inhibidor de tripsina pancreática, fluorofosfato de diisopropilo (DEP) y trasilol. Los liposomas incluyen emulsiones agua-en-aceite-en-agua además de liposomas convencionales.

50 El compuesto activo también se puede administrar en dispersiones preparadas en glicerol, polietilenglicoles líquidos y/o sus mezclas y en aceites. En condiciones ordinarias de almacenamiento y uso, estas preparaciones contienen un conservante para evitar el crecimiento de microorganismos.

55 Las formas farmacéuticas adecuadas para uso inyectable incluyen soluciones (cuando son hidrosolubles) o dispersiones acuosas estériles y polvos estériles para la preparación extemporánea de soluciones o dispersiones inyectables estériles. En todos los casos la forma debe ser estéril y debe ser fluida en la medida que permita que sea inyectada fácilmente. Debe ser estable en las condiciones de manufactura y almacenamiento y se debe preservar contra la acción contaminante de microorganismos tales como bacterias y hongos. El vehículo puede ser un solvente o medio de dispersión que contenga, por ejemplo, agua, etanol, poliol (por ejemplo, glicerol, propilenglicol y polietilenglicol líquido y similares), mezclas adecuadas de estos y aceites vegetales. Se puede mantener la fluidez apropiada, por ejemplo, por el uso de un recubrimiento tal como lecitina, por mantenimiento del tamaño de partícula requerido en el caso de una dispersión y por el uso de tensioactivos. La prevención de la acción de los microorganismos se puede lograr con diversos agentes antibacterianos y antifúngicos, por ejemplo parabenos, clorobutanol, fenol, ácido sórbico, timerosal y similares. En muchos casos, será preferible incluir agentes isotónicos, por ejemplo azúcares o cloruro de sodio. La absorción prolongada de las composiciones inyectables se puede lograr con el uso en las composiciones de agentes que retardan la absorción, por ejemplo monoestearato de aluminio y gelatina.

Las soluciones inyectables estériles se preparan por incorporación del compuesto activo en la cantidad requerida del solvente apropiado con varios de los otros componentes enumerados anteriormente, según corresponda, seguida de esterilización por filtración. En general, las dispersiones se preparan por incorporación de varios componentes activos esterilizados en un vehículo estéril que contiene el medio de dispersión básico y los otros componentes necesarios entre los enumerados anteriormente. En el caso de los polvos estériles para la preparación de soluciones inyectables estériles, los métodos preferidos de preparación son secado al vacío y la técnica de liofilización que produce un polvo del componente activo más cualquier componente deseado adicional a partir de la solución previamente esterilizada por filtración del mismo. Cuando la  $\alpha$ -N-acetilglucosaminidasa de la presente invención se protege en forma adecuada como se describió anteriormente, la composición se puede administrar por vía oral, por ejemplo, con un diluyente inerte o con un vehículo comestible asimilable, o se puede encerrar en una cápsula de gelatina dura o blanda, o se puede compactar para dar comprimidos o se puede incorporar directamente con el alimento de la dieta. Para la administración terapéutica oral, el compuesto activo se puede incorporar con excipientes y usar en forma de comprimidos ingeribles, comprimidos bucales, pastillas de disolución oral, cápsulas, elixires, suspensiones, jarabes, obleas y similares. Tales composiciones y preparaciones deben contener al menos 1% en peso del compuesto activo. El porcentaje de las composiciones y preparaciones, obviamente, se pueden variar y en forma conveniente puede estar entre aproximadamente 5 a aproximadamente 80% del peso de la unidad. La proporción de compuesto activo en las composiciones de vacuna es tal que se obtenga una dosis adecuada. Las composiciones o preparaciones preferidas de acuerdo con la presente invención se preparan, de modo que una forma unitaria de dosificación oral contenga entre aproximadamente 0,5  $\mu$ g y 20 mg de compuesto activo.

Los comprimidos, pastillas de disolución oral, píldoras, cápsulas y similares también pueden contener los siguientes: un aglutinante tal como goma de tragacanto, acacia, almidón de maíz o gelatina; excipientes tales como fosfato dicálcico; un agente desintegrante tal como almidón de maíz, almidón de patata, ácido algínico y similares; un lubricante tal como estearato de magnesio; y se puede agregar un agente edulcorante tal como sacarosa, lactosa o sacarina o un agente saborizante tal como menta, aceite de gaulteria, o saborizante de cerezas. Cuando la forma unitaria de dosificación es una cápsula, puede contener, además de los materiales del tipo anterior, un vehículo líquido. Otros materiales diversos pueden estar presentes como recubrimientos o modificar de otro modo la forma física de la unidad de dosificación. Por ejemplo, los comprimidos, píldoras o cápsulas se pueden recubrir con goma laca, azúcar o ambos. Un jarabe o elixir puede contener el compuesto activo, sacarosa como agente edulcorante, metil y propilparabenos como conservantes, un colorante y saborizante tal como sabor a cerezas o naranja. Obviamente, cualquier material usado en la preparación de cualquier unidad de dosificación debe ser farmacéuticamente puro y sustancialmente no tóxico en las proporciones empleadas. Además, el compuesto activo se puede incorporar en las preparaciones y formulaciones de liberación sostenida.

Como se usa en la presente memoria "vehículos y/o diluyentes farmacéuticamente aceptables" incluyen cualquiera y todos los solventes, medios de dispersión, soluciones acuosas, recubrimientos, agentes antibacterianos y antifúngicos, agentes isotónicos y retardadores de la absorción y similares. El uso de tales medios y agentes para las sustancias activas farmacéuticas es bien conocido en la técnica. Excepto en la medida que cualquier medio o agente convencional sea incompatible con el componente activo, se contempla su uso en las composiciones farmacéuticas. También se pueden incorporar componentes activos suplementarios en las composiciones.

La presente invención además se refiere al uso de  $\alpha$ -N-acetilglucosaminidasa o fragmento activo, mutante o derivado de la misma en la fabricación de un medicamento para el tratamiento de pacientes que padecen una deficiencia de la enzima natural (por ej., MPS-IIIB).

La presente invención se describe además con referencia a los siguientes Ejemplos no limitantes.

#### Ejemplo 1

##### *Purificación de $\alpha$ -N-acetilglucosaminidasa*

Se purificó  $\alpha$ -N-acetilglucosaminidasa de acuerdo con el método descrito en Weber *et al.* (1996). La enzima se purificó hasta homogeneidad a partir de placenta humana. La evidencia de pureza se muestra después de una SDS/PAGE que se representa en la Figura 1. Todas las muestras se redujeron con ditiotreitól antes de la electroforesis.

#### Ejemplo 2

##### *Caracterización de $\alpha$ -N-acetilglucosaminidasa*

Los resultados presentados en la Figura 1 muestran dos polipéptidos de aproximadamente 82 kDa y 77 kDa de peso molecular, que corresponden a polipéptidos de  $\alpha$ -N-acetilglucosaminidasa purificados de placenta humana de acuerdo con el Ejemplo 1.

## ES 2 319 828 T3

### Ejemplo 3

#### *Determinación de la Secuencia de Aminoácidos*

5 Se determinaron las secuencias de aminoácidos N-terminales de los polipéptidos de  $\alpha$ -N-acetilglucosaminidasa de 77 kDa y 82 kDa, además de la secuencia de aminoácidos de un producto de escisión interna con CNBr de estos péptidos, utilizando los métodos de Weber *et al.* (1996).

10 Las secuencias de aminoácidos se muestran en la Tabla 4.

TABLA 4

15 *Secuencias de aminoácidos N-terminales (SEQ ID NO: 4 y SEQ ID NO: 5) y secuencia del péptido obtenido utilizando CNBr (SEQ ID NO: 6), determinadas a partir de  $\alpha$ -N-acetilglucosaminidasa purificada humana*

20 polipéptido de 82 kDa	DEAREAAVRLVARLLGPG
polipéptido de 77 kDa	KPGLDTYSLGGGGAAX <sup>1</sup> VR
péptido obtenido utilizando CNBr	WRLLLTSAPSLX <sup>1</sup> TX <sup>1</sup> P

25 X<sup>1</sup>: no se pudo identificar ningún resto para esta posición, lo que indica que este resto podría estar fosforilado o glicosilado.

