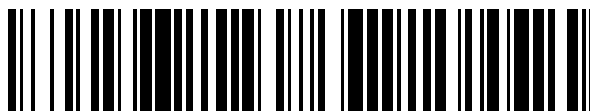


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 373 658**

51 Int. Cl.:

C12N 9/56 (2006.01)

C12N 9/50 (2006.01)

C12N 9/64 (2006.01)

C12N 9/00 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Número de solicitud europea: **04741841 .3**

96 Fecha de presentación: **18.06.2004**

97 Número de publicación de la solicitud: **1633865**

97 Fecha de publicación de la solicitud: **15.03.2006**

54 Título: **NUEVAS ENTIDADES BIOLÓGICAS Y EL USO DE LAS MISMAS.**

30 Prioridad:
18.06.2003 EP 03013819
10.11.2003 EP 03025851
11.11.2003 EP 03025871

45 Fecha de publicación de la mención BOPI:
07.02.2012

45 Fecha de la publicación del folleto de la patente:
07.02.2012

73 Titular/es:
Bayer Pharma Aktiengesellschaft
Müllerstrasse 178
13353 Berlin, DE

72 Inventor/es:
HAUPTS, Ulrich;
KOLTERMANN, Andre;
SCHEIDIG, Andreas;
VOETSMEIER, Christian y
KETTLING, Ulrich

74 Agente: **Carpintero López, Mario**

ES 2 373 658 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

Nuevas entidades biológicas y el uso de las mismas

La presente divulgación proporciona enzimas obtenidas por ingeniería genética comprendidas por un armazón proteico y Regiones Determinantes de Especificidad, la producción de tales enzimas y el uso de las mismas para fines terapéuticos, de investigación, de diagnóstico, de cuidado nutricional, de cuidado personal e industriales.

Antecedentes

La investigación académica e industrial busca continuamente proteínas funcionales para usarlas como agentes terapéuticos, de investigación, de diagnóstico, nutricionales, de cuidado personal o industriales. En la actualidad, tales proteínas funcionales pueden clasificarse principalmente en dos categorías: proteínas naturales y proteínas obtenidas por ingeniería. Las proteínas naturales, por un lado, se descubren en la naturaleza, por ejemplo cribando aislados naturales o secuenciando genomas de diversas especies. Las proteínas obtenidas por ingeniería genética, por otro lado, se basan normalmente en proteínas conocidas y se alteran para adquirir funcionalidades modificadas. En el presente documento se desvelan proteínas obtenidas por ingeniería genética con nuevas funciones en comparación con los componentes de partida. Tales proteínas se denominan NBE (Nuevas Entidades Biológicas). Las NBE desveladas son enzimas obtenidas por ingeniería genética con nuevas especificidades de sustrato o proteínas de fusión de tales enzimas obtenidas por ingeniería genética con otros componentes funcionales.

La especificidad es un elemento esencial de la función enzimática. Una célula consiste en miles de catalizadores altamente reactivos diferentes. Aún así la célula es capaz de mantener un metabolismo coordinado y una estructura tridimensional altamente organizada. Esto se debe en parte a la especificidad de las enzimas, es decir la conversión selectiva de sus sustratos respectivos. La especificidad es una propiedad cualitativa y cuantitativa: la especificidad de una enzima particular puede variar ampliamente, variando de solamente un tipo particular de moléculas diana a todos los tipos moleculares con ciertas subestructuras químicas. En la naturaleza, la especificidad de las enzimas de un organismo ha evolucionado a tenor de las necesidades particulares del organismo. Es poco probable encontrar especificidades arbitrarias con alto valor para aplicaciones terapéuticas, de investigación, de diagnóstico, nutricionales o industriales en el repertorio enzimático de cualquier organismo debido al gran espacio de posibles especificidades. El único modo realista de obtener tales especificidades es su generación *de novo*.

Cuando se comparan enzimas con agentes de unión, se proporciona un paradigma de especificidad por anticuerpos que reconocen epítopos individuales como estructuras pequeñas definidas dentro de moléculas grandes. El vasto intervalo de origen natural de especificidades de anticuerpo se atribuye a la diversidad generada por el sistema inmune en combinación con la selección natural. Varios mecanismos contribuyen al vasto repertorio de especificidad de anticuerpos y aparecen en diferentes etapas de la generación de respuesta inmune y maduración de anticuerpos (Janeway, C y col. (1999) Immunobiology, Elsevier Science Ltd., Garland Publishing, Nueva York). Específicamente, los anticuerpos contienen regiones determinantes de complementariedad (CDR) que interaccionan con el antígeno de una manera altamente específica y permiten la diferenciación incluso entre epítopos muy similares. La cadena ligera así como la pesada del anticuerpo contribuyen cada una con tres CDR al dominio de unión. En la naturaleza se usa la recombinación de diversos segmentos génicos combinados con mutagénesis adicional en la generación de CDR. Como resultado, las secuencias de los seis bucles de CDR son altamente variables en composición y longitud y esto forma la base de la diversidad de especificidades de unión en los anticuerpos. No se conoce un principio similar para la generación de una diversidad de especificidades catalíticas en la naturaleza.

La catálisis, es decir el aumento de la velocidad de una reacción química específica, es además de la unión la función proteica más importante. Las propiedades catalíticas, es decir enzimas, se clasifican de acuerdo con la reacción química que catalizan.

Las transferasas son enzimas que transfieren un grupo, por ejemplo, el grupo metilo o un grupo glucosilo, de un compuesto (generalmente considerado como donador) a otro compuesto (generalmente considerado como aceptor). Por ejemplo, las glucosiltransferasas (EC 2.4) transfieren restos glucosilo de una molécula donadora a una aceptor. Algunas de las glucosiltransferasas también catalizan la hidrólisis, que puede considerarse como transferencia de un grupo glucosilo del donador al agua. La subclase se subdivide adicionalmente en hexosiltransferasas (EC 2.4.1), pentosiltransferasas (EC 2.4.2) y las que transfieren otros grupos glucosilo (EC 2.4.99, Comité de Nomenclatura de la Unión Internacional de Bioquímica y Biología Molecular (NC-IUBMB)).

Las oxidoreductasas catalizan óxido-reducciones. El sustrato que se oxida se considera donador de hidrógeno o electrones. Las oxidoreductasas se clasifican como deshidrogenasas, oxidasas, mono- y dioxigenasas. Las deshidrogenasas transfieren hidrógeno de un donador de hidrógeno a una molécula aceptor de hidrógeno. Las oxidasas reaccionan con oxígeno molecular como aceptor de hidrógeno y producen productos oxidados así como peróxido de hidrógeno o agua. Las monooxigenasas transfieren un átomo de oxígeno del oxígeno molecular al sustrato y uno se reduce a agua. Por el contrario, las dioxigenasas catalizan la inserción de ambos átomos de oxígeno del oxígeno molecular al sustrato.

Las liasas catalizan reacciones de eliminación y de este modo generan dobles enlaces o, en la dirección inversa, cataliza las adiciones en dobles enlaces. Las isomerasas catalizan reordenamientos intramoleculares. Las ligasas

catalizan la formación de enlaces químicos a costa de consumo de ATP.

Finalmente, las hidrolasas son enzimas que catalizan la hidrólisis de enlaces químicos como C-O o C-N. La clasificación E. C. de estas enzimas generalmente las clasifica según la naturaleza del enlace hidrolizado y según la naturaleza del sustrato. Las hidrolasas tales como lipasas y proteasas desempeñan un papel importante en la naturaleza así como en aplicaciones técnicas de biocatalizadores. Las proteasas hidrolizan un enlace peptídico dentro de un contexto de un oligo- o polipéptido. Dependiendo del mecanismo catalítico las proteasas se agrupan en proteasas aspárticas, serina, cisteína, metalo- y treonina proteasas (Handbook of proteolytic enzymes. (1998) Eds: Barret, A; Rawling, N.; Woessner, J.; Academic Press, Londres). Esta clasificación se basa en las cadenas laterales de aminoácidos que son responsables de la catálisis y que se presentan normalmente en el sitio activo en orientación muy similar entre sí. El enlace escindible del sustrato se hace coincidir con los restos catalíticos debido a interacciones específicas entre las cadenas laterales de aminoácidos del sustrato y regiones complementarias de la proteasa (Perona, J. & Craik, C (1995) Protein Science, 4, 337-360). Los restos del extremo N- y C- terminal del enlace escindible se denominan habitualmente P₁, P₂, P₃ etc. y P'₁, P'₂, P'₃ y los bolsillos de unión complementarios al sustrato S₁, S₂, S₃ y S'₁, S'₂, S'₃, respectivamente (nomenclatura de acuerdo con Schlechter & Berger, Biochem. Biophys. Res. Commun. 27 (1967) 157-162). La selectividad de las proteasas puede variar ampliamente de ser prácticamente no selectivas, por ejemplo, las subtilisinas, pasando por una preferencia estricta en la posición P₁, por ejemplo, la Tripsina que corta selectivamente en el extremo C-terminal de restos de arginina o lisina, hasta proteasas altamente específicas, por ejemplo activador de plasminógeno de tipo tisular humano (t-PA) que escinde en el extremo C-terminal de la arginina en la secuencia CPGRWG (Ding, L y col. (1995) Proc. Natl. Acad. Sci. USA 92, 7627-7631; Coombs, G y col. (1996) J. Biol. Chem. 271, 4461-4467).

La especificidad de las proteasas, es decir su capacidad para reconocer e hidrolizar preferentemente ciertos sustratos peptídicos, puede expresarse de forma cualitativa y cuantitativa. La especificidad cualitativa se refiere al tipo de restos aminoácidos que se aceptan por una proteasa en ciertas posiciones del sustrato peptídico. Por ejemplo, la tripsina y t-PA se relacionan con respecto a su especificidad cualitativa, puesto que ambas de ellas requieren en la posición P₁ una arginina o un resto similar. Por otro lado, la especificidad cuantitativa se refiere al número relativo de sustratos peptídicos que se aceptan como sustratos por la proteasa, o de forma más precisa, a las relaciones de k_{cat}/k_M relativas de la proteasa para los diferentes péptidos que se aceptan por la proteasa. Las proteasas que aceptan solamente una parte pequeña de todos los péptidos posibles tiene una alta especificidad, mientras que la especificidad de proteasas que, en caso extremo, escinden cualquier sustrato peptídico serían teóricamente cero.

La comparación de la estructura primaria, secundaria así como la terciaria de proteasas (Fersht, A., Enzyme Structure and Mechanism, W. H. Freeman and Company, Nueva York, 1995) permite la identificación de clases que muestran un alto grado de conservación (Rawlings, N.D. & Barrett, A.J. (1997) En: Proteolysis in Cell Functions Eds. Hopsu-Havu, V.K.; Jarvinen, M.; Kirschke, H, pág. 13-21, IOS Press, Ámsterdam). Un esquema ampliamente aceptado para la clasificación de proteasas se ha propuesto por Rawlings & Barrett (Handbook of proteolytic enzymes. (1998) Eds: Barret, A; Rawling, N.; Woessner, J.; Academic Press, Londres). Por ejemplo, la familia de serina proteasas puede subdividirse en clases estructurales con pliegues de quimiotripsina (clase S1), subtilisina (clase S8) y carboxipeptidasa (clase SC) cada una de las cuales incluye proteasas no específicas así como específicas (Rawlings, N.D. & Barrett, A.J. (1994) Methods Enzymol. 244, 19-61). Esto se aplica a otras familias de proteasa de forma análoga. Puede realizarse una distinción adicional de acuerdo con la localización relativa del enlace escindido en el sustrato. Las carboxi- y aminopeptidasas escinden aminoácidos de los extremos C- y N-terminal, respectivamente, mientras que las endopeptidasas cortan en cualquier lugar a lo largo del oligopéptido.

Muchas aplicaciones serían concebibles si estuvieran disponibles enzimas con un espectro de especificidades básicamente ilimitado. Sin embargo, el uso de tales enzimas con especificidad baja, alta o cualquiera definida está limitado en la actualidad a las que pueden aislarse de fuentes naturales. El campo de aplicación de estas enzimas varía de fines terapéuticos, de investigación, de diagnóstico, nutricionales a cuidado personal e industriales.

Los aditivos enzimáticos en detergentes han llegado a constituir casi un tercio del mercado de enzimas industriales completo. Las enzimas de detergentes incluyen proteinasas para retirar manchas orgánicas, lipasas para retirar manchas grasas, amilasas para retirar restos de alimentos con almidón y celulasas para restaurar superficie lisa de la fibra. La enzima de detergente mejor conocida es probablemente la proteinasa subtilisina no específica, aislada de diversas especies de *Bacillus*.

Las enzimas de almidón, tales como amilasas, ocupan la mayoría de las usadas en el procesamiento de alimentos. Mientras que las enzimas de almidón incluyen productos que son importantes para desencolado textil, fermentación alcohólica, procesamiento de papel y pulpa y aditivos de detergentes de lavandería, la mayor aplicación es para la producción de jarabe de maíz alto en fructosa. La producción de jarabe de maíz de almidón por medio de enzimas industriales fue una alternativa exitosa a la hidrólisis ácida.

Aparte del procesamiento de almidón, las enzimas se usan para una serie creciente de aplicaciones en alimentos. Las enzimas en los alimentos pueden mejorar la textura, apariencia y valor nutricional o pueden generar sabores y aromas deseables. Enzimas alimentarias usadas actualmente en panadería son amilasa, amiloglucosidasas, pentosanasa para rotura de pentosan y producción de gluten reducida o glucosa oxidasas para aumentar la

estabilidad de la masa. Son enzimas habituales para la industria láctea el cuajo (proteasa) como coagulante en la producción de queso, lactasa para hidrólisis de lactosa, proteasa para hidrólisis de proteínas del suero o catalasa para la retirada de peróxidos de hidrógeno. Las enzimas usadas en el procedimiento de fermentación son las anteriormente nombradas amilasas, pero también celulasas o proteasas para aclarar la cerveza de proteínas suspendidas. En vinos y zumos de frutas, la turbidez está provocada más habitualmente por almidón y pectinas de modo que las amilasas y pectinasas aumentan el rendimiento y clarificación. La papaína y otras proteinasas se usan para ablandar la carne.

También se han desarrollado enzimas para ayudar a los animales en la digestión del pienso. En el hemisferio occidental, el almidón es una fuente principal de alimento para vacas, cerdos y pollos. Para mejorar la biodisponibilidad de fosfato del maíz, se añade habitualmente fitasa (Wyss, M. y col. Biochemical characterization of fungal phytases (myo-inositol hexakisphosphate phosphohydrolases): Catalytic properties. Applied & Environmental Microbiology 65, 367-373 (1999)). Además, se ha mostrado que la hidrólisis del fitato proporciona mejoras en la capacidad de digestión de proteína y absorción de minerales tales como calcio (Bedford, M. R. & Schulze, H. EXOGENOUS ENZYMES FOR PIGS AND POULTRY [Revisión]. Nutrition Research Reviews 11, 91-114 (1998)). Otra enzima de pienso principal es xilanasas. Esta enzima es particularmente útil como un complemento para pienso alimentario que comprende más de aproximadamente 10 % de maíz, cebada, o centeno, debido a su contenido en fibra soluble relativamente alto. Las xilanasas provocan dos acciones importantes: reducción de viscosidad de los contenidos intestinales por hidrólisis de los arabinoxilanos de alto peso molecular de tipo gel en el pienso (Murphy, T., C., Bedford, M. R. & McCracken, K. J. Effect of a range of new xylanases on *in vitro* viscosity and on performance of broiler diets. British Poultry Science 44, S16-S18 (2003)) y rotura de polímeros en paredes celulares que mejora la biodisponibilidad de la proteína y almidón.

Los laboratorios de investigación y desarrollo de biotecnología usan de forma rutinaria enzimas especiales en cantidades pequeñas junto con muchos otros reactivos. Esas enzimas crean un mercado significativo para diversas enzimas. Enzimas como fosfatasa alcalina, peroxidasa de rábano rusticano y luciferasa son sólo algunos ejemplos. Las polimerasas de ADN termoestables como Taq polimerasa o endonucleasas de restricción revolucionaron el trabajo en laboratorio. Las enzimas terapéuticas son una clase particular de fármacos, clasificados por la FDA como agentes biológicos, con muchas ventajas en comparación con otros agentes farmacéuticos especialmente no biológicos. Son ejemplos de enzimas terapéuticas exitosas los factores de coagulación humanos como factor VIII y factor IX para tratamiento de seres humanos. Además, las enzimas digestivas se usan para diversas deficiencias en procesos digestivos humanos. Otros ejemplos son t-PA y estreptoquinasa para el tratamiento de enfermedad cardiovascular, beta-glucocerebrosidasa para el tratamiento de enfermedad de Gaucher de Tipo I, L-asparaginasa para el tratamiento de leucemia linfoblástica aguda y ADNasa para el tratamiento de fibrosis quística. Un problema importante en la aplicación de proteínas como agentes terapéuticos es su potencial inmunogenicidad. Para reducir este riesgo, se preferirían enzimas de origen humano, lo que estrecha el conjunto de enzimas disponibles. La provisión de enzimas diseñadas, preferentemente de origen humano, con nuevas especificidades adaptadas permitiría la modificación específica de sustratos diana a voluntad, minimizando el riesgo de inmunogenicidad. Una ventaja adicional de enzimas altamente específicas como agentes terapéuticos sería su menor riesgo de efectos secundarios. Debido a la posibilidad limitada de interacciones específicas entre una molécula pequeña y una proteína, la unión a proteínas no diana y por lo tanto los efectos secundarios son bastante comunes y con frecuencia provocan la terminación de un compuesto candidato de otro modo prometedor. Las enzimas específicas, por otro lado, proporcionan muchos sitios de contacto y mecanismos para diferenciar el sustrato y por lo tanto permiten una mayor especificidad y de este modo menos actividades secundarias.

Las proteasas representan una clase importante de agentes terapéuticos (Drugs of today, 33, 641-648 (1997)). Sin embargo, en la actualidad la proteasa terapéutica es habitualmente un sustituto de actividad insuficiente de las proteasas del propio cuerpo. Por ejemplo, el factor VII puede administrarse en ciertos casos de deficiencias de coagulación de hemorragias o durante la cirugía (Heuer L.; Blumenberg D. (2002) Anaesthetist 51:388). Se aplica activador de plasminógeno de tipo tisular (t-PA) en infarto cardíaco agudo, lo que inicia la disolución de coágulos de fibrina a través de escisión específica y activación de plasminógeno (Verstraete, M. y col. (1995) Drugs, 50, 29-41). Hasta el momento se genera una proteasa con especificidad adaptada para proporcionar un agente terapéutico que activa o inactiva específicamente una proteína diana relacionada con la enfermedad.

Los anticuerpos monoclonales representan otra clase biológica importante de sustancias con capacidades terapéuticas. Una de las principales dianas de anticuerpo son los factores de necrosis tumoral (TNF) que pertenecen a la familia de las citocinas. Los TNF desempeñan un papel principal en el proceso de inflamación. Como homotrímeros podrían unirse a receptores de casi cada célula. Activan una multiplicidad de genes celulares, múltiples mecanismos de transducción de señal, quinasas y factor de transcripción. Los TNF más importantes son TNF-alfa y TNF-beta. TNF-alfa se produce por macrófagos, monocitos y otras células. TNF-alfa es un mediador de inflamación. Por lo tanto, la investigación de la última década se ha centrado en inhibidores de TNF-alfa como anticuerpos monoclonales como posibles agentes terapéuticos para diferentes indicaciones terapéuticas como Artritis Reumatoide, enfermedad de Crohn o psoriasis (Hamilton y col. (2000) Expert Opin Pharmacother, 1 (5): 1041-1052). Una de las principales desventajas de los anticuerpos monoclonales son sus altos costes, de modo que son de gran importancia nuevas alternativas biológicas.

Existen muchos ejemplos de enzimas modificadas por ingeniería genética en la bibliografía. Fulani y col. (Fulani F. y

col. (2003) *Protein Engineering* 16, 515-519) describen una rodanasa (tiosulfato:cianuro sulfurtransferasa) de *Azotobacter vinelandii* que tiene un dominio catalítico estructuralmente relacionado con la subunidad catalítica de las enzimas sulfatasas Cdc25. La diferencia del mecanismo catalítico depende del diferente tamaño del sitio activo. Tanto rodanasa como fosfatasa son altamente específicas en diferentes sustratos (sulfato frente a fosfato). El mecanismo catalítico de la rodanasa podría desplazarse hacia serina/treonina fosfatasa por inserción de resto sencillo. Por lo tanto, Fulani y col. proporciona un ejemplo sencillo para el cambio de un mecanismo catalítico por comparación estructural y alineamiento de secuencia de enzimas conocidas en la naturaleza de diferentes clases enzimáticas pero carecen de una indicación de cómo generar una especificidad de sustrato definible por el usuario manteniendo el mismo mecanismo catalítico.

La tioredoxina reductasa descrita por Briggs y col. (documento WO 02/090300 A2) tiene una especificidad de cofactor alterada que se une preferentemente a NADPH en comparación con NADH. Por lo tanto, ambas enzimas, el punto de partida así como la enzima obtenida por ingeniería genética resultante son altamente específicas para sustratos diferentes. Los métodos para conseguir una especificidad de sustrato alterada tal son procedimientos de procesamiento computacional o alineamiento de secuencia de proteínas relacionadas para definir restos variables y conservados. Todos tienen en común que se basan en la comparación de estructuras y secuencias de proteínas con especificidades conocidas seguido de la transferencia de las mismas a otra cadena principal.

Existen otros ejemplos de enzima con especificidad modificada por ingeniería genética y, en particular, de proteasas que se han publicado en la bibliografía. Ninguno de estos ejemplos, sin embargo, proporciona un medio para generar nuevas especificidades en comparación con la especificidad del material de partida usado dentro de los procedimientos descritos. Los procedimientos varían de mutaciones de punto sencillo dirigidas a la estructura (Kurth, T. y col. (1998) *Biochemistry* 37, 11434-11440; Ballinger, M y col. (1996) *Biochemistry*, 35:13579-13585), intercambio de bucles superficiales entre dos proteasas específicas (Horrevoets y col. (1993) *J. Biol. Chem.* 268, 779-782), a mutagénesis aleatoria regio-selectiva o a lo largo del gen completo en combinación con selección *in vivo* o *in vitro* (Sices, H. & Kristie, T. (1998) *Proc. Natl. Acad. Sci. USA*, 95, 2828-2833).

El diseño racional de la especificidad de proteasa se limita a muy pocos ejemplos. Este enfoque está gravemente limitado por el insuficiente entendimiento de las complejidades que gobiernan el plegamiento y la dinámica así como las relaciones de función-estructura en las proteínas (Corey, M.J. & Corey, E. (1996) *Proc. Natl. Acad. Sci. USA*, 93:11428-11434). Es por lo tanto difícil alterar la secuencia de aminoácidos primaria de una proteasa para cambiar su actividad o especificidad de una manera predecible. En un ejemplo exitoso, Kurth y col. modificaron por ingeniería genética tripsina para que mostrara una preferencia por un motivo dibásico (Kurth, T. y col. (1998) *Biochemistry*, 37:11434-11440). En otro ejemplo, Hedstrom y col. convirtieron la especificidad del sustrato de S₁ de tripsina a la de quimiotripsina (Hedstrom, L. y col. (1992) *Science*, 255:1249-1253). Este es un ejemplo en el que una propiedad conocida se transfirió de una cadena principal a otra.

Ballinger y col. (documento WO 96/27671) describen variantes de subtilisina con mutaciones de combinación (N62D/G166D y opcionalmente Y104D) que tiene un desplazamiento de especificidad de sustrato hacia sustratos peptídicos o polipeptídicos con aminoácidos básicos en las posiciones P1, P2 y P4 del sustrato. Los sustratos adecuados de la subtilisina variante se revelaron buscando en una biblioteca de partículas de fago (fago sustrato) que contenía cinco restos de seleccionados de forma aleatoria contiguos. Estas variantes de subtilisina son útiles para escindir proteínas de fusión con engarces de sustrato básico y procesar hormonas u otras proteínas (*in vitro* o *in vivo*) que contienen sitios de escisión básicos. Los problemas asociados con el rediseño racional de enzimas puede superarse parcialmente por evolución dirigida (como se desvela en el documento PCT/EP03/04864). Estos estudios pueden clasificarse por sus sistemas de expresión y selección. La selección genética significa producir dentro de un organismo una enzima, por ejemplo una proteasa, que es capaz de escindir una proteína precursora que a su vez da como resultado una alteración del comportamiento de crecimiento del organismo productor. A partir de una población de organismos con diferentes proteasas pueden seleccionarse los que tengan un comportamiento de crecimiento alterado. Este principio se indicó por ejemplo por Davis y col. (documentos US 5258289, WO 96/21009). La producción de un sistema de fago depende de la escisión de una proteína de fago que sólo puede activarse en presencia de una enzima proteolítica que es capaz de escindir la proteína de fago. Otros enfoques usan un sistema indicador que permite una selección por cribado en lugar de una selección genética, pero tampoco pueden superar la insuficiencia intrínseca de la caracterización intracelular de las enzimas.

También se han indicado sistemas para generar enzimas con especificidades alteradas con enzimas auto-secretoras. Duff y col. (documento WO 98/11237) describen un sistema de expresión para una proteasa auto-secretora. Un elemento esencial del diseño experimental es que la reacción catalítica actúa en la proteasa en sí misma por un procesamiento autoproteolítico de la molécula precursora unida a membrana para liberar la proteasa madura de la membrana celular al ambiente extracelular. Por lo tanto, debe construirse una proteína de fusión en la que la secuencia peptídica diana reemplace al sitio de escisión natural para autoproteólisis. Las limitaciones de un sistema tal son que las proteasas identificadas de forma segura tendrá la capacidad de escindir una cierta secuencia de aminoácidos pero también pueden escindir muchas otras secuencias peptídicas. Por lo tanto, no se puede conseguir alta especificidad de sustrato. Además, un sistema tal no es capaz de controlar que las proteasas seleccionadas se escindan en una posición específica en una secuencia de aminoácidos definida y no permite una caracterización precisa de las constantes cinéticas de las proteasas seleccionadas (k_{cat} , K_M).

