

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 531 372**

51 Int. Cl.:

C07K 1/107 (2006.01)

C12N 15/62 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **12.10.2007 E 07821275 (0)**

97 Fecha y número de publicación de la concesión europea: **10.12.2014 EP 2076527**

54 Título: **Enzimas de procesamiento fusionadas a marcadores de proteínas básicas**

30 Prioridad:

13.10.2006 EP 06122266

45 Fecha de publicación y mención en BOPI de la traducción de la patente:

13.03.2015

73 Titular/es:

**NOVO NORDISK HEALTH CARE AG (100.0%)
Thurgauerstrasse 36/38
8050 Zürich, CH**

72 Inventor/es:

SHAW, ALLAN CHRISTIAN

74 Agente/Representante:

CARPINTERO LÓPEZ, Mario

Observaciones :

Véase nota informativa (Remarks) en el folleto original publicado por la Oficina Europea de Patentes

ES 2 531 372 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

Enzimas de procesamiento fusionadas a marcadores de proteínas básicas

Campo de la invención

5 La presente invención se refiere a enzimas de procesamiento novedosas que comprenden un marcador proteico alcalino obtenido a partir de bacterias termofílicas.

Antecedentes de la invención

10 La expresión de proteínas recombinantes en células hospedadoras microbianas se utiliza extensamente para la producción industrial de proteínas terapéuticas humanas. Una pareja de fusión puede estar fijada en el extremo N-terminal (o C-terminal) de la proteína diana y sirve para aumentar el nivel de expresión, la solubilidad o el plegamiento correcto de la proteína diana y para facilitar la purificación durante el procesamiento aguas abajo. La eliminación de la pareja de fusión de la proteína diana es necesaria principalmente para conservar la actividad biológica de la proteína diana. El procesamiento de la proteína de fusión se puede realizar o bien después de la expresión o después de una o varias etapas de purificación inicial, utilizando proteasas adecuadas capaces de escindir la pareja de fusión para liberar la proteína diana. La proteína diana recombinante con o sin pareja de fusión también puede ser 15 que sea necesario modificarla después de la traducción para obtener ciertas características biológicas, usando una o varias etapas enzimáticas. Tal modificación incluye la eliminación de glicosilaciones no deseadas con azúcares de tipo manosa, fucosa o xilosa o la introducción de glicosilaciones. Las modificaciones también pueden incluir desamidaciones, amidaciones, metilaciones, fosforilaciones, desfosforilaciones, sulfataciones, acetilaciones o transaminaciones. Para disminuir el coste de las etapas de procesamiento aguas abajo es fundamental que estas enzimas de procesamiento tengan un bajo coste de producción, por ejemplo, que se puedan expresar en grandes cantidades y purificar con un número reducido de etapas cromatográficas. Además, es deseable que las enzimas de procesamiento utilizadas para la escisión de proteínas de fusión se puedan eliminar fácilmente después del procesamiento de la proteína diana. También puede ser ventajoso que las enzimas de procesamiento se puedan inmovilizar en columnas de purificación.

25 Gråslund et al. (J. Biotechnol. 96 (2002) 93-102) describen una proteasa de fusión que comprende un controlador de la purificación Zbasic fusionado con el extremo N-terminal de una polimerasa de ADN Klenow y con una proteasa 3C. El dominio Zbasic es un agrupamiento de tres hélices antiparalelas de 58 aminoácidos, desarrollado a partir de uno de los cinco dominios que constituyen la proteína A de estafilococos.

30 Bi et al. (Prot. Expr. and Purif. 47 (2006) 234-240) describen una proteína de fusión de NTL9, una secuencia de escisión del Factor Xa y un fragmento de 21 aminoácidos HP21 del subdominio helicoidal de vilina en la parte anterior (HP36), cuya proteína de fusión es útil para la purificación de proteínas pequeñas.

El documento WO 2006/108826 responde a las condiciones del Art. 54(3) y describe proteínas ribosómicas básicas como marcadores para la purificación.

35 La presente invención proporciona nuevas enzimas de procesamiento que son muy adecuadas para la modificación de proteínas diana, incluyendo la escisión de parejas de fusión de una proteína de fusión. Además, las nuevas enzimas de procesamiento se pueden separar fácilmente de la proteína diana después de la etapa de procesamiento.

Compendio de la invención

40 En un aspecto, la invención proporciona enzimas de procesamiento que comprenden un marcador fijado en el extremo N-terminal, en donde el marcador consiste en una secuencia peptídica procedente de la familia L9 de proteínas ribosómicas procedentes de bacterias termófilas, en donde el marcador tiene de 15 a 250 aminoácidos, en donde el marcador comprende al menos 15% de residuos de aminoácidos básicos, Lys y Arg, y en donde el marcador tiene un pI superior a 9.

En una realización, el marcador se fija a la proteína de procesamiento a través de una secuencia enlazadora.

En otra realización, el marcador no comprende residuos de cisteína.

45 En otra realización, el marcador comprende de 50 a 250 o de 75 a 200, de 100 a 250 o de 100 a 200 residuos de aminoácidos.

En otra realización, el marcador comprende de 20 a 50% de residuos de aminoácidos básicos.

En otra realización, el marcador se selecciona entre el grupo de secuencias peptídicas que consiste en

MKVILLRDVPGKIGKKGEIKEVSDGYARNYLIPRGFAKEYTEGLERAIKHEKEIEKRKKERERE
SEKILKELKKRTHVVKVKGEGGKIFGAVTAATVAEEISKTTGLKLDKRWFKLDKPIKELGEY
SLEVSLPGGVKDTIKIRVEREE(SEQ ID NO:10);

MKVILLEPLENLGDVGVVDVKPGYARNYLLPRGLAVLATESNLKALEARIRAQAKR
LAERKAEERLKEILENLTLTIPVRAGETKIYGSVTAKDIAEALSRQHGITIDPKRLALEKPIKEL
GEYVLTYPHPEVPIQLKVSVAQE(SEQ ID NO: 41);

MKVIFLKDVKGKGGKGEIKDVADGYANNFLFKQGLAIEATPANIKALEAQKQKEQRQAAEEL
ANAKLKEELEKLTVEIPAKAGEGGRLFGSITSKQIAEALQAQHGLKLDKRKIELADAIRSLGY
TNVPVKLHPEVTATLKVHVKEQK(SEQ ID NO: 42) y

5

MKVLLKDVS KIGKKGEIKN VSDGYARNYLIPKGLALEATPRVLKRLEAEKR-
KKEEEKIQ IKTQNEELLKMLKFLYKIP VKAGESGKLFGALTNSDIK AVEKIADVNI
DKKFIVLEKPIKEIGMYDVLVRLPEGVSGK IKVEVIQEGKN(SEQ ID NO: 43).

En otra realización, la enzima de procesamiento se selecciona entre oxidorreductasas, transferasas, hidrolasas, liasas, isomerasas y ligasas.

10 En un segundo aspecto, la presente invención proporciona una estructura artificial de ácido nucleico que codifica la enzima de procesamiento de acuerdo con la invención.

En otro aspecto, la presente invención proporciona un método para preparar una enzima de procesamiento de acuerdo con la invención, que comprende;

- i) la expresión de la enzima de procesamiento de acuerdo con la invención,
- ii) cargar dicha enzima de procesamiento expresada sobre una columna de intercambio catiónico, y
- 15 iii) eluir dicha enzima de procesamiento con un eluyente adecuado.

En una realización del método, la enzima de procesamiento es una proteasa seleccionada entre el grupo que consiste en proteasa HRV14 3C, proteasa HAV18 3C, proteasa TEV y RHDV 3C.

20 En otro aspecto, la presente invención proporciona un método para modificar un precursor de proteína diana que comprende las etapas de hacer reaccionar una enzima de procesamiento marcada en el extremo N-terminal de acuerdo con la invención, con el precursor de la proteína diana y separar la proteína diana de la enzima de procesamiento.

En una realización de este método, la enzima de procesamiento marcada en el extremo N-terminal y el precursor de la proteína diana se coexpresan en un hospedador adecuado para permitir el aislamiento de la proteína diana después de la expresión.

25 En otra realización de este método, dicho hospedador se selecciona a partir de las cepas de *Bacillus*, *Streptomyces* y *Escherichia coli*.

30 El marcador facilitará la expresión de una enzima de procesamiento marcada soluble con un rendimiento elevado. El marcador también permitirá una pureza elevada después de una etapa de purificación mediante cromatografía de intercambio catiónico aislada, así como una eliminación fácil de la enzima de procesamiento marcada después del procesamiento del precursor de la proteína diana. Los marcadores también permiten el procesamiento en la columna empleando una matriz de cromatografía de intercambio catiónico como inmovilizador para la enzima de procesamiento marcada.

Descripción de los dibujos

35 La Fig. 1: describe un mapa vectorial del vector obtenido a partir de pET11a que codifica la enzima de procesamiento pACSH238 que consiste en SEQ ID NO: 10 como marcador N-terminal para la purificación, SEQ ID NO: 28 como enlazador y SEQ ID NO: 35 como la proteasa HRV14 3C. Los sitios de clonación NdeI, XhoI y BamHI se indican, así como la secuencia que codifica lacl, la secuencia que codifica ampicilina (bla), el origen de replicación de pBR322, la región promotora de T7, la región terminadora de T7 y

40 la Fig 2 describe un espectro de masas extraído y deconvolucionado de pACSH294 incubado con pACSH239 durante una noche a 25 grados Celsius. Con una relación de 1:100 entre la enzima y el sustrato, el péptido S661 (4802,1 Da) se detecta claramente, así como el marcador liberado + enlazador (10168,4 Da). No se detectó proteína de fusión intacta, lo que indica que la proteína de fusión se había escindido completamente en la parte del marcador y el péptido S661. El eje Y indica la intensidad de los iones medidos. Eje X indica la masa molecular en Dalton (Da).

Descripción de la invención

- La presente invención proporciona nuevas enzimas de procesamiento que se pueden utilizar para la modificación de proteínas, p. ej., para escindir proteínas de fusión expresadas en células hospedadoras microbianas. Además, la presente invención proporciona un método para la expresión y purificación eficaz de tales enzimas de procesamiento marcadas.
- Las enzimas de procesamiento de acuerdo con la presente invención comprenderán marcadores fusionados en el extremo N-terminal que facilitan la expresión soluble de la enzima de procesamiento marcada con un rendimiento elevado y permiten una purificación eficaz empleando una sola etapa de cromatografía de intercambio catiónico, así como una eliminación fácil de la enzima de procesamiento marcada después del uso.
- Los marcadores incrementarán la solubilidad y de este modo también la estabilidad de la enzima de procesamiento marcada. Esto supone una ventaja ya que las enzimas poco solubles requieren mayor volumen de reacción, tiempo de incubación y condiciones de reacción más específicas.
- Los marcadores también facilitarán la unión a una columna catiónica y permitirán de este modo una escisión en la columna empleando una matriz cromatográfica de intercambio catiónico como inmovilizador para la enzima de procesamiento marcada.
- Algunos de los marcadores utilizados de acuerdo con la presente invención se describen en el documento de solicitud de patente en tramitación PCT/EP2006/061493.
- Con la expresión "una proteína altamente básica" se entiende una proteína que tiene un porcentaje elevado de los residuos de aminoácidos básicos Lys y Arg, por ejemplo, al menos aproximadamente el 15% del número total de residuos de aminoácidos en la proteína.
- Por "proteínas ribosómicas" se entiende subunidades peptídicas o polipeptídicas del ribosoma, que son las partículas que catalizan la síntesis de proteínas dirigida por el ARNm en todos los organismos. Las proteínas ribosómicas se definen basándose en su secuencia a través de secuencias distintivas ribosómicas como se describen en las bases de datos de dominios, tales como InterPro y Prosite. La ventaja de usar proteínas ribosómicas es que en su mayoría están cargadas muy positivamente.
- En el presente contexto "microorganismos termófilos" significa organismos que crecen de manera óptima desde 50°C a 100°C. Esto está en contraste con los mesófilos que en general crecen de manera óptima a temperaturas desde 30-37°C. La expresión "bacterias termófilas" incluye en este contexto también a las bacterias hipertermófilas.
- En una realización, el marcador no contiene residuos de cisteína.
- En una realización adicional, el marcador comprende de 15 a 250, de 15 a 225, de 15 a 200, de 15 a 175, de 15 a 150, de 15 a 75 o de 15 a 50 residuos de aminoácidos.
- En una realización, el marcador tiene de 15 a 250 residuos de aminoácidos.
- En una realización, el marcador tiene de 15 a 200 residuos de aminoácidos.
- En una realización, el marcador tiene de 15 a 150 residuos de aminoácidos.
- En una realización, el marcador tiene de 15 a 100 residuos de aminoácidos.
- En una realización, el marcador tiene de 15 a 50 residuos de aminoácidos.
- En una realización adicional, el marcador comprende de 20 a 120, de 20 a 100, de 20 a 90, de 20 a 75 residuos de aminoácidos o de 20 a 50 residuos de aminoácidos.
- En una realización, el marcador tiene de 20 a 120 residuos de aminoácidos.
- En una realización, el marcador tiene de 20 a 100 residuos de aminoácidos.
- En una realización, el marcador tiene de 20 a 90 residuos de ácido amino.
- En una realización, el marcador tiene de 20 a 75 residuos de ácido amino.
- En una realización, el marcador tiene de 20 a 50 residuos de ácido amino.
- En una realización, el marcador tiene de 50 a 250 residuos de aminoácidos.
- En una realización, el marcador tiene de 50 a 225 residuos de aminoácidos.
- En una realización, el marcador tiene de 50 a 200 residuos de aminoácidos.

- En una realización, el marcador tiene de 50 a 175 residuos de aminoácidos.
- En una realización, el marcador tiene de 50 a 150 residuos de aminoácidos.
- En una realización, el marcador tiene de 50 a 125 residuos de aminoácidos.
- En una realización, el marcador tiene de 50 a 100 residuos de aminoácidos.
- 5 En una realización, el marcador tiene de 50 a 75 residuos de ácido amino.
- En una realización, el marcador tiene de 75 a 250 residuos de aminoácidos.
- En una realización, el marcador tiene de 75 a 225 residuos de aminoácidos.
- En una realización, el marcador tiene de 75 a 200 residuos de aminoácidos.
- En una realización, el marcador tiene de 75 a 175 residuos de aminoácidos.
- 10 En una realización, el marcador tiene de 75 a 150 residuos de aminoácidos.
- En una realización, el marcador tiene de 75 a 125 residuos de aminoácidos.
- En una realización, el marcador tiene de 75 a 100 residuos de aminoácidos.
- En una realización, el marcador tiene de 100 a 250 residuos de aminoácidos.
- En una realización, el marcador tiene de 100 a 225 residuos de aminoácidos.
- 15 En una realización, el marcador tiene de 100 a 200 residuos de aminoácidos.
- En una realización, el marcador tiene de 100 a 175 residuos de aminoácidos.
- En una realización, el marcador tiene de 100 a 150 residuos de aminoácidos.
- En una realización, el marcador tiene de 100 a 125 residuos de aminoácidos.
- 20 El marcador contendrá típicamente de 15 a 50%, de 15 a 45%, de 15 a 40%, de 15 a 35%, de 15 a 30%, de 15 a 25% o de 15 a 20% de residuos de aminoácidos básicos.
- En otra realización, el marcador comprende de 20 a 50%, de 20 a 40% o de 20 a 30% de residuos de aminoácidos básicos.
- En otra realización, el marcador comprende de 40% a 50% de residuos de aminoácidos básicos.
- En otra realización, el marcador comprende de 40% a 60% de residuos de aminoácidos básicos, Lys y Arg.
- 25 En una realización, el marcador comprende de 15 a 50% de residuos de aminoácidos básicos.
- En otra realización, el marcador comprende de 15 a 45% de residuos de aminoácidos básicos.
- En otra realización, el marcador comprende de 15 a 40% de residuos de aminoácidos básicos.
- En otra realización, el marcador comprende de 15 a 35% de residuos de aminoácidos básicos.
- En otra realización, el marcador comprende de 15 a 30% de residuos de aminoácidos básicos.
- 30 En otra realización, el marcador comprende de 15 a 25% de residuos de aminoácidos básicos.
- En otra realización, el marcador comprende del 15 al 20% de residuos de aminoácidos básicos.
- Los marcadores se obtienen a partir de la familia L9 de la proteína ribosómica, tal y como se describe en un ejemplo de la base de datos Prosite (Hulo N., Bairoch A., Bulliard V., Cerutti L., De Castro E., Langendijk-Genevaux P.S., Pagni M., Sigrist C.J.A. "The PROSITE database". Nucleic Acids Res. 34:D227-D230 (2006).
- 35 En una realización de la invención, el marcador se selecciona a partir del grupo de secuencias peptídicas que consiste en
- MSKTIVRKNESIDDALRRFKRAVSKTGTGLQEVKREFYKPSVRRKKKSEAARKRK (SEQ ID NO:1);
- MGKKTGVGKKRLAKAYKQNRAPWITVKTKRSVFGSPKRRHWRRSKLKV (SEQ ID NO:2);
- MKRTYQPSRRKRKRTHGFLARKRTPGGRRVLKNRRRKGWRWLV (SEQ ID NO:3);

MGKGDRRTRRGKIWRGTYGKYRPRKKK (SEQ ID NO:4);

MAKVKMKTNRSAARFKVTAKGKIKRWKSGGAHYNTKKSSKRKRHLRKHTYVKD
NMLKHVKALLKEF (SEQ ID NO:5);

MPKHSKRYLEARKLVDRTKYDDLDEAIELVKKTATAKFDETIELHIQTGIDYRKPEQHI
RGTIVLPHGTGKEVKLVFAKGEKAKEALEAGADYVGAEDLVEKIEKEGFLDFDVAIATPDM
MRIIGRLGKILGPRGLMPSPKSGTVTQEVAAEAVKEFKKGRIEVRTDKTGNIHIPVGKRSFDNE
KLKENIAAIKQIMQMMPAGVKGQFIKKVVLASTMGPGLNQLSLLKE(SEQ ID NO:6),

5

MAQVDLLNVKGEKVGTLISDFVFNIDPNYDVMWRYVDMQLSNRRAGTASTKTRG
EVSGGGRKPWPQKHTGRARHGSIRSPIWRHGGVVHGPKPRDWSKKNKMKKLALRSAL
SVKYRENKLLVDDLKLERPKTKSLKEILQNLQLSDKKTLLVLPWKEEGYMNVLKSGRNLPDV
KVIIADNPNSKNGEKAVRIDGLNVFDMLKYDYLVLRDMVSKIEEVLGNEAGKALTA(SEQ
ID NO:7),

MRYEYVPLKDQYEKEIVPALMKEFNYKNIHQVPKLVKIVINMGIGEGSRNYDLIERHA
NELAKITGQKPIVTRARKSISNFKIRKGMPIGLKVTLRGARMYNFLYKLINIVLPKVRDFRGLDP
NSFDGRGNYSFGLSEQLVPELNPDEVRRIQGMDITIVTTAKTDQEARLLELFGMPFKRG
(SEQ ID NO:8),

MSRLAKKPIVLPQGVTVEIKDNVVKVKGPKGELSQEFLPYVKIEVEGNEV
WVRPNEEQIIRKSDWRKVKMFQGTYSWLIRNMVGVTEGYKKELEIVGIGYRAQLQGNTLV
MNLGYAHPVVEIPSDVKIEVPAPNRIIVSGIDKQRVGQVAEIRAFRPPNVYTGKGIRYVGE
VVRQKEGKKA(SEQ ID NO:9),

