



19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA

11 Número de publicación: **2 289 288**

51 Int. Cl.:

**C12Q 1/37** (2006.01)

**C12N 9/50** (2006.01)

**C12N 9/60** (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Número de solicitud europea: **03727460 .2**

86 Fecha de presentación : **09.05.2003**

87 Número de publicación de la solicitud: **1504117**

87 Fecha de publicación de la solicitud: **09.02.2005**

54

Título: **Procedimiento para generar proteasas específicas de secuencia mediante evolución dirigida.**

30

Prioridad: **10.05.2002 EP 02010576**

45

Fecha de publicación de la mención BOPI:  
**01.02.2008**

45

Fecha de la publicación del folleto de la patente:  
**01.02.2008**

73

Titular/es: **Direvo Biotech AG.**  
**Nattermannallee 1**  
**50829 Köln, DE**

72

Inventor/es: **Koltermann, André;**  
**Kettling, Ulrich;**  
**Haupts, Ulrich;**  
**Tebbe, Jan;**  
**Scholz, Peter;**  
**Pilling, Jens;**  
**Werner, Susanne y**  
**Rarbach, Markus**

74

Agente: **Carpintero López, Francisco**

ES 2 289 288 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**

Procedimiento para generar proteasas específicas de secuencia mediante evolución dirigida.

5 Se describe un procedimiento para generar proteasas específicas de secuencia mediante evolución dirigida basada en exploración. El uso del procedimiento proporciona proteasas que reconocen y rompen secuencias de aminoácidos que pueden definirse por el usuario con alta especificidad de secuencia. Las proteasas que pueden obtenerse mediante el procedimiento pueden usarse en una variedad de aplicaciones médicas, diagnósticas e industriales.

10 **Antecedentes de la invención**

Las enzimas proteolíticas o proteasas son una clase de enzimas que tiene una posición destacada entre las diferentes enzimas, ya que la reacción catalizada mediante proteasas es la rotura de enlaces peptídicos en otras proteínas. Las proteasas no son sólo enzimas muy comunes en la naturaleza, sino que pertenecen a las enzimas más importantes para el uso médico e industrial. De las ventas totales a nivel mundial de enzimas, que se estima que son más de 1000 millones de dólares estadounidenses al año, las proteasas suponen aproximadamente el 60%. Basándose en el grupo funcional presente en el sitio activo, las proteasas se clasifican en cuatro grupos, es decir, serina proteasas (EC 3.4.21), cisteína proteasas (EC 3.4.22), aspártico proteasas (EC 3.4.23), y metaloproteasas (EC 3.4.24). La clasificación en uno de los cuatro grupos se realiza normalmente mediante determinación experimental de la sensibilidad hacia diferentes tipos de inhibidores de proteasas. Además, las proteasas de los cuatro grupos difieren en sus propiedades bioquímicas. Por ejemplo, las serina proteasas son sensibles a los inhibidores 3,4-DCI, DFP, PMSF y TLCK, y tienen un valor óptimo de pH entre pH 7 y 11. Las aspártico proteasas se inhiben mediante pepstatina, DAN y EPNP, y predominantemente tienen un valor óptimo de pH entre pH 3 y 4. Las cisteína proteasas son sensibles a inhibidores de sulfhidrilo tales como PCMB, y excepto unas pocas excepciones, tienen valores óptimos de pH neutros. Las metaloproteasas se caracterizan por el requisito de un ión metálico divalente para su actividad. Por tanto, las metaloproteasas se inhiben mediante agentes quelantes tales como EDTA, y tienen valores óptimos de pH neutros o alcalinos. Entre estos cuatro grupos, normalmente se realiza una clasificación adicional basándose en similitudes estructurales.

Además de una clasificación bioquímica y estructural combinada de este tipo, las proteasas pueden agruparse según su espectro de sustratos. Los dos grupos más generales que han de distinguirse son las exoproteasas y endoproteasas. Las exoproteasas sólo rompen enlaces peptídicos en el extremo de un péptido, mientras que las endoproteasas catalizan la rotura de enlaces en cualquier parte de una cadena peptídica. La especificidad de las proteasas, es decir su capacidad para reconocer e hidrolizar específicamente ciertos sustratos peptídicos mientras que otros permanecen sin romper, puede expresarse cualitativa y cuantitativamente. La especificidad cualitativa se refiere al tipo de restos de aminoácido que son aceptados por una proteasa en ciertas posiciones del sustrato peptídico. Por ejemplo, la tripsina y el activador del plasminógeno de tipo tisular están relacionados con respecto a su especificidad cualitativa, ya que los dos requieren en la posición P1 una arginina o un resto similar (nomenclatura de posiciones del sustrato peptídico según la nomenclatura de Schlechter y Berger (Biochem. Biophys. Res. Commun. 27 (1967) 157-162). Por otro lado, la especificidad cuantitativa se refiere al número relativo de sustratos peptídicos que se aceptan como sustratos. La especificidad cuantitativa puede expresarse mediante el término

$$s = -\log(Q),$$

45 en el que Q es la proporción de todos los sustratos peptídicos aceptados frente a todos los sustratos peptídicos posibles. Se muestran a modo de ejemplo especificidades cuantitativas de diversas proteasas en la tabla I. El cálculo de especificidades cuantitativas se basa en los veinte aminoácidos que se producen de manera natural, y en la suposición de que todas las combinaciones de estos veinte aminoácidos son factibles. Por consiguiente, las proteasas que aceptan sólo una pequeña parte de todos los péptidos posibles tienen una especificidad alta, mientras que la especificidad de proteasas que, como extremo, rompen cualquier sustrato peptídico sería teóricamente cero.

TABLA I

*Especificidades cuantitativas de diferentes proteasas*

55

60

65

| proteasa | Requisitos de sustrato |    |    |    |    |       |     |     |     | Especificidad cuantitativa |            |
|----------|------------------------|----|----|----|----|-------|-----|-----|-----|----------------------------|------------|
|          | P6                     | P5 | P4 | P3 | P2 | P1    | P1' | P2' | P3' | Q                          | s = -log Q |
|          | X                      | X  | X  | X  | X  | X     | X   | X   | X   | 1,00E+00                   | 0          |
| quimo-   | X                      | X  | X  | X  | X  | F/Y/W | X   | X   | X   | 1,50E-01                   | 0,82       |

## ES 2 289 288 T3

|    |                 |   |   |         |   |       |       |       |   |   |          |      |
|----|-----------------|---|---|---------|---|-------|-------|-------|---|---|----------|------|
|    | tripsina        |   |   |         |   |       |       |       |   |   |          |      |
| 5  | papaína         | X | X | X       | X | F/V/L | X     | X     | X | X | 1,50E-01 | 0,82 |
|    | tripsina        | X | X | X       | X | X     | K/R   | X     | X | X | 1,00E-01 | 1,00 |
|    | pepsina         | X | X | X       | X | X     | F/Y/L | W/F/Y | X | X | 2,25E-02 | 1,65 |
| 10 | VGT             | E | X | X       | Y | X     | Q     | S/G   | X | X | 1,25E-05 | 4,90 |
|    | fibrinolisisina | X | X | K/V/I/F | X | F/Y/W | R/K   | N     | A | X | 7,50E-06 | 5,12 |
|    | trombina        | X | X | L/I/V/F | X | P     | R     | N     | A | X | 1,25E-06 | 5,90 |
| 15 | t-PA            | X | X | C       | P | G     | R     | V     | V | G | 7,81E-10 | 9,11 |

(Los restos de aminoácido están abreviados tal como se muestra en la tabla II. X se refiere a cualquier resto de aminoácido).

La especificidad cuantitativa de las proteasas varía a lo largo de un amplio intervalo. Se conocen proteasas muy inespecíficas, tales como la papaína que rompe todos los polipéptidos que contienen un resto de fenilalanina, uno de valina o uno de leucina ( $s = 0,82$ ), o la tripsina que rompe todos los polipéptidos que contienen un resto de arginina o uno de lisina ( $s = 1,0$ ). Por otro lado, se conocen proteasas muy específicas, tales como el activador del plasminógeno de tipo tisular (t-PA) que rompe el plasminógeno sólo en una única secuencia específica ( $s = 9,11$ ). Las proteasas con especificidad de sustrato alta desempeñan un papel importante en la regulación de las funciones de las proteínas en organismos vivos. La rotura específica de sustratos polipeptídicos, por ejemplo, activa proteínas precursoras o desactiva enzimas o proteínas activas, regulando de ese modo sus funciones. Varias proteasas con especificidades de sustrato altas se usan en aplicaciones médicas. Los ejemplos farmacéuticos para la activación o desactivación mediante rotura de sustratos polipeptídicos específicos son la aplicación del t-PA en el infarto agudo de miocardio que activa el plasminógeno para resolver coágulos de fibrina, o la aplicación de Ancrod en el accidente cerebrovascular que desactiva el fibrinógeno, disminuyendo así la viscosidad de la sangre y aumentando su capacidad de transporte. Mientras que el t-PA es una proteasa humana con una actividad necesaria en la regulación sanguínea humana, el Ancrod es una proteasa no humana. Se aisló de la víbora *Agkistrodon rhodostoma*, y comprende el componente principal del veneno de serpiente. Por tanto, existen algunas pocas proteasas no humanas con aplicabilidad terapéutica. Sin embargo, normalmente su identificación es muy fortuita.

El tratamiento de enfermedades administrando fármacos se basa normalmente en un mecanismo molecular iniciado por el fármaco que activa o inactiva una función proteica específica en el cuerpo del paciente, ya sea una proteína endógena o una proteína de un virus o microbio infeccioso. Aunque la acción de los fármacos químicos sobre estas dianas es todavía difícil de entender o de predecir, los fármacos proteicos pueden reconocer específicamente estas proteínas diana entre millones de otras proteínas. Los ejemplos destacados de proteínas que tienen la posibilidad intrínseca de reconocer otras proteínas son anticuerpos, receptores y proteasas. Aunque hay un número enorme de posibles proteínas diana, sólo muy pocas proteasas están disponibles actualmente para acceder a estas proteínas diana. Debido a su actividad proteolítica, las proteasas son particularmente adecuadas para la inactivación o activación de dianas proteicas. Si se consideran sólo las proteínas humanas, el número de posibles proteínas diana aún es enorme. Se estima que el genoma humano comprende entre 30.000 y 100.000 genes, cada uno de los cuales codifica una proteína diferente. Muchas de estas proteínas están implicadas en enfermedades humanas y son por tanto posibles dianas farmacéuticas. Las proteasas que reconocen y rompen estas proteínas diana con una especificidad alta son por consiguiente de gran valor como posibles fármacos. Sin embargo, la aplicación médica de tales proteasas está restringida por su aparición. Por ejemplo, teóricamente hay 25 mil millones de posibilidades diferentes para una especificidad de  $s = 10,4$  (que corresponden al reconocimiento específico de una secuencia única de ocho restos de aminoácido). Es muy improbable encontrar una proteasa de este tipo con una especificidad cualitativa particular explorando aislados naturales.

Los sistemas de selección para proteasas de especificidad conocida se conocen en la técnica, por ejemplo, a partir de Smith *et al.*, Proc. Natl. Acad. Sci. USA, vol. 88 (1991). Tal como se muestra a modo de ejemplo, el sistema comprende el factor de transcripción de levadura GAL4 como marcador de selección, una secuencia diana definida y rompible insertada en GAL4 junto con la proteasa de VGT. La rotura separa el dominio de unión a ADN del dominio de activación de transcripción y de este modo inactiva el factor de transcripción. La incapacidad fenotípica de las células resultantes para metabolizar la galactosa puede detectarse mediante un ensayo colorimétrico o mediante la selección en el sustrato suicida 2-desoxigalactosa.

Además, la selección puede realizarse mediante el uso de sustratos peptídicos con modificaciones tales como, por ejemplo, restos fluorogénicos basados en grupos tales como ACC, descritos previamente por Harris *et al.* (documento US 2002/022243).

En principio se conocen técnicas de laboratorio para generar enzimas proteolíticas con especificidades de secuencia alteradas. Pueden clasificarse mediante sus sistemas de expresión y de selección. La selección genética significa producir una proteasa dentro de un organismo, proteasa que puede romper una proteína precursora lo que a su vez da como resultado una alteración del comportamiento de crecimiento del organismo productor. De una población de organismos con diferentes proteasas pueden seleccionarse aquellos con un comportamiento de crecimiento alterado. Este principio se notificó por Davis *et al.* (documento US 5258289, documento WO 96/21009). La producción de un sistema de fago depende de la rotura de una proteína de fago que sólo puede activarse en presencia de una enzima proteolítica o anticuerpo que puede romper la proteína de fago. Las enzimas proteolíticas o anticuerpos seleccionados tendrían la capacidad de romper una secuencia de aminoácidos para la activación de la producción de fagos. Además, no hay control de la especificidad de las proteasas que se seleccionan. El sistema no selecciona proteasas con actividades bajas para otros péptidos distintos del sustrato peptídico usado. Adicionalmente, este sistema no permite una caracterización precisa de las constantes cinéticas de las proteasas seleccionadas ( $k_{cat}$ ,  $K_M$ ). Se notifican otros varios sistemas con expresión de proteasa intracelular pero todos experimentan las desventajas mencionadas anteriormente. Algunos de ellos usan un sistema indicador genético que permite una selección mediante exploración el lugar de una selección genética, pero tampoco pueden superar la insuficiencia intrínseca de la caracterización intracelular de las proteasas.

Se notifica un sistema para generar las enzimas proteolíticas con especificidades de secuencia alteradas con proteasas unidas a membrana. Iverson *et al.* (documento WO 98/49286) describe un sistema de expresión para una proteasa unida a membrana que se presenta en la superficie de células. Un elemento esencial del diseño experimental es que la reacción catalítica ha de realizarse en la superficie celular, es decir, los sustratos y productos deben permanecer asociados con la bacteria que expresa la enzima en la superficie. Esta restricción limita la generación de enzimas proteolíticas con especificidades de secuencia alteradas y no permite una caracterización precisa de las constantes cinéticas de las proteasas seleccionadas ( $k_{cat}$ ,  $K_M$ ). Además, el procedimiento no permite el control de la posición en la que se rompe el péptido. Adicionalmente, las proteasas identificadas positivamente tendrán la capacidad de romper una cierta secuencia de aminoácidos (aa) pero también podrán romper muchas otras secuencias de aa. Por tanto, no hay control de la especificidad de las proteasas que se seleccionan.

También se conoce un sistema para generar enzimas proteolíticas con especificidades de secuencia alteradas con proteasas autosecretantes. Duff *et al.* (documento WO 98/11237) describe un sistema de expresión para una proteasa autosecretante. Un elemento esencial del diseño experimental es que la reacción catalítica actúa sobre la propia proteasa mediante un procesamiento autoproteolítico de la molécula precursora unida a membrana para liberar la proteasa madura de la membrana celular al entorno extracelular. Por tanto, debe construirse una proteína de fusión en la que la secuencia peptídica diana sustituye al sitio de rotura natural para la autoproteolisis. Las limitaciones de un sistema de este tipo son que las proteasas identificadas positivamente tendrán la capacidad de romper una cierta secuencia de aa pero también pueden romper muchas otras secuencias peptídicas. Por tanto, no puede conseguirse una especificidad de sustrato alta con un enfoque de este tipo. Adicionalmente, un sistema de este tipo no puede controlar que las proteasas seleccionadas rompan en una posición específica en una secuencia de aa definida y no permite una caracterización precisa de las constantes cinéticas de las proteasas seleccionadas ( $k_{cat}$ ,  $K_M$ ).

Broad *et al.* (WO 99/11801) describe un sistema de células heterólogas adecuado para la alteración de la especificidad de proteasas. El sistema comprende un precursor del factor de transcripción en el que el factor de transcripción está unido a un dominio de anclaje a membrana por medio de un sitio de rotura de proteasa. La rotura en el sitio de rotura de proteasa mediante una proteasa libera el factor de transcripción, lo que a su vez inicia la expresión de un gen diana que está bajo el control del promotor respectivo. El diseño experimental de la alteración de la especificidad consiste en la inserción de sitios de rotura de proteasa con secuencias modificadas y el sometimiento de la proteasa a mutagénesis. Las nuevas proteasas obtenidas pueden reconocer la secuencia modificada, cuyo efecto se controla mediante la expresión del gen diana. Un sistema de este tipo tampoco permite un control preciso de las propiedades bioquímicas de las proteasas seleccionadas.

La mayoría de estos enfoques aplican procedimientos de evolución dirigida para la generación de enzimas proteolíticas con especificidades de secuencia alteradas. En otra parte se notifican y se describen varios procedimientos de mutación y recombinación diferentes para generar bibliotecas genéticas. Todos los diferentes procedimientos experimentan la carencia de una selección precisa de variantes de proteasa positivas de grandes bibliotecas. En primer lugar, estos procedimientos no pueden distinguir entre recambios único y múltiple de sustratos peptídicos lo que es necesario con el fin de prevenir la selección de variantes de  $k_{cat}$  baja. En segundo lugar, no es posible desencadenar la concentración de sustrato y enzima para seleccionar variantes de proteasa para una  $K_M$  menor. En tercer lugar, ninguno de estos sistemas permite la selección de una proteasa con una actividad aumentada sobre el sustrato peptídico deseado mediante lo cual disminuye la actividad sobre el sustrato peptídico original.

Los procedimientos que cumplen los tres criterios de selección mencionados anteriormente ( $k_{cat}$ ,  $K_M$  y especificidad de sustrato) para generar enzimas proteolíticas con especificidad de secuencia alta aplicando evolución dirigida basada en exploración no han estado disponibles hasta la fecha.

