



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



① Número de publicación: **2 212 695**

② Número de solicitud: 200102090

⑤ Int. Cl.7: **C12N 9/64**
C12Q 1/37
C07K 16/40
A61K 38/48

⑫

SOLICITUD DE PATENTE

A1

⑫ Fecha de presentación: **14.09.2001**

⑬ Fecha de publicación de la solicitud: **16.07.2004**

⑭ Fecha de publicación del folleto de la solicitud:
16.07.2004

⑰ Solicitante/s: **Universidad de Oviedo**
Plaza del Riego, 4 Edificio Histórico
33003 Asturias, ES

⑱ Inventor/es: **Cal Miguel, Santiago;**
Obaya González, Álvaro Jesús;
Llamazares Prada, María;
Garabaya Fernández, Cecilia y
López Otín, Carlos

⑳ Agente: **No consta**

㉔ Título: **Procedimiento de identificación de la proteína humana ADAMTS-13.**

㉖ Resumen:

Procedimiento de identificación de la proteína humana ADAMTS-13.

La invención consiste en identificar fragmentos de genes humanos similares a secuencias de genes de proteínas ADAM, amplificarlos mediante PCR de ARN de tejidos humanos, extender la secuencia de los fragmentos obtenidos hacia los extremos 5' y 3' y determinar la secuencia de los clones de ADNc generados. La secuencia identificada es SEQ ID NO :1 y se ha denominado ADAMTS-13. La aplicación de dicha secuencia está relacionada fundamentalmente: con la diagnosis y el tratamiento de anomalías en los procesos de angiogénesis, hemostasis, adhesión celular y remodelación tisular.

ES 2 212 695 A1

DESCRIPCIÓN

Procedimiento de identificación de la proteína humana ADAMTS-13.

Campo de la invención

La invención se adscribe al campo de los procesos biológicos de adhesión celular y remodelación tisular, incluyendo los asociados a condiciones fisiológicas como la respuesta inmune, la angiogénesis, la coagulación, la cicatrización de heridas, los procesos reproductivos, la implantación embrionaria, o el desarrollo fetal, así como procesos patológicos incluyendo los tumorales, artríticos, cardiovasculares, hematológicos y neurodegenerativos. En concreto, la presente invención versa sobre una proteína humana que contiene dominios de adhesión celular y metaloproteasa, sobre el gen que la codifica, y sobre sus posibles inhibidores. Más particularmente, la presente invención aborda la identificación de la proteína humana llamada ADAMTS-13, y el análisis de su estructura y de sus posibles funciones normales y patológicas.

Estado de la técnica

Las proteínas denominadas ADAMs (a disintegrin and metalloproteinase domain) o desintegrinas celulares, son una familia de enzimas que han adquirido una notable importancia dada su capacidad de participar en procesos biológicos que implican fenómenos de adhesión celular y proteólisis extracelular (Cell 90, 589, (1997)). Estas proteínas poseen una peculiar organización estructural con dominios de proenzima, metaloproteasa, desintegrina, rico en cisteína, factor de crecimiento epidérmico, transmembrana, y citoplasmático. Algunos de estos dominios son semejantes a los encontrados en una familia de proteínas aisladas de venenos de serpientes (Methods Enzymol. 248, 345, (1995)). Estas proteínas de serpientes junto con las ADAMs, constituyen la superfamilia de las reprotinas, caracterizadas por la presencia de una secuencia HEXXHXXGXXHD en su dominio catalítico.

Las ADAMs han sido identificadas en una variedad de tejidos de mamíferos, así como en otros organismos eucariotas como *Xenopus laevis*, *Drosophila melanogaster* y *Caenorhabditis elegans*, pero no en plantas, levaduras o bacterias. Inicialmente, las ADAMs se asociaron a procesos reproductivos, pero posteriormente su espectro de funciones se ha extendido considerablemente (Curr. Opin. Cell Biol. 10, 654, (1998)). Así, la meltrina- α (ADAM-12) se ha implicado en fusión de mioblastos. Las meltrinas α y β también participan en procesos de diferenciación y actividad osteoblástica. Otras ADAMs como las denominadas MS2 y decisina, participan en distintos procesos de la respuesta inmune. Además, estudios recientes han permitido caracterizar las propiedades enzimáticas y especificidad de sustrato de varias ADAMs como ADAM-9, ADAM-10 o ADAM-17 que actúan como proteasas implicadas en el procesamiento proteolítico de sustratos celulares relevantes, incluyendo precursores de citoquinas y factores de crecimiento.

La complejidad estructural y funcional de esta familia de proteínas se ha extendido considerablemente tras el reciente hallazgo de una serie de nuevas proteasas relacionadas con las ADAMs y caracterizadas por la presencia en su secuencia de aminoácidos de varias copias de dominios trombospondina (J. Biol. Chem. 274, 25555, (1999)). El primer miem-

bro de esta subfamilia, denominado ADAMTS-1, se identificó como consecuencia de su asociación con el desarrollo de caquexia tumoral y de varios procesos inflamatorios. Posteriormente, se identificó la ADAMTS-2, con actividad de procolágeno I amino-proteasa y cuya deficiencia origina el síndrome de Ehlers-Danlos tipo VIIC (Am. J. Hum. Gen. 65, 308, (1999)). Otros miembros de la familia son las proteínas denominadas ADAMTS-4 y ADAMTS-11, las cuales poseen la actividad agreganasa responsable de la degradación del cartílago articular en enfermedades artríticas. Por otra parte la ADAMTS-8 y la ADAMTS-1 han sido identificadas como proteínas capaces de inhibir los procesos de angiogénesis. Finalmente, otras proteínas como las ADAMTS-3, ADAMTS-5, ADAMTS-6, ADAMTS-7 y ADAMTS-12 sólo se han caracterizado al nivel estructural y sus funciones todavía no han sido aclaradas. Todas estas proteínas tienen una organización similar en dominios, pero difieren sustancialmente de la estructura prototipo de las ADAMs. Así, las ADAMTS-s carecen del dominio de factor de crecimiento epidérmico, la región transmembrana, y la cola citoplasmática características de las ADAMs, pero contienen una serie de copias de trombospondina, que representan la característica distintiva de los miembros de esta familia de proteínas. El hallazgo de que las ADAMTS pueden estar implicadas en una amplia variedad de procesos biológicos y patológicos ha estimulado la búsqueda de nuevos componentes de la familia.