### 30 Ejemplo 4

#### *Clonación de ADNc de $\alpha$ -N-Acetilglucosaminidasa*

35 Se prepararon sondas de oligonucleótidos basadas en las secuencias de aminoácidos parciales obtenidas para los polipéptidos de  $\alpha$ -N-acetilglucosaminidasa purificados (Ejemplo 3). Las sondas se usaron posteriormente para seleccionar una genoteca de ADNc en leucocitos de sangre periférica humana. Se aisló un clon de ADNc de aproximadamente 2,6 kpb que codifica la mayor parte de la secuencia de  $\alpha$ -N-acetilglucosaminidasa humana (SEQ ID NO: 1).

40 La secuencia codificadora de  $\alpha$ -N-acetilglucosaminidasa restante se obtuvo de la secuencia de nucleótidos del gen genómico correspondiente (SEQ ID NO: 3), aislado por hibridación con una genoteca del cromosoma 17 humano (Weber *et al.* 1996). El marco de lectura abierto completo tiene 2232 nucleótidos de largo y codifica una proteína de 743 aminoácidos (más el codón de detención). La masa molecular predicha de la proteína madura más larga (menos los 23 aminoácidos N-terminales del péptido señal) es de aproximadamente 79.622 dalton.

45 La secuencia de aminoácidos de una  $\alpha$ -N-acetilglucosaminidasa se muestra en la SEQ ID NO: 2. El peso molecular deducido de la secuencia de aminoácidos deseada de una  $\alpha$ -N-acetilglucosaminidasa es aproximadamente 70 kDa. El sitio probable de escisión por peptidasa del péptido señal está entre los aminoácidos 23 y 24. Existen siete sitios de N-glicosilación potenciales en la secuencia.

50 La secuencia de nucleótidos del gen genómico correspondiente de  $\alpha$ -N-acetilglucosaminidasa (SEQ ID NO: 3) comprende 10380 pb que incluyen 889 pb de la secuencia 5' corriente arriba que corresponde al menos a una parte de la secuencia promotora de la  $\alpha$ -N-acetilglucosaminidasa, además de las secuencias de nucleótidos de los intrones I, II, II, IV, V, además de 1326 pb de la secuencia 3' no traducida.

### Ejemplo 5

#### *Construcción de un vector de expresión que comprende la secuencia de ADNc de $\alpha$ -N-acetilglucosaminidasa*

60 El inserto de ADNc del clon  $\lambda$  pbl 33, que contiene las bases 107 a 2575 del ADNc de  $\alpha$ -N-acetilglucosaminidasa se escindió con *EcoRI* y se subclonó en pBluescript II SK-(Stratagene). Un fragmento *XmaI* de 178 pb (bases 1 a 178 del ADNc de  $\alpha$ -N-acetilglucosaminidasa) proveniente del subclon cósmido 6.3, que contiene el codón de iniciación, se clonó en el subclon pBluescript para producir una secuencia de ADNc de longitud completa además de 101 pb de la secuencia 5' no traducida como así también 245 pb de la región 3' no traducida, incluyendo el sitio de poliadenilación, la cola poliA y el ADN conector. El ADNc de longitud completa se clonó en forma direccional en el vector de expresión pCDNA3 (Invitrogen) por medio de los sitios *EcoRI* y *BamHI*.

## ES 2 319 828 T3

### Ejemplo 6

#### *Expresión de $\alpha$ -N-acetilglucosaminidasa recombinante*

5 Se transfectaron células de ovario de hámster chino (CHO) con un vector de expresión usando el reactivo de transfección DOTAP (Boehringer Mannheim) de acuerdo con las instrucciones del fabricante. Las células se cultivaron en medio de Ham F12, 10% (v/v) de suero fetal de ternero, penicilina y sulfato de estreptomicina a razón de 100  $\mu$ g/ml cada uno. Las células se cultivaron en medio no selectivo durante 48 h y luego se incubaron en un medio que contenía 750  $\mu$ g/ml de G418 sulfato (Geneticina) hasta que aparecieron las colonias resistentes.

10

Los clones de células únicas se cultivaron y 26 de ellos se analizaron para determinar la expresión de  $\alpha$ -N-acetilglucosaminidasa recombinante con un sustrato fluorogénico de  $\alpha$ -N-acetilglucosaminidasa (es decir, N-acetilglucosamina conectada en  $\alpha$  a 4-metilumbeliferona).

15

### Ejemplo 7

#### *Producción de $\alpha$ -N-acetilglucosaminidasa en gran escala*

20 Se embebieron 2 g de perlas del microportador Cytodex en 250 ml de PBS durante 3 h a 37°C con tres cambios de PBS y luego se autoclavaron durante 15 minutos a 120°C (ciclo húmedo). Las perlas luego se lavaron con medio de crecimiento estéril (Coons/DMEM, 10% v/v de suero fetal de ternero, penicilina y sulfato de estreptomicina a razón de 100  $\mu$ g/ml cada uno y 0,1% p/v de Pluronic F68) y se transfirieron a un matraz de cultivo con agitador Techne. Las perlas del microportador se inocularon con siete matraces confluentes 175 del clon celular que mostraba la expresión más alta de  $\alpha$ -N-acetilglucosaminidasa recombinante. Se agregó medio de crecimiento hasta 200 ml y el cultivo se incubó con una velocidad de agitación de 20 rpm para obtener una distribución uniforme de las células sobre las perlas. Las células se dejaron fijar a las perlas durante 16 h a baja velocidad y luego se agregó medio hasta 500 ml y la velocidad de agitación aumentó a 30 rpm. Después de una fase de crecimiento de 48 a 72 h con aireación diaria para permitir el intercambio gaseoso, las perlas estaban completamente cubiertas de células y el medio se cambió por medio de producción (Coons/DMEM, sin suero fetal de ternero, penicilina y sulfato de estreptomicina a razón de 100  $\mu$ g/ml cada uno, 0,1% p/v de Pluronic F68 y  $\text{NH}_4\text{Cl}$  5 mM). La concentración de glucosa se controló diariamente y se reemplazó el medio, cuando la glucosa descendió por debajo de 5 mM cada 2 o 3 días. El medio recolectado contenía aproximadamente 2 mg de proteína de  $\alpha$ -N-acetilglucosaminidasa por  $\text{dm}^3$  de medio de producción.

35

### Ejemplo 8

#### *Purificación de $\alpha$ -N-acetilglucosaminidasa recombinante*

40 El medio de producción se dializó contra NaAc 50 mM de pH 5,5 y se cargó en una columna de heparina-agarosa equilibrada en el mismo tampón. Después de lavar con tampón de NaAc y NaAc/NaCl 50 mM, la columna se eluyó con NaCl 75 mM en tampón de NaAc. El eluato se dializó contra Tris/HCl 20 mM de pH 7,5, se cargó en una columna DEAE Sephacel, se lavó con NaCl 25 mM en Tris/HCl 20 mM y luego se eluyó con NaCl 50 y 75 mM en mM/HCl Tris 20, respectivamente.

45

La SDS-PAGE de los dos eluatos mostró dos bandas del polipéptido asociadas con la actividad enzimática con pesos moleculares aparentes de 79 y 89 kDa. La  $\alpha$ -N-acetilglucosaminidasa más pequeña eluyó predominantemente en la fracción de NaCl 50 mM mientras que el polipéptido de 89 kDa de  $\alpha$ -N-acetilglucosaminidasa se enriqueció en la fracción de NaCl 75 mM (Fig. 2).

50

La diferencia entre el peso molecular aparente de los polipéptidos de la  $\alpha$ -N-acetilglucosaminidasa recombinante se debe a la presencia de cadenas laterales de carbohidratos adicionales, debido a que un digesto con PNGasa F, que escinde los grupos de N-glicosilación, redujo ambos polipéptidos de 79 kDa y 89 kDa a la banda del polipéptido que tiene un peso molecular aparente de aproximadamente 70 kDa (Fig. 2), que corresponde al peso molecular aproximado deducido de los datos de la secuencia de aminoácidos primaria (SEQ ID NO: 2).

55

### Ejemplo 9

#### *Características de la $\alpha$ -N-acetilglucosaminidasa recombinante*

No se observaron diferencias entre las actividades enzimáticas de los polipéptidos de 79 y 89 kDa de  $\alpha$ -N-acetilglucosaminidasa recombinante producidos en las células CHO de acuerdo con los Ejemplos 7 y 8. Con el sustrato fluorogénico N-acetilglucosamina conectada en  $\alpha$  a 4-metilumbeliferona (4-MU), la enzima tiene un pH óptimo de 4,6 con una  $k_M$  de 5,34 mM y una  $V_{\text{máx}}$  de  $3,97 \times 10^6$  pmol/min/mg. Para un sustrato disacárido marcado con  $^3\text{H}$ , tendría un pH óptimo de 4,1 con una  $k_M$  de 0,0166 mM y una  $V_{\text{máx}}$  de  $4,48 \times 10^4$  pmol/min/mg.