Se ha descrito un procedimiento que se dirige a la generación de nuevas actividades catalíticas y especificidades dentro de las proteínas de barril α/β (documento WO 01/42432; Fersht y col, Methods of producing novel encimes; Altamirano y col (2000) Nature 403, 617-622). Las proteínas de barril α/β comprenden una gran superfamilia de proteínas que representan una fracción grande de todas las enzimas conocidas. La estructura de las proteínas se compone de un barril α/β rodeado de hélices α . Los bucles que conectan las cadenas β y las hélices comprenden la llamada estructura de tapa que incluye los restos de sitio activo. El procedimiento se basa en la clasificación de proteínas de barril α/β en dos clases basándose en la estructura de la tapa catalítica. Una comparación exhaustiva de estructuras proteicas de barril α/β condujo a los autores a la conclusión de que la unión de sustrato y especificidad se define principalmente por la estructura de barril mientras que la especificidad de la reacción química reside dentro de los bucles. Se sugiere que los barriles y estructuras de tapa de diferentes enzimas pueden combinarse para generar nuevas actividades enzimáticas y para proporcionar un punto de partida para ajustar las propiedades por mutagénesis dirigida o aleatoria y selección. El procedimiento no proporciona la generación de especificidad definida por el usuario.

En resumen, resulta evidente que existen muchas aplicaciones posibles en los campos de agentes terapéuticos, investigación y diagnóstico, enzimas industriales, procesamiento de alimentos y piensos, cosmética y otras áreas que resultarían posibles mediante la disponibilidad de enzimas con una nueva especificidad de sustrato. Sin embargo, solamente se ha identificado un número limitado de enzimas específicas de fuentes naturales hasta la fecha. Los procedimientos de diseño racional para modificar, alterar, convertir o transferir especificidad de secuencia así como enfoques aleatorios descritos anteriormente no permitieron la generación de una nueva especificidad definible por el usuario que no estaba presente en el material de partida empleado.

Por lo tanto, ninguno de los procedimientos actualmente disponibles puede proporcionar enzimas con una nueva especificidad de secuencia definible por el usuario. Por el contrario, la presente invención proporciona tales enzimas así como procedimientos para generarlas.

Sumario de la invención

El objetivo es proporcionar proteínas modificadas por ingeniería genética con nuevas funciones que no existen en los componentes usados para la ingeniería genética de tales proteínas. En particular, la divulgación proporciona enzimas con especificidades definibles por el usuario. La especificidad definible por el usuario significa que se proporcionan enzimas con especificidades que no existen en los componentes usados para la ingeniería genética de tales enzimas. Las especificidades pueden seleccionarse por el usuario de modo que se reconozca preferentemente uno o más sustratos diana pretendidos y se convierten por las enzimas. Además, la divulgación proporciona enzimas que poseen esencialmente secuencias idénticas a proteínas humanas pero tienen especificidades diferentes. En una realización particular, la divulgación proporciona proteasas con especificidades definibles por el usuario.

Además, la presente divulgación se refiere a enzimas obtenidas por ingeniería genética que se fusionan con uno o más componentes funcionales adicionales. Estos componentes adicionales pueden ser componentes proteicos que preferentemente tienen propiedades de unión y son del grupo que consiste en dominios de unión a sustrato, anticuerpos, receptores o fragmentos de los mismos. Además, estos componentes adicionales pueden ser componentes funcionales adicionales, preferentemente seleccionándose del grupo que consiste en polietilenglicoles, carbohidratos, lípidos, ácidos grasos, ácidos nucleicos, metales, quelados metálicos y fragmentos o derivados de los mismos. Las proteínas de fusión resultantes se entienden como enzimas con especificidades definibles por el usuario.

Además, la divulgación se refiere a la aplicación de tales enzimas con nuevas especificidades definibles por el usuario para fines terapéuticos, de investigación, diagnósticos, nutricionales, de cuidado personal o industriales. Además, la divulgación se refiere a un procedimiento para generar enzimas obtenidas por ingeniería genética con especificidades definibles por el usuario. En particular, la divulgación se refiere a generar enzimas que posean secuencias esencialmente idénticas a enzimas humanas pero que tengan especificidades diferentes.

Este problema se ha resuelto por las realizaciones especificadas en la descripción posterior y en las reivindicaciones. La presente divulgación se refiere por lo tanto a

(1) una enzima proteolítica con actividad catalítica de especificidad definida no conferida por el armazón proteico y caracterizada por una combinación de los siguientes componentes:

(a) un armazón proteico que tiene al menos 90 % de homología con la tripsina humana I que tiene la secuencia de aminoácidos mostrada en SEC ID N^o: 1 y que es capaz de catalizar al menos una escisión peptídica en al menos un sustrato peptídico diana y

(b) una o más regiones determinantes de especificidad insertadas o sustituidas con el armazón proteico en sitios en el armazón proteico que permiten que la enzima proteolítica resultante distinga el sustrato diana en tantos sitios como sea necesario para hidrolizar preferentemente el sustrato diana frente a uno o más otros sustratos y sustituyéndose o insertándose las regiones determinantes de especificidad en una o más posiciones del grupo de posiciones que corresponden estructuralmente o por homología de secuencia de aminoácidos a las regiones 18-

25, 38-48, 54-63, 73-86, 122-130, 148-156, 165-171 y 194-204 en tripsina humana I que tiene la secuencia de aminoácidos mostrada en SEC ID N°: 1, y siendo las regiones determinantes de especificidad secuencias peptídicas que tienen una longitud de menos de 50 restos aminoacídicos;

5 (2) el uso de una enzima proteolítica como se ha definido en (1) anteriormente para fines terapéuticos, de investigación, diagnósticos, nutricionales, de cuidado personal o industriales;

(3) un procedimiento para generar una enzima proteolítica como se ha definido en (1) anteriormente que tiene especificidad definida para al menos un sustrato diana, no estando presente dicha especificidad en los componentes de partida individuales, que comprende al menos las siguientes etapas:

10 (a) proporcionar un armazón proteico que tiene al menos 90 % de homología con tripsina humana I que tiene la secuencia de aminoácidos mostrada en SEC ID N°: 1, que cataliza al menos una reacción química en al menos un sustrato diana,

15 (b) generar una biblioteca de enzimas proteolíticas o enzimas proteolíticas aisladas combinando un polinucleótido que codifica el armazón proteico de la etapa (a) mediante inserción o sustitución con de 1 a 11 secuencias oligonucleotídicas sintéticas parcial o completamente aleatorias que codifican secuencias peptídicas con una longitud de menos de 50 restos aminoacídicos en una o más posiciones del grupo de posiciones dentro del polinucleótido que codifica el armazón proteico que corresponde estructuralmente o por homología de secuencia de aminoácidos a las regiones 18-25, 38-48, 54-63, 73-86, 122-130, 148-156, 165-171 y 194-204 en tripsina humana I que tiene la secuencia de aminoácidos mostrada en SEC ID N°: 1, que expresa dichas enzimas y (c) seleccionando de la biblioteca de enzimas proteolíticas generadas en la etapa (b) una o más enzimas que tienen especificidades definidas no conferidas por el armazón proteico proporcionado en la etapa (a) para al menos un sustrato diana;

20 (4) una proteína de fusión que está comprendida por al menos una enzima proteolítica como se ha definido en (1) anteriormente y

25 (i) al menos un componente proteico adicional, seleccionándose preferentemente del grupo que consiste en dominios de unión, receptores, anticuerpos, dominios de regulación, pro-secuencias y fragmentos de los mismos y/o

30 (ii) al menos un componente funcional adicional, seleccionándose preferentemente del grupo que consiste en polietilenglicoles, carbohidratos, lípidos, ácidos grasos, ácidos nucleicos, metales y quelados metálicos;

(5) una composición o composición farmacéutica que comprende una o más enzimas proteolíticas como se ha definido en (1) anteriormente o una proteína de fusión como se ha definido en (4) anteriormente, pudiendo comprender opcionalmente dicha composición farmacéutica un vehículo, excipiente y/o agente adyuvante aceptable;

35 (6) un ácido nucleico que codifica una enzima proteolítica como se ha definido en (1) anteriormente o una proteína de fusión como se ha definido en (4) anteriormente;

(7) un vector que comprende el ácido nucleico como se ha definido en (6) anteriormente;

40 (8) una célula huésped u organismo transgénico que se transforma/transfecta con un vector como se ha definido en (7) anteriormente o que comprende el ácido nucleico como se ha definido en (6) anteriormente y;

(9) un procedimiento para producir la enzima proteolítica como se ha definido en (1) anteriormente o una proteína de fusión como se ha definido en (4) anteriormente que comprende cultivar una célula u organismo como se ha definido en (8) anteriormente y opcionalmente aislar la enzima del caldo de cultivo.

Breve descripción de las figuras

45 Las siguientes figuras se proporcionan para explicar adicionalmente la presente invención en complemento a la descripción detallada:

La Figura 1 ilustra la estructura tridimensional de tripsina humana I con los restos del sitio activo mostrados en representación de "bolas y varillas" e indicando las regiones marcadas sitios de inserción de SDR potenciales.

50 La Figura 2 muestra el alineamiento de la secuencia de aminoácidos primaria de tres miembros de la familia S1 de la clase de serina proteasa: tripsina humana I, alfa trombina humana y enteropeptidasa humana (véase también SEC ID N°: 1, 5 y 6).

La Figura 3 ilustra la estructura tridimensional de subtilisina mostrándose los restos del sitio activo en representación de "bolas y varillas" indicando las regiones numeradas sitios de inserción de SDR

potenciales.

La Figura 4 muestra el alineamiento de las secuencias de aminoácidos primarias de cuatro miembros de la familia de la clase S8 de serina proteasa: subtilisina E, furina, PC1 y PC5 (véase también SEC ID N°: 7-10).

5 La Figura 5 ilustra la estructura tridimensional de pepsina mostrándose los restos de sitio activo en representación de “bolas y varillas” indicado en las regiones numeradas en sitios de inserción de SDR potenciales.

La Figura 6 muestra el alineamiento de las secuencias de aminoácidos primarias de tres miembros de la familia de proteasa de ácido aspártico A1: pepsina, β -secretasa y catepsina D (véase también SEC ID N°: 11-13).

10 La Figura 7 ilustra la estructura tridimensional de caspasa 7 mostrándose los restos del sitio activo en representación “bolas y varillas” indicando las regiones numeradas sitios de inserción de SDR potenciales.

La Figura 8 muestra la secuencia de aminoácidos primaria de caspasa 7 como un miembro de la familia de clase C14 de cisteína proteasa (véase también SEC ID N°: 14).

La Figura 9 representa esquemáticamente el tercer aspecto de la divulgación.

15 La Figura 10 muestra un análisis de transferencia de Western de un sobrenadante de cultivo de células que expresan variantes de tripsina humana I con SDR1 y SDR2, en comparación con controles negativos.

La Figura 11 muestra el transcurso en el tiempo de la escisión proteolítica de un sustrato diana por tripsina humana I.

20 La Figura 12 muestra las actividades relativas de tres variantes de enzimas proteolíticas obtenidas por ingeniería genética en comparación con tripsina humana I en dos sustratos peptídicos diferentes.

La Figura 13 muestra las especificidades relativas de tripsina humana I y variantes de enzimas proteolíticas obtenidas por ingeniería genética con una o dos SDR, respectivamente.

25 La Figura 14 muestra las especificidades relativas de tripsina humana I y de variantes de enzimas proteolíticas obtenidas por ingeniería genética que son específicas para TNF-alfa humano con este armazón en péptidos con una secuencia diana de TNF-alfa humano.

La Figura 15 muestra la reducción de citotoxicidad inducida por TNF-alfa cuando se incubaba el TNF-alfa con sobrenadante concentrado de cultivos que expresan las enzimas proteolíticas obtenidas por ingeniería genética que son específicas para TNF-alfa humano.

30 La Figura 16: muestra la reducción de citotoxicidad inducida por TNF-alfa cuando se incubaba el TNF-alfa con enzima proteolítica obtenida por ingeniería genética purificada que es específica para TNF-alfa humano.

La Figura 17 compara la actividad de enzimas proteolíticas obtenidas por ingeniería genética que son específicas para TNF-alfa humano con la actividad de tripsina humana I en dos sustratos proteicos: (a) TNF-alfa humano; (b) mezcla de proteínas del suero humanas.

35 La Figura 18 muestra la actividad específica de una enzima proteolítica obtenida por ingeniería genética con especificidad para VEGF humano.

Definiciones

En el marco de la presente invención se usan los siguientes términos y definiciones.

40 El término “proteasa” significa cualquier molécula proteica que es capaz de hidrolizar enlaces peptídicos. Esto incluye enzimas proteolíticas artificiales o de origen natural, así como variantes de las mismas obtenidas por mutagénesis dirigida o aleatoria o cualquier otro procedimiento de ingeniería proteica, cualquier fragmento activo de una enzima proteolítica o cualquier complejo molecular o proteína de fusión que comprende una de las proteínas anteriormente mencionadas. Una “quimera de proteasas” significa una proteína de fusión de dos o más fragmentos derivados de diferentes proteasas parentales.

45 El término “sustrato” significa cualquier molécula que pueda convertirse catalíticamente por una enzima. La expresión “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 por una proteasa. El enlace peptídico que se hidroliza se denomina “sitio de escisión”. La numeración de las posiciones en el sustrato se realiza de acuerdo con el sistema introducido por Schlechter & Berger (Biochem. Biophys. Res Commun. 27 (1967) 157-162). Los restos aminoacídicos adyacentes N-terminal al sitio de escisión se numeran P₁, P₂, P₃, etc., mientras que los restos adyacentes C-terminal al sitio de escisión se numeran P₁’, P₂’, P₃’, etc.

La expresión “sustrato diana” describe un sustrato definido por el usuario que se reconoce específicamente y se convierte por una enzima de acuerdo con la invención. La expresión “sustrato peptídico diana” describe un sustrato peptídico definido por el usuario. La expresión “especificidad de diana” describe la especificidad cualitativa y cuantitativa de una enzima que es capaz de reconocer y convertir un sustrato diana. Las propiedades catalíticas de las enzimas 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”, de acuerdo con las definiciones de Michaelis y Menten (Fersht, A., Enzyme Structure and Mechanism, W. H. Freeman and Company, Nueva York, 1995). La expresión “actividad catalítica” describe cuantitativamente la conversión de un sustrato dado en condiciones de reacción definidas.

El término “especificidad” significa la capacidad de una enzima para reconocer y convertir preferentemente ciertos sustratos. La especificidad puede expresarse cualitativa y cuantitativamente. “Especificidad cualitativa” se refiere a la naturaleza química de los restos de sustrato que se reconocen por una enzima. “Especificidad cuantitativa” se refiere al número de sustratos que se aceptan como sustratos. La especificidad cuantitativa puede expresarse por el término s , que se define como el logaritmo negativo del número de todos los sustratos aceptados dividido por el número de todos los posibles sustratos. Las proteasas, por ejemplo, que aceptan preferentemente una parte pequeña de todos los posibles sustratos peptídicos tienen una “alta especificidad”. Las proteasas que aceptan casi cualquier sustrato peptídico tienen una “baja especificidad”. Se realizan definiciones de acuerdo con el documento WO 03/095670. Las proteasas con muy baja especificidad también se denominan “proteasas no específicas”. La expresión “especificidad definida” se refiere a un cierto tipo de especificidad, es decir, a un cierto sustrato diana o un conjunto de ciertos sustratos diana que se convierten preferentemente frente a otros sustratos.

La expresión “obtenido por ingeniería genética” en combinación con el término “enzima” describe una enzima que está comprendida por diferentes componentes y que tiene características no conferidas por los componentes individuales por sí solos.

La expresión “armazón proteico” o “proteína de armazón” se refiere a una diversidad de estructuras polipeptídicas primarias, secundarias y terciarias.

La expresión “secuencia peptídica” indica cualquier secuencia peptídica usada para inserción o sustitución en o combinación con un armazón proteico. Las secuencias peptídicas se obtienen habitualmente por expresión de secuencias de ADN que pueden sintetizarse de acuerdo con técnicas bien establecidas o pueden obtenerse de fuentes naturales. La inserción, sustitución o combinación de secuencias peptídicas con el armazón proteico se generan por inserción, sustitución o combinación de oligonucleótidos en o con un polinucleótido que codifica el armazón proteico. El término “sintético” en combinación con la expresión “secuencia peptídica” se refiere a secuencias peptídicas que no están presentes en el armazón proteico en el que las secuencias peptídicas se insertan o sustituyen o con las que se combinan.

El término “componentes” en combinación con la expresión “enzima obtenida por ingeniería genética” se refiere a secuencias peptídicas o polipeptídicas que se combinan en la modificación por ingeniería genética de tales enzimas. Tales componentes pueden entre otros comprender uno o más armazones proteicos y una o más secuencias peptídicas sintéticas. La expresión “biblioteca de enzimas obtenidas por ingeniería genética” describe una mezcla de enzimas obtenidas por ingeniería genética, por lo que cada enzima obtenida por ingeniería genética sencilla se codifica por una secuencia polinucleotídica diferente. La expresión “genoteca” indica una biblioteca de polinucleótidos que codifica la biblioteca de enzimas obtenidas por ingeniería genética. El término “SDR” o “región determinante de especificidad” se refiere a una secuencia peptídica sintética que proporciona la especificidad definida cuando se combina con el armazón proteico en sitios que permiten que las enzimas resultantes diferencien entre el sustrato diana y uno o más otros sustratos. Tales sitios se denominan “sitios SDR”.

Las expresiones “estructura terciaria similar a la estructura de” y “estructura terciaria similar” en combinación con los términos “enzima” o “proteína” se refieren a proteínas en las que el tipo, secuencia, conectividad y orientación relativa de los elementos estructurales secundarios típicos de una proteína, por ejemplo, hélices alfa, láminas beta, giros beta y bucles, son similares y las proteínas se agrupan por lo tanto en la misma clase o pliegue estructural o topológico. Esto incluye proteínas que tienen elementos estructurales alterados, adicionales o suprimidos de cualquier tipo pero una topología en lo demás no cambiada. Los ejemplos de tales clases estructurales son la superfamilia de TNF, el pliegue S1 o el pliegue S8 dentro de las serina proteasas, los GPCR o el pliegue de barril α/β .

La expresión “posiciones que corresponden estructuralmente” indica aminoácidos en proteínas de estructura terciaria similar que se corresponden estructuralmente entre sí, es decir se localizan habitualmente dentro del mismo elemento estructural o topológico de la estructura. Dentro del elemento estructural poseen las mismas posiciones relativas con respecto al comienzo y al final del elemento estructural. Si, por ejemplo, la comparación topológica de dos proteínas revela dos secuencias estructuralmente correspondientes de diferente longitud, entonces los aminoácidos dentro de, por ejemplo 20 % y 40 % de las longitudes de región respectivas, corresponden estructuralmente entre sí.

La expresión “biblioteca de enzimas obtenidas por ingeniería genética” se refiere a una multiplicidad de enzimas o

variantes de enzima, que pueden existir como una mezcla o en forma aislada.

Los restos aminoacídicos se abrevian de acuerdo con la siguiente Tabla 1 en código de una o tres letras.

Tabla 1: Abreviaturas de aminoácidos

| Abreviaturas | | Aminoácidos |
|--------------|-----|-----------------|
| 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 |

Descripción detallada de la invención

- 5 La presente divulgación proporciona proteínas obtenidas por ingeniería genética con nuevas funciones. En particular, la divulgación proporciona enzimas con especificidades definibles por el usuario. En una realización particular, la divulgación proporciona proteasas con especificidades definibles por el usuario. Además, la divulgación proporciona aplicaciones de tales enzimas con nuevas especificidades definibles por el usuario para fines terapéuticos, de investigación, de diagnóstico, nutricionales, de cuidado personal o industriales. Además, la divulgación proporciona un procedimiento para generar enzimas con especificidades que no están presentes en los componentes usados para la ingeniería genética de tales enzimas. En particular, la divulgación se refiere a la generación de enzimas que tienen secuencias que son esencialmente idénticas a enzimas de mamífero especialmente humanas pero que tienen diferentes especificidades. Además, la divulgación proporciona bibliotecas de enzimas obtenidas por ingeniería genética específicas con especificidades correspondientes codificadas genéticamente, un procedimiento para la generación de bibliotecas de enzimas obtenidas por ingeniería genética específicas con especificidades correspondientes codificadas genéticamente y la aplicación de tales bibliotecas para fines técnicos, diagnósticos, nutricionales, de cuidado personal o de investigación.

Un primer aspecto desvela enzimas obtenidas por ingeniería genética con especificidades definidas. Estas enzimas obtenidas por ingeniería genética se caracterizan por los siguientes componentes:

- 20 (a) un armazón proteico capaz de catalizar al menos una reacción química en un sustrato y
 (b) uno o más regiones determinantes de especificidad (SDR) localizadas en sitios en el armazón proteico que permiten que la proteína obtenida por ingeniería genética resultante diferencie entre al menos un sustrato diana y uno o más diferentes sustratos, siendo las SDR esencialmente secuencias peptídicas sintéticas.
- 25 Preferentemente, dicha especificidad definida de las enzimas obtenidas por ingeniería genética no se confiere por el

almazón proteico.

En principio, el almacén proteico puede tener una diversidad de estructuras primarias, secundarias y terciarias. La estructura primaria, es decir, la secuencia de aminoácidos, puede ser una secuencia obtenida por ingeniería genética o puede derivarse de cualquier origen viral, procariota o eucariota. Para uso terapéutico humano, sin embargo, el almacén proteico es preferentemente de origen mamífero y más preferentemente de origen humano. Además, el almacén proteico es capaz de catalizar una o más reacciones químicas y tiene preferentemente solamente una especificidad baja.

Preferentemente, se usan derivados del almacén proteico que tienen secuencias de aminoácidos modificadas que confieren características mejoradas para la aplicabilidad como almacenes proteicos. Tales características mejoradas comprenden, pero sin limitación, estabilidad; rendimiento de expresión o secreción; plegamiento, en particular después de combinación del almacén proteico con SDR; aumento o disminución de la sensibilidad a reguladores tales como activadores o inhibidores; inmunogenicidad; velocidad catalítica; K_M o afinidad de sustrato.

Las enzimas obtenidas por ingeniería genética revelan su especificidad cuantitativa de las secuencias peptídicas sintéticas que se combinan con el almacén proteico. Por lo tanto, las secuencias peptídicas obtenidas por ingeniería genética actúan como regiones determinantes de especificidad o SDR. El número, la longitud y las posiciones de tales SDR pueden variar en un amplio intervalo. El número de SDR dentro del almacén es al menos una, preferentemente más de una, más preferentemente entre dos y once, más preferentemente entre dos y seis. Las SDR tienen una longitud entre uno y 50 restos aminoácidos, preferentemente una longitud de entre uno y 15 restos aminoácidos, más preferentemente una longitud de entre uno y seis restos aminoácidos. Como alternativa, las SDR tienen una longitud de entre dos y 20 restos aminoácidos, preferentemente una longitud de entre dos y diez restos aminoácidos, más preferentemente una longitud de entre tres y ocho restos aminoácidos.

Las enzimas obtenidas por ingeniería genética pueden describirse adicionalmente como moléculas proteicas de tipo anticuerpo que comprenden regiones constantes y variables, pero que tienen una cadena principal no de inmunoglobulina y que tienen un sitio activo (actividad catalítica) en la región constante, por lo que la especificidad de sustrato del sitio activo se modula por la región variable. Preferentemente, como en la estructura de inmunoglobulina, las regiones variables son bucles de longitud y composición variable que interaccionan con una molécula diana.

En particular, las enzimas obtenidas por ingeniería genética tienen actividad hidrolasa. En una variante preferida, las enzimas obtenidas por ingeniería genética tienen actividad proteolítica. Son almacenes proteicos particularmente preferidos para esta variante proteasas no específicas o partes de proteasas no específicas o derivan de otro modo de proteasas no específicas. Las expresiones "derivado de" o "un derivado del mismo" con respecto a esto y en las siguientes variantes y realizaciones se refieren a derivados de proteínas que mutan en una o más posiciones aminoácidas y/o tienen una homología de al menos el 70 %, preferentemente del 90 %, más preferentemente del 95 % y más preferentemente del 99 %, que se procesan proteolíticamente, que tienen un patrón de glucosilación alterado, que se unen covalentemente a sustancias no proteicas, que se fusionan con dominios proteicos adicionales, que tienen truncamientos C terminales y/o N terminales, y/o que tienen inserciones, sustituciones y/o deleciones específicas. Como alternativa, "derivado de" puede referirse a derivados que son combinaciones o quimeras de dos o más fragmentos de dos o más proteínas, cada una de los cuales comprende opcionalmente cualquiera o todas las modificaciones anteriormente mencionadas. La estructura terciaria del almacén proteico puede ser de cualquier tipo. Preferentemente, sin embargo, la estructura terciaria pertenece a una de las siguientes clases estructurales: clase S1 (pliegue de quimiotripsina de la familia de serina proteasas), clase S8 (pliegue de subtilisina de la familia de serina proteasas), clase SC (pliegue de carboxipeptidasa de la familia de serina proteasas) clase A1 (pliegue de pepsina A de las proteasas aspárticas) o clase C14 (pliegue de caspasa-1 de las cisteína proteasas). Son ejemplos de proteasas que pueden actuar como el almacén proteico de enzimas proteolíticas obtenidas por ingeniería genética para el uso como agentes terapéuticos humanos o derivan de tripsina humana, trombina humana, quimiotripsina humana, pepsina humana, endotiapepsina humana, caspasas humanas 1 a 14 y/o furina humana.

La especificidad definida de las enzimas proteolíticas obtenidas por ingeniería genética es una medida de su capacidad para diferenciar entre al menos un péptido diana o sustratos proteicos y uno o más sustratos peptídicos o proteicos adicionales. Preferentemente, la especificidad definida se refiere a la capacidad para diferenciar sustratos peptídicos o proteicos que difieren en otras posiciones distintas del sitio P1, más preferentemente, la especificidad definida se refiere a la capacidad para diferenciar sustratos peptídicos o proteicos que difieren en otras posiciones distintas del sitio P1 y el sitio P1'. Más preferentemente, las enzimas proteolíticas obtenidas por ingeniería genética distinguen sustratos peptídicos o proteicos diana en tantos sitios como sea necesario para hidrolizar preferentemente el sustrato diana frente a otras proteínas. Como ejemplo, una enzima proteolítica obtenida por ingeniería genética terapéuticamente útil aplicada por vía intravenosa en el cuerpo humano debería ser suficientemente específica para diferenciar entre el sustrato diana y cualquier otra proteína en el suero humano. Preferentemente, una enzima proteolítica obtenida por ingeniería genética tal reconoce y diferencia sustratos peptídicos en tres o más posiciones de aminoácidos, más preferentemente en cuatro o más posiciones e incluso más preferentemente en cinco o más posiciones de aminoácidos. Estas posiciones pueden estar adyacentes o no adyacentes.