Una de las estrategias para la identificación de nuevas ADAMTS humanas consistiría en la aplicación de métodos de clonación por homología. Una de las múltiples formas de abordar este método, persigue en un primer paso la búsqueda en bancos de datos accesibles públicamente, de fragmentos de secuencias denucleótidos de genes humanos generados de manera aleatoria y que tengan similitud con las secuencias de los genes de las desintegrinas ya conocidas. Tras su identificación, los hipotéticos fragmentos homólogos se pueden amplificar mediante PCR de ARN total de tejidos humanos en los que se sospeche la expresión de dichos genes, y utilizarlos como sondas para hibridar genotecas de ADNc preparadas a partir de ARN de los mismos tejidos. Alternativamente, se puede extender la secuencia hacia los extremos 5' o 3' mediante técnicas de amplificación rápida de los extremos de los ADNcs (RACE). Finalmente, la secuenciación y posterior caracterización de los clones humanos aislados mediante técnicas estándar de Biología Molecular, permitiría confirmar la identificación de nuevas ADAMTS y definir el posible papel de las proteínas codificadas por dichos clones en procesos normales y patológicos de adhesión celular o proteólisis. Basándose en esta idea, los autores de la invención, tras los pertinentes estudios experimentales, han llegado a los objetivos antes enumerados que constituyen los diversos aspectos de la presente invención.

Breve descripción de la invención

Un objeto de la presente invención es identificar el gen humano que codifica una nueva proteína humana denominada ADAMTS-13.

Un segundo objeto de la invención es analizar la expresión en tejidos humanos del gen de la ADAMTS-13.

Un tercer objeto de la invención es analizar la expresión del gen de la ADAMTS-13 en tumores humanos.

Descripción detallada de la invención

El primer objeto de la invención consistió en la identificación de un gen humano que pudiera codificar una nueva ADAM humana. Para ello la secuencia de aminoácidos de regiones conservadas en las ADAMs descritas se comparó con la división de *Expressed Sequence Tags* (ESTs) de la base de datos GenBank utilizando el programa TBLASTN (J. Mol. Biol. 215, 403, (1990)). Se identificó una EST humana, cuyo número de acceso es AI806237. A partir de la secuencia de nucleótidos de esta EST, se diseñaron y sintetizaron dos oligonucleótidos, AD-1 (5'-CTGCCGCGGCTGCAGGGGA-3') y AD-2 (5'-GGACCCGTCCTGGGGCTCA-3'). Estos oligonucleótidos fueron utilizados para amplificar el fragmento de ADNc correspondiente, empleando como molde DNA total aislado a partir de una genoteca de ADNc humano de hígado fetal. Para ello se utilizaron 20 pmoles de cada oligonucleótido, aproximadamente 1 microgramo de ADNc, 0,2 mM dNTPs y 1,25 U de Taq DNA polimerasa en un volumen total de 50 microlitros de tampón ExpandLong 3 (Boehringer Mannheim). La amplificación se llevó a cabo en un aparato GeneAmp2400 de Perkin-Elmer, y consistió en 400 ciclos de desnaturalización (15 s, 94°C), hibridación (20 s, 64°C) y extensión (20 s, 72°C). El fragmento de DNA resultante, de 460 pares de bases (pb) se purificó mediante electroforesis en gel de agarosa y posterior extracción con GeneClean. La identidad del fragmento amplificado con la secuencia parcial de la EST-AI806237 se verificó tras subclonarlo en pUC18 y determinar su secuencia de nucleótidos mediante técnicas estándar de Biología Molecular.

Con el fin de obtener una secuencia de ADNc que contuviera la información codificante de la proteína completa, el producto de PCR obtenido con los oligonucleótidos AD-1 y AD-2, se marcó radiactivamente y se utilizó como sonda para hibridar una genoteca de ADNc de pulmón fetal humano. Tras el escrutinio de aproximadamente 1×10^6 unidades formadoras de placa, se obtuvo un clon con hibridación positiva cuya secuencia de nucleótidos se determinó mediante métodos convencionales. El análisis de dicha secuencia reveló que este clon contenía un inserto de 1.4 kb, que parecía estar incompleto en su extremo 5' y 3'. Dicha región se extendió mediante técnicas de amplificación rápida de extremos de ADNcs utilizando ARN de hígado fetal humano y el método Marathon de Clontech. Tras una serie de amplificaciones sucesivas se obtuvo un fragmento que contenía un codón de terminación en la misma fase de lectura que el resto del ADNc identificado. Finalmente, el ADNc codificante completo se obtuvo por amplificación con los oligonucleótidos ADTS13F (5'-ATGCACCAGCGTCACCCCGG-3') y ADTS13R (5'-GGTTCCTTCCCTTCCCTTCCAG-3'). El análisis informático de la secuencia obtenida reveló la existencia de una fase abierta de lectura, que codifica una proteína de 1340 aminoácidos a la que denominamos ADAMTS-13. Su secuencia de aminoácidos, así como la secuencia nucleotídica que la codifica se muestra como SEQ ID NO :1. La comparación de esta secuencia de aminoácidos con todas las secuencias presentes en los bancos de datos accesibles públicamente demostró la existencia de un grado significativo de similitud con otras ADAMs y más específicamente con miembros de la subfamilia de las ADAMTS (J. Biol. Chem. 274, 25555, (1999)). Así, la protef-

na presenta todos los motivos característicos de estas enzimas incluyendo la secuencia señal, el propéptido, los dominios metaloproteasa, desintegrina y rico en cisteína, así como diversas repeticiones tipo trombospondina (TS).

Un análisis más detallado de la secuencia de aminoácidos deducida para la ADAMTS-13, confirmó que ésta posee el predominio más corto entre todos los miembros de la familia. En él se localizan dos residuos cisteína (posiciones 27 y 37) que podrían estar implicados en el mantenimiento de la latencia enzimática. Este predominio termina en un motivo dibásico que podía corresponder al sitio de activación por furina, que poseen estas enzimas. El dominio catalítico incluye la secuencia HEXXHXXGXXHD (posiciones 224-235) implicado en la coordinación del átomo de zinc en el centro activo de las metaloproteasas, y con el residuo de ácido aspártico que permite distinguir las reprotinas de las MMPs. Este dominio también posee el residuo de metionina (posición 249) que contribuye a formar la estructura Met-giro presente en reprotinas y MMPs. Tras el dominio catalítico puede reconocerse el dominio desintegrina, similar en tamaño al de otras ADAMTS y con las ocho cisteínas altamente conservadas en dicha región. Finalmente, el dominio rico en cisteínas muestra un alto porcentaje de identidades (alrededor del 50%) con el dominio equivalente presente en otras ADAMTS incluyendo los diez residuos de cisteína conservados en todas ellas. Por todo ello, podemos concluir que la proteína identificada pertenece a la familia de las ADAMTS y ha sido denominada ADAMTS-13 La secuencia fue depositada en el banco de datos EMBL con el número de acceso AJ305314). Tanto el ADN aislado como el polipéptido codificado, representados en SEQ ID NO:1, como secuencias parciales obtenidas de ambos, pueden sintetizarse químicamente también.