65

## Ejemplo 10

*Análisis de las mutaciones de los pacientes de Sanfilippo B*

5 Se aísla ADN genómico a partir de fibroblastos de piel cultivados de pacientes por extracción con fenol/cloroformo y se utiliza para amplificar los ocho exones y las secuencias intrónicas adyacentes en forma individual por PCR.

10 Las secuencias cebadoras usadas en la reacción de amplificación se determinan fácilmente a partir de las secuencias de nucleótidos del ADNc de  $\alpha$ -N-acetilglucosaminidasa y los clones genómicos expuestos en la SEC. ID No: 1 o SEQ ID NO: 3. Las condiciones de amplificación también se determinan fácilmente sin experimentación indebida. Los procedimientos para el diseño de los cebadores de PCR y las condiciones de amplificación se describen en detalle, por ejemplo, en McPherson *et al.* (1991). Las diferencias en la secuencia primaria se pueden identificar separando los productos de PCR en un gel de poliacrilamida en condiciones no desnaturalizantes (geles de SSCP). Los cambios, inserciones y supresiones de bases llevarán a un patrón diferente de banda, comparado con el de tipo salvaje en la mayoría de los casos, el que se puede visualizar tanto por autorradiografía del gel después de la marcación del ADN durante la PCR como por tinción del ADN no marcado en el gel con plata. Los productos de PCR que muestran un patrón de banda diferente son secuencias para identificar el cambio. Otras muestras de pacientes se pueden analizar para determinar mutaciones y polimorfismos que se hallaron por hibridación con oligonucleótidos de tipo salvaje y específicos de la mutación (ASO).

**Referencias**

25 1. Ausubel, F. M., Brent, R., Kingston, RE, Moore, D.D., Seidman, J.G., Smith, IA., y 5 Struhl, K. (1987). En: Current Protocols en Molecular Biology. Wiley Interscience (ISBN 047150338).

2. Hopwood JJ (1989) En: “Heparin: Chemical and Biological Properties, Clinical Applications” (Lane DW y Lindahl U, eds.), 190-229, Edward Arnold, London.

30 3. McKusick V y Neufeld E (1983) En: “The Metabolic Basis of Inherited Disease”. (Stanbury JB, Wyngaarden JB, Fredrickson DS, Goldstein JL y Brown MS, eds), 5th Ed., 751-771, McGraw-Hill, New York.

4. McPherson, MI, Quirke, P. y Taylor, G.R, (1991) En: PCR A Practical Approach. Oxford University Press, Oxford. (ISBN 0-19-96322L-X).

35 5. O’Brien Js, (1972) *Proc. Natl. Acad. Sci. USA* 69: 1720-1722.

40 6. Sambrook, J., Fritsch, E. y Maniatis, T. (1989) En: “Molecular Cloning” a laboratory manual, Cold Spring Harbour.

7. Von Figura, K y Kresse H, (1972) *Biochem Biophys. Res. Commun*, 48: 262-269.

45 8. Weber B, Scott H, Blanch L, Clements P, Morris CP, Anson D, Hopwood J, (1996) *Nature Genetics* (presentado).

50

55

60

65

REIVINDICACIONES

5 1. Un método para la producción de  $\alpha$ -N-acetilglucosaminidasa o un fragmento o derivado de la misma en donde la  $\alpha$ -N-acetilglucosaminidasa o fragmento o derivado de la misma hidroliza restos de  $\alpha$ -N-acetilglucosamina del extremo terminal no reductor del heparán sulfato, que comprende:

10 (i) transfectar células de mamífero con un vector de expresión que comprende una molécula de ADN que codifica  $\alpha$ -N-acetilglucosaminidasa o un fragmento o derivado de la misma al menos 40% idéntica a la SEQ ID NO:1 o SEQ ID NO:3;

(ii) cultivar las células transfectadas, por lo que se produce la proteína deseada y se secreta al medio; y

15 (iii) recuperar la proteína deseada del medio.

2. El método de acuerdo con la reivindicación 1, en donde la molécula de ADN que codifica  $\alpha$ -N-acetilglucosaminidasa o un fragmento o derivado de la misma es el ADNc que tiene la secuencia SEQ ID NO:1.

20 3. El método de acuerdo con la reivindicación 1, en donde la proteína recuperada del medio comprende los aminoácidos de la SEQ ID NO:2.

4. El método de acuerdo con la reivindicación 3, en donde la proteína recuperada del medio comprende los aminoácidos 24-743 de la SEQ ID NO:2.

25 5. El método de acuerdo con la reivindicación 1, en donde las células de mamífero son células CHO.

6. El método de acuerdo con la reivindicación 1, en donde la proteína recuperada del medio está en forma glicosilada.

30 7. El método de acuerdo con la reivindicación 6, en donde el peso molecular de la forma glicosilada de la proteína recuperada del medio determinado mediante SDS/PAGE es al menos aproximadamente 79 kDa.