En una primera realización, el armazón proteico tiene una estructura terciaria o pliegue igual o similar a la estructura terciaria o pliegue de la subclase estructural S1 de la serina proteasas, es decir el pliegue de quimiotripsina y/o tiene al menos un 70 % de identidad en el nivel de aminoácidos con una proteína de la subclase estructural S1 de serina proteasas. Se prefiere que las SDR se inserten en el armazón proteico en una o más posiciones del grupo de posiciones que corresponden estructuralmente o por homología de secuencia de aminoácidos a las regiones 18-25, 38-48, 54-63, 73-86, 122-130, 148-156, 165-171 y 194-204 en tripsina humana I y más preferentemente en una o más posiciones del grupo de posiciones que corresponden estructuralmente o por homología de secuencia de aminoácidos a las regiones 20-23, 41-45, 57-60, 76-83, 125-128, 150-153, 167-169 y 197-201 (numeración de aminoácidos de acuerdo con SEC ID N°: 1). El número de SDR a combinar con este tipo de armazón proteico está preferentemente entre 1 y 10 y más preferentemente entre 2 y 4. Preferentemente, el armazón proteico es igual a o es un derivado u homólogo de una o más de las siguientes proteínas: quimiotripsina, granzime, calicreína, tripsina, mesotripsina, elastasa de neutrófilos, elastasa pancreática, enteropeptidasa, catepsina, trombina, ancrod, factor de coagulación IXa, factor de coagulación VIIa, factor de coagulación Xa, proteína activada C, uroquinasa, activador del plasminógeno de tipo tisular, plasmina, activador de plasminógeno de tipo Desmodus. Más preferentemente, el armazón proteico es tripsina o trombina o es un derivado u homólogo de tripsina o trombina. Para el uso como un agente terapéutico humano, el armazón de tripsina o trombina es más preferentemente de origen humano para minimizar el riesgo de una respuesta inmune o una reacción alérgica.

Preferentemente, se usan derivados con características mejoradas derivados de tripsina humana I o de proteínas con estructura terciaria similar. Los ejemplos preferidos de tales derivados derivan de tripsina humana I (SEC ID N°: 1) y comprenden una o más de las siguientes sustituciones de aminoácidos E56G; R78W; Y131F; A146T; C183R.

Se prefiere que se inserten al menos una de dos SDR en tripsina humana I, o un derivado de la misma, entre los restos 42 y 43 (SDR 1) y entre 123 y 124 (SDR 2), respectivamente (numeración de aminoácidos de acuerdo con SEC ID N°: 1). Además la SDR 1 tiene una longitud preferida de 6 y la SDR 2 tiene una longitud preferida de 5 aminoácidos, respectivamente. En una variante preferida de esta realización, las secuencias SDR 1 y SDR 2 comprenden una de las secuencias de aminoácidos enumeradas en la Tabla 2. Tales enzimas proteolíticas obtenidas por ingeniería genética tienen especificidad por el sustrato diana B como se ejemplifica en el ejemplo IV.

En una realización adicional el armazón proteico pertenece a la subclase estructural S8 de serina proteasas y/o tiene una estructura terciaria similar a subtilisina E de *Bacillus subtilis* y/o tiene al menos un 70 % de identidad en el nivel de aminoácidos con una proteína de la subclase estructural S8 de serina proteasas. Preferentemente, el armazón pertenece a la familia de subtilisina o las convertasas de pro-proteína humanas. Se prefiere que las SDR se inserten en el armazón proteico en una o más posiciones del grupo de posiciones que corresponden estructuralmente o por homología de secuencia de aminoácidos a las regiones 6-17, 25-29, 47-55, 59-69, 101-111, 117-125, 129-137, 139-154, 158-169, 185-195 y 204-225 en subtilisina E de *Bacillus subtilis* y más preferentemente en uno o más posiciones del grupo de posiciones que corresponden estructuralmente o por homología de secuencia de aminoácidos a las regiones de 59-69, 101-111, 129-137, 158-169 y 204-225 (numeración de aminoácidos de acuerdo con SEC ID N°: 7). Se prefiere que el armazón proteico sea igual o sea un derivado u homólogo de una o más de las siguientes proteínas: subtilisina Carlsberg; subtilisina E de *B. subtilis*; subtilisina BPN'; subtilisina de *B. licheniformis*; subtilisina de *B. lentus*; proteasa alcalina de *Bacillus alcalophilus*; proteinasa K; kexina; convertasa de pro-proteína humana; furina humana. En una variante preferida, se usa subtilisina BPN' o una de las proteínas SPC 1 a 7 como el armazón proteico.

En una realización adicional el armazón proteico pertenece a la familia de proteasas aspárticas y/o tiene una estructura terciaria similar a pepsina humana. Preferentemente, el armazón pertenece a la clase A1 de proteasas y/o tiene al menos un 70 % de identidad en el nivel de aminoácidos con una proteína de la clase A1 de proteasas. Se prefiere que las SDR se inserten en el armazón proteico en una o más posiciones del grupo de posiciones que corresponden estructuralmente o por homología de secuencia de aminoácidos a las regiones 6-18, 49-55, 74-83, 91-97, 112-120, 126-137, 159-164, 184-194, 242-247, 262-267 y 277-300 en pepsina humana y más preferentemente en una o más posiciones del grupo de posiciones que corresponden estructuralmente o por homología de secuencia de aminoácidos a las regiones 10-15, 75-80, 114-118, 130-134, 186-191 y 280-296 (numeración de aminoácidos de acuerdo con SEC ID N°: 11). Se prefiere que el armazón proteico sea igual a o sea un derivado u homólogo de una o más de las siguientes proteínas: pepsina, quimosina, renina, catepsina, yapsina. Preferentemente, se usa pepsina o endotiopepsina o un derivado u homólogo de las mismas como el armazón proteico.

En una realización adicional el armazón proteico pertenece a la familia de cisteína proteasa y/o tiene una estructura terciaria similar a caspasa humana 7. Preferentemente el armazón pertenece a la clase C14 de cisteína proteasas o tiene al menos un 70 % de identidad en el nivel de aminoácidos con una proteína de la clase C14 de cisteína proteasas. Se prefiere que las SDR se inserten en el armazón proteico en una o más posiciones del grupo de posiciones que corresponden estructuralmente o por homología de secuencia de aminoácidos a las regiones 78-91, 144-160, 186-198, 226-243 y 271-291 en la caspasa humana 7 y más preferentemente en una o más posiciones del grupo de posiciones que corresponden estructuralmente o por homología de secuencia de aminoácidos a las regiones 80-86, 149-157, 190-194 y 233-238 (numeración de aminoácidos de acuerdo con SEC ID N°: 14). Se prefiere que el armazón proteico sea igual a o sea un derivado u homólogo de una de las caspasas 1 a 9.

En una realización adicional el armazón proteico pertenece a la clase S11 de serina proteasas o tiene al menos un

- 70 % de identidad en el nivel de aminoácidos con una proteína de la clase S11 de serina proteasas y/o tiene una estructura terciaria similar a D-alanil-D-alanina transpeptidasa de la especie de *Streptomyces* K15. Se prefiere que las SDR se inserten en el armazón proteico en una o más posiciones del grupo de posiciones que corresponden estructuralmente o por homología de secuencia de aminoácidos a las regiones 67-79, 137-150, 191-206, 212-222 y 251-241 en D-alanil-D-alanina transpeptidasa de la especie de *Streptomyces* K15 y más preferentemente en una o más posiciones del grupo de posiciones que corresponden estructuralmente o por homología de secuencia de aminoácidos a las regiones 70-75, 141-147, 195-202 y 216-220 (numeración de aminoácidos de acuerdo con SEC ID N°: 15). Se prefiere que la D-alanil-D-alanina transpeptidasa de la especie de *Streptomyces* K15 o un derivado u homólogo de la misma se use como el armazón.
- En una realización adicional el armazón proteico pertenece a la clase S21 de serina proteasas o tiene al menos un 70 % de identidad en el nivel de aminoácidos con una proteína de la clase S21 de serina proteasas y/o tiene una estructura terciaria similar a ensamblina de citomegalovirus humano. Se prefiere que las SDR se inserten en el armazón proteico en una o más posiciones del grupo de posiciones que corresponden estructuralmente o por homología de secuencia de aminoácidos a las regiones 25-33, 64-69, 134-155, 162-169 y 217-244 en ensamblina de citomegalovirus humano y más preferentemente en una o más posiciones del grupo de posiciones que corresponden estructuralmente o por homología de secuencia de aminoácidos a las regiones 27-31, 164-168 y 222-239 (numeración de aminoácidos de acuerdo con SEC ID N°: 16). Se prefiere que la ensamblina de citomegalovirus humano o un derivado u homólogo de la misma se use como el armazón.
- En una realización adicional el armazón proteico pertenece a la clase S26 de serina proteasas o tiene al menos un 70 % de identidad en el nivel de aminoácidos con una proteína de la clase S26 de serina proteasas y/o tiene una estructura terciaria similar a la peptidasa señal de *Escherichia coli*. Se prefiere que las SDR se inserten en el armazón proteico en una o más posiciones del grupo de posiciones que corresponden estructuralmente o por homología de secuencia de aminoácidos a las regiones 8-14, 57-68, 125-134, 239-254, 200-211 y 228-239 en peptidasa de señal de *Escherichia coli* y más preferentemente en una o más posiciones del grupo de posiciones que corresponden estructuralmente o por homología de secuencia de aminoácidos a las regiones 9-13, 60-67, 127-132 y 203-209 (numeración de aminoácidos de acuerdo con SEC ID N°: 17). Se prefiere que la peptidasa señal de *Escherichia coli* o un derivado u homólogo de la misma se use como el armazón.
- En una realización adicional el armazón proteico que pertenece a la clase S33 de serina proteasas o tiene al menos un 70 % de identidad en el nivel de aminoácidos con una proteína de la clase S33 de serina proteasas y/o tiene una estructura terciaria similar a la prolil aminopeptidasa de *Serratia marcescens*. Se prefiere que las SDR se inserten en el armazón proteico en una o más posiciones del grupo de posiciones que corresponden estructuralmente o por homología de secuencia de aminoácidos a las regiones 47-54, 152-160, 203-212 y 297-302 en prolil aminopeptidasa de *Serratia marcescens* y más preferentemente en una o más posiciones del grupo de posiciones que corresponden estructuralmente o por homología de secuencia de aminoácidos a las regiones, 50-53, 154-158 y 206-210 (numeración de aminoácidos de acuerdo con SEC ID N°: 18). Se prefiere que la prolil aminopeptidasa de *Serratia marcescens* o un derivado u homólogo de la misma se use como el armazón.
- En una realización adicional el armazón proteico pertenece a la clase S51 de serina proteasas o tiene al menos un 70 % de identidad en el nivel de aminoácidos con una proteína de la clase S51 de serina proteasas y/o tiene una estructura terciaria similar a aspartil dipeptidasa de *Escherichia coli*. Se prefiere que las SDR se inserten en el armazón proteico en una o más posiciones del grupo de posiciones que corresponden estructuralmente o por homología de secuencia de aminoácidos a las regiones 8-16, 38-46, 85-92, 132-140, 159-170 y 205-211 en aspartil dipeptidasa de *Escherichia coli* y más preferentemente en una o más posiciones del grupo de posiciones que corresponden estructuralmente o por homología de secuencia de aminoácidos a las regiones 10-14, 87-90, 134-138 y 160-165 (numeración de aminoácidos de acuerdo con SEC ID N°: 19). Se prefiere que la aspartil dipeptidasa de *Escherichia coli* o un derivado u homólogo de la misma se use como el armazón.
- En una realización adicional el armazón proteico pertenece a la clase A2 de proteasas aspárticas o tiene al menos un 70 % de identidad en el nivel de aminoácidos con una proteína de la clase A2 de proteasas aspárticas y/o tiene una estructura terciaria similar a la proteasa del virus de la inmunodeficiencia humana. Se prefiere que las SDR se inserten en el armazón proteico en una o más posiciones del grupo de posiciones que corresponden estructuralmente o por homología de secuencia de aminoácidos a las regiones 5-12, 17-23, 27-30, 33-38 y 77-83 en proteasa del virus de la inmunodeficiencia humana y más preferentemente en una o más posiciones del grupo de posiciones que corresponden estructuralmente o por homología de secuencia de aminoácidos a las regiones 7-10, 18-21, 34-37 y 79-82 (numeración de aminoácidos de acuerdo con SEC ID N°: 20). Se prefiere que la proteasa del virus de inmunodeficiencia humana, preferentemente proteasa de VIH-1, o un derivado u homólogo de la misma se use como el armazón.
- En una realización adicional el armazón proteico pertenece a la clase A26 de proteasas aspárticas o tiene al menos un 70 % de identidad en el nivel de aminoácidos con una proteína de la clase A26 de proteasas aspárticas y/o tiene una estructura terciaria similar a la omptina de *Escherichia coli*. Se prefiere que las SDR se inserten en el armazón proteico en una o más posiciones del grupo de posiciones que corresponden estructuralmente o por homología de secuencia de aminoácidos a las regiones 28-40, 86-98, 150-168, 213-219 y 267-278 en omptina de *Escherichia coli* y más preferentemente en una o más posiciones del grupo de posiciones que corresponden estructuralmente o por

homología de secuencia de aminoácidos a las regiones 33-38, 161-168 y 273-277 (numeración de aminoácidos de acuerdo con SEC ID N°: 21). Se prefiere que la omptina de *Escherichia coli* o un derivado u homólogo de la misma se use como el armazón.

5 En una realización adicional el armazón proteico pertenece a la clase C1 de cisteína proteasas o tiene al menos un 70 % de identidad en el nivel de aminoácidos con una proteína de la clase C1 de cisteína proteasas y/o tiene una estructura terciaria similar a la papaína de *Carica papaya*. Se prefiere que las SDR se inserten en el armazón proteico en una o más posiciones del grupo de posiciones que corresponden estructuralmente o por homología de secuencia de aminoácidos a las regiones 17-24, 61-68, 88-95, 135-142, 153-158 y 176-184 en papaína de *Carica papaya* y más preferentemente en una o más posiciones del grupo de posiciones que corresponden estructuralmente o por homología de secuencia de aminoácidos a las regiones 63-66, 139-136 y 177-181 (numeración de aminoácidos de acuerdo con SEC ID N°: 22). Se prefiere que la papaína de *Carica papaya* o un derivado u homólogo de la misma se use como el armazón.

15 En una realización adicional el armazón proteico pertenece a la clase C2 de cisteína proteasas o tiene al menos un 70 % de identidad en el nivel de aminoácidos con una proteína de la clase C2 de cisteína proteasas y/o tiene una estructura terciaria similar a calpaína-2 humana. Se prefiere que las SDR se inserten en el armazón proteico en una o más posiciones del grupo de posiciones que corresponden estructuralmente o por homología de secuencia de aminoácidos a las regiones 90-103, 160-172, 193-199, 243-260, 286-294 y 316-322 en calpaína-2 humana y más preferentemente en una o más posiciones del grupo de posiciones que corresponden estructuralmente o por homología de secuencia de aminoácidos a las regiones 92-101, 245-250 y 287-291 (numeración de aminoácidos de acuerdo con SEC ID N°: 23). Se prefiere que la calpaína-2 humana o un derivado u homólogo de la misma se use como el armazón.

25 En una realización adicional el armazón proteico pertenece a la clase C4 de cisteína proteasas o tiene al menos un 70 % de identidad en el nivel de aminoácidos con una proteína de la clase C4 de cisteína proteasas y/o tiene una estructura terciaria similar a proteasa NIa del virus del grabado del tabaco. Se prefiere que las SDR se inserten en el armazón proteico en una o más posiciones del grupo de posiciones que corresponden estructuralmente o por homología de secuencia de aminoácidos a las regiones 23-31, 112-120, 144-150, 168-176 y 205-218 en NIa proteasa del virus del grabado del tabaco y más preferentemente en una o más posiciones del grupo de posiciones que corresponden estructuralmente o por homología de secuencia de aminoácidos a las regiones 145-149, 169-174 y 212-218 (numeración de aminoácidos de acuerdo con SEC ID N°: 24). Se prefiere que la NIa proteasa del virus del grabado del tabaco (proteasa de VGT) o un derivado u homólogo de la misma se use como el armazón.

35 En una realización adicional el armazón proteico pertenece a la clase C10 de cisteína proteasas o tiene al menos un 70 % de identidad en el nivel de aminoácidos con una proteína de la clase C10 de cisteína proteasas y/o tiene una estructura terciaria similar a la estreptopaína de *Streptococcus pyogenes*. Se prefiere que las SDR se inserten en el armazón proteico en una o más posiciones del grupo de posiciones que corresponden estructuralmente o por homología de secuencia de aminoácidos a las regiones 81-90, 133-140, 150-164, 191-199, 219-229, 246-256, 306-312 y 330-337 en estreptopaína de *Streptococcus pyogenes* y más preferentemente en una o más posiciones del grupo de posiciones que corresponden estructuralmente o por homología de secuencia de aminoácidos a las regiones 82-87, 134-138, 250-254 y 331-335 (numeración de aminoácidos de acuerdo con SEC ID N°: 25). Se prefiere que la estreptopaína de *Streptococcus pyogenes* o un derivado u homólogo de la misma se use como el armazón.

45 En una realización adicional el armazón proteico pertenece a la clase C19 de cisteína proteasas o tiene al menos un 70 % de identidad en el nivel de aminoácidos con una proteína de la clase C19 de cisteína proteasas y/o tiene una estructura terciaria similar a la proteasa específica de ubiquitina humana 7. Se prefiere que las SDR se inserten en el armazón proteico en una o más posiciones del grupo de posiciones que corresponden estructuralmente o por homología de secuencia de aminoácidos a las regiones 3-15, 63-70, 80-86, 248-256, 272-283 y 292-304 en proteasa específica de ubiquitina humana 7 y más preferentemente en una o más posiciones del grupo de posiciones que corresponden estructuralmente o por homología de secuencia de aminoácidos a las regiones 10-15, 251-255, 277-281 y 298-304 (numeración de aminoácidos de acuerdo con SEC ID N°: 26). Se prefiere que la proteasa específica de ubiquitina humana 7 o un derivado u homólogo de la misma se use como el armazón.

50 En una realización adicional el armazón proteico pertenece a la clase C47 de cisteína proteasas o tiene al menos un 70 % de identidad en el nivel de aminoácidos con una proteína de la clase C47 de cisteína proteasas y/o tiene una estructura terciaria similar a la estafopaína de *Staphylococcus aureus*. Se prefiere que las SDR se inserten en el armazón proteico en una o más posiciones del grupo de posiciones que corresponden estructuralmente o por homología de secuencia de aminoácidos a las regiones 15-23, 57-66, 108-119, 142-149 y 157-164 en estafopaína de *Staphylococcus aureus* y más preferentemente en una o más posiciones del grupo de posiciones que corresponden estructuralmente o por homología de secuencia de aminoácidos a las regiones 17-22, 111-117, 143-147 y 159-163 (numeración de aminoácidos de acuerdo con SEC ID N° 27). Se prefiere que la estafopaína de *Staphylococcus aureus* o un derivado u homólogo de la misma se use como el armazón.

60 En una realización adicional el armazón proteico pertenece a la clase C48 de cisteína proteasas o tiene al menos un 70 % de identidad en el nivel de aminoácidos con una proteína de la clase C48 de cisteína proteasas y/o tiene una

- 5 estructura terciaria similar a la endopeptidasa Ulp1 de *Saccharomyces cerevisiae*. Se prefiere que las SDR se inserten en el armazón proteico en una o más posiciones del grupo de posiciones que corresponden estructuralmente o por homología de secuencia de aminoácidos a las regiones 40-51, 108-115, 132-141, 173-179 y 605-597 en endopeptidasa Ulp1 de *Saccharomyces cerevisiae* y más preferentemente en una o más posiciones del grupo de posiciones que corresponden estructuralmente o por homología de secuencia de aminoácidos a las regiones 43-49, 110-113, 133-137 y 175-178 (numeración de aminoácidos de acuerdo con SEC ID N° 28). Se prefiere que la endopeptidasa Ulp1 de *Saccharomyces cerevisiae* o un derivado u homólogo de la misma se use como el armazón.
- 10 En una realización adicional el armazón proteico pertenece a la clase C56 de cisteína proteasas o tiene al menos un 70 % de identidad en el nivel de aminoácidos con una proteína de la clase C56 de cisteína proteasas y/o tiene una estructura terciaria similar a la endopeptidasa Pfpl de *Pyrococcus horikoshii*. Se prefiere que las SDR se inserten en el armazón proteico en una o más posiciones del grupo de posiciones que corresponden estructuralmente o por homología de secuencia de aminoácidos a las regiones 8-16, 40-47, 66-73, 118-125 y 147-153 en endopeptidasa Pfpl de *Pyrococcus horikoshii*, y más preferentemente en una o más posiciones del grupo de posiciones que corresponden estructuralmente o por homología de secuencia de aminoácidos a las regiones 9-14, 68-71, 120-123 y 148-151 (numeración de aminoácidos de acuerdo con SEC ID N° 29). Se prefiere que la endopeptidasa Pfpl de *Pyrococcus horikoshii* o un derivado u homólogo de la misma se use como el armazón.
- 15 En una realización adicional el armazón proteico pertenece a la clase M4 de metalo proteasas o tiene al menos un 70 % de identidad en el nivel de aminoácidos con una proteína de la clase M4 de metalo proteasas y/o tiene una estructura terciaria similar a termolisina de *Bacillus thermoproteolyticus*. Se prefiere que las SDR se inserten en el armazón proteico en una o más posiciones del grupo de posiciones que corresponden estructuralmente o por homología de secuencia de aminoácidos a las regiones 106-118, 125-130, 152-160, 197-204, 210-213 y 221-229 en termolisina de *Bacillus thermoproteolyticus*, y más preferentemente en una o más posiciones del grupo de posiciones que corresponden estructuralmente o por homología de secuencia de aminoácidos a las regiones 108-115, 126-129, 199-203 y 223-227 (numeración de aminoácidos de acuerdo con SEC ID N° 30). Se prefiere que la termolisina de *Bacillus thermoproteolyticus*, o un derivado u homólogo de la misma se use como el armazón.
- 20 En una realización adicional el armazón proteico pertenece a la clase M10 de metalo proteasas o tiene al menos un 70 % de identidad en el nivel de aminoácidos con una proteína de la clase M10 de metalo proteasas y/o tiene una estructura terciaria similar a colagenasa humana. Se prefiere que las SDR se inserten en el armazón proteico en una o más posiciones del grupo de posiciones que corresponden estructuralmente o por homología de secuencia de aminoácidos a las regiones 2-7, 68-79, 85-90, 107-111 y 135-141 en colagenasa humana y más preferentemente en una o más posiciones del grupo de posiciones que corresponden estructuralmente o por homología de secuencia de aminoácidos a las regiones 3-67, 1-78 y 136-140 (numeración de aminoácidos de acuerdo con SEC ID N° 31). Se prefiere que la colagenasa humana o un derivado u homólogo de la misma se use como el armazón.
- 30 Se prefiere adicionalmente que las enzimas obtenidas por ingeniería genética tengan actividad glucosidasa. Un armazón proteico especialmente adecuado para esta variante es una glucosilasa o deriva de una glucosilasa. Preferentemente, la estructura terciaria pertenece a una de las siguientes clases estructurales: barril (beta/alfa) 8 de clase GH13, GH7, GH12, GH11, GH10, GH28, GH26 y GH18.
- 35 En una primera realización el armazón proteico pertenece a la clase GH13 de glucosilasas o tiene al menos un 70 % de identidad en el nivel de aminoácidos con una proteína de la clase GH13 de glucosilasas y/o tiene una estructura terciaria similar a alfa amilasa pancreática humana. Se prefiere que las SDR se inserten en el armazón proteico en una o más posiciones del grupo de posiciones que corresponden estructuralmente o por homología de secuencia de aminoácidos a las regiones 50-60, 100-110, 148-167, 235-244, 302-310 y 346-359 en alfa amilasa pancreática humana y más preferentemente en una o más posiciones del grupo de posiciones que corresponden estructuralmente o por homología de secuencia de aminoácidos a las regiones 51-58, 148-155 y 303-309 (numeración de aminoácidos de acuerdo con SEC ID N° 32). Se prefiere que la alfa amilasa pancreática humana o un derivado u homólogo de la misma se use como el armazón.
- 40 En una realización adicional el armazón proteico pertenece a la clase GH7 de glucosilasas o tiene al menos un 70 % de identidad en el nivel de aminoácidos con una proteína de la clase GH7 de glucosilasas y/o tiene una estructura terciaria similar a la celulasa de *Trichoderma reesei*. Se prefiere que las SDR se inserten en el armazón proteico en una o más posiciones del grupo de posiciones que corresponden estructuralmente o por homología de secuencia de aminoácidos a las regiones 47-56, 93-104, 173-182, 215-223, 229-236 y 322-334 en celulasa de *Trichoderma reesei* y más preferentemente en una o más posiciones del grupo de posiciones que corresponden estructuralmente o por homología de secuencia de aminoácidos a las regiones 175-180, 218-222 y 324-332 (numeración de aminoácidos de acuerdo con SEC ID N° 33). Se prefiere que la celulasa de *Trichoderma reesei* o un derivado u homólogo de la misma se use como el armazón.
- 45 En una realización adicional el armazón proteico pertenece a la clase GH12 de glucosilasas o tiene al menos un 70 % de identidad en el nivel de aminoácidos con una proteína de la clase GH12 de glucosilasas y/o tiene una estructura terciaria similar a celulasa de *Aspergillus niger*. Se prefiere que las SDR se inserten en el armazón proteico en una o más posiciones del grupo de posiciones que corresponden estructuralmente o por homología de
- 50
- 55
- 60

secuencia de aminoácidos a las regiones 18-28, 55-60, 106-113, 126-132 y 149-159 en celulasa de *Aspergillus niger*, y más preferentemente en una o más posiciones del grupo de posiciones que corresponden estructuralmente o por homología de secuencia de aminoácidos a las regiones 20-26, 56-59, 108-112 y 151-156 (numeración de aminoácidos de acuerdo con SEC ID N° 34). Se prefiere que la celulasa de *Aspergillus niger* o un derivado u homólogo de la misma se use como el armazón.

En una realización adicional el armazón proteico pertenece a la clase GH11 de glucosilasas o tiene al menos un 70 % de identidad en el nivel de aminoácidos con una proteína de la clase GH11 de glucosilasas y/o tiene una estructura terciaria similar a xilanasas de *Aspergillus niger*. Se prefiere que las SDR se inserten en el armazón proteico en una o más posiciones del grupo de posiciones que corresponden estructuralmente o por homología de secuencia de aminoácidos a las regiones 7-14, 33-39, 88-97, 114-126 y 158-167 en xilanasas de *Aspergillus niger* y más preferentemente en una o más posiciones del grupo de posiciones que corresponden estructuralmente o por homología de secuencia de aminoácidos a las regiones 20-26, 56-59, 108-112 y 151-156 (numeración de aminoácidos de acuerdo con SEC ID N° 35). Se prefiere que la xilanasas de *Aspergillus niger* o un derivado u homólogo de la misma se use como el armazón.