El segundo objeto de la invención es analizar la expresión en tejidos humanos del gen de la ADAMTS-13. Con este fin, tres membranas conteniendo ARN poliadenilado procedente de múltiples tejidos humanos adultos (leucocitos, colon, intestino delgado, ovario, testículo, próstata, timo, bazo, páncreas, riñón, músculo esquelético, hígado, pulmón, placenta, cerebro y corazón) y fetales (cerebro, pulmón, hígado y riñón) se hibridaron con una sonda específica de ADAMTS-13 (276 pb, correspondiente a la región comprendida entre los aminoácidos 81 a 172) marcada radiactivamente. Tras una prehibridación a 42°C durante tres horas en formamida al 40%, 5x PBS/EDTA (1x=NaCl 150 mM, NaH₂PO₄ 10 mM, EDTA 1mM, pH 7,4), 10x disolución de Denhardt (1x= Albúmina de suero bovino, 0,02%, polivinilpirrolidona, 0,02%, ficoll, 0,02%), SDS 2% y DNA de esperma de salmón 100 mg/ml, se añadió la sonda y se hibridó durante 20 horas en las mismas condiciones. Los filtros se lavaron con 1x SSC (NaCl 150 mM, citrato sódico 15 mM, pH 7,0) conteniendo SDS al 0,1% durante 2 horas a 50°C, y finalmente se expusieron a autorradiografía (Fig. 1). Como puede observarse en la figura 1, tras hibridación con la sonda de ADAMTS-13, se detectó un ARN mensajero de aproximadamente 4.5 en hígado, tanto fetal como adulto, y un segundo transcrito de 2.4 kilobases mayoritariamente en hígado fetal, aunque su expresión también se localiza en otros tejidos como riñón y pulmón fetales, placenta y músculo de tejidos adultos, y en diversos tipos celulares incluyendo células endoteliales.

El tercer objeto de la invención consistió en el estudio de la expresión del gen de la ADAMTS-13 en muestras obtenidas de carcinomas humanos. Estos estudios se llevaron a cabo mediante análisis Northern utilizando ARN total extraído de diversas líneas tumorales humanas. Las correspondientes membranas se hibridaron con una sonda específica de ADAMTS-13 (276 pb) marcada radiactivamente utilizando las mismas condiciones que las descritas para el análisis de expresión de ADAMTS-13 en tejidos normales. Como puede observarse en la figura 2, tras hibridación con la sonda de ADAMTS-13, se detectó un ARN mensajero de aproximadamente 2.4 kilobases en algunas líneas tumorales como las denominadas SW480 y G361 derivadas de carcinoma de colon y de melanoma maligno humanos respectivamente. Estos resultados demuestran que que la ADAMTS-13 se produce en carcinomas derivados de tejidos en los que en condiciones normales no se expresa dicho gen. Estos resultados confirman la propuesta de que la ADAMTS-13, a través de posibles funciones como

proteasa, y/o como molécula reguladora de procesos de adhesión celular, hemostasis y angiogénesis, puede también asociarse a patologías que implican destrucción tisular como el cáncer.

Descripción de las figuras

Figura 1. Análisis Northern de la expresión del gen ADAMTS-13 en tejidos humanos adultos y fetales. Dos microgramos de ARN poliadenilado de los tejidos indicados se hibridaron con un fragmento del ADNc de ADAMTS-13. El origen del ARN de los distintos tejidos, así como el tamaño de los distintos ARNs utilizados como marcadores quedan indicados. Las membranas se hibridaron con una sonda de actina para asegurar la igualdad en la cantidad de ARN cargado en las distintas calles.

Figura 2. Análisis Northern de la expresión del gen ADAMTS-13 en líneas tumorales humanas. 20 microgramos de ARN total de las líneas tumorales indicadas se hibridaron con un fragmento del ADNc de ADAMTS-13. A la derecha se indica el tamaño de los ARNs utilizados como marcadores.

25

30

35

40

45

50

55

60

65

REIVINDICACIONES

1. Procedimiento de identificación de la proteína humana ADAMTS-13 **caracterizado** porque comprende las siguientes etapas:

- a) Comparar la secuencia de nucleótidos de regiones conservadas en proteínas ADAMTS con las secuencias parciales de nucleótidos presentes en las bases de datos de genes expresados.
- b) Identificar fragmentos homólogos y amplificarlos mediante PCR de RNA total de tejidos humanos en los que se puedan expresar dichas secuencias génicas.
- c) Utilizar los fragmentos amplificados como sondas para hibridar genotecas de ADNc humano o como moldes informativos para extender la secuencia hacia los extremos 5' o 3'.
- d) Aislar los clones de ADNc obtenidos y determinar su secuencia completa de nucleótidos.

2. Procedimiento de identificación de acuerdo con la reivindicación 1 **caracterizado** porque la secuencia génica identificada codifica una proteína humana denominada ADAMTS-13.

3. Procedimiento de identificación de acuerdo con cualquiera de las reivindicaciones anteriores **ca-**

racterizado porque la secuencia génica identificada y su secuencia de aminoácidos deducida son SEQ ID NO :1.

4. Secuencia génica de SEQ ID NO :1 y sus polimorfismos, transcritos alternativos, mutaciones, derivados o secuencias parciales, que codifiquen un enzima con actividad proteolítica o de regulación de procesos de adhesión celular, homeostasis y angiogénesis.

5. Utilización de la secuencia SEQ ID NO :1 en el diseño de inhibidores de la actividad de la ADAMTS-13.

6. Utilización de la secuencia SEQ ID NO :1 en la producción de proteínas recombinantes o sintéticas.

7. Utilización de la secuencia SEQ ID NO :1 en la producción de anticuerpos.

8. Utilización de la secuencia SEQ ID NO :1 en la producción de sistemas de detección de proteínas con alguna de las actividades descritas para las ADAMTS-s y/o de los genes que codifican para las mismas.