10

MKVILLRDVPKIGKKGEIKEVSDGYARNYLIPRGFAKEYTEGLERAIKHEKEIEKRKKE
RERESEKILKELKKRTHVVVKVAGEGGKIFGAVTAATVAEEISKTTGLKLDKRWFKLDKPIK
ELGEYSLEVSLPGGVKDTIKIRVEREE(SEQ ID NO:10),

MLTRQQKELIVKEMSEIFKKTSLILFADFLGFTVADLTELSRLREKYGDGARFRV
KNTLLNLALKNAEYEGYEEFLKGPTAVLYVTEGDPVEAVKIIYNFYKDKKADLSRLKGGFLEG
KKFTAEEVENIAKLPSKEELYAMLVGRVKA PITGLVFALS GILRNLVYVLNAIKEKKSE(SEQ
ID NO:11),

15

MARYFPVQKTTMIKPEEVERKWYVVDASGKVLGRLATRIAKILMGKHKPN
YTPHVDTDGYVIVVNADKVVLTGKKLDQKVYYWHSGYPGGLKSLTARQML
EKHPERLIWLAVKRMLPKNRKGRKMLKRLKVYASPEHPHQAKPEPIEL(SEQ ID NO:12),

MRLEDLRPTPGAMKKRKRVRGRPGSGHGKTSRGRGHKGQKARGSGKVHIWFEGG
QTPLQRRLPKRGFKNINKKVYAVVNVKLEERFEANEEVTPEKLIERKIIKDLKDGVKILGDGE
LTKPLVVKAHAFSKSAVEKIESAGGKAEVI(SEQ ID NO:13),

MRHRVKRHKLGRYGSHRKSLLRNLRSREIVEHGSIVTTAKAKALKTFFMDKLVSKAIE
AATDDRARSVHLRRQINAVLGDRLTNKLVDEIAKNYVGRGGYVRVLRIGFRRGDAAEM
SLVQLVEASSQEG(SEQ ID NO:14),

20

MDHLVKIIEKKYEKKEIPDFRPGDTRVHVVKVIEGDRERTQVFEGIVIAKRGSGINKT
FTVRRIGSHGVGVERIFPVHSPVVEKIEVVRKGVVRAKLYLNRNVRGKIRIKERRD(SEQ ID
NO:15),

MRVKRAVHAKKKRKKYLKAAKGYRGALSRRYKLAKQMYVRSKWYSYVGRKQKK
RDMRKLWITRINIAARNEGLKYSELIHGLKLAGVSINRKMLSELAVNDPEAFKEYVKIAKEALA
S(SEQ ID NO:16),

5 MLYAIVETAGRQYRVEEGKILYTEKQKDYSYVDRVVFVRKDGEVLVGPYV
EGAKVVGKVLHAKARKVKTKYRPRKNSKVEKGRQWYTAIKIEKIEL(SEQ ID NO:17),

MKQEKLSLHDVLRPIITEKALILREQRKYVFEVNPLANKNLVKEAVEKLFNVKVEKV
NILNMKPKPKRRGIFEGKTRSWKKAVVTLKEGYTIKELEGEH(SEQ ID NO:18),

MAHKKSGGVAKNGRDSLPHYLVGVGDGQIVKAGNILVRQRGTRFYPGKNVGMG
RDFTLFALKDGRVKFETKNNKYSVYEE(SEQ ID NO:19),

10 MKASELRNYTDEELKNLLEEKRRQLMELRFQLAMGQLKNTSLIKLTKRDI
ARIKTILRERELGIRR(SEQ ID NO:20),

MPKKLKIKLVKSPIGYSWDQKDTVKRLGLKKNQVVIKDDLPIRGMIRK
VKHLVEVEEIEEGGSNA(SEQ ID NO:21),

15 MPKVKTNRSAKRFRITKNGKIMRNHAYRSHKTGKKRRNALRALRKKDVVSSADK
NRVLRLLGKK(SEQ ID NO:22),

MGQKVHPRGFRLGLSADWQAKWFNEKNYKEWLLDEEIRKIIKNKYYHAGISEIYV
ERPDAERINITVKTARPGIIGRKGSEITSLREELERKFNRRVINIEEIKTPELDAQLVAESIAS
RIEKRASVYKAMKRAIMNAMRKGAGQIKVMVAGRLGGAEIARREWYLRGRLPLQKIKAIIDY
GTATAWTKYGTIGIKVWIYKGDADI(SEQ ID NO:23),

METQGVMEIQYEEFEKIIIRRTSKVTKGGKNLSFRVVAIVGNKNGKVLGIGIKAR
EVPEAIRKAISAARNIVEVPVINGTIPHEVIGRQDASKVLLKPAAPGTGIIAGGTVRAVELAG
IQNILTKSLGSTNPLNLALATMNGLNLLDPRKVAKLRDISVEEVFKGVRRENN(A SEQ ID
NO:24),

20 MVSLDPEKKNEIIEFQIHENDTGSVEVQIALLTARIKHLTEHLRKHHPKDFHSRRGLM
KMIGRRRKMLKYLRHKKPEVYRELIKLGIRK(SEQ ID NO:25),

MGRSRKKGYPVDRKLEKIRKLNKNETGEKKVIKTWSRSMIPEMVGHTIAVYNGMK
HIPVYITENMIGHRLGEFAPTRRFGGHADKKAKKGELKK(SEQ ID NO:26) . y

MPNIKSAKRVRVSEKRRLRNKAYKTFKRNRIKEVLKAIENKEPKEVVLELTRKAQA
AIDKAVSKGVIHKNQGARRKARLFEKVNEYLRTLETTQE(SEQ ID: NO 27).

En otra realización, el marcador se selecciona a partir del grupo que consiste en

25 MKQEKLSLHDVLRPIITEKALILREQRKYVFEVNPLANKNLVKEAVEKLFNVKVEKVNILNMK
PKPKRRGIFEGKTRSWKKAVVTLKEGYTIKELEGEH(SEQ ID NO:18) . y

MAHKKSGGVAKNGRDSLPHYLVGVGDGQIVKAGNILVRQRGTRFYPGKNVGMGRDFTLF
ALKDGRVKFETKNNKYSVYEE(SEQ ID NO:19).

En otra realización, el marcador se selecciona a partir del grupo que consiste en

MKVILLRDVDPKIGKKGEIKEVSDGYARNYLIPRGFAKEYTEGLERAIKHEKEIEKRKKERERE
SEKILKELKKRTHVVKVKGEGGKIFGAVTAATVAEEISKTTGLKLDKRWFKLDKPIKELGEY
SLEVSLPGGVKDTIKIRVEREE(SEQ ID NO:10); MDHLVKIIEKKYEKKEIPDFRPGD-
TVRVHVKVIEGDRERTQVFEGIVIAKRSGINKTFTVRRIGSHGVGVERIFPVHSPVVEKIEVV
RKGKVRRAKLYYLRNVRGKIRIKERRD(SEQ ID NO:15) and MSKTIVRKNESIDDALRRFKR
AVSKTGTLQEVKREFYEKPSVRRKKKSEAARKRK (SEQ ID NO:1).

En otra realización, el marcador se selecciona a partir del grupo que consiste en

MPKHSKRYLEARKLVDRTKYDLDDEAIELVKKTATAKFDETIELHIQTGIDYRKPEQ
HIRGTIVLPHGTGKEVKVLVFAKGEKAKEALEAGADYVGAEDLVEKIEKEGFLDFDVAIATPD
MMRIIGRLGKILGPRGLMPSKSGTVTQEVAVKEFKKGRIEVRTDKTGNIHIPVGKRSFDN
EKLKENIIAAIKQIMQMKPAGVKGQFIKKVVLASTMGPGIKLNLSLLKE(SEQ ID NO:6);

MAQVDLLNVKGEKVGTLAISDFVFNIDPNYDVMWRYVDMQLSNRRAGTASTKTRG
EVSGGGRKPWPQKHTGRARHGSIRSPIWRHGGVHVGPKPRDWSKLNKMKKLALRSAL
SVKYRENKLLVLDLKLERPCTKSLKEILQNLQSDKKTLLVLPWKEEGYMNKLSGRNLPDV
KVIIADNPNSKNGEKAVRIDGLNVFDMLKYDYLVLTRDMVSKIEEVLGNEAGKALTA(SEQ
ID NO:7),

5

MSRLAKKPIVLPQGVTVVEIKDNVVKVKGPKGELSQEFLPYVKIEVEGNEV
WVRPNEEQIIRKSDWRKVKMFQGTYSLIRNMVVGVTGYKKELEIVGIGYRAQLQGNTLV
MNLGYAHPVVEIPSDVKIEVPAPNRIVSGIDKQRVGQVAEIRAFRPPNVYTGKIRYVGE
VVRQKEGKKA(SEQ ID NO:9),

MGQKVHPRGFRLGLSADWQAKWFNEKNYKEWLLDEEIRKIIKNKYHAGISEIYV
ERPDAERINIVKTARPGIIGRKGSEITSLREELERKFNRVVIINIEIKTPELDAQLVAESIAS
RIEKRASVYKAMKRAIMNAMRGAQGIKVMVAGRLGGAEIARREWYLRGRLPLQKIKAIIDY
GTATAWTKYGTIGIKVWIYKGDADI(SEQ ID NO:23) y

10

METQGVMEIKIYEEFEKIIERRTSKVTKGGKNLSFRVVAIVGNKNGK/VGLGIGKA
REVPEAIRKAISAAKRNIVEVPVINGTIPHEVIGRQDASKVLLKPAAPGTGIIAGGTVRAVELA
GIQNILTKSLGSTNPLNLALATMNGLNLLDPRKVAKLKRDISVEEVFKGVRRENNNA(SEQ ID
NO:24),

En otra realización, el marcador se selecciona a partir del grupo que consiste en

MKVILLRDVDPKIGKKGEIKEVSDGYARNYLIPRGFAKEYTEGLERAIKHEKEIEKRKKE
RERESEKILKELKKRTHVVKVKGEGGKIFGAVTAATVAEEISKTTGLKLDKRWFKLDKPIK
ELGEYSLEVSLPGGVKDTIKIRVEREE(SEQ ID NO:10);

MKVILLEPLENLGDVGVQVVDVKPGYARNYLLPRGLAVLATESNLKALEARIRAQAKR
LAERKAEAERLKEILENLTLTIPVRAGETKIYGSVTAKDIAEALSRQHGITIDPKRLALEKPIKEL
GEYVLTYPHPEVPIQLKVSVAQE(SEQ ID NO: 41);

15

20

MKVIFLKDVKGKGGKKEIKDVADGYANNFLFKQGLAIEATPANIKALEAQKQKEQRQAAEEL
 ANAKLKEELEKLTVEIPAKAGEGGRLFGSITSKQIAEALQAQHGLKLDKRKIELADAIRSLGY
 TNVPVKLHPEVTATLKVHVKEQK(SEQ ID NO: 42) y

MKVLLKDVSKIGKKGEIKNVSDGYARNYLIPKGLALEATPRVLKRLEAEKR-
 KKEEEKIQIKTQNEELLKMLKKFLYKIPVKAGESGKLFALTNSDIAKAVEKIADVNI
 DKKFIVLEKPIKEIGMYDVLVRLPEGVSGKIKVEVIQEGKN(SEQ ID NO: 43).

5 En una realización adicional, el marcador tiene la secuencia MKQEKLSDHDLIRPIITEKALIL
 REQRKYVFEVNPANKNLVKEAVEKLFNVKVEKVNILNMKPKPKRRGIFEGKTRSWKKAVVT LKEGYTIKELEGEH
 (SEQ ID NO: 18).

En una realización adicional, el marcador tiene la secuencia MAHKKSGGVAKNGRDSLPL
 KYLGVKVGDGQIVKAGNILVRQRGTRFYPGKNVGMGRDFTLFALKDGRVKFETKNNKYYVS VYEE (SEQ ID NO:
 19).

10 En una realización adicional, el marcador tiene la secuencia MDHLVKIIEKKYEKKEIPD
 FRPGDTRVHVVKVIEGDRERTQVFEGIVIAKRGSGINKTFTVRRIGSHGVGVERIFPVHSPVV
 EKIEVVRKGVRRAKLYLRLNVRGKIRIKERRD (SEQ ID NO: 15).

15 En una realización adicional, el marcador tiene la secuencia
 MKVILLRDVDPKIGKKGEIKEVSDGYARNYLIPRGFAKEYTEGLERAIKHEKEIEKRKKERERE
 SEKILKELKKRTHVVKVAGEGGKIFGAVTAATVAEEISKTTGLKLDKRWFKLDPKIKELGEY SLEVSLPGGVKDTI-
 KIRVEREE (SEQ ID NO: 10).

En una realización adicional, el marcador tiene la secuencia
 MKVILLEPLENLGDVGVVDVDPGYARNYLLPRGLAVLATESNLKALEARIRAQAKRLAERK
 AEAERLKEILENLTIPVRAGETKIYGSVTAKDIAEALSRQHGITIDPKRLALEKPIKELGEYVL TYKPHPEV-
 PIQLKVSVAQE (SEQ ID NO: 41).

20 En una realización adicional, el marcador tiene la secuencia
 MKVIFLKDVKGKGGKKEIKDVADGYANNFLFKQGLAIEATPANIKALEAQKQKEQRQAAEEL
 ANAKLKEELEKLTVEIPAKAGEGGRLFGSITSKQIAEALQAQHGLKLDKRKIELADAIRSLGY TNVPVKLHPEV-
 TATLKVHVKEQK (SEQ ID NO: 42).

25 En una realización adicional, el marcador tiene la secuencia MKVLLKDVSKIGKKGEIKNV
 VSDGYARNYLIPKGLALEATPRVLKRLEAEKRKKEEEKIQIKTQNEELLKMLKKFLYKIPVKAGESGKLFALTNSDIAK
 AVEKIADVNI DKKFIVLEKPIKEIGMYDVLVRLPEGVSGKIKVEVIQEGKN (SEQ ID NO: 43).

30 El marcador se puede fijar directamente al aminoácido N-terminal en la enzima de procesamiento en cuestión o a
 través de un grupo enlazador. Tal grupo enlazador puede tener una función polivalente es decir, facilitar la fijación
 del marcador a la enzima de procesamiento y colocar el marcador y la enzima de procesamiento en una posición
 relativa entre sí que permita el plegamiento correcto de la enzima de procesamiento. El enlazador puede servir tam-
 bién para presentar el marcador de una manera favorable a la matriz de purificación utilizada para purificar la enzima
 de procesamiento marcada.

El enlazador puede tener de 1-30, de 1-25, de 1-20, de 1-15, de 1-14, de 1-13, de 1-12, de 1-11, de 1-10, de 1-9, de
 1-8, de 1-7, de 1-6, de 1-5, de 1-4 o de 1-3 residuos de aminoácidos.

35 En una realización adicional, el enlazador tendrá de 2-30, de 2-25, de 2-20, de 2-15, de 2-14, de 2-13, de 2-12, de 2-
 11, 2-10, de 2-9, de 2-8, de 2-7, de 2-6, de 2-5 o de 2-4 residuos de aminoácidos.

En una realización adicional, el enlazador tendrá de 3-30, de 3-25, de 3-20, de 3-15, de 3-14, de 3-13, de 3-12, de 3-
 11, de 3-10, de 3-9, de 3-8, de 3-7, de 3-6, de 3-5 o de 3-4 residuos de aminoácidos.

40 En una realización adicional, el enlazador tendrá de 4-30, de 4-25, de 4-20, de 4-15, de 4-14, de 4-13, de 4-12, de 4-
 11, de 4-10, de 4-9, de 4-8, de 4-7, de 4-6 o de 4-5 residuos de aminoácidos.

En una realización adicional, el enlazador tendrá de 5-30, de 5-25, de 5-20, de 5-15, de 5-14, de 5-13, de 5-12, de 5-
 11, de 5-10, de 5-9, de 5-8, de 5-7 o de 5-6 residuos de aminoácidos.

En una realización adicional, el enlazador tendrá de 6-30, de 6-25, de 6-20, de 6-15, de 6-14, de 6-13, de 6-12, de 6-
 11, de 6-10, de 6-9, de 6-8 o de 6-7 residuos de aminoácidos.

45 En una realización adicional, el enlazador tendrá de 7-30, de 7-25, de 7-20, de 7-15, de 7-14, de 7-13, de 7-12, de 7-
 11, de 7-10, de 7-9 o de 6-8 residuos de aminoácidos.

- En una realización, el enlazador tiene 20 residuos de aminoácidos.
- En otra realización, el enlazador tiene 15 residuos de aminoácidos.
- En otra realización, el enlazador tiene 12 residuos de aminoácidos.
- En otra realización, el enlazador tiene 10 residuos de aminoácidos.
- 5 En otra realización, el enlazador tiene 9 residuos de aminoácidos.
- En otra realización, el enlazador tiene 8 residuos de aminoácidos.
- En otra realización, el enlazador tiene 7 residuos de aminoácidos.
- En otra realización, el enlazador tiene 6 residuos de aminoácidos.
- En otra realización, el enlazador tiene 5 residuos de aminoácidos.
- 10 Los residuos de aminoácidos en el enlazador pueden ser cualquier residuo de aminoácido y pueden ser iguales o diferentes, dependiendo de la función prevista para el enlazador.
- En una realización, el enlazador comprenderá residuos de aminoácidos que hacen que el enlazador sea una estructura flexible, tales como residuos de Ser y Gly alternos. Otros residuos de aminoácidos dan como resultado una estructura rígida, es decir, aminoácidos que pueden formar estructuras helicoidales alfa, tales como Ala y Leu.
- 15 Además, ciertos residuos de aminoácidos añaden una curvatura estructural al enlazador, por ejemplo, Pro.
- Ejemplos no limitantes de grupos enlazadores son SSSGSSGSSGSS (SEQ ID NO: 28), GGSSGGSS (SEQ ID NO: 29), SSSGSGSG (SEQ ID NO: 30), ALALALA (SEQ ID NO: 31), ALALALAPA (SEQ ID NO: 32), SSSALALALA (SEQ ID NO: 33); SGSGSGSGS (SEQ ID NO: 34), SSSGSGSGSG (SEQ ID NO: 47) y GSSGSGS (SEQ ID NO: 48).
- En una realización, el enlazador tiene la secuencia SSSGSSGSSGSS (SEQ ID NO: 28).
- 20 En otra realización, el enlazador tiene la secuencia GGSSGGSS (SEQ ID NO: 29).
- En otra realización, el enlazador tiene la secuencia SSSGSGSG (SEQ ID NO: 30).
- En otra realización, el enlazador tiene la secuencia ALALALA (SEQ ID NO: 31).
- En otra realización, el enlazador tiene la secuencia ALALALAPA (SEQ ID NO: 32)
- En otra realización, el enlazador tiene la secuencia SSSALALALA (SEQ ID NO: 33).
- 25 En otra realización, el enlazador tiene la secuencia SGSGSGSGS (SEQ ID NO: 34).
- En otra realización, el enlazador tiene la secuencia SSSGSGSGSG (SEQ ID NO: 47).
- En otra realización, el enlazador tiene la secuencia GSSGSGS (SEQ ID NO: 48).
- En una realización, los marcadores descritos anteriormente pueden ser una combinación de dos marcadores que pueden ser iguales o diferentes. También el marcador puede ser un análogo de los marcadores descritos específicamente. Por un "análogo" se entiende una secuencia que tiene hasta 10% de mutaciones en comparación con la secuencia original. Tales mutaciones comprenderán adiciones, deleciones o adiciones o cualquier combinación de las mismas, siempre y cuando tales mutaciones no alteren la funcionalidad original del marcador.
- 30 En una realización, la invención incluye enzimas de procesamiento marcadas en donde el marcador es un análogo truncado del marcador original. La expresión "análogo truncado" se entiende que incluye una secuencia del marcador original, en la que hasta 10% de los residuos de aminoácidos originales del extremo N-terminal o del extremo C-terminal se han delecionado, con la condición de que el análogo truncado haya conservado la funcionalidad de la secuencia sin modificar.
- 35 Si el marcador comprende una combinación de dos marcadores, en una realización estos marcadores estarán conectados a través de un enlazador.
- 40 En una realización, el enlazador que conecta dos marcadores se selecciona a partir del grupo que consiste en SSSGSSGSSGSS (SEQ ID NO: 28); GGSSGGSS (SEQ ID NO: 29), SSSGSGSG (SEQ ID NO: 30), ALALALA (SEQ ID NO: 31), ALALALAPA (SEQ ID NO: 32), SSSALALALA (SEQ ID NO: 33); SGSGSGSGS (SEQ ID NO: 34), SSSGSGSGSG (SEQ ID NO: 47) y GSSGSGS (SEQ ID NO: 48).
- 45 La invención incluirá cualquier combinación de las realizaciones del marcador, el enlazador y la enzima de procesamiento tal y como se describen en esta memoria.