En una realización adicional el armazón proteico pertenece a la clase GH10 de glucosilasas o tiene al menos un 70 % de identidad en el nivel de aminoácidos con una proteína de la clase GH10 de glucosilasas y/o tiene una estructura terciaria similar a xilanasas de *Streptomyces lividans*. Se prefiere que las SDR se inserten en el armazón proteico en una o más posiciones del grupo de posiciones que corresponden estructuralmente o por homología de secuencia de aminoácidos a las regiones 21-29, 42-50, 84-92, 130-136, 206-217 y 269-278 en xilanasas de *Streptomyces lividans* y más preferentemente en una o más posiciones del grupo de posiciones que corresponden estructuralmente o por homología de secuencia de aminoácidos a las regiones 43-49, 86-90, 208-213 y 271-276 (numeración de aminoácidos de acuerdo con SEC ID N° 36). Se prefiere que la xilanasas de *Streptomyces lividans* o un derivado u homólogo de la misma se use como el armazón.

En una realización adicional el armazón proteico pertenece a la clase GH28 de glucosilasas o tiene al menos un 70 % de identidad en el nivel de aminoácidos con una proteína de la clase GH28 de glucosilasas y/o tiene una estructura terciaria similar a pectinasas de *Aspergillus niger*. Se prefiere que las SDR se inserten en el armazón proteico en una o más posiciones del grupo de posiciones que corresponden estructuralmente o por homología de secuencia de aminoácidos a las regiones 82-88, 118-126, 171-178, 228-236, 264-256 y 289-299 en pectinasas de *Aspergillus niger* y más preferentemente en una o más posiciones del grupo de posiciones que corresponden estructuralmente o por homología de secuencia de aminoácidos a las regiones 116-124, 174-178 y 291-296 (numeración de aminoácidos de acuerdo con SEC ID N° 37). Se prefiere que la pectinasa de *Aspergillus niger* o un derivado u homólogo de la misma se use como el armazón.

En una realización adicional el armazón proteico pertenece a la clase GH26 de glucosilasas o tiene al menos un 70 % de identidad en el nivel de aminoácidos con una proteína de la clase GH26 de glucosilasas y/o tiene una estructura terciaria similar a mananasa de *Pseudomonas cellulosa*. Se prefiere que las SDR se inserten en el armazón proteico en una o más posiciones del grupo de posiciones que corresponden estructuralmente o por homología de secuencia de aminoácidos a las regiones 75-83, 113-125, 174-182, 217-224, 247-254, 324-332 y 325-340 en mananasa de *Pseudomonas cellulosa* y más preferentemente en una o más posiciones del grupo de posiciones que corresponden estructuralmente o por homología de secuencia de aminoácidos a las regiones 115-123, 176-180, 291-286 y 328-337 (numeración de aminoácidos de acuerdo con SEC ID N° 38). Se prefiere que la mananasa de *Pseudomonas cellulosa* o un derivado u homólogo de la misma se use como el armazón.

En una realización adicional el armazón proteico pertenece a la clase de barril (beta/alfa) 8 de GH18 de glucosilasas o tiene al menos un 70 % de identidad en el nivel de aminoácidos con una proteína de la clase GH18 de glucosilasas y/o tiene una estructura terciaria similar a quitinasas de *Bacillus circulans*. Se prefiere que las SDR se inserten en el armazón proteico en una o más posiciones del grupo de posiciones que corresponden estructuralmente o por homología de secuencia de aminoácidos a las regiones 21-29, 57-65, 130-136, 176-183, 221-229, 249-257 y 327-337 en quitinasas de *Bacillus circulans* y más preferentemente en una o más posiciones del grupo de posiciones que corresponden estructuralmente o por homología de secuencia de aminoácidos a las regiones 59-63, 178-181, 250-254 y 330-336 (numeración de aminoácidos de acuerdo con SEC ID N° 39). Se prefiere que la quitinasa de *Bacillus circulans* o un derivado u homólogo de la misma se use como el armazón.

Se prefiere adicionalmente que las enzimas obtenidas por ingeniería genética tengan actividad esterhidrolasa. Preferentemente, el armazón proteico para esta variante tiene actividad de lipasa, fosfatasa, fitasa o fosfodiesterasa.

En una primera realización el armazón proteico pertenece a la clase GX de esterases o tiene al menos un 70 % de identidad en el nivel de aminoácidos con una proteína de la clase GX de esterases y/o tiene una estructura terciaria similar a la estructura de la lipasa B de *Candida antarctica*. Preferentemente, el armazón tiene actividad lipasa. Se prefiere que las SDR se inserten en el armazón proteico en una o más posiciones del grupo de posiciones que corresponden estructuralmente o por homología de secuencia de aminoácidos a las regiones 139-148, 188-195, 216-224, 256-266, 272-287 en lipasa B de *Candida antarctica* y más preferentemente en una o más posiciones del grupo de posiciones que corresponden estructuralmente o por homología de secuencia de aminoácidos a las regiones 141-146, 218-222, 259-263 y 275-283 (numeración de aminoácidos de acuerdo con SEC ID N° 40). Se

prefiere que la lipasa B de *Candida antarctica* o un derivado u homólogo de la misma se use como el armazón.

En una realización adicional el armazón proteico pertenece a la clase GX de esterasas o tiene al menos un 70 % de identidad en el nivel de aminoácidos con una proteína de la clase GX de esterasas y/o tiene una estructura terciaria similar a la lipasa pancreática de cobaya. Preferentemente, el armazón tiene actividad lipasa. Se prefiere que las SDR se inserten en el armazón proteico en una o más posiciones del grupo de posiciones que corresponden estructuralmente o por homología de secuencia de aminoácidos a las regiones 78-90, 91-100, 112-120, 179-186, 207-218, 238-247 y 248-260 en lipasa pancreática de cobaya y más preferentemente en una o más posiciones del grupo de posiciones que corresponden estructuralmente o por homología de secuencia de aminoácidos a las regiones 80-87, 114-118, 209-215 y 239-246 (numeración de aminoácidos de acuerdo con SEC ID N° 41). Se prefiere que la lipasa pancreática de cobaya o un derivado u homólogo de la misma se use como el armazón.

En una realización adicional el armazón proteico tiene una estructura terciaria similar a la estructura de la fosfatasa alcalina de *Escherichia coli* o tiene al menos un 70 % de identidad en el nivel de aminoácidos con una proteína que tiene una estructura terciaria similar a la estructura de la fosfatasa alcalina de *Escherichia coli*. Preferentemente, el armazón tiene actividad fosfatasa. Se prefiere que las SDR se inserten en el armazón proteico en una o más posiciones del grupo de posiciones que corresponden estructuralmente o por homología de secuencia de aminoácidos a las regiones 110-122, 187-142, 170-175, 186-193, 280-287 y 425-435 en fosfatasa alcalina de *Escherichia coli* y más preferentemente en una o más posiciones del grupo de posiciones que corresponden estructuralmente o por homología de secuencia de aminoácidos a las regiones 171-174, 187-191, 282-286 y 426-433 (numeración de aminoácidos de acuerdo con SEC ID N° 42). Se prefiere que la fosfatasa alcalina de *Escherichia coli* o un derivado u homólogo de la misma se use como el armazón.

En una realización adicional el armazón proteico tiene una estructura terciaria similar a la estructura de la desoxirribonucleasa pancreática bovina I o tiene al menos un 70 % de identidad en el nivel de aminoácidos con una proteína que tiene una estructura terciaria similar a la estructura de la desoxirribonucleasa pancreática bovina I. Preferentemente, el armazón tiene actividad fosfodiesterasa. Más preferentemente, una nucleasa, y más preferentemente, una endonucleasa no específica o un derivado de la misma se usa como el armazón. Se prefiere que las SDR se inserten en el armazón proteico en una o más posiciones del grupo de posiciones que corresponden estructuralmente o por homología de secuencia de aminoácidos a las regiones 14-21, 41-47, 72-77, 97-111, 135-143, 171-178, 202-209 y 242-251 en desoxirribonucleasa pancreática bovina I y más preferentemente en una o más posiciones del grupo de posiciones que corresponden estructuralmente o por homología de secuencia de aminoácidos a las regiones 16-19, 42-46, 136-141 y 172-176 (numeración de aminoácidos de acuerdo con SEC ID N° 43). Se prefiere que la desoxirribonucleasa pancreática bovina I o desoxirribonucleasa humana I o un derivado u homólogo de la misma se use como el armazón.

Se prefiere adicionalmente que la enzima obtenida por ingeniería genética tenga actividad transferasa. Un armazón proteico particularmente adecuado para esta variante es una glucosil, fosfo o metiltransferasa o un derivado de las mismas. Son armazones proteicos particularmente preferidos para esta variante las glucosiltransferasas o derivan de glucosiltransferasas. La estructura terciaria del armazón proteico puede ser de cualquier tipo. Preferentemente, sin embargo, la estructura terciaria pertenece a una de las siguientes clases estructurales: GH13 y GT1.

En una primera realización el armazón proteico pertenece a la clase GH13 de transferasas o tiene al menos un 70 % de identidad en el nivel de aminoácidos con una proteína de la clase GH13 de transferasas y/o tiene una estructura terciaria similar a la estructura de la ciclomaltodextrin glucanotransferasa de *Bacillus circulans*. Preferentemente, el armazón tiene actividad transferasa y más preferentemente se usa una glucosil transferasa como el armazón. Se prefiere que las SDR se inserten en el armazón proteico en una o más posiciones del grupo de posiciones que corresponden estructuralmente o por homología de secuencia de aminoácidos a las regiones 38-48, 85-94, 142-154, 178-186, 259-266, 331-340 y 367-377 en ciclomaltodextrin glucanotransferasa de *Bacillus circulans* y más preferentemente en una o más posiciones del grupo de posiciones que corresponden estructuralmente o por homología de secuencia de aminoácidos a las regiones 87-92, 180-185, 261-264 y 269-275 (numeración de aminoácidos de acuerdo con SEC ID N° 44). Se prefiere que la ciclomaltodextrin glucanotransferasa de *Bacillus circulans* o un derivado u homólogo de la misma se use como el armazón.

En una realización adicional el armazón proteico pertenece a la clase GT1 de transferasas o tiene al menos un 70 % de identidad en el nivel de aminoácidos con una proteína de la clase GT1 de transferasas y/o tiene una estructura terciaria similar a la estructura de la glucosil transferasa de *Amycolatopsis orientalis* A82846. Preferentemente el armazón tiene actividad transferasa y más preferentemente actividad glucosil transferasa. Se prefiere que las SDR se inserten en el armazón proteico en una o más posiciones del grupo de posiciones que corresponden estructuralmente o por homología de secuencia de aminoácidos a las regiones 58-74, 130-138, 185-193, 228-236 y 314-323 en glucosil transferasa de *Amycolatopsis orientalis* A82846 y más preferentemente en una o más posiciones del grupo de posiciones que corresponden estructuralmente o por homología de secuencia de aminoácidos a las regiones 61-71, 230-234 y 316-321 (numeración de aminoácidos de acuerdo con SEC ID N° 45). Se prefiere que la glucosil transferasa de *Amycolatopsis orientalis* A82846 o un derivado u homólogo de la misma se use como el armazón.

Se prefiere adicionalmente que las enzimas obtenidas por ingeniería genética tengan actividad oxidorreductasa. Un

almazón proteico particularmente adecuado para esta variante es una monooxigenasa, una dioxigenasa o un alcohol deshidrogenasa o un derivado de las mismas. La estructura terciaria del almacén proteico puede ser de cualquier tipo.

5 En una primera realización el almacén proteico tiene una estructura terciaria similar a la estructura de la 2,3-difidroxibifenil dioxigenasa de *Pseudomonas sp.* o tiene al menos un 70 % de identidad en el nivel de aminoácidos con una proteína que tiene una estructura terciaria similar a la estructura de la 2,3-difidroxibifenil dioxigenasa de *Pseudomonas sp.* Preferentemente, el almacén tiene actividad dioxigenasa. Se prefiere que las SDR se inserten en el almacén proteico en una o más posiciones del grupo de posiciones que corresponden estructuralmente o por homología de secuencia de aminoácidos a las regiones 172-185, 198-206, 231-237, 250-259 y 282-287 en 2,3-difidroxibifenil dioxigenasa de *Pseudomonas sp.* y más preferentemente en una o más posiciones del grupo de posiciones que corresponden estructuralmente o por homología de secuencia de aminoácidos a las regiones 175-182, 200-204, 252-257 y 284-287 (numeración de aminoácidos de acuerdo con SEC ID N° 46). Se prefiere que la 2,3-difidroxibifenil dioxigenasa de *Pseudomonas sp.* o un derivado u homólogo de la misma se use como el almacén.

15 En una realización adicional el almacén proteico tiene una estructura terciaria similar a la estructura de la catecol dioxigenasa de *Acinetobacter sp.* o tiene al menos un 70 % de identidad en el nivel de aminoácidos con una proteína que tiene una estructura terciaria similar a la estructura de la catecol dioxigenasa de *Acinetobacter sp.* Preferentemente, el almacén tiene actividad dioxigenasa y más preferentemente actividad catecol dioxigenasa. Se prefiere que las SDR se inserten en el almacén proteico en una o más posiciones del grupo de posiciones que corresponden estructuralmente o por homología de secuencia de aminoácidos a las regiones 66-72, 105-112, 156-171 y 198-207 en catecol dioxigenasa de *Acinetobacter sp.* y más preferentemente en una o más posiciones del grupo de posiciones que corresponden estructuralmente o por homología de secuencia de aminoácidos a las regiones 107-110, 161-171 y 201-205 (numeración de aminoácidos de acuerdo con SEC ID N° 47). Se prefiere que la catecol dioxigenasa de *Acinetobacter sp.* o un derivado u homólogo de la misma se use como el almacén.

25 En una realización adicional el almacén proteico tiene una estructura terciaria similar a la estructura de la canfor-5-monooxigenasa de *Pseudomonas putida* o tiene al menos un 70 % de identidad en el nivel de aminoácidos con una proteína que tiene una estructura terciaria similar a la estructura de la canfor-5-monooxigenasa de *Pseudomonas putida*. Preferentemente, el almacén tiene actividad monooxigenasa y más preferentemente actividad canfor monooxigenasa. Se prefiere que las SDR se inserten en el almacén proteico en una o más posiciones del grupo de posiciones que corresponden estructuralmente o por homología de secuencia de aminoácidos a las regiones 26-31, 57-63, 84-98, 182-191, 242-256, 292-299 y 392-399 en canfor-5-monooxigenasa de *Pseudomonas putida* y más preferentemente en una o más posiciones del grupo de posiciones que corresponden estructuralmente o por homología de secuencia de aminoácidos a las regiones 85-96, 183-188, 244-253, 293-298 y 393-398 (numeración de aminoácidos de acuerdo con SEC ID N° 48). Se prefiere que la canfor-5-monooxigenasa de *Pseudomonas putida* o un derivado u homólogo de la misma se use como el almacén.

35 En una realización adicional el almacén proteico tiene una estructura terciaria similar a la estructura de la alcohol deshidrogenasa de *Equus callabus* o tiene al menos un 70 % de identidad en el nivel de aminoácidos con una proteína que tiene una estructura terciaria similar a la estructura de la alcohol deshidrogenasa de *Equus callabus*. Preferentemente, el almacén tiene actividad alcohol deshidrogenasa. Se prefiere que las SDR se inserten en el almacén proteico en una o más posiciones del grupo de posiciones que corresponden estructuralmente o por homología de secuencia de aminoácidos a las regiones 49-63, 111-112, 294-301 y 361-369 en alcohol deshidrogenasa de *Equus callabus* y más preferentemente en una o más posiciones del grupo de posiciones que corresponden estructuralmente o por homología de secuencia de aminoácidos a las regiones 51-61 y 295-299 (numeración de aminoácidos de acuerdo con SEC ID N° 49). Se prefiere que la alcohol deshidrogenasa de *Equus callabus* o un derivado u homólogo de la misma se use como el almacén.

45 Se prefiere adicionalmente que las enzimas obtenidas por ingeniería genética tengan actividad liasa. Un almacén proteico particularmente adecuado para esta variante es una liasa oxoácida o es un derivado de la misma. Son almacenes proteicos particularmente preferidos para esta variante aldolasas o sintetetasas o derivan de las mismas. La estructura terciaria del almacén proteico puede ser de cualquier tipo, pero se prefiere una estructura de barril (beta/alfa) 8.

50 En una primera realización el almacén proteico tiene una estructura terciaria similar a la estructura de la ácido N-acetil-d-neurámico aldolasa de *Escherichia coli* o tiene al menos un 70 % de identidad en el nivel de aminoácidos con una proteína que tiene una estructura terciaria similar a la estructura de la ácido N-acetil-d-neurámico aldolasa de *Escherichia coli*. Preferentemente, el almacén tiene actividad aldolasa. Se prefiere que las SDR se inserten en el almacén proteico en una o más posiciones del grupo de posiciones que corresponden estructuralmente o por homología de secuencia de aminoácidos a las regiones 45-55, 78-87, 105-113, 137-146, 164-171, 187-193, 205-210, 244-255 y 269-276 en ácido N-acetil-d-neurámico aldolasa de *Escherichia coli*, y más preferentemente en una o más posiciones del grupo de posiciones que corresponden estructuralmente o por homología de secuencia de aminoácidos a las regiones 45-52, 138-144, 189-192, 247-253 y 271-275 (numeración de aminoácidos de acuerdo con SEC ID N° 50). Se prefiere que la ácido N-acetil-d-neurámico aldolasa de *Escherichia coli* o un derivado u homólogo de la misma se use como el almacén.

60

En una realización adicional el armazón proteico tiene una estructura terciaria similar a la estructura de la triptófano sintasa de *Salmonella typhimurium* o tiene al menos un 70 % de identidad en el nivel de aminoácidos con una proteína que tiene una estructura terciaria similar a la estructura de la triptófano sintasa de *Salmonella typhimurium*. Preferentemente, el armazón tiene actividad sintasa. Se prefiere que las SDR se inserten en el armazón proteico en una o más posiciones del grupo de posiciones que corresponden estructuralmente o por homología de secuencia de aminoácidos a las regiones 56-63, 127-134, 154-161, 175-193, 209-216 y 230-240 en triptófano sintasa de *Salmonella typhimurium* y más preferentemente en una o más posiciones del grupo de posiciones que corresponden estructuralmente o por homología de secuencia de aminoácidos a las regiones 57-62, 155-160, 178-190 y 210-215 (numeración de aminoácidos de acuerdo con SEC ID N° 51). Se prefiere que la triptófano sintasa de *Salmonella typhimurium* o un derivado u homólogo de la misma se use como el armazón.

Se prefiere adicionalmente que las enzimas obtenidas por ingeniería genética tengan actividad isomerasa. Un armazón proteico particularmente adecuado para esta variante es una aldosa de conversión o una cetosa de conversión o es un derivado de las mismas.

En una primera realización, el armazón proteico tiene una estructura terciaria similar a la estructura de la xilosa isomerasa de *Actinoplanes missouriensis* o tiene al menos un 70 % de identidad en el nivel de aminoácidos con una proteína que tiene una estructura terciaria similar a la estructura de la xilosa isomerasa de *Actinoplanes missouriensis*. Se prefiere que las SDR se inserten en el armazón proteico en una o más posiciones del grupo de posiciones que corresponden estructuralmente o por homología de secuencia de aminoácidos a las regiones 18-31, 92-103, 136-147, 178-188 y 250-257 en xilosa isomerasa de *Actinoplanes missouriensis* y más preferentemente en una o más posiciones del grupo de posiciones que corresponden estructuralmente o por homología de secuencia de aminoácidos a las regiones de 20-27, 92-99 y 180-186 (numeración de aminoácidos de acuerdo con SEC ID N° 52). Se prefiere que la xilosa isomerasa de *Actinoplanes missouriensis* o un derivado u homólogo de la misma se use como el armazón.

Se prefiere adicionalmente que las enzimas obtenidas por ingeniería genética tengan actividad ligasa. Un armazón proteico particularmente adecuado para esta variante es una ADN ligasa o es un derivado de la misma.

En una primera realización, el armazón proteico tiene una estructura terciaria similar a la estructura de la ADN ligasa del bacteriófago T7 o tiene al menos un 70 % de identidad en el nivel de aminoácidos con una proteína que tiene una estructura terciaria similar a la estructura de la ADN ligasa de bacteriófago T7. Se prefiere que las SDR se inserten en el armazón proteico en una o más posiciones del grupo de posiciones que corresponden estructuralmente o por homología de secuencia de aminoácidos a las regiones, 52-60, 94-108, 119-131, 241-248, 255-263 y 302-318 en ADN ligasa de bacteriófago T7 y más preferentemente en una o más posiciones del grupo de posiciones que corresponden estructuralmente o por homología de secuencia de aminoácidos a las regiones 96-106, 121-129, 256-262 y 304-316 (numeración de aminoácidos de acuerdo con SEC ID N° 53). SE prefiere que la ADN ligasa de bacteriófago T7 o un derivado u homólogo de la misma se use como el armazón.

Un segundo aspecto se refiere a la aplicación de enzimas obtenidas por ingeniería genética con especificidades para fines terapéuticos, de investigación, diagnósticos, nutricionales, de cuidado personal o industriales. La aplicación comprende al menos las siguientes etapas:

- (a) identificación de un sustrato peptídico diana cuya hidrólisis tiene un efecto positivo en relación con el fin pretendido, tal como curar una enfermedad, diagnosticar una enfermedad, procesamiento de ingredientes para nutrición humana o animal u otros procedimientos técnicos;
- (b) provisión de una enzima obtenida por ingeniería genética, siendo específica la enzima para el péptido diana identificado en la etapa (a); y
- (c) uso de la enzima como se proporciona en la etapa (b) para el fin pretendido.

En una primera variante de este aspecto, la enzima obtenida por ingeniería genética se usa como un medio terapéutico para inactivar un sustrato diana relacionado con la enfermedad. Esta aplicación comprende al menos las siguientes etapas:

- (a) identificación de un sustrato diana cuya función está relacionada con una enfermedad y cuya inactivación tiene un efecto positivo en relación con la enfermedad y determinación de un sitio diana dentro del sustrato diana caracterizado por el hecho de que una modificación en el sitio diana conduce a la inactivación del sustrato diana;
- (b) provisión de una enzima obtenida por ingeniería genética, siendo específica la enzima del sitio diana identificado en la etapa (a); y
- (c) uso de la enzima para la inactivación del sustrato diana dentro o fuera del cuerpo humano.

En una realización preferida el armazón de la enzima obtenida por ingeniería genética proporcionada en la etapa (c) es de origen humano para evitar o reducir la inmunogenicidad o efectos alérgenos asociados con la aplicación de la enzima en el cuerpo humano. En una realización más preferida de esta variante, el armazón es de una proteasa humana y la modificación es hidrólisis de un sitio diana en una diana proteica. Preferentemente, la hidrólisis conduce a la activación o inactivación de la diana peptídica o proteica. Las dianas peptídicas o proteicas potenciales incluyen:

citocinas, factores de crecimiento, hormonas peptídicas, interleucinas, interferones, enzimas de la cascada de coagulación, serpinas, inmunoglobulinas, receptores unidos a membrana o solubles, proteínas de superficie celular o viral, fármacos peptídicos, fármacos proteicos.

5 Una realización particularmente preferida se basa en el hallazgo de que la enzima obtenida por ingeniería genética es capaz de escindir el factor de necrosis tumoral alfa (TNF- α) humano. Las enzimas obtenidas por ingeniería genética o la proteína de fusión puede por lo tanto usarse para preparar medicamentos para el tratamiento de enfermedades inflamatorias (así como otras enfermedades relacionadas con TNF- α). Preferentemente, dicha enzima obtenida por ingeniería genética o dicha proteína de fusión es capaz de inactivar específicamente el factor de necrosis tumoral alfa humano (hTNF- α), más preferentemente dicha enzima obtenida por ingeniería genética o dicha proteína de fusión es capaz de hidrolizar el enlace peptídico entre las posiciones 31/32, 32/33, 44/45, 87/88, 128/129 y/o 141/142 (más preferentemente entre las posiciones 31/32 y 32/33) en hTNF- α (SEC ID N° 96).

15 En realizaciones adicionales, el sustrato diana es un profármaco que se activa por la enzima obtenida por ingeniería genética. En una realización particular de esta variante, la enzima obtenida por ingeniería genética tiene actividad proteolítica y el sustrato diana es una diana proteica que se activa de forma proteolítica. Son ejemplos de tales profármacos pro-proteínas tales como las formas inactivadas de factores de coagulación. En otra variante particular, la enzima obtenida por ingeniería genética es una oxidorreductasa y el sustrato diana es un agente químico que puede activarse por oxidación.

20 En una segunda variante de este aspecto, la enzima obtenida por ingeniería genética se usa como un medio técnico para catalizar una reacción industrial o nutricionalmente relevante con especificidad definida. En una realización particular de esta variante la enzima obtenida por ingeniería genética tiene actividad proteolítica, la reacción catalizada es un procesamiento proteolítico y la enzima obtenida por ingeniería genética hidroliza específicamente uno o más sustratos proteicos industrial o nutricionalmente relevantes. En una realización preferida de esta variante la enzima obtenida por ingeniería genética hidroliza uno o más sustratos proteicos industrial o nutricionalmente relevantes en sitios específicos, conduciendo de este modo a propiedades de producto industrial o nutricionalmente deseadas tales como textura, sabor o características de precipitación. En una realización particular adicional de esta variante, la enzima obtenida por ingeniería genética cataliza la hidrólisis de enlaces glucosídicos (actividad glicosidasa o glucosilasa). Después, preferentemente, la reacción catalizada es un procesamiento de polisacáridos y la enzima obtenida por ingeniería genética hidroliza específicamente uno o más sustratos polisacáridos industrial, técnica o nutricionalmente relevantes. En una realización particular adicional de esta variante, la enzima obtenida por ingeniería genética cataliza la hidrólisis de ésteres de triglicéridos o lípidos (actividad lipasa).

30 Después, preferentemente, la reacción catalizada es una etapa de procesamiento de lípidos y la enzima obtenida por ingeniería genética hidroliza específicamente uno o más sustratos lipídicos industrial, técnica o nutricionalmente relevantes. En una variante particular adicional de esta realización, la enzima obtenida por ingeniería genética cataliza la oxidación o reducción de sustratos (actividad oxidorreductasa). Después, preferentemente, la enzima obtenida por ingeniería genética oxida o reduce específicamente uno o más sustratos químicos industrial, técnica o nutricionalmente relevantes.