9. Utilización de la secuencia SEQ ID NO : 1 en la producción de composiciones activas en el tratamiento de procesos patológicos mediados por ADAMTS- s, y/o por genes que codifican para las mismas.

10. Secuencia de aminoácidos completa o partes de la misma, reflejadas en SEQ ID NO:1.

30

35

40

45

50

55

60

65

FIGURA 1

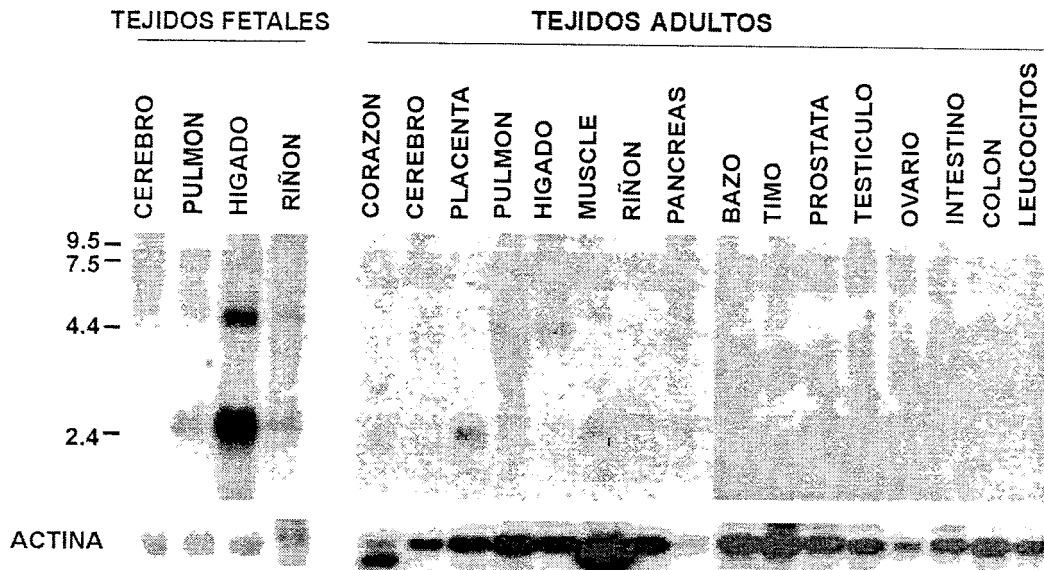
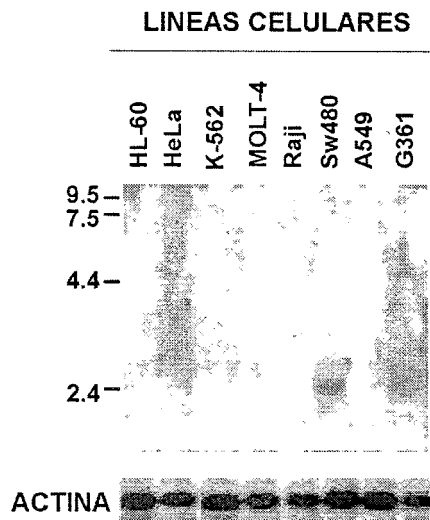


FIGURA 2



ES 2 212 695 A1

LISTA DE SECUENCIAS

INFORMACIÓN GENERAL:

5 SOLICITANTE:

NOMBRE: Universidad de Oviedo
CALLE: San Francisco, 3
CIUDAD: Oviedo
10 PAÍS: España
CODIGO PÓSTAL: 33003
TELÉFONO: 34 (9)8 510 4058
FACSIMIL: 34 (9)8 522 7126

15 TÍTULO DE LA INVENCION:

Procedimiento de identificación de la proteína humana ADAMTS-13

20 NÚMERO DE SECUENCIAS: 1

FORMA LEGIBLE POR ORDENADOR:

25 TIPO DE MEDIO: Disco flexible
ORDENADOR: PC IBM compatible
SISTEMA OPERATIVO: Windows 97
SOPORTE LÓGICO: Microsoft Word 7.0

DATOS DE LA SOLICITUD ACTUAL

30 NÚMERO DE LA SOLICITUD:

DATOS DE LA SOLICITUD ANTERIOR

35 NÚMERO DE LA SOLICITUD:
FECHA DE PRESENTACIÓN:

INFORMACIÓN CONCERNIENTE A SEQ ID NO: 1

40 CARACTERÍSTICAS DE LA SECUENCIA:

LONGITUD: 4076
TIPO: ácido nucleico
NÚMERO DE HEBRAS: doble
45 CONFIGURACIÓN: lineal

TIPO DE MOLÉCULA: ADNc a ARNm
FUENTE DE ORIGEN:

50 ORGANISMO: Homo Sapiens
TIPO DE CELULA:

FUENTE INMEDIATA:

55 GENOTECA: hígado fetal
CLON:

CARACTERÍSTICA:

60 NOMBRE/CLAVE: codón de iniciación
LOCALIZACIÓN: 21..24

CARACTERÍSTICA:

65 NOMBRE/CLAVE: secuencia codificante
LOCALIZACIÓN: 21..4040

ES 2 212 695 A1

CARACTERÍSTICA:

NOMBRE/CLAVE: codón de parada
LOCALIZACIÓN: 4041..4043

5 DESCRIPCIÓN DE LA SECUENCIA: SEQ ID NO : 1 (Depositada en el GeneBank-Database
el 6 de Febrero de 2001, con número de acceso AJ305314)

10 TCGGGCTGGC TCTCCTGAGG ATG CAC CAG CGT CAC CCC CGG GCA
44

15 Met His Gln Arg His Pro Arg Ala
1 5

20 AGA TGC CCT CCC CTC TGT GTG GCC GGA ATC CTT GCC TGT GGC TTT CTC
92
Arg Cys Pro Pro Leu Cys Val Ala Gly Ile Leu Ala Cys Gly Phe Leu
10 15 20

25 CTG GGC TGC TGG GGA CCC TCC CAT TTC CAG CAG AGT TGT CTT CAG GCT
140
Leu Gly Cys Trp Gly Pro Ser His Phe Gln Gln Ser Cys Leu Gln Ala
25 30 35 40

35 TTG GAG CCA CAG GCC GTG TCT TCT TAC TTG AGC CCT GGT GCT CCC TTA
188
Leu Glu Pro Gln Ala Val Ser Ser Tyr Leu Ser Pro Gly Ala Pro Leu
45 50 55