En una realización, la enzima de procesamiento marcada comprenderá una extensión N-terminal que comprende una combinación del marcador y del enlazador, seleccionada a partir del siguiente grupo: SEQ ID NO: 1 y SEQ ID NO: 28, SEQ ID NO: 3 y SEQ ID NO: 28, SEQ ID NO: 6 y SEQ ID NO: 28, SEQ ID NO: 9 y SEQ ID NO: 28, SEQ ID NO: 10 y SEQ ID NO: 28, SEQ ID NO: 11 y SEQ ID NO: 28, SEQ ID NO: 12 y SEQ ID NO: 28, SEQ ID NO: 13 y SEQ ID NO: 28, SEQ ID NO: 15 y SEQ ID NO: 28, SEQ ID NO: 16 y SEQ ID NO: 28, SEQ ID NO: 20 y SEQ ID NO: 28, SEQ ID NO: 21 y SEQ ID NO: 28, SEQ ID NO: 24 y SEQ ID NO: 28, SEQ ID NO: 27 y SEQ ID NO: 28, SEQ ID NO: 18 y SEQ ID NO: 28, SEQ ID NO: 19 y SEQ ID NO: 28, SEQ ID NO: 41 y SEQ ID NO: 28, SEQ ID NO: 42 y SEQ ID NO: 28, y SEQ ID NO: 43 y SEQ ID NO: 28.

En una realización, la enzima de procesamiento marcada comprenderá una extensión N-terminal que comprende una combinación del marcador y del enlazador, seleccionada a partir del siguiente grupo: SEQ ID NO: 1 y SEQ ID NO: 47, SEQ ID NO: 3 y SEQ ID NO: 47, SEQ ID NO: 6 y SEQ ID NO: 47, SEQ ID NO: 9 y SEQ ID NO: 47, SEQ ID NO: 10 y SEQ ID NO: 47, SEQ ID NO: 11 y SEQ ID NO: 47, SEQ ID NO: 12 y SEQ ID NO: 47, SEQ ID NO: 13 y SEQ ID NO: 47, SEQ ID NO: 15 y SEQ ID NO: 47, SEQ ID NO: 16 y SEQ ID NO: 47, SEQ ID NO: 20 y SEQ ID NO: 47, SEQ ID NO: 21 y SEQ ID NO: 47, SEQ ID NO: 24 y SEQ ID NO: 47, SEQ ID NO: 27 y SEQ ID NO: 47, SEQ ID NO: 18 y SEQ ID NO: 47, SEQ ID NO: 19 y SEQ ID NO: 47, SEQ ID NO: 41 y SEQ ID NO: 47, SEQ ID NO: 42 y SEQ ID NO: 47, y SEQ ID NO: 43 y SEQ ID NO: 47.

En una realización, la enzima de procesamiento marcada comprenderá una extensión N-terminal que comprende una combinación del marcador y del enlazador, seleccionada a partir del siguiente grupo: SEQ ID NO: 1 y SEQ ID NO: 48, SEQ ID NO: 3 y SEQ ID NO: 48, SEQ ID NO: 6 y SEQ ID NO: 48, SEQ ID NO: 9 y SEQ ID NO: 48, SEQ ID NO: 10 y SEQ ID NO: 48, SEQ ID NO: 11 y SEQ ID NO: 48, SEQ ID NO: 12 y SEQ ID NO: 48, SEQ ID NO: 13 y SEQ ID NO: 48, SEQ ID NO: 15 y SEQ ID NO: 48, SEQ ID NO: 16 y SEQ ID NO: 48, SEQ ID NO: 20 y SEQ ID NO: 48, SEQ ID NO: 21 y SEQ ID NO: 48, SEQ ID NO: 24 y SEQ ID NO: 48, SEQ ID NO: 27 y SEQ ID NO: 48, SEQ ID NO: 18 y SEQ ID NO: 48, SEQ ID NO: 19 y SEQ ID NO: 48, SEQ ID NO: 41 y SEQ ID NO: 48, SEQ ID NO: 42 y SEQ ID NO: 48, y SEQ ID NO: 43 y SEQ ID NO: 48.

En una realización el marcador se selecciona a partir del grupo que consiste en SEQ ID NO: 10, SEQ ID NO: 41, SEQ ID NO: 42 y SEQ ID NO: 43 y el enlazador se selecciona a partir del grupo que consiste en SEQ ID NO: 28, SEQ ID NO: 29, SEQ ID NO: 30, SEQ ID NO: 34, SEQ ID NO: 47 y SEQ ID NO: 48.

En otro aspecto, la presente invención se refiere a un método para preparar una enzima de procesamiento de acuerdo con la invención, que comprende;

- iv) la expresión de la enzima de procesamiento de acuerdo con la invención,
- v) cargar dicha enzima de procesamiento expresada sobre una columna de intercambio catiónico, y
- vi) eluir dicha enzima de procesamiento con un eluyente adecuado.

En otro aspecto, la invención se refiere a un método en el que la enzima de procesamiento marcada en el extremo N-terminal y el precursor de la proteína diana se coexpresan en un hospedador adecuado para permitir el aislamiento de la proteína diana después de la expresión.

Las enzimas de procesamiento se seleccionan típicamente a partir de oxidorreductasas, transferasas, hidrolasas, liasas, isomerasas y ligasas.

En una realización, la enzima de procesamiento es una proteasa que se utiliza para escindir proteínas de fusión.

En una realización, la proteasa de procesamiento se selecciona a partir del grupo que consiste en aspartoproteasas, metaloproteasas, proteasas de serina y de cisteína.

En una realización adicional las proteasas de serina incluyen tripsinas, quimotripsinas, enterocinasa (EK), trombina, factor Xa y calicreínas.

En una realización, las proteasas de cisteína se seleccionan a partir del grupo de proteasas 3C y proteasas 2A codificadas por los genomas víricos del grupo de *Coronaviridae*, *Picornaviridae*, *Retroviridae*, *Adenoviridae* o *Flaviviridae*. Otras realizaciones preferidas incluyen proteasas 3C y 2A procedentes de *Picornaviridae* y *Coronaviridae*.

En una realización adicional, las proteasas 3C se seleccionan a partir de la proteasa 3C del rinovirus humano con la secuencia
 GPNTEFALSLLRKNIMTITTSKGEFTGLGIHDRVCVIPTH-
 AQPGLDVLVNGQKIRVKDKYKLVDPENINLELTVLTLDRNEKFRDIRGFISEDLEGVDATLVV
 HSNFTNTILEVGPVTMAGLINLSSTPTNRMIRYDYATKTGQCGGVLCATGKIFGIHVGGNG RQGFSACLKQYF-
 VEKQ (SEQ ID NO: 35); la proteasa 3C de la cepa 18 del virus de la Hepatitis A con la secuencia
 STLEIAGLVRKNLVQFVGEKNGCVRWVMNALGVKDDWLLVPSHAYKFE-
 KDYEMMEFYFNRRGGTYYSISAGNVVIQSLDVGFDVVLKVPPTIPKFRDITEHFIIKKGDVPR
 ALNRLATLVTVNGTPMLISEGPLKMEEKATYVHKKNDGTTVDLTVDQAWRGKGEGLPGM
 CGGALVSSNQSIQNAILGIHVAGGNSILVAKLVTQEMFQNIKDKIESQ (SEQ ID NO: 45) o la proteasa similar a 3C del

virus de la enfermedad hemorrágica del conejo con la secuencia
GLPGFMRHNGSGWMIHIGNGLYISNTH TARSSCSEIVTCSPTTDLCLVKGEAIRSVAQIAE
GTPVCDWKKSPISTYGIKKTLS DSTKIDVLAYD GCTQTTHGDCGLPLYDSSGKIVAIHTGK LLGFSKMCTLIDL-
TITKGVYE (SEQ ID NO: 46).

- 5 En una realización adicional, la proteasa es de un virus vegetal. Las proteasas normalmente procederán de las siguientes familias de virus vegetales: *Bromoviridae*, *Partitiviridae*, *Geminiviridae*, *Potyviridae*, *Tombusviridae*, *Comoviridae*, *Rhabdoviridae*, *Reoviridae*, *Potyviridae*, *Sequiviridae* y *Bunyaviridae*.

En una realización, la proteasa procedente de un virus vegetal es la proteasa madura del virus del jaspeado del tabaco (TEV) con la secuencia de aminoácidos:

GESLFKGRDYNPISSTICHLT NESDGHTTSLYGIGFGPFIITNKHLFRRNNGTLLVQ
SLHGVFKVKNTTTLQQLIDGRDMIIRMPKDFPPFPQKLFREPQREERICLVTTNFQTKSM
SSMVS DTSCTFPSSDGIFWKHWIQTKD GQCGSPLVSTRDGFIVGIHSASNFTNTNNYFTSVP
KNFMELLTNQEAQQWVSGWRLNADSVLWGGHKVFMVKPEEPFQPVKEATQLMN (SEQ ID
10 NO: 38)

- 15 La proteína diana que se puede producir a partir de una proteína precursora modificada con la enzima de procesamiento de acuerdo con la invención, se puede seleccionar a partir de proteínas humanas y sus análogos tales como aprotinina, inhibidor de la vía del factor tisular u otros inhibidores de proteasas, insulina o análogos de insulina, hormona del crecimiento humana o bovina, interleucina, glucagón, GLP-1, GLP-2, IGF-I, IGF-II, activador del plasminógeno tisular, factor de crecimiento transformante α o β , factor de crecimiento obtenido a partir de plaquetas, GRF (factor de liberación de la hormona de crecimiento), inmunoglobulinas, EPO, TPA, proteína C, factores de coagulación de la sangre tales como FVII, FVIII, FIV y FXIII, exendina-3, exentidina-4 y enzimas o análogos funcionales de las mismas.

- 20 En un aspecto adicional de la invención, la proteína diana se selecciona a partir de péptidos antagonistas o agonistas de la insulina, tales como S661 u hormonas peptídicas tales como la amilina o análogos funcionales de las mismas.

En un aspecto adicional de la invención, la proteína diana se selecciona a partir de enzimas adecuadas para cualquier modificación de proteínas, tales como transglutaminasas, transferasas y racemasas.

- 25 En una realización, la invención se refiere a un procedimiento para la eliminación de un prodominio procedente de la protransglutaminasa (proTGasa) utilizando una enzima de procesamiento marcada de acuerdo con la invención.

- 30 En otra realización, la invención se refiere a un procedimiento para la escisión de una protransglutaminasa (proTGasa) con una proteasa HRV14 3C marcada, una proteasa HAV18 3C marcada o una proteasa RHDV 3C marcada. En esta realización, el marcador se puede seleccionar a partir de SEQ ID NO: 10, SEQ ID NO: 41, SEQ ID NO: 42 o SEQ ID NO: 43 y el enlazador se puede seleccionar a partir del grupo que consiste en SEQ ID NO: 28, SEQ ID NO: 29, SEQ ID NO: 30, SEQ ID NO: 34, SEQ ID NO: 47 y SEQ ID NO: 48.

- 35 En otra realización, la invención se refiere a un procedimiento para la eliminación de parejas de fusión de la hormona de crecimiento humana o de análogos de la misma, con una proteasa HRV14 3C marcada, una proteasa HAV18 3C marcada, una proteasa TEV marcada o una proteasa RHDV 3C marcada. En esta realización, el marcador se puede seleccionar a partir de SEQ ID NO: 10, SEQ ID NO: 41, SEQ ID NO: 42 o SEQ ID NO: 43 y el enlazador se puede seleccionar a partir del grupo que consiste en SEQ ID NO: 28, SEQ ID NO: 29, SEQ ID NO: 30, SEQ ID NO: 34, SEQ ID NO: 47 y SEQ ID NO: 48.

- 40 En otra realización, la invención se refiere a un procedimiento para la eliminación de parejas de fusión de amilina humana o análogos de la misma con una proteasa HRV14 3C marcada, una proteasa HAV18 3C marcada, una proteasa TEV marcada o una proteasa RHDV 3C marcada. En esta realización, el marcador se puede seleccionar a partir de SEQ ID NO: 10, SEQ ID NO: 41, SEQ ID NO: 42 o SEQ ID NO: 43 y el enlazador se puede seleccionar a partir del grupo que consiste en SEQ ID NO: 28, SEQ ID NO: 29, SEQ ID NO: 30, SEQ ID NO: 34, SEQ ID NO: 47 y SEQ ID NO: 48.

- 45 Estas marcadores se obtienen a partir de cepas de bacterias termófilas, con temperaturas óptimas que oscilan desde ~50°C hasta por encima del punto de ebullición del agua. Las cepas que sobreviven a temperaturas extremadamente altas se denominan hipertermófilas o termófilas y tienen una temperatura óptima de 80°C (176°F) o superior. Las bacterias termófilas se producen de forma natural en aguas termales, suelos calientes, respiraderos geotérmicos y otros lugares en los que están presentes temperaturas elevadas. *Bacillus stearothermophilus* a partir del cual se ha clonado RS21_BACST (SEQ ID NO: 1) y que se utiliza como marcador, se ha observado que, por ejemplo, crece a una temperatura superior a 65°C en muchos suelos. Con el fin de poder sobrevivir a las temperaturas elevadas,
50 estos organismos han desarrollado proteínas que son más estables que las de bacterias mesófilas.

- Las marcadores de acuerdo con la presente invención obtenidos a partir de tales bacterias termófilas son en general solubles, altamente estables y tienen un pI muy básico debido a una gran cantidad de residuos de Arg y Lys presentes en la secuencia de aminoácidos. El pI de los marcadores es superior a 9. El pI de los marcadores estará generalmente entre 9 y 13,5. En una realización, el marcador tiene un pI entre aproximadamente 9 y 12,5. En otra realización, el pI del marcador estará entre aproximadamente 10 y aproximadamente 12,5 y en un aspecto adicional, el pI es de aproximadamente 10.
- Se cree que la solubilidad se obtiene a partir de la carga superficial generalmente elevada de las proteínas.
- Las enzimas de procesamiento de acuerdo con la presente invención se pueden usar para procesar proteínas expresadas de manera recombinante, para proporcionar una proteína modificada deseada.
- 10 Las enzimas de procesamiento se seleccionan a partir del siguiente grupo de enzimas: oxidorreductasas, que pueden catalizar reacciones de oxidación/reducción, transferasas, que pueden transferir un grupo funcional (p. ej., grupos fosfato), hidrolasas que catalizan la hidrólisis de diversos enlaces (incluyendo proteasas que hidrolizan enlaces peptídicos), liasas, que escinden enlaces por medios distintos de la hidrólisis y la oxidación, isomerasas, que catalizan cambios dentro de una molécula aislada, o ligasas, que unen dos moléculas a través de enlaces covalentes.
- 15 Estas enzimas pueden ser o bien enzimas de tipo silvestre o variantes o análogos de las mismas, modificadas mediante etapas de mutagénesis bien conocidas.
- Ejemplos de proteasas marcadas que se pueden utilizar para escindir parejas de fusión a partir de una proteína diana o eliminar prodominios o recortar aminoácidos en el extremo N-terminal o C-terminal de proteínas diana, son aspartoproteasas, metaloproteasas, proteasas de serina y cisteína.
- 20 Ejemplos no limitantes de proteasas de serina incluyen tripsinas, quimotripsinas, enterocinasa (EK), trombina, factor Xa y calicreínas.
- Ejemplos no limitantes de proteasas de cisteína incluyen proteasas víricas obtenidas a partir del grupo de *Coronaviridae*, *Picornaviridae*, *Retroviridae*, *Adenoviridae*, *Caliciviridae* o *Flaviviridae*. Otros ejemplos incluyen proteasas procedentes del virus herpes, citomegalovirus, virus de la hepatitis A-D, virus del dengue, rinovirus, virus de la polio, virus Cocksackie, virus del síndrome respiratorio agudo grave (SRAG), virus mengo, virus de la polio, virus de la enfermedad aftosa, virus de la inmunodeficiencia humana (VIH), virus T-linfotrópico humano (HTLV), virus espumoso humano y adenovirus. Otros ejemplos no limitantes incluyen proteasas de cisteína obtenidas a partir de bacterias tales como *Leishmania sp.* o microbios eucariotas tales como *Trypanosoma sp.*
- 25 Ejemplos no limitantes de proteasas de cisteína se seleccionan a partir del grupo de proteasas 3C y proteasas 2A codificadas por los genomas víricos del grupo de *Coronaviridae*, *Picornaviridae*, *Retroviridae*, *Adenoviridae* o *Flaviviridae*. Otras realizaciones preferidas incluyen proteasas 3C y 2A procedentes de *Picornaviridae* y *Coronaviridae*. Otros ejemplos de proteasas de cisteína incluyen caspasas tales como las caspasas humanas 1-10.
- Una revisión de las proteasas 3C se encuentra en Krausslich, H.G. & Wimmer, E., *Viral proteinases*, Annual Review of Biochemistry (1988) 57, 701-754.
- 30 Ejemplos no limitantes de metaloproteasas incluyen metaloproteasas matriciales tales como matrixinas o metaloproteasas unidas a la membrana, tales como MT1-MMP.
- Otras realizaciones incluyen aspartoproteasas tales como pepsina y catepsina, p. ej., catepsina D y quimosina.
- Otros ejemplos de enzimas incluyen lipasas, fosfatasas, glicosilhidrolasas (p. ej., manosidasas, xilosidasas, fucosidasas), cinasas, mono o dioxidasas, peroxidasas, transaminasas, carboxipeptidasas, amidasas, esterasas y fosfatasas.
- 40 El precursor de la proteína diana que se puede modificar a través de la enzima de procesamiento marcada puede ser cualquier proteína y ejemplos ilustrativos son aprotinina, inhibidor de la vía del factor tisular u otros inhibidores de proteasas, insulina o precursores de insulina, hormona de crecimiento humana o bovina, interleucina, glucagón, GLP-1, GLP-2, IGF-I, IGF-II, activador del plasminógeno tisular, factor de crecimiento transformante α o β , factor de crecimiento obtenido a partir de plaquetas, GRF (factor de liberación de la hormona de crecimiento), inmunoglobulinas, EPO, TPA, proteína C, factores de coagulación de la sangre tales como FVII, FVIII, FIV y FXIII, exendina-3, exentidina-4 y enzimas o análogos funcionales de las mismas.
- 45 Otros ejemplos de proteínas diana son el factor de crecimiento transformante α (TGF- α), el factor de crecimiento transformante β (TGF- β), el factor de crecimiento epidérmico (EGF), el factor de crecimiento endotelial vascular (VEGF), la trombopoyetina (TPO), el interferón, la prourocinasa, la urocinasa, el inhibidor 1 del activador del plasminógeno, el inhibidor 2 del activador del plasminógeno, el factor de von Willebrandt, una citocina, por ejemplo, una interleucina tal como interleucina (IL) 1, IL-1Ra, IL-2, IL-4, IL-5, IL-6, IL-9, IL-11, IL-12, IL-13, IL-15, IL-16, IL-17, IL-18, IL-20, IL-21, IL-22, IL-23, IL-24, IL-25, IL-26, IL-27, IL-28, IL-29, IL-30, IL-31, IL-32 o IL-33, un factor estimulador de colonias (CSF), tal como GM-CSF, un factor de células madre, un factor de necrosis tumoral como TNF- α , linfo-
- 50