## 65 Sumario de la invención

Por tanto, el problema técnico subyacente a la presente invención es proporcionar un procedimiento para generar nuevas proteasas con especificidades de sustrato definidas por el usuario aplicando evolución dirigida. En particular,

la invención se refiere a un procedimiento para la evolución de proteasas novedosas hacia el reconocimiento selectivo y la rotura solamente de secuencias de aminoácidos específicas. Este problema técnico se ha resuelto mediante las realizaciones de la invención especificadas a continuación y en las reivindicaciones adjuntas. Por tanto la presente invención se refiere a

5

(1) un procedimiento para identificar proteasas específicas de secuencia con especificidades de sustrato objetivo que comprende las etapas siguientes

10

(a) proporcionar una población de proteasas comprendida por variantes de una primera proteasa o por variantes o quimeras de dos o más primeras proteasas, teniendo dichas primeras proteasas una especificidad de sustrato para una secuencia de aminoácidos particular de un primer sustrato peptídico;

15

(b) poner en contacto dicha población de proteasas con uno o más segundos sustratos, que comprenden al menos una secuencia de aminoácidos específica que se parece a la secuencia de aminoácidos del sustrato peptídico diana pero sin estar presente dentro del primer sustrato peptídico; y

20

(c) seleccionar una o más variantes de proteasa de la población de proteasas proporcionada en la etapa (a) que tienen especificidad para dicha secuencia de aminoácidos específica de los segundos sustratos proporcionada en la etapa (b) en condiciones que permiten la identificación de proteasas que reconocen e hidrolizan preferiblemente dicha una secuencia de aminoácidos específica dentro de los segundos sustratos, en el que la exploración para determinar la actividad proteasa se consigue añadiendo en exceso péptidos distintos del segundo péptido, usando de ese modo los péptidos añadidos como competidores;

25

(2) en una realización preferida de (1) anterior sólo se usa un segundo sustrato en el uno o más ciclos (a) a (c), es decir, el segundo sustrato es idéntico al sustrato diana;

30

(3) en una realización preferida adicional de (1) anterior se usan diferentes segundos sustratos, y los segundos sustratos tienen un rasgo intermedio con respecto al primer sustrato y el sustrato diana, y el último segundo sustrato que se usa es idéntico al sustrato diana; y

(4) en una realización preferida particular de (1) a (3) anteriores la proteasa objetivo tiene una especificidad similar al activador de plasminógeno de tipo tisular y rompe el sustrato diana CPGR<sup>1</sup>VVGG.

35

La identificación y selección de proteasas que han evolucionado hacia la especificidad objetivo se realiza explorando para determinar actividades catalíticas sobre diferentes sustratos peptídicos, o bien explorando para determinar un aumento de la afinidad, o usando dos sustratos en comparación, o usando péptidos no específicos como competidores, o usando sustratos peptídicos intermedios.

40

La siguiente descripción detallada describirá las características preferidas, las ventajas y la utilidad de la presente invención.

### Breve descripción de las figuras

45

Las siguientes figuras se proporcionan con el fin de explicar adicionalmente la presente invención en complemento a la descripción detallada:

La figura 1 representa esquemáticamente las dos alternativas A y B del procedimiento de la invención.

50

La figura 2 distingue las dos alternativas A y B del procedimiento de la invención mostrando esquemáticamente los cambios cualitativos y cuantitativos en la especificidad durante la evolución hacia la especificidad objetivo.

La figura 3 ilustra esquemáticamente cómo evolucionan las proteasas con actividades catalíticas cambiadas usando las dos alternativas A y B del procedimiento de la invención.

55

La figura 4 representa esquemáticamente de dos formas diferentes el enfoque intermedio como un aspecto particular de la invención que usa sustratos intermedios.

60

La figura 5 ilustra esquemáticamente cómo evolucionan, según la invención, las proteasas con actividades catalíticas cambiadas usando el enfoque intermedio.

La figura 6 muestra a modo de ejemplo un vector de expresión para *S. cerevisiae* que puede usarse para el procedimiento de la invención.

65

La figura 7 muestra a modo de ejemplo la hidrólisis de un sustrato peptídico mediante la proteasa de virus del grabado del tabaco.

La figura 8 muestra a modo de ejemplo una distribución de actividades catalíticas obtenidas mediante exploración usando espectroscopía de fluorescencia confocal.

## ES 2 289 288 T3

La figura 9 muestra a modo de ejemplo la disminución de  $K_M$  durante la evolución hacia una afinidad mayor.

La figura 10 muestra a modo de ejemplo el cambio en la especificidad durante la evolución de proteasas hacia la especificidad de t-PA.

La figura 11 representa esquemáticamente una variante preferida del enfoque intermedio.

La figura 12 muestra a modo de ejemplo la conversión de sustrato dependiente del tiempo de una proteasa de partida en comparación con una de las variantes evolucionadas.

### Descripción detallada de la invención

En el contexto de esta invención se usan los términos y definiciones siguientes.

El término “proteasa” significa cualquier molécula proteica que actúa en la hidrólisis de enlaces peptídicos. Incluye enzimas proteolíticas que se producen de manera natural, así como variantes de las mismas obtenidas mediante mutagénesis dirigida al sitio o al azar o cualquier otro procedimiento de ingeniería de proteínas, cualquier fragmento de una enzima proteolítica, o cualquier complejo molecular o proteína de fusión que comprende una de las proteínas mencionadas anteriormente. Una “quimera de proteasas” significa una proteína de fusión compuesta por dos o más fragmentos derivados de diferentes proteasas originales.

El término “sustrato” o “sustrato peptídico” significa cualquier péptido, oligopéptido, o molécula proteica de cualquier composición, secuencia o longitud de aminoácidos, que contiene un enlace peptídico que puede hidrolizarse catalíticamente mediante una proteasa. El enlace peptídico que se hidroliza se denomina el “sitio de rotura”. La numeración de las posiciones en el sustrato se realiza según el sistema introducido por Schlechter y Berger (Biochem. Biophys. Res. Commun. 27 (1967) 157-162). Los restos de aminoácido adyacentes en N-terminal al sitio de rotura se numeran P1, P2, P3, etc., mientras que los restos adyacentes en C-terminal al sitio de rotura se numeran P1', P2', P3', etc.

El término “especificidad” significa la capacidad de una proteasa para reconocer e hidrolizar selectivamente ciertos sustratos peptídicos mientras que otros permanecen sin romper. La especificidad puede expresarse cualitativa y cuantitativamente. La “especificidad cualitativa” se refiere al tipo de restos de aminoácido que se aceptan por una proteasa en ciertas posiciones del sustrato peptídico. La “especificidad cuantitativa” se refiere al número de sustratos peptídicos que se aceptan como sustratos. La especificidad cuantitativa puede expresarse mediante el término  $s$ , que es el logaritmo negativo del número de todos los sustratos peptídicos aceptados dividido entre el número de todos los sustratos peptídicos posibles. Las proteasas que aceptan solamente una pequeña parte de todos los sustratos peptídicos posibles tiene una “especificidad alta” ( $s \gg 1$ ). Las proteasas que aceptan casi cualquier sustrato peptídico tienen una “especificidad baja”. Las proteasas con especificidad muy baja ( $s \leq 1$ ) se denominan también “proteasas inespecíficas”.

El término “primera proteasa” describe cualquier proteasa usada en la etapa (a) de esta invención como punto de partida con el fin de generar poblaciones de variantes de proteasa que están relacionadas con esta primera proteasa. El término “primer sustrato” o “primer sustrato peptídico” describe un sustrato que se reconoce e hidroliza por la primera proteasa. El término “primera especificidad” describe la especificidad cualitativa y cuantitativa de la primera proteasa.

El término “proteasa evolucionada” describe cualquier proteasa que se genera mediante el uso del procedimiento de la invención. El término “sustrato diana” o “sustrato peptídico diana” describe un sustrato que se reconoce e hidroliza por la proteasa evolucionada. El término “especificidad objetivo” describe la especificidad cualitativa y cuantitativa de la proteasa evolucionada que ha de generarse mediante el uso del procedimiento de la invención. Por tanto, la especificidad objetivo define la especificidad de la proteasa evolucionada para el sustrato peptídico diana mientras que otros sustratos no se reconocen ni hidrolizan o sólo débilmente.

El término “producto intermedio” o “sustrato intermedio” describe cualquier sustrato que tiene un rasgo intermedio entre otros dos sustratos. El rasgo intermedio puede basarse en la composición de aminoácidos, la secuencia de aminoácidos, las propiedades de los restos de aminoácido contenidos en los sustratos, o una combinación de estas características.

Las propiedades catalíticas de las proteasas se expresan usando los parámetros cinéticos “ $K_M$ ” o “constante de Michaelis-Menten”, “ $k_{cat}$ ” o “constante de velocidad catalítica”, y “ $k_{cat}/K_M$ ” o “eficacia catalítica”, según las definiciones de Michaelis y Menten (Fersht, A., Enzyme Structure and Mechanism, W. H. Freeman and Company, Nueva York, 1995). El término “actividad catalítica” describe la tasa de conversión del sustrato en condiciones definidas.

Los aminoácidos se abrevian según la tabla II siguiente en un código de una o bien de tres letras.

TABLA II

*Abreviaturas de aminoácidos*

5

10

15

20

25

30

35

40

| Abreviaturas |     | Aminoácido      |
|--------------|-----|-----------------|
| A            | Ala | Alanina         |
| C            | Cys | Cisteína        |
| D            | Asp | Ácido aspártico |
| E            | Glu | Ácido glutámico |
| F            | Phe | Fenilalanina    |
| G            | Gly | Glicina         |
| H            | His | Histidina       |
| I            | Ile | Isoleucina      |
| K            | Lys | Lisina          |
| L            | Leu | Leucina         |
| M            | Met | Metionina       |
| N            | Asn | Asparagina      |
| P            | Pro | Prolina         |
| Q            | Gln | Glutamina       |
| R            | Arg | Arginina        |
| S            | Ser | Serina          |
| T            | Thr | Treonina        |
| V            | Val | Valina          |
| W            | Trp | Triptófano      |
| Y            | Tyr | Tirosina        |

45

50

Tal como se expuso anteriormente, la presente invención se refiere a un procedimiento para generar proteasas específicas de secuencia con una especificidad de sustrato objetivo aplicando principios de evolución molecular. Según la invención, esto se consigue proporcionando una población de proteasas que estén relacionadas entre sí, así como un sustrato peptídico que se parezca al sustrato diana, y seleccionando una o más variantes de proteasa de la población de proteasas con respecto a su especificidad para el sustrato proporcionado. La selección se realiza en condiciones que permiten la identificación de proteasas que reconocen e hidrolizan preferiblemente la secuencia diana.

En particular, la realización (1) de la invención se refiere a un procedimiento para generar proteasas específicas de secuencia con especificidades de sustrato objetivo, en el que se llevan a cabo las etapas siguientes:

55

(a) proporcionar una población de proteasas, en la que cada variante está relacionada con una o más primeras proteasas, teniendo estas primeras proteasas una primera especificidad de sustrato;

60

(b) proporcionar uno o más sustratos peptídicos que comprenden al menos una secuencia de aminoácidos que se parece al sustrato peptídico diana;

(c) seleccionar una o más variantes de proteasa de la población de proteasas proporcionada en la etapa (a) con respecto a su especificidad para el sustrato proporcionado en la etapa (b) en condiciones que permiten la identificación de proteasas que reconocen y rompen preferiblemente la secuencia diana;

65

y en el que las etapas (a) a (c) se llevan a cabo de manera cíclica hasta que se identifican una o más variantes de proteasa con la especificidad de sustrato objetivo.

## ES 2 289 288 T3

Cuando se repiten las etapas (a) a (c), se usan la una o más proteasas seleccionadas en la etapa (c) de un ciclo como la una o más primeras proteasas en la etapa (a) del ciclo siguiente.

5 En una alternativa de la invención, la una o más primeras proteasas que sirven como puntos de partida en la etapa (a) del procedimiento tienen una especificidad de secuencia alta que se mantiene alta durante la evolución dirigida hacia la especificidad objetivo.

10 En otra alternativa del procedimiento, la una o más primeras proteasas que sirven como puntos de partida en la etapa (a) del procedimiento tienen una especificidad de secuencia baja, que aumenta durante la evolución dirigida hacia la especificidad objetivo.

15 Las etapas (a) a (c) del procedimiento anterior se llevan a cabo durante al menos un ciclo. Preferiblemente, sin embargo, estas etapas se llevan a cabo durante varios ciclos, siendo cada una o más variantes de proteasa seleccionadas en un ciclo el origen de la población de variantes de proteasa en el ciclo siguiente. Se llevan a cabo preferiblemente más de un ciclo y menos de cien ciclos, más preferiblemente más de dos ciclos y menos de cincuenta ciclos, de manera particularmente preferible más de tres ciclos y menos de veinte ciclos, de manera especialmente preferible más de cuatro ciclos y menos de diez ciclos, y lo más preferiblemente cinco ciclos de las etapas (a) a (c) hasta que se identifica una variante de proteasa o más variantes de proteasa con la especificidad de sustrato objetivo.

20 La invención aplica medios de evolución tal como se describe con mucho detalle en el documento WO9218645 incorporándose ese documento en su totalidad para todos los fines.

25 Para una visión global sobre la aplicación de principios de evolución a la biotecnología molecular, que usualmente se denomina "evolución dirigida" o "biotecnología de evolución", véase la revisión por Koltermann y Kettling (Biophys. Chem. 66 (1997) 159-177).

30 Parte de la invención es proporcionar poblaciones de variantes de proteasa en las que cada variante está relacionada con una o más primeras proteasas. En principio, puede haber un gran número de estas primeras proteasas, siendo todas juntas el origen para el primer ciclo del procedimiento. Sin embargo se prefiere que estas primeras proteasas comprendan cincuenta o menos proteasas diferentes, más preferiblemente diez o menos proteasas diferentes, de manera especialmente preferible dos o menos proteasas diferentes. Lo más preferiblemente, sólo se emplea una primera proteasa.

35 Según la invención, puede usarse cualquier proteasa como primera proteasa. Preferiblemente, se usa una endoproteasa como primera proteasa. Se prefiere que la proteasa pertenezca al grupo de proteasas constituido por serina proteasas (EC 3.4.21), cisteína proteasas (EC 3.4.22), aspártico proteasas (EC 3.4.23), y metaloproteasas (EC 3.4.24). Las primeras proteasas se caracterizan por su capacidad para reconocer e hidrolizar sustratos peptídicos con una cierta especificidad cualitativa y cuantitativa. Las primeras proteasas pueden tener una especificidad en el mismo intervalo que la especificidad de la proteasa que ha de generarse. Los ejemplos de proteasas con especificidades relativamente altas son proteasa de VGT, proteasa de VIH-1, proteasa BAR1, factor Xa, trombina, activador del plasminógeno de tipo tisular, proteasa Kex2, proteasa de VMVT, proteasa de VRS, proteasa de VLM, proteasa de VMPPM, proteasa de VTMM, proteasa de VLB, proteasa de VAIE, proteasa de VISmac. Como alternativa, las primeras proteasas tienen una especificidad menor que la especificidad de la proteasa que ha de generarse. Como un ejemplo extremo de las últimas, se emplean proteasas con especificidad de secuencia muy baja, por ejemplo proteasas tales como papaína, tripsina, quimotripsina, subtilisina, SET (la serina proteasa de tipo tripsina de *Streptomyces erythraeus*), elastasa, catepsina G o quimasa.

50 Una proteasa particularmente adecuada es proteasa BAR1 sp IP12630 (precursor de pepsina de barrera BAR1\_YEAST (EC 3.4.23.35) (proteína de "barrera" extracelular) (proteína BAR) de *S. cerevisiae* (véase la SEC ID NO:8).

55 Proporcionar poblaciones de proteasas se realiza esencialmente tal como se describe en el documento WO9218645. Según la invención, los genes que codifican variantes de proteasa se unen en un vector de expresión adecuado mediante técnicas de clonación molecular convencionales (Sambrook, J.F.; Fritsch, E.F.; Maniatis, T.; Cold Spring Harbor Laboratory Press, segunda edición, 1989, Nueva York). El vector se introduce en una célula huésped de expresión adecuada, que expresa la variante de proteasa correspondiente. Los huéspedes de expresión particularmente adecuados son los huéspedes de expresión bacterianos tales como *Escherichia coli* o *Bacillus subtilis*, o huéspedes de expresión de levadura tales como *Saccharomyces cerevisiae* o *Pichia pastoris*, o huéspedes de expresión de mamífero tales como líneas celulares de ovario de hámster chino (CHO, *Chinese Hamster Ovary*) o riñón de hámster recién nacido (BHK, *Baby Hamster Kidney*), o sistemas de expresión virales tales como el sistema de baculovirus. Como alternativa, pueden usarse sistemas para la expresión de proteínas *in vitro*.

65 En una realización preferida de la invención, los genes se unen en el vector de expresión detrás de una secuencia señal adecuada que conduce a la secreción de las variantes de proteasa en el espacio extracelular, permitiendo de este modo la detección directa de la actividad proteasa en el sobrenadante celular. Las secuencias señal particularmente adecuadas para *Escherichia coli* son HlyA, para *Bacillus subtilis* AprE, NprB, Mpr, AmyA, AmyE, Blac, SacB, y para *S. cerevisiae* Bar1, Suc2, Mata, Inu1A, Ggplp.

## ES 2 289 288 T3

En otra realización preferida de la invención, las variantes de proteasa se expresan intracelularmente y los sustratos peptídicos se expresan también intracelularmente. Preferiblemente, esto se realiza esencialmente tal como se describe en el documento WO 0212543, usando un sustrato peptídico de fusión que comprende dos proteínas autofluorescentes unidas mediante la secuencia de aminoácidos del sustrato.

5 En otra realización preferida de la invención, las variantes de proteasa se expresan intracelularmente, o se secretan al espacio periplasmático usando secuencias señal tales como DsbA, PhoA, PelB, OmpA, OmpT o gIII para *Escherichia coli*, seguido de una etapa de permeabilización o lisis para liberar las variantes de proteasa al sobrenadante. La destrucción de la barrera de membrana puede forzarse mediante el uso de medios mecánicos tales como ultrasonidos, prensa francesa, o el uso de enzimas que digieren la membrana tales como lisozima.

10 Como alternativa adicional, los genes que codifican las variantes de proteasa se expresan libres de células mediante el uso de un sistema de expresión libre de células adecuado. En una realización particularmente preferida, el extracto S30 de células de *Escherichia coli* se usa para este fin tal como se describe por Lesly *et al.* (Methods in Molecular Biology 37 (1995) 265-278).

15 La relación con la una o más primeras proteasas puede conseguirse mediante diversos procedimientos. Por ejemplo, los genes que codifican la una o más primeras proteasas se modifican mediante procedimientos para mutagénesis al azar de ácidos nucleicos. En una realización preferida de la invención, la mutagénesis al azar se consigue mediante el uso de una polimerasa tal como se describe en el documento WO 9218645. Según esta realización, el uno o más genes que codifican la una o más primeras proteasas se amplifican mediante el uso de una polimerasa con una tasa de error alta, o en condiciones que aumentan la tasa de incorporaciones erróneas, conduciendo de este modo a una población de genes en la que cada gen codifica una proteasa que está relacionada con la una o más primeras proteasas. Por ejemplo puede emplearse el procedimiento según Cadwell, R.C y Joyce, G.F. (PCR Methods Appl. 2 (1992) 28-33). Otros procedimientos para la mutagénesis al azar que pueden emplearse hacen uso de cepas mutadoras, radiación UV o mutágenos químicos. Lo más preferiblemente, se introducen errores en el gen en o cerca pero por debajo del umbral de error tal como se describe en el documento WO 9218645.