40 AAA GGC CGC CCT CCT TCC CCT GGC TTC CAG AGG CAG AGG CAG AGG CAG
236
Lys Gly Arg Pro Pro Ser Pro Gly Phe Gln Arg Gln Arg Gln Arg Gln
60 65 70

50 AGG CGG GCT GCA GGC GGC ATC CTA CAC CTG GAG CTG CTG GTG GCC GTG
284
Arg Arg Ala Ala Gly Gly Ile Leu His Leu Glu Leu Leu Val Ala Val
75 80 85

55 GGC CCC GAT GTC TTC CAG GCT CAC CAG GAG GAC ACA GAG CGC TAT GTG
332
Gly Pro Asp Val Phe Gln Ala His Gln Glu Asp Thr Glu Arg Tyr Val
90 95 100

65

ES 2 212 695 A1

CTC ACC AAC CTC AAC ATC GGG GCA GAA CTG CTT CGG GAC CCG TCC CTG
 380
 Leu Thr Asn Leu Asn Ile Gly Ala Glu Leu Leu Arg Asp Pro Ser Leu
 5 105 110 115 120
 GGG GCT CAG TTT CGG GTG CAC CTG GTG AAG ATG GTC ATT CTG ACA GAG
 10 428
 Gly Ala Gln Phe Arg Val His Leu Val Lys Met Val Ile Leu Thr Glu
 15 125 130 135
 CCT GAG GGT GCC CCA AAT ATC ACA GCC AAC CTC ACC TCG TCC CTG CTG
 20 476
 Pro Glu Gly Ala Pro Asn Ile Thr Ala Asn Leu Thr Ser Ser Leu Leu
 140 145 150
 AGC GTC TGT GGG TGG AGC CAG ACC ATC AAC CCT GAG GAC GAC ACG GAT
 25 524
 Ser Val Cys Gly Trp Ser Gln Thr Ile Asn Pro Glu Asp Asp Thr Asp
 30 155 160 165
 CCT GGC CAT GCT GAC CTG GTC CTC TAT ATC ACT AGG TTT GAC CTG GAG
 35 572
 Pro Gly His Ala Asp Leu Val Leu Tyr Ile Thr Arg Phe Asp Leu Glu
 170 175 180
 TTG CCT GAT GGT AAC CGG CAG GTG CGG GGC GTC ACC CAG CTG GGC GGT
 40 620
 Leu Pro Asp Gly Asn Arg Gln Val Arg Gly Val Thr Gln Leu Gly Gly
 45 185 190 195 200
 GCC TGC TCC CCA ACC TGG AGC TGC CTC ATT ACC GAG GAC ACT GGC TTC
 50 668
 Ala Cys Ser Pro Thr Trp Ser Cys Leu Ile Thr Glu Asp Thr Gly Phe
 55 205 210 215
 GAC CTG GGA GTC ACC ATT GCC CAT GAG ATT GGG CAC AGC TTC GGC CTG
 60 716
 Asp Leu Gly Val Thr Ile Ala His Glu Ile Gly His Ser Phe Gly Leu
 220 225 230
 GAG CAC GAC GGC GCG CCC GGC AGC GGC TGC GGC CCC AGC GGA CAC GTG
 65