35

40

45

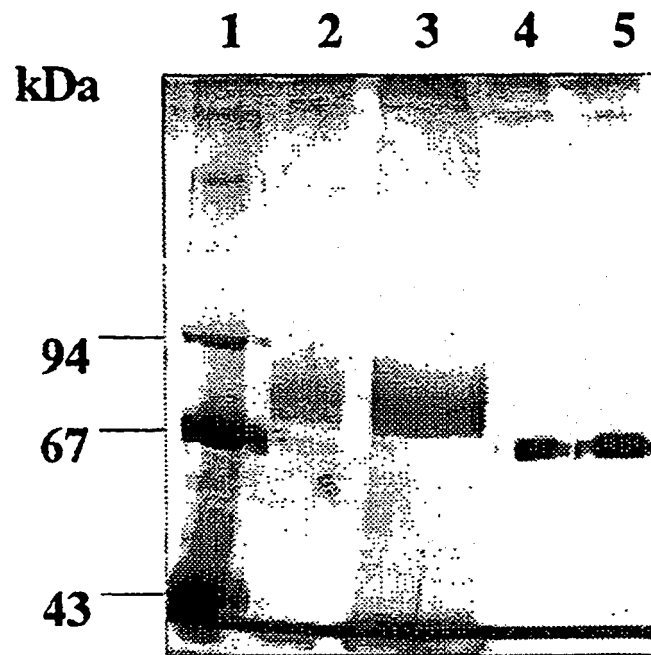
50

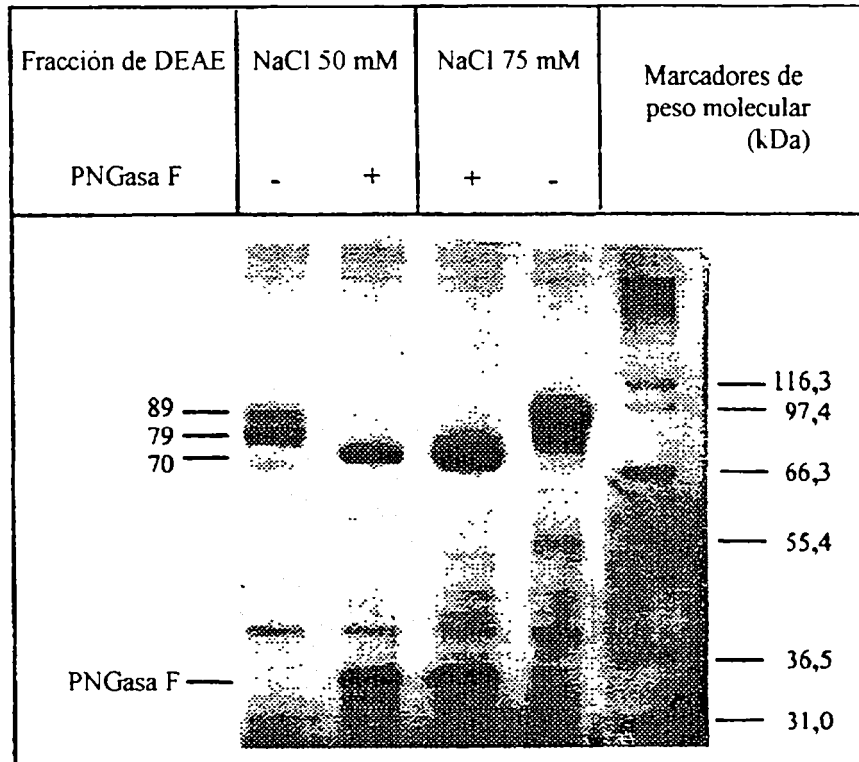
55

60

65

**Fig. 1**





**FIGURA 2**

# ES 2 319 828 T3

## LISTA DE SECUENCIAS

### (1) INFORMACIÓN GENERAL:

- 5 (i) SOLICITANTE: WOMEN'S AND CHILDREN'S HOSPITAL  
(SÓLO EE.UU.: HOPWOOD JJ, WEBER B SCOTT HS, BLANCH LC, ANSON D)
- (ii) TÍTULO DE LA INVENCION:  $\alpha$ -N-ACETILGLUCOSAMINIDASA SINTÉTICA DE MAMÍFERO Y  
10 SECUENCIAS GENÉTICAS QUE LA CODIFICAN
- (iii) NÚMERO DE SECUENCIAS: 6
- (iv) DIRECCIÓN PARA LA CORRESPONDENCIA:
- 15 (A) DESTINATARIO: DAVIES COLLISON CAVE  
(B) CALLE: 1 LITTLE COLLINS STREET  
(C) CIUDAD: MELBOURNE  
(D) ESTADO: VICTORIA  
20 (E) PAÍS: AUSTRALIA  
(F) CÓDIGO: 3000
- (v) FORMA LEGIBLE POR ORDENADOR:
- 25 (A) TIPO DE MEDIO: disquete  
(B) ORDENADOR: PC IBM compatible  
(C) SISTEMA OPERATIVO: PC-DOS/MS-DOS  
30 (D) PROGRAMA: Patentln, Release #1.0; Version #1.25
- (vi) DATOS DE LA SOLICITUD ANTERIOR:
- (A) NÚMERO DE SOLICITUD: AU PN6748/95  
35 (B) FECHA DE PRESENTACIÓN: 23-NOV-1995
- (vi) DATOS DE LA SOLICITUD ACTUAL:
- (A) NÚMERO DE SOLICITUD: PCT INTERNATIONAL  
40 (B) FECHA DE PRESENTACIÓN: 22-NOV-1996
- (viii) INFORMACIÓN SOBRE EL ABOGADO/AGENTE:
- (A) NOMBRE: HUGHES, DR. E JOHN L  
(C) REFERENCIA/NÚMERO DE EXPEDIENTE: MRO/EJH/JMC
- 45 (ix) INFORMACIÓN PARA TELECOMUNICACIONES:
- (A) TELÉFONO: +61 3 9254 2777  
(B) TELEFAX: +61 3 9254 2770

### 50 (2) INFORMACIÓN PARA LA SEQ ID NO:1:

- (i) CARACTERÍSTICAS DE LA SECUENCIA:
- 55 (A) LONGITUD: 2575 pares de bases  
(B) TIPO: ácido nucleico  
(C) CATENARIEDAD: sencilla  
(D) TOPOLOGÍA: lineal
- 60 (ii) TIPO DE MOLÉCULA: ADNc
- (vi) FUENTE ORIGINAL:
- (A) ORGANISMO: *Homo sapiens*  
65 (F) TIPO DE TEJIDO: Sangre periférica  
(G) TIPO DE CÉLULA: Leucocito

## ES 2 319 828 T3

(ix) CARACTERÍSTICAS:

(A) NOMBRE/CLAVE: CDS

(B) LOCALIZACIÓN: 102..2333

5

(xi) DESCRIPCIÓN DE LA SECUENCIA: SEQ ID NO:1:

10	CCGGGGCTTA GCCTTCGGGT CCACGTGGCC GGAGGCCGGC AGCTGATTGG ACGCGGGCCG	60
15	CCCCACCCCC TGGCCGTCGC GGGACCCGCA GGACTGAGAC C ATG GAG GCG GTG <span style="padding-left: 100px;">Met Glu Ala Val</span>	113
	<span style="padding-left: 100px;">1</span>	
20	GCG GTG GCC GCG GCG GTG GGG GTC CTT CTC CTG GCC GGG GCC GGG GGC Ala Val Ala Ala Ala Val Gly Val Leu Leu Leu Ala Gly Ala Gly Gly	161
	<span style="padding-left: 10px;">5</span> <span style="padding-left: 40px;">10</span> <span style="padding-left: 40px;">15</span> <span style="padding-left: 40px;">20</span>	
25	GCG GCA GGC GAC GAG GCC CGG GAG GCG GCG GCC GTG CGG GCG CTC GTG Ala Ala Gly Asp Glu Ala Arg Glu Ala Ala Ala Val Arg Ala Leu Val	209
	<span style="padding-left: 100px;">25</span> <span style="padding-left: 40px;">30</span> <span style="padding-left: 40px;">35</span>	
30	GCC CGG CTG CTG GGG CCA GGC CCC GCG GCC GAC TTC TCC GTG TCG GTG Ala Arg Leu Leu Gly Pro Gly Pro Ala Ala Asp Phe Ser Val Ser Val	257
	<span style="padding-left: 100px;">40</span> <span style="padding-left: 40px;">45</span> <span style="padding-left: 40px;">50</span>	
35	GAG CGC GCT CTG GCT GCC AAG CCG GGC TTG GAC ACC TAC AGC CTG GGC Glu Arg Ala Leu Ala Ala Lys Pro Gly Leu Asp Thr Tyr Ser Leu Gly	305
	<span style="padding-left: 10px;">55</span> <span style="padding-left: 40px;">60</span> <span style="padding-left: 40px;">65</span>	
40	GGC GGC GGC GCG GCG CGC GTG CGG GTG GCG GGC TCC ACG GGC GTG GCG Gly Gly Gly Ala Ala Arg Val Arg Val Arg Gly Ser Thr Gly Val Ala	353
	<span style="padding-left: 10px;">70</span> <span style="padding-left: 40px;">75</span> <span style="padding-left: 40px;">80</span>	
45	GCC GCC GCG GGG CTG CAC CGC TAC CTG GCG GAC TTC TGT GGC TGC CAC Ala Ala Ala Gly Leu His Arg Tyr Leu Arg Asp Phe Cys Gly Cys His	401
	<span style="padding-left: 10px;">85</span> <span style="padding-left: 40px;">90</span> <span style="padding-left: 40px;">95</span> <span style="padding-left: 40px;">100</span>	

45

50

55

60

65









# ES 2 319 828 T3

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

# ES 2 319 828 T3

	Val	Leu	Pro	Ala	Phe	Ala	Gly	His	Val	Pro	Glu	Ala	Val	Thr	Arg	Val	
					245					250					255		
5	Phe	Pro	Gln	Val	Asn	Val	Thr	Lys	Met	Gly	Ser	Trp	Gly	His	Phe	Asn	
				260					265					270			
10	Cys	Ser	Tyr	Ser	Cys	Ser	Phe	Leu	Leu	Ala	Pro	Glu	Asp	Pro	Ile	Phe	
			275					280					285				
15	Pro	Ile	Ile	Gly	Ser	Leu	Phe	Leu	Arg	Glu	Leu	Ile	Lys	Glu	Phe	Gly	
		290					295					300					
20	Thr	Asp	His	Ile	Tyr	Gly	Ala	Asp	Thr	Phe	Asn	Glu	Met	Gln	Pro	Pro	
	305					310					315				320		
25	Ser	Ser	Glu	Pro	Ser	Tyr	Leu	Ala	Ala	Ala	Thr	Thr	Ala	Val	Tyr	Glu	
					325					330					335		
30	Ala	Met	Thr	Ala	Val	Asp	Thr	Glu	Ala	Val	Trp	Leu	Leu	Gln	Gly	Trp	
				340					345					350			
35	Leu	Phe	Gln	His	Gln	Pro	Gln	Phe	Trp	Gly	Pro	Ala	Gln	Ile	Arg	Ala	
			355					360					365				
40	Val	Leu	Gly	Ala	Val	Pro	Arg	Gly	Arg	Leu	Leu	Val	Leu	Asp	Leu	Phe	
		370					375					380					
45	Ala	Glu	Ser	Gln	Pro	Val	Tyr	Thr	Arg	Thr	Ala	Ser	Phe	Gln	Gly	Gln	
					390						395				400		
50	Pro	Phe	Ile	Trp	Cys	Met	Leu	His	Asn	Phe	Gly	Gly	Asn	His	Gly	Leu	
					405					410					415		
55	Phe	Gly	Ala	Leu	Glu	Ala	Val	Asn	Gly	Gly	Pro	Glu	Ala	Ala	Arg	Leu	
				420					425					430			
60	Phe	Pro	Asn	Ser	Thr	Met	Val	Gly	Thr	Gly	Met	Ala	Pro	Glu	Gly	Ile	
								440						445			
65	Ser	Gln	Asn	Glu	Val	Val	Tyr	Ser	Leu	Met	Ala	Glu	Leu	Gly	Trp	Arg	
							455					460					
70	Lys	Asp	Pro	Val	Pro	Asp	Leu	Ala	Ala	Trp	Val	Thr	Ser	Phe	Ala	Ala	
	465					470					475				480		
75	Arg	Arg	Tyr	Gly	Val	Ser	His	Pro	Asp	Ala	Gly	Ala	Ala	Trp	Arg	Leu	
					485					490					495		
80	Leu	Leu	Arg	Ser	Val	Tyr	Asn	Cys	Ser	Gly	Glu	Ala	Cys	Arg	Gly	His	
					500					505					510		
85	Asn	Arg	Ser	Pro	Leu	Val	Arg	Arg	Pro	Ser	Leu	Gln	Met	Asn	Thr	Ser	
									520					525			