20 En otra realización preferida de la invención, ciertas partes del gen que codifica las variantes de proteasa se aleatorizan completamente con respecto a la secuencia de aminoácidos, y vuelven a introducirse en el gen como un casete de oligonucleótido. Esta técnica se denomina usualmente mutagénesis con casete (Oliphant, A.R. *et al.*, Gene 44 (1986) 177-183; Horwitz, M.S., *et al.* Genome 31 (1989) 112-117). En una realización particularmente preferida de la invención, la parte del gen que codifica restos de aminoácido que son esenciales para el reconocimiento del sustrato se aleatoriza por medio de mutagénesis con casete. Estos restos pueden identificarse a partir de estudios estructurales. En particular, la mutagénesis con casete se dirige a restos que comprenden partes del bolsillo de unión al sustrato. Como alternativa, la sustitución de cada resto de aminoácido por una alanina, y el análisis de si hay un efecto sobre la actividad catalítica puede identificar tales restos. Como alternativa adicional, estos restos pueden identificarse introduciendo en primer lugar mutaciones al azar en el gen, explorando para detectar un efecto sobre la especificidad, la afinidad o actividad catalítica, y determinando después la posición de las mutaciones en las variantes que representan la especificidad alterada, la afinidad o actividad catalítica alterada. Como un extremo de este enfoque, la secuencia aleatorizada completamente puede tener la longitud de sólo un nucleótido. Este enfoque se denomina normalmente mutagénesis de saturación de sitio.

30 En otra realización preferida de la invención, se introducen al azar secuencias de ácido nucleico en, o se delecionan del, uno o más primeros genes de proteasa con el fin de proporcionar una población de proteasas. Este enfoque se denomina mutagénesis por inserción y/o delección. Para la mutagénesis por inserción, se introducen al azar secuencias aleatorias de longitud definida o aleatoria en un gen. Como ejemplo, el procedimiento descrito por Hallet *et al.* (Nucleic Acids Res. 1997, vol. 25, págs. 1866ss) puede usarse para introducir al azar una secuencia de 15 nt ("nucleótidos") aleatoria en un gen. Como alternativa, pueden insertarse al azar secuencias definidas, por ejemplo una secuencia que codifica un motivo de estructura secundaria de una proteína específica, en un gen. Como alternativa, pueden insertarse secuencias aleatorias de longitud definida o aleatoria en sitios específicos en un gen. Esto puede realizarse usando sitios de restricción o mediante procedimientos de extensión del solapamiento de oligonucleótidos tales como el procedimiento descrito por Horton (Gene 1989, vol. 77, págs. 61ss). Para la mutagénesis por delección, se delecionan al azar secuencias de longitud definida o aleatoria de un gen. En una realización particular de la invención, se combinan la mutagénesis por delección y por inserción de tal modo que las inserciones en un sitio pueden combinarse potencialmente, y de este modo compensarse posiblemente, mediante la delección en otro sitio.

45 En una realización preferida adicional de la invención, se usan procedimientos para la recombinación homóloga *in vitro* para proporcionar poblaciones de proteasa. Los ejemplos de procedimientos que pueden aplicarse son la reacción en cadena de recombinación (RCR) según el documento WO 0134835, el procedimiento de intercambio de ADN según el documento WO 9522625, el procedimiento de extensión escalonado según el documento WO 9842728, o la recombinación de cebado al azar según el documento WO9842728. Además, también pueden aplicarse procedimientos para la recombinación no homóloga tales como el procedimiento de Itchy (Ostermeier, M. *et al.*, Nature Biotechnology 17 (1999) 1205-1209). Todas las referencias mencionadas anteriormente se incorporan al presente documento por referencia en su totalidad para todos los fines.

65 En realizaciones adicionales de la invención, los procedimientos mencionados anteriormente se combinan entre sí. En una realización particularmente preferida, se combina la reacción en cadena de recombinación con mutagénesis al

## ES 2 289 288 T3

azar tales como PCR propensa a error según Cadwell, R.C y Joyce, G.F. (PCR Methods Appl. 2 (1992) 28-33) con el fin de desacoplar las mutaciones seleccionadas en la vuelta anterior y para introducir simultáneamente un número definido de nuevas mutaciones al azar en la población.

5 El acoplamiento del genotipo y fenotipo de las proteasas se consigue mediante el uso de portamuestras que posibilitan la compartimentación de muestras, y la distribución de genotipos en portamuestras se realiza a una multiplicidad por compartimento que permite una diferenciación suficiente de fenotipos.

10 La una o más primeras proteasas que sirven como punto de partida del procedimiento tienen una especificidad que está en el intervalo de la especificidad objetivo que ha de generarse mediante el procedimiento o bien tienen una especificidad menor que la especificidad objetivo. Por consiguiente, el procedimiento de la invención se realiza en condiciones que mantienen cuantitativamente la especificidad y la alteran cualitativamente (alternativa A), o bien el procedimiento de la invención se realiza en condiciones que mantienen cualitativamente la especificidad y la aumentan cuantitativamente (alternativa B). Además, ambos enfoques pueden combinarse. Estas tres alternativas principales se muestran esquemáticamente en la figura 2.

15 En una realización preferida de la invención que corresponde a la alternativa A, la una o más primeras proteasas tienen una primera especificidad que está cuantitativamente en el intervalo de la especificidad objetivo, pero es cualitativamente distinta de la especificidad objetivo. Las proteasas que tienen la especificidad de sustrato objetivo se consiguen usando el procedimiento de la invención seleccionando variantes de proteasa en condiciones que permiten la identificación de proteasas que reconocen y rompen preferiblemente la secuencia diana.

20 En otra realización preferida de la invención que corresponde a la alternativa B, la especificidad de la una o más primeras proteasas es cuantitativamente menor cuando se compara con la especificidad objetivo. Esto significa que aceptan e hidrolizan un gran número de sustratos peptídicos. Esta primera especificidad baja se aumenta posteriormente mediante el procedimiento de la invención hasta que está en el intervalo de la especificidad objetivo. Como variante preferida de esta realización, la primera especificidad está cualitativamente relacionada con la especificidad objetivo. Por tanto, el gran número de sustratos peptídicos que se acepta e hidroliza incluye ya el sustrato diana. Por consiguiente, los restos de aminoácido que son esenciales en el primer sustrato siguen siendo restos esenciales en el sustrato diana. Entonces, las proteasas que tienen la especificidad de sustrato objetivo se consiguen usando el procedimiento de la invención seleccionando variantes de proteasa en condiciones que permiten la identificación de proteasas que reconocen y rompen preferiblemente la secuencia diana.

25 Otra parte de la invención es proporcionar sustratos peptídicos que se parecen al sustrato diana y usar estos sustratos para explorar variantes de proteasa con respecto a su actividad catalítica.

30 En una realización preferida de la invención, se sintetizan sustratos peptídicos adecuados por medio del enfoque de síntesis de péptidos en fase sólida de Merrifield *et al* (Nature. 207 (1965) 522-523). Estos sustratos peptídicos se incuban entonces durante cierto tiempo en un tampón de muestra que contiene la variante de proteasa que ha de someterse a prueba. Entonces se analiza la hidrólisis del péptido mediante un procedimiento adecuado. Por ejemplo, la cantidad de péptidos fragmentados puede analizarse mediante cromatografía. En particular, los fragmentos de péptido se analizan ventajosamente en un sistema de HPLC en fase inversa. Como alternativa, el sustrato peptídico se modifica de cualquier manera para posibilitar el análisis de la hidrólisis de péptido. En particular, el sustrato peptídico puede llevar grupos funcionales que posibilitan la detección de la hidrólisis del sustrato. Tales grupos funcionales incluyen, pero no se limitan a, los siguientes:

35 uno o más fluoróforos o cromóforos, cuyas propiedades espectroscópicas cambian tras la hidrólisis del péptido, realizándose la exploración mediante la determinación del cambio en las propiedades espectroscópicas; o

40 dos fluoróforos que pueden distinguirse por sus propiedades de fluorescencia y que están unidos a extremos opuestos del segundo sustrato, realizándose la exploración mediante espectroscopía de fluorescencia confocal a concentraciones de fluoróforo inferiores a 1  $\mu\text{M}$ ; o

45 dos fluoróforos que forman un par de transferencia de energía por resonancia de fluorescencia (FRET) y que están unidos a extremos opuestos del segundo sustrato, realizándose la exploración mediante la determinación de la disminución en la transferencia de energía entre los dos fluoróforos; o

50 una primera y segunda proteína autofluorescente que flanquean el segundo sustrato, realizándose la exploración mediante espectroscopía de fluorescencia confocal a concentraciones de sustrato inferiores a 1  $\mu\text{M}$ ; o

55 un fluoróforo y una molécula inhibidora que están unidos a extremos opuestos del segundo sustrato, realizándose la exploración mediante la determinación de la disminución en la inhibición del fluoróforo; o

60 un fluoróforo o un cromóforo y un resto de unión que están unidos a extremos opuestos del segundo sustrato, realizándose la exploración mediante la determinación de la unión del resto de unión a una pareja de unión específica; o

65 un marcador radiactivo y un resto de unión que están unidos a extremos opuestos del segundo sustrato, realizándose la exploración mediante el uso de un ensayo de proximidad de centelleo; o

cualquier combinación de los mismos.

Con respecto a los grupos funcionales mencionados anteriormente, puede unirse un grupo químico al péptido que altera sus propiedades cuando se hidroliza el péptido. Por ejemplo, puede usarse un grupo para-nitrofenilo para este fin. Como otro ejemplo, uno o más fluoróforos y/o una molécula inhibidora están unidas al péptido, y la cantidad de péptido fragmentado se analiza midiendo una diferencia en la fluorescencia de los fluoróforos. Por ejemplo, dos fluoróforos que son adecuados para formar un par de FRET (transferencia de energía por resonancia de fluorescencia) están unidos al péptido en extremos opuestos, y la hidrólisis del péptido se mide mediante una disminución en la transferencia de energía entre los dos fluoróforos. Por ejemplo, pueden usarse rodamina verde (Molecular Probes Inc., Oregón, EE.UU.) y tetrametilrodamina (Molecular Probes Inc., Oregón, EE.UU.) como fluoróforos que son adecuados para formar un par de FRET de este tipo.

En una realización particularmente preferida de la invención, dos fluoróforos que no forman un par de FRET sustancial están unidos a extremos opuestos de sustratos peptídicos sintéticos. Como ejemplo, pueden usarse rodamina verde (Molecular Probes Inc., Oregón, EE.UU.) y Cy-5 (Amersham Biosciences Europe GmbH, Friburgo) para este fin, y puede conseguirse la unión covalente del tinte por medio de un enlace de éster succinimidílico a un grupo amino primario del péptido. La hidrólisis de estos péptidos se analiza preferiblemente mediante medios de espectroscopía de fluorescencia confocal según las solicitudes de patente WO 9416313 y WO9613744, que se incorporan en el presente documento por referencia en su totalidad para todos los fines. Debido a la sensibilidad alta de espectroscopía de fluorescencia confocal, se usan sustratos en concentraciones inferiores a uno micromolar, más preferiblemente inferiores a cien nanomolar, y lo más preferiblemente inferiores a diez nanomolar. Por tanto, la exploración según esta realización se realiza sustancialmente por debajo de la  $K_M$  de proteasas típicas.

En otra realización particularmente preferida de la invención, se usan como sustratos peptídicos proteínas de fusión que comprenden una primera proteína autofluorescente, un péptido y una segunda proteína autofluorescente. Según el documento WO0212543, que se incorpora en el presente documento por referencia en su totalidad para todos los fines, autofluorescente incluye la proteína fluorescente verde GFP (*Green Fluorescent Protein*) y sus mutantes, así como dsRED y sus mutantes. Las proteínas de fusión pueden producirse mediante la expresión de un gen de fusión adecuado en *E. coli*, lisis de células y purificación de la proteína de fusión mediante procedimientos convencionales tales como cromatografía de intercambio iónico o cromatografía de afinidad.

Es una parte esencial de la invención que las proteasas con la especificidad de sustrato objetivo se generan seleccionando variantes de proteasa en condiciones que permiten la identificación de proteasas que reconocen y rompen preferiblemente la secuencia diana. Esta selección puede conseguirse según los diferentes aspectos de la invención tal como se explica resumidamente a continuación.

En un primer aspecto de la invención, se identifican proteasas que reconocen y rompen preferiblemente la secuencia diana explorando para determinar proteasas con una afinidad alta para la secuencia de sustrato diana. Una afinidad alta corresponde a una  $K_M$  baja que se selecciona mediante exploración a concentraciones de sustrato diana sustancialmente por debajo de la  $K_M$  de la primera proteasa. Este aspecto se denomina el “enfoque de afinidad”.

En una realización preferida de este aspecto de la invención, el sustrato peptídico proporcionado en la etapa (b) está unido a uno o más fluoróforos que posibilitan la detección de la hidrólisis del sustrato peptídico a concentraciones inferiores a 10  $\mu$ M, preferiblemente inferiores a 1  $\mu$ M, más preferiblemente inferiores a 100 nM, y lo más preferiblemente inferiores a 10 nM.

En un segundo aspecto de la invención, se identifican las proteasas que sólo reconocen y rompen la secuencia diana proporcionando dos o más sustratos peptídicos en la etapa (b) y explorando para determinar la actividad sobre estos dos o más sustratos peptídicos en comparación. Este aspecto se denomina el “enfoque de comparación”.

En una realización preferida de este aspecto de la invención, los dos o más sustratos peptídicos proporcionados en la etapa (b) están unidos a diferentes moléculas marcadoras, posibilitando de este modo la detección de la rotura de los dos o más sustratos peptídicos consecutivamente o en paralelo. En una realización particularmente preferida de la invención, se proporcionan dos sustratos peptídicos en la etapa (b), teniendo un sustrato peptídico una secuencia de aminoácidos idéntica, o que se parece, al primer sustrato peptídico posibilitando de ese modo el control de la actividad original de las primeras proteasas, y teniendo el otro sustrato peptídico una secuencia de aminoácidos idéntica, o que se parece, a la secuencia de sustrato diana posibilitando de ese modo el control de la actividad sobre el sustrato diana. En una realización especialmente preferida de la invención, estos dos sustratos peptídicos están unidos a moléculas marcadoras fluorescentes, y las propiedades fluorescentes de los dos sustratos peptídicos son suficientemente diferentes con el fin de distinguir ambas actividades cuando se miden consecutivamente o en paralelo. Por ejemplo, puede usarse una proteína de fusión que comprende una primera proteína autofluorescente, un péptido, y una segunda proteína autofluorescente según la solicitud de patente WO 0212543 para este fin. Como alternativa, los fluoróforos tales como las rodaminas están unidos químicamente a los sustratos peptídicos.

En un tercer aspecto de la invención, se identifican las proteasas que reconocen y rompen preferiblemente la secuencia diana proporcionando en la etapa (b) uno o más sustratos peptídicos que se parecen al péptido diana junto con sustratos peptídicos competidores en gran exceso. Entonces se realiza la exploración con respecto a la actividad sobre los sustratos que se parecen al sustrato diana en presencia de los sustratos competidores. Las proteasas que tienen

una especificidad que corresponde cualitativamente a la especificidad objetivo, pero que tienen sólo una especificidad cuantitativa baja se identifican como muestras negativas en una exploración de este tipo. Mientras que las proteasas que tienen una especificidad que corresponde cualitativa y cuantitativamente a la especificidad objetivo se identifican positivamente. Este aspecto se denomina el “enfoque de competidor”.

En una realización preferida de este aspecto de la invención, el uno o más sustratos peptídicos que se parecen al sustrato diana están unidos a moléculas marcadoras, posibilitando de ese modo la detección de su hidrólisis, mientras que los sustratos peptídicos competidores no llevan moléculas marcadoras. Los sustratos peptídicos competidores tienen una secuencia de aminoácidos idéntica, o que se parece, al primer sustrato peptídico, o tienen secuencias de aminoácidos aleatorias, actuando de este modo como inhibidores competidores para la hidrólisis de los sustratos peptídicos que llevan marcadores.

En un cuarto aspecto de la invención, se identifican las proteasas que reconocen y rompen preferiblemente la secuencia diana usando sustratos intermedios para hacer evolucionar la proteasa hacia la especificidad de sustrato objetivo. Este aspecto se denomina también a continuación en el presente documento el “enfoque intermedio”. En una primera variante de este aspecto de la invención, esto se consigue proporcionando en ciclos diferentes sustratos peptídicos diferentes, teniendo cada sustrato peptídico un rasgo intermedio con respecto al ciclo anterior y el sustrato peptídico diana. Según esta variante, las proteasas evolucionan gradualmente hacia la especificidad objetivo. La figura 4 representa esquemáticamente el principio básico de esta variante del enfoque intermedio.

Más generalmente, una primera variante de este aspecto de la invención se refiere a un procedimiento para generar proteasas específicas de secuencia con una especificidad de sustrato objetivo, en el que se llevan a cabo las etapas siguientes:

(a) proporcionar una población de proteasas, en la que cada variante está relacionada con una o más primeras proteasas, teniendo estas primeras proteasas especificidad para un espectro de sustratos peptídicos o un único sustrato peptídico;

(b) proporcionar uno o más sustratos peptídicos que tienen un rasgo intermedio con respecto al primer sustrato peptídico y al sustrato diana;

(c) seleccionar una o más variantes de proteasa de la población de proteasas proporcionada en la etapa (a) con respecto a su especificidad para el sustrato proporcionado en la etapa (b);

(d) repetir las etapas (a) a (c) hasta que se identifican una o más variantes de proteasa con actividad para el sustrato intermedio proporcionado en la etapa (b);

(e) sustituir el primer sustrato peptídico en las etapas (a) y (b) por el sustrato intermedio, y las primeras proteasas en la etapa (a) por las variantes de proteasa seleccionadas en la etapa (c);

y repetir las etapas (a) a (e) hasta que se identifican una o más variantes de proteasa con la especificidad de sustrato objetivo.