ES 2 212 695 A1

764
 Glu His Asp Gly Ala Pro Gly Ser Gly Cys Gly Pro Ser Gly His Val
 235 240 245

5
 ATG GCT TCG GAC GGC GCC GCG CCC CGC GCC GGC CTC GCC TGG TCC CCC
 812

10
 Met Ala Ser Asp Gly Ala Ala Pro Arg Ala Gly Leu Ala Trp Ser Pro
 250 255 260

15
 TGC AGC CGC CGG CAG CTG CTG AGC CTG CTC AGC GCC AAC GAG CAG TGC
 860

20
 Cys Ser Arg Arg Gln Leu Leu Ser Leu Leu Ser Ala Asn Glu Gln Cys
 265 270 275 280

25
 CGC GTG GCC TTC GGC CCC AAG GCT GTC GCC TGC ACC TTC GCC AGG GAG
 908

30
 Arg Val Ala Phe Gly Pro Lys Ala Val Ala Cys Thr Phe Ala Arg Glu
 285 290 295

35
 CAC CTG GAT ATG TGC CAG GCC CTC TCC TGC CAC ACA GAC CCG CTG GAC
 956

40
 His Leu Asp Met Cys Gln Ala Leu Ser Cys His Thr Asp Pro Leu Asp
 300 305 310

45
 CAA AGC AGC TGC AGC CGC CTC CTC GTT CCT CTC CTG GAT GGG ACA GAA
 1004

50
 Gln Ser Ser Cys Ser Arg Leu Leu Val Pro Leu Leu Asp Gly Thr Glu
 315 320 325

55
 TGT GGC GTG GAG AAG TGG TGC TCC AAG GGT CGC TGC CGC TCC CTG GTG
 1052

60
 Cys Gly Val Glu Lys Trp Cys Ser Lys Gly Arg Cys Arg Ser Leu Val
 330 335 340

65
 GAG CTG ACC CCC ATA GCA GCA GTG CAT GGG CGC TGG TCT AGC TGG GGT
 1100

70
 Glu Leu Thr Pro Ile Ala Ala Val His Gly Arg Trp Ser Ser Trp Gly
 345 350 355 360

75
 CCC CGA AGT CCT TGC TCC CGC TCC TGC GGA GGA GGT GTG GTC ACC AGG
 1148

ES 2 212 695 A1

		Pro Arg Ser	Pro Cys Ser Arg Ser Cys	Gly Gly Gly	Val Val Thr Arg
		365		370	375
5					
	1196	AGG CGG CAG TGC AAC AAC CCC AGA CCT GCC TTT GGG GGG CGT GCA TGT			
10		Arg Arg Gln Cys Asn Asn Pro Arg Pro Ala Phe Gly Gly Arg Ala Cys			
		380		385	390
15		GTT GGT GCT GAC CTC CAG GCC GAG ATG TGC AAC ACT CAG GCC TGC GAG			
	1244	Val Gly Ala Asp Leu Gln Ala Glu Met Cys Asn Thr Gln Ala Cys Glu			
20		395		400	405
		AAG ACC CAG CTG GAG TTC ATG TCG CAA CAG TGC GCC AGG ACC GAC GGC			
25	1292	Lys Thr Gln Leu Glu Phe Met Ser Gln Gln Cys Ala Arg Thr Asp Gly			
		410		415	420
30		CAG CCG CTG CGC TCC TCC CCT GGC GGC GCC TCC TTC TAC CAC TGG GGT			
	1340	Gln Pro Leu Arg Ser Ser Pro Gly Gly Ala Ser Phe Tyr His Trp Gly			
35		425		430	435
		440			
40	1388	GCT GCT GTA CCA CAC AGC CAA GGG GAT GCT CTG TGC AGA CAC ATG TGC			
		Ala Ala Val Pro His Ser Gln Gly Asp Ala Leu Cys Arg His Met Cys			
45		445		450	455
		CGG GCC ATT GGC GAG AGC TTC ATC ATG AAG CGT GGA GAC AGC TTC CTC			
50	1436	Arg Ala Ile Gly Glu Ser Phe Ile Met Lys Arg Gly Asp Ser Phe Leu			
		460		465	470
55		GAT GGG ACC CGG TGT ATG CCA AGT GGC CCC CGG GAG GAC GGG ACC CTG			
	1484	Asp Gly Thr Arg Cys Met Pro Ser Gly Pro Arg Glu Asp Gly Thr Leu			
60		475		480	485
		AGC CTG TGT GTG TCG GGC AGC TGC AGG ACA TTT GGC TGT GAT GGT AGG			
65	1532				

ES 2 212 695 A1

		Ser Leu Cys Val Ser Gly Ser Cys Arg Thr Phe Gly Cys Asp Gly Arg			
		490	495	500	
5		ATG GAC TCC CAG CAG GTA TGG GAC AGG TGC CAG GTG TGT GGT GGG GAC			
	1580				
10		Met Asp Ser Gln Gln Val Trp Asp Arg Cys Gln Val Cys Gly Gly Asp			
		505	510	515	520
15		AAC AGC ACG TGC AGC CCA CGG AAG GGC TCT TTC ACA GCT GGC AGA GCG			
	1628				
20		Asn Ser Thr Cys Ser Pro Arg Lys Gly Ser Phe Thr Ala Gly Arg Ala			
		525	530	535	
25		AGA GAA TAT GTC ACA TTT CTG ACA GTT ACC CCC AAC CTG ACC AGT GTC			
	1676				
30		Arg Glu Tyr Val Thr Phe Leu Thr Val Thr Pro Asn Leu Thr Ser Val			
		540	545	550	
35		TAC ATT GCC AAC CAC AGG CCT CTC TTC ACA CAC TTG GCG GTG AGG ATC			
	1724				
40		Tyr Ile Ala Asn His Arg Pro Leu Phe Thr His Leu Ala Val Arg Ile			
		555	560	565	
45		GGA GGG CGC TAT GTC GTG GCT GGG AAG ATG AGC ATC TCC CCT AAC ACC			
	1772				
50		Gly Gly Arg Tyr Val Val Ala Gly Lys Met Ser Ile Ser Pro Asn Thr			
		570	575	580	
55		ACC TAC CCC TCC CTC CTG GAG GAT GGT CGT GTC GAG TAC AGA GTG GCC			
	1820				
60		Thr Tyr Pro Ser Leu Leu Glu Asp Gly Arg Val Glu Tyr Arg Val Ala			
		585	590	595	600
65		CTC ACC GAG GAC CGG CTG CCC CGC CTG GAG GAG ATC CGC ATC TGG GGA			
	1868				
70		Leu Thr Glu Asp Arg Leu Pro Arg Leu Glu Glu Ile Arg Ile Trp Gly			
		605	610	615	
75		CCC CTC CAG GAA GAT GCT GAC ATC CAG GTT TAC AGG CGG TAT GGC GAG			
	1916				