## ES 2 319 828 T3

Ile Trp Tyr Asn Arg Ser Asp Val Phe Glu Ala Trp Arg Leu Leu Leu  
 530 535 540

5 Thr Ser Ala Pro Ser Leu Ala Thr Ser Pro Ala Phe Arg Tyr Asp Leu  
 545 550 555 560

10 Leu Asp Leu Thr Arg Gln Ala Val Gln Glu Leu Val Ser Leu Tyr Tyr  
 565 570 575

15 Glu Glu Ala Arg Ser Ala Tyr Leu Ser Lys Glu Leu Ala Ser Leu Leu  
 580 585 590

20 Arg Ala Gly Gly Val Leu Ala Tyr Glu Leu Leu Pro Ala Leu Asp Glu  
 595 600 605

25 Val Leu Ala Ser Asp Ser Arg Phe Leu Leu Gly Ser Trp Leu Glu Gln  
 610 615 620

30 Ala Arg Ala Ala Ala Val Ser Glu Ala Glu Ala Asp Phe Tyr Glu Gln  
 625 630 635 640

35 Asn Ser Arg Tyr Gln Leu Thr Leu Trp Gly Pro Glu Gly Asn Ile Leu  
 645 650 655

40 Asp Tyr Ala Asn Lys Gln Leu Ala Gly Leu Val Ala Asn Tyr Tyr Thr  
 660 665 670

45 Pro Arg Trp Arg Leu Phe Leu Glu Ala Leu Val Asp Ser Val Ala Gln  
 675 680 685

50 Gly Ile Pro Phe Gln Gln His Gln Phe Asp Lys Asn Val Phe Gln Leu  
 690 695 700

55 Glu Gln Ala Phe Val Leu Ser Lys Gln Arg Tyr Pro Ser Gln Pro Arg  
 705 710 715 720

Gly Asp Thr Val Asp Leu Ala Lys Lys Ile Phe Leu Lys Tyr Tyr Pro  
 725 730 735

60 Gly Trp Val Ala Gly Ser Trp  
 740

55 (2) INFORMACIÓN PARA LA SEQ ID NO:3:

(i) CARACTERÍSTICAS DE LA SECUENCIA:

- (A) LONGITUD: 10380 pares de bases
- (B) TIPO: ácido nucleico
- (C) CATENARIEDAD: sencilla
- (D) TOPOLOGÍA: lineal

65 (ii) TIPO DE MOLÉCULA: ADN (genómico)

(vi) FUENTE ORIGINAL:

- (A) ORGANISMO: *Homo sapiens*

## ES 2 319 828 T3

(viii) POSICIÓN DEL GENOMA:

(A) CROMOSOMA/SEGMENTO: Cromosoma 17

(ix) CARACTERÍSTICA:

(A) NOMBRE/CLAVE: exón 1

(B) LOCALIZACIÓN: 990..1372

(ix) CARACTERÍSTICA:

(A) NOMBRE/CLAVE: exón 2

(B) LOCALIZACIÓN: 2115..2262

(ix) CARACTERÍSTICA:

(A) NOMBRE/CLAVE: exón 3

(B) LOCALIZACIÓN: 3056..3202

(ix) CARACTERÍSTICA:

(A) NOMBRE/CLAVE: exón 4

(B) LOCALIZACIÓN: 3387...3472

(ix) CARACTERÍSTICA:

(A) NOMBRE/CLAVE: exón 5

(B) LOCALIZACIÓN: 5667..5923

(ix) CARACTERÍSTICA:

(A) NOMBRE/CLAVE: exón 6

(B) LOCALIZACIÓN: 7745..8955

(xi) DESCRIPCIÓN DE LA SECUENCIA: SEQ ID NO:3:

35	ATAATGAGCA GTGAGGACGA TCAGAGGTCA CCTTCCTGTC TTGGTTTTGG CAGGTTTTGA	60
	CCAGTTTCTT TGCTGCATTC TGTTTTATCA GCGGGTCTT GTGACCTTTT ATCTTGTGCT	120
40	GACCTCCTGT CTCATCCTGT GACGAAGGCC TAACCTCCTG GGAATTCAGC CCAGCAGGTC	180
	TCTGCCTCAT TTTACCCAGC CCCTGTTCOA GATGGAGTCG CTCIGGTTGG AAACCTTCGA	240
45	CAAAATGACA GCTCCTGTTA TGTTGCTGCT GCTGCCGCCA ATGGACAGCC TTTAACGTGC	300
	CCGCCAGCCC TGCTCCACCG CCGGCCTGGG CTCACATGGC CCCATCCCTC CTCGAACCTC	360
50	CTAGCCTGTT AGTTACTCAA ATCTGCAAGC TCTCTGCCTT CTCAGGGCCT TCAATAAATG	420
	CATTTCTTCT GTCTGGAAGG CTCTTCCTTT CCCTCTTCTA GCCAATTCCT ATTCATCCCT	480
55	GAGTTTCAGA TTAAAAGTCA CTTCTTTGG AAACCTTACT TCGCTACTTC GCTACTTACT	540
	GCACTACTTC GCAGCATCAC AACTATGATG GAAATCCTTA CTTACGTTAA ATATCTGGTT	600