En esta primera variante de este aspecto de la invención, la evolución de la especificidad de proteasa se dirige por medio de selección consecutiva sobre un cierto número de sustratos peptídicos intermedios, pareciéndose todos los sustratos peptídicos cada vez más a la secuencia peptídica diana. Este enfoque se basa en el hallazgo de que las proteasas que aceptan sustratos relacionados también están usualmente relacionadas entre sí. La relación de las proteasas en el contexto de esta invención es una medición de la homología de las secuencias de aminoácidos de dos o más enzimas. Además, este enfoque se basa en el descubrimiento sorprendente de que los subsitios que pueden distinguirse en un sitio activo de proteasa pueden evolucionar por separado y que su estructura molecular puede atribuirse a restos diferentes de un sustrato peptídico (Schlechter y Berger, *Biochem. Biophys. Res. Commun.* 27 (1967) 157-162).

Los sustratos intermedios pueden realizarse sustituyendo restos de aminoácido en una o más posiciones de la primera secuencia peptídica con restos de aminoácido en las mismas posiciones de la secuencia peptídica diana. Tales productos intermedios se denominan “productos intermedios de composición de aminoácido”. Adicionalmente, un sustrato peptídico intermedio puede incluir uno o más restos de aminoácido en una o más posiciones que no son ni los restos de la primera secuencia peptídica ni los restos de la secuencia peptídica diana en esa posición, sino que son restos de aminoácido con un rasgo intermedio con respecto a los restos en el primer sustrato y el sustrato diana. Tales productos intermedios se denominan “productos intermedios de propiedades de aminoácido”. El rasgo intermedio de este tipo de productos intermedios puede basarse en uno o más parámetros físicos y químicos, que incluyen, pero no se limitan a, la superficie del resto, su volumen, el punto isoeléctrico, el pKa de cadena lateral, la polaridad, la capacidad para formar enlaces de hidrógeno o la hidrofobicidad. En la tabla siguiente, se clasifican los veinte restos de aminoácido que se producen de manera natural según estos parámetros.

# ES 2 289 288 T3

TABLA III

*Clasificación de los 20 restos de aminoácido que se producen de manera natural*

| Resto de aminoácido | Tipo | Superficie <sup>a</sup><br>[Å <sup>2</sup> ] | Volumen <sup>b</sup><br>[Å <sup>3</sup> ] | pK <sub>a</sub> de cadena lateral <sup>c</sup><br>(carga) <sup>d</sup> | Hidrofobicidad relativa <sup>e</sup> | Aceptor o donador de enlace de hidrógeno |   |
|---------------------|------|--|---|--|--------------------------------------|--|---|
| A                   | Ala  | alifático                                    | 115                                       | 88,6   | -                                    | 0,62                                     |   |
| C                   | Cys  | alifático                                    | 135                                       | 108,5  | 9,1-9,5                              | 0,68                                     | + |
| D                   | Asp  | alifático                                    | 150                                       | 111,1  | 4,5 (-)                              | 0,03                                     | + |
| E                   | Glu  | alifático                                    | 190                                       | 138,4  | 4,6 (-)                              | 0,04                                     | + |
| F                   | Phe  | aromático                                    | 210                                       | 189,9  | -                                    | 1,00                                     |   |
| G                   | Gly  | alifático                                    | 75  | 60,1   | -                                    | 0,50                                     |   |
| H                   | His  | alifático                                    | 195                                       | 153,2  | 6,2                                  | 0,17                                     | + |
| I                   | Ile  | alifático                                    | 175                                       | 166,7  | -                                    | 0,94                                     |   |
| K                   | Lys  | alifático                                    | 200                                       | 168,6  | 10,4 (+)                             | 0,28                                     | + |
| L                   | Leu  | alifático                                    | 170                                       | 166,7  | -                                    | 0,94                                     |   |
| M                   | Met  | alifático                                    | 185                                       | 162,9  | -                                    | 0,74                                     |   |
| N                   | Asn  | alifático                                    | 160                                       | 114,1  | -                                    | 0,24                                     | + |
| P                   | Pro  | alifático                                    | 145                                       | 112,7  | -                                    | 0,71                                     |   |
| Q                   | Gln  | alifático                                    | 180                                       | 143,8  | -                                    | 0,25                                     | + |
| R                   | Arg  | alifático                                    | 225                                       | 173,4  | ~ 12 (+)                             | 0,00                                     | + |
| S                   | Ser  | alifático                                    | 115                                       | 89,0   | -                                    | 0,36                                     | + |
| T                   | Thr  | alifático                                    | 140                                       | 116,1  | -                                    | 0,45                                     | + |
| V                   | Val  | alifático                                    | 155                                       | 140,0  | -                                    | 0,83                                     |   |
| W                   | Trp  | aromático                                    | 255                                       | 227,8  | -                                    | 0,88                                     | + |
| Y                   | Tyr  | aromático                                    | 230                                       | 193,6  | 9,7                                  | 0,88                                     | + |

<sup>a</sup> Chothia, C., J. Mol. Biol., 105 (1975) 1-14; <sup>b</sup> Zamyatin, A.A., Prog. Biophys. Mol. Biol., 24 (1972) 107-123; <sup>c</sup> Tanford, C., Adv. Prot. Chem., 17 (1962) 69-165; <sup>d</sup> carga a pH fisiológico; <sup>e</sup> Black, S.D, Mould, D.R, Anal. Biochem., 193 (1991) 72-82.

Si, por ejemplo, el primer sustrato fuera ALY y el sustrato diana fuera NRF, los sustratos intermedios con respecto a la composición de aminoácidos serían, por ejemplo, ALF, NRY o ARF (se indican las modificaciones con respecto al primer sustrato). Un sustrato intermedio con respecto a las propiedades de aminoácido sería, por ejemplo, AQF, pareciéndose el resto de glutamina a la leucina original en el sentido de que no está cargada, pero pareciéndose más al resto de arginina del sustrato diana con respecto a la hidrofobicidad y su capacidad para formar enlaces de hidrógeno. Un ejemplo adicional para este enfoque sería SLY en el que S se parece a A con respecto al volumen y la superficie pero es más similar al N objetivo en cuanto a la hidrofobicidad y el enlace de hidrógeno.

Además, el número de sustratos peptídicos consecutivos que ha de usarse depende de la relación de la primera secuencia peptídica y la secuencia peptídica diana así como la especificidad cuantitativa de la una o más primeras proteasas. Cuanto menos relacionadas estén la primera secuencia peptídica y la secuencia peptídica diana, y cuando mayor sea la especificidad de la una o más primeras proteasas, más sustratos peptídicos intermedios consecutivos se requieren.

En una segunda variante de este aspecto de la invención, las proteasas diferentes que tienen especificidad para productos intermedios diferentes se seleccionan en paralelo en una primera etapa del procedimiento. En una segunda etapa, las proteasas que tienen la especificidad objetivo se seleccionan entonces de una población que contiene quimeras recombinadas al azar de la proteasas seleccionadas en la primera etapa. Preferiblemente, la recombinación de las diferentes proteasas seleccionadas en paralelo se consigue empleando una técnica de recombinación homóloga *in vitro*, tal como la reacción en cadena de recombinación descrita en la solicitud de patente WO 0134835. Ambas formas de productos intermedios pueden usarse para esta variante. Sin embargo, se emplean preferiblemente productos intermedios de composición de aminoácido. La figura 11 muestra esquemáticamente el principio básico de esta variante del cuarto aspecto de la invención.

Los diferentes productos intermedios empleados en la primera etapa de esta variante se eligen preferiblemente de una manera que la suma de todas las modificaciones introducidas en estos productos intermedios es igual o se parece a las características del sustrato diana. Como ejemplo, si el primer sustrato fuera ALY y el sustrato diana fuera NRF, los productos intermedios de composición de aminoácido adecuados para esta realización serían NLY, ARY, y ALF. Las proteasas que tienen especificidad para estos tres sustratos se recombinarían entonces al azar y se explorarían para determinar la especificidad hacia el sustrato diana de este NRF de ejemplo. Otros ejemplos, incluyendo los productos intermedios con más de una modificación, han de construirse de manera análoga.

En aspectos adicionales de la invención, dos, tres o los cuatro diferentes aspectos mencionados anteriormente se combinan entre sí. En una combinación preferida, la exploración para determinar proteasas con constantes de Michaelis-Menten reducidas se combina con el uso de sustratos intermedios. En otra combinación preferida, la exploración para determinar proteasas con constantes de Michaelis-Menten reducidas se combina con la exploración de dos o más sustratos consecutivos o en paralelo. En una combinación preferida adicional, la exploración para determinar proteasas con constantes de Michaelis-Menten reducidas se combina con el uso de un exceso de sustrato competidor, no marcado. En una combinación particularmente preferida, se combinan entre sí los cuatro aspectos, la exploración para determinar proteasas con constantes de Michaelis-Menten reducidas, la exploración de dos o más sustratos en paralelo, el uso de un exceso de sustrato competidor, no marcado y el uso de sustratos intermedios.

En una realización particularmente preferida de la invención la proteasa objetivo tiene una especificidad similar al activador del plasminógeno de tipo tisular y rompe el sustrato diana CPGR<sup>1</sup>VVGG. Tal proteasa objetivo puede generarse, entre otros, mediante el procedimiento definido anteriormente de la invención cuando la proteasa de partida es proteasa BAR1 de *S. cerevisiae*. En tal procedimiento se utilizan preferiblemente los siguientes sustratos segundos/intermedios:

- (i) WLGLVPGG
- (ii) WLGQVPGG
- (iii) WLGRVPGG
- (iv) WLGRVVGG
- (v) CPGRVVGG.

La proteasa específica de secuencia que puede obtenerse mediante los procedimientos descritos anteriormente en el presente documento tiene preferiblemente una especificidad similar al activador del plasminógeno de tipo tisular y rompe el sustrato diana CPGR<sup>1</sup>VVGG. La proteasa de partida es preferiblemente proteasa BAR1 que incluye, pero no se limita a, la representada en SEC ID NO: 8. Adicionalmente, pueden usarse como proteasas de partida proteasas BAR1 modificadas mediante truncamiento de hasta 200 aa, preferiblemente en el intervalo de 100 aa a 200 aa, más preferiblemente en el intervalo de 120 aa a 180 aa, lo más preferiblemente en el intervalo de 140 aa a 160 aa en el extremo C- o N-terminal. Incluso más preferiblemente la proteasa específica de secuencia se deriva de dicha proteasa derivada de BAR1 y tiene al menos una mutación seleccionada del grupo que comprende las modificaciones L33I, Y45D, T47A, T59I, N82D, E96V, M107I, N123D, E143D, N151V, I152F, K161E, A163T, T165A, R178S, T221I, E231V, D321N, D367G, M369L, V370I, A399S, K404R y S440L. Se prefieren particularmente entre dichas proteasas las que tienen al menos una de D367G, V370I, M107I, I152F, E143D y E231V. Los mutantes particularmente preferidos de proteasa BAR1 pueden modificarse adicionalmente por ejemplo mediante truncamiento de hasta 10 aa en los extremos C- o N-terminales de los mismos o mediante delección, inserción o sustitución de hasta 50 aa, preferiblemente hasta 20 aa, lo más preferiblemente hasta 10 aa dentro de su secuencia.

## 60 Descripción detallada de las figuras

La figura 1 representa esquemáticamente las dos alternativas A y B del procedimiento de la invención. Comenzando con una primera proteasa, el objetivo de la invención es la generación de una proteasa evolucionada con una especificidad alta por un sustrato peptídico diana que se caracteriza por su secuencia de aminoácidos.

Para el fin de esta figura, diferentes formas representan diferentes restos de aminoácido, y el perfil inverso de las formas representa los sitios de reconocimiento de proteasa, respectivamente. Las formas con una línea ondulada en la parte superior representan cualquier resto de aminoácido en esa posición. El sitio activo de la enzima se

indica mediante un asterisco, y la flecha indica el sitio de rotura en el sustrato. El tipo de la una o más proteasas usadas como primeras proteasas define si ha de emplearse la alternativa A o la alternativa B. En la alternativa A, la primera proteasa se caracteriza por una ya alta especificidad hacia un primer sustrato definido. Según el procedimiento de la invención, esta especificidad ha de cambiarse cualitativamente en la especificidad objetivo. En la alternativa B, la primera proteasa tiene una especificidad relativamente baja, es decir, no distingue entre un conjunto de sustratos que se diferencian, por ejemplo, en las posiciones P2, P1', y/o P2'. Mediante el procedimiento de la invención, sólo se aumenta la especificidad cuantitativa de aquellas proteasas hacia el valor de la especificidad objetivo.

La figura 2 distingue las dos alternativas A y B del procedimiento de la invención mostrando esquemáticamente los cambios cualitativos y cuantitativos en especificidad durante la evolución hacia la especificidad objetivo. La especificidad cuantitativa  $s$ , tal como se define en el contexto de esta invención, se refiere a la proporción entre todos los sustratos aceptados y todos los posibles. La especificidad cualitativa se refiere a la secuencia y composición de aminoácidos de los sustratos aceptados. Las especificidades de las primeras proteasas (círculos vacíos) y la proteasa evolucionada (círculo relleno) se indican esquemáticamente. En la alternativa A, la primera proteasa tiene una especificidad cuantitativa en el intervalo de la especificidad objetivo, pero una especificidad cualitativa que difiere de la especificidad objetivo. Con el fin de generar la especificidad objetivo, la especificidad se cambia sólo cualitativamente. En la alternativa B, la primera proteasa tiene la especificidad cualitativa del sustrato objetivo, pero una especificidad cuantitativa que es muy inferior a la especificidad objetivo. Con el fin de generar la especificidad objetivo, la especificidad se cambia sólo cuantitativamente. Cuando se combinan ambas alternativas, la primera proteasa no tiene ni la especificidad cualitativa ni la cuantitativa del sustrato objetivo. Con el fin de generar la especificidad objetivo, la especificidad se cambia cuantitativa y cualitativamente.

La figura 3 ilustra esquemáticamente cómo proteasas con actividades catalíticas cambiadas se hacen evolucionar usando las dos alternativas A y B del procedimiento de la invención. Según el procedimiento de la invención, la actividad catalítica (abreviada como A) puede usarse como un parámetro de selección. En la alternativa A, la primera proteasa hidroliza sólo el sustrato 1, mientras que otros sustratos incluyendo el sustrato diana (T) no se hidrolizan o sólo se hidrolizan muy lentamente. Mediante el uso del procedimiento de la invención, las proteasas se hacen evolucionar de modo que hidrolizan específicamente el sustrato diana, mientras que otros sustratos incluyendo el primer sustrato no se hidrolizan o sólo se hidrolizan muy lentamente. Estas proteasas evolucionadas se seleccionan mediante un aumento de la actividad catalítica sobre el sustrato diana y una disminución de actividad catalítica sobre el primer sustrato (enfoque de comparación). Como alternativa, la selección puede basarse en la afinidad hacia el sustrato diana (enfoque de afinidad). En la alternativa B, la primera proteasa hidroliza todos los sustratos incluyendo el sustrato diana (T). Mediante el uso del procedimiento de la invención, las proteasas se hacen evolucionar de modo que hidrolizan específicamente el sustrato diana, mientras que otros sustratos incluyendo el primer sustrato no se hidrolizan o sólo se hidrolizan muy lentamente.

Estas proteasas se seleccionan mediante exploración con un exceso de sustratos competidores (enfoque de competidor) o exploración para determinar mayor afinidad de sustrato (enfoque de afinidad). En general, la proteasa evolucionada puede identificarse mediante la comparación de la actividad catalítica hacia los sustratos ofrecidos, incluyendo los primeros sustratos y el sustrato diana.

La figura 4 representa esquemáticamente de dos formas diferentes el enfoque intermedio como un aspecto particular de la invención. Para la descripción de los símbolos, véase la figura 1. El enfoque intermedio usa uno o más sustratos intermedios para guiar la evolución de la especificidad gradualmente hacia la especificidad objetivo en etapas tan pequeñas como sea necesario. Los sustratos intermedios son sustratos que tienen un rasgo intermedio si se comparan con el primer sustrato y el sustrato diana. Los productos intermedios pueden clasificarse de dos formas. En primer lugar, pueden proporcionarse sustratos intermedios sustituyendo al menos uno pero menos de todos los restos de aminoácido del primer sustrato por restos de aminoácido del sustrato diana (producto intermedio con respecto a composición de aminoácido, enfoque 1). En segundo lugar, pueden proporcionarse sustratos intermedios introduciendo selectivamente en posiciones definidas del sustrato restos de aminoácido cuyas propiedades oscilan entre las del resto de aminoácido correspondiente en el primer sustrato y el sustrato diana (producto intermedio con respecto a propiedades de aminoácido, enfoque 2). Como alternativa adicional, pueden combinarse ambos enfoques intermedios. Preferiblemente, tal como se muestra en la figura, se implementa el segundo enfoque en el primer enfoque siempre que la etapa entre dos productos intermedios sea demasiado grande.

La figura 5 ilustra esquemáticamente cómo, según la invención, proteasas con actividades catalíticas cambiadas se hacen evolucionar usando el enfoque intermedio. La primera proteasa tiene una alta actividad sobre un primer sustrato (1) y ninguna o muy baja actividad sobre todos los otros sustratos incluyendo el sustrato diana (T). La siguiente etapa esencial es proporcionar un sustrato intermedio (2) tal como se ilustra en la figura 4. Explorando para determinar la actividad catalítica sobre este sustrato, se seleccionan variantes de proteasa con una actividad aumentada sobre este sustrato intermedio. Puede repetirse esta etapa intermedia con una variación gradual del sustrato intermedio hacia la naturaleza del sustrato diana, hasta que se aísla una proteasa evolucionada que muestra actividad catalítica sólo frente al sustrato diana y ninguna o muy poca actividad sobre el primer sustrato y otros sustratos.

La figura 6 muestra esquemáticamente el vector lanzadera pPDE que puede usarse para el procedimiento de la invención. El vector comprende un origen de *S. cerevisiae* (ori 2  $\mu$ ), un origen de *E. coli* (ori pMB1), un marcador de *S. cerevisiae* (URA3), un marcador de *E. coli* (AmpR), y el casete de expresión que está compuesto por un promotor

## ES 2 289 288 T3

(GAL) de *S. cerevisiae* inducible por galactosa, una secuencia señal para la secreción de la proteína expresada (señal), un sitio KpnI y un sitio de reconocimiento XhoI para insertar el gen de interés, y un terminador (Cyc1).