- toxina α , linfotóxina β , CD40L o CD30L, un inhibidor de proteasas, por ejemplo, aprotinina, una enzima tal como superóxido dismutasa, asparaginasa, arginasa, arginina desaminasa, adenosina desaminasa, ribonucleasa, catalasa, uricasa, bilirrubina oxidasa, tripsina, papaína, fosfatasa alcalina, β -glucuronidasa, fosforilasa de nucleósido purina o batroxobina, un opioide, por ejemplo, endorfinas, encefalinas u opioides no naturales, una hormona o neuropéptido, por ejemplo, calcitonina, glucagón, gastrinas, hormona adrenocorticotrópica (ACTH), colecistoquininas, hormona luteinizante, hormona de liberación de gonadotropina, gonadotropina coriónica, factor de liberación de corticotropina, vasopresina, oxitocina, hormonas antiuréticas, hormona estimulante de la tiroides, hormona liberadora de tirotrópina, relaxina, prolactina, péptido YY, neuropéptido Y, polipéptido pancreático, leptina, CART (transcrito regulado por cocaína y anfetamina), un péptido relacionado con CART, perilipina, melanocortinas (hormonas estimulantes de melanocitos), tales como MC-4, hormonas que concentran melanina, péptidos natriuréticos, adrenomedulina, endotelina, secretina, amilina, péptido intestinal vasoactivo (VIP), polipéptido activador de la adenilato ciclasa de la pituitaria (PACAP), bombesina, péptidos similares a bombesina, timosina, proteína que se une a heparina, CD4 soluble, factor de liberación hipotalámico y melanotoninas.
- En otra realización de la invención, la proteína diana puede ser péptidos agonistas o antagonistas del receptor de insulina u otros péptidos diseñados para interactuar con otros receptores de la membrana celular.
- Una realización de la presente invención incluirá la coexpresión de la enzima de procesamiento marcada de acuerdo con la invención y el precursor de la proteína diana, facilitando de este modo el procesamiento *in vivo* del precursor de la proteína diana y una eliminación facilitada de la enzima de procesamiento durante las etapas iniciales de la purificación.
- La enzima de procesamiento marcada expresada, producida por el método de acuerdo con la invención se puede recuperar del medio de cultivo mediante procedimientos convencionales que incluyen la separación de las células hospedadoras del medio mediante centrifugación o filtración, la liberación de la enzima de procesamiento marcada mediante rotura mecánica de las células, tal mediante ultrasonidos o presión, la precipitación de los componentes proteínicos del material sobrenadante o el filtrado por medio de una sal, por ejemplo, sulfato de amonio.
- Cualquier matriz de intercambio catiónico adecuada, disponible procedente de proveedores comerciales se puede utilizar en el método de acuerdo con la invención y una lista no limitante de material para la columna de intercambio catiónico adecuada es: SP Sefarosa Fast Flow, SP Sefarosa XL, Streamline SP XL, Streamline Direct CST, Obelix SP, S-Support Unosphere, SP Sefarosa High Performance, Source30S y Toyopearl SP650S TosoHaas.
- Las etapas siguientes a la etapa de procesamiento, por ejemplo, la escisión de una proteína de fusión, pueden incluir una purificación en columna de intercambio catiónico como en la primera etapa. Las etapas de purificación después de la escisión pueden comprender también otros medios de purificación tales como cromatografía de intercambio aniónico, cromatografía de interacción hidrófoba y cromatografía de filtración en gel (véase, por ejemplo, Scopes, R., Protein Purification, Springer-Verlag, N.Y., 1982).
- La estructura artificial de ácido nucleico que codifica la enzima de procesamiento marcada de acuerdo con la invención puede ser de forma adecuada de origen genómico o ADNc, por ejemplo, obtenida preparando una genoteca genómica o de ADNc y detectando secuencias de ADN que codifican toda o parte de la proteína de fusión mediante hibridación, usando sondas de oligonucleótidos sintéticas de acuerdo con técnicas convencionales (cf. Sambrook et al, Molecular Cloning: A Laboratory Manual, 2ª ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, Nueva York, 1989).
- La estructura artificial de ácido nucleico que codifica la enzima de procesamiento marcada también se puede preparar sintéticamente mediante métodos convencionales establecidos, por ejemplo, el método de la fosforamidita descrito por Beaucage y Caruthers, Tetrahedron Letters 22 (1981), 1859-1869, o el método descrito por Matthes et al, EMBO Journal 3 (1984), 801-805. Según el método de la fosforamidita, los oligonucleótidos se sintetizan, por ejemplo, en un sintetizador de ADN automático, se purifican, se reasocian, se ligan y se clonan en vectores adecuados. Las secuencias de ADN que codifican la proteína de fusión también se pueden preparar por reacción en cadena de la polimerasa, tal como corte y empalme mediante PCR con extensión solapada usando cebadores específicos, por ejemplo, tal y como se describe en el documento US 4.683.202, Saiki et al, Science 239 (1988), 487-491, o Sambrook et al., supra.
- Además, la estructura artificial de ácido nucleico puede ser de origen mixto sintético y genómico, mixto sintético y ADNc o de origen mixto genómico y ADNc, preparada ligando fragmentos de origen sintético, genómico o de ADNc (según sea apropiado), en donde los fragmentos se corresponden a varias partes de la estructura artificial de ácido nucleico completa, de acuerdo con técnicas convencionales.
- Las secuencias de ADN que codifican la enzima de procesamiento marcada, por lo general se insertan en un vector recombinante que puede ser cualquier vector, que se puede someter convenientemente a procedimientos de ADN recombinante, y la elección del vector dependerá frecuentemente de la célula hospedadora en la que se va a introducir. Por lo tanto, el vector puede ser un vector con replicación autónoma, es decir, un vector, que existe como una entidad extracromosómica, cuya replicación es independiente de la replicación cromosómica, por ejemplo, un plásmido. Alternativamente, el vector puede ser uno que, cuando se introduce en una célula hospedadora, se integra

en el genoma de la célula hospedadora y se replica junto con el(los) cromosoma(s) en el(los) que ha sido integrado.

El vector es preferiblemente un vector de expresión en el que la secuencia de ADN que codifica la enzima de procesamiento marcada está ligada funcionalmente a segmentos adicionales requeridos para la transcripción del ADN. En general, el vector de expresión se obtiene a partir de ADN plasmídico o vírico, o puede contener elementos de ambos. La expresión, "ligado funcionalmente" indica que los segmentos están dispuestos de modo que actúan conjuntamente para sus fines previstos, por ejemplo, la transcripción se inicia en un promotor y procede a través de la secuencia de ADN que codifica la proteína de fusión.

Los vectores de expresión para uso en la expresión de la enzima de procesamiento marcada comprenderán un promotor capaz de dirigir la transcripción de un gen clonado o un ADNc. El promotor puede ser cualquier secuencia de ADN que muestre actividad transcripcional en la célula hospedadora seleccionada y se puede obtener a partir de genes que codifican proteínas que sean homólogas o heterólogas para la célula hospedadora.

Ejemplos de promotores adecuados para uso en células hospedadoras bacterianas incluyen el promotor del gen de la amilasa maltogénica de *Bacillus stearothermophilus*, el gen de la alfa-amilasa de *Bacillus licheniformis*, el gen de la amilasa BAN de *Bacillus amyloliquefaciens*, el gen de la proteasa alcalina de *Bacillus subtilis* o el gen de la xilosidasa de *Bacillus pumilus*, o los promotores P_R o P_L del fago Lambda o promotores usados para la expresión en *E. coli*, por ejemplo, lac, trp, phoA, araBAD, tac, bacteriófago T7 y cspA.

El vector también puede comprender un marcador seleccionable, por ejemplo, un producto génico que complementa un defecto en la célula hospedadora, tal como el gen que codifica la dihidrofolato reductasa (DHFR) o un gen marcador que confiere resistencia frente a un fármaco, por ejemplo, ampicilina, kanamicina, tetraciclina, cloranfenicol, neomicina, higromicina o metotrexato.

Las secuencias de ADN que codifican la enzima de procesamiento marcada también pueden estar conectadas funcionalmente, si es necesario, a un terminador adecuado, tal como el terminador de la hormona de crecimiento humana (Palmiter et al., Science 222, 1983, págs. 809-814) o el terminador TPI1 (Alber y Kawasaki, J. Mol. Appl. Gen. 1, 1982, págs. 419-434) o ADH3 (McKnight et al., The EMBO J. 4, 1985, págs. 2093-2099). Los vectores de expresión también pueden contener un conjunto de sitios de corte y empalme del ARN, localizados aguas abajo del promotor y aguas arriba del sitio de inserción para la secuencia del polipéptido de fusión mismo. Tales sitios de corte y empalme del ARN se pueden obtener a partir de genes de adenovirus y/o de inmunoglobulinas. Una señal de poliadenilación localizada aguas abajo del sitio de inserción también está contenida en los vectores de expresión. Ejemplos de señales de poliadenilación incluyen la señal de poliadenilación temprana o tardía de SV40 (Kaufman y Sharp, *ibid.*), la señal de poliadenilación procedente de la región de adenovirus 5 Elb, el terminador del gen de la hormona de crecimiento humana (DeNoto et al. Nucl. Acids Res. 9:3719-3730, 1981). Los vectores de expresión también pueden incluir una secuencia líder vírica no codificante, tal como el líder tripartito 2 de adenovirus, situado entre el promotor y los sitios de corte y empalme del ARN; y secuencias potenciadoras, tales como el potenciador de SV40.

La célula hospedadora en la que se introduce la estructura artificial de ADN que codifica la enzima de procesamiento marcada, puede ser cualquier célula bacteriana que sea capaz de producir la enzima de procesamiento marcada.

Ejemplos de células hospedadoras bacterianas son bacterias grampositivas tales como cepas de *Bacillus*, tales como cepas de *B. subtilis*, *B. licheniformis*, *B. lentus*, *B. brevis*, *B. stearothermophilus*, *B. alkalophilus*, *B. amyloliquefaciens*, *B. coagulans*, *B. circulans*, *B. lautus*, *B. megatherium* o *B. thuringiensis*, o cepas de *Streptomyces*, tales como *S. lividans* o *S. murinus*, o bacterias gramnegativas tales como cepas de *Escherichia coli*. La transformación de las bacterias se puede efectuar mediante transformación de protoplastos o usando células competentes de una manera conocida en sí misma (cf. Sambrook et al., supra).

Cuando una proteína se expresa en bacterias tales como *E. coli*, la proteína puede estar retenida en el citoplasma, típicamente como gránulos insolubles (conocidos como cuerpos de inclusión), o se puede dirigir al espacio periplásmico a través de una secuencia de secreción bacteriana. En el primer caso, las células se lisan y los gránulos se recuperan y se desnaturalizan, después de lo cual el polipéptido se repliega mediante dilución del agente desnaturalizante. En el último caso, la proteína diana se puede clonar con una secuencia de péptido señal fuerte tal como phoA, degQ, degS, degP, OmpA, OmpF, OmpH, OmpP, OmpT, lamb o pelB (de *Erwania carotovora*) y el polipéptido se puede recuperar a partir del espacio periplásmico mediante la rotura de las células, por ejemplo, por tratamiento con ultrasonidos o choque osmótico, para liberar el contenido del espacio periplásmico y recuperar el polipéptido.

A continuación, la célula hospedadora transformada o transfectada se cultiva en un medio nutriente adecuado en condiciones que permiten la expresión de la enzima de procesamiento marcada, después de lo cual todo o parte del péptido resultante se puede recuperar a partir del cultivo. El medio usado para cultivar las células puede ser cualquier medio convencional adecuado para el crecimiento de las células hospedadoras, tal como medios mínimos o complejos que contienen complementos apropiados. Los medios adecuados están disponibles en proveedores comerciales o se pueden preparar según recetas publicadas (por ejemplo, en catálogos de la American Type Culture Collection).

"Aplicación" significa una muestra que contiene la proteína de fusión que se carga en una columna de purificación.

"Flujo continuo" significa la parte de la aplicación que contiene proteínas de la célula hospedadora y contaminantes que no se unen a la columna de purificación.

"Pico principal" se refiere al pico en un cromatograma de purificación que tiene la mayor intensidad de UV y que contiene la proteína de fusión.

5 "mAU" es unidades de miliabsorbancia.

"Intensidad UV 280" es la absorbancia a una longitud de onda de 280 nm a la que se absorberán las proteínas, medida en unidades de miliabsorbancia.

"UV214" es la absorbancia a una longitud de onda de 214 nm a la que se absorberán las proteínas, medida en unidades de miliabsorbancia.

10 "IPTG" es isopropil-β-D-tiogalactopiranosido.

"TIC" es recuento iónico total.

"HPLC" es cromatografía líquida de alta resolución.

El término "proteína" incluirá tanto péptidos como polipéptidos.

15 Un "precursor de proteína diana" es una proteína que es un precursor de la proteína diana y que debe ser procesada con la enzima de procesamiento para obtener la proteína diana. Ejemplos incluyen una proteína de fusión a partir de la cual se tiene que eliminar la pareja de fusión a través de una enzima de procesamiento para obtener la proteína diana, proteínas a partir de las cuales se tienen que eliminar modificaciones postraduccionales (por ejemplo, glicosilaciones) con una enzima de procesamiento para obtener la proteína diana o proteínas a las que hay que añadir modificaciones postraduccionales mediante una enzima de procesamiento o combinaciones de las anteriores.

20 Una "proteína diana" es una proteína que ha sido procesada por una enzima de procesamiento para llegar a su forma deseada.

"SOE PCR" significa corte y empalme mediante PCR de extensión solapada.

"LC-MS" se refiere a espectrometría de masas con cromatografía líquida.

25 "% de solubilidad" se define como la cantidad de proteína soluble procedente de material lisado de la célula hospedadora, dividida por la cantidad de proteína de fusión soluble + insoluble procedente del material lisado de la célula hospedadora X 100.

"% de pureza" se define como la cantidad de proteína de interés, dividida por la cantidad de proteína de interés + la cantidad de contaminantes de la célula hospedadora X 100.

SDS-PAGE es electroforesis en gel de poliacrilamida con dodecil sulfato sódico.

30 En el presente contexto se ha empleado la designación de tres letras o de una letra de los aminoácidos con su significado convencional, como se indica en la tabla 1. Salvo que se indique explícitamente, los aminoácidos mencionados en esta memoria son L-aminoácidos. Además, los extremos izquierdo y derecho de una secuencia de aminoácidos de un péptido son, respectivamente, los extremos N-terminal y C-terminal, a menos que se especifique lo contrario.

35 **Tabla 1: Abreviaturas para aminoácidos:**

Aminoácido	Código de tres letras	Código de una letra
Glicina	Gly	G
Prolina	Pro	P
Alanina	Ala	A
Valina	Val	V
Leucina	Leu	L
Isoleucina	Ile	I
Metionina	Met	M
Cisteína	Cys	C
Fenilalanina	Phe	F
Tirosina	Tyr	Y

Aminoácido	Código de tres letras	Código de una letra
Triptófano	Trp	W
Histidina	His	H
Lisina	Lys	K
Arginina	Arg	R
Glutamina	Gln	Q
Asparagina	Asn	N
Ácido glutámico	Glu	E
Ácido aspártico	Asp	D
Serina	Ser	S
Treonina	Thr	T

Todos los títulos y subtítulos se utilizan en esta memoria únicamente para ofrecer una mayor claridad y no deben interpretarse de ningún modo como limitantes de la invención.

5 El uso de cualquiera y de todos los ejemplos, o el lenguaje de ejemplos (por ejemplo, "tal como") proporcionado en este documento, pretende meramente clarificar mejor la invención y no plantea una limitación del alcance de la invención a menos que se reivindique lo contrario. Ninguna expresión en la memoria descriptiva debería interpretarse como una indicación de que cualquier elemento no reivindicado sea esencial para la puesta en práctica de la invención.

10 Las citas y la incorporación de documentos de patentes en esta memoria se realizan solo por motivos prácticos y no refleja ningún punto de vista de la validez, la patentabilidad y/o la aplicabilidad de tales documentos de patentes.

Ejemplos

Ejemplo 1

Síntesis de una secuencia que codifica la proteasa HRV14 3C usando PCR de solapamiento

15 La secuencia de aminoácidos de la proteasa 3C madura de la cepa 14 de rinovirus humano (HRV14) es del modo siguiente:

GPNTEFALSLLRKNIMTITTSKGEFTGLGIHDRVCVIPHTHAQPGDDVLVNGQKIRVK
 DKYKLVDPENINLELTVLTLDRNEKFRDIRGFISEDLEGVDATLVVHSSNFTNTILEVGPVTMA
 GLINLSSTPTNRMIRYDYATKTGQCGVLCATGKIFGIHVGGNGRQGFSAQLKKQYFVEKQ
 (SEQ ID NO: 35)

20 Se generó una secuencia de ADN que codificaba SEQ ID NO: 35 y un codón optimizado para la expresión en *E. coli*. El fragmento génico sintético de la proteasa HRV14 3C madura se obtuvo mediante el uso de metodología de corte y empalme mediante PCR con extensión solapada, usando dos etapas consecutivas de PCR. En la primera reacción PCR, 16 cebadores solapantes (ocho cebadores directos y ocho cebadores inversos) se utilizaron para generar la secuencia completa de HRV14, incluyendo una región en el extremo 5' que codificaba un enlazador flexible (SEQ ID NO: 28). Los cebadores tenían ~50 pb de tamaño y el solapamiento entre los cebadores era de -20 pb. En la segunda reacción PCR, se emplearon cebadores terminales directos e inversos que comprendían los sitios de clonación XhoI y BamHI, respectivamente, para amplificar toda la secuencia usando el producto de la primera reacción PCR como molde. Se utilizó el sistema de PCR Phusion (Finnzymes). La primera reacción de PCR se configuró del modo siguiente con cantidades adecuadas de los 16 cebadores, esencialmente de acuerdo con las instrucciones del fabricante. Las condiciones de la PCR para la primera reacción fueron las siguientes;

30 98°C 30 s (desnaturalización inicial)
 98°C 10 s (desnaturalización)
 50°C 30 s (reasociación)
 72°C 15 s (elongación)
 durante 10 ciclos
 72°C 5 min (elongación final)

35 El producto de la PCR de la primera reacción se purificó directamente a partir de la solución de PCR, utilizando el kit GFX® PCR DNA and Gel Band Purification (GE Healthcare).

En la segunda reacción PCR, los cebadores directos e inversos más terminales se utilizaron para amplificar toda la secuencia que codificaba HRV14 3C a partir del producto de la PCR de la primera reacción. Las condiciones fueron las mismas que las descritas para la primera reacción, excepto que como temperatura de reasociación se empleó 55°C y tuvieron lugar 20 ciclos en lugar de 10.