60

65

# ES 2 319 828 T3

	TCTAGGTCAC CTCCTGACG GGGACGGTAG GGACCGTCTT CTCGTTTCATC AGTAGGGAAG	660
	TAGCTATGGC AGTGCCTGAT ACAAATAAAA CTCCAAATGT GTATTTATTA GATGGTTGGA	720
5	TGGAAGTTAT TTGCGTGTGA AAGCGCGTTT TACCCGAAGG CGCTCTGTGA GGGCCAGCGG	780
	GTCCCCTTCG GCCCTGGAGC CGGGGTCACA CGCTCCCCAC CGCGTGCGGT CACGAGACGC	840
10	CCCCAAGGGA GTATCCTGGT ACCCGGAAGC CGCGACTCCT GGCCCTGAGC CCGGGCTTAG	900
	CCTTCGGGTC CACGTGGCCG GAGCCGGCAG CTGATTGGAC GCGGGCCGCC CCACCCCCTG	960
15	GCCGTGCGGG GACCCGCAGG ACTGAGACCA TGGAGGCGGT GCGGTGGCC GCGGCGGTGG	1020
	GGGTCTTCT CCTGGCCGGG GCCGGGGGCG CGGCAGGCGA CGAGGCCCGG GAGGCGGCGG	1080
20	CCGTGCGGGC GTCGTGGCC CGGCTGCTGG GGCCAGGCC CGCGGCCGAC TTCTCCGTGT	1140
	CGGTGGAGCG CGCTCTGGCT GCCAAGCCGG GCTTGGACAC CTACAGCCTG GCGGCGGCG	1200
25	GCGCGGCGCG CGTGCGGGTG CGCGGCTCCA CGGGCGTGGC GGCCGCCGCG GGGCTGCACC	1260
	GCTACCTGCG CCACTTCTGT GGCTGCCACG TGGCTGGTC CGGCTCTCAG CTGCGCCTGC	1320
30	CGCGGCCACT GCCAGCCGTG CCGGGGGAGC TGACCGAGGC CACGCCAAC AGGTACCGCC	1380
	CCGAAGCTTC CCCGCGTCCG CCCGAGGCSC TTACCCCTC CCGGAGCCGC TGCCACCCAA	1440
35	ATCGGGAGGC TGAGCGGGGA GCGCTGGCCG GAAGGCCAG CTGCGCCGCC TCCAGCAGCT	1500
	GTGTGGCCTT GAGCCAGCCA CTCTGCCTTT CAGAGCCTCG GCTGGCCAC CTGAAAAACG	1560
40	GAAAGAAGAC GCCTACCGTG CAGTGTATT GTGAGGATTT GCACGATGAT GGGCATAGAA	1620
	TTTGTGGTGC ACAATTGGTG ATGAGTGAAT TTTCTTGCT TCCTCCCCCA CCTTCTCTTT	1680
45	GAACCTGCGG ACTGAGGAAG GACGCCTCCA TCCCCACCC TACAGGCCTG TGTTCAGCG	1740
	CCTGCCACAC TATGGAGTGA TGTGTTTACA CAGCTGTCTT CCCCTGCCA TCTGTTAGAC	1800
50	TGTGGGGGCA GGGATTCCCC GTTCCAGGAA AACACCGTGC AGAGGAGGGG CTCTGGCAGT	1860
	GTGGCATGAA AGTGAATAT GCCACCCAAA TACCCGCCAG GCTAGAGGGC CCTGGGAGAG	1920
55	TGCAGGGGAC GAGTGCCTCA GAAGCCGAG CCCGGTACCT GGTCTCAGCT CCACCTGGGG	1980
	TGGGTCCCAG TGTGAGCAG AAGGGCCGAG TTTGGAGCCC CTCCCCTCTC CTCTAGGTGG	2040
60	GGGATGGGGG ATTTGTTCCA GGGCCGTGGA CCCTCCAGGG TGGGATGCGC CCCTGCTCAT	2100
	GACACTGCCC GCAGGTACCG CTATTACCAG AATGTGTGCA CGCAAAGCTA CTCCTTCGTG	2160
65	TGGTGGGACT GGGCCCCTG GGAGCGAGAG ATAGACTGGA TGGCGTGAA TGGCATCAAC	2220

# ES 2 319 828 T3

	CTGGCACTGG CCTGGAGCGG CCAGGAGGCC ATCTGGCAGC GGGTGCCTGC CCACTGTCCC	2280
	TTCCCCACCC TCCTCTATGG CGGGAGCCAC CGTAGGTGTT TTCACCCGCC CCCCAGCATG	2340
5	GGCGCAGTGT CTCTCTCTAG AAGTGCTTTC AGCGTGCACA GTGGCTTGGG CCTCCTAAAA	2400
	ACTGAGGCTT CCGGCCGGGC GCGGTGGCTC ACGCCTGTCA TCCCAGCACT TCGGGAGGCC	2460
10	TAGGCGGGCG GATCAGGAGT TCAGGAGATC GAGACCATCC TGGCCAACAT TGTGAAACCC	2520
	CGTCTCTACT AAAATACAAA GAAATAGCAA CCTGGGCAAC AGAGCGAGAC TCTGTCTAAA	2580
15	AAAAAAAAAA AAAAAAAGT AGGCTTCCAG TTTGAGGAGT GGGGCTCCTT CCCCATCTC	2640
	CCCTATGCAG CCAATCACCT GGTCCCTTGG ATCCAACCTCA TGGGCAGCTC TAGATCTGCC	2700
20	TCCCTGGAAG CTTCTGTGCT GCAATGGCTG CTCCAGGCTC TGCTTAAGCT CTTACACAG	2760
	TTGCCCTGCC CTTCCATCTG GCACTCTTGC TCCATGAAGC CTTCTAAGGC CTTCTGTG	2820
25	GGGAAAGCC CCTTGTGCTC CCATCTCCTC ACCCATGCGA CAAAGGCAAC ACAGTGAAC	2880
	CACCTACTCA CAGGCTCTT TCCTCTGGGC TGTGGGCTCC TTGATGGCAG CGTTCGGATT	2940
30	TTGTCTCAGT AGCCCTAGCA CCCAGCACAA AGAAGCAATG AGTGAATGGT TGTGAATGA	3000
	ATGAATGAAT GAATGAAGAT GAATATATTT CTATGTGTGG GCCCTTCTTC CTCAGGTGTA	3060
35	CCTGGCCTTG GGCCTGACCC AGGCAGAGAT CAATGAGTTC TTTACTGGTC CTGCCTTCCT	3120
	GGCCTGGGGG CGAATGGGCA ACCTGCACAC CTGGGATGGC CCCCTGCCCC CCTCCTGGCA	3180
40	CATCAAGCAG CTTTACCTGC AGGTAAAAGG ATGGAAAAGG GAAGGGGCAG AATCGGTGAT	3240
	AGATGGTCAT GGGCCCAGGA AGGGTGGTAT TAGGCCGGCC CCAGGGCTCT TAACTGAGGC	3300
45	GGGGGGCTGC GTGTATCCTG GGAGATGAGG GCCTTCTCAT AGGACAGCAG TGGCCATGCT	3360
	CACCACCCTT CTTTCTGTTC CTCCAGCACC GGGTCTGGA CCAGATGCGC TCCTTCGGCA	3420
50	TGACCCAGT GCTGCCTGCA TTCGCGGGGC ATGTTCCCGA GGCTGTCACC AGGTGAGGTT	3480
	CCGCTCACCC CCTCCACTTA GCTCAGAGAG GGAATTTTAT TCCCTTCTAG AACATGACTT	3540
55	AAAACTTAA GCTCTGGGCC GGGCGCAGTG GCTCACGCCT GTAATCCCAG CACTTTGGGA	3600
	GGCCGAGTTG GGCGGATCAC CTGAGGTCAG GAGTTCGAGA CCAGCCTGGC CAACATGGTG	3660
60	AAACCCTGTC TCTACTAAAA ATATAAAAAT TAGCTGGGCA TGGTGGCACG CGCCTGTAAT	3720
	CCCATCTACT TAGGAGGCTG AGACAGGAGA ATTGCTTAAA CCTGGGAGGC AGACGTTGCA	3780
65	GTGAGTCAAG ATCACGCCAT TGCACTCCAG CCTGGGTGAC GAGCGAAACT CTGTCTCAA	3840