La figura 7 muestra a modo de ejemplo la hidrólisis de un sustrato peptídico catalizada por la proteasa de virus del grabado del tabaco controlada mediante espectroscopía de fluorescencia confocal de correlación cruzada (cc-FCS). El sustrato peptídico con la secuencia ENLYFQS se reconoce y se hidroliza específicamente mediante la proteasa de VGT. Se incubó el péptido doblemente marcado (Alexa 488, Cy5) 100 nM con (cuadrados rellenos) y sin (círculos vacíos) adición de proteasa 0,01 U/ $\mu$ l en tampón de ensayo que contenía Tris-HCl 50 mM pH 8,0, EDTA 0,5 mM, DTT 10 mM, glicerol al 0,05%.

La figura 8 muestra a modo de ejemplo una distribución de actividades catalíticas obtenida explorando una población de variantes de proteasa sobre el sustrato WLGLVPGG (producto intermedio 1, véase el ejemplo VI) usando espectroscopía de fluorescencia confocal. Se muestra la frecuencia N con la que se identifica determinada actividad catalítica (rendimiento, unidades arbitrarias). Los valores bajos representan bajas actividades catalíticas, mientras que valores altos representan altas actividades catalíticas sobre el sustrato. Se aíslan los genes que codifican variantes que tienen los mayores valores de rendimiento y se evalúan con respecto a su especificidad. Luego se usan estas variantes como primeras proteasas para el ciclo siguiente. Se repite este procedimiento hasta que haya variantes de proteasa identificadas que tengan la especificidad objetivo.

La figura 9 muestra a modo de ejemplo la disminución de  $K_M$  durante la evolución hacia mayor afinidad usando el enfoque de afinidad de la invención. La proteasa usada como primera proteasa (tipo natural) en este experimento fue subtilisina E de *B. subtilis* que tenía una  $K_M$  de 194  $\mu$ M. Esta  $K_M$  disminuyó gradualmente mediante el uso del procedimiento de la invención por un factor de 7,5 hasta 26  $\mu$ M.

La figura 10 muestra a modo de ejemplo el cambio en especificidad durante la evolución de proteasas hacia la especificidad de t-PA. Se evaluó la actividad de las variantes 1, 2, y 3 usando los sustratos intermedio 1, intermedio 2, e intermedio 3 del ejemplo VI. La disminución de la concentración de sustrato corresponde a actividad proteolítica. Cuanto más rápida es la disminución, mayor es la actividad catalítica de la variante de proteasa. Mientras que la primera proteasa tiene actividad muy baja en el producto intermedio 1, y ninguna actividad en los productos intermedios 2 ó 3, las variantes evolucionadas muestran diversas actividades sobre los tres sustratos intermedios.

La figura 11 representa esquemáticamente una variante preferida del enfoque intermedio de la invención (cuarto aspecto, véase más adelante), en el que se seleccionan proteasas en una primera etapa según su especificidad para diferentes sustratos intermedios en paralelo. Luego se seleccionan proteasas según su especificidad para el sustrato diana a partir de una población que contiene variantes recombinadas de las variantes de proteasa seleccionadas en la primera etapa.

La figura 12 muestra a modo de ejemplo curvas de progresión cinética para la primera proteasa en comparación con una proteasa evolucionada obtenida en la vuelta 5 del procedimiento de optimización según la invención. En el caso del primer sustrato la actividad de la proteasa evolucionada es inferior en comparación con la primera proteasa. Esto se invierte en el caso del primer y cuarto producto intermedio, en los que la primera proteasa muestra un recambio del sustrato muy limitado y ninguno, respectivamente.

La invención se explica adicionalmente mediante los siguientes ejemplos. Se entiende que los ejemplos y las realizaciones descritas en la misma son sólo para fines ilustrativos y que los expertos en la técnica sugerirán diversos cambios o modificaciones a la luz de la misma y han de incluirse dentro del espíritu y ámbito de esta solicitud y se consideran dentro del alcance de las reivindicaciones adjuntas. Todas las publicaciones, patentes y solicitudes de patente citadas en el presente documento se incorporan al presente documento por referencia en su totalidad para todos los fines.

### Ejemplos

En los siguientes ejemplos, se proporcionan materiales y procedimientos de la presente invención incluyendo la determinación de propiedades catalíticas de enzimas obtenidas por el procedimiento. Debe entenderse que estos ejemplos tienen solamente fin ilustrativo y no deben interpretarse como que limitan esta invención de ninguna manera.

En los ejemplos experimentales descritos a continuación, se usaron técnicas habituales de tecnología de ADN de recombinación que se describieron en diversas publicaciones, por ejemplo Sambrook *et al.* (1989), Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, o Ausubel *et al.* (1987), Current Protocols in Molecular Biology 1987-1988, Wiley Interscience, Methods in Yeast Genetics (1994) A Cold Spring Harbour Laboratory Manual, que se incorporan por referencia al presente documento en su totalidad. A menos que se indique lo contrario, se usaron enzimas de restricción, polimerasas y otras enzimas así como kits de purificación de ADN según las especificaciones de los fabricantes.

## Ejemplo I

*Clonación molecular de genes que codifican variantes de proteasa*

5 Se clonaron genes que codifican variantes de proteasa en un vector adecuado para la expresión extracelular de proteínas mediante la levadura *Saccharomyces cerevisiae*. El vector usado es un derivado del plásmido pYES2, que está disponible comercialmente de Invitrogen, Inc. Se muestra un mapa del plásmido en la figura 6. El vector contiene un origen 2  $\mu$  para la amplificación en *S. cerevisiae*, un origen pMB1 para la amplificación en *E. coli*, un marcador URA para la selección en *S. cerevisiae*, un marcador de resistencia a la ampicilina para la selección en *E. coli*, así como un promotor GAL y un terminador de transcripción Cyc1 para la expresión inducible en *S. cerevisiae*. Se introdujo un fragmento de 90 pb que contiene la secuencia líder que codifica el péptido señal del gen BAR1 de *S. cerevisiae* detrás del promotor GAL1. Los sitios de restricción KpnI y XhoI sirvieron como sitios de inserción para los genes heterólogos que iban a expresarse. Se realizó la clonación de los genes que codifican variantes de proteasa tal como sigue: se amplificó la secuencia codificante de la proteína madura mediante PCR usando cebadores que introducían un sitio KpnI en el extremo 5' y un sitio XhoI en el extremo 3'. Se clonó el fragmento de PCR en los sitios apropiados del vector y se confirmó la identidad mediante secuenciación.

## Ejemplo II

20 *Proporcionar poblaciones de variantes de proteasa*

Se proporcionó una población de variantes de proteasa mediante modificación al azar de genes que codifican proteasas con especificidades de sustrato conocidas, seguido por la expresión de las variantes de proteasa codificadas por estos genes modificados usando *S. cerevisiae* como un organismo huésped adecuado. En primer lugar, se amplificaron por PCR los genes que codifican variantes de proteasa con especificidades de sustrato conocidas en condiciones propensas a error, esencialmente tal como se describe por Cadwell, R.C y Joyce, G.F. (PCR Methods Appl. 2 (1992) 28-33). Se realizó la PCR propensa a error usando 30 pmol de cada cebador, 20 nmol de dGTP y dATP, 100 nmol de dCTP y dTTP, 20 fmol de plantilla, y 5 U de ADN polimerasa de Taq en Tris HCl 10 mM pH 7,6, KCl 50 mM, MgCl<sub>2</sub> 7 mM, MnCl<sub>2</sub> 0,5 mM, gelatina al 0,01% durante 20 ciclos de 1 min. a 94°C, 1 min. a 65°C y 1 min. a 72°C. Se purificó la biblioteca de ADN resultante usando el kit de purificación de PCR Qiaquick siguiendo las instrucciones del proveedor. Se digirieron los productos de PCR con las enzimas de restricción *XhoI* y *KpnI* y se purificaron tal como se describe en el ejemplo I. Después se unieron los productos de PCR en el vector que se digirió con *XhoI* y *KpnI*, se purificaron en gel y se desfosforilaron. Se transformaron los productos de unión dentro de *E. coli*, se amplificaron en LB que contenía ampicilina como marcador, y se purificaron los plásmidos usando el kit de purificación de plásmidos de Qiagen siguiendo las instrucciones del proveedor. Se transformaron los plásmidos resultantes dentro de células de *S. cerevisiae*. Se proporcionaron poblaciones de variantes de proteasa induciendo la expresión en las células de *S. cerevisiae* transformadas añadiendo galactosa al 2% al medio. Como alternativa, se recombinaron estadísticamente genes que codifican variantes de proteasa con especificidades de sustrato conocidas en posiciones homólogas mediante el uso de la reacción en cadena de recombinación, esencialmente tal como se describe en el documento WO 0134835. Se purificaron productos de PCR de los genes que codifican las variantes de proteasa usando el kit de purificación de PCR QIAquick siguiendo las instrucciones del proveedor, se comprobó el tamaño correcto mediante electroforesis en gel de agarosa y se mezclaron juntos en cantidades equimolares. Se calentaron 80  $\mu$ g de esta mezcla de PCR en TrisHCl 150 mM pH 7,6, MgCl<sub>2</sub> 6,6 mM durante 5 min. a 94°C y posteriormente se enfriaron hasta 37°C a 0,05°C/segundo con el fin de volver a asociar las cadenas y producir así heterodúplex de una manera estocástica. Luego se añadieron 2,5 U de exonucleasa III por  $\mu$ g de ADN y se incubaron durante 20, 40 ó 60 min. a 37°C con el fin de digerir diferentes longitudes desde ambos extremos 3' de los heterodúplex. Volvieron a llenarse los productos de PCR parcialmente digeridos con 0,6 U de polimerasa de Pfu por  $\mu$ g de ADN mediante incubación durante 15 min. a 72°C en dNTPs 0,17 mM y tampón de polimerasa de Pfu según las instrucciones del proveedor. Tras realizar un único ciclo de PCR, se purificó el ADN resultante usando el kit de purificación de PCR QIAquick siguiendo las instrucciones del proveedor, se digirió con *KpnI* y *XhoI* y se unió en el vector linealizado. Se transformaron los productos de unión dentro de *E. coli*, se amplificaron en LB que contenía ampicilina como marcador, y se purificaron los plásmidos usando el kit de purificación de plásmidos de Qiagen siguiendo las instrucciones del proveedor. Se transformaron los plásmidos resultantes dentro de células de *S. cerevisiae*. Se proporcionaron poblaciones de variantes de proteasa induciendo la expresión en las células de *S. cerevisiae* transformadas añadiendo galactosa al 2% al medio.

## Ejemplo III

60 *Proporcionar sustratos peptídicos semejantes al sustrato diana*

Se sintetizaron todos los sustratos peptídicos en un sintetizador de péptidos usando el enfoque de Merrifield *et al.* (Nature. 207 (1965) 522-523). Se diseñaron sustratos peptídicos semejantes al sustrato diana sustituyendo los restos de aminoácido en una o más posiciones del primer sustrato peptídico por los restos de aminoácido en la una o más posiciones del sustrato diana. Como alternativa, se sustituyeron los restos de aminoácido en una o más posiciones del primer sustrato peptídico por restos de aminoácido que tienen un rasgo intermedio con respecto a los restos de aminoácido del primer sustrato peptídico y los restos de aminoácido del sustrato peptídico diana. Véase la tabla III para la determinación del rasgo intermedio de los restos de aminoácido. Se unieron fluoróforos marcadores a los sustratos peptídicos mediante el grupo amino del extremo N terminal o bien mediante el grupo carboxilo del extremo C terminal.

## ES 2 289 288 T3

Como alternativa, se añadió un resto de cisteína al extremo N terminal o bien al extremo C terminal del péptido, y se unió químicamente el fluoróforo marcador al grupo tiol del resto de cisteína.

5 Normalmente se usaron Alexa 488 (Molecular Probes Inc., Oregón, EE.UU.) y Cy-5 (Amersham Biosciences Europe GmbH, Friburgo, Alemania) como marcadores fluoróforos. Se controló la rotura por proteasa del sustrato peptídico mediante FCS de correlación cruzada (Proc. Natl. Acad. Sci. USA. 95 (1998) 1416-1420). Como un ejemplo, en la figura 7 se muestra la rotura de un sustrato peptídico que contiene el sustrato diana para la proteasa del virus del grabado del tabaco (proteasa de VGT) y tiene el fluoróforo Alexa 488 unido al extremo C terminal del péptido y el fluoróforo Cy-5 unido al extremo N terminal del péptido. La proteasa de VGT ya tiene una especificidad relativamente alta ( $s = 4,9$ , véase la tabla I). Se realizó la rotura a una concentración peptídica de 100 nM añadiendo proteasa de VGT 0,01 U/ $\mu$ l en tampón de ensayo que contenía Tris-HCl 50 mM pH 8,0, EDTA 0,5 mM, DTT 10 mM, y glicerol al 0,05%.

### Ejemplo IV

15 *Procedimiento de exploración*

Con el fin de identificar las variantes de enzima que tenían la especificidad de sustrato deseada, se utilizó un enfoque de exploración basado en una configuración de espectroscopía de fluorescencia confocal tal como se describe en el documento WO 9416313. Se utilizó la suspensión celular de un cultivo de *S. cerevisiae* directamente, o bien una alícuota del sobrenadante libre de células, como la muestra que contenía la variante de proteasa secretada. Tras añadir el sustrato a la muestra e incubar durante un periodo de tiempo determinado, se sometieron las muestras a medición mediante espectroscopía de fluorescencia confocal. Si fue necesario, se repitió este procedimiento varias veces con el fin de medir la cinética de la rotura proteolítica. Por consiguiente, se clasificaron las muestras según su actividad proteolítica, y se identificaron las muestras que superaban un umbral de actividad determinado con el fin de aislar el gen que codifica la variante de proteasa correspondiente. En la figura 8 se muestra la distribución de actividades proteolíticas de variantes de proteasa obtenidas mediante este procedimiento.

### Ejemplo V

30 *Generar proteasas específicas de secuencia con afinidad aumentada hacia el sustrato peptídico diana seleccionando a bajas concentraciones de sustrato*

Se generaron variantes de proteasa que tienen una afinidad aumentada hacia el sustrato peptídico diana mediante el procedimiento de la invención basándose en la exploración a bajas concentraciones de sustrato. Por medio de PCR propensa a error (según Cadwell, R.C y Joyce, G.F., PCR Methods Appl. 2 (1992) 28-33), se generó una población de variantes de proteasa que está relacionada con la proteasa alcalina subtilisina E de *Bacillus subtilis*, que tiene una especificidad relativamente baja ( $s = 0,82$ ). Esto está correlacionado con la  $K_M$  relativamente alta que está en el intervalo de 150 - 200  $\mu$ M. Se exploró la población de variantes de proteasa a una complejidad de  $10^6$  variantes mediante espectroscopía de fluorescencia confocal empleando concentraciones de sustrato en el intervalo de 10 nM. Se usaron las variantes aisladas en esta primera exploración como primeras proteasas en un segundo ciclo para proporcionar otra población de variantes de proteasa. De manera análoga se usaron variantes aisladas en ciclos posteriores como primeras proteasas en el ciclo siguiente. Se generó la población de variantes proporcionada en el segundo ciclo y todos los ciclos posteriores mediante una combinación de PCR propensa a error (véase anteriormente) y recombinación homóloga *in vitro* (según el documento WO 0134835). Se analizaron cinéticamente las variantes aisladas de los cuatro primeros ciclos de este procedimiento. El incremento en la afinidad hacia el sustrato durante las cuatro vueltas corresponde a la disminución de la  $K_M$  de los que tenían mejor rendimiento de cada ciclo, lo que se muestra en la figura 9.

### Ejemplo VI

*Generar proteasas específicas de secuencia con una especificidad de sustrato objetivo semejante a la especificidad de activador del plasminógeno de tipo tisular*

55 Se generaron proteasas mediante el procedimiento de la invención que tenían una especificidad que se alteró hacia la especificidad de activador del plasminógeno de tipo tisular (t-PA). Se usó la proteasa BAR1 de *Saccharomyces cerevisiae* (SEC ID NO: 8) como primera proteasa. Esta proteasa pertenece al grupo de las aspártico proteinasas (Mac Kay *et al.*; Structure and Function of the Aspartic Proteinases (1991) 161-172). Es específica para sustratos peptídicos que contienen la secuencia de aminoácidos WLQLKPGQ, y cataliza la rotura en el enlace peptídico entre la segunda leucina y el resto de lisina. Se generaron poblaciones de variantes de proteasa que estaban relacionadas con la proteasa BAR 1 o proteasas aisladas en ciclos de exploración posteriores por medio de PCR propensa a error (según Cadwell, R.C y Joyce, G.F., PCR Methods Appl. 2 (1992) 28-33) y recombinación homóloga *in vitro* usando la reacción en cadena de recombinación (documento WO 0134835). Se exploraron las variantes de proteasa para determinar su actividad proteolítica a complejidades de  $10^6$  variantes mediante espectroscopía de fluorescencia confocal. La proteasa BAR 1 usada como primera proteasa ya tenía una especificidad relativamente alta que estaba en el intervalo de la especificidad objetivo. Por tanto, se usó una combinación del enfoque de afinidad y del enfoque intermedio. La exploración a bajas concentraciones mantuvo alta la especificidad de la proteasa, mientras que la exploración en sustratos intermedios permitió la evolución hacia la nueva especificidad. Se construyeron cuatro sustratos intermedios. El sustrato

intermedio 1 tenía la secuencia de aminoácidos WLGLVPPG, el sustrato intermedio 2 la secuencia de aminoácidos WLQVPPG, el sustrato intermedio 3 la secuencia de aminoácidos WLGRVPPG, y el producto intermedio cuatro tenía la secuencia WLGRVVGG. La especificidad de sustrato objetivo de t-PA se dirige hacia CPGRVVGG con rotura entre el resto de arginina y el primer resto de valina. En la tabla IV se muestran todos los sustratos.