35 Un tercer aspecto se refiere a un procedimiento para generar enzimas obtenida por ingeniería genética con especificidades que son cualitativa y/o cuantitativamente nuevas en combinación con el armazón proteico. El procedimiento comprende al menos las siguientes etapas:

- 40 (a) proporcionar un armazón proteico capaz de catalizar al menos una reacción química en al menos un sustrato diana.
 (b) generar una biblioteca de enzimas obtenidas por ingeniería genética o enzimas obtenidas por ingeniería genética aisladas combinando el armazón proteico de la etapa (a) con una o más secuencias peptídicas completa o parcialmente aleatorias en sitios en el armazón proteico que permiten que la enzima obtenida por ingeniería genética resultante diferencie entre al menos un sustrato diana y uno o más sustratos diferentes y
 45 (c) seleccionar de la biblioteca de enzimas obtenidas por ingeniería genética generadas en la etapa (b) una o más enzimas que tienen especificidades definidas para al menos un sustrato diana.

En una primera variante de este aspecto, el procedimiento comprende al menos las siguientes etapas:

- 50 (a) proporcionar un armazón proteico capaz de catalizar al menos una reacción química en al menos un sustrato diana,
 (b) generar una biblioteca de enzimas obtenidas por ingeniería genética o enzimas obtenidas por ingeniería genética aisladas insertando en el armazón proteico de la etapa (a) una o más secuencias peptídicas completa o parcialmente aleatorias en sitios del armazón proteico que permiten que la enzima obtenida por ingeniería genética resultante diferencie entre al menos un sustrato diana y uno o más sustratos diferentes y
 55 (c) seleccionar de la biblioteca de enzimas obtenidas por ingeniería genética generadas en la etapa (b) una o más enzimas que tienen especificidades definidas para al menos un sustrato diana.

Preferentemente, las posiciones en las que la o las secuencias peptídicas completa o parcialmente aleatorias se combinan con o se insertan en el armazón proteico se identifican antes de la combinación o inserción.

5 El número de inserciones u otras combinaciones de secuencias peptídicas completa o parcialmente aleatorias así como su longitud puede variar a lo largo de un intervalo amplio El número es al menos uno, preferentemente más de uno, más preferentemente entre dos y once, más preferentemente entre dos y seis. La longitud de tales secuencias peptídicas completa o parcialmente aleatorias es habitualmente menos de 50 restos aminoacídicos. Preferentemente, la longitud está entre uno y 15 restos aminoacídicos, más preferentemente entre uno y seis restos aminoacídicos. Como alternativa, la longitud está entre dos y 20 restos aminoacídicos, preferentemente entre dos y diez restos aminoacídicos, más preferentemente entre tres y ocho restos aminoacídicos.

10 Preferentemente tales inserciones u otras combinaciones se realizan en el nivel de ADN, usando polinucleótidos que codifican tales armazones proteicos y polinucleótidos u oligonucleótidos que codifican tales secuencias peptídicas completa o parcialmente aleatorias.

15 Opcionalmente, las etapas (a) a (c) se repiten cíclicamente, por lo que las enzimas seleccionadas en la etapa (c) actúan como el armazón proteico en la etapa (a) de un ciclo adicional y se insertan secuencias peptídicas seleccionadas de forma aleatoria o, como alternativa, sustituyen a secuencias peptídicas que se han insertado en ciclos anteriores. De este modo, el número de secuencias peptídicas insertadas es constante o aumenta a lo largo de los ciclos. Los ciclos se repiten hasta que se generan una o más enzimas con las especificidades pretendidas.

20 Además, durante o después de uno o más ciclos de las etapas (a) a (c), el armazón puede mutarse en una o más posiciones para hacer el armazón más aceptable para la combinación con secuencias SDR, para aumentar la actividad catalítica a un pH y temperatura específicos, para cambiar el patrón de glucosilación, para reducir la sensibilidad hacia inhibidores de enzima y/o para cambiar la estabilidad enzimática.

En una segunda variante de este aspecto, el procedimiento comprende al menos las siguientes etapas:

- 25 (a) proporcionar un primer fragmento de armazón proteico,
- (b) conectar dicho fragmento de armazón proteico mediante un engarce peptídico con una primera SDR, opcionalmente
- (c) conectar el producto de la etapa (b) mediante un engarce peptídico con un péptido SDR adicional o con un fragmento de armazón proteico adicional y opcionalmente
- (d) repetir la etapa (c) durante tantos ciclos como sea necesario para generar una enzima suficientemente específica, y
- 30 (e) seleccionar de la población generada en las etapas (a) - (d) una o más enzimas que tengan las especificidades deseadas para el o los sustratos diana.

El fragmento de armazón proteico significa una parte de la secuencia de un armazón proteico. Un armazón proteico está comprendido por al menos dos fragmentos de armazón proteico.

35 En una tercera variante de este aspecto, el armazón proteico, las SDR y la enzima obtenida por ingeniería genética se codifican por una secuencia de ADN y se usa un sistema de expresión para producir la proteína. En una variante alternativa, el armazón proteico, las SDR y/o la enzima obtenida por ingeniería genética se sintetizan químicamente a partir de componentes básicos peptídicos.

En una cuarta variante de este aspecto, el procedimiento comprende al menos las siguientes etapas:

- 40 (a) proporcionar un polinucleótido que codifica un armazón proteico capaz de catalizar una o más reacciones químicas en uno o más sustratos diana;
- (b) combinar una o más secuencias oligonucleotídicas parcial o completamente aleatorias con el polinucleótido que codifica el armazón proteico, localizándose las secuencias oligonucleotídicas parcial o completamente aleatorias en sitios en el polinucleótido que permiten que la enzima obtenida por ingeniería genética codificada diferencie entre el o los sustratos diana y uno o más sustratos diferentes; y
- 45 (c) seleccionar de la población generada en la etapa (b) uno o más polinucleótidos que codifican enzimas que tienen las especificidades definidas para el o los sustratos diana.

50 Cualquier enzima puede actuar como el armazón proteico en la etapa (a). Puede ser una enzima de origen natural, una variante o un derivado truncado de las mismas o una enzima obtenida por ingeniería genética. Para uso terapéutico humano, el armazón proteico es preferentemente una enzima de mamífero y más preferentemente una enzima humana. En ese aspecto, se refiere a un procedimiento para la generación de enzimas esencialmente de mamífero, especialmente esencialmente humanas con especificidades que son diferentes de las especificidades de cualquier enzima codificada en genomas de mamífero o en el genoma humano, respectivamente.

55 El armazón proteico proporcionado en la etapa (a) de este aspecto requiere ser capaz de catalizar una o más reacciones químicas en un sustrato diana. Por lo tanto, se selecciona un armazón proteico del grupo de armazones proteicos potenciales por su actividad en el sustrato diana.

En una variante preferida de este aspecto, se usa un armazón proteico con actividad hidrolasa. Preferentemente, se usa un armazón proteico con actividad proteolítica y más preferentemente se usa una proteasa con especificidad muy baja que tiene actividad básica en el sustrato diana como el armazón proteico. Los ejemplos de proteasas de diferentes clases estructurales con especificidad de sustrato baja son papaína, tripsina, quimiotripsina, subtilisina, SET (serina proteasa de tipo tripsina de *Streptomyces erythraeus*), elastasa, catepsina G o quimasa. Antes de emplearse como el armazón proteico, la secuencia de aminoácidos de la proteasa puede modificarse para cambiar las propiedades proteicas distintas de especificidad, por ejemplo actividad catalítica, estabilidad, sensibilidad al inhibidor o rendimiento de la expresión, esencialmente como se describe en el documento WO 92/18645 o para cambiar la especificidad, esencialmente como se describe en el documento EP 02020576.3 y el documento PCT/EP03/04864.

Otra opción para un armazón proteico factible son las lipasas. La lipasa hepática, lipasa lipoproteica y lipasa pancreática pertenecen a la "superfamilia de lipasa lipoproteica", que a su vez es un ejemplo de la clase GX de lipasas (M. Fischer, J. Pleiss (2003), Nucl. Acid. Res., 31, 319-321). La especificidad de sustrato de las lipasas puede caracterizarse por su actividad relativa hacia ésteres de triglicerol de ácidos grasos y fosfolípidos, que portan un grupo principal cargado. Como alternativa, otras hidrolasas tales como esterasas, glucosilasas, amidasas o nitrilasas pueden usarse como armazones.

Las transferasas también son armazones proteicos factibles. Las glucosiltransferasas están implicadas en muchas síntesis biológicas que implican una diversidad de donadores y aceptores.

Como alternativa, el armazón proteico puede tener actividad ligasa, liasa, oxidorreductasa o isomerasa.

En una primera realización, la o las secuencias peptídicas completa o parcialmente aleatorias se insertan en sitios específicos en el armazón proteico. Estos sitios de inserción se caracterizan por el hecho de que las secuencias peptídicas insertadas pueden actuar como diferenciadores entre diferentes sustratos, es decir como regiones determinantes de especificidad o SDR. Tales sitios de inserción pueden identificarse por varios enfoques. Preferentemente, los sitios de inserción se identifican por análisis de la estructura tridimensional de los armazones proteicos, por análisis comparativo de las secuencias primarias del armazón proteico con otras enzimas que tienen diferentes especificidades cuantitativas o experimentalmente por técnicas tales como cribado de alanina, mutagénesis aleatoria o delección aleatoria o por cualquier combinación de las mismas.

Un primer enfoque para identificar sitios de inserción para SDR se basa en la estructura tridimensional del armazón proteico como puede obtenerse por cristalografía de rayos x o por estudios de resonancia magnética nuclear. El alineamiento estructural del armazón proteico en comparación con otras enzimas de la misma clase estructural pero que tiene diferentes especificidades cuantitativas revela regiones de alta similitud estructural y regiones con baja similitud estructural. Un análisis tal puede por ejemplo realizarse usando software público tal como Swiss PDB (Guex, N. y Peitsch, M.C. (1997) Electrophoresis 18, 2714-2723). Las regiones de baja similitud estructural son sitios de inserción de SDR preferidos.

En un segundo enfoque para identificar sitios de inserción para SDR, se analizan las estructuras tridimensionales de la proteína del armazón en complejo con inhibidores competitivos o análogos de sustrato. Se asume que el sitio de unión de un inhibidor competitivo solapa significativamente con el sitio de unión del sustrato. En ese caso, es probable que los átomos de la proteína que están dentro de una cierta distancia de átomos del inhibidor estén a una distancia similar del sustrato también. La selección de una distancia corta, por ejemplo, $< 5 \text{ \AA}$, dará como resultado un conjunto de átomos de proteínas que están en contacto cercano con el sustrato. Estos restos constituirán los primeros contactos de superficie y son por lo tanto sitios de inserción preferidos para SDR. Una vez que se han identificado los primeros contactos de superficie, pueden hallarse segundos contactos de superficie repitiendo el análisis de distancia comenzando desde los primeros átomos de superficie. En otra alternativa más, el análisis de distancia descrito anteriormente se realiza empezando desde los restos del sitio activo.

En un tercer enfoque para identificar sitios de inserción para SDR, la secuencia primaria de la proteína del armazón se alinea con otras enzimas de la misma clase estructural pero que tienen diferentes especificidades cuantitativas usando un algoritmo de alineamiento. Están publicados ejemplos de tales algoritmos de alineamiento (Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. (1990) J. Mol Biol. 215:403-410; "Statistical methods in Bioinformatics: an introduction" por Ewens, W. & Grant, G.R. 2001, Springer, Nueva York). Un alineamiento tal puede revelar regiones conservadas y no conservadas con diversa homología de secuencia y, en particular, elementos de secuencia adicionales en una o más enzimas en comparación con la proteína del armazón. Es más probable que las regiones conservadas contribuyan a fenotipos compartidos entre las diferentes proteínas, por ejemplo, estabilización del pliegue tridimensional. Las regiones no conservadas y, en particular, secuencias adicionales en enzimas con especificidad cuantitativamente mayor (Turner, R. y col (2002) J. Biol. Chem., 277, 33068-33074) son sitios de inserción preferidos para las SDR.

Para proteasas se conocen actualmente cinco familias, concretamente proteasas aspárticas, cisteína proteasas, serina proteasas, metalo proteasas y treonina proteasas. Cada familia incluye grupos de proteasas que comparten un pliegue similar. Las estructuras cristalográficas de miembros de estos grupos se han resuelto y son accesibles a través de bases de datos públicas, por ejemplo, la base de datos de proteínas Brookhaven (H. M. Berman y col,

Nucleic Acids Research, 28 pág. 235-242 (2000)). Tales bases de datos también incluyen homólogos estructurales en otras clases de enzima y proteínas no activas enzimáticamente de cada clase. Están disponibles varias herramientas para buscar en bases de datos públicas homólogos estructurales: SCOP, una clasificación estructural de base de datos de proteínas para la investigación de secuencias y estructuras. (Murzin A. G. y col (1995) J. Mol. Biol. 247, 536-540); CATH, superfamilia de clase, arquitectura, topología y homóloga: una clasificación jerárquica de estructuras de dominio proteico (Orengo y col (1997) Structure 5(8) 1093-1108); FSSP, clasificación de pliegues basada en el alineamiento de estructura-estructura de proteínas (Holm y Sander (1998). Nucl. Acids Res. 26 316-319); o VAST, herramienta de búsqueda de alineamiento de vector (Gibrat, Madej y Bryant (1996) Current Opinión in Structural Biology 6, 377-385).

5 En los enfoques anteriormente descritos, se comparan miembros de clases estructurales para identificar sitios de inserción para SDR.

En una variante preferida de estos enfoques se comparan serina proteasas de la clase estructural S1 entre sí. La tripsina representa un miembro con baja especificidad de sustrato, puesto que requiere solamente un resto de arginina o lisina en la posición P₁. Por otro lado, la trombina, el activador de plasminógeno de tipo tisular o la enteroquinasa tienen todos una alta especificidad para sus secuencias sustrato, es decir, (L/I/V/F)XPR^{NA}, CPGR^{WGG} y DDDK^A, respectivamente (Perona, J. & Craik, C. (1997) J. Biol. Chem., 272, 29987-29990; Perona, J. & Craik, C (1995) Protein Science, 4, 337-360). Se describe un alineamiento de las secuencias de aminoácidos de estas proteasas en el ejemplo 1 (Figura 2) junto con la identificación de SDR.

Un ejemplo adicional dentro de la familia de serina proteasas lo constituyen los miembros de la clase estructural S8 (pliegue de subtilisina). La subtilisina es la proteasa tipo para esta clase y representa una proteasa no específica (Ottesen, M. & Svendsen, A. (1998) Methods Enzymol. 19, 199-215). Furina, PC1 y PC5 son proteasas de la misma clase estructural implicadas en el procesamiento de pro péptidos y tienen una alta especificidad de sustrato (Seidah, N. & Chretien, M. (1997) Curr. Opin. Biotech., 8:602-607; Bergeron, F. y col (2000) J. Mol Endocrin., 24:1-22). En una variante preferida del enfoque se usan alineamientos de las secuencias de aminoácidos primarias (Figura 4) para identificar once tramos de secuencia mayores de tres aminoácidos que las proteasas específicas tienen además en comparación con subtilisina y son por lo tanto regiones determinantes de especificidad potenciales. En una variante adicional del enfoque puede usarse información de la estructura tridimensional de subtilisina para restringir adicionalmente la selección (Figura 3). De los once tramos de secuencia insertados, tres están especialmente cerca de los restos del sitio activo, concretamente el número de tramo 7, 8 y 11 que son inserciones en PC5, PC1 y las tres proteasas específicas, respectivamente (Figura 3). En una variante preferida, pueden insertarse uno o varios tramos de aminoácidos de longitud y composición variable en la secuencia de subtilisina en una o varias de las once posiciones. En una variante más preferida del enfoque la inserción se realiza en las regiones 7, 8 u 11 o cualquier combinación de las mismas. En otra variante preferida del enfoque se usan armazones de proteasa distintos de subtilisina de la clase estructural S8.

En una variante preferida adicional de este enfoque, se analizan proteasas de ácido aspártico de la clase estructural A1 (Rawlings, N.D. & Barrett, A.J. (1995). Methods Enzymol. 248, 105-120; Chitpintyol, S. & Crabbe, MJ. (1998), Food Chemistry, 61, 395-418). Son ejemplos de la clase estructural A1 de proteasas aspárticas pepsina con baja especificidad así como beta-secretasa (Gruninger-Leitch, F y col (2002) J. Biol. Chem. 277, 4687-4693) y renina (Wang, W. & Liang, TC. (1994) Biochemistry, 33, 14636-14641) con especificidades de sustrato relativamente altas. Las proteasas retrovirales también pertenecen a esta clase, aunque la enzima activa es un dímero de dos subunidades idénticas. Las proteasas virales son esenciales para el correcto procesamiento del precursor poliproteico para generar proteínas funcionales que requiere una alta especificidad de sustrato en cada caso (Wu, J. y col (1998) Biochemistry, 37, 4518-4526; Pettit, S. y col (1991) J. Biol. Chem., 266, 14539-14547). La pepsina es la proteasa tipo para esta clase y representa una proteasa no específica (Kageyama, T. (2002) Cell. Mol. Life Sci. 59, 288-306). B-secretasa y catepsina D (Aguilar, C. F. y col (1995) Adv. Exp. Med. Biol., 362, 155-166) son proteasas de la misma clase estructural y tienen una alta especificidad de sustrato. En una variante preferida del enfoque se usan alineamientos de las secuencias de aminoácidos primarias (Figura 6) para identificar seis tramos de secuencia mayores de tres aminoácidos que se insertan en las proteasas específicas en comparación con pepsina y son por lo tanto regiones determinantes de especificidad potenciales. En una variante adicional del enfoque puede usarse información de la estructura tridimensional de b-secretasa para restringir la selección. De los seis tramos de secuencia insertados, tres están especialmente cerca de los restos del sitio activo, concretamente los números de tramo 1, 3 y 4 que son inserciones en catepsina D y beta-secretasa, respectivamente (Figura 5). En una variante preferida del enfoque, pueden insertarse uno o varios tramos de aminoácidos de longitud y composición variable en la secuencia de pepsina en una o más de las seis posiciones. En una realización más preferida la inserción se realiza en las posiciones 1, 3 ó 4 o cualquier combinación de las mismas. En otra realización preferida se usan armazones de proteasa distintos de pepsina.

Existen casos en los que una cierta clase estructural no incluye miembros conocidos de alta y baja especificidad. Esto se ejemplifica por la clase C14 de caspasas que pertenecen a la familia de cisteína proteasa (Rawlings, N.D. & Barrett, A.J. (1994) Methods Enzymol. 244, 461-486) y que muestran todas alta especificidad para las posiciones P₄ a P₁. Por ejemplo, caspasa-1, caspasa-3 y caspasa-9 reconocen las secuencias YVAD^A, DEVD^A o LEHD^A, respectivamente. La identificación de las regiones que difieren entre las caspasas incluirán las regiones responsables de las diferencias en especificidad de sustrato (Figuras 7 y 8).

Finalmente, las proteínas no enzimáticas del mismo pliegue que el armazón enzimático también pueden contribuir a la identificación de sitios de inserción para SDR. Por ejemplo, haptoglobina (Arcoletto, J. & Greer, J.; (1982) J. Biol. Chem. 257, 10063-10068) y azurocidina (Almeida, R. y col (1991) Biochem. Biophys. Res Commun. 177, 688-695) comparten el mismo pliegue de tipo quimiotripsina con todas las proteasas S1. Debido a sustituciones en los restos del sitio activo estas proteínas no poseen ninguna función proteolítica, pero muestran alta homología con proteasas activas. Las diferencias entre estas proteínas y las proteasas específicas incluyen regiones que pueden actuar como sitios de inserción para SDR.

En un cuarto enfoque, se identifican sitios de inserción para SDR experimentalmente por técnicas tales como cribado de alanina, mutagénesis aleatoria, inserción aleatoria o delección aleatoria. A diferencia del enfoque desvelado anteriormente, este enfoque no requiere conocimiento detallado acerca de la estructura tridimensional del armazón proteico. En una variante preferida de este enfoque, puede usarse mutagénesis aleatoria de enzimas con especificidad relativamente alta de la misma clase estructural que el armazón proteico y cribado con respecto a pérdida de cambio de especificidad para identificar sitios de inserción para SDR en el armazón proteico.

La mutagénesis aleatoria, cribado de alanina, inserción aleatoria o delección aleatoria se realizan todas en el nivel de los polinucleótidos que codifica las enzimas. Existe una diversidad de protocolos conocidos en la bibliografía (por ejemplo, Sambrook, J. F. Fritsch, E.F.; Maniatis, T.; Cold Spring Harbor Laboratory Press, segunda edición, 1989, Nueva York). Por ejemplo, puede conseguirse mutagénesis aleatoria mediante el uso de una polimerasa como se describe en la patente WO 9218645. De acuerdo con esta patente, el o los genes que codifican la o las proteasas se amplifican por el uso de una ADN polimerasa con una alta tasa de error o en condiciones que aumentan la tasa de incorporaciones erróneas. Por ejemplo puede emplearse el procedimiento de Cadwell y Joyce (Cadwell, R.C. y Joyce, G.F., PCR methods. Apl. 2 (1992) 28-33). También pueden emplearse otros procedimientos de mutagénesis aleatoria tales como, pero sin limitación, el uso de cepas mutantes, mutágenos químicos o radiación UV. Como alternativa, pueden usarse oligonucleótidos para mutagénesis que sustituyen restos aminoácidos distribuidos aleatoriamente con una alanina. Este procedimiento se denomina generalmente mutagénesis de cribado de alanina (Fersht, A.R. Biochemistry (1989) 8031-8036). Como una alternativa adicional, pueden emplearse modificaciones de la mutagénesis de cribado de alanina tales como mutagénesis binominal (Gregoret, L.M. y Sauer, R.T. PNAS (1993) 4246-4250) o cribado de alanina combinatoria (Weiss y col, PNAS (2000) 8950-8954).

Para expresar enzimas obtenidas por ingeniería genética, el ADN que codifica tales proteínas obtenidas por ingeniería genética se liga en un vector de expresión adecuado por técnicas de clonación molecular convencionales (por ejemplo, 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 enzimática obtenida por ingeniería genética correspondiente. Son huéspedes de expresión particularmente adecuados huéspedes de expresión bacterianos tales como *Escherichia coli* o *Bacillus subtilis*, huéspedes de expresión de levaduras tales como *Saccharomyces cerevisiae* o *Pichia pastoris*, huéspedes de expresión de mamíferos tales como las líneas celulares de ovario de hámster chino (CHO) o riñón de cría de hámster (BHK), sistemas de expresión virales tales como bacteriófagos como M13 o Lambda o virus tales como sistema de expresión de baculovirus. Como una alternativa adicional, pueden usarse sistemas para expresión de proteínas *in vitro*. Normalmente, el ADN se liga en un vector de expresión detrás de una secuencia señal adecuada que conduce a secreción de las variantes de enzima al espacio extracelular, permitiendo de este modo la detección directa de la actividad proteasa en el sobrenadante celular. Son secuencias señal particularmente adecuadas para *Escherichia coli* HlyA, para *Bacillus subtilis* AprE, NprB, Mpr, AmyA, AmyE, Blac, SacB y para *S. cerevisiae* Bar1, Suc2, Mat α , Inu1A, Ggplp. Como alternativa, las variantes de enzima se expresan intracelularmente y los sustratos se expresan también intracelularmente. Preferentemente, esto se realiza esencialmente como se ha descrito en la solicitud de patente WO 0212543, usando un sustrato peptídico de fusión que comprende dos proteínas autofluorescentes ligadas por la secuencia de aminoácidos sustrato. Como una alternativa adicional, después de expresión intracelular de las variantes de enzima o secreción al espacio periplásmico usando secuencias señal tales como DsbA, PhoA, PelB, OmpA, OmpT o gIII para *Escherichia coli*, una etapa de permeabilización o lisis libera las variantes de enzima 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. Como otra alternativa adicional, los genes que codifican las variantes de enzima se expresan sin células mediante el uso de un sistema de expresión sin células adecuado. Por ejemplo, el extracto S30 de células de *Escherichia coli* se usa para este fin como se describe en Lesly y col (Methods in Molecular Biology 37 (1995) 265-278).

El conjunto de variantes génicas generado y expresado por cualquiera de los procedimientos anteriores se analiza con respecto a su afinidad, especificidad de sustrato o actividad por ensayo apropiado y procedimientos de cribado como se describe en detalle por ejemplo en la solicitud de patente PCT/EP03/04864. Se analizan genes de variantes catalíticamente activas que tienen especificidad reducida en comparación con la enzima original por secuenciación. Los sitios en los que se producen mutaciones, inserciones y/o delecciones son sitios de inserción preferidos en los que pueden insertarse SDR de forma específica de sitio.

En una segunda realización, la o las secuencias peptídicas parcial o completamente aleatorias se insertan en sitios aleatorios en el armazón proteico. Esta modificación se realiza habitualmente en el nivel de polinucleótidos, es decir insertando secuencias de nucleótidos en el gen que codifica el armazón proteico. Están disponibles varios procedimientos que permiten la inserción aleatoria de secuencias de nucleótidos. Los sistemas que pueden usarse

para inserción aleatoria son por ejemplo sistemas basados en ligación (Murakami y col. Nature Biotechnology 20 (2002) 76-81), sistemas basados en polimerización de ADN y sistemas basados en transposones (por ejemplo, sistema de mutagénesis GPS-MTM, NEB Biolabs; sistema de generación de mutaciones MGSTM, Finnzymes). Los procedimientos basados en transposón emplean una inserción mediada por transposasa de un gen marcado seleccionable que contiene en sus extremos secuencias de reconocimiento para la transposasa así como dos sitios para una endonucleasa de restricción de corte poco habitual. Usando la segunda endonucleasa habitualmente se libera el marcador de selección y después de la religación se obtiene una inserción. En lugar de realizar la religación se puede insertar como alternativa un fragmento que tiene secuencias de reconocimiento terminales para una o dos endonucleasas de restricción de corte exterior así como un marcador seleccionable. Después de la ligación, se libera este fragmento usando la o las dos endonucleasas de corte exterior. Después de crear extremos romos por procedimientos convencionales se insertan fragmentos aleatorios de extremos romos en posiciones aleatorias en el gen.

En una realización referida adicional, se usan procedimientos para recombinación *in vitro* homóloga para combinar las mutaciones introducidas por los procedimientos anteriormente mencionados para generar poblaciones de enzima. Son ejemplos de procedimientos que pueden aplicarse la Reacción en Cadena de Recombinación (RCR) de acuerdo con la Solicitud de Patente WO 0134835, el procedimiento de Barajado de ADN de acuerdo con la Solicitud de Patente WO 9522625, el procedimiento de Extensión Escalonada de acuerdo con la Patente WO 9842728, o la recombinación de Cebadores Aleatorios de acuerdo con la Solicitud de Patente WO9842728. Además, también pueden aplicarse procedimientos para recombinación no homóloga tales como el procedimiento de Itchy (Ostermeier, M. y col. Nature Biotechnology 17 (1999) 1205-1209).