- 5 El producto de la PCR de la segunda reacción tenía un tamaño de ~600 pares de bases como se esperaba, determinado por electroforesis en gel de agarosa. La banda de la PCR se escindió de los geles de agarosa y el producto de la PCR se purificó utilizando el kit GFX™ PCR DNA and Gel Band Purification (GE Healthcare) según las instrucciones del fabricante. El producto de la PCR purificado se subclonó en plásmidos usando sistemas de clonación TOPO establecidos (Invitrogen) según las instrucciones del fabricante y los productos de la ligación se transformaron en células TOP10 de *E. coli* (Invitrogen) para la propagación del plásmido. La secuencia correcta de insertos XhoI/BamHI de los plásmidos obtenidos a partir de clones aislados de TOP10, se verificó mediante secuenciación del ADN.

Ligación de la secuencia de HRV14 con diferentes marcadores proteicos alcalinos termoestables en vectores pET11a

- 15 El fragmento XhoI/BamHI que comprendía la región que codificaba la proteasa HRV14 3C se escindió de los vectores plasmídicos TOPO (Invitrogen). Los fragmentos XhoI/BamHI se ligaron a continuación en vectores pET11a (Novagen) de modo que los insertos que codificaban diferentes marcadores alcalinos termoestables N-terminales se enlazaron con el extremo 5' de la región que codificaba el enlazador y la proteasa. Se generaron dieciséis vectores pET11a que codificaban diferentes marcadores fusionados con el extremo N-terminal de la proteasa HRV14 3C con un enlazador intercalado, como se ejemplifica en la Figura 1.

La verificación de la secuencia de ADN se realizó mediante la escisión con enzimas de restricción y/o la secuenciación del ADN.

Se produjeron las siguientes enzimas marcadas:

Nombre del producto	Marcador	Enlazador	Enzima de procesamiento
pACSH240	SEQ ID NO: 1	SEQ ID NO: 28	SEQ ID NO: 35
pACSH329	SEQ ID NO: 3	SEQ ID NO: 28	SEQ ID NO: 35
pACSH319	SEQ ID NO: 6	SEQ ID NO: 28	SEQ ID NO: 35
pACSH320	SEQ ID NO: 8	SEQ ID NO: 28	SEQ ID NO: 35
pACSH321	SEQ ID NO: 9	SEQ ID NO: 28	SEQ ID NO: 35
pACSH238	SEQ ID NO: 10	SEQ ID NO: 28	SEQ ID NO: 35
pACSH322	SEQ ID NO: 11	SEQ ID NO: 28	SEQ ID NO: 35
pACSH323	SEQ ID NO: 12	SEQ ID NO: 28	SEQ ID NO: 35
pACSH324	SEQ ID NO: 13	SEQ ID NO: 28	SEQ ID NO: 35
pACSH325	SEQ ID NO: 15	SEQ ID NO: 28	SEQ ID NO: 35
pACSH326	SEQ ID NO: 16	SEQ ID NO: 28	SEQ ID NO: 35
pACSH239	SEQ ID NO: 19	SEQ ID NO: 28	SEQ ID NO: 35
pACSH327	SEQ ID NO: 20	SEQ ID NO: 28	SEQ ID NO: 35
pACSH328	SEQ ID NO: 21	SEQ ID NO: 28	SEQ ID NO: 35
pACSH331	SEQ ID NO: 24	SEQ ID NO: 28	SEQ ID NO: 35
pACSH332	SEQ ID NO: 27	SEQ ID NO: 28	SEQ ID NO: 35

25 **Expresión de las proteasas HRV14 3C marcadas**

La estructura artificial del vector pET11a que codificaba la proteasa HRV14 3C con diferentes marcadores N-terminales, se transformó en la cepa Rosetta (DE3) (Novagen) de *E. coli* mediante métodos convencionales basados en choque térmico o químicos. Las células transformadas se propagaron en placas de agar con medio LB (Luria-Bertani) que contenía 150 microgramos por mililitro de ampicilina y 34 microgramos por mililitro de cloranfenicol, durante una noche. El medio LB líquido con ampicilina y cloranfenicol fue inoculado con las placas de cultivo y se cultivó en matraces de agitación a 37 grados C y 240 rpm. La temperatura se bajó a 30 grados C cuando se alcanzó una DO₆₀₀: ~0,4-0,8. La inducción de la proteína se inició 20-30 minutos después, mediante la adición de IPTG 0,5 mM al medio de cultivo. La DO₆₀₀ final obtenida después de 3 horas de inducción estaba en el intervalo de 1,6-2,8.

30

Las células se recogieron mediante centrifugación. El SDS-PAGE de material lisado del sedimento que contenía células inducidas y no inducidas, así como fracciones solubles e insolubles de células inducidas (obtenidas por tratamiento con ultrasonidos y centrifugación en un tampón que consistía en NaPO₄ 25 mM pH 7), se realizó para evaluar el nivel de expresión y la solubilidad de las variantes de la proteasa HRV14 3C. Aunque se observaron diferencias en los niveles de expresión, se observó una solubilidad muy alta (~80% o más) para las versiones más marcadas de la proteasa HRV14 3C.

Purificación de la proteasa HRV14 3C marcada en posición N-terminal con SEQ ID NO: 19 (pACSH239)

El sedimento de cultivos celulares (procedente de 40 ml de cultivo) de la proteasa pACSH239 se rompió por tratamiento con ultrasonidos en un tampón que contenía NaPO₄ 25 mM pH 7. Los residuos celulares se centrifugaron a 10.000 g. El material sobrenadante se filtró de forma estéril antes de la purificación, que se llevó a cabo utilizando un sistema de purificación AKTA explorer 100 (GE Healthcare). Una columna de SP Sefarosa FF HiTrap preempaquetada con un volumen de columna de 5 ml (GE Healthcare) se utilizó para la separación con un caudal de 3 ml/min, usando los siguientes tampones:

Tampón A: fosfato de sodio 50 mM, pH 7 + EDTA 1 mM

Tampón B: fosfato de sodio 50 mM, pH 7 + NaCl 1 M + EDTA 1 mM

La columna se equilibró inicialmente con 7 volúmenes de columna de tampón A. Después de cargar la aplicación, la proteína no unida se eliminó por lavado usando 7 volúmenes de columna de tampón A. Se empleó un gradiente lineal de 0-100% de tampón B con 20 volúmenes de columna, para eluir la proteasa pACSH239 de la columna. La proteína de fusión se eluyó de la columna a una concentración salina de 0,4-0,5 M de NaCl. Algunas proteínas se perdieron en el flujo a través. Sin embargo, la proteasa pACSH239 se pudo purificar con pureza elevada (~85-90%) tal como se determinó mediante métodos convencionales de HPLC y SDS-PAGE, en una etapa aislada de cromatografía de intercambio catiónico. Mediante LC-MS también se verificó que la proteasa pACSH239 eluida tenía un peso molecular como se esperaba.

Purificación de la proteasa HRV14 3C marcada en posición N-terminal con SEQ ID NO: 10 (pACSH238)

El material sobrenadante listo para la purificación se preparó a partir de sedimentos de cultivos de células sometidas a ultrasonidos en NaPO₄ 25 mM pH 7, a partir de 80 ml de cultivo celular tal y como se ha descrito para pACSH239. Una columna HiTrap de SP Sefarosa FF preempaquetada de 5 ml (GE Healthcare) se cargó con la aplicación. El caudal era de 3 ml/min y los tampones utilizados para la purificación eran:

Tampón A: fosfato de sodio 50 mM, pH 7

Tampón B: fosfato de sodio 50 mM, pH 7 + NaCl 1 M

La columna se equilibró inicialmente con 7 volúmenes de columna de tampón A. La muestra no unida se eliminó en una etapa de lavado usando 7 volúmenes de columna de tampón A y la proteasa pACSH238 se eluyó de la columna utilizando un gradiente lineal de 0-100% de tampón B con 20 volúmenes de columna.

La proteasa pACSH238 se eluyó de la columna a una concentración de ~0,2 M de NaCl. En contraste con la proteasa pACSH239, la mayor parte de la proteasa pACSH238 se unía a la columna con estas condiciones de purificación. Mediante LC-MS también se verificó que la proteasa eluida tenía un peso molecular como el esperado. La recuperación fue 4 veces mayor después de la primera etapa de purificación, en comparación con la proteasa pACSH239 y la pureza era del ~90% tal y como se determinó con HPLC y SDS PAGE.

Evaluación de la actividad de la proteasa HRV14_3C (pACSH239) marcada con SEQ ID NO: 19

La actividad de la proteasa pACSH239 se estimó utilizando diferentes cantidades de la proteasa para escindir una proteína de fusión (pACSH294) que comprendía el péptido antagonista del receptor de insulina S661 con un marcador N-terminal (SEQ ID NO: 19) y un enlazador intercalado con un sitio de escisión modificado para HRV14 3C SSSGGSEVLFFQ (SEQ ID NO: 36).

El péptido S661 tiene la siguiente secuencia: GSLDESFYDWFERQLGGGSGSSLEEEWAQIQCEVWGRGCPYSY (SEQ ID NO: 37).

Concentraciones proteicas comparables del sustrato de la proteína de fusión pACSH294 y la proteasa pACSH239 se calcularon mediante la integración de las áreas pico de la absorbancia UV214 nm procedentes del análisis de los cromatogramas de HPLC. La enzima y el sustrato se mezclaron en un volumen de reacción de 200 ul y se incubaron 6 o 12 horas a 25°C en un tampón de escisión que contenía Tris HCl 50 mM pH 7,5, NaCl 150 mM utilizando diferentes relaciones de enzima a sustrato (ajustadas según la concentración). Estudios de LC-MS mostraron que solo se observó el péptido S661 (SEQ ID NO: 37) y el marcador liberado (SEQ ID NO: 19) + el enlazador (SEQ ID NO: 36) después de 12 horas de incubación, utilizando una relación 1:100 de enzima a sustrato (Figura 2). Solo permanecieron cantidades muy pequeñas de la proteína de fusión después de 12 horas usando una relación de 1:200. Por tanto, los resultados indican que una relación enzima/sustrato de 1:100-1:200 es suficiente para el procesamiento completo de la proteína de fusión pACSH294 en una incubación durante una noche.

El péptido S661 liberado, presente en el producto de la digestión, se pudo purificar posteriormente hasta tener más del 98% de pureza usando una columna de cromatografía de intercambio aniónico (HiTrap Q sefarosa High Performance, 1 ml, GE Healthcare). Los tampones y las condiciones generales de purificación eran como se han descrito para la purificación de la proteasa pACSH239. Aunque el péptido S661 se unía a esta columna, la proteasa pACSH239 no se eluyó de la columna con ninguna concentración salina dentro del gradiente de elución, pero solo se pudo observar en el flujo a través tal y como se determinó por el análisis LC-MS de fracciones de flujo a través. Por lo tanto, la proteasa pACSH239 se eliminó de manera eficaz de la proteína diana en la segunda etapa de purificación.

Evaluación de la actividad de la proteasa HRV14 3C (pACSH238) marcada con SEQ ID NO: 10

Se utilizaron cantidades idénticas de proteasa pACSH238 o pACSH239 para procesar el sustrato de pACSH294 en un tampón que consistía en Tris HCl 50 mM pH 7,5, NaCl 150 mM, en un experimento paralelo. Las relaciones entre enzima y sustrato de 1:25 o 1:50 se evaluaron para las dos enzimas. La cantidad de péptido antagonista del receptor de la insulina S661 liberada a partir de la proteína de fusión pACSH294 después de 3 horas de incubación a 30°C, era exactamente la misma para las dos enzimas, lo que indica que tienen la misma actividad específica a pesar del diferente tamaño de los marcadores fijados al extremo N-terminal. Como con la proteasa pACSH239, la proteasa pACSH238 no se unió a la columna de intercambio aniónico, indicando de este modo que se podía eliminar de manera eficaz mediante una etapa de purificación adecuada, después de la digestión del precursor de la proteína diana.

Ejemplo 2

Clonación de las proteasas marcadas del virus jaspeado del tabaco

La proteasa madura del virus jaspeado del tabaco (TEV) tiene la siguiente secuencia de aminoácidos:

GESLFKGRDYNPISSTICHLTNE SDGHTTSLYGIGFGPFIITNKHLFRRNNGTLLVQ
 SLHG VFKVKN TTTLQQHLIDGRDMIIRMPKDFPPFPQKLFREPQREERICLVTTNFQTKSM
 SSMVSDTSCTFPSSDGIFWKHWIQT KDGC GSPLVSTRDGFIVGIHSASNFTNTN NYFTSVP
 KNFMELLTNQEAQQWVSGWRLNADSVLWGGHKVFMVKPEEPFQPVKEATQLMN (SEQ ID
 NO: 38)

Un inserto que comprendía la secuencia de TEV con una mutación estabilizadora (S219V) se obtuvo con la metodología convencional de PCR utilizando el sistema Phusion PCR (Finnzymes), cebadores directos e inversos que comprendían el sitio de clonación XhoI y BamHI, respectivamente, y un plásmido que codificaba TEV (Ser219Val) existente como molde. El extremo 5' del producto de la PCR comprendía una secuencia que codificaba un enlazador flexible (SEQ ID NO: 28).

El producto de la PCR se purificó por GFX (GE Healthcare) y se subclonó en vectores TOPO de acuerdo con las instrucciones de los fabricantes (Invitrogen) y como se ha descrito en el Ejemplo 1. La secuencia correcta que codificaba TEV, de los insertos XhoI/BamHI procedentes de plásmidos TOPO, fue verificada mediante secuenciación del ADN. Los insertos correctos se escindieron de los plásmidos TOPO con las enzimas XhoI/BamHI y se fusionaron con el extremo 3' de insertos que codificaban marcadores alcalinos termoestables en vectores de expresión pET11a (Novagen). Se generaron las siguientes estructuras artificiales:

Nombre del producto	Marcador	Enlazador	Enzima de procesamiento
pACSH307	SEQ ID NO: 10	SEQ ID NO: 28	SEQ ID NO: 38
pACSH308	SEQ ID NO: 18	SEQ ID NO: 28	SEQ ID NO: 38
pACSH309	SEQ ID: NO:19	SEQ ID NO: 28	SEQ ID NO: 38

Expresión de variantes de la proteasa TEV marcada

Los plásmidos que codificaban las proteasas pACSH307, pACSH308 y pACSH309 se transformaron en cepas Rosetta (DE3) (Novagen) de *E. coli*. Las células transformadas se propagaron durante una noche en placas de agar con medio LB (Luria-Bertani) que contenía 150 microgramos por mililitro de ampicilina y 34 microgramos por ml de cloranfenicol. Medio LB líquido con ampicilina y cloranfenicol fue inoculado con placas de cultivo y se cultivó en matraces de agitación a 37 grados C y 240 rpm. La temperatura se bajó a 30°C cuando se alcanzó una DO₆₀₀: ~ 0,4-0,5. Después de 30 minutos de adaptación a esta temperatura, se inició la inducción de proteínas añadiendo IPTG 0,5 mM al medio de cultivo. La DO₆₀₀ final obtenida después de 3 horas de inducción estaba en el campo de 1,6. Las células se recogieron por centrifugación. Se realizó un SDS-PAGE del material lisado procedente de sedimentos que contenían células inducidas y no inducidas, así como fracciones solubles e insolubles de células inducidas (obteni-

das por tratamiento con ultrasonidos y centrifugación en un tampón que consistía en NaPO₄ 10 mM pH 7), para evaluar el nivel de expresión y la solubilidad de las variantes de proteasas TEV. Los niveles de expresión fueron similares para las tres estructuras artificiales. Se estimó que la solubilidad más elevada era la de la variante TEV marcada con SEQ ID NO: 10 (~70%) (pACSH307).

5 **Purificación de la proteasa pACSH307**

La purificación de la proteasa pACSH307 a partir de material sobrenadante filtrado de forma estéril procedente de sedimentos de cultivos (de un cultivo de 80 ml), tratado con ultrasonidos en NaPO₄ 25 mM pH 7, se llevó a cabo usando el sistema AKTA Explorer 100 (GE Healthcare) con exactamente las mismas condiciones y tampones descritos para pACSH238 en el Ejemplo 1. Cuando la separación se realizó en una columna HiTrap de SP Sefarosa FF de 5 ml, un pico eluyó claramente a aproximadamente 0,25 M de NaCl. Un análisis LC-MS de la fracción procedente del punto más alto del pico, proporcionó una masa molecular exactamente tal y como se esperaba para esta estructura artificial y la pureza del pico principal fue del ~85 a 90%, como se determinó por análisis SDS-PAGE y HPLC

Evaluación de la actividad de la proteasa TEV (pACSH307) monomarcada con SEQ ID NO: 10

La actividad de la proteasa pACSH307 se estimó utilizando diferentes cantidades de la proteasa para escindir una proteína de fusión (pACSH248) que comprendía el péptido antagonista de insulina S661 (SEQ ID NO: 37), como se ha descrito en el ejemplo 1 con el mismo marcador N-terminal (SEQ ID NO: 19), pero con un enlazador que comprendía un sitio de reconocimiento para la escisión del virus jaspeado del tabaco, SSSGGSENLYFQ (SEQ ID NO: 39). Los experimentos de digestión se realizaron en un tampón de escisión que contenía Tris HCl 50 mM pH 8, EDTA 0,5 mM. Los estudios de LC-MS mostraron que después de incubar durante 8 horas con una relación de enzima y sustrato de 1:1, la mayor parte del péptido S661 se liberaba de la proteína de fusión pACSH248. Esto verifica que la proteasa TEV marcada se expresa y se purifica como una enzima activa y se puede utilizar para eliminar el marcador SEQ ID NO: 19 de la proteína de fusión pACSH248.

Ejemplo 3

Clonación y expresión de protGasa.

El gen de la transglutaminasa (tGASA) consiste en un péptido señal y un prodominio seguido de la secuencia de tGASA madura. Una protGASA (sin péptido señal, pero con un residuo Met N-terminal) se preparó mediante PCR a partir de ADN genómico de *Streptomyces mobaraensis* usando metodología convencional. El prodominio se modificó, de manera que un sitio de escisión de HRV14 3C se incluyó justo antes de una extensión N-terminal Gly53-Pro54 de la secuencia de tGASA madura. La secuencia de protGASA se clonó en el vector pET11a (pNNC130) usando técnicas de clonación convencionales y la proteína codificada por este plásmido tenía la siguiente secuencia (la secuencia de tGASA madura comienza en el residuo 55):

MDNGAGEETKSYAETYRLTADDVANINALNESAPAASSAGPSFRAPLEVLFGQPD
 SDDRVTTPPAEPLDRMPDPYRPSYGRAETVNNYIRKWQQVYSHRDGRKQQMTEEQREWL
 SYGCVGVTWVNSGQYPTNRLAFASFDEDRFKNELKNGRPRSGETRAEFEGRVAKESFDEE
 KGFQRAREVASVMNRALENAHDESAYLDNLKKELANGNDALRNEDARSPFYALSALRNTPSF
 KERNGGNHDP SRMKAVIYSKHFWSGQDRSSSADKRKYGDPDAFRPAPGTGLVDM SRDRN
 IPRSPTSPGEGFVNFYDYGWFGAQTEADAKTVWTHGNHYHAPNGSLGAMHVVYESKFRNW
 SEGYSDFDRGAYVITFIPKSWNTAPDKVKQGW(PSEQ ID NO 40).

