



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



① Número de publicación: **2 211 256**

② Número de solicitud: 200102192

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: **28.09.2001**

⑬ Fecha de publicación de la solicitud: **01.07.2004**

⑭ Fecha de publicación del folleto de la solicitud:  
**01.07.2004**

⑰ Solicitante/s: **Universidad de Oviedo.**  
**Plaza del Riego, 4 Edf. Histórico**  
**33003 Oviedo, 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**

⑳ Título: **Procedimiento de identificación de la proteína humana ADAMTS-19.**

㉑ Resumen:

Procedimiento de identificación de la proteína humana ADAMTS-19.

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-19. 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 2 211 256 A1

## DESCRIPCIÓN

Procedimiento de identificación de la proteína humana ADAMTS-19.

### 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-19, 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- $\alpha$  (ADAM-12) se ha implicado en fusión de mioblastos. Las meltrinas  $\alpha$  y  $\beta$  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 miembro de esta fami-

lia, denominado ADAMTS-1, se identificó como consecuencia de su asociación con el desarrollo de caque-  
xia 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 agreca-  
nasa 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 ADAM-TS12 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 de nucleótidos 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-19.

Un segundo objeto de la invención es analizar la expresión en tejidos humanos del gen de la ADAMTS-19.

Un tercer objeto de la invención es analizar la expresión del gen de la ADAMTS-19 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 en ADN genómico humano secuencias que podrían corresponder a regiones metaloproteasa y disintegrina de una nueva ADAMTS human. Para llevar a cabo su amplificación se sintetizaron dos oligonucleótidos, AD-1 (5'-ACCTCCTCCACAAGTGGCATC-3') y AD-2 (5'-GCTTACTTGAGTGGGAATGTGTAG-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 pulmón 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 40 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 se verificó mediante su clonación en el vector pUC18 y posterior secuenciación de nucleótidos mediante técnicas estándar de Biología Molecular. La traducción conceptual del fragmento clonado indicó que se trataba de un nuevo miembro de la familia ADAMTS.

Con el fin de obtener una secuencia de ADNc que contuviera la información codificante de la proteína completa, a partir del producto de PCR obtenido con los oligonucleótidos AD-1 y AD-2 se llevó a cabo la extensión de sus extremos 5' y 3' mediante técnicas de amplificación rápida de extremos de ADNcs utilizando ARN de pulmón 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 ADTS19F (5'-ATGCGCCTGACTCACATCTGC-3') y ADTS19R (5'-TTATAAATAATGTAATTGCCA-3'). El análisis informático de la secuencia obtenida reveló la existencia de una fase abierta de lectura, que codifica una proteína de 1224 aminoácidos a la que denominamos ADAMTS-19. 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 familia de las ADAMTS (J. Biol. Chem. 274, 25555, (1999)). Así, la proteína presenta todos los motivos característicos de estas enzimas incluyendo la secuencia señal, el propéptido, los dominios metaloproteasa, disintegrina y rico en cisteína, así como diversas repeticiones tipo trombospondina

(TS).

Un análisis más detallado de la secuencia de aminoácidos deducida para la ADAMTS-19 determinó la existencia de un prodominio en el que se localiza un residuo de cisteína (posición 243) que podrían estar implicado en el mantenimiento de la latencia enzimática. Este prodominio termina en un motivo dibásico que podía corresponder al sitio de activación por furina, que poseen estas enzimas. El dominio catalítico incluye la secuencia HEXXXHXXGXXHD (posiciones 431-442) 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 445) que contribuye a formar la estructura Met-giro presente en reprotinas y MMPs. Tras el dominio catalítico puede reconocerse el dominio disintegrina, 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-19. La secuencia fue depositada en el banco de datos EMBL con el número de acceso AJ311904. 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-19. Con este fin, se realizaron reacciones de amplificación mediante técnicas de PCR de ADNc de diversas genotecas de tejidos humanos adultos (próstata, cerebros, mama, glándula submaxilar, endotelio, placenta, hígado, aorta, ovario) y fetales (corazón, pulmón, hígado y riñón). Para ellos se utilizaron 20 pmoles los oligonucleótidos específicos AD-1 y AD2. 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 40 ciclos de desnaturalización (15 s, 94°C), hibridación (20 s, 64°C) y extensión (20 s, 72°C). Como puede observarse en la figura 1, tras hibridación con la sonda de ADAMTS-19, se detectó un producto de amplificación en la genoteca de ADNc de pulmón fetal. La confirmación de que se trataba de ADAMTS-19 se hizo mediante la secuenciación directa del producto de amplificación y posterior traducción conceptual de la secuencia obtenida.

El tercer objeto de la invención consistió en el estudio de la expresión del gen de la ADAMTS-19 en muestras obtenidas de tumores humanos. Se realizó de forma similar a la anterior, utilizando ADNc de genotecas de carcinoma mamario y de osteosarcoma. En la figura 1 se muestra la amplificación del producto esperado en la genoteca de osteosarcoma.

### Descripción de las figuras

Figura 1. Análisis de la expresión de ADAMTS-19 en las diversas genotecas de ADNc analizadas.

### REIVINDICACIONES

1. Procedimiento de identificación de la proteína humana ADAMTS-19 **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-19.