Tras inserción aleatoria de una secuencia de nucleótidos en el armazón proteico se obtiene una biblioteca de diferentes genes que codifican variantes de enzima. La biblioteca polinucleotídica se transfiere posteriormente a un vector de expresión apropiado. Tras la expresión en un huésped adecuado o mediante el uso de un sistema de expresión *in vitro*, se obtiene una biblioteca de enzimas que contienen tramos de aminoácidos insertados de forma aleatoria.

De acuerdo con la etapa (b) de este tercer aspecto, se insertan una o más secuencias peptídicas completa o parcialmente aleatorias en el armazón proteico. El número real de tales SDR insertadas se determina por la especificidad cuantitativa pretendida que sigue la relación: cuanto mayor sea la especificidad pretendida más SDR se insertan. Mientras que una SDR sencilla permite la generación de enzimas moderadamente específicas, dos SDR permiten ya la generación de enzimas significativamente específicas. Sin embargo, pueden insertarse hasta seis y más SDR en un armazón proteico. Una relación similar es válida para la longitud de las SDR: cuanto mayor sea la especificidad pretendida, más largas serán las SDR que deban insertarse. Las SDR pueden ser tan cortas como de uno a cuatro restos aminoacídicos. Pueden, sin embargo, también ser tan largas como 50 restos aminoacídicos. Puede generarse ya especificidad significativa mediante el uso de SDR de una longitud de cuatro a seis restos aminoacídicos.

Las secuencias peptídicas que se insertan pueden ser completa o parcialmente aleatorias. En este contexto, completamente aleatorias significa que se inserta un conjunto de secuencias en paralelo que incluye secuencias que difieren entre sí en todas y cada una de las posiciones. Parcialmente aleatorias significa que se inserta un conjunto de secuencias en paralelo que incluye secuencias que difieren entre sí en al menos una posición. Esta diferencia puede ser por pares o con respecto a una secuencia sencilla. Por ejemplo, con respecto a una inserción de la longitud de cuatro aminoácidos, parcialmente aleatorio puede ser un conjunto (i) que incluye AGGG, GVGG, GGLG, GGGI, o (ii) que incluye AGGG, VGGG, LGGG y IGGG. Como alternativa, las secuencias aleatorias también comprenden secuencias que difieren entre sí en la longitud. La selección aleatoria de las secuencias peptídicas se consigue seleccionando aleatoriamente las secuencias de nucleótidos que se insertan en el gen en los sitios respectivos. De este modo, la selección aleatoria puede conseguirse empleando mezclas de bases nitrogenadas como monómeros durante la síntesis química de los oligonucleótidos. Una mezcla particularmente preferida de monómeros para un codón completamente aleatorio que además minimiza la probabilidad de codones de parada es NN(GTC). Como alternativa, pueden obtenerse oligonucleótidos aleatorios por fragmentación de ADN en fragmentos cortos que se insertan en el gen en los sitios respectivos. La fuente del ADN a fragmentar puede ser un oligonucleótido sintético pero como alternativa puede originarse de genes clonados, ADNc o ADN genómico. Preferentemente, el ADN es un gen que codifica una enzima. La fragmentación puede, por ejemplo, conseguirse por digestión endonucleolítica aleatoria de ADN. Preferentemente, se emplea una endonucleasa no específica tal como ADNasa I (por ejemplo de páncreas bovino) para la digestión endonucleolítica.

Si las etapas (a) - (c) del procedimiento se repiten de forma cíclica, existen diferentes alternativas para obtener secuencias peptídicas aleatorias que se insertan en ciclos consecutivos. Preferentemente, las SDR que se identificó en un ciclo que conducían a especificidad aumentada de enzima se usan como moldes para las secuencias peptídicas aleatorias que se insertan en el ciclo siguiente.

En una alternativa preferida, las secuencias seleccionadas en un ciclo se analizan y se generan oligonucleótidos seleccionados de forma aleatoria basándose en estas secuencias. Esto puede, por ejemplo, conseguirse usando además del nucleótido original con un cierto porcentaje de mezclas de los otros tres monómeros nucleotídicos en cada posición en la síntesis de oligonucleótido. Si, por ejemplo, en un primer ciclo se identifica una SDR que tiene la

secuencia de aminoácidos ARLT, por ejemplo, codificada por la secuencia de nucleótidos GCG CGC CTT ACC, una secuencia peptídica aleatoria insertada en este sitio de SDR podría codificarse por un oligonucleótido con G 70 %, A 10 %, T 10 % y C 10 % en la primera posición, C 70 %, G 10 %, T 10 % y A 10 % en la segunda posición, etc. Esto conduce en cada posición aproximadamente en uno de cada tres casos al aminoácido molde y en dos de cada tres casos a otro aminoácido.

En otra alternativa preferida, las secuencias seleccionadas en un ciclo se analizan y se genera una biblioteca consenso basándose en estas secuencias. Esto puede, por ejemplo, conseguirse usando mezclas definidas de nucleótidos en cada posición en la síntesis de oligonucleótidos de modo que conduzca a mezclas de los restos aminoacídicos que se identificaron en cada posición de la SDR seleccionada en el ciclo anterior. Si, por ejemplo, en un primer ciclo se identifican dos SDR que tienen las secuencias de aminoácidos ARLT y VPGS, una biblioteca consenso insertada en ese sitio SDR en el siguiente ciclo podría codificarse por un oligonucleótido con la secuencia G(C/T)G C(G/C)C (G/T)(G/T)G (A/T)CC. Esto correspondería a la secuencia peptídica aleatoria (A/V)(R/P)(L/G/V/W)(T/S), permitiendo de este modo todas las combinaciones de los restos aminoacídicos identificados en el primer ciclo y, debido a la degeneración del código genético, permitiendo además en menor grado restos aminoacídicos alternativos en algunas posiciones.

En otra alternativa preferida, estas secuencias seleccionadas en un ciclo se recombinan, sin análisis previo, usando procedimientos para la recombinación *in vitro* de polinucleótidos, tales como los procedimientos descritos en el documento WO 01/34835 (lo siguiente también proporciona detalles del octavo y noveno aspecto).

Después de la inserción de las secuencias parcial o completamente aleatorias en el gen que codifica la proteína de armazón y con el tiempo la ligación del gen resultante en un vector de expresión adecuado usando 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 enzimática correspondiente. Son huéspedes de expresión particularmente adecuados los huéspedes de expresión bacterianos tales como *Escherichia coli* o *Bacillus subtilis*, huéspedes de expresión de levaduras tales como *Saccharomyces cerevisiae* o *Pichia pastoris*, huéspedes de expresión de mamíferos tales como líneas celulares de Ovario de Hámster Chino (CHO) o Riñón de Cría de Hámster (BHK), sistemas de expresión viral tales como bacteriófagos como fago M13T7 o Lambda o virus tales como el sistema de expresión de Baculovirus. Como una alternativa adicional, pueden usarse sistemas para expresión proteica *in vitro*. Normalmente el ASDN se liga en un vector de expresión detrás de una secuencia señal adecuada que conduce a secreción de las variantes enzimáticas al espacio intracelular, permitiendo de este modo la detección directa de actividad enzimática en el sobrenadante celular. Son secuencias señal particularmente adecuadas para *Escherichia coli* ompA, pelB, HlyA, para *Bacillus subtilis* AprE, NprB, Mpr, AmyA, AmyE, Blac, SacB, y para *S. cerevisiae* Bar1, Suc2, Mat α , Inu1A, Ggplp. Como alternativa, las variantes de enzima se expresan intracelularmente y los sustratos se expresan también intracelularmente. De acuerdo con variantes de proteasa esto se realiza esencialmente como se describe en la Solicitud de Patente WO 0212543, usando un sustrato peptídico de fusión que comprende dos proteínas auto-fluorescentes ligadas por la secuencia de aminoácidos sustrato. Como una alternativa adicional, después de la expresión intracelular de las variantes de enzima o secreción al espacio periplásmico usando secuencias señal tales como DsbA, PhoA, PelB, OmpA, OmpT o gIII para *Escherichia coli*, una etapa de permeabilización o lisis libera las variantes de enzima al sobrenadante. La destrucción de la barrera de membrana puede forzarse por el uso de medios mecánicos tales como ultrasonidos, prensa Francesa o el uso de enzimas de digestión de membrana tales como lisozima. Como otra alternativa adicional, los genes que codifican las variantes enzimáticas se expresan sin células mediante el uso de un sistema de expresión sin células adecuado. Por ejemplo, el extracto S30 de células de *Escherichia coli* se usa para este fin como se describe en Lesly y col. (Methods in Molecular Biology 37 (1995) 265-278).

Después de la introducción del vector en células huésped, estas células se criban con respecto a la expresión de enzimas con especificidad para el sustrato diana pretendido. Dicho cribado se realiza normalmente separando las células entre sí, para permitir la correlación de genotipo y fenotipo y ensayando la actividad de cada clon celular después de un periodo de crecimiento y expresión. Dicha separación puede por ejemplo realizarse por distribución de las células en los compartimentos de vehículos de muestras, por ejemplo como se describe en el documento WO 01/24933. Como alternativa, las células se separan sembrando en estrías en placas de agar, incluyendo en un polímero tal como agarosa, llenando capilares o por procedimientos similares.

La identificación de variantes con la especificidad pretendida puede realizarse por diferentes enfoques. En el caso de proteasas, preferentemente se emplean ensayos que usan sustratos peptídicos esencialmente como se describe en el documento PCT/EP03/04864.

Independientemente del formato de expresión, la selección de variantes de enzima se realiza en condiciones que permiten la identificación de enzimas que reconocen y convierten la secuencia diana preferentemente. Como una primera alternativa, se identifican enzimas que reconocen y convierten la secuencia diana preferentemente por cribado con respecto a enzimas con una alta afinidad para la secuencia sustrato diana. La alta afinidad corresponde a una baja K_M que se selecciona cribando a concentraciones de sustrato diana sustancialmente por debajo de la K_M de la primera enzima. Preferentemente, los sustratos que se usan están ligados a uno o más fluoróforos que permite la detección de la modificación del sustrato en concentraciones por debajo de 10 μ M, preferentemente por debajo de

1 μ M, más preferentemente por debajo de 100 nM y más preferentemente por debajo de 10 nM.

Como una segunda alternativa, se identifican enzimas que reconocen y convierten el sustrato diana preferentemente empleando dos o más sustratos en el ensayo y cribando con respecto a actividad en estos dos o más sustratos en comparación. Preferentemente, los dos o más sustratos empleados están ligados a diferentes moléculas marcadoras, permitiendo de este modo la detección de la modificación de los dos o más sustratos consecutivamente o en paralelo. En caso de proteasas, particularmente preferentemente se emplean dos sustratos peptídicos, un sustrato peptídico que tiene una secuencia de aminoácidos seleccionada de forma arbitraria o incluso parcial o completamente aleatoria permitiendo de este modo controlar la actividad en un sustrato arbitrario y teniendo el otro sustrato peptídico una secuencia de aminoácidos idéntica o semejante a la secuencia sustrato diana pretendida permitiendo de este modo controlar la actividad en el sustrato diana. Especialmente preferentemente, estos dos sustratos peptídicos están ligados a moléculas marcadoras fluorescentes y las propiedades fluorescentes de los dos sustratos peptídicos son suficientemente diferentes para distinguir ambas actividades cuando se miden consecutivamente o en paralelo. Por ejemplo, puede usarse una proteína de fusión que comprende una primera proteína auto-fluorescente, un péptido y una segunda proteína auto-fluorescente de acuerdo con la Solicitud de Patente WO 0212543 para este fin. Como alternativa, se ligan químicamente fluoróforos tales como rodaminas a los sustratos peptídicos.

Como una tercera alternativa, se identifican enzimas que reconocen y convierten un sustrato diana preferentemente empleando uno o más sustratos que se asemejan al sustrato diana junto con sustratos que compiten en alto exceso. Se realiza después cribado con respecto a actividad en los sustratos que se asemejan a sustrato diana en presencia de los sustratos que compiten. Las enzimas que tiene una especificidad que corresponde cualitativamente a la especificidad diana, pero que tienen solamente una especificidad cuantitativa baja se identifican como muestras negativas en un cribado tal, mientras que las enzimas que tienen una especificidad que corresponde cualitativa y cuantitativamente a la especificidad diana se identifican positivamente. Preferentemente, el o los sustratos que se asemejan al sustrato diana se ligan a moléculas marcadoras, permitiendo de este modo la detección de sus modificaciones, mientras que los sustratos que compiten no portan moléculas marcadoras. Los sustratos que compiten tiene secuencias de aminoácidos seleccionadas arbitrariamente o aleatorias, actuando de este modo como inhibidores competitivos para la hidrólisis de los sustratos que portan marcadores. Por ejemplo, los hidrolizados proteicos tales como Trypton pueden actuar como sustratos de competición para enzimas proteolíticas obtenidas por ingeniería genética.

Como una cuarta alternativa, se identifican y seleccionan enzimas que reconocen y convierten el sustrato diana preferentemente por una etapa de selección acoplada a amplificación o acoplada a crecimiento. Además, la actividad puede medirse intracelularmente y la selección puede realizarse por un clasificador celular, tal como un clasificador celular activado por fluorescencia.

Como una alternativa adicional, se identifican enzimas que reconocen y convierten el sustrato diana seleccionando primero enzimas que se unen preferentemente al sustrato diana y en segundo lugar seleccionando de este subgrupo de variantes de enzima las enzimas que convierten el sustrato diana. La selección de enzimas que se unen preferentemente al sustrato diana puede realizarse por selección de agentes de unión al sustrato diana o por el contrario selección de enzimas que se unen a otros sustratos. Se conocen en la técnica procedimientos para la selección de agentes de unión o para la contra selección de agentes no de unión. Tales procedimientos normalmente requieren acoplamiento fenotipo-genotipo que puede resolverse usando procedimientos de expresión de presentación superficial. Tales procedimientos incluyen, por ejemplo, presentación de fagos o virus, presentación de superficie celular y presentación *in vitro*. La presentación de fagos o viral normalmente implica fusión de la proteína de interés con una proteína viral/de fago. La presentación de superficie celular, es decir presentación de célula bacteriana de eucariota, normalmente implica fusión de la proteína de interés con un péptido o proteína que se localiza en la superficie celular. En presentación *in vitro*, la proteína se prepara normalmente *in vitro* y se liga directa o indirectamente al ARNm que codifica la proteína (documento DE 19646372).

La divulgación también proporciona una composición o composición farmacéutica que comprende una o más enzimas obtenidas por ingeniería genética de acuerdo con el primer aspecto como se ha definido anteriormente en el presente documento. La composición puede comprender opcionalmente un vehículo, excipiente y/o agente adyuvante aceptable. Las composiciones no farmacéuticas como se define en el presente documento son composición de investigación, composición nutricional, composición de limpieza, composición de desinfección, composición cosmética o composición para cuidado personal. Además, también se proporcionan secuencias de ADN que codifican la enzima obtenida por ingeniería genética como se ha definido anteriormente en el presente documento y vectores que contienen dichas secuencias de ADN. Finalmente, también se contemplan células huésped transformadas (procariotas o eucariotas) u organismos transgénicos que contiene tales secuencias de ADN y/o vectores, así como un procedimiento que utiliza tales células huésped o animales transgénicos para producir la enzima obtenida por ingeniería genética del primer aspecto.

Descripción detallada de las figuras

Figura 1: Estructura tridimensional de tripsina humana I con los restos de sitio activo mostrados en representación de "bolas y varillas" e indicando las reacciones marcadas sitios de inserción de SDR potenciales.

Figura 2: Alineamiento de las secuencias de aminoácidos primarias de las proteasa humanas tripsina I, alfa-trombina y enteropeptidasa todas las cuales pertenecen a la clase estructural S1 de la familia de serina proteasa. La tripsina presenta una proteasa no específica de esta clase estructural, mientras que alfa-trombina y enteropeptidasa son proteasas con alta especificidad de sustrato. En comparación con tripsina se ven varias regiones de inserciones de tres o más aminoácidos en la secuencia primaria de α -trombina y enteroquinasa. La región marcada con (-1-) y la región marcada con (-3-) son sitios de inserción de SDR preferidos. En la estructura terciaria de alfa-trombina ambas regiones están en las cercanías del sitio de unión a sustrato. Estas regiones cumplen por lo tanto dos criterios para seleccionarse como candidatos para SDR: en primer lugar, representan inserciones en las proteasas específicas en comparación con la no específica y en segundo lugar, están cerca del sitio de unión de sustrato. Se proporciona la representación de la estructura tridimensional en la Figura 3.

Figura 3: Estructura tridimensional de subtilisina mostrándose los restos de sitio activo en representación de "bolas y varillas" indicando las regiones numeradas sitios de inserción de SDR potenciales.

Figura 4: Alineamiento de las secuencias de aminoácidos primarias de subtilisina E, furina, PC1 y PC5 todas las cuales pertenecen a la clase estructural S8 de la familia de serina proteasa. La subtilisina E representa una proteasa no específica de esta clase estructural, mientras que furina, PC1 y PC5 son proteasas con alta especificidad de sustrato. En comparación con subtilisina se ven varias regiones de inserciones de tres o más aminoácidos en la secuencia primaria de furina, PC1 y PC5. Las reacciones marcadas con (-4-), (-5-), (-7-), (-9-) y (-11-) son sitios de inserción de SDR preferidos. Estos tramos de región cumplen dos criterios para seleccionarse como candidatos para SDR: en primer lugar, representan inserciones en las proteasas específicas en comparación con la no específica y, en segundo lugar, están cerca de los restos del sitio activo.

Figura 5: Estructura tridimensional de beta-secretasa mostrándose los restos de sitio activo en representación de "bolas y varillas" e indicando las regiones numeradas sitios de inserción de SDR potenciales.

Figura 6: Alineamiento de las secuencias de aminoácidos primarias de pepsina, b-secretasa y catepsina D, todas las cuales pertenecen a la clase estructural A1 de la familia de proteasa aspártica. La pepsina representa una proteasa no específica de esta clase estructural, mientras que b-secretasa y catepsina D son proteasas con alta especificidad de sustrato. En comparación con pepsina se ven varias regiones de inserciones de tres o más aminoácidos en la secuencia primaria de b-secretasa y catepsina D. las regiones marcadas con -1- a -11- corresponden a posibles sitios de combinación de SDR y también están marcados en la Figura S.

Figura 7: Ilustra la estructura tridimensional de caspasa 7 mostrándose los restos del sitio activo en representación de "bolas y varillas" e indicando las regiones numeradas sitios de inserción de SDR potenciales.

Figura 8: Muestra la secuencia de aminoácidos primarios de caspasa 7 como un miembro de la familia de cisteína proteasa de clase C14 (véase también SEC ID N°: 14).

Figura 9: Representación esquemática del procedimiento de acuerdo con el tercer aspecto.

Figura 10: Análisis de transferencia de Western de expresión de tripsina. El sobrenadante de cultivos celulares que expresan variantes de tripsina se compara con controles negativos. Carril 1: patrón de peso molecular; carril 2: control negativo; carril 3: sobrenadante de variante a; carril 4: control negativo; carril 5: sobrenadante de variante b. Se usó un anticuerpo primario específico para la proteína expresada y un anticuerpo secundario para generación de la señal.

Figura 11: Curso temporal de la escisión proteolítica de un sustrato diana. El sobrenadante de células que contienen el vector con el gen para tripsina humana y el de células que contienen el vector sin el gen se incubó con el sustrato peptídico descrito en el texto. La escisión del péptido dio como resultado una disminución del valor de lectura. La actividad proteolítica se confirma para el clon positivo.

Figura 12: Actividad relativa de tres enzimas proteolíticas obtenidas por ingeniería genética en comparación con tripsina humana I en dos sustratos peptídicos diferentes. Se realizó un ciclo temporal de la digestión proteolítica de los dos sustratos y se evaluó. Se usó sustrato B para el cribado y el sustrato A es una secuencia cercanamente relacionada. La actividad relativa de las tres variantes se normalizó a la actividad de tripsina humana I. La variante 1 y 2 claramente muestran aumento de la especificidad hacia el sustrato diana. La variante 3, por otro lado sirve como control negativo con actividades similares a la tripsina I humana.

Figura 13: Especificidades relativas de tripsina y variantes de enzimas proteolíticas obtenidas por ingeniería genética con una o dos SDR, respectivamente. La actividad de las proteasas se determinó en presencia y ausencia del sustrato competidor, es decir peptona a una concentración de 10 mg/ml. Los ciclos temporales para la escisión proteolítica se registraron y se determinaron las constantes de tiempo k. las relaciones entre las constantes de tiempo con y sin competidor se formaron y representan una medida cuantitativa para especificidad de la proteasa. Las relaciones se normalizaron a tripsina. La especificidad de la variante que contenía dos SDR es 2,5 veces mayor que la de la variante con SDR2 solamente.

Figura 14: Muestra las especificidades relativas de variantes de proteasa en ausencia y presencia de sustrato

competidor. Las variantes de proteasa que contiene dos insertos con diferentes secuencias y la tripsina humana I de armazón no modificado se expresaron en un huésped adecuado. La actividad de las variantes de proteasa se determinó como la velocidad de escisión de un péptido con la secuencia diana deseada de TNF-alfa en ausencia y presencia de sustrato competidor. La especificidad se expresa como la relación de velocidades de escisión en presencia o ausencia de competidor.

Figura 15: La figura muestra la reducción de citotoxicidad inducida por TNF-alfa humano cuando se incubó el TNF-alfa humano con sobrenadante concentrado de cultivos que expresan las enzimas proteolíticas obtenidas por ingeniería genética que son específicas para TNF-alfa humano. Esto indica la eficacia de las enzimas proteolíticas obtenidas por ingeniería genética.

Figura 16: La figura muestra la reducción de citotoxicidad inducida por TNF-alfa humano cuando se incubó el TNF-alfa humano con diferentes concentraciones de enzima proteolítica obtenida por ingeniería genética purificada que es específica para TNF-alfa humano. La variante G comprende SEC ID N°: 72 como SDR1 y SEC ID N°: 73 como SDR2. Esto indica la eficacia de las enzimas proteolíticas obtenidas por ingeniería genética.

Figura 17: La figura compara la actividad de enzimas proteolíticas obtenidas por ingeniería genética que son específicas para TNF-alfa humano con la actividad de tripsina humana I en dos sustratos proteicos: (a) TNF-alfa humano; (b) mezcla de proteínas de suero humano. Esto indica la seguridad de las enzimas proteolíticas obtenidas por ingeniería genética. La variante x corresponde a SEC ID N°: 75 que comprende las SDR de acuerdo con SEC ID N° 89 (SDR1) y 95 (SDR2). Las variantes xi y xii corresponden a derivados de la misma que comprende las mismas secuencias de SDR.

Figura 18: Hidrólisis específica de VEGF humano por una enzima proteolítica obtenida por ingeniería genética derivada de tripsina humana.

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. Debería entenderse que estos ejemplos son para fines ilustrativos solamente y no deben interpretarse como limitantes de la presente invención de ninguna manera.

En los ejemplos experimentales descritos posteriormente, se usaron técnicas convencionales de tecnología de ADN recombinante que se han descrito en diversas publicaciones, por ejemplo Sambrook y col. (1989), *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory, o Ausubel y col. (1987), *Current Protocols in Molecular Biology 1987-1988*, Wiley Interscience. A no ser que se indique de otro modo, las enzimas de restricción, polimerasas y otras enzimas así como kits de purificación de ADN se usaron de acuerdo con las especificaciones de los fabricantes.

Ejemplo I: Identificación de sitios de SDR en tripsina humana

Se han identificado sitios de inserción para SDR en la serina proteasa tripsina humana I (clase estructural S1) por comparación con miembros de la misma clase estructural que tienen una mayor especificidad de secuencia. La tripsina representa un miembro con especificidad de sustrato baja, puesto que requiere solamente un resto de arginina o lisina en la posición P₁. Por otro lado, la trombina, activador de plasminógeno de tipo tisular o enteroquinasa tiene todas una alta especificidad para sus secuencias sustrato, es decir (L/I/V/F)XPR[^]NA, CPGR[^]VVGG y DDDK[^], respectivamente. Las secuencias primarias y estructuras terciarias de estas y adicionales serina proteasa S1 se han alineado para determinar regiones de baja y alta homología de secuencia y estructura y especialmente regiones que corresponden a inserciones en las secuencias de las proteasas más específicas (Figura 2). Se han identificado varias regiones de inserciones iguales o mayores de tres aminoácidos que representan sitios de SDR potenciales como se indica en la Figura 1. Estas regiones se seleccionaron como sitios diana para la inserción de SDR en los ejemplos posteriores, por ejemplo SDR1 (región 1 en la Figura 2, después del aminoácido 42 de acuerdo con SEC ID N°: 1) con una longitud de 6 y SDR2 (región 3 en la Figura 2, después del aminoácido 123 de acuerdo con SEC ID N°: 1) con una longitud de 5 aminoácidos, respectivamente.

Ejemplo II: Clonación molecular del gen de tripsina humana I para usarlo como proteína de armazón y expresión de la proteasa madura en *B. subtilis*

El gen que codifica la proteasa no específica tripsinógeno humano I se clonó en el vector pUC18. La clonación se realizó como sigue: la secuencia codificante de la proteína se codificó por PCR usando cebadores que introdujeron un sitio KpnI en el extremo 5' y un sitio BamHI en el extremo 3'. Este fragmento de PCR se clonó en los sitios apropiados del vector pUC18. La identidad se confirmó por secuenciación. Después de secuenciar la secuencia codificante de la proteína madura se amplificó por PCR usando cebadores que introdujeron diferentes sitios BglII en el extremo 5' y el extremo 3'. Este fragmento de PCR se clonó en los sitios apropiados de un vector lanzadera de *E. coli* - *B. subtilis*. El vector contiene un origen pMB1 para la amplificación en *E. coli*, un marcador de resistencia a neomicina para selección en *E. coli* así como un promotor P43 para la expresión constitutiva en *B. subtilis*. Se introdujo un fragmento de 87 pb que contiene la secuencia líder que codifica el péptido señal del gen sacB de *B.*

subtilis detrás del promotor P43. Diferentes sitios de restricción BglI actúan como sitios de inserción para que se expresen genes heterólogos.