ES 2 212 695 A1

Pro Leu Gln Glu Asp Ala Asp Ile Gln Val Tyr Arg Arg Tyr Gly Glu
 620 625 630

5

GAG TAT GGC AAC CTC ACC CGC CCA GAC ATC ACC TTC ACC TAC TTC CAG
 1964

Glu Tyr Gly Asn Leu Thr Arg Pro Asp Ile Thr Phe Thr Tyr Phe Gln
 635 640 645

10

CCT AAG CCA CGG CAG GCC TGG GTG TGG GCC GCT GTG CGT GGG CCC TGC
 15 2012

Pro Lys Pro Arg Gln Ala Trp Val Trp Ala Ala Val Arg Gly Pro Cys
 650 655 660

20

TCG GTG AGC TGT GGG GCA GGG CTG CGC TGG GTA AAC TAC AGC TGC CTG
 2060

Ser Val Ser Cys Gly Ala Gly Leu Arg Trp Val Asn Tyr Ser Cys Leu
 665 670 675 680

25

GAC CAG GCC AGG AAG GAG TTG GTG GAG ACT GTC CAG TGC CAA GGG AGC
 30 2108

Asp Gln Ala Arg Lys Glu Leu Val Glu Thr Val Gln Cys Gln Gly Ser
 685 690 695

35

CAG CAG CCA CCA GCG TGG CCA GAG GCC TGC GTG CTC GAA CCC TGC CCT
 2156

Gln Gln Pro Pro Ala Trp Pro Glu Ala Cys Val Leu Glu Pro Cys Pro
 700 705 710

40

CCC TAC TGG GCG GTG GGA GAC TTC GGC CCA TGC AGC GCC TCC TGT GGG
 45 2204

Pro Tyr Trp Ala Val Gly Asp Phe Gly Pro Cys Ser Ala Ser Cys Gly
 715 720 725

50

GGT GGC CTG CGG GAG CGG CCA GTG CGC TGC GTG GAG GCC CAG GGC AGC
 2252

Gly Gly Leu Arg Glu Arg Pro Val Arg Cys Val Glu Ala Gln Gly Ser
 730 735 740

55

CTT CTG AAG ACA TTG CCC CCA GCC CGG TGC AGA GCA GGG GCC CAG CAG
 60 2300

Leu Leu Lys Thr Leu Pro Pro Ala Arg Cys Arg Ala Gly Ala Gln Gln
 745 750 755 760

65

ES 2 212 695 A1

CCA GCT GTG GCG CTG GAA ACC TGC AAC CCC CAG CCC TGC CCT GCC AGG
 2348
 5 Pro Ala Val Ala Leu Glu Thr Cys Asn Pro Gln Pro Cys Pro Ala Arg
 765 770 775
 TGG GAG GTG TCA GAG CCC AGC TCA TGC ACA TCA GCT GGT GGA GCA GGC
 10 2396
 Trp Glu Val Ser Glu Pro Ser Ser Cys Thr Ser Ala Gly Gly Ala Gly
 15 780 785 790
 CTG GCC TTG GAG AAC GAG ACC TGT GTG CCA GGG GCA GAT GGC CTG GAG
 20 2444
 Leu Ala Leu Glu Asn Glu Thr Cys Val Pro Gly Ala Asp Gly Leu Glu
 795 800 805
 GCT CCA GTG ACT GAG GGG CCT GGC TCC GTA GAT GAG AAG CTG CCT GCC
 25 2492
 Ala Pro Val Thr Glu Gly Pro Gly Ser Val Asp Glu Lys Leu Pro Ala
 30 810 815 820
 CCT GAG CCC TGT GTC GGG ATG TCA TGT CCT CCA GGC TGG GGC CAT CTG
 35 2540
 Pro Glu Pro Cys Val Gly Met Ser Cys Pro Pro Gly Trp Gly His Leu
 825 830 835 840
 GAT GCC ACC TCT GCA GGG GAG AAG GCT CCC TCC CCA TGG GGC AGC ATC
 40 2588
 Asp Ala Thr Ser Ala Gly Glu Lys Ala Pro Ser Pro Trp Gly Ser Ile
 45 845 850 855
 AGG ACG GGG GCT CAA GCT GCA CAC GTG TGG ACC CCT GCG GCA GGG TCG
 50 2636
 Arg Thr Gly Ala Gln Ala Ala His Val Trp Thr Pro Ala Ala Gly Ser
 860 865 870
 TGC TCC GTC TCC TGC GGG CGA GGT CTG ATG GAG CTG CGT TTC CTG TGC
 55 2684
 Cys Ser Val Ser Cys Gly Arg Gly Leu Met Glu Leu Arg Phe Leu Cys
 60 875 880 885
 ATG GAC TCT GCC CTC AGG GTG CCT GTC CAG GAA GAG CTG TGT GGC CTG
 65 2732

ES 2 212 695 A1

		Met	Asp	Ser	Ala	Leu	Arg	Val	Pro	Val	Gln	Glu	Glu	Leu	Cys	Gly	Leu	
			890					895						900				
5																		
	2780	GCA	AGC	AAG	CCT	GGG	AGC	CGG	CGG	GAG	GTC	TGC	CAG	GCT	GTC	CCG	TGC	
10		Ala	Ser	Lys	Pro	Gly	Ser	Arg	Arg	Glu	Val	Cys	Gln	Ala	Val	Pro	Cys	
		905					910					915					920	
15																		
	2828	CCT	GCT	CGG	TGG	CAG	TAC	AAG	CTG	GCG	GCC	TGC	AGC	GTG	AGC	TGT	GGG	
20		Pro	Ala	Arg	Trp	Gln	Tyr	Lys	Leu	Ala	Ala	Cys	Ser	Val	Ser	Cys	Gly	
						925						930				935		
25																		
	2876	AGA	GGG	GTC	GTG	CGG	AGG	ATC	CTG	TAT	TGT	GCC	CGG	GCC	CAT	GGG	GAG	
30		Arg	Gly	Val	Val	Arg	Arg	Ile	Leu	Tyr	Cys	Ala	Arg	Ala	His	Gly	Glu	
					940					945					950			
35																		
	2924	GAC	GAT	GGT	GAG	GAG	ATC	CTG	TTG	GAC	ACC	CAG	TGC	CAG	GGG	CTG	CCT	
40		Asp	Asp	Gly	Glu	Glu	Ile	Leu	Leu	Asp	Thr	Gln	Cys	Gln	Gly	Leu	Pro	
				955						960				965				
45																		
	2972	CGC	CCG	GAA	CCC	CAG	GAG	GCC	TGC	AGC	CTG	GAG	CCC	TGC	CCA	CCT	AGG	
50		Arg	Pro	Glu	Pro	Gln	Glu	Ala	Cys	Ser	Leu	Glu	Pro	Cys	Pro	Pro	Arg	
				970						975				980				
55																		
	3020	TGG	AAA	GTC	ATG	TCC	CTT	GGC	CCA	TGT	TCG	GCC	AGC	TGT	GGC	CTT	GGC	
60		Trp	Lys	Val	Met	Ser	Leu	Gly	Pro	Cys	Ser	Ala	Ser	Cys	Gly	Leu	Gly	
		985					990					995					1000	
65																		
	3068	ACT	GCT	AGA	CGC	TCG	GTG	GCC	TGT	GTG	CAG	CTC	GAC	CAA	GGC	CAG	GAC	
70		Thr	Ala	Arg	Arg	Ser	Val	Ala	Cys	Val	Gln	Leu	Asp	Gln	Gly	Gln	Asp	
						1005						1010				1015		
75																		
	3116	GTG	GAG	GTG	GAC	GAG	GCG	GCC	TGT	GCG	GCG	CTG	GTG	CGG	CCC	GAG	GCC	
80		Val	Glu	Val	Asp	Glu	Ala	Ala	Cys	Ala	Ala	Leu	Val	Arg	Pro	Glu	Ala	
						1020						1025				1030		

ES 2 212 695 A1

3164 AGT GTC CCC TGT CTC ATT GCC GAC TGC ACC TAC CGC TGG CAT GTT GGC
 5 Ser Val Pro Cys Leu Ile Ala Asp Cys Thr Tyr Arg Trp His Val Gly
 1035 1040 1045

10 ACC TGG ATG GAG TGC TCT GTT TCC TGT GGG GAT GGC ATC CAG CGC CGG
 3212
 15 Thr Trp Met Glu Cys Ser Val Ser Cys Gly Asp Gly Ile Gln Arg Arg
 1050 1055 1060

20 CGT GAC ACC TGC CTC GGA CCC CAG GCC CAG GCG CCT GTG CCA GCT GAT
 3260
 Arg Asp Thr Cys Leu Gly Pro Gln Ala Gln Ala Pro Val Pro Ala Asp
 1065 1070 1075 1080

25 TTC TGC CAG CAC TTG CCC AAG CCG GTG ACT GTG CGT GGC TGC TGG GCT
 3308
 30 Phe Cys Gln His Leu Pro Lys Pro Val Thr Val Arg Gly Cys Trp Ala
 1085 1090 1095

35 GGG CCC TGT GTG GGA CAG GGT GCC TGT GGC AGG CAG CAC CTT GAG CCA
 3356
 Gly Pro Cys Val Gly Gln Gly Ala Cys Gly Arg Gln His Leu Glu Pro
 1100 1105 1110