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

**rizado** 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-19.

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

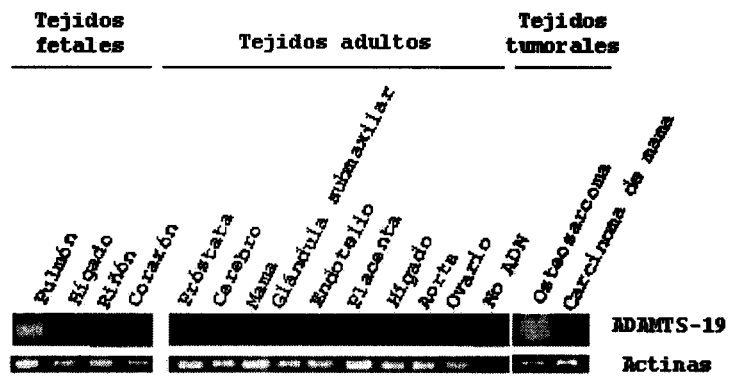
50

55

60

65

FIGURA 1



# ES 2 211 256 A1

## LISTA DE SECUENCIAS

### INFORMACIÓN GENERAL:

5 SOLICITANTE:

NOMBRE: Universidad de Oviedo  
CALLE: San Francisco, 3  
CIUDAD: Oviedo  
10 PAIS: España  
CÓDIGO POSTAL: 33003  
TELÉFONO: 34 (9)8 510 4058  
FACSÍMIL: 34 (9)8 522 7126

15 TITULO DE LA INVENCION:

Procedimiento de identificación de la proteína humana ADAMTS-19

NÚMERO DE SECUENCIAS: 1

20 FORMA LEGIBLE POR ORDENADOR:

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

DATOS DE LA SOLICITUD ACTUAL

NÚMERO DE LA SOLICITUD:

30

DATOS DE LA SOLICITUD ANTERIOR

NÚMERO DE LA SOLICITUD:  
FECHA DE PRESENTACIÓN:

35

INFORMACIÓN CONCERNIENTE A SEQ ID NO: 1

CARACTERÍSTICAS DE LA SECUENCIA:

40

LONGITUD: 3696  
TIPO: ácido nucleico  
NÚMERO DE HEBRAS: doble  
CONFIGURACIÓN: lineal

45

TIPO DE MOLÉCULA: ADNc a ARNm

FUENTE DE ORIGEN:

50

ORGANISMO: Homo Sapiens  
TIPO DE CÉLULA:

FUENTE INMEDIATA:

55

GENOTECA: pulmón fetal  
CLON:

CARACTERÍSTICA:

60

NOMBRE/CLAVE: codón de iniciación  
LOCALIZACIÓN: 1..3

CARACTERÍSTICA:

65

NOMBRE/CLAVE: secuencia codificante  
LOCALIZACIÓN: 1..3672

## ES 2 211 256 A1

CARACTERÍSTICA:

NOMBRE/CLAVE: codón de parada  
 LOCALIZACIÓN: 3673..3675

5 DESCRIPCIÓN DE LA SECUENCIA: SEQ ID NO : 1 (Depositada en el GeneBank Database el 30 de julio de 2001, con número de acceso AJ311904)

5	ATG CGC CTG	ACT CAC ATC	TGC TGC TGC	CTC CTT TAC	CAG CTG GGG	48
10	Met Arg Leu	Thr His Ile	Cys Cys Cys	Cys Leu Leu	Tyr Gln Leu Gly	
	1	5	10	15		
15	TTC CTG TCG	AAT GGG ATC	GTT TCA GAG	CTG CAG TTC	GCC CCC GAC	96
	Phe Leu Ser	Asn Gly Ile	Val Ser Glu	Leu Gln Phe	Ala Pro Asp Arg	
		20	25	30		
20	GAG GAG TGG	GAA GTC GTG	TTT CCT GCG	CTC TGG CGC	CGG GAG CCG	144
	Glu Glu Trp	Glu Val Val	Phe Pro Ala	Leu Trp Arg	Arg Glu Pro Val	
25		35	40	45		
30	GAC CCG GCT	GGC GGC AGC	GGG GGC AGC	GCG GAC CCG	GGC TGG GTG	192
	Asp Pro Ala	Gly Gly Ser	Gly Gly Ser	Ala Asp Pro	Gly Trp Val Arg	
	50	55	60			
35	GGC GTT GGG	GGC GGC GGA	AGC GCC CGG	GCG CAG GCT	GCC GGC AGC	240
	Gly Val Gly	Gly Gly Gly	Ser Ala Arg	Ala Gln Ala	Ala Gly Ser Ser	
	65	70	75	80		
40	CGC GAG GTG	CGC TCT GTG	GCT CCG GTG	CCT TTG GAG	GAG CCC GTG	288
	Arg Glu Val	Arg Ser Val	Ala Pro Val	Pro Leu Glu	Glu Glu Pro Val	Glu
		85	90	95		
45	GGC CGA TCA	GAG TCC CGG	CTC CGG CCC	CCG CCG CCG	TCG GAG GGT	336
	Gly Arg Ser	Glu Ser Arg	Leu Arg Pro	Pro Pro Pro	Ser Glu Gly Glu	
50		100	105	110		
55	GAG GAC GAG	GAG CTC GAG	TCG CAG GAG	CTG CCG CGG	GGA TCC AGC	384
	Glu Asp Glu	Glu Leu Glu	Ser Gln Glu	Leu Pro Arg	Gly Ser Ser Gly	
	115	120	125			
60	GCT GCC GCC	TTG TCC CCG	GGC GCC CCG	GCC TCG TGG	CAG CCG CCG	432
	Ala Ala Ala	Leu Ser Pro	Gly Ala Pro	Ala Ser Trp	Gln Pro Pro Pro	
	130	135	140			
65	CCC CCG CAG	CCG CCC CCG	TCC CCG CCC	CCG GCC CAG	CAT GCC GAG	480
					CCG	



















OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA

① ES 2 211 256

② Nº de solicitud: 200102192

③ Fecha de presentación de la solicitud: **28.09.2001**

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
X	WO 0159133 A (MERCK PATENT GMBH) 16.08.2001	4,10

#### 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

03.05.2004

Examinador

J. Manso Tomico

Página

1/1