# ES 2 319 828 T3

	CAAACAAACA AGCTCTGGAC GTAGGCCTGG GTTTGATTTT TGA CTCTGCT ACTAATTAGC	3900
5	TGTGTGACTT CGGGCAGATG ACATGACTGC TCTGTGCCTC AGTTTCCTTA CTTGTAAAAT	3960
	GGGATCTCTA CCCACTTCGC TGTAGGGTTT GTAATTATCT CTCGATCTAT CTGTGACTTT	4020
10	GCACAGAGTG CTAGCAAATG GCAGCCCTTG GGAGTGGCAG CAGGGGTGCT CCAGTGTCCC	4080
	TTGTCCCTCC TGTTCCTCTG TGCTTCCCAG CCATCCTCTC ACATGTGGTT GGGAAAAGTC	4140
15	TTCAAGGCTC ACCTGAGACC TCCCCTCCTT CAGGAAGCCT TGCTAGTGCC CCGCATGACC	4200
	TCCTTTGCAC CTGCTAATGT CTGGCTCCCA TACTCTCGTA GGACTTAATG CATGCCAGTG	4260
20	GCCTCCCTGC CCGCCTCTTT GCCCCCATCA CCAGGTGGCA GGAAACTCAC TCATTCATTC	4320
	AATAAACTTG GTCCAGCTGT CTGAGGCTGC CAGAACTGGC TGTGCTGGGT CCTGGGAGGC	4380
25	GGCAAGAAAG GTGCCCAAGG GCTTACCCCT GATAGGAGAG ATATGTTGGC TGAAGGATAC	4440
	AATGTGGGGA CAAGGACAGG AATATATGTG GGTTCGCTC TCCTCTGCCG GGAGAGAGGG	4500
30	GCAGGAAGGG CTCAGGGCAG AGCCAGCCT TGA AAAATGA GTGTTGCTTG GACGGACGCT	4560
	TGGCTAATGC TTGTAATCCT AGCGTTTTGG GAGGCTGAGG CGTATGGATC ACCTGCGGTC	4620
35	AGGAGTTAAA GACCAGCCTG GCCAACATGG CGAAACCCCA TCTCTACTAA AAGTACAAA	4680
	ATTAGCCAGG CGTGGTGGCG GGCTCCTGTA ATCCAGCTA CTCGGTAGGC TGAGGCATGA	4740
40	GAATCTCTTG AAGCCAGGGG CCAGAGACTG CAGTGAGCCG AGATCACACC ACTTCACTCC	4800
	AGCCTGGGTG ACAGAGTGAG ACTCCGTCTC AAAAAAAAAA AAAAAAAAAAG GAAAGAAAAT	4860
45	TAAACACCTC ATGTTCTCAC TCATAGTGGG AGTTGAACAA TGAGAACAAC ATGGACACAG	4920
	GAAGGGGAAC ATCACACACC GGGGCCTTTC GCGGTGTGGG GGTCAAGGGG AGGAGTAGCA	4980
50	TTGGGACAGA TACTTAATGC ATGCGGGGCT GAAAACCTAG ATGATGGGTT GATGGGTGCA	5040
	GCAAACCACC ATGGCACATG TATACCTATG CAACAAACCT GCATGTTCTG CACAGAACTG	5100
55	AACTGAAAGT ATAATTA AAAA AAAAAA AAGCTGGGTG CGGTGGCCCA CACCTGTAAT	5160
	CCCAGCACTT TGGGAGGCCG AGACGGGCGG ATCACAAGGT CAGCAGATCG AGACCATCCT	5220
60	GGCTAACACA GTGAAACTCA GTCTCTACTA AAAATACAAA AAATTAGCCG GGTGTGGTGG	5280
	CGGGCACCTG TAGTCCCAGC TACTAGGGAG GCTGAGGCAG GAGAATGGCA TGAACCTGGG	5340
65	AGGCAGAGCT TGCAGTGAGC TGAGAATGCG CCACTGCACT CCAGCCTGGG GGACAGAGTG	5400
	AGACTCTGCC TCAAAAAAAAA AAAAAAAAAAG AAAGAAAAG GAGCGTTGCT TGTTTCAGGC	5460

# ES 2 319 828 T3

	CACAGGAAGG GGAGAGATAG TGAAAGTTTT TCAGAGAAGG TGGCCAGGGA AGGAGAAGAA	5520
5	AGGACTGTAG GCAGAGAGCA TAGCCTGTAC AAAGCCATAG AGGCAAGAGA AACCAGGAGC	5580
	TGTAGAGAAG TTGGCAAGGC TGTGAAACAC TATGGTGAAC ACTATGGCGG CTTCCATGAA	5640
10	ATATCTGAGC TTTTGCTCCC CACTAGGGTG TTCCTCAGG TCAATGTCAC GAAGATGGGC	5700
	AGTTGGGGCC ACTTAACTG TTCCTACTCC TGCTCCTTCC TTCTGGCTCC GGAAGACCCC	5760
15	ATATCCCCA TCATCGGGAG CCTCTCCTG CGAGAGCTGA TCAAAGAGTT TGGCACAGAC	5820
	CACATCTATG GGGCCGACAC TTTCAATGAG ATGCAGCCAC CTTCTCAGA GCCCTCCTAC	5880
20	CTTGCCGAG CCACCACTGC CGTCTATGAG GCCATGACTG CAGGTACAGT GCCTGGGTGG	5940
	GGTGGGAGAG CCCCCAGAC CCTCAAAAAG AAGGGAGTAG CAGATGTCAG TAGGGGTAGG	6000
25	CAGAGGGACT GGAATAATGC CTCGCCATAA CACACAGTAC TTTATAGTTT ACCAAGCACG	6060
	TGTACACATG CGTTGTCTCA GTGAATCCCA CTGTGGTTGA GAGGTGAGCT CTGGAAGCCA	6120
30	ACAACCTGGG TCACACCTCG CGCTCCTATT TCCTGGCCGT GTGACTTATG ACTCATGACC	6180
	TCCTTCCCAG TGCTCGTTT GCTTTTCTG TAACTGGGA CTACCTCATA GGTAGAATAA	6240
35	CGCCTGGCCC AGAGCAAAGG CCACTAAGAG CTAGCTATGA ACAAGGATTI TGTTTCATCT	6300
	CTGCGTGGTT GCTGAAGTAG GCACTGCAGG CAGGAGGTGA GTGGATGTGC CTAAAGGCAC	6360
40	TAAGTGCACA TCCTGCTACA AAACTGTGAA GCCAGGGCTC CTCCTGCCA CTTAAAGGAG	6420
	GAGTGGAGCA GAGGGCGCCC AAGTCAGGAA TGACTTAGTG GAGAGGCGTC TGTGTTGGCC	6480
45	AGGAAGGGAA CAGATCAGCT CAGCCTTCT TGAGCAGTAC TGCTCCAAGT GTGACCCAAA	6540
	ACCAGCAGCA GCAGCAGCAG CAGCCCGAGC TGTGAGATGG CAAATTCTCA GGCCCTACCC	6600
50	AAGACCTGAA GGAGAAGCTA CTTTTTTTTT TTTTTGAGA CAGATTTTAC TCTGTTGCTG	6660
	AGGCTGGAGC ACAGTGGCAC AATCTCATCT CACTGCAACC TTCGTCTCCT AGGTTCAAGC	6720
55	GATTCTCCTG CCTCAGCCTC CCGAGTAGCT GGGACTATAG GCACCCGCCA CCACGCCCGG	6780
	CAATTTTGT TTGTTTTGAG ATAGAGTCTC GCTCTGTCAC CCAGGCTGGA GTGCAGTGGC	6840
60	ACGATCTCAG TTCACTGCAA CCTCTGCTTC CTGAGTTCAA GCGATTCTCC TGCCCTAGCC	6900
	TCCTGAGTAG CTGGGATTAC AGGCGCCCCC CAACCACACT CGGCTAATTT TTGTATTTTT	6960
65	AGTAGAGACG GGGTTTCGCT ATGTAGGTCA AGCTGGTTTC AAACTCCTGA CCTCAAATGA	7020
	TTCGCCACT TCAGCCTCCC AAAGTGCTGG GATTACAGGT GTGAGCCACC TTGCCTGGCC	7080