El plásmido pNNC130 se transformó en BL21 (DE3) (Novagen) y se cultivó durante una noche en medio LB que contenía 150 microgramos por mililitro de ampicilina. Medio líquido LB que contenía ampicilina fue inoculado con placas de agar de cultivo durante una noche y se cultivó a 37°C a 240 rpm hasta que la DO₆₀₀ llegó a ~0,8. Después de 20 min de adaptación a 30°C, la expresión se indujo con IPTG 0,1 mM. Las células se recogieron después de 4 horas de inducción de las proteínas con una DO₆₀₀ final de aproximadamente 2. Los análisis con SDS-PAGE mostraban que la mayoría de la protGASA se encontraba en la fracción soluble en el material sobrenadante obtenido después de tratamiento con ultrasonidos en un tampón que contenía NaPO₄ 10 mM pH 8 y de centrifugación de los residuos celulares.

Purificación de protGASA:

Los sedimentos celulares procedentes de 200 ml de cultivo celular inducido se disolvieron en 40 ml de tampón NaPO₄ 10 mM pH 8. Después de tratar con ultrasonidos y centrifugar a 10.000 g, el material sobrenadante se filtró de forma estéril y el volumen de la aplicación final se ajustó a 50 ml con tampón NaPO₄ 10 mM pH 8.

La purificación se realizó usando la instrumentación descrita en el Ejemplo 1 y 2, empleando un caudal de 3 ml/min y los tampones siguientes:

Tampón A: tampón NaPO₄ 25 mM pH 8,0
 Tampón B: tampón NaPO₄ 25 mM pH 8,0, NaCl 0,5 M

5 Una columna de 5 ml de HiTrap Q Sefarosa High Performance (HP) (GE Healthcare) se equilibró con 7 volúmenes de columna (VC) de tampón A. La aplicación que contenía protGASA se cargó en la columna y la muestra no unida se eliminó por lavado con tampón A con 10 VC. Un gradiente lineal de 0-100% de tampón B con 30 volúmenes de columna se utilizó para eluir la protGASA. Un pico que contenía protGASA se eluyó de la columna con aproximadamente 0,05 M de concentración de NaCl. Las fracciones recogidas se analizaron mediante SDS PAGE y la identidad de protGASA se confirmó mediante análisis LC-MS, que proporcionó una masa muy próxima a la masa teórica esperada de la proteína

10 **Eliminación del prodominio usando la proteasa pACSH238 o pACSH239**

15 Las fracciones procedentes de la primera etapa de purificación que contenían protGASA se agruparon y el tampón se cambió a un tampón de escisión con Tris HCl 50 mM pH 7,5, NaCl 150 mM y DTT 1 mM utilizando Vivaspin 20 (10000 MWCO) (Satorius) para la diafiltración. La cantidad de protGASA purificada y proteasa pACSH239 se ajustó estimando las áreas de los picos de absorbancia a UV214 nm de los cromatogramas de HPLC. Las pruebas de escisión con relaciones diferentes entre la enzima y el sustrato (1:10, 1:50, 1:100 y 1:200) se realizaron con o sin DTT 1 mM en el tampón de escisión a 25°C durante una noche. Los experimentos mostraron que la proteasa pACSH239 era capaz de eliminar completamente el prodominio (para obtener una tGASA madura extendida con Gly-Pro) usando una relación enzima sustrato de 1:200 y DTT 1 mM en el tampón de disociación. La masa de la tGASA madura liberada y la protGASA se determinó utilizando LC-MS.

20 En pruebas similares, se observó que la proteasa pACSH238 también era capaz de eliminar por escisión de forma eficaz el prodominio de la protGASA.

Purificación de la tGASA madura:

25 La tGASA madura extendida con Gly-Pro obtenida después de la digestión con la proteasa pACSH239, se diluyó hasta 48 ml en NaPO₄ 25 mM pH 6, DTT 1 mM (tampón A) y se separó en una columna de 5 ml de SP sefarosa High Performance (HP) HiTrap (GE Healthcare) utilizando los siguientes tampones:

Tampón A: tampón NaPO₄ 25 mM pH 6,0 + DTT 1 mM
 Tampón B: tampón NaPO₄ 25 mM pH 6,0 + DTT 1 mM + NaCl 0,5 M

La purificación se realizó utilizando un caudal de 3 ml/min y una elución con gradiente lineal de 0-100% de tampón B con 30 volúmenes de columna.

30 Los cromatogramas UV280 de la purificación indicaron que un solo pico principal se eluía de la columna de SP Sefarosa HP con una concentración de NaCl, de aprox. 0,1 M. El análisis LC-MS confirmó que la identidad de la proteína eludida era tGASA madura. La proteasa pACSH239, debido a la carga elevada del marcador SEQ ID NO: 19 fusionado en el extremo N-terminal, eluía de la columna mucho más tarde que la tGASA madura y por lo tanto se podía eliminar de manera eficaz de la proteína diana después de una segunda etapa de purificación.

35 **Ejemplo 4**

Síntesis de una secuencia que codifica la proteasa HAV18 3C usando PCR de solapamiento:

La secuencia de aminoácidos de la proteasa 3C de la cepa 18 madura del virus de la Hepatitis A (HAV18) es del modo siguiente:

STLEIAGLVRKNLVQFGVGEKNGCVRWWMNALGVKDDWLLVPSHAYKFEKDYEM
 MEFYFNRGGTYYSISAGNVVIQSLDVGFDVVLKVPPTIPKFRDITEHFIKKGDVPRALNRLA
 TLVTTVNGTPMLISEGPLKMEEKATYVHKKNDGTTVDLTDQAWRGKGEGLPGMCGGALV
 SSNQSISQNAILGIHVAGGNSILVAKLVTQEMFQNIKKIESQ (SEQ ID NO: 45).

40 Se obtuvo un fragmento de ADN con codones optimizados del gen de la proteasa HAV18 3C madura usando metodología de corte y empalme mediante PCR de extensión solapada con dos etapas consecutivas de PCR, como se ha descrito en el ejemplo 1 para la proteasa HRV14 3C. El fragmento final con la secuencia correcta contenía un sitio XhoI en el extremo 5' y un sitio BamHI en el extremo 3', lo que permitió la ligación en vectores pET11a (Novagen) para que la parte que codificaba SEQ ID NO: 10 estuviera unida al extremo 5' de la región que codificaba el enlazador y la proteasa.

Nombre del producto	Marcador	Enlazador	Enzima de procesamiento
pACSH362	SEQ ID NO: 10	SEQ ID NO: 47	SEQ ID NO: 45

Expresión de las proteasas HAV18 3C (pACSH362) marcadas con SEQ ID NO: 10

La expresión en cepas Rosetta (DE3) de *E. coli* se realizó esencialmente como se ha descrito en el ejemplo 1. Se pudo conseguir un nivel de expresión similar a la proteasa HRV14 3C marcada con SEQ ID NO: 10 y al menos 80% de la proteína estaba en la fracción soluble después de la rotura celular y la centrifugación.

5 Purificación de la proteasa HAV18 3C (pACSH362) marcada con SEQ ID NO: 10

Los sedimentos de cultivos celulares (procedentes de 80 ml de cultivo) de la proteasa pACSH362 se rompieron por tratamiento con ultrasonidos en un tampón que contenía NaPO₄ 25 mM pH 7. Los restos celulares se centrifugaron mediante centrifugación a 10.000 g. El material sobrenadante se filtró de forma estéril antes de la purificación, que se llevó a cabo utilizando un sistema de purificación AKTA explorer 100 (GE Healthcare). Una columna de SP Sefarosa FF HiTrap preempaquetada con un volumen de columna de 5 ml (GE Healthcare) se utilizó para la separación con un caudal de 3 ml/min, usando los siguientes tampones:

Tampón A: fosfato de sodio 50 mM, pH 7

Tampón B: fosfato de sodio 50 mM, pH 7 + NaCl 1 M

La columna se equilibró inicialmente con 7 volúmenes de columna de tampón A. Después de cargar la aplicación, la proteína no unida se eliminó mediante lavado usando 7 volúmenes de columna de tampón A. Se utilizó un gradiente lineal de 0-100% de tampón B con 20 volúmenes de columna para eluir la proteasa pACSH362 de la columna. La proteasa pACSH362 se eluyó de la columna con una concentración salina de 0,2-0,3 M de NaCl. La mayor parte de la proteasa pACSH362 se unía a la columna con esas condiciones de purificación. La pureza de la proteasa pACSH362 eluida se estimó que era del ~85-90% según se determinó por métodos convencionales de HPLC y SDS-PAGE, después de una etapa aislada de cromatografía de intercambio catiónico. Mediante LC-MS también se verificó que la proteasa eluida tenía un peso molecular tal y como se esperaba.

Evaluación de la actividad de la proteasa HAV18 3C (pACSH362) marcada con SEQ ID NO: 10

La proteasa HAV18 3C (pACSH362) marcada se utilizó para procesar la proteína de fusión pACSH360 que comprendía el péptido antagonista del receptor de insulina S661 (SEQ ID NO: 37) con un marcador N-terminal (SEQ ID NO: 19) y un enlazador intercalado con un sitio de escisión modificado para HAV18 3C, SSSGGSELRTQ (SEQ ID NO: 49).

Una relación de enzima a sustrato de 1:10 o 1:50 de la proteasa pACSH362 y la proteína de fusión pACSH360 fue evaluada mediante LC-MS después de incubar durante 7 horas a 37°C. Se observó un fragmento que representaba el péptido S661 liberado en ambas relaciones, como en el ejemplo 1, mostrando que la enzima HAV18 marcada está activa.

Ejemplo 5**Clonación de la proteasa marcada similar a 3C del virus de la enfermedad hemorrágica del conejo**

Los marcadores proteicos para la purificación procedentes de la familia L9 de proteínas ribosómicas, procedentes de *Thermus thermophilus*, *Geobacillus kaustophilus*, *Thermosiphon melanesiensis*, B1429 y *Acidothermus cellulolyticus* 11 B se sometieron a optimización de codones para la expresión en *E. coli* y se obtuvieron como genes sintéticos a partir GENEART GmbH, Regensburg, Alemania. Los fragmentos génicos que codificaban los marcadores tenían un sitio NdeI en el extremo 5' y un sitio XhoI en el extremo 3'. Un fragmento con codón optimizado con un sitio XhoI en el extremo 5' y un sitio BamHI en el extremo 3' que codificaba la proteasa similar a 3C del virus de la enfermedad hemorrágica del conejo (RHDV) también se obtuvo (GENEART GmbH, Regensburg, Alemania).

Las siguientes estructuras artificiales fueron ensambladas mediante técnicas convencionales de clonación en pET11a, un vector de expresión de *E. coli*, con el fin de generar las siguientes variantes de la proteasa RHDV:

Nombre del producto	Marcador	Enlazador	Enzima de procesamiento
pACSH433	SEQ ID NO: 41	SEQ ID NO: 48	SEQ ID NO: 46
pACSH434	SEQ ID NO: 42	SEQ ID NO: 48	SEQ ID NO: 46
pACSH435	SEQ ID: NO:43	SEQ ID NO: 48	SEQ ID NO: 46
pACSH436	SEQ ID: NO:44	SEQ ID NO: 48	SEQ ID NO: 46

Expresión de la proteasa marcada similar a 3C del virus de la enfermedad hemorrágica del conejo

La expresión en cepas BL21 (DE3) de *E. coli* se realizó esencialmente como se ha descrito en el ejemplo 1. Se obtuvieron niveles de expresión muy similares para las cuatro estructuras artificiales y el intervalo de solubilidad de aproximadamente 80%-90% tal y como se determinó mediante SDS-PAGE del material sobrenadante y el sedimento

obtenidos después de la rotura celular y la centrifugación.

Purificación de las proteasas HAV18 3C marcadas con SEQ ID NO: 41, 42, 43 y 44

5 El sedimento de los cultivos celulares (procedente de 80 ml de cultivo) de cada una de las proteasas RHDV marcadas se rompió mediante tratamiento con ultrasonidos. El material sobrenadante se preparó y se separó utilizando una columna preempaquetada de SP Sefarosa FF de 5 ml (GE Healthcare) usando los mismos tampones y condiciones que los descritos anteriormente.

Después de comparar las cantidades recuperadas de proteasa después de la etapa de captura, el mejor resultado se logró con la proteasa pACSH433 que comprendía el marcador SEQ ID NO: 41, que unía de ~2 - 3,5 veces más proteína a la columna que las variantes restantes.

10 **Ejemplo 6**

Coexpresión de la variante de la proteasa HRV14 3C marcada junto con protGASA (pNNC130) o pACSH294 (proteína de fusión S661)

15 El plásmido pACYCDuet-1 (Novagen) tiene un número de copias menor que el plásmido pET11a (Novagen) y la cantidad de proteína producida a partir de este plásmido debe ser inferior que a partir del plásmido pET11a. Además, el plásmido contiene un marcador de selección diferente (resistencia a cloranfenicol) que el de pET11a, lo que permite el mantenimiento y la coexpresión de estos dos plásmidos en la misma cepa hospedadora. Se evaluó si un método para procesar precursores de proteínas diana que emplea enzimas de procesamiento marcadas era factible utilizando el vector pACYCDuet-1 para la expresión de la proteasa pACSH238 y el vector pET11a para la expresión de protGASA (pNNC130) o la proteína de fusión S661 (pACSH294) que se pueden procesar con la proteasa pACSH238.

20 Un fragmento procedente del plásmido pET11a que codificaba la proteasa pACSH238 (Ejemplo 1) y una parte de la estructura principal del vector aguas arriba del codón de inicio, se escindió con las enzimas de restricción MluI y BamHI y se ligó con la estructura principal del plásmido pACYCDuet-1 escindido con MluI/BamHI (Novagen).

25 El plásmido pACYCDuet-1 que expresaba la proteasa pACSH238 se transformó en BL21 (DE3) junto con el plásmido pET11a que codificaba protGASA (pNNC130) o la proteína de fusión S661 (pACSH294). Por lo tanto, una cepa puede coexpresar la proteasa pACSH238 (a partir del plásmido pACYCDuet-1) junto con la protGASA y la otra cepa expresa la proteasa pACSH238 junto con la proteína de fusión S661 (pACSH294).

30 Las cepas que se coexpresaban se cultivaron en medio LB que contenía 150 microgramos por mililitro de ampicilina y 34 microgramos por mililitro de cloranfenicol, de forma paralela con cepas de control que expresaban la protGASA o solo la proteína de fusión S661. Las condiciones de la expresión eran como se han descrito para la proteasa pACSH238 en el ejemplo 1. Después de 3 horas de expresión a 30°C, se recogieron las cepas que se coexpresaban y sus controles. Basándose en el SDS-PAGE, era evidente que la coexpresión daba como resultado un peso molecular menor para la protGASA inducida o la proteína de fusión S661, en comparación con los controles. La proteína troncada inducida presente en la banda del gel SDS procedente de la cepa que coexpresaba la proteasa pACSH238/protGASA fue identificada como tGASA usando métodos convencionales basados en espectrometría de masas, seguidos de digestión en gel. Por tanto, los resultados muestran que la proteasa pACSH238 coexpresada era capaz de separar por escisión el prodominio de la protGASA y el marcador de purificación SEQ ID NO: 19 de la proteína de fusión S661 *in vivo*.

LISTA DE SECUENCIAS

40 <110> Novo Nordisk A/S Shaw, Allan

<120> Enzimas de procesamiento

<130> 7490.204-WO

<160> 49

<170> PatentIn versión 3.3

45 <210> 1

< 211> 56

< 212> PRT

< 213> Bacillus stearothermophilus

ES 2 531 372 T3

<400> 1

Met Ser Lys Thr Ile Val Arg Lys Asn Glu Ser Ile Asp Asp Ala Leu
1 5 10 15

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

Arg Lys Arg Glu Phe Tyr Glu Lys Pro Ser Val Arg Arg Lys Lys Lys
35 40 45

Ser Glu Ala Ala Arg Lys Arg Lys
50 55

<210> 2

< 211> 50

< 212> PRT

< 213> Archaeoglobus fulgidus

5

<400> 2

Met Gly Lys Lys Thr Val Gly Val Lys Lys Arg Leu Ala Lys Ala Tyr
1 5 10 15

Lys Gln Asn Arg Arg Ala Pro Val Trp Ile Thr Val Lys Thr Lys Arg
20 25 30

Ser Val Phe Gly Ser Pro Lys Arg Arg His Trp Arg Arg Ser Lys Leu
35 40 45

Lys Val
50

<210> 3

< 211> 44

< 212> PRT

< 213> Thermotoga maritima

10

<400> 3

Met Lys Arg Thr Tyr Gln Pro Ser Arg Arg Lys Arg Lys Arg Thr His
1 5 10 15

Gly Phe Leu Ala Arg Lys Arg Thr Pro Gly Gly Arg Arg Val Leu Lys
20 25 30

Asn Arg Arg Arg Lys Gly Arg Trp Arg Leu Thr Val
35 40

15

<210> 4

< 211> 27

< 212> PRT

< 213> Thermus thermophilus

<400> 4

Met Gly Lys Gly Asp Arg Arg Thr Arg Arg Gly Lys Ile Trp Arg Gly
1 5 10 15

Thr Tyr Gly Lys Tyr Arg Pro Arg Lys Lys Lys
20 25

20

ES 2 531 372 T3

<210> 5
<211> 67
<212> PRT
<213> Thermotoga maritima

5

<400> 5
Met Ala Lys Val Lys Met Lys Thr Asn Arg Ser Ala Ala Lys Arg Phe
1 5 10 15