La expresión de tripsina humana I se confirmó por medición de la actividad proteolítica en sobrenadante de células que contiene el vector con el gen en comparación con un control negativo. Se seleccionó un péptido que incluía un sitio de escisión de arginina como un sustrato. El péptido se biotiniló en el extremo N-terminal y se marcó con un fluoróforo en el extremo C-terminal. Después de incubación del péptido con sobrenadante de cultivo se añadió estreptavidina. El péptido no escindido se asocia con estreptavidina y conduce a un alto valor de lectura mientras que la escisión da como resultado bajos valores de lectura. La Figura 11 muestra el ciclo temporal de una digestión proteolítica de células *B. subtilis* que contiene el vector con el gen de tripsina I en comparación con células de *B. subtilis* que contiene el vector sin el gen de tripsina I (control negativo). Como confirmación adicional de la expresión de la proteasa, se analizaron los sobrenadantes de células que contenían el vector con el gen y células de control por electroforesis en gel de poliacrilamida y posterior transferencia de Western usando un anticuerpo específico para la proteasa diana. El procedimiento se realizó de acuerdo con procedimientos convencionales (Sambrook, J.F.; Fritsch, E.F.; Maniatis, T.; Cold Spring Harbor Laboratory Press, Segunda edición, 1989, Nueva York). La Figura 8 confirma la expresión de la proteína solamente en las células que albergan el vector con el gen para tripsina.

Ejemplo III: Proporcionar una proteína de almacén

En este ejemplo, la tripsina humana I se usó como la proteína de almacén. El gen se usó en su forma natural o, como alternativa, se modificó para dar como resultado una proteína de almacén con actividad catalítica aumentada o características mejoradas adicionales.

La modificación se realizó por modificación aleatoria del gen, seguido de expresión de la enzima y posterior selección con respecto a actividad aumentada. En primer lugar, el gen se amplificó por PCR en condiciones propensas a errores, esencialmente como se describe en Cadwell, R.C y Joyce, G.F. (PCR Methods Appl. 2 (1992) 28-33). Se realizó PCR propensa errores usando 30 pmol de cada cebador, 20 nmol de dGTP y dATP, 100 nmol de dCTP y dTTP, 20 fmol de molde y 5 U de ADN polimerasa Taq en Tris HCl 10 mM pH 7,6, KCl 50 mM, MgCl₂ 7 mM, MnCl₂ 0,5 mM, gelatina 0,01 % durante 20 ciclos de 1 min a 94 °C, 1 min a 65 °C y 1 min a 72 °C. La biblioteca de ADN resultante se purificó usando el Kit de Purificación de PCR Qiaquick siguiendo las instrucciones de los proveedores. El producto de PCR se digirió con la enzima de restricción BglI y se purificó. A continuación, el producto de PCR se ligó en el vector lanzadera *E. coli* - *B. subtilis* descrito anteriormente que se digirió con BglI y se desfosforiló. Los productos de ligación se transformaron en *E. coli*, se amplificaron en LB y los plásmidos se purificaron usando el Kit de Purificación de plásmidos Qiagen siguiendo las instrucciones de los proveedores. Los plásmidos resultantes se transformaron en células *B. subtilis*.

Como alternativa, o además de la mutagénesis aleatoria, las variantes del gen se recombinaron estadísticamente en posiciones homólogas mediante el uso de reacción en cadena de recombinación, esencialmente como se describe en el documento WO 0134935. Los productos de PCR de los genes que codifican las variantes de proteasa se purificaron usando el Kit de Purificación de PCR QIAquick siguiendo las instrucciones de los proveedores, se comprobaron con respecto a tamaño correcto mediante electroforesis en gel de agarosa y se mezclaron entre sí en cantidades equimolares. Se calentaron 80 µg de esta mezcla de PCR en TrisHCl 150 mM, pH 7,6, MgCl₂ 6,6 mM durante 5 min a 94 °C y se enfrió posteriormente hasta 37 °C a 0,05 °C/s para rehibridar las cadenas y producir de este modo heterodúplex de una manera estocástica. Después se añadieron 2,5 U de Exonucleasa III por µg de ADN y se incubó durante 20, 40 o 60 min a 37 °C para digerir diferentes longitudes de ambos extremos 3' de los heterodúplex. Los productos de PCR parcialmente digeridos se recargaron con 0,6 U de Pfu polimerasa por µg de ADN incubando durante 15 minutos a 72 °C en dNTP 0,17 mM y tampón de Pfu polimerasa de acuerdo con las instrucciones de los proveedores. Después de realizar un ciclo de PCR sencillo, el ADN resultante se purificó usando el Kit de Purificación de PCR QIAquick siguiendo las instrucciones de los proveedores, se digirió con BglI y se ligó en el vector linealizado. Los productos de ligación se transformaron en *E. coli*, se amplificaron en LB que contenía ampicilina como marcador y los plásmidos se purificaron usando el Kit de Purificación de Plásmidos Qiagen siguiendo las instrucciones de los proveedores. Los plásmidos resultantes se transformaron en células de *B. subtilis*.

Ejemplo IV: Inserción de SDR en el almacén proteico de tripsina humana I y generación de una enzima proteolítica obtenida por ingeniería genética con especificidad para un sustrato peptídico que tiene la secuencia KKWLGRVGGPV

Para crear sitios de inserción para SDR en tripsina humana I se introdujeron dos pares de sitios de restricción diferentes en el gen en sitios que se identificaron como sitios de SDR potenciales (véase Ejemplo I anterior) sin cambiar la secuencia de aminoácidos. La inserción de los sitios de restricción se realizó por PCR de extensión de solapamiento. Se usaron los cebadores restr1 y restr2 para la introducción de los sitios de restricción SacII y BamHI, se usaron restr3 y restr4 para la introducción de los sitios de restricción KpnI y NheI. Las secuencias de los cebadores fueron las siguientes:

Sitio de unión para restr1 y restr2 y la secuencia de aminoácidos correspondiente (SEC ID N°: 54):

5'-GGTGGTATCAGCAGGCCACTGCTACAAGTCCCCGCATCCAGGT-3'

V V S A G H C Y K S R I Q

Cebador directo restr1 (SEC ID N°: 56):

5'-GGTGGTATCCGCGGGCCACTGCTACAAGTCCCGGATCCAGGT-3'

Cebador inverso restr2 (SEC ID N°: 57):

5'-ACCTGGATCCGGGACTTGTAGCAGTGGCCCGCGGATACCACC-3'

Sitio de unión para restr3 y restr4 y la secuencia de aminoácidos correspondiente (SEC ID N°: 58):

5'-CCACTGGCACGAAGTGCCTCATCTCTGGCTGGGGCAACACTGCGAGCTCT-3'

T G T K C L I S G W G N T A S S

Cebador directo restr3 (SEC ID N°: 60):

5'-CCACTGGCACGAAGTGCCTCATCTCTGGCTGGGGCAACACTGCGAGCTCT-3'

Cebador inverso restr4 (SEC ID N°: 61):

5'-AGAGCTAGCAGTGTTGCCCCAGCCAGAGATGAGGCACTTGGTACCAGTGG-3'

En una primera PCR de extensión de solapamiento, se introdujeron los sitios SacII/BamHI permitiendo insertar SDR1 y en una segunda PCR de extensión de solapamiento los sitios KpnI/NheI permitiendo la inserción de SDR2. El producto de la PCR de extensión de solapamiento se amplificó usando los cebadores pUC directo y pUC inverso. Las secuencias de pUC directo y pUC inverso son las siguientes:

pUC directo (SEC ID N°: 62): 5'-GGGGTACCCACCACCATGAATCCACTCCT-3'

pUC inverso (SEC ID N°: 63): 5'-CGGGATCCGGTATAGAGACTGAAGAGATAC-3'

Los sitios de restricción generados de este modo se usaron posteriormente para insertar oligonucleótidos definidos o aleatorios en los sitios de inserción SDR1 SDR2 por procedimientos de restricción y ligación convencionales. Normalmente, se hibridaron dos oligonucleótidos 5'-fosforilados sintéticos complementarios y se ligaron en un vector que portaba el gen de tripsina humana I modificado que se escindió con las enzimas de restricción respectivas. Se insertaron oligonucleótidos que codificaban SDR1 mediante los sitios SacII/BamHI mientras que se insertaron oligonucleótidos que codificaban SDR2 mediante los sitios KpnI/NheI. Para cada inserción se usó un par de oligonucleótidos de acuerdo con las secuencias generales siguientes (indicando [P] fosforilación 5', indicando N y X cualquier resto nucleotídico o aminoacídico, respectivamente):

oligox-SDR1f (SEC ID N°: 64):

5'-[P]-GGGCCACTGCTACNNNNNNNNNNNNNNNNNNNAAGTCCCG-3'

oligox-SDR1r (SEC ID N°: 66):

3'-CGCCGGTGACGATGNNNNNNNNNNNNNNNNNNNTTCAGGGCCTAG-[P]-5'

G H C Y X X X X X X K S

oligox-SDR2f (SED ID N°: 67):

5'-[P]-CAAGTGCCTCATCTCTGGCTGGGGCAACNNNNNNNNNNNNNNNACTG-3'

oligox-SDR2r (SEC ID N°: 69):

3'-CATGGTTCACGGAGTAGAGACCGACCCCGTTGNNNNNNNNNNNNNNNTGACGATC-[P]-5'

K C L I S G W G N X X X X X T

Como alternativa al procedimiento anterior, se usó un procedimiento basado en PCR para la integración de secuencias aleatorias en los sitios de inserción de SDR1 y SDR2 en la tripsina humana modificada 1. Para cada SDR, se usó un cebador en el que la región SDR se ha seleccionado de forma completamente aleatoria. Las secuencias de los cebadores fueron las siguientes (N = A/C/G/T, B = C/G/T, V = A/C/G):

Cebador SDR1-mutnbn-directo (SEC ID N°: 70):

5'-TGGTATCCGCGGGCCACTGCTACNNBNNBNNBNNBNNBNNBAAGTCCCGGATCCAGGTG-3'

Cebador SDR2-mutnbnb-inverso (SEC ID N°: 71):

5'-GGCGCCAGAGCTAGCAGTVNNVNNVNNVNNVNNNGTTGCCCCAGCCAGAGATG-3'

El codón NNB o VNN en la hebra inversa, permite que se realicen los 20 aminoácidos pero reduce la probabilidad de codificar un codón de parada de 0,047 a 0,021.

5 Como una alternativa adicional, después de la identificación de SDR que conducen a especificidad aumentada, estas SDR se usaron como moldes para selección aleatoria adicional. De este modo, se insertaron secuencias peptídicas aleatorias que se seleccionaron de forma parcialmente aleatoria en cada posición y parcialmente idénticas en cada posición a la secuencia original.

10 Como ejemplo, se insertaron secuencias peptídicas aleatorias que tenían en aproximadamente en uno de cada tres casos el resto aminoacídico del molde y en aproximadamente dos de cada tres casos cualquier otro resto aminoacídico en cada posición en los dos sitios de inserción de SDR de la tripsina humana modificada humana I. Para este fin, se usaron cebadores que contenían en cada posición de nucleótido de la SDR aproximadamente el 70 % de las bases del molde y el 30 % de una mezcla de las tres otras bases.

15 Con cada par de cebadores se realizó una PCR en condiciones convencionales usando el gen de tripsina humana I como molde. El ADN restante se purificó usando el Kit de Purificación de PCR QIAquick siguiendo las instrucciones de los proveedores y se digirió con SacII y NheI. Después de la digestión el ADN se purificó y se ligó en el vector digerido con SacII y NheI y desfosforilado. Los productos de ligación se transformaron en *E. coli*, se amplificaron en LB que contenía el marcador respectivo y los plásmidos se purificaron usando el Kit de Purificación de Plásmidos Qiagen siguiendo las instrucciones de los proveedores. Los plásmidos resultantes se transformaron en células de *B. subtilis*. Estas células se separaron después a células sencillas, se dejaron crecer a clones y después de la expresión del gen de proteasa se cribaron con respecto a actividad proteolítica.

20 Se emplearon los siguientes sustratos para cribado con respecto a actividad proteolítica (SEC ID N°: 76 y 77):

| | | | | | | | | | | | | |
|------------|---|---|---|---|---|---|---|---|---|---|---|---|
| Sustrato A | L | L | W | L | G | R | V | V | G | G | P | V |
| Sustrato B | K | K | W | L | G | R | V | V | G | G | P | V |

25 Se cribaron variantes de proteasa en el sustrato B a complejidades de 10^6 variantes por espectroscopia de fluorescencia confocal. El sustrato fue un péptido biotinilado en el extremo N-terminal y marcado con fluorescencia en el extremo C-terminal. Después de la incubación del péptido con sobrenadante de células que expresaban diferentes variantes de la proteasa, se añadió estreptavidina y las muestras se analizaron mediante fluorimetría confocal. La concentración baja del péptido (20 nM) conduce a escisión preferente por proteasas con un alto valor de k_{cat}/K_M , es decir proteasas con alta especificidad para la secuencia diana.

30 Se evaluaron adicionalmente variantes seleccionadas en el procedimiento de cribado con respecto a su especificidad para el sustrato B y el sustrato cercanamente relacionado A midiendo los ciclos temporales de la digestión proteolítica y determinando las constantes de velocidad que son proporcionales a los valores de k_{cat}/K_M . claramente, en comparación con la tripsina humana que se usó como proteína de armazón, la actividad específica de las variantes 1 y 2 está desplazada (SEC ID N°: 2 y 3, respectivamente) hacia el sustrato B. La variante 3 (SEC ID N°: 4), por otro lado, actúa como un control negativo con actividades similares a las de tripsina humana 1. La secuenciación de los genes de las tres variantes reveló las siguientes secuencias de aminoácidos en las SDR.

Tabla 2: Secuencias de las dos SDR en tres variantes diferentes seleccionadas para hidrólisis específica de sustrato B (SEC ID N°: 78-82).

| | SDR1 | | | | | | SDR2 | | | | |
|------------|------|---|---|---|---|---|------|---|---|---|---|
| Tripsina | - | - | - | - | - | - | - | - | - | - | - |
| Variante 1 | D | A | V | G | R | D | T | I | T | N | S |
| Variante 2 | N | G | R | D | L | E | V | R | G | T | W |
| Variante 3 | G | F | V | M | F | N | R | S | P | L | T |

40 En un experimento adicional se cribó un grupo de variantes que contenían diferentes números de SDR por gen con respecto a especificidad aumentada usando una mezcla del sustrato definido y peptona como un sustrato competidor. Se han analizado adicionalmente las variantes que contienen una o dos SDR por gen. Como una medida de la especificidad la actividad en el ensayo de escisión de péptidos se comparó con y sin la presencia del sustrato competidor. La concentración del sustrato competidor fue de 10 mg/ml. En estas condiciones, las proteasas no específicas muestran, en comparación con proteasas específicas, una reducción más fuerte de la actividad con

concentraciones de competidor crecientes (intervalo entre 0 y 100 mg/ml). La relación de actividad proteolítica con y sin sustrato es una medida cuantitativa de la especificidad de la proteasas. La Figura 9 muestra las actividades relativas con o sin sustrato competidor. La tripsina humana 1 que se usó como la proteína de armazón y dos variantes, una que contenía solamente SDR2 y una que contenía ambas SDR, se compararon. La especificidad de la variante con ambas SDR es mayor en un factor de 2,5 que la de la variante con SDR2 solamente, lo que confirma que existe una relación directa entre el número de SDR y la especificidad cuantitativa de enzimas proteolíticas de enzimas proteolíticas obtenidas por ingeniería genética resultantes.

Ejemplo V: Generación de una enzima proteolítica obtenida por ingeniería genética que inactiva específicamente TNF-alfa humano

La tripsina humana alfa I o un derivado que comprende una o más de las siguientes sustituciones de aminoácidos E56G; R78W; Y131F; A146T; C183R se usó como armazón proteico para la generación de una enzima proteolítica obtenida por ingeniería genética con alta especificidad para TNF-alfa humano. La identificación de sitios SDR en tripsina humana I o derivados de la misma se realizó como se ha descrito anteriormente. Se seleccionaron dos sitios de inserción dentro del armazón para SDR. Las variantes de proteasa que contenían dos insertos con diferentes secuencias y también la tripsina humana I en sí misma sin insertos se expresaron en células de *Bacillus Subtilis*. Las células con proteasa variante se separaron en clones celulares sencillos y las variantes que expresaban proteasa se cribaron con respecto a actividad proteolítica en péptidos con la secuencia diana deseada de TNF-alfa. La actividad de las variantes de proteasa se determinó como la velocidad de escisión de un péptido con la secuencia diana deseada de TNF-alfa en ausencia y presencia de sustrato competidor. La especificidad se expresa como la relación de velocidades de escisión en presencia o ausencia de competidor (Figura 14).

Tabla 3: Especificidad relativa de variantes de enzimas proteolíticas obtenidas por ingeniería genética con diferentes secuencias SDR en ausencia y presencia de sustrato competidor (SEC ID N°: 84-95)

| | K con comp./k sin comp. | Secuencia de SDR1 | Secuencia de SDR2 |
|-------------------|-------------------------|-------------------|-------------------|
| Armazón (sin SDR) | 0,092 | --- | --- |
| variante a | 0,130 | RPWDPS | VHPTS |
| variante b | 0,187 | GFVMFN | RSPLT |
| variante c | 0,235 | EIANRE | RGART |
| variante d | 0,310 | KAVVGT | RTPIS |
| variante e | 0,374 | VNIMAA | TTARK |
| variante f | 0,487 | AAFNGD | RKDFW |

El efecto antagonista de tres variantes de proteasa en TNF-alfa humano se muestra en la Figura 15. Mediante el uso de las variantes, la inducción de la apoptosis se elimina casi completamente lo que indica la eficacia antiinflamatoria de las proteasas para iniciar rotura de TNF-alfa. Se ha incubado TNF-alfa con un sobrenadante concentrado de cultivos que expresan las variantes i a iii durante 2 horas. El TNF-alfa resultante se ha incubado con células no modificadas durante 4 horas. El efecto de la actividad de TNF-alfa restante se determinó como el alcance de la inducción de apoptosis por detección de caspasa-3 activada como marcador de células apoptóticas. Para los controles no se añade proteasa con el TNF-alfa humano (células muertas) o se usó tampón en lugar de TNF-alfa humano (células vivas), respectivamente. Se muestra un experimento análogo en la Figura 16 que usa variante purificada xiii. Se incubó TNF-alfa con diferentes concentraciones de la variante de proteasa purificada.

Para demostrar la especificidad de las variantes de proteasa, se han incubado proteínas de suero sanguíneo humano o TNF-alfa humano purificado con tripsina humana I o las variantes de enzima proteolítica obtenida por ingeniería genética, respectivamente. Aquí, la variante x corresponde a SEC ID N°: 75 que comprende las mismas SDR que la variante f, es decir las SDR de acuerdo con SEC ID N°: 89 (SDR1) y 95 (SDR2). Las variantes xi y xii corresponden a derivados de las mismas que comprenden las mismas secuencias de SDR. Se determinó la proteína intacta restante en función del tiempo. Aunque las variantes así como tripsina humana I digieren TNF-alfa humano, solamente la tripsina muestra actividad en proteína del suero (Figura 17 a y b). Esto demuestra la alta especificidad de TNF-alfa de las enzimas proteolíticas e indica su seguridad y en consecuencia sus bajos efectos secundarios para uso terapéutico.

Ejemplo VI: Generación de una enzima proteolítica obtenida por ingeniería genética que hidroliza específicamente VEGF humano

Se usó tripsina humana I como armazón proteico para la generación de una enzima proteolítica obtenida por ingeniería genética con alta especificidad para VEGF humano. La identificación de los sitios SDR en tripsina humana

5 I se realizó como se ha descrito anteriormente. Se seleccionaron dos sitios de inserción dentro del armazón para las SDR. Las variantes de proteasa que contienen dos insertos con diferentes secuencias se expresaron en células *Bacillus subtilis*. Las células de proteasa variante se separaron en clones celulares sencillos y las variantes que expresaban proteasa se cribaron como se ha descrito anteriormente. La actividad de las variantes de proteasa se determinó como la velocidad de escisión de VEGF. Se incubaron 4 µg de VEGF165 humano recombinante con 0,18 µg de proteasa purificada en PBS / pH 7,4 a temperatura ambiente. Se tomaron alícuotas en los puntos temporales indicados y se analizaron en un gel de poliacrilamida. La extensión de la escisión se cuantificó por análisis densitométrico de las bandas. La actividad se representa frente al tiempo de incubación en la Figura 18. La escisión específica se controló mediante análisis en gel de poliacrilamida de SDS adicionales.

10 LISTADO DE SECUENCIAS

<110> DIREVO Biotech AG

<120> NUEVAS ENTIDADES BIOLÓGICAS Y USO DE LAS MISMAS

15

<130> 041480wo JH/cw

<160> 96

20 <170> PatentIn versión 3.1

<210> 1

<211> 224

<212> PRT

25 <213> Homo sapiens

<400> 1

ES 2 373 658 T3

```

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

```

```

          180          185          190
Trp Gly Asp Gly Cys Ala Gln Lys Asn Lys Pro Gly Val Tyr Thr Lys
          195          200          205
Val Tyr Asn Tyr Val Lys Trp Ile Lys Asn Thr Ile Ala Ala Asn Ser
          210          215          220

```

<210> 2

<211> 235

<212> PRT

5 <213> secuencia artificial

<220>

<223> variante 1 de tripsina

<400> 2

10

ES 2 373 658 T3

```

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

```

210

215

220

Lys Trp Ile Lys Asn Thr Ile Ala Ala Asn Ser

225

230

235

<210> 3

<211> 235

5 <212> PRT

<213> secuencia artificial

<220>

<223> variante 2 de tripsina

10 <400> 3

ES 2 373 658 T3

```

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

```

225

230

235

<210> 4

<211> 235

5 <212> PRT

<213> secuencia artificial

<220>

<223> variante 3 de tripsina

10 <400> 4

ES 2 373 658 T3

```

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

```

<210> 5
 <211> 259
 <212> PRT
 <213> Homo sapiens

5

<400> 5

ES 2 373 658 T3

Ile Val Glu Gly Ser Asp Ala Glu Ile Gly Met Ser Pro Trp Gln Val
 1 5 10 15
 Met Leu Phe Arg Lys Ser Pro Gln Glu Leu Leu Cys Gly Ala Ser Leu
 20 25 30
 Ile Ser Asp Arg Trp Val Leu Thr Ala Ala His Cys Leu Leu Tyr Pro
 35 40 45
 Pro Trp Asp Lys Asn Phe Thr Glu Asn Asp Leu Leu Val Arg Ile Gly
 50 55 60
 Lys His Ser Arg Thr Arg Tyr Glu Arg Asn Ile Glu Lys Ile Ser Met
 65 70 75 80
 Leu Glu Lys Ile Tyr Ile His Pro Arg Tyr Asn Trp Arg Glu Asn Leu
 85 90 95
 Asp Arg Asp Ile Ala Leu Met Lys Leu Lys Lys Pro Val Ala Phe Ser
 100 105 110
 Asp Tyr Ile His Pro Val Cys Leu Pro Asp Arg Glu Thr Ala Ala Ser
 115 120 125
 Leu Leu Gln Ala Gly Tyr Lys Gly Arg Val Thr Gly Trp Gly Asn Leu
 130 135 140
 Lys Glu Thr Trp Thr Ala Asn Val Gly Lys Gly Gln Pro Ser Val Leu
 145 150 155 160
 Gln Val Val Asn Leu Pro Ile Val Glu Arg Pro Val Cys Lys Asp Ser
 165 170 175
 Thr Arg Ile Arg Ile Thr Asp Asn Met Phe Cys Ala Gly Tyr Lys Pro
 180 185 190
 Asp Glu Gly Lys Arg Gly Asp Ala Cys Glu Gly Asp Ser Gly Gly Pro
 195 200 205
 Phe Val Met Lys Ser Pro Phe Asn Asn Arg Trp Tyr Gln Met Gly Ile
 210 215 220
 Val Ser Trp Gly Glu Gly Cys Asp Arg Asp Gly Lys Tyr Gly Phe Tyr
 225 230 235 240
 Thr His Val Phe Arg Leu Lys Lys Trp Ile Gln Lys Val Ile Asp Gln
 245 250 255
 Phe Gly Glu

<210> 6
 <211> 235
 <212> PRT
 <213> Homo sapiens

5

<400> 6

ES 2 373 658 T3

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

<210> 7
 <211> 275
 5 <212> PRT
 <213> Bacillus subtilis
 <400> 7

ES 2 373 658 T3

Ile Ala His Glu Tyr Ala Gln Ser Val Pro Tyr Gly Ile Ser Gln Ile
1 5 10 15
Lys Ala Pro Ala Leu His Ser Gln Gly Tyr Thr Gly Ser Asn Val Lys
20 25 30
Val Ala Val Ile Asp Ser Gly Ile Asp Ser Ser His Pro Asp Leu Asn
35 40 45
Val Arg Gly Gly Ala Ser Phe Val Pro Ser Glu Thr Asn Pro Tyr Gln
50 55 60
Asp Gly Ser Ser His Gly Thr His Val Ala Gly Thr Ile Ala Ala Leu
65 70 75 80
Asn Asn Ser Ile Gly Val Leu Gly Val Ser Pro Ser Ala Ser Leu Tyr
85 90 95
Ala Val Lys Val Leu Asp Ser Thr Gly Ser Gly Gln Tyr Ser Trp Ile
100 105 110
Ile Asn Gly Ile Glu Trp Ala Ile Ser Asn Asn Met Asp Val Ile Asn
115 120 125
Met Ser Leu Gly Gly Pro Thr Gly Ser Thr Ala Leu Lys Thr Val Val
130 135 140
Asp Lys Ala Val Ser Ser Gly Ile Val Val Ala Ala Ala Ala Gly Asn
145 150 155 160
Glu Gly Ser Ser Gly Ser Thr Ser Thr Val Gly Tyr Pro Ala Lys Tyr
165 170 175
Pro Ser Thr Ile Ala Val Gly Ala Val Asn Ser Ser Asn Gln Arg Ala
180 185 190
Ser Phe Ser Ser Ala Gly Ser Glu Leu Asp Val Met Ala Pro Gly Val
195 200 205
Ser Ile Gln Ser Thr Leu Pro Gly Gly Thr Tyr Gly Ala Tyr Asn Gly
210 215 220
Thr Ser Met Ala Thr Pro His Val Ala Gly Ala Ala Ala Leu Ile Leu
225 230 235 240
Ser Lys His Pro Thr Trp Thr Asn Ala Gln Val Arg Asp Arg Leu Glu
245 250 255
Ser Thr Ala Thr Tyr Leu Gly Asn Ser Phe Tyr Tyr Gly Lys Gly Leu
260 265 270
Ile Asn Val
275

5 <210> 8
<211> 320
<212> PRT
<213> Murinae gen. sp.