# ES 2 319 828 T3

	AATTTTTGTA	TTTTTAGTAG	AAACAGGTTT	CACCATGGTG	GCCAGACTGG	TCTCAAACCTC	7140
	CTGACCTCAG	GTGAACTGCC	CACCTCAGCC	TCCCAAAGTA	CTGGTATTAC	AGGCGTGATC	7200
5							
	CACTGCGACT	GGCCTTGATT	TTGTTTTTGA	GACAGAATCT	TACTCTGTCTG	CCCAGACTGG	7260
	AGTGCAGTGG	CACAATCTCA	GCTCACTGCA	ACTTCTGCCT	CATGGGTTC	AGTGATTCTT	7320
10							
	GTGCCTCTAC	CTCCCGAGTA	GCCGGGATTA	CAGGCACCTG	CCATTACGCT	AGGCTAATTT	7380
	TTGTATTTTT	AGTATAGACA	GGGTTTCCCC	ACATTGGCCA	GGCTGGTCTG	GAACTCCTGG	7440
15							
	GCTCAAGTGA	TCCACCTGCT	TCAGCCCCTC	AGAGTACTGG	GATTATAGGT	GTGGGCCACC	7500
	ACGCCCATTG	AGAAACCTCC	ATGTTTTAAG	GAGCCCTCTG	GGTAACTCTC	ATGTTCACCC	7560
20							
	AAGCTGCTGA	ACCCTGTCCT	GGAGTTTCA	GAGGGACGCG	TATGTGCCAC	AGAGCGTCCC	7620
	GCTGGTGGGG	GTCATGGGAA	GCCATGACCT	GGGATAGACA	GTCGTCTGTA	GAGTGGGGTG	7680
25							
	AACATTCCCT	GGGCCCTCTG	TTTCATCACT	CCTCTTCTCT	GTTCCCCCTA	CCTCCTGTCC	7740
	ACAGTGGATA	CTGAGGCTGT	GTGGCTGCTC	CAAGGCTGGC	TCTTCCAGCA	CCAGCCGCAG	7800
30							
	TTCTGGGGGC	CCGCCAGAT	CAGGGCTGTG	CTGGGAGCTG	TGCCCGTGG	CCGCCTCCTG	7860
	GTTCTGGACC	TGTTTGCTGA	GAGCCAGCCT	GTGTATACCC	GCACTGCCTC	CTTCCAGGGC	7920
35							
	CAGCCCTTCA	TCTGGTGCAT	GCTGCACAAC	TTTGGGGGAA	ACCATGGTCT	TTTTGGAGCC	7980
	CTAGAGGCTG	TGAACGGAGG	CCCAGAAGCT	GCCCGCCTCT	TCCCAACTC	CACCATGGTA	8040
40							
	GGCACGGGCA	TGGCCCCCGA	GGGCATCAGC	CAGAACGAAG	TGGTCTATTG	CCTCATGGCT	8100
	GAGCTGGGCT	GGCGAAAGGA	CCCAGTGCCA	GATTTGGCAG	CCTGGGTGAC	CAGCTTTGCC	8160
45							
	GCCCCGCGGT	ATGGGGTCTC	CCACCCGGAC	GCAGGGGCAG	CGTGGAGGCT	ACTGCTCCGG	8220
	AGTGTGTACA	ACTGCTCCGG	GGAGGCCTGC	AGGGGCCACA	ATCGTAGCCC	GCTGGTCAGG	8280
50							
	CGGCCGTCCC	TACAGATGAA	TACCAGCATC	TGGTACAACC	GATCTGATGT	GTTTGAGGCC	8340
	TGGCGGCTGC	TGCTCACATC	TGCTCCCTCC	CTGGCCACCA	GCCCCGCCTT	CCGCTACGAC	8400
55							
	CTGCTGGACC	TCACTCGGCA	GGCAGTGCCAG	GAGCTGGTCA	GCTTGTAATA	TGAGGAGGCA	8460
	AGAAGCGCCT	ACCTGAGCAA	GGAGCTGGCC	TCCCTGTTGA	GGGCTGGAGG	CGTCCTGGCC	8520
60							
	TATGAGCTGC	TGCCGGCACT	GGACGAGGTG	CTGGCTAGTG	ACAGCCGCTT	CTGCTGGGC	8580
	AGCTGGCTAG	AGCAGGCCCG	AGCAGCGGCA	GTCAGTGAGG	CCGAGGCCGA	TTTCTACGAG	8640
65							
	CAGAACAGCC	GCTACCAGCT	GACCTTGTGG	GGGCCAGAAG	GCAACATCCT	GGAATATGCC	8700

# ES 2 319 828 T3

	AACAAGCAGC TGGCGGGGTT GGTGGCCAAC TACTACACCC CTCGCTGGCG GCTTTTCCTG	8760
	GAGGCGCTGG TTGACAGTGT GGCCAGGGC ATCCCTTCC AACAGCACCA GTTTGACAAA	8820
5	AATGTCTTCC AACTGGAGCA GGCCTTCGTT CTCAGCAAGC AGAGGTACCC CAGCCAGCCG	8880
	CGAGGAGACA CTGTGGACCT GGCCAAGAAG ATCTTCCTCA AATATTACCC CGGCTGGGTG	8940
10	GCCGGCTCTT GGTGATAGAT TCGCCACCAC TGGGCCTTGT TTTCCGCTAA TTCCAGGGCA	9000
	GATTCCAGGG CCCAGAGCTG GACAGACATC ACAGGATAAC CCAGGCCTGG GAGGAGGCC	9060
15	CACGGCCTGC TGGTGGGGTC TGACCTGGGG GGATTGGAGG GAAATGACCT GCCCTCCACC	9120
	ACCACCCAAA GTGTGGGATT AAAGTACTGT TTTCTTTCCA CTTAAACTGA TGAGTCCCCT	9180
20	GGGTCTGTCA AAATGAGAAG GTCACTGCTG CCACGCTTGG GAGGACTCAG GGCTATAGCA	9240
	TGGCCCTGGG GTGGGACCTG TTCTCCCATC CCTTGCCTCA CGTCCCTGTT TTTGTTTTGT	9300
25	TGTTTGTITG TGACGGAGCC TTGGTCTGTT GCCCAGGCTT GAGTACAATG GCACAGTCTC	9360
	GGCTCACTGC AACCTCCGCC TCCTGGGTTC AAGCAATTCT TGTGCCTCAG CCTCCCCGGT	9420
30	AGCTGGGACT ATAGGCATGC ACCACCACAC CAGGCTAATT TTTTTTTTC CAAGATGGAG	9480
	TCTTGCTCTG TCGCCAGGT TGGAGTTTAG TGGCACCATA TTGGTTTACT GCAACCTCTG	9540
35	CCTCCCCGGT TCAAGCAATT CTCCTGCCTC AGTCTACCAG GGAGTTAGGA CTACGGGCCT	9600
	GTGCCATCAC GCCTGGCTAA TTTTTGTATT TTTCATAGAG ATAAGGTTTC ACCATGTTGG	9660
40	CCAGGCTGGT CTTTAACTCC TGAACTCAAG TGATCCACCT GCCTCGGCCT TCCAAAGTGC	9720
	TGGGATTACA GGAGTGAGCC ACCGTGCCCG GCCATGTCTC TCTTTTAACT ACTAATGTTA	9780
45	CCCTGACCTT TGAACGTAGA ATGCCCTTCT GTTGCAGGAA AACCTCTTTT CAAACCATGT	9840
	TTGTCTTTG CTGGCATGCC ACAGCAACAG TCACCAACAC AGAAGACTTC TGTGACCAA	9900
50	TATTTGGAGG ATTTTCCCA CACACACCAA GCAGCAGACA TCAGCTGGGT GTCCTCCAAT	9960
	TCAGTTCCAA TGTAATCAAC CAGAGACAGC ATCAGATCCC ACAGGGTTAG GGTGCAGATC	10020
	CATGAGACCA CCCCCTCCTT CCCAACGGTT ACAAGTCCTG ATCCCTGGAA CTTCTGACTA	10080
55	ACTGGCTTCA AGTTGGAGTT CCCATGACCC CCTTCCCCTC TTTGGAGTCA ACTCATTGTC	10140
	GACAGTGACC CACGAAACAC AGGGAAACCC TTATTATGTT TATTGCTTTA TTACAGAGGA	10200
60	AAAAAATTTT TTTCTTCTT TTTTGAGACA GGGTCTCACT CTGTCAATCA GAATGACTGC	10260
	AGTGGCAGSA TCTGGCTCCG TCACCCAGGC TGGAGTGACG TGGCATGATC TCGGCTCACT	10320
65	ACAGCCTCCA TCCCCCCTAA ACCCCACGCC TCAGCGCCCC ACCCCGCAAG TGGCTGGGAC	10380

## ES 2 319 828 T3

### (2) INFORMACIÓN PARA LA SEQ ID NO:4:

(i) CARACTERÍSTICAS DE LA SECUENCIA:

- (A) LONGITUD: 20 aminoácidos  
(B) TIPO: aminoácido  
(D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: péptido

(iii) HIPOTÉTICA: NO

(v) TIPO DE FRAGMENTO: N-terminal

(vi) FUENTE ORIGINAL:

- (A) ORGANISMO: *Homo sapiens*

(ix) CARACTERÍSTICAS:

- (A) NOMBRE/CLAVE: Sitio-modificada  
(B) LOCALIZACIÓN: 10

(xi) DESCRIPCIÓN DE LA SECUENCIA: SEQ ID NO:4:

Asp Glu Ala Arg Glu Ala Ala Ala Val Arg Ala Leu Val Ala Arg Leu  
1 5 10 15

Leu Gly Pro Gly  
20

### (2) INFORMACIÓN PARA LA SEQ ID NO:5:

(i) CARACTERÍSTICAS DE LA SECUENCIA:

- (A) LONGITUD: 18 aminoácidos  
(B) TIPO: aminoácido  
(D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: péptido

(iii) HIPOTÉTICA: NO

(v) TIPO DE FRAGMENTO: N-terminal

(vi) FUENTE ORIGINAL:

- (A) ORGANISMO: *Homo sapiens*

(ix) CARACTERÍSTICAS:

- (A) NOMBRE/CLAVE: Sitio-modificada, glicosilada o fosforilada en donde Xaa puede ser cualquier resto de aminoácido, con preferencia Arg  
(B) LOCALIZACIÓN: 16

(xi) DESCRIPCIÓN DE LA SECUENCIA: SEQ ID NO:5:

Lys Pro Gly Leu Asp Thr Tyr Ser Leu Gly Gly Gly Gly Ala Ala Xaa Val  
1 5 10 15

Arg