Lys Val Thr Ala Lys Gly Lys Ile Lys Arg Trp Lys Ser Gly Gly Ala
 20 25 30

His Tyr Asn Thr Lys Lys Ser Ser Lys Arg Lys Arg His Leu Arg Lys
 35 40 45

His Thr Tyr Val Lys Asp Asn Met Leu Lys His Val Lys Ala Leu Leu
 50 55 60

Lys Glu Phe
65

10

<210> 6
<211> 233
<212> PRT
<213> Thermotoga maritima

ES 2 531 372 T3

<400> 6

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

5

<210> 7

<211> 235

<212> PRT

<213> *Thermotoga maritima*

ES 2 531 372 T3

<400> 7

Met Ala Gln Val Asp Leu Leu Asn Val Lys Gly Glu Lys Val Gly Thr
 1 5 10 15

Leu Glu Ile Ser Asp Phe Val Phe Asn Ile Asp Pro Asn Tyr Asp Val
 20 25 30

Met Trp Arg Tyr Val Asp Met Gln Leu Ser Asn Arg Arg Ala Gly Thr
 35 40 45

Ala Ser Thr Lys Thr Arg Gly Glu Val Ser Gly Gly Gly Arg Lys Pro
 50 55 60

Trp Pro Gln Lys His Thr Gly Arg Ala Arg His Gly Ser Ile Arg Ser
 65 70 75 80

Pro Ile Trp Arg His Gly Gly Val Val His Gly Pro Lys Pro Arg Asp
 85 90 95

Trp Ser Lys Lys Leu Asn Lys Lys Met Lys Lys Leu Ala Leu Arg Ser
 100 105 110

Ala Leu Ser Val Lys Tyr Arg Glu Asn Lys Leu Leu Val Leu Asp Asp
 115 120 125

Leu Lys Leu Glu Arg Pro Lys Thr Lys Ser Leu Lys Glu Ile Leu Gln
 130 135 140

Asn Leu Gln Leu Ser Asp Lys Lys Thr Leu Ile Val Leu Pro Trp Lys
 145 150 155 160

Glu Glu Gly Tyr Met Asn Val Lys Leu Ser Gly Arg Asn Leu Pro Asp
 165 170 175

Val Lys Val Ile Ile Ala Asp Asn Pro Asn Asn Ser Lys Asn Gly Glu
 180 185 190

Lys Ala Val Arg Ile Asp Gly Leu Asn Val Phe Asp Met Leu Lys Tyr
 195 200 205

Asp Tyr Leu Val Leu Thr Arg Asp Met Val Ser Lys Ile Glu Glu Val
 210 215 220

Leu Gly Asn Glu Ala Gly Lys Ala Leu Thr Ala
 225 230 235

5

<210> 8

<211> 184

<212> PRT

<213> *Thermotoga maritima*

ES 2 531 372 T3

<400> 8

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

5

<210> 9

<211> 184

<212> PRT

<213> Thermotoga maritima

ES 2 531 372 T3

<400> 9

Met Ser Arg Leu Ala Lys Lys Pro Ile Val Leu Pro Gln Gly Val Thr
 1 5 10 15

Val Glu Ile Lys Asp Asn Val Val Lys Val Lys Gly Pro Lys Gly Glu
 20 25 30

Leu Ser Gln Glu Phe Leu Pro Tyr Val Lys Ile Glu Val Glu Gly Asn
 35 40 45

Glu Val Trp Val Arg Pro Asn Glu Glu Gln Ile Ile Arg Lys Ser Asp
 50 55 60

Trp Arg Lys Val Lys Met Phe Gln Gly Thr Tyr Trp Ser Leu Ile Arg
 65 70 75 80

Asn Met Val Val Gly Val Thr Glu Gly Tyr Lys Lys Glu Leu Glu Ile
 85 90 95

Val Gly Ile Gly Tyr Arg Ala Gln Leu Gln Gly Asn Thr Leu Val Met
 100 105 110

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

Lys Ile Glu Val Pro Ala Pro Asn Arg Ile Ile Val Ser Gly Ile Asp
 130 135 140

Lys Gln Arg Val Gly Gln Val Ala Ala Glu Ile Arg Ala Phe Arg Pro
 145 150 155 160

Pro Asn Val Tyr Thr Gly Lys Gly Ile Arg Tyr Val Gly Glu Val Val
 165 170 175

Arg Gln Lys Glu Gly Lys Lys Ala
 180

5

<210> 10

<211> 149

<212> PRT

<213> Thermotoga maritima

ES 2 531 372 T3

<400> 10

Met Lys Val Ile Leu Leu Arg Asp Val Pro Lys Ile Gly Lys Lys Gly
1 5 10 15

Glu Ile Lys Glu Val Ser Asp Gly Tyr Ala Arg Asn Tyr Leu Ile Pro
20 25 30

Arg Gly Phe Ala Lys Glu Tyr Thr Glu Gly Leu Glu Arg Ala Ile Lys
35 40 45

His Glu Lys Glu Ile Glu Lys Arg Lys Lys Glu Arg Glu Arg Glu Glu
50 55 60

Ser Glu Lys Ile Leu Lys Glu Leu Lys Lys Arg Thr His Val Val Lys
65 70 75 80

Val Lys Ala Gly Glu Gly Gly Lys Ile Phe Gly Ala Val Thr Ala Ala
85 90 95

Thr Val Ala Glu Glu Ile Ser Lys Thr Thr Gly Leu Lys Leu Asp Lys
100 105 110

Arg Trp Phe Lys Leu Asp Lys Pro Ile Lys Glu Leu Gly Glu Tyr Ser
115 120 125

Leu Glu Val Ser Leu Pro Gly Gly Val Lys Asp Thr Ile Lys Ile Arg
130 135 140

Val Glu Arg Glu Glu
145

5

<210> 11

< 211> 179

< 212> PRT

< 213> *Thermotoga maritima*

ES 2 531 372 T3

<400> 11

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

5

<210> 12

<211> 149

<212> PRT

<213> *Thermotoga maritima*

ES 2 531 372 T3

<400> 12

Met Ala Arg Tyr Phe Pro Val Gln Lys Thr Thr Met Ile Lys Pro Glu
1 5 10 15

Glu Val Glu Arg Lys Trp Tyr Val Val Asp Ala Ser Gly Lys Val Leu
20 25 30

Gly Arg Leu Ala Thr Arg Ile Ala Lys Ile Leu Met Gly Lys His Lys
35 40 45

Pro Asn Tyr Thr Pro His Val Asp Thr Gly Asp Tyr Val Ile Val Val
50 55 60

Asn Ala Asp Lys Val Val Leu Thr Gly Lys Lys Leu Asp Gln Lys Val
65 70 75 80

Tyr Tyr Trp His Ser Gly Tyr Pro Gly Gly Leu Lys Ser Leu Thr Ala
85 90 95

Arg Gln Met Leu Glu Lys His Pro Glu Arg Leu Ile Trp Leu Ala Val
100 105 110

Lys Arg Met Leu Pro Lys Asn Arg Lys Gly Arg Lys Met Leu Lys Arg
115 120 125

Leu Lys Val Tyr Ala Ser Pro Glu His Pro His Gln Ala Gln Lys Pro
130 135 140

Glu Pro Ile Glu Leu
145

<210> 13

<211> 147

<212> PRT

<213> *Thermotoga maritima*

5

ES 2 531 372 T3

<400> 13

Met Arg Leu Glu Asp Leu Arg Pro Thr Pro Gly Ala Met Lys Lys Arg
1 5 10 15

Lys Arg Val Gly Arg Gly Pro Gly Ser Gly His Gly Lys Thr Ser Gly
20 25 30

Arg Gly His Lys Gly Gln Lys Ala Arg Gly Ser Gly Lys Val His Ile
35 40 45

Trp Phe Glu Gly Gly Gln Thr Pro Leu Gln Arg Arg Leu Pro Lys Arg
50 55 60

Gly Phe Lys Asn Ile Asn Lys Lys Val Tyr Ala Val Val Asn Val Lys
65 70 75 80

Val Leu Glu Glu Arg Phe Glu Ala Asn Glu Glu Val Thr Pro Glu Lys
85 90 95

Leu Ile Glu Arg Lys Ile Ile Lys Asp Leu Lys Asp Gly Val Lys Ile
100 105 110

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

Phe Ser Lys Ser Ala Val Glu Lys Ile Glu Ser Ala Gly Gly Lys Ala
130 135 140

Glu Val Ile
145

<210> 14

<211> 131

<212> PRT

<213> *Thermotoga maritima*

5

ES 2 531 372 T3

<400> 17

Met Leu Tyr Ala Ile Val Glu Thr Ala Gly Arg Gln Tyr Arg Val Glu
1 5 10 15

Glu Gly Lys Ile Leu Tyr Thr Glu Lys Gln Lys Asp Tyr Ser Pro Gly
20 25 30

Asp Glu Ile Val Phe Asp Arg Val Val Phe Val Arg Lys Asp Gly Glu
35 40 45

Val Leu Val Gly Lys Pro Tyr Val Glu Gly Ala Lys Val Val Gly Lys
50 55 60

Val Leu Glu His Ala Lys Ala Arg Lys Val Lys Thr Val Lys Tyr Arg
65 70 75 80

Pro Arg Lys Asn Ser Lys Val Glu Lys Gly His Arg Gln Trp Tyr Thr
85 90 95

Ala Ile Lys Ile Glu Lys Ile Glu Leu
100 105

<210> 18

<211> 100

<212> PRT

<213> *Thermotoga maritima*

5

<400> 18

Met Lys Gln Glu Lys Leu Ser Leu His Asp Val Leu Ile Arg Pro Ile
1 5 10 15

Ile Thr Glu Lys Ala Leu Ile Leu Arg Glu Gln Arg Lys Tyr Val Phe
20 25 30

Glu Val Asn Pro Leu Ala Asn Lys Asn Leu Val Lys Glu Ala Val Glu
35 40 45

Lys Leu Phe Asn Val Lys Val Glu Lys Val Asn Ile Leu Asn Met Lys
50 55 60

Pro Lys Pro Lys Arg Arg Gly Ile Phe Glu Gly Lys Thr Arg Ser Trp
65 70 75 80

Lys Lys Ala Val Val Thr Leu Lys Glu Gly Tyr Thr Ile Lys Glu Leu
85 90 95

Glu Gly Glu His
100

10

<210> 19

<211> 83

<212> PRT

<213> *Thermotoga maritima*

ES 2 531 372 T3

<400> 19

Met Ala His Lys Lys Ser Gly Gly Val Ala Lys Asn Gly Arg Asp Ser
1 5 10 15

Leu Pro Lys Tyr Leu Gly Val Lys Val Gly Asp Gly Gln Ile Val Lys
20 25 30

Ala Gly Asn Ile Leu Val Arg Gln Arg Gly Thr Arg Phe Tyr Pro Gly
35 40 45

Lys Asn Val Gly Met Gly Arg Asp Phe Thr Leu Phe Ala Leu Lys Asp
50 55 60

Gly Arg Val Lys Phe Glu Thr Lys Asn Asn Lys Lys Tyr Val Ser Val
65 70 75 80

Tyr Glu Glu

<210> 20

< 211> 66

< 212> PRT

< 213> *Thermotoga maritima*

5

<400> 20

Met Lys Ala Ser Glu Leu Arg Asn Tyr Thr Asp Glu Glu Leu Lys Asn
1 5 10 15

Leu Leu Glu Glu Lys Lys Arg Gln Leu Met Glu Leu Arg Phe Gln Leu
20 25 30

Ala Met Gly Gln Leu Lys Asn Thr Ser Leu Ile Lys Leu Thr Lys Arg
35 40 45

Asp Ile Ala Arg Ile Lys Thr Ile Leu Arg Glu Arg Glu Leu Gly Ile
50 55 60

Arg Arg
65

10

<210> 21

< 211> 67

< 212> PRT

< 213> *Thermotoga maritima*

ES 2 531 372 T3

<400> 21

Met Pro Lys Lys Leu Lys Ile Lys Leu Val Lys Ser Pro Ile Gly Tyr
1 5 10 15

Ser Trp Asp Gln Lys Asp Thr Val Lys Arg Leu Gly Leu Lys Lys Leu
20 25 30

Asn Gln Val Val Ile Lys Asp Asp Leu Pro Gln Ile Arg Gly Met Ile
35 40 45

Arg Lys Val Lys His Leu Val Glu Val Glu Glu Ile Glu Glu Gly Gly
50 55 60

Ser Asn Ala
65

<210> 22

< 211> 65

< 212> PRT

< 213> *Thermotoga maritima*

5

<400> 22

Met Pro Lys Val Lys Thr Asn Arg Ser Ala Ala Lys Arg Phe Arg Ile
1 5 10 15

Thr Lys Asn Gly Lys Ile Met Arg Asn His Ala Tyr Arg Ser His Lys
20 25 30

Thr Gly Lys Lys Arg Arg Asn Ala Leu Arg Ala Leu Arg Lys Lys Asp
35 40 45

Val Val Ser Ser Ala Asp Lys Asn Arg Val Leu Arg Leu Leu Gly Lys
50 55 60

Lys
65

10

<210> 23

< 211> 209

< 212> PRT

< 213> *Thermotoga maritima*

ES 2 531 372 T3

<400> 23

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

Asp Trp Gln Ala Lys Trp Phe Asn Glu Lys Asn Tyr Lys Glu Trp Leu
20 25 30

Leu Glu Asp Glu Glu Ile Arg Lys Ile Ile Lys Asn Lys Tyr Tyr His
35 40 45

Ala Gly Ile Ser Glu Ile Tyr Val Glu Arg Pro Asp Ala Glu Arg Ile
50 55 60

Asn Ile Thr Val Lys Thr Ala Arg Pro Gly Ile Ile Ile Gly Arg Lys
65 70 75 80

Gly Ser Glu Ile Thr Ser Leu Arg Glu Glu Leu Glu Arg Lys Phe Asn
85 90 95

Arg Arg Val Val Ile Asn Ile Glu Glu Ile Lys Thr Pro Glu Leu Asp
100 105 110

Ala Gln Leu Val Ala Glu Ser Ile Ala Ser Arg Ile Glu Lys Arg Ala
115 120 125

Ser Tyr Lys Val Ala Met Lys Arg Ala Ile Met Asn Ala Met Arg Lys
130 135 140

Gly Ala Gln Gly Ile Lys Val Met Val Ala Gly Arg Leu Gly Gly Ala
145 150 155 160

Glu Ile Ala Arg Arg Glu Trp Tyr Leu Arg Gly Arg Leu Pro Leu Gln
165 170 175

Lys Ile Lys Ala Ile Ile Asp Tyr Gly Thr Ala Thr Ala Trp Thr Lys
180 185 190

Tyr Gly Thr Ile Gly Ile Lys Val Trp Ile Tyr Lys Gly Asp Ala Asp
195 200 205

Ile

5

<210> 24

<211> 178

<212> PRT

<213> *Thermotoga maritima*

ES 2 531 372 T3

<400> 24

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

5

<210> 25

< 211> 90

< 212> PRT

< 213> *Thermotoga maritima*

ES 2 531 372 T3

<400> 25

Met Val Ser Leu Asp Pro Glu Lys Lys Asn Glu Ile Ile Lys Glu Phe
1 5 10 15

Gln Ile His Glu Asn Asp Thr Gly Ser Val Glu Val Gln Ile Ala Leu
20 25 30

Leu Thr Ala Arg Ile Lys His Leu Thr Glu His Leu Arg Lys His Pro
35 40 45

Lys Asp Phe His Ser Arg Arg Gly Leu Met Lys Met Ile Gly Arg Arg
50 55 60

Arg Lys Met Leu Lys Tyr Leu Arg His Lys Lys Pro Glu Val Tyr Arg
65 70 75 80

Glu Leu Ile Ala Lys Leu Gly Ile Arg Lys
85 90

<210> 26

<211> 95

<212> PRT

<213> *Thermotoga maritima*

5

<400> 26

Met Gly Arg Ser Arg Lys Lys Gly Pro Tyr Val Asp Arg Lys Leu Leu
1 5 10 15

Glu Lys Ile Arg Lys Leu Asn Glu Thr Gly Glu Lys Lys Val Ile Lys
20 25 30

Thr Trp Ser Arg Ala Ser Met Ile Ile Pro Glu Met Val Gly His Thr
35 40 45

Ile Ala Val Tyr Asn Gly Met Lys His Ile Pro Val Tyr Ile Thr Glu
50 55 60

Asn Met Ile Gly His Arg Leu Gly Glu Phe Ala Pro Thr Arg Arg Phe
65 70 75 80

Gly Gly His Ala Asp Lys Lys Ala Lys Lys Gly Glu Leu Lys Lys
85 90 95

10

<210> 27

<211> 96

<212> PRT

<213> *Thermotoga maritima*

ES 2 531 372 T3

<400> 27

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

Arg Arg Leu Arg Asn Lys Ala Tyr Lys Thr Phe Phe Lys Asn Arg Ile
20 25 30

Lys Glu Val Leu Lys Ala Ile Glu Asn Lys Glu Pro Lys Glu Val Val
35 40 45

Leu Glu Leu Thr Arg Lys Ala Gln Ala Ala Ile Asp Lys Ala Val Ser
50 55 60

Lys Gly Val Ile His Lys Asn Gln Gly Ala Arg Arg Lys Ala Arg Leu
65 70 75 80

Phe Glu Lys Val Asn Glu Tyr Leu Arg Thr Leu Glu Thr Thr Gln Glu
85 90 95

<210> 28

< 211> 12

< 212> PRT

< 213> Artificial

5

<220>

< 223> Enlazador 1

<400> 28

Ser Ser Ser Gly Ser Ser Gly Ser Ser Gly Ser Ser
1 5 10

10

<210> 29

< 211> 8

< 212> PRT

< 213> Artificial

15

<220>

< 223> Enlazador 2

<400> 29

Gly Gly Ser Ser Gly Gly Ser Ser
1 5

<210> 30

< 211> 8

< 212> PRT

< 213> Artificial

20

<220>

< 223> Enlazador 3

<400> 30

Ser Ser Ser Gly Ser Gly Ser Gly
1 5

25

<210> 31

< 211> 7

< 212> PRT

< 213> Artificial

30

ES 2 531 372 T3

<220>
< 223> Enlazador 4

<400> 31
Ala Leu Ala Leu Ala Leu Ala
1 5

5 <210> 32
< 211> 9
< 212> PRT
< 213> Artificial

10 <220>
< 223> Enlazador 5

<400> 32
Ala Leu Ala Leu Ala Leu Ala Pro Ala
1 5

15 <210> 33
< 211> 10
< 212> PRT
< 213> Artificial

<220>
< 223> Enlazador 6

<400> 33
Ser Ser Ser Ala Leu Ala Leu Ala Leu Ala
1 5 10

20 <210> 34
< 211> 9
< 212> PRT
< 213> Artificial

25 <220>
< 223> Enlazador 7

<400> 34
Ser Gly Ser Gly Ser Gly Ser Gly Ser
1 5

30 <210> 35
< 211> 182
< 212> PRT
< 213> Rinovirus humano 14

ES 2 531 372 T3

<400> 35

Gly Pro Asn Thr Glu Phe Ala Leu Ser Leu Leu Arg Lys Asn Ile Met
 1 5 10 15

Thr Ile Thr Thr Ser Lys Gly Glu Phe Thr Gly Leu Gly Ile His Asp
 20 25 30

Arg Val Cys Val Ile Pro Thr His Ala Gln Pro Gly Asp Asp Val Leu
 35 40 45

Val Asn Gly Gln Lys Ile Arg Val Lys Asp Lys Tyr Lys Leu Val Asp
 50 55 60

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

Glu Lys Phe Arg Asp Ile Arg Gly Phe Ile Ser Glu Asp Leu Glu Gly
 85 90 95

Val Asp Ala Thr Leu Val Val His Ser Asn Asn Phe Thr Asn Thr Ile
 100 105 110

Leu Glu Val Gly Pro Val Thr Met Ala Gly Leu Ile Asn Leu Ser Ser
 115 120 125

Thr Pro Thr Asn Arg Met Ile Arg Tyr Asp Tyr Ala Thr Lys Thr Gly
 130 135 140

Gln Cys Gly Gly Val Leu Cys Ala Thr Gly Lys Ile Phe Gly Ile His
 145 150 155 160

Val Gly Gly Asn Gly Arg Gln Gly Phe Ser Ala Gln Leu Lys Lys Gln
 165 170 175

Tyr Phe Val Glu Lys Gln
 180

5

<210> 36
 <211> 11
 <212> PRT
 <213> Rinovirus humano 14

10

<400> 36
 Ser Ser Ser Gly Gly Ser Glu Val Leu Phe Gln
 1 5 10

<210> 37
 <211> 43
 <212> PRT
 <213> Homo sapiens

ES 2 531 372 T3

<400> 37

Gly Ser Leu Asp Glu Ser Phe Tyr Asp Trp Phe Glu Arg Gln Leu Gly
1 5 10 15

Gly Gly Ser Gly Gly Ser Ser Leu Glu Glu Glu Trp Ala Gln Ile Gln
20 25 30

Cys Glu Val Trp Gly Arg Gly Cys Pro Ser Tyr

35 40

5

<210> 38

< 211> 236

< 212> PRT

< 213> Virus del jaspeado del tabaco

<400> 38

Gly Glu Ser Leu Phe Lys Gly Pro Arg Asp Tyr Asn Pro Ile Ser Ser
1 5 10 15

Thr Ile Cys His Leu Thr Asn Glu Ser Asp Gly His Thr Thr Ser Leu
20 25 30

Tyr Gly Ile Gly Phe Gly Pro Phe Ile Ile Thr Asn Lys His Leu Phe
35 40 45

Arg Arg Asn Asn Gly Thr Leu Leu Val Gln Ser Leu His Gly Val Phe
50 55 60

Lys Val Lys Asn Thr Thr Thr Leu Gln Gln His Leu Ile Asp Gly Arg
65 70 75 80

Asp Met Ile Ile Ile Arg Met Pro Lys Asp Phe Pro Pro Phe Pro Gln
85 90 95

Lys Leu Lys Phe Arg Glu Pro Gln Arg Glu Glu Arg Ile Cys Leu Val
100 105 110

Thr Thr Asn Phe Gln Thr Lys Ser Met Ser Ser Met Val Ser Asp Thr
115 120 125

Ser Cys Thr Phe Pro Ser Ser Asp Gly Ile Phe Trp Lys His Trp Ile
130 135 140

Gln Thr Lys Asp Gly Gln Cys Gly Ser Pro Leu Val Ser Thr Arg Asp
145 150 155 160

Gly Phe Ile Val Gly Ile His Ser Ala Ser Asn Phe Thr Asn Thr Asn
165 170 175

Asn Tyr Phe Thr Ser Val Pro Lys Asn Phe Met Glu Leu Leu Thr Asn
180 185 190

10

ES 2 531 372 T3

Gln Glu Ala Gln Gln Trp Val Ser Gly Trp Arg Leu Asn Ala Asp Ser
 195 200 205

Val Leu Trp Gly Gly His Lys Val Phe Met Val Lys Pro Glu Glu Pro
 210 215 220

Phe Gln Pro Val Lys Glu Ala Thr Gln Leu Met Asn
 225 230 235

5

<210> 39
 <211> 12
 <212> PRT
 <213> Virus del jaspeado del tabaco

<400> 39
 Ser Ser Ser Gly Gly Ser Glu Asn Leu Tyr Phe Gln
 1 5 10

10

<210> 40
 <211> 385
 <212> PRT
 <213> Streptomyces mobaraensis

<400> 40
 Met Asp Asn Gly Ala Gly Glu Glu Thr Lys Ser Tyr Ala Glu Thr Tyr
 1 5 10 15