40 ACA GGA ACC ATT GAC ATG CGA GGC CCA GGG CAG GCA GAC TGT GCA GTG
 3404
 45 Thr Gly Thr Ile Asp Met Arg Gly Pro Gly Gln Ala Asp Cys Ala Val
 1115 1120 1125

50 GCC ATT GGG CGG CCC CTC GGG GAG GTG GTG ACC CTC CGC GTC CTT GAG
 3452
 55 Ala Ile Gly Arg Pro Leu Gly Glu Val Val Thr Leu Arg Val Leu Glu
 1130 1135 1140

60 AGT TCT CTC AAC TGC AGT GCG GGG GAC ATG TTG CTG CTT TGG GGC CGG
 3500
 Ser Ser Leu Asn Cys Ser Ala Gly Asp Met Leu Leu Leu Trp Gly Arg
 1145 1150 1155 1160

65

ES 2 212 695 A1

CTC ACC TGG AGG AAG ATG TGC AGG AAG CTG TTG GAC ATG ACT TTC AGC
 3548
 5 Leu Thr Trp Arg Lys Met Cys Arg Lys Leu Leu Asp Met Thr Phe Ser
 1165 1170 1175

 TCC AAG ACC AAC ACG CTG GTG GTG AGG CAG CGC TGC GGG CGG CCA GGA
 10 3596
 Ser Lys Thr Asn Thr Leu Val Val Arg Gln Arg Cys Gly Arg Pro Gly
 1180 1185 1190
 15

 GGT GGG GTG CTG CTG CGG TAT GGG AGC CAG CTT GCT CCT GAA ACC TTC
 20 3644
 Gly Gly Val Leu Leu Arg Tyr Gly Ser Gln Leu Ala Pro Glu Thr Phe
 1195 1200 1205
 25

 TAC AGA GAA TGT GAC ATG CAG CTC TTT GGG CCC TGG GGT GAA ATC GTG
 30 3692
 Tyr Arg Glu Cys Asp Met Gln Leu Phe Gly Pro Trp Gly Glu Ile Val
 1210 1215 1220
 35

 AGC CCC TCG CTG AGT CCA GCC ACG AGT AAT GCA GGG GGC TGC CGG CTC
 35 3740
 Ser Pro Ser Leu Ser Pro Ala Thr Ser Asn Ala Gly Gly Cys Arg Leu
 1225 1230 1235 1240
 40

 TTC ATT AAT GTG GCT CCG CAC GCA CGG ATT GCC ATC CAT GCC CTG GCC
 45 3788
 Phe Ile Asn Val Ala Pro His Ala Arg Ile Ala Ile His Ala Leu Ala
 1245 1250 1255
 50

 ACC AAC ATG GGC GCT GGG ACC GAG GGA GCC AAT GCC AGC TAC ATC TTG
 50 3836
 Thr Asn Met Gly Ala Gly Thr Glu Gly Ala Asn Ala Ser Tyr Ile Leu
 1260 1265 1270
 55

 ATC CGG GAC ACC CAC AGC TTG AGG ACC ACA GCG TTC CAT GGG CAG CAG
 60 3884
 Ile Arg Asp Thr His Ser Leu Arg Thr Thr Ala Phe His Gly Gln Gln
 1275 1280 1285
 65

ES 2 212 695 A1

3932 GTG CTC TAC TGG GAG TCA GAG AGC AGC CAG GCT GAG ATG GAG TTC AGC
 5 Val Leu Tyr Trp Glu Ser Glu Ser Ser Gln Ala Glu Met Glu Phe Ser
 1290 1295 1300

10 3980 GAG GGC TTC CTG AAG GCT CAG GCC AGC CTG CGG GGC CAG TAC TGG ACC
 15 Glu Gly Phe Leu Lys Ala Gln Ala Ser Leu Arg Gly Gln Tyr Trp Thr
 1305 1310 1315 1320

20 4028 CTC CAA TCA TGG GTA CCG GAG ATG CAG GAC CCT CAG TCC TGG AAG GGA
 25 Leu Gln Ser Trp Val Pro Glu Met Gln Asp Pro Gln Ser Trp Lys Gly
 1325 1330 1335

30 AAG GAA GGA ACC TGAGGGTCAT TGAACATTTG TTCCGTGTCT GGCCAG
 Lys Glu Gly Thr
 1340

30 Para la secuencia descrita, se han encontrado varios transcritos alternativos como el que presenta la inserción de 93 nucleótidos en la posición 842

35 AGC GCA GGA CGG GCG CGC TGC GTG TGG GAC CCG CCG CGG CCT CAA CCC
 Ser Ala Gly Arg Ala Arg Cys Val Trp Asp Pro Pro Arg Pro Gln Pro

40 GGG TCC GCG GGG CAC CCG CCG GAT GCG CAG CCT GGC CTC TAC TAC
 Gly Ser Ala Gly His Pro Pro Asp Ala Gln Pro Gly Leu Tyr Tyr

que se traduciría a una proteína de 1371 residuos. También se ha determinado la existencia de transcritos que se traducirían a proteínas truncadas por la presencia de un codón de terminación (TGA) en posición 843..84

45

50

55

60

65



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA

① ES 2 212 695

② Nº de solicitud: 200102090

③ Fecha de presentación de la solicitud: **14.09.2001**

④ Fecha de prioridad:

INFORME SOBRE EL ESTADO DE LA TÉCNICA

⑤ **Int. Cl.7:** C12N 9/64, C12Q 1/37, C07K 16/40, A61K 38/48

DOCUMENTOS RELEVANTES

Categoría	Documentos citados	Reivindicaciones afectadas
X	WO 0053774 A2 (NEUROCRINE BIOSCIENCE) 14.09.2000, descripción; reivindicaciones.	1-10
X	EP 1004647 A (KUREHA CHEMICAL IND) 31.05.2000, páginas 16-20; reivindicaciones.	1-10
X	WO 0134785 A1 (YAMANOUCHI PHAR. CO. LTD) 17.05.2001	1-10

Categoría de los documentos citados

X: de particular relevancia

Y: de particular relevancia combinado con otro/s de la misma categoría

A: refleja el estado de la técnica

O: referido a divulgación no escrita

P: publicado entre la fecha de prioridad y la de presentación de la solicitud

E: documento anterior, pero publicado después de la fecha de presentación de la solicitud

El presente informe ha sido realizado

para todas las reivindicaciones

para las reivindicaciones nº:

Fecha de realización del informe

06.02.2004

Examinador

J. Manso Tomico

Página

1/1