TABLA IV

|                       | P4 | P3 | P2 | P1 | P1' | P2' | P3' | P4' |
|-----------------------|----|----|----|----|-----|-----|-----|-----|
| Primer sustrato       | W  | L  | Q  | L  | K   | P   | G   | Q   |
| Producto intermedio 1 | W  | L  | G  | L  | V   | P   | G   | G   |
| Producto intermedio 2 | W  | L  | G  | █  | V   | P   | G   | G   |
| Producto intermedio 3 | W  | L  | G  | R  | V   | P   | G   | G   |
| Producto intermedio 4 | W  | L  | G  | R  | V   | V   | G   | G   |
| Sustrato diana        | C  | P  | G  | R  | V   | V   | G   | G   |

El producto intermedio 1 era un producto intermedio de composición de aminoácido debido al hecho de que contenía en las posiciones P4, P3, P1 y P2' los mismos restos de aminoácido que el primer sustrato, y en las posiciones P2, P1', y P4' los mismos restos que el sustrato diana. El producto intermedio 2 era un producto intermedio de propiedad de aminoácido con respecto al producto intermedio 1 y al sustrato diana. Era semejante al producto intermedio 1 pero contenía en la posición P1 un resto de glutamina que tiene un rasgo intermedio comparado con el resto de leucina presente en esa posición en el primer sustrato y el resto de arginina presente en esa posición en el sustrato diana. El producto intermedio 3, como otro producto intermedio de composición de aminoácido, se basaba en restos de aminoácido que provienen tanto del primer sustrato como del sustrato diana, al igual que el sustrato intermedio 1, pero, a diferencia de éste último, compartía una posición adicional con el sustrato diana. En comparación con el producto intermedio 3, el producto intermedio 4 comparte un aminoácido más con la secuencia diana en la posición P2'. En la figura 10 se muestran las especificidades cambiadas de diferentes variantes que se generaron mediante este procedimiento.

El aumento de especificidad de sustrato puede medirse también como la conversión dependiente del tiempo de los sustratos, tal como se demuestra a modo de ejemplo en la figura 12. La conversión de sustrato se presenta como la fracción de sustrato no convertido con el tiempo. Tal como en la figura 12, la primera proteasa y una variante evolucionada de la vuelta 5 se diferencian en su actividad proteolítica sobre el primer sustrato, producto intermedio 1 y producto intermedio 4, respectivamente. En el caso del primer sustrato, la actividad de la proteasa evolucionada es inferior en comparación con la primera proteasa. Esto se invierte en el caso del primer y cuarto producto intermedio, en los que la primera proteasa muestra un recambio del sustrato muy limitado y ninguno, respectivamente, mientras que la proteasa evolucionada muestra actividad considerable sobre ambos sustratos.

De este modo se generan proteasas según el procedimiento de la invención, que tienen una especificidad de sustrato similar al activador del plasminógeno de tipo tisular humano. La proteasas generadas tienen al menos una mutación en una posición del grupo: 33, 45, 47, 59, 82, 96, 107, 123, 143, 151, 152, 161, 163, 165, 178, 221, 231, 321, 367, 369, 370, 399, 404, 440 (basándose en la numeración de la secuencia de aminoácidos de la proteasa BAR1 enumerada como SEC ID NO:8). Preferiblemente, una variante de proteasa evolucionada a partir de la proteasa BAR1 de tipo natural hacia especificidad del activador del plasminógeno de tipo tisular humano tiene al menos una mutación del grupo: D367G, M369L, V370I, M107I, I152F, E143D, E231V, L33I, Y45D, T47A, T59I, N82D, E96V, N123D, N151V, K161E, A163T, T165A, R178S, T221I, D321N, A399S, K404R y S440L.

La figura 12 presenta el comportamiento catalítico de una proteasa evolucionada según el procedimiento de la invención en comparación con la proteasa de partida (primera). Comenzando con proteasa BAR1 (SEC ID NO: 8) se obtienen variantes con diferentes mutaciones. La figura 12 muestra representaciones gráficas que reflejan el aumento de especificidad de sustrato de una variante de la vuelta 5. Las investigaciones realizadas sobre la secuencia de aminoácidos de la variante a modo de ejemplo de la vuelta 5 revelaron una combinación particular de sustituciones de aminoácidos (con numeración equivalente a la numeración de la proteasa Bar1) como Y45D, T47A, N82D, M107I, E143D, I152F, T165A, E231V, D367G, V370I.

REIVINDICACIONES

5 1. Un procedimiento para identificar proteasas específicas de secuencia con especificidades de sustrato objetivo que comprende las etapas siguientes

(a) proporcionar una población de proteasas comprendida por variantes de una primera proteasa o de variantes o quimeras de dos o más primeras proteasas, teniendo dichas primeras proteasas una especificidad de sustrato para una secuencia de aminoácidos particular de un primer sustrato peptídico;

10 (b) poner en contacto dicha población de proteasas con uno o más segundos sustratos, que comprenden al menos una secuencia de aminoácidos específica que es idéntica al sustrato peptídico diana o que tiene un rasgo intermedio con respecto al primer sustrato y el sustrato diana pero sin estar presente dentro del primer sustrato peptídico; y

15 (c) seleccionar una o más variantes de proteasa de la población de proteasas proporcionada en la etapa (a) que tienen especificidad para dicha secuencia de aminoácidos específica de los segundos sustratos proporcionados en la etapa (b) en condiciones que permiten la identificación de proteasas que reconocen e hidrolizan preferiblemente dicha una secuencia de aminoácidos específica dentro de los segundos sustratos, en el que la exploración para determinar la actividad proteasa se consigue añadiendo en exceso péptidos distintos del segundo péptido, usando de ese modo los péptidos añadidos como competidores.

20 2. El procedimiento según la reivindicación 1, en el que las condiciones de selección en la etapa (c) comprenden además

25 (i) explorar para determinar la actividad proteasa en concentraciones de sustrato bajas, aumentando de ese modo la afinidad para el segundo sustrato, y/o

(ii) explorar para determinar la actividad proteasa usando dos o más sustratos en comparación, aumentando de ese modo la selectividad de la enzima.

30 3. El procedimiento según la reivindicación 1 ó 2, en el que las etapas (a) a (c) se repiten de manera cíclica hasta que se identifican una o más variantes de proteasa con especificidad para el segundo sustrato, y en el que variantes de proteasa seleccionadas en un ciclo se usan como primeras proteasas en el ciclo siguiente, y en el que se realizan al menos un ciclo y menos de 100 ciclos.

35 4. El procedimiento según una cualquiera de las reivindicaciones 1 a 3, en el que se usa sólo un segundo sustrato en el uno o más ciclos, y en el que el segundo sustrato es idéntico al sustrato diana.

40 5. El procedimiento según una cualquiera de las reivindicaciones 1 a 3, en el que se usan diferentes segundos sustratos, y en el que los segundos sustratos tienen un rasgo intermedio con respecto al primer sustrato y al sustrato diana, y en el que el segundo sustrato que se usa en el último ciclo es idéntico al sustrato diana.

6. El procedimiento según la reivindicación 5, en el que se usan diferentes segundos sustratos en ciclos consecutivos, y en el que cada segundo sustrato tiene un rasgo intermedio con respecto al segundo sustrato usado anteriormente y el sustrato diana.

45 7. El procedimiento según la reivindicación 5 ó 6, ejecutándose en al menos un ciclo las etapas (b) a (c) con diferentes segundos sustratos en paralelo, y en el que se combinan las variantes de proteasa aisladas de dicha manera paralela y se usan como primeras proteasas en el ciclo siguiente.

50 8. El procedimiento según cualquiera de las reivindicaciones 5 a 7, en el que el rasgo intermedio de los sustratos intermedios se basa en

(i) la composición de aminoácidos,

55 (ii) la secuencia de aminoácidos,

(iii) las propiedades físicas y/o las químicas de los restos de aminoácido dentro de la secuencia de aminoácidos específica, usándose preferiblemente una o más propiedades del grupo constituido por las propiedades de aminoácido siguientes: la superficie, el volumen, el punto isoeléctrico, el pka de cadena lateral, la carga, la polaridad, la hidrofobicidad, o

60 (iv) cualquier combinación de los mismos.

9. El procedimiento según una cualquiera de las reivindicaciones 1 a 5, en el que los segundos sustratos difieren de los primeros sustratos porque de 1 a 5 restos de aminoácido dentro de la secuencia de aminoácidos específica están intercambiados.

65 10. El procedimiento según una cualquiera de las reivindicaciones 1 a 9, en el que los segundos sustratos llevan grupos funcionales que posibilitan la detección de la hidrólisis del sustrato, siendo dichos grupos funcionales

## ES 2 289 288 T3

- (i) uno o más fluoróforos o cromóforos, cuyas propiedades espectroscópicas cambian tras la hidrólisis del péptido, realizándose la exploración mediante la determinación del cambio en las propiedades espectroscópicas; o
- 5 (ii) dos fluoróforos que pueden distinguirse por sus propiedades de fluorescencia y que están unidos a extremos opuestos del segundo sustrato, realizándose la exploración mediante espectroscopía de fluorescencia confocal a concentraciones de fluoróforo inferiores a 1  $\mu\text{M}$ ; o
- 10 (iii) dos fluoróforos que forman un par de transferencia de energía por resonancia de fluorescencia (FRET) y que están unidos a extremos opuestos del segundo sustrato, realizándose la exploración mediante la determinación de la disminución en la transferencia de energía entre los dos fluoróforos; o
- 15 (iv) una primera y segunda proteína autofluorescente que flanquean el segundo sustrato, realizándose la exploración mediante espectroscopía de fluorescencia confocal a concentraciones de sustrato inferiores a 1  $\mu\text{M}$ ; o
- 20 (v) un fluoróforo y una molécula inhibidora que están unidos a extremos opuestos del segundo sustrato, realizándose la exploración mediante la determinación de la disminución en la inhibición del fluoróforo; o
- 25 (vi) un fluoróforo o un cromóforo y un resto de unión que están unidos a extremos opuestos del segundo sustrato, realizándose la exploración mediante la determinación de la unión del resto de unión a una pareja de unión específica; o
- (vii) un marcador radiactivo y un resto de unión que están unidos a extremos opuestos del segundo sustrato, realizándose la exploración mediante el uso de un ensayo de proximidad de centelleo; o
- (viii) cualquier combinación de los mismos.

11. El procedimiento según una cualquiera de las reivindicaciones 1 a 10, en el que

- 30 (i) la población de proteasas se obtiene mediante mutagénesis al azar de ácidos nucleicos, mutagénesis con casete, mutagénesis de saturación de sitio, mutagénesis por delección y/o inserción al azar o específica de sitio, recombinación homóloga *in vitro*, recombinación homóloga *in vivo*, recombinación no homóloga o una combinación de las mismas; y/o
- 35 (ii) la expresión de la población de proteasas se realiza mediante el uso de células huésped, preferiblemente de origen bacteriano, de levadura, de insecto, viral o de mamífero, o se realiza mediante el uso de sistemas de expresión de proteínas libres de células, y/o
- 40 (iii) el acoplamiento del genotipo y fenotipo de las proteasas se consigue mediante el uso de portamuestras que posibilitan la compartimentación de muestras, y la distribución de genotipos en portamuestras se realiza a una multiplicidad por compartimento que permite una diferenciación suficiente de fenotipos.

12. El procedimiento según cualquiera de las reivindicaciones 1 a 11, en el que se selecciona la primera proteasa del grupo de proteasas constituido por serina proteasas, cisteína proteasas, aspártico proteasas y metaloproteasas, y en el que la primera proteasa se selecciona preferiblemente del grupo de proteasas constituido por papaína, bromelaína, tripsina, pepsina, quimotripsina, subtilisina, SET, elastasa humana, catepsina, quimasa, PEP I de *Saccharomycopsis fibuligera*, caliceína, uroquinasa, termolisina, colagenasa, elastasa de *Pseudomonas aeruginosa*, proteasa de VGT, proteasa de VIH-1, proteasa BAR1, factor Xa, trombina, activador del plasminógeno de tipo tisular, proteasa Kex2, proteasa de VMVT, proteasa de VRS, proteasa de VLM, proteasa de VMPPM, proteasa de VTMM, proteasa de VLB, proteasa de VAIE, proteasa de VISmac.

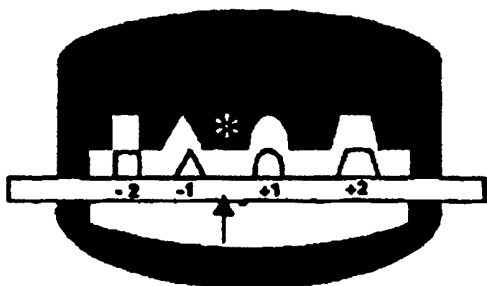
13. El procedimiento según una cualquiera de las reivindicaciones 1 a 12, preferiblemente según una cualquiera de las reivindicaciones 5 a 12, en el que la proteasa objetivo tiene una especificidad similar al activador del plasminógeno de tipo tisular y rompe el sustrato diana CPGR<sup>1</sup>VVGG.

55 14. El procedimiento según la reivindicación 13, en el que la proteasa de partida es proteasa BAR1 de *S. cerevisiae* y preferiblemente se utilizan los sustratos segundos/intermedios siguientes:

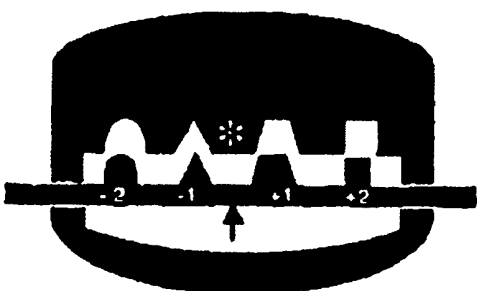
- (i) WLGLVPGG
- 60 (ii) WLGQVPGG
- (iii) WLGRVPGG
- (iv) WLGRVVGG
- 65 (v) CPGRVVGG.

**Alternativa A**

**primera proteasa**

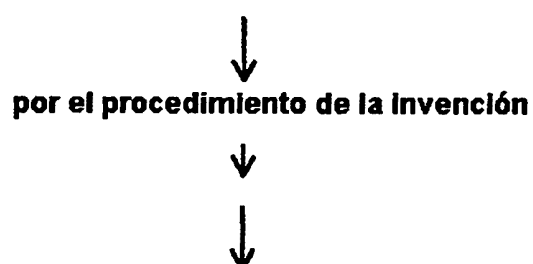
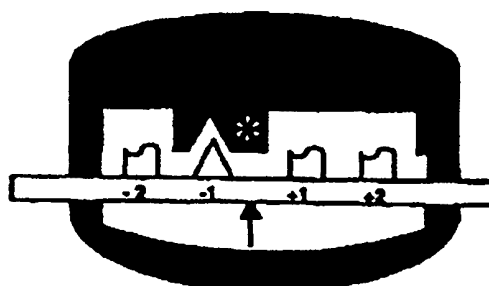