Arg Leu Thr Ala Asp Asp Val Ala Asn Ile Asn Ala Leu Asn Glu Ser
 20 25 30

Ala Pro Ala Ala Ser Ser Ala Gly Pro Ser Phe Arg Ala Pro Leu Glu
 35 40 45

Val Leu Phe Gln Gly Pro Asp Ser Asp Asp Arg Val Thr Pro Pro Ala
 50 55 60

Glu Pro Leu Asp Arg Met Pro Asp Pro Tyr Arg Pro Ser Tyr Gly Arg
 65 70 75 80

Ala Glu Thr Val Val Asn Asn Tyr Ile Arg Lys Trp Gln Gln Val Tyr
 85 90 95

15

Ser His Arg Asp Gly Arg Lys Gln Gln Met Thr Glu Glu Gln Arg Glu
 100 105 110

ES 2 531 372 T3

Trp Leu Ser Tyr Gly Cys Val Gly Val Thr Trp Val Asn Ser Gly Gln
 115 120 125

Tyr Pro Thr Asn Arg Leu Ala Phe Ala Ser Phe Asp Glu Asp Arg Phe
 130 135 140

Lys Asn Glu Leu Lys Asn Gly Arg Pro Arg Ser Gly Glu Thr Arg Ala
 145 150 155 160

Glu Phe Glu Gly Arg Val Ala Lys Glu Ser Phe Asp Glu Glu Lys Gly
 165 170 175

Phe Gln Arg Ala Arg Glu Val Ala Ser Val Met Asn Arg Ala Leu Glu
 180 185 190

Asn Ala His Asp Glu Ser Ala Tyr Leu Asp Asn Leu Lys Lys Glu Leu
 195 200 205

Ala Asn Gly Asn Asp Ala Leu Arg Asn Glu Asp Ala Arg Ser Pro Phe
 210 215 220

Tyr Ser Ala Leu Arg Asn Thr Pro Ser Phe Lys Glu Arg Asn Gly Gly
 225 230 235 240

Asn His Asp Pro Ser Arg Met Lys Ala Val Ile Tyr Ser Lys His Phe
 245 250 255

Trp Ser Gly Gln Asp Arg Ser Ser Ser Ala Asp Lys Arg Lys Tyr Gly
 260 265 270

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

Ser Arg Asp Arg Asn Ile Pro Arg Ser Pro Thr Ser Pro Gly Glu Gly
 290 295 300

Phe Val Asn Phe Asp Tyr Gly Trp Phe Gly Ala Gln Thr Glu Ala Asp
 305 310 315 320

Ala Asp Lys Thr Val Trp Thr His Gly Asn His Tyr His Ala Pro Asn
 325 330 335

Gly Ser Leu Gly Ala Met His Val Tyr Glu Ser Lys Phe Arg Asn Trp
 340 345 350

Ser Glu Gly Tyr Ser Asp Phe Asp Arg Gly Ala Tyr Val Ile Thr Phe
 355 360 365

Ile Pro Lys Ser Trp Asn Thr Ala Pro Asp Lys Val Lys Gln Gly Trp
 370 375 380

Pro
 385

ES 2 531 372 T3

<210> 41
 < 211> 148
 < 212> PRT
 < 213> Thermus thermophilus

5

<400> 41
 Met Lys Val Ile Leu Leu Glu Pro Leu Glu Asn Leu Gly Asp Val Gly
 1 5 10 15
 Gln Val Val Asp Val Lys Pro Gly Tyr Ala Arg Asn Tyr Leu Leu Pro
 20 25 30
 Arg Gly Leu Ala Val Leu Ala Thr Glu Ser Asn Leu Lys Ala Leu Glu
 35 40 45
 Ala Arg Ile Arg Ala Gln Ala Lys Arg Leu Ala Glu Arg Lys Ala Glu
 50 55 60
 Ala Glu Arg Leu Lys Glu Ile Leu Glu Asn Leu Thr Leu Thr Ile Pro
 65 70 75 80
 Val Arg Ala Gly Glu Thr Lys Ile Tyr Gly Ser Val Thr Ala Lys Asp
 85 90 95
 Ile Ala Glu Ala Leu Ser Arg Gln His Gly Ile Thr Ile Asp Pro Lys
 100 105 110
 Arg Leu Ala Leu Glu Lys Pro Ile Lys Glu Leu Gly Glu Tyr Val Leu
 115 120 125
 Thr Tyr Lys Pro His Pro Glu Val Pro Ile Gln Leu Lys Val Ser Val
 130 135 140
 Val Ala Gln Glu
 145

10

<210> 42
 < 211> 149
 < 212> PRT
 < 213> Geobacillus kaustophilus

ES 2 531 372 T3

<400> 42

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

<210> 43

<211> 151

<212> PRT

<213> *Thermosipho melanesiensis*

5

ES 2 531 372 T3

<400> 43

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

<210> 44

<211> 149

<212> PRT

<213> Acidothermus cellulolyticus

<400> 44

Met Lys Leu Ile Leu Thr Gln Glu Val Ala Gly Leu Gly Gly Pro Gly
 1 5 10 15
 Asp Val Val Glu Val Arg Asp Gly Tyr Gly Arg Asn Tyr Leu Leu Pro
 20 25 30

5

ES 2 531 372 T3

Lys Arg Leu Ala Met Pro Ala Ser Pro Gly Ala Val Lys Gln Val Ala
 35 40 45

Leu Ile Lys Arg Ala Arg Glu Val Arg Glu Ile Arg Asp Leu Asp Gln
 50 55 60

Ala Arg Ala Leu Arg Asp Gln Leu Glu Ala Leu Pro Val Thr Leu Pro
 65 70 75 80

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

Asp Ile Ala Ala Ala Val His Ala Ala Gly Gly Pro Lys Leu Asp Lys
 100 105 110

Arg Arg Ile Glu Ile Ser Gly Pro Ile Lys Thr Ile Gly Ser His Gln
 115 120 125

Val Thr Val Arg Leu His Pro Glu Val Ser Ala Thr Val Ser Val Glu
 130 135 140

Val Val Pro Ala Ser
 145

<210> 45
 <211> 219
 <212> PRT
 <213> homo sapiens

5

<400> 45
 Ser Thr Leu Glu Ile Ala Gly Leu Val Arg Lys Asn Leu Val Gln Phe
 1 5 10 15

Gly Val Gly Glu Lys Asn Gly Cys Val Arg Trp Val Met Asn Ala Leu
 20 25 30

Gly Val Lys Asp Asp Trp Leu Leu Val Pro Ser His Ala Tyr Lys Phe
 35 40 45

Glu Lys Asp Tyr Glu Met Met Glu Phe Tyr Phe Asn Arg Gly Gly Thr
 50 55 60

ES 2 531 372 T3

Tyr Tyr Ser Ile Ser Ala Gly Asn Val Val Ile Gln Ser Leu Asp Val
65 70 75 80

Gly Phe Gln Asp Val Val Leu Met Lys Val Pro Thr Ile Pro Lys Phe
85 90 95

Arg Asp Ile Thr Glu His Phe Ile Lys Lys Gly Asp Val Pro Arg Ala
100 105 110

Leu Asn Arg Leu Ala Thr Leu Val Thr Thr Val Asn Gly Thr Pro Met
115 120 125

Leu Ile Ser Glu Gly Pro Leu Lys Met Glu Glu Lys Ala Thr Tyr Val
130 135 140

His Lys Lys Asn Asp Gly Thr Thr Val Asp Leu Thr Val Asp Gln Ala
145 150 155 160

Trp Arg Gly Lys Gly Glu Gly Leu Pro Gly Met Cys Gly Gly Ala Leu
165 170 175

Val Ser Ser Asn Gln Ser Ile Gln Asn Ala Ile Leu Gly Ile His Val
180 185 190

Ala Gly Gly Asn Ser Ile Leu Val Ala Lys Leu Val Thr Gln Glu Met
195 200 205

Phe Gln Asn Ile Asp Lys Lys Ile Glu Ser Gln
210 215

<210> 46

<211> 143

<212> PRT

<213> Virus de la enfermedad hemorrágica del conejo

5

<400> 46

Gly Leu Pro Gly Phe Met Arg His Asn Gly Ser Gly Trp Met Ile His
1 5 10 15

Ile Gly Asn Gly Leu Tyr Ile Ser Asn Thr His Thr Ala Arg Ser Ser
20 25 30

Cys Ser Glu Ile Val Thr Cys Ser Pro Thr Thr Asp Leu Cys Leu Val
35 40 45

ES 2 531 372 T3

Lys Gly Glu Ala Ile Arg Ser Val Ala Gln Ile Ala Glu Gly Thr Pro
50 55 60

Val Cys Asp Trp Lys Lys Ser Pro Ile Ser Thr Tyr Gly Ile Lys Lys
65 70 75 80

Thr Leu Ser Asp Ser Thr Lys Ile Asp Val Leu Ala Tyr Asp Gly Cys
85 90 95

Thr Gln Thr Thr His Gly Asp Cys Gly Leu Pro Leu Tyr Asp Ser Ser
100 105 110

Gly Lys Ile Val Ala Ile His Thr Gly Lys Leu Leu Gly Phe Ser Lys
115 120 125

Met Cys Thr Leu Ile Asp Leu Thr Ile Thr Lys Gly Val Tyr Glu
130 135 140

5 <210> 47
< 211> 10
< 212> PRT
< 213> artificial

<220>
< 223> Enlazador 8

10 <400> 47
Ser Ser Ser Gly Ser Gly Ser Gly Ser Gly
1 5 10

<210> 48
< 211> 7
< 212> PRT
< 213> Artificial

15 <220>
< 223> Enlazador 9

<400> 48
Gly Ser Ser Gly Ser Gly Ser
1 5

20 <210> 49
< 211> 11
< 212> PRT
< 213> Artificial

<220>
< 223> Enlazador 10

25 <400> 49
Ser Ser Ser Gly Gly Ser Glu Leu Arg Thr Gln
1 5 10

REIVINDICACIONES

1. Enzima de procesamiento que comprende un marcador fijado en el extremo N-terminal, en donde el marcador consiste en una secuencia peptídica procedente de la familia L9 de proteínas ribosómicas procedentes de bacterias termófilas, en donde el marcador tiene de 15 a 250 aminoácidos, en donde el marcador comprende al menos 15% de residuos de aminoácidos básicos, Lys y Arg, y en donde el marcador tiene un pI superior a 9.
2. Enzima de procesamiento según la reivindicación 1, en la que el marcador se fija a la enzima de procesamiento a través de una secuencia enlazadora.
3. Enzima de procesamiento según cualquiera de las reivindicaciones 1-2, en la que el marcador no comprende residuos de cisteína.
4. Enzima de procesamiento según cualquiera de las reivindicaciones 1-3, en la que el marcador comprende de 50 a 250, o de 75 a 200, de 100 a 250 o de 100 a 200 residuos de aminoácidos.
5. Enzima de procesamiento según cualquiera de las reivindicaciones 1-4, en donde el marcador comprende de 20 a 50% de residuos de aminoácidos básicos.
6. Enzima de procesamiento según cualquiera de las reivindicaciones 1-5, en donde el marcador se selecciona entre el grupo de secuencias peptídicas que consisten en
 MKVILLRDVPIKIGKKGEIKEVSDGYARNYLIPRGFAKEYTEGLERAIKHEKEIEKRKKERERESEKI
 LKELKKRTHVVKVKGEGGKIFGAVTAATVAEEISKTTGLKLDKRWFKLDKPIKELGEYSLEVSLPG
 GVKDTIKIRVEREE(SEQ ID NO:10);
 MKVILLEPLENLGDVGQWVDVKPGYARNYLLPRGLAVLATESNLKALEARIRAQAKR
 LAERKAEAEERLKEILENLTLTIPVRAGETKIYGSVTAKDIAEALSRQHGITIDPKRLALEKPIKELGEYV
 LTYKPHPEVPIQLKVSVAQE(SEQ ID NO: 41);
 MKVIFLKDVKGKGGKGEIKDVADGYANNFLFKQGLAIEATPANIKALEAQKQKEQRQAAEEL
 ANAKLKEELEKLTVEIPAKAGEGGRLFGSITSKQIAEALQAQHGLKLDKRIELADAIRSLGY
 TNVPVKLHPEVTATLKVHVKEQK(SEQ ID NO: 42) y
 MKVVLLKDVS KIGKKGEIKN VSDGYARNYLIPKGLALEATPRVLKRLEAEKR-KKEEEKIQ
 IKTQNEELLKMLKFLYKIP VKAGESGKLFGALTNSDIK AVEKIADVNI
 DKKFIVLEKPIKEIGMYDVLVRLPEGVSGK IKVEVIQEGKN(SEQ ID NO: 43).
7. Enzima de procesamiento según cualquiera de las reivindicaciones 2-6, en donde la secuencia del enlazador tiene de 1 a 30, de 1 a 25, de 1 a 20 o de 1 a 15 residuos de aminoácidos.
8. Enzima de procesamiento según la reivindicación 7, en la que el enlazador se selecciona a partir del grupo que consiste en SEQ ID NO: 28, SEQ ID NO: 29, SEQ ID NO: 30, SEQ ID NO: 31, SEQ ID NO: 32, SEQ ID NO: 33, SEQ ID NO: 34, SEQ ID NO: 47 y SEQ ID NO: 48.
9. Enzimas de procesamiento según cualquiera de las reivindicaciones 1-8, que se seleccionan a partir de oxidoreductasas, transferasas, hidrolasas, liasas, isomerasas y ligasas.
10. Estructura artificial de ácido nucleico que codifica la enzima de procesamiento según cualquiera de las reivindicaciones 1-9.
11. Método para la preparación de una enzima de procesamiento según cualquiera de las reivindicaciones 1-9, que comprende;
 - i) la expresión de la enzima de procesamiento según cualquiera de las reivindicaciones 1-9,
 - ii) cargar dicha enzima de procesamiento expresada sobre una columna de intercambio catiónico, y
 - iii) eluir dicha enzima de procesamiento con un eluyente adecuado.
12. Método según la reivindicación 11, en donde la enzima de procesamiento es una proteasa seleccionada entre el grupo que consiste en proteasa HRV14 3C, proteasa HAV18 3C, proteasa TEV y RHDV 3C.

13. Método para modificar un precursor de proteína diana que comprende las etapas de hacer reaccionar una enzima de procesamiento marcada en el extremo N-terminal según cualquiera de las reivindicaciones 1-9 con el precursor de la proteína diana y separar la proteína diana de la enzima de procesamiento.
- 5 14. Método según la reivindicación 13, en donde la enzima de procesamiento marcada en el extremo N-terminal y el precursor de la proteína diana se coexpresan en un hospedador adecuado que permite el aislamiento de la proteína diana después de la expresión.
15. Método según la reivindicación 14, en donde dicho hospedador se selecciona a partir de las cepas de *Bacillus*, *Streptomyces* y *Escherichia coli*.

Fig. 1

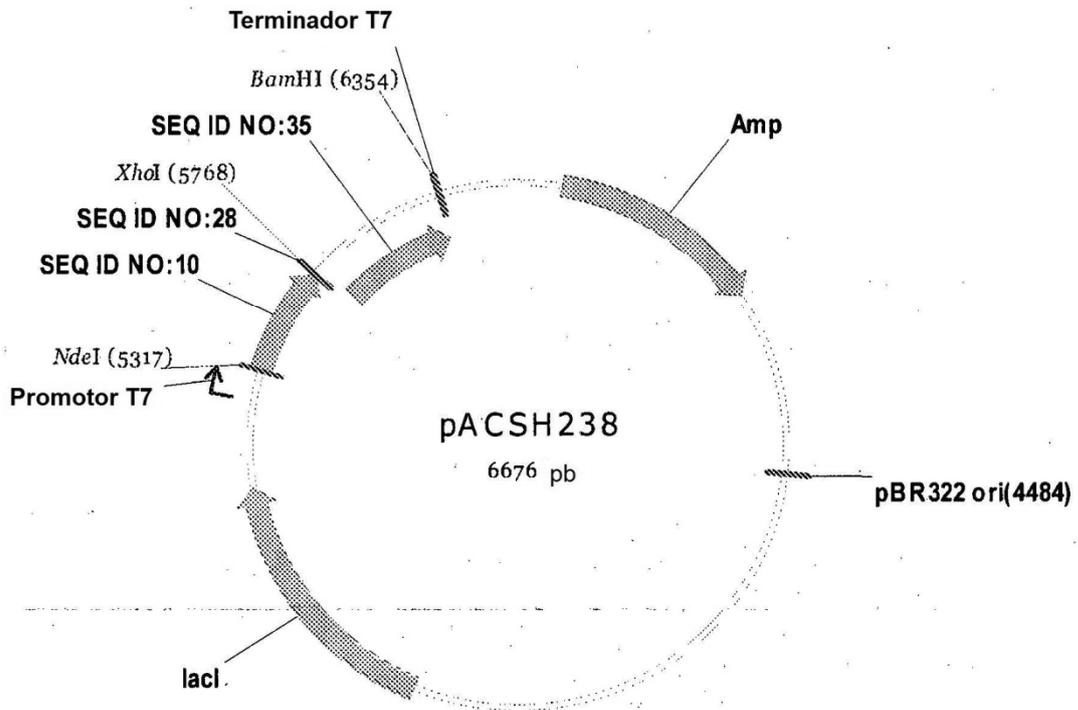


Fig. 2