**proteasa evolucionada**

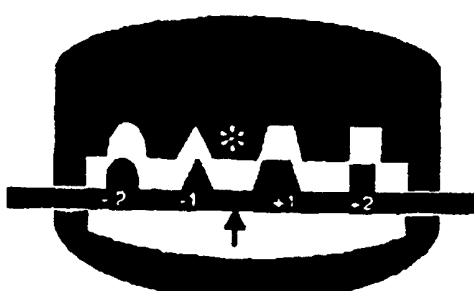


**Alternativa B**

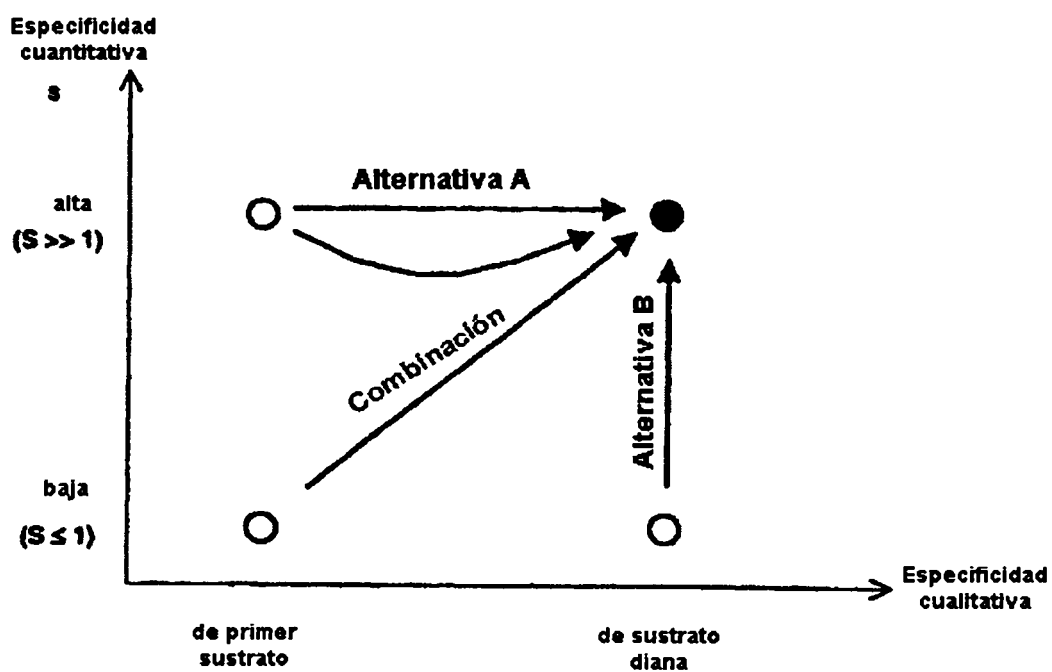
**primera proteasa**



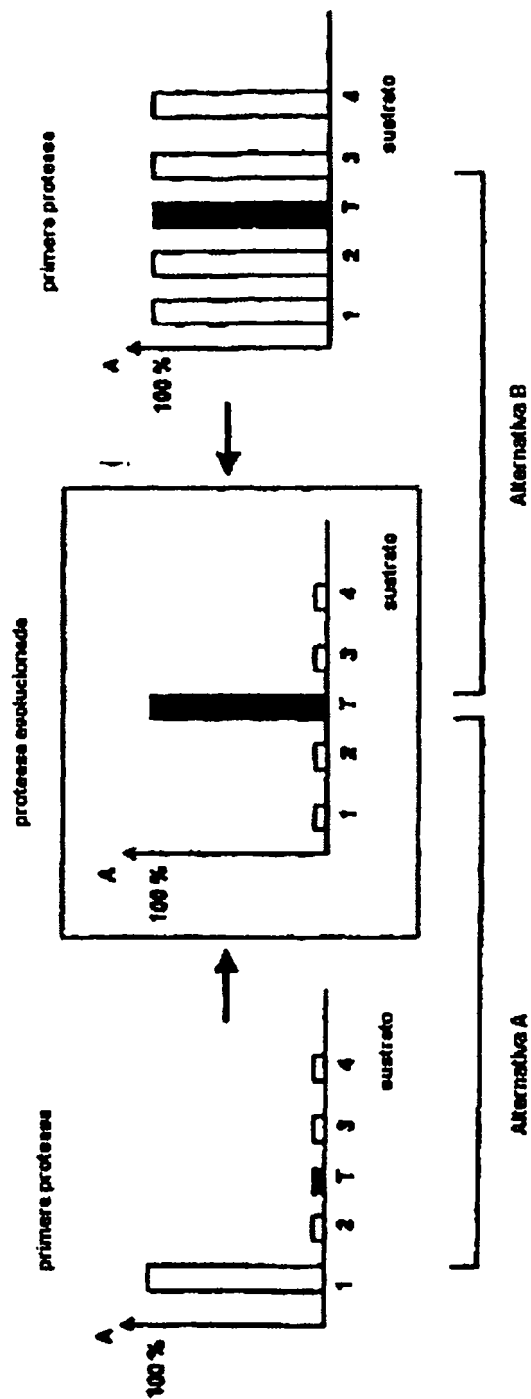
**proteasa evolucionada**



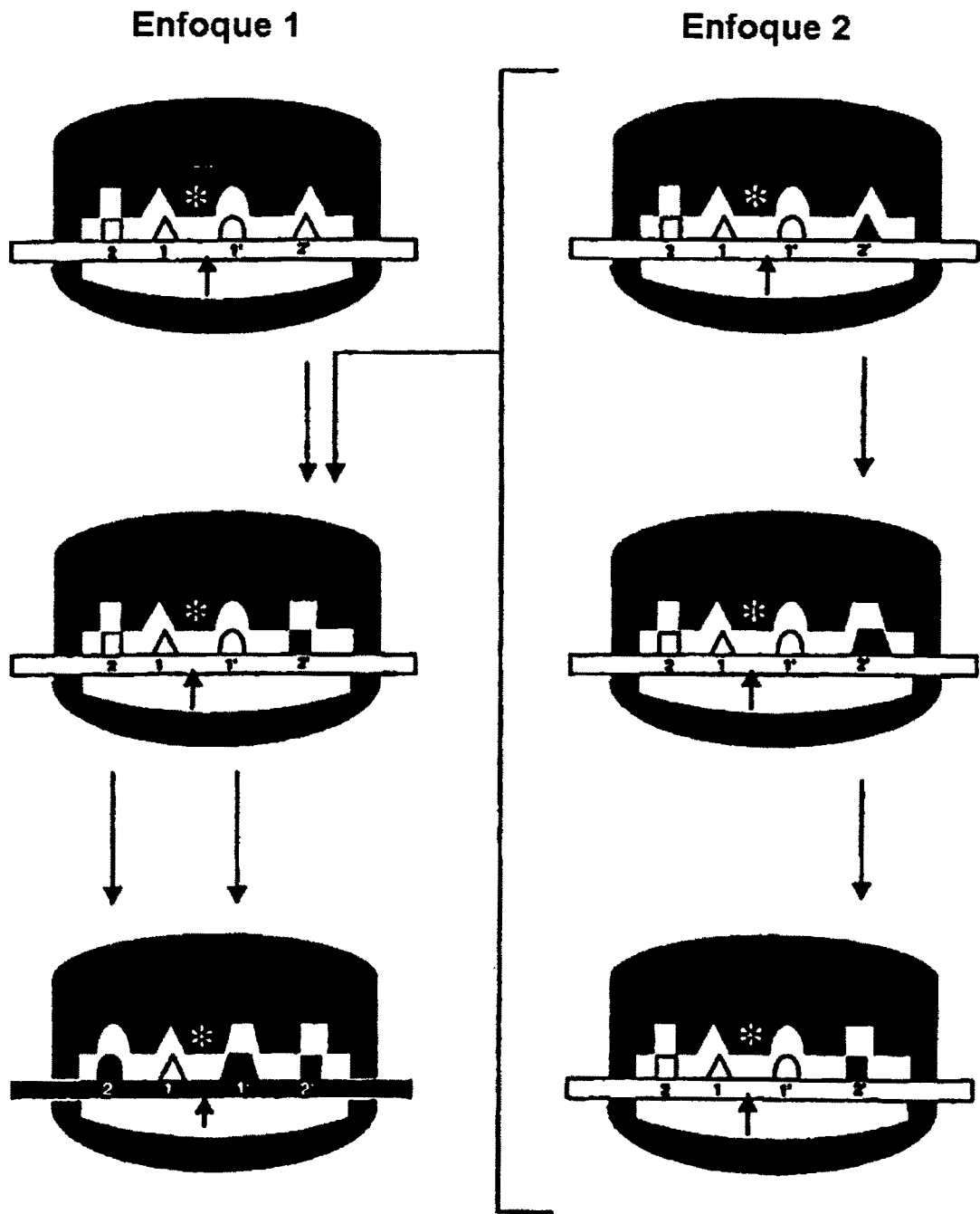
**Fig.1**



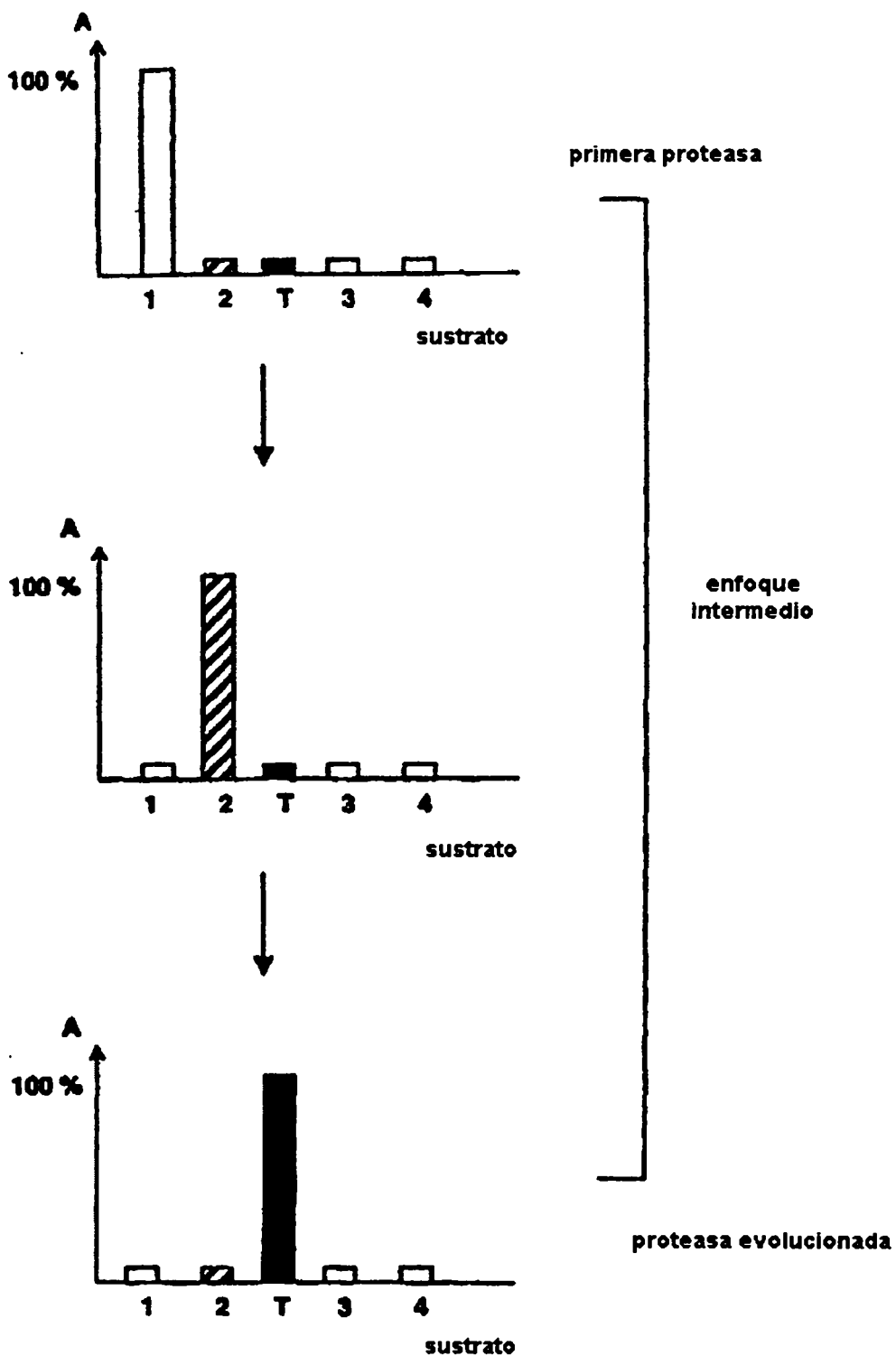
**Fig.2**



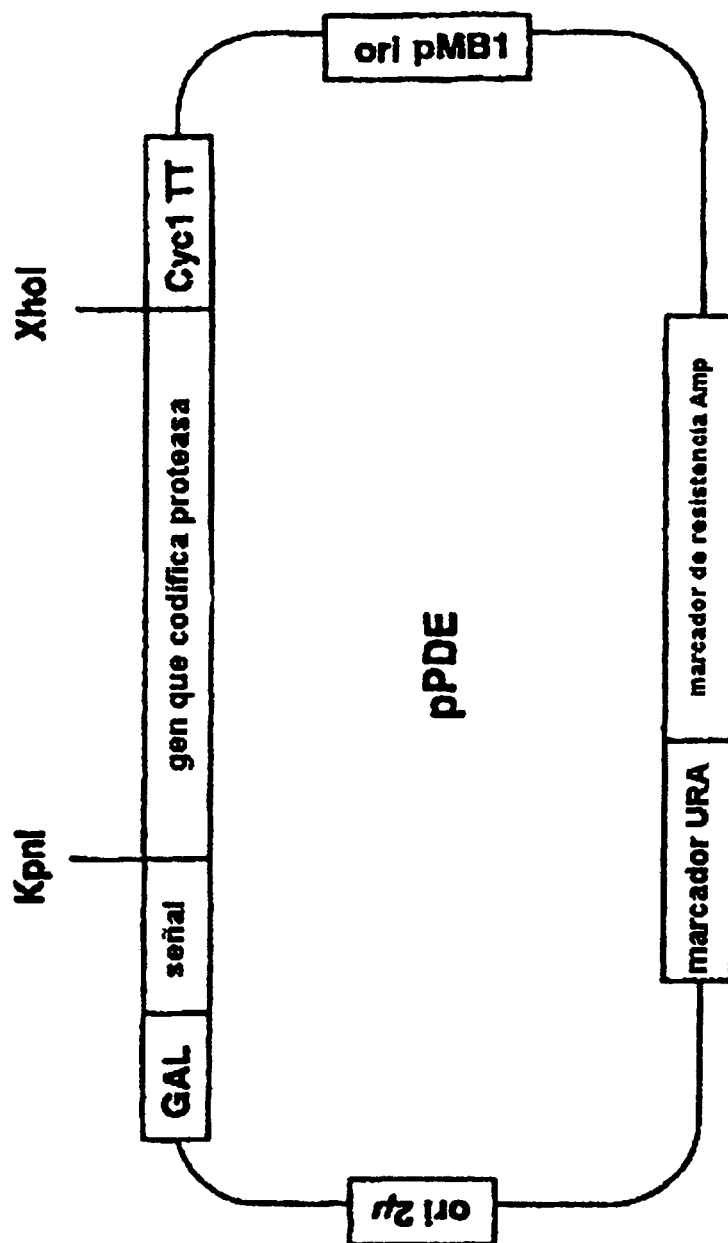
**Fig.3**



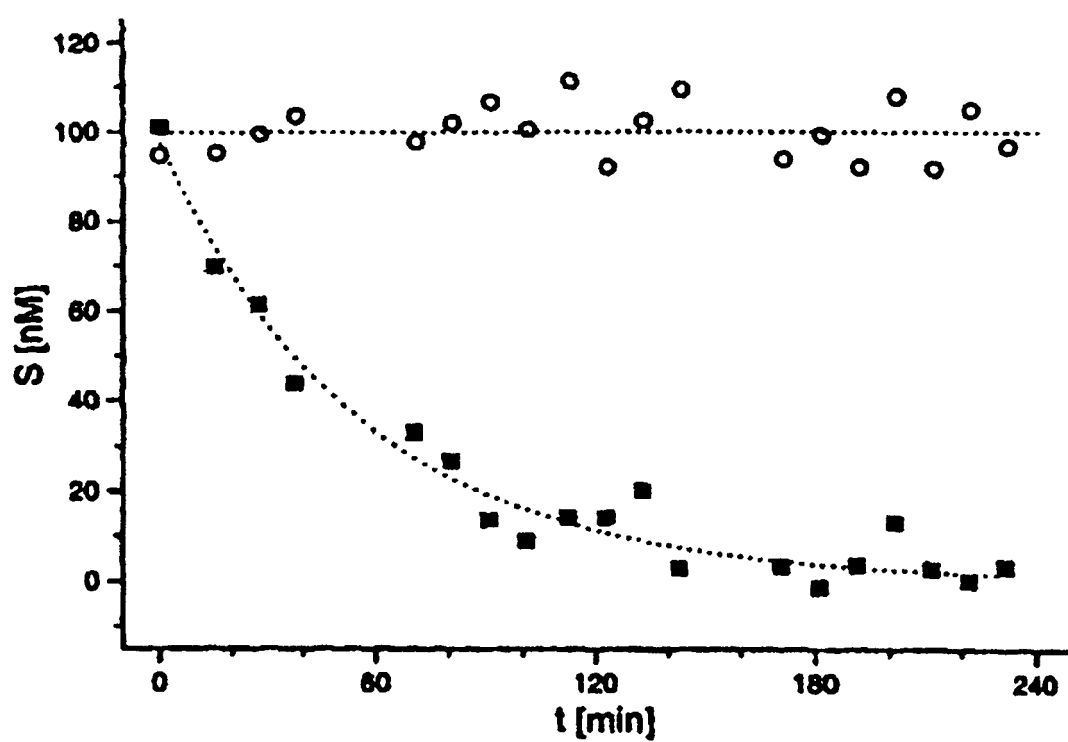
**Fig.4**



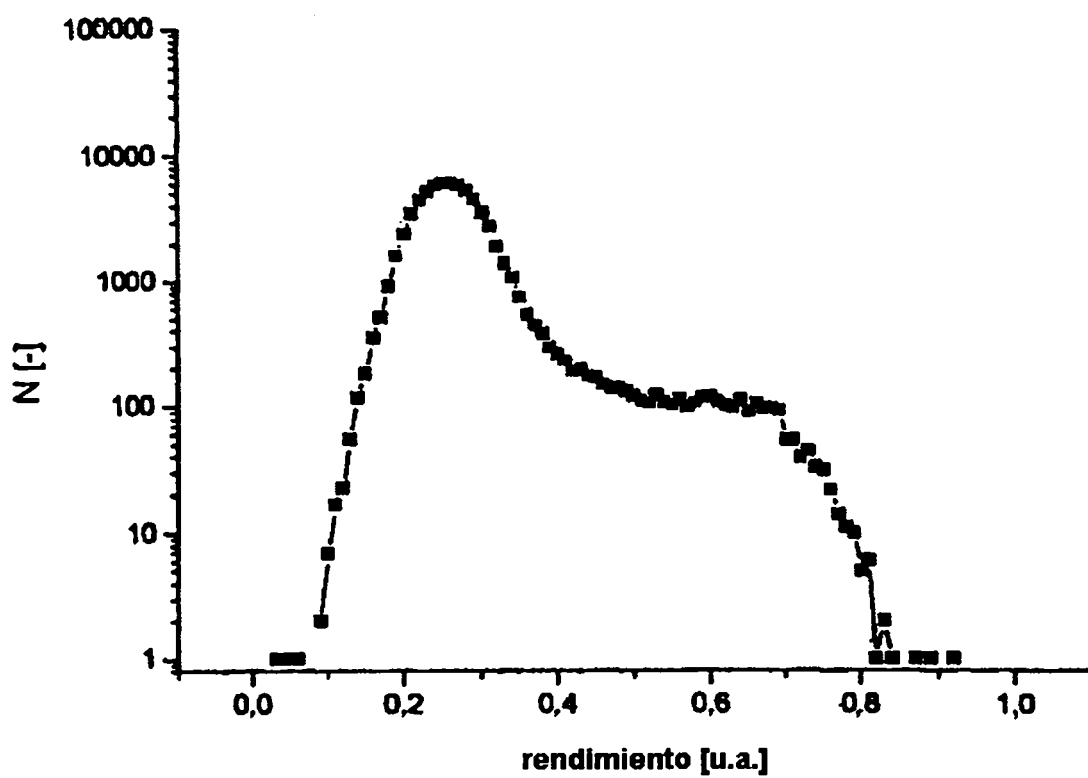
**Fig.5**



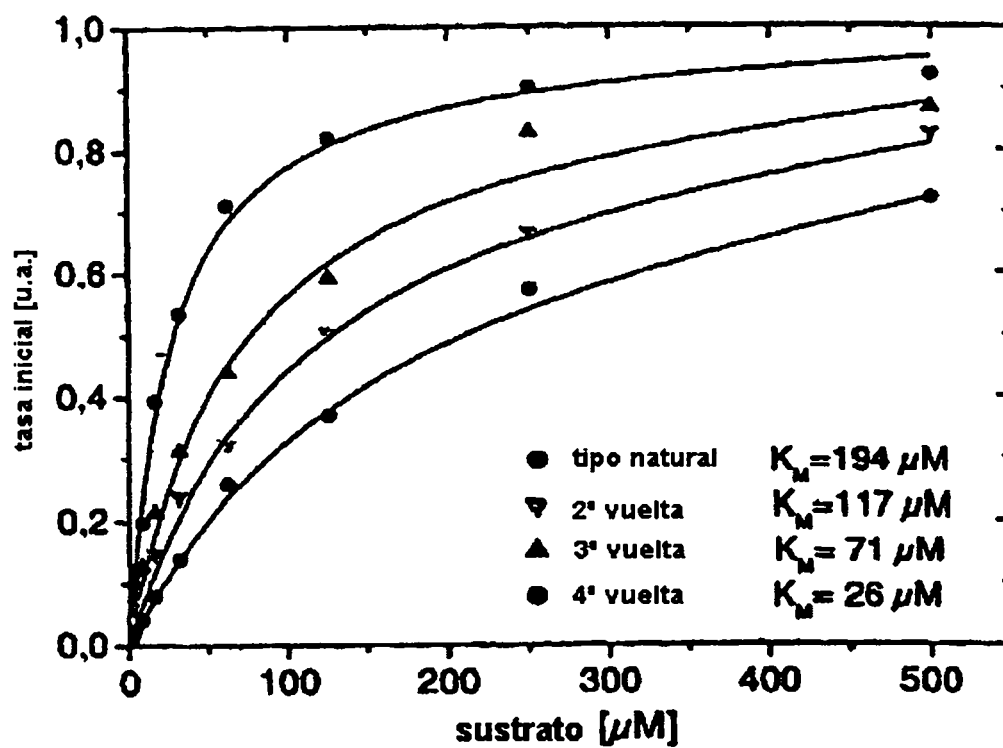
**Fig.6**



**Fig.7**



**Fig.8**



**Fig.9**

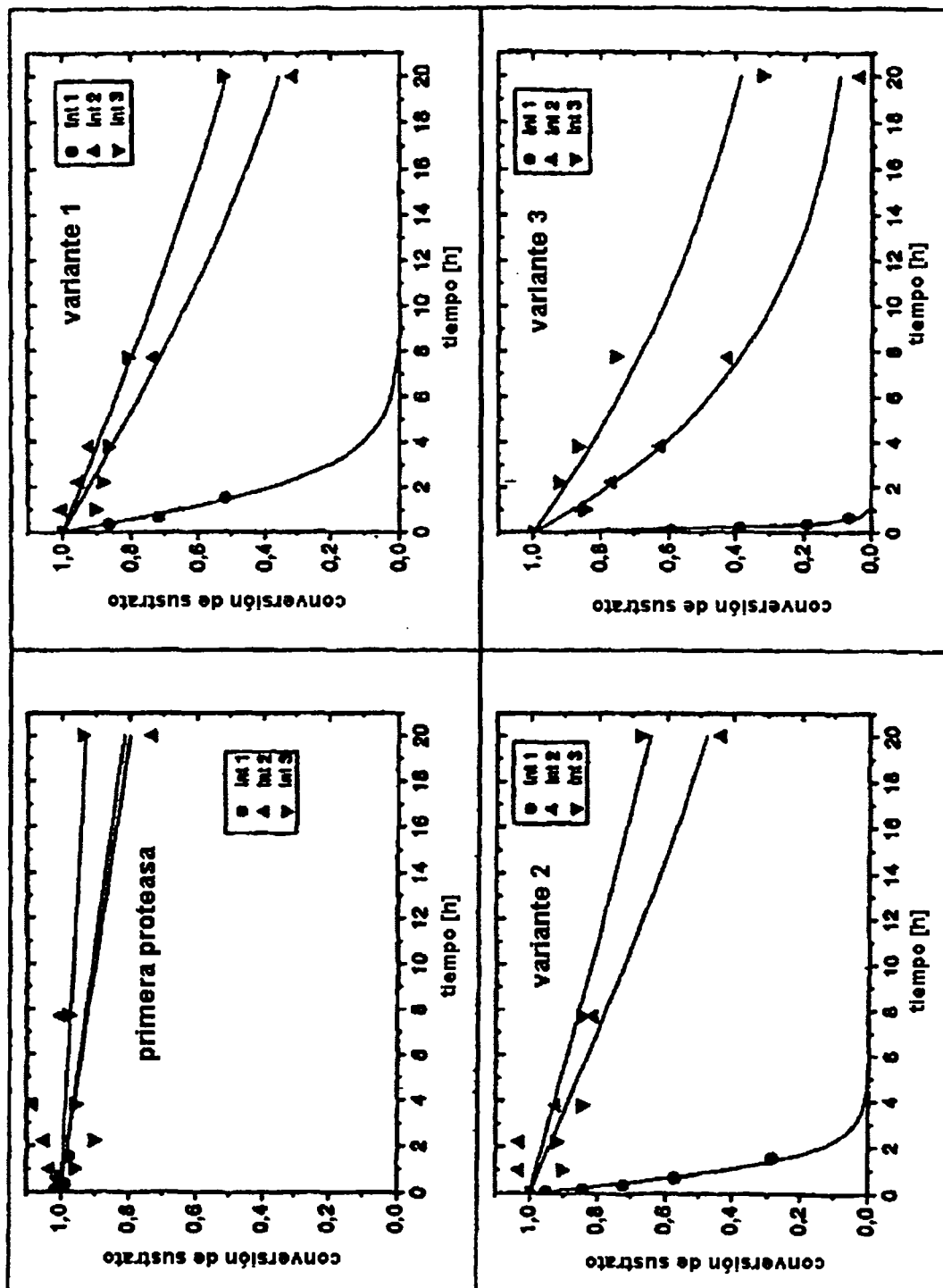
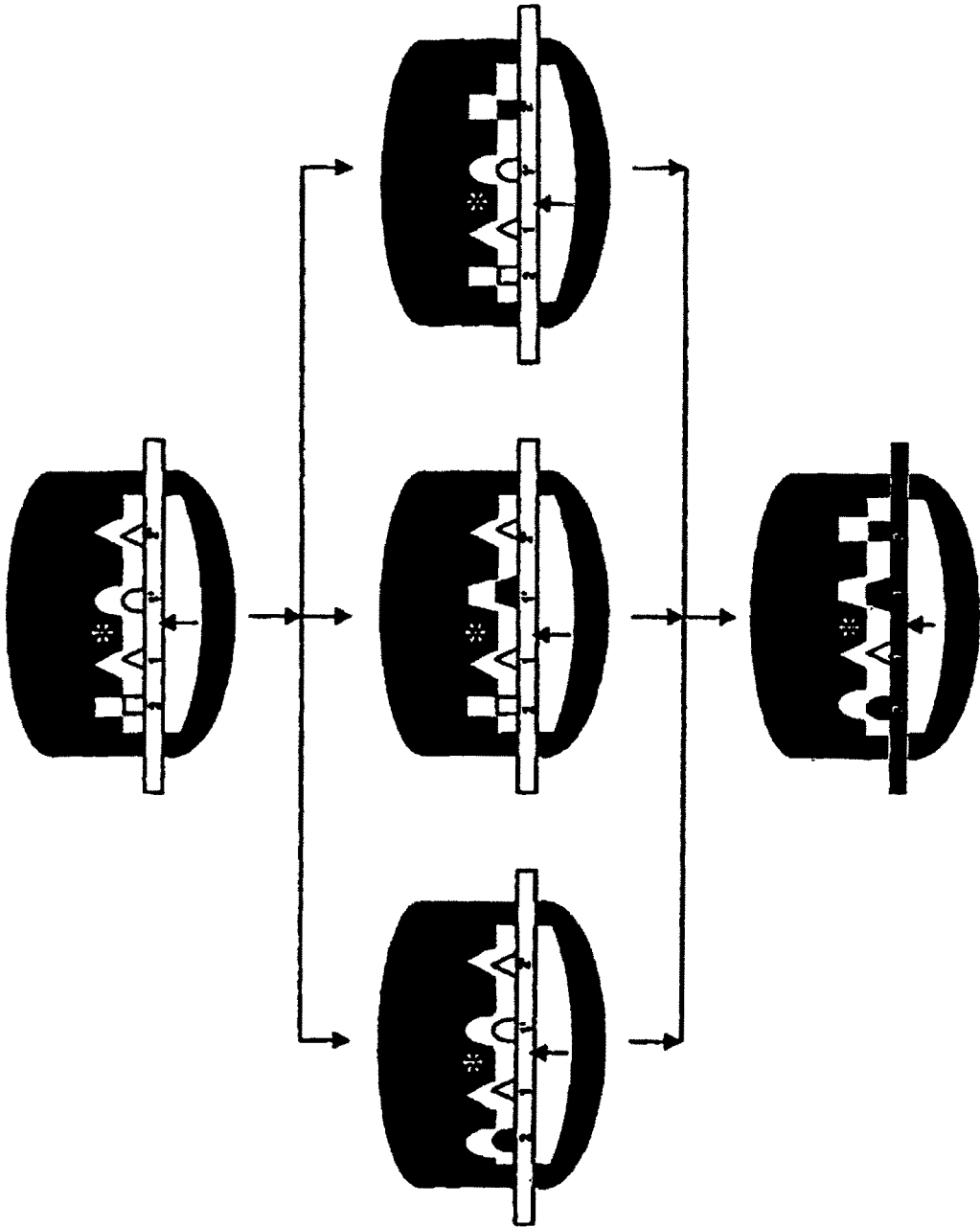


Fig.10



**Fig.11**

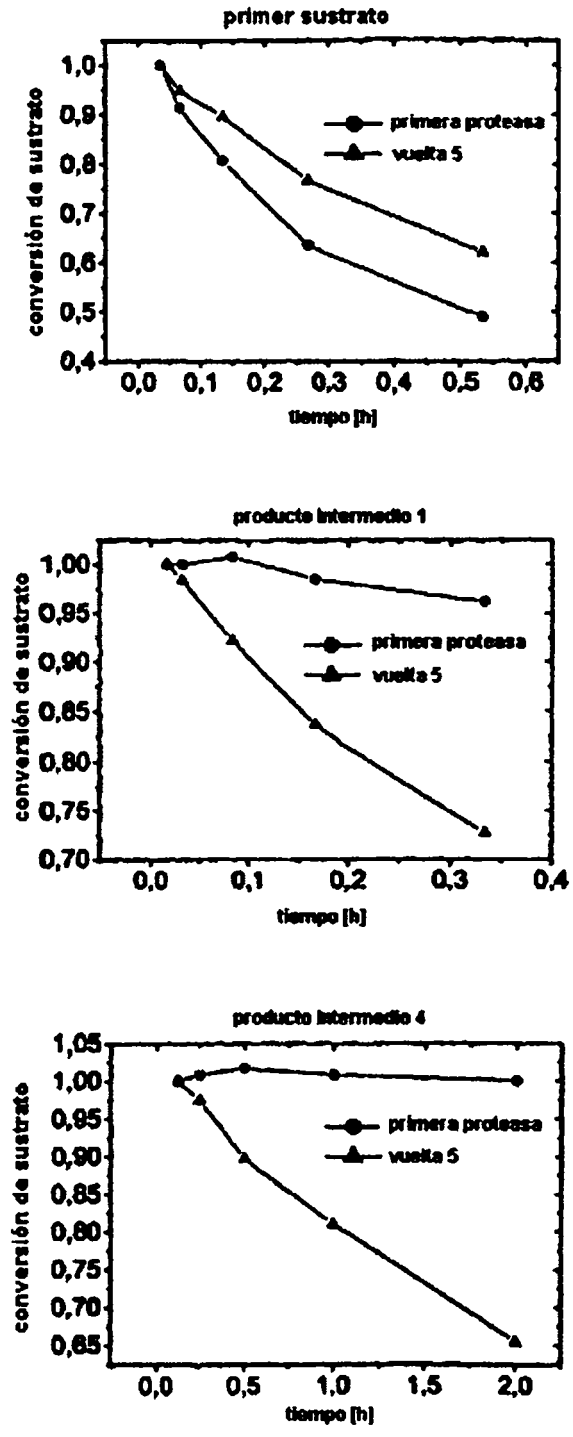


Fig.12

# ES 2 289 288 T3

## LISTA DE SECUENCIAS

- <110> DIREVO Biotech AG
- 5 <120> Procedimiento para generar proteasas específicas de secuencia mediante evolución dirigida y uso del mismo  
<130> 031139wo/JH/ml  
<140>  
<141>
- 10 <160> 8  
<170> PatentIn Ver. 2.1  
<210> 1
- 15 <211> 8  
<212> PRT  
<213> Secuencia artificial  
<220>
- 20 <223> Descripción de secuencia artificial: sustrato de proteasa  
  
<400> 1
- 25        Trp Leu Gln Leu Lys Pro Gly Gln  
          1                            5
- <210> 2
- 30 <211> 8  
<212> PRT  
<213> Secuencia artificial  
<220>
- 35 <223> Descripción de secuencia artificial: sustrato de proteasa  
  
<400> 2
- 40        Trp Leu Gly Leu Val Pro Gly Gly  
          1                            5
- <210> 3
- 45 <211> 8  
<212> PRT  
<213> Secuencia artificial  
<220>
- 50 <223> Descripción de secuencia artificial: sustrato de proteasa  
  
<400> 3
- 55        Trp Leu Gly Gln Val Pro Gly Gly  
          1                            5
- <210> 4
- 60 <211> 8  
<212> PRT  
<213> Secuencia artificial  
<220>
- 65 <223> Descripción de secuencia artificial: sustrato de proteasa

ES 2 289 288 T3

<400> 4

Trp Leu Gly Arg Val Pro Gly Gly  
1 5

5

<210> 5

<211> 8

<212> PRT

10

<213> Secuencia artificial

<220>

<223> Descripción de secuencia artificial: sustrato de proteasa

15

<400> 5

Trp Leu Gly Arg Val Val Gly Gly  
1 5

20

<210> 6

<211> 8

<212> PRT

25

<213> Secuencia artificial

<220>

<223> Descripción de secuencia artificial: sustrato de proteasa

30

<400> 6

Cys Pro Gly Arg Val Val Gly Gly  
1 5

35

<210> 7

<211> 7

<212> PRT

40

<213> Secuencia artificial

<220>

<223> Descripción de secuencia artificial: sustrato de proteasa de VGT

45

<400> 7

Glu Asn Leu Tyr Phe Gln Ser  
1 5

50

<210> 8

<211> 587

<212> PRT

55

<213> *Saccharomyces cerevisiae*

<400> 8

60

**Met Ser Ala Ile Asn His Leu Cys Leu Lys Leu Ile Leu Ala Ser Phe**  
**1 5 10 15**

65

**Val Ile Ile Asn Thr Ile Thr Ala Leu Thr Asn Asp Gly Thr Gly His**  
**20 25 30**

ES 2 289 288 T3

Leu Glu Phe Leu Leu Gln His Glu Glu Glu Met Tyr Tyr Ala Thr Thr  
 35 40 45  
 5 Leu Asp Ile Gly Thr Pro Ser Gln Ser Leu Thr Val Leu Phe Asp Thr  
 50 55 60  
 Gly Ser Ala Asp Phe Trp Val Met Asp Ser Ser Asn Pro Phe Cys Leu  
 65 70 75 80  
 10 Pro Asn Ser Asn Thr Ser Ser Tyr Ser Asn Ala Thr Tyr Asn Gly Glu  
 85 90 95  
 Glu Val Lys Pro Ser Ile Asp Cys Arg Ser Met Ser Thr Tyr Asn Glu  
 100 105 110  
 15 His Arg Ser Ser Thr Tyr Gln Tyr Leu Glu Asn Gly Arg Phe Tyr Ile  
 115 120 125  
 20 Thr Tyr Ala Asp Gly Thr Phe Ala Asp Gly Ser Trp Gly Thr Glu Thr  
 130 135 140  
 Val Ser Ile Asn Gly Ile Asp Ile Pro Asn Ile Gln Phe Gly Val Ala  
 145 150 155 160  
 25 Lys Tyr Ala Thr Thr Pro Val Ser Gly Val Leu Gly Ile Gly Phe Pro  
 165 170 175  
 Arg Arg Glu Ser Val Lys Gly Tyr Glu Gly Ala Pro Asn Glu Tyr Tyr  
 180 185 190  
 30 Pro Asn Phe Pro Gln Ile Leu Lys Ser Glu Lys Ile Ile Asp Val Val  
 195 200 205  
 Ala Tyr Ser Leu Phe Leu Asn Ser Pro Asp Ser Gly Thr Gly Ser Ile  
 210 215 220  
 35 Val Phe Gly Ala Ile Asp Glu Ser Lys Phe Ser Gly Asp Leu Phe Thr  
 225 230 235 240  
 40 Phe Pro Met Val Asn Glu Tyr Pro Thr Ile Val Asp Ala Pro Ala Thr  
 245 250 255  
 Leu Ala Met Thr Ile Gln Gly Leu Gly Ala Gln Asn Lys Ser Ser Cys  
 260 265 270  
 45 Glu His Glu Thr Phe Thr Thr Thr Lys Tyr Pro Val Leu Leu Asp Ser  
 275 280 285  
 Gly Thr Ser Leu Leu Asn Ala Pro Lys Val Ile Ala Asp Lys Met Ala  
 290 295 300  
 50 Ser Phe Val Asn Ala Ser Tyr Ser Glu Glu Glu Gly Ile Tyr Ile Leu  
 305 310 315 320  
 55 Asp Cys Pro Val Ser Val Gly Asp Val Glu Tyr Asn Phe Asp Phe Gly  
 325 330 335  
 Asp Leu Gln Ile Ser Val Pro Leu Ser Ser Leu Ile Leu Ser Pro Glu  
 340 345 350  
 60 Thr Glu Gly Ser Tyr Cys Gly Phe Ala Val Gln Pro Thr Asn Asp Ser  
 355 360 365  
 65

ES 2 289 288 T3

4et Val Leu Gly Asp Val Phe Leu Ser Ser Ala Tyr Val Val Phe Asp  
 370 375 380  
 5 Leu Asp Asn Tyr Lys Ile Ser Leu Ala Gln Ala Asn Trp Asn Ala Ser  
 385 390 395 400  
 10 Flu Val Ser Lys Lys Leu Val Asn Ile Gln Thr Asp Gly Ser Ile Ser  
 405 410 415  
 15 Gly Ala Lys Ile Ala Thr Ala Glu Pro Trp Ser Thr Asn Glu Pro Phe  
 420 425 430  
 20 Thr Val Thr Ser Asp Ile Tyr Ser Ser Thr Gly Cys Lys Ser Arg Pro  
 435 440 445  
 Phe Leu Gln Ser Ser Thr Ala Ser Ser Leu Ile Ala Glu Thr Asn Val  
 450 455 460  
 25 3ln Ser Arg Asn Cys Ser Thr Lys Met Pro Gly Thr Arg Ser Thr Thr  
 465 470 475 480  
 Val Leu Ser Lys Pro Thr Gln Asn Ser Ala Met His Gln Ser Thr Gly  
 485 490 495  
 30 Ala Val Thr Gln Thr Ser Asn Glu Thr Lys Leu Glu Leu Ser Ser Thr  
 500 505 510  
 Met Ala Asn Ser Gly Ser Val Ser Leu Pro Thr Ser Asn Ser Ile Asp  
 515 520 525  
 35 Lys Glu Phe Glu His Ser Lys Ser Gln Thr Thr Ser Asp Pro Ser Val  
 530 535 540  
 40 Ala Glu His Ser Thr Phe Asn Gln Thr Phe Val His Glu Thr Lys Tyr  
 545 550 555 560  
 Arg Pro Thr His Lys Thr Val Ile Thr Glu Thr Val Thr Lys Tyr Ser  
 565 570 575  
 45 Thr Val Leu Ile Asn Val Cys Lys Pro Thr Tyr  
 580 585  
 50  
 55  
 60  
 65