<400> 8

ES 2 373 658 T3

Val Ala Lys Arg Arg Ala Lys Arg Asp Val Tyr Gln Glu Pro Thr Asp
 1 5 10 15
 Pro Lys Phe Pro Gln Gln Trp Tyr Leu Ser Gly Val Thr Gln Arg Asp
 20 25 30
 Leu Asn Val Lys Glu Ala Trp Ala Gln Gly Phe Thr Gly His Gly Ile
 35 40 45
 Val Val Ser Ile Leu Asp Asp Gly Ile Glu Lys Asn His Pro Asp Leu
 50 55 60
 Ala Gly Asn Tyr Asp Pro Gly Ala Ser Phe Asp Val Asn Asp Gln Asp
 65 70 75 80
 Pro Asp Pro Gln Pro Arg Tyr Thr Gln Met Asn Asp Asn Arg His Gly
 85 90 95
 Thr Arg Cys Ala Gly Glu Val Ala Ala Val Ala Asn Asn Gly Val Cys
 100 105 110
 Gly Val Gly Val Ala Tyr Asn Ala Arg Ile Gly Gly Val Arg Met Leu
 115 120 125
 Asp Gly Glu Val Thr Asp Ala Val Glu Ala Arg Ser Leu Gly Leu Asn
 130 135 140
 Pro Asn His Ile His Ile Tyr Ser Ala Ser Trp Gly Pro Glu Asp Asp
 145 150 155 160
 Gly Lys Thr Val Asp Gly Pro Ala Arg Leu Ala Glu Glu Ala Phe Phe
 165 170 175
 Arg Gly Val Ser Gln Gly Arg Gly Gly Leu Gly Ser Ile Phe Val Trp
 180 185 190
 Ala Ser Gly Asn Gly Gly Arg Glu His Asp Ser Cys Asn Cys Asp Gly
 195 200 205
 Tyr Thr Asn Ser Ile Tyr Thr Leu Ser Ile Ser Ser Ala Thr Gln Phe
 210 215 220
 Gly Asn Val Pro Trp Tyr Ser Glu Ala Cys Ser Ser Thr Leu Ala Thr
 225 230 235 240
 Thr Tyr Ser Ser Gly Asn Gln Asn Glu Lys Gln Ile Val Thr Thr Asp
 245 250 255
 Leu Arg Gln Lys Cys Thr Glu Ser His Thr Gly Thr Ser Ala Ser Ala
 260 265 270
 Pro Leu Ala Ala Gly Ile Ile Ala Leu Thr Leu Glu Ala Asn Lys Asn
 275 280 285
 Leu Thr Trp Arg Asp Met Gln His Leu Val Val Gln Thr Ser Lys Pro
 290 295 300
 Ala His Leu Asn Ala Asp Asp Trp Ala Thr Asn Gly Val Gly Arg Lys
 305 310 315 320

ES 2 373 658 T3

<210> 9
 <211> 330
 <212> PRT
 <213> Homo sapiens

5

<400> 9

```

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

ES 2 373 658 T3

```

                245                250                255
Tyr Thr Asp Gln Arg Ile Thr Ser Ala Asp Leu His Asn Asp Cys Thr
                260                265                270
Glu Thr His Thr Gly Thr Ser Ala Ser Ala Pro Leu Ala Ala Gly Ile
                275                280                285
Phe Ala Leu Ala Leu Glu Ala Asn Pro Asn Leu Thr Trp Arg Asp Met
                290                295                300
Gln His Leu Val Val Trp Thr Ser Glu Tyr Asp Pro Leu Ala Asn Asn
305                310                315                320
Pro Gly Trp Lys Lys Asn Gly Ala Gly Leu
                325                330

```

- <210> 10
- <211> 297
- 5 <212> PRT
- <213> Homo sapiens
- <400> 10

```

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

```

10

ES 2 373 658 T3

```

                180                185                190
Ser Ile Ser Ser Thr Ala Glu Ser Gly Lys Lys Pro Trp Tyr Leu Glu
                195                200                205
Glu Cys Ser Ser Thr Leu Ala Thr Thr Tyr Ser Ser Gly Glu Ser Tyr
                210                215                220
Asp Lys Lys Ile Ile Thr Thr Asp Leu Arg Gln Arg Cys Thr Asp Asn
225                230                235                240
His Thr Gly Thr Ser Ala Ser Ala Pro Met Ala Ala Gly Ile Ile Ala
                245                250                255
Leu Ala Leu Glu Ala Asn Pro Phe Leu Thr Trp Arg Asp Val Gln His
                260                265                270
Val Ile Val Arg Thr Ser Arg Ala Gly His Leu Asn Ala Asn Asp Trp
                275                280                285
Lys Thr Asn Ala Ala Gly Phe Lys Val
                290                295

```

<210> 11
 <211> 328
 <212> PRT
 <213> Homo sapiens

5

<400> 11

```

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

```

10

ES 2 373 658 T3

145 150 155 160
Asp Lys Ser Gly Ser Val Val Ile Phe Gly Gly Ile Asp Ser Ser Tyr
165 170 175
Tyr Thr Gly Ser Leu Asn Trp Val Pro Val Thr Val Glu Gly Tyr Trp
180 185 190
Gln Ile Thr Val Asp Ser Ile Thr Met Asn Gly Glu Thr Ile Ala Cys
195 200 205
Ala Glu Gly Cys Gln Ala Ile Val Asp Thr Gly Thr Ser Leu Leu Thr
210 215 220
Gly Pro Thr Ser Pro Ile Ala Asn Ile Gln Ser Asp Ile Gly Ala Ser
225 230 235 240
Glu Asn Ser Asp Gly Asp Met Val Val Ser Cys Ser Ala Ile Ser Ser
245 250 255
Leu Pro Asp Ile Val Phe Thr Ile Asn Gly Val Gln Tyr Pro Val Pro
260 265 270
Pro Ser Ala Tyr Ile Leu Gln Ser Glu Gly Ser Cys Ile Ser Gly Phe
275 280 285
Gln Gly Met Asn Val Pro Thr Glu Ser Gly Glu Leu Trp Ile Leu Gly
290 295 300
Asp Val Phe Ile Arg Gln Tyr Phe Thr Val Phe Asp Arg Ala Asn Asn
305 310 315 320
Gln Val Gly Leu Ala Pro Val Ala
325

<210> 12
<211> 358
5 <212> PRT
<213> Homo sapiens

<400> 12

Glu Met Val Asp Asn Leu Arg Gly Lys Ser Gly Gln Gly Tyr Tyr Val
1 5 10 15
Glu Met Thr Val Gly Ser Pro Pro Gln Thr Leu Asn Ile Leu Val Asp
20 25 30
Thr Gly Ser Ser Asn Phe Ala Val Gly Ala Ala Pro His Pro Phe Leu
35 40 45
His Arg Tyr Tyr Gln Arg Gln Leu Ser Ser Thr Tyr Arg Asp Leu Arg
50 55 60
Lys Gly Val Tyr Val Pro Tyr Thr Gln Gly Lys Trp Glu Gly Glu Leu
65 70 75 80
Gly Thr Asp Leu Val Ser Ile Pro His Gly Pro Asn Val Thr Val Arg

10

ES 2 373 658 T3

| | | | | | | | | | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| | | | | 85 | | | | | 90 | | | | 95 | | |
| Ala | Asn | Ile | Ala | Ala | Ile | Thr | Glu | Ser | Asp | Lys | Phe | Phe | Ile | Asn | Gly |
| | | | 100 | | | | | 105 | | | | | 110 | | |
| Ser | Asn | Trp | Glu | Gly | Ile | Leu | Gly | Leu | Ala | Tyr | Ala | Glu | Ile | Ala | Arg |
| | | | 115 | | | | 120 | | | | | 125 | | | |
| Pro | Asp | Asp | Ser | Leu | Glu | Pro | Phe | Phe | Asp | Ser | Leu | Val | Lys | Gln | Thr |
| | | | 130 | | | 135 | | | | | 140 | | | | |
| His | Val | Pro | Asn | Leu | Phe | Ser | Leu | Gln | Leu | Cys | Gly | Ala | Gly | Phe | Pro |
| 145 | | | | 150 | | | | | | 155 | | | | 160 | |
| Leu | Asn | Gln | Ser | Glu | Val | Leu | Ala | Ser | Val | Gly | Gly | Ser | Met | Ile | Ile |
| | | | | 165 | | | | | 170 | | | | | 175 | |
| Gly | Gly | Ile | Asp | His | Ser | Leu | Tyr | Thr | Gly | Ser | Leu | Trp | Tyr | Thr | Pro |
| | | | 180 | | | | | 185 | | | | | 190 | | |
| Ile | Arg | Arg | Glu | Trp | Tyr | Tyr | Glu | Val | Ile | Ile | Val | Arg | Val | Glu | Ile |
| | | | 195 | | | | 200 | | | | | 205 | | | |
| Asn | Gly | Gln | Asp | Leu | Lys | Met | Asp | Cys | Lys | Glu | Tyr | Asn | Tyr | Asp | Lys |
| | | | 210 | | | 215 | | | | | 220 | | | | |
| Ser | Ile | Val | Asp | Ser | Gly | Thr | Thr | Asn | Leu | Arg | Leu | Pro | Lys | Lys | Val |
| 225 | | | | | 230 | | | | | 235 | | | | 240 | |
| Phe | Glu | Ala | Ala | Val | Lys | Ser | Ile | Lys | Ala | Ala | Ser | Ser | Thr | Glu | Lys |
| | | | | 245 | | | | | 250 | | | | | 255 | |
| Phe | Pro | Asp | Gly | Phe | Trp | Leu | Gly | Glu | Gln | Leu | Val | Cys | Trp | Gln | Ala |
| | | | 260 | | | | 265 | | | | | | 270 | | |
| Gly | Thr | Thr | Pro | Trp | Asn | Ile | Phe | Pro | Val | Ile | Ser | Leu | Tyr | Leu | Met |
| | | | 275 | | | | 280 | | | | | 285 | | | |
| Gly | Glu | Val | Thr | Asn | Gln | Ser | Phe | Arg | Ile | Thr | Ile | Leu | Pro | Gln | Gln |
| | | | 290 | | | 295 | | | | | 300 | | | | |
| Tyr | Leu | Arg | Pro | Val | Glu | Asp | Val | Ala | Thr | Ser | Gln | Asp | Asp | Cys | Tyr |
| 305 | | | | | 310 | | | | | 315 | | | | 320 | |
| Lys | Phe | Ala | Ile | Ser | Gln | Ser | Ser | Thr | Gly | Thr | Val | Met | Gly | Ala | Val |
| | | | | 325 | | | | | 330 | | | | 335 | | |
| Ile | Met | Glu | Gly | Phe | Tyr | Val | Val | Phe | Asp | Arg | Ala | Arg | Lys | Arg | Ile |
| | | | 340 | | | | | 345 | | | | | 350 | | |
| Gly | Phe | Ala | Val | Ser | Ala | | | | | | | | | | |
| | | | 355 | | | | | | | | | | | | |

<210> 13
 <211> 351
 5 <212> PRT
 <213> Homo sapiens

<400> 13

ES 2 373 658 T3

Pro Ala Val Thr Glu Gly Pro Ile Pro Glu Val Leu Lys Asn Tyr Met
 1 5 10 15
 Asp Ala Gln Tyr Tyr Gly Glu Ile Gly Ile Gly Thr Pro Pro Gln Cys
 20 25 30
 Phe Thr Val Val Phe Asp Thr Gly Ser Ser Asn Leu Trp Val Pro Ser
 35 40 45
 Ile His Cys Lys Leu Leu Asp Ile Ala Cys Trp Ile His His Lys Tyr
 50 55 60
 Asn Ser Asp Lys Ser Ser Thr Tyr Val Lys Asn Gly Thr Ser Phe Asp
 65 70 75 80
 Ile His Tyr Gly Ser Gly Ser Leu Ser Gly Tyr Leu Ser Gln Asp Thr
 85 90 95
 Val Ser Val Pro Cys Gln Ser Ala Ser Ser Ala Ser Ala Leu Gly Gly
 100 105 110
 Val Lys Val Glu Arg Gln Val Phe Gly Glu Ala Thr Lys Gln Pro Gly
 115 120 125
 Ile Thr Phe Ile Ala Ala Lys Phe Asp Gly Ile Leu Gly Met Ala Tyr
 130 135 140
 Pro Arg Ile Ser Val Asn Asn Val Leu Pro Val Phe Asp Asn Leu Met
 145 150 155 160
 Gln Gln Lys Leu Val Asp Gln Asn Ile Phe Ser Phe Tyr Leu Ser Arg
 165 170 175
 Asp Pro Asp Ala Gln Pro Gly Gly Glu Leu Met Leu Gly Gly Thr Asp
 180 185 190
 Ser Lys Tyr Tyr Lys Gly Ser Leu Ser Tyr Leu Asn Val Thr Arg Lys
 195 200 205
 Ala Tyr Trp Gln Val His Leu Asp Gln Val Glu Val Ala Ser Gly Leu
 210 215 220
 Thr Leu Cys Lys Glu Gly Cys Glu Ala Ile Val Asp Thr Gly Thr Ser
 225 230 235 240
 Leu Met Val Gly Pro Val Asp Glu Val Arg Glu Leu Gln Lys Ala Ile
 245 250 255
 Gly Ala Val Pro Leu Ile Gln Gly Glu Tyr Met Ile Pro Cys Glu Lys
 260 265 270
 Val Ser Thr Leu Pro Ala Ile Thr Leu Lys Leu Gly Gly Lys Gly Tyr
 275 280 285
 Lys Leu Ser Pro Glu Asp Tyr Thr Leu Lys Val Ser Gln Ala Gly Lys
 290 295 300
 Thr Leu Cys Leu Ser Gly Phe Met Gly Met Asp Ile Pro Pro Pro Ser
 305 310 315 320
 Gly Pro Leu Trp Ile Leu Gly Asp Val Phe Ile Gly Arg Tyr Tyr Thr

ES 2 373 658 T3

325 330 335
 Val Phe Asp Arg Asp Asn Asn Arg Val Gly Phe Ala Glu Ala Ala
 340 345 350

5 <210> 14
 <211> 305
 <212> PRT
 <213> Homo sapiens
 <400> 14

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

10

ES 2 373 658 T3

245 250 255
 Asp Leu Glu Ile Met Gln Ile Leu Thr Arg Val Asn Asp Arg Val Ala
 260 265 270
 Arg His Phe Glu Ser Gln Ser Asp Asp Pro His Phe His Glu Lys Lys
 275 280 285
 Gln Ile Pro Cys Val Val Ser Met Leu Thr Lys Glu Leu Tyr Phe Ser
 290 295 300
 Gln
 305

<210> 15

<211> 262

5 <212> PRT

<213> Streptomyces sp. K15

<400> 15

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

10

ES 2 373 658 T3

```

                195                200                205
Ala Ile Gly Val Lys Thr Gly Ser Gly Pro Glu Ala Lys Tyr Cys Leu
      210                215                220
Val Phe Ala Ala Thr Arg Gly Gly Lys Thr Val Ile Gly Thr Val Leu
225                230                235                240
Ala Ser Thr Ser Ile Pro Ala Arg Glu Ser Asp Ala Thr Lys Ile Met
                245                250                255
Asn Tyr Gly Phe Ala Leu
                260

```

<210> 16
 <211> 256
 5 <212> PRT
 <213> Citomegalovirus humano

<400> 16

```

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

```

10

ES 2 373 658 T3

```

                195                200                205
Ala Ser Gly Asp Pro Phe Arg Ser Asp Ser Tyr Gly Leu Leu Gly Asn
    210                215                220
Ser Val Asp Ala Leu Tyr Ile Arg Glu Arg Leu Pro Lys Leu Arg Tyr
225                230                235                240
Asp Lys Gln Leu Val Gly Val Thr Glu Arg Glu Ser Tyr Val Lys Ala
                245                250                255

```

<210> 17
 <211> 248
 5 <212> PRT
 <213> Escherichia coli

 <400> 17

```

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

```

ES 2 373 658 T3

210 215 220
 Trp Met Ser Phe Asp Lys Gln Glu Gly Glu Trp Pro Thr Gly Leu Arg
 225 230 235 240
 Leu Ser Arg Ile Gly Gly Ile His
 245

- <210> 18
- <211> 317
- 5 <212> PRT
- <213> *Serratia marcescens*
- <400> 18

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

ES 2 373 658 T3

```

225                230                235                240
Phe Leu Glu Ser Asp Asp Gln Leu Leu Arg Asn Val Pro Leu Ile Arg
                245                250                255
His Ile Pro Ala Val Ile Val His Gly Arg Tyr Asp Met Ala Cys Gln
                260                265                270
Val Gln Asn Ala Trp Asp Leu Ala Lys Ala Trp Pro Glu Ala Glu Leu
                275                280                285
His Ile Val Glu Gly Ala Gly His Ser Tyr Asp Glu Pro Gly Ile Leu
                290                295                300
His Gln Leu Met Ile Ala Thr Asp Arg Phe Ala Gly Lys
305                310                315

```

<210> 19
 <211> 229
 5 <212> PRT
 <213> Escherichia coli

<400> 19

```

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

```

ES 2 373 658 T3

```

                180                185                190
Gly Asn Trp Ile Gln Val Ser Asn Gly Gln Ala Val Leu Gly Gly Pro
      195                200                205
Asn Thr Thr Trp Val Phe Lys Ala Gly Glu Glu Ala Val Ala Leu Glu
      210                215                220
Ala Gly His Arg Phe
225

```

- <210> 20
- <211> 99
- 5 <212> PRT
- <213> virus de la inmunodeficiencia humana
- <400> 20

```

Pro Gln Ile Thr Leu Trp Gln Arg Pro Leu Val Thr Val Lys Ile Gly
1                5                10                15
Gly Gln Leu Arg Glu Ala Leu Leu Asp Thr Gly Ala Asp Asp Thr Val
      20                25                30
Leu Glu Asp Ile Asn Leu Pro Gly Lys Trp Lys Pro Lys Met Ile Gly
      35                40                45
Gly Ile Gly Gly Phe Ile Lys Val Arg Gln Tyr Asp Gln Ile Leu Ile
      50                55                60
Glu Ile Cys Gly Lys Lys Ala Ile Gly Thr Val Leu Val Gly Pro Thr
65                70                75                80
Pro Val Asn Ile Ile Gly Arg Asn Met Leu Thr Gln Ile Gly Cys Thr
      85                90                95
Leu Asn Phe

```

- 10
- <210> 21
- <211> 297
- <212> PRT
- 15 <213> Escherichia coli
- <400> 21

```

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

```

ES 2 373 658 T3

| | | | | | | | | | | | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|--|-----|
| | 35 | | | | | | 40 | | | | | | | 45 | | | |
| Ala | Ala | Ile | Ile | Lys | Gly | Ala | Ile | Asn | Trp | Asp | Leu | Met | Pro | Gln | Ile | | |
| | 50 | | | | | | 55 | | | | | | | 60 | | | |
| Ser | Ile | Gly | Ala | Ala | Gly | Trp | Thr | Thr | Leu | Gly | Ser | Arg | Gly | Gly | Asn | | |
| 65 | | | | | | | 70 | | | | | | | 75 | | | 80 |
| Met | Val | Asp | Gln | Asp | Trp | Met | Asp | Ser | Ser | Asn | Pro | Gly | Thr | Trp | Thr | | |
| | | | | | | | 85 | | | | | | | 90 | | | 95 |
| Asp | Glu | Ala | Arg | His | Pro | Asp | Thr | Gln | Leu | Asn | Tyr | Ala | Asn | Glu | Phe | | |
| | | | | | | | 100 | | | | | | | 105 | | | 110 |
| Asp | Leu | Asn | Ile | Lys | Gly | Trp | Leu | Leu | Asn | Glu | Pro | Asn | Tyr | Arg | Leu | | |
| | | | | | | | 115 | | | | | | | 120 | | | 125 |
| Gly | Leu | Met | Ala | Gly | Tyr | Gln | Glu | Ser | Arg | Tyr | Ser | Phe | Thr | Ala | Arg | | |
| | | | | | | | 130 | | | | | | | 135 | | | 140 |
| Gly | Gly | Ser | Tyr | Ile | Tyr | Ser | Ser | Glu | Glu | Gly | Phe | Arg | Asp | Asp | Ile | | |
| 145 | | | | | | | 150 | | | | | | | 155 | | | 160 |
| Gly | Ser | Phe | Pro | Asn | Gly | Glu | Arg | Ala | Ile | Gly | Tyr | Lys | Gln | Arg | Phe | | |
| | | | | | | | 165 | | | | | | | 170 | | | 175 |
| Lys | Met | Pro | Tyr | Ile | Gly | Leu | Thr | Gly | Ser | Tyr | Arg | Tyr | Glu | Asp | Phe | | |
| | | | | | | | 180 | | | | | | | 185 | | | 190 |
| Glu | Leu | Gly | Gly | Thr | Phe | Lys | Tyr | Ser | Gly | Trp | Val | Glu | Ser | Ser | Asp | | |
| | | | | | | | 195 | | | | | | | 200 | | | 205 |
| Asn | Asp | Glu | His | Tyr | Asp | Pro | Lys | Gly | Arg | Ile | Thr | Tyr | Arg | Ser | Lys | | |
| | | | | | | | 210 | | | | | | | 215 | | | 220 |
| Val | Lys | Asp | Gln | Asn | Tyr | Tyr | Ser | Val | Ala | Val | Asn | Ala | Gly | Tyr | Tyr | | |
| 225 | | | | | | | 230 | | | | | | | 235 | | | 240 |
| Val | Thr | Pro | Asn | Ala | Lys | Val | Tyr | Val | Glu | Gly | Ala | Trp | Asn | Arg | Val | | |
| | | | | | | | 245 | | | | | | | 250 | | | 255 |
| Thr | Asn | Lys | Lys | Gly | Asn | Thr | Ser | Leu | Tyr | Asp | His | Asn | Asn | Asn | Thr | | |
| | | | | | | | 260 | | | | | | | 265 | | | 270 |
| Ser | Asp | Tyr | Ser | Lys | Asn | Gly | Ala | Gly | Ile | Glu | Asn | Tyr | Asn | Phe | Ile | | |
| | | | | | | | 275 | | | | | | | 280 | | | 285 |
| Thr | Thr | Ala | Gly | Leu | Lys | Tyr | Thr | Phe | | | | | | | | | |
| | | | | | | | 290 | | | | | | | 295 | | | |

5 <210> 22
 <211> 212
 <212> PRT
 <213> Carica papaya

<400> 22

ES 2 373 658 T3

Ile Pro Glu Tyr Val Asp Trp Arg Gln Lys Gly Ala Val Thr Pro Val

```

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

```

<210> 23
 <211> 699
 <212> PRT
 <213> Homo sapiens

5

<400> 23

ES 2 373 658 T3

Ala Gly Ile Ala Ala Lys Leu Ala Lys Asp Arg Glu Ala Ala Glu Gly
1 5 10 15
Leu Gly Ser His Glu Arg Ala Ile Lys Tyr Leu Asn Gln Asp Tyr Glu
 20 25 30
Ala Leu Arg Asn Glu Cys Leu Glu Ala Gly Thr Leu Phe Gln Asp Pro
 35 40 45
Ser Phe Pro Ala Ile Pro Ser Ala Leu Gly Phe Lys Glu Leu Gly Pro

ES 2 373 658 T3

| | | | | | | | | | | | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|--|--|
| 50 | | | | | | 55 | | | | | | 60 | | | | | |
| Tyr | Ser | Ser | Lys | Thr | Arg | Gly | Met | Arg | Trp | Lys | Arg | Pro | Thr | Glu | Ile | | |
| 65 | | | | | 70 | | | | | 75 | | | | 80 | | | |
| Cys | Ala | Asp | Pro | Gln | Phe | Ile | Ile | Gly | Gly | Ala | Thr | Arg | Thr | Asp | Ile | | |
| | | | | 85 | | | | | 90 | | | | | 95 | | | |
| Cys | Gln | Gly | Ala | Leu | Gly | Asp | Cys | Trp | Leu | Leu | Ala | Ala | Ile | Ala | Ser | | |
| | | | 100 | | | | | 105 | | | | | 110 | | | | |
| Leu | Thr | Leu | Asn | Glu | Glu | Ile | Leu | Ala | Arg | Val | Val | Pro | Leu | Asn | Gln | | |
| | | 115 | | | | | 120 | | | | | 125 | | | | | |
| Ser | Phe | Gln | Glu | Asn | Tyr | Ala | Gly | Ile | Phe | His | Phe | Gln | Phe | Trp | Gln | | |
| | 130 | | | | | 135 | | | | | 140 | | | | | | |
| Tyr | Gly | Glu | Trp | Val | Glu | Val | Val | Val | Asp | Asp | Arg | Leu | Pro | Thr | Lys | | |
| 145 | | | | | 150 | | | | | 155 | | | | | 160 | | |
| Asp | Gly | Glu | Leu | Leu | Phe | Val | His | Ser | Ala | Glu | Gly | Ser | Glu | Phe | Trp | | |
| | | | | 165 | | | | | 170 | | | | | 175 | | | |
| Ser | Ala | Leu | Leu | Glu | Lys | Ala | Tyr | Ala | Lys | Ile | Asn | Gly | Cys | Tyr | Glu | | |
| | | | 180 | | | | | 185 | | | | | 190 | | | | |
| Ala | Leu | Ser | Gly | Gly | Ala | Thr | Thr | Glu | Gly | Phe | Glu | Asp | Phe | Thr | Gly | | |
| | | 195 | | | | | 200 | | | | | 205 | | | | | |
| Gly | Ile | Ala | Glu | Trp | Tyr | Glu | Leu | Lys | Lys | Pro | Pro | Pro | Asn | Leu | Phe | | |
| | 210 | | | | | 215 | | | | | | 220 | | | | | |
| Lys | Ile | Ile | Gln | Lys | Ala | Leu | Gln | Lys | Gly | Ser | Leu | Leu | Gly | Cys | Ser | | |
| 225 | | | | | 230 | | | | | 235 | | | | 240 | | | |
| Ile | Asp | Ile | Thr | Ser | Ala | Ala | Asp | Ser | Glu | Ala | Ile | Thr | Phe | Gln | Lys | | |
| | | | | 245 | | | | | 250 | | | | | 255 | | | |
| Leu | Val | Lys | Gly | His | Ala | Tyr | Ser | Val | Thr | Gly | Ala | Glu | Glu | Val | Glu | | |
| | | | 260 | | | | | 265 | | | | | 270 | | | | |
| Ser | Asn | Gly | Ser | Leu | Gln | Lys | Leu | Ile | Arg | Ile | Arg | Asn | Pro | Trp | Gly | | |
| | | 275 | | | | | 280 | | | | | 285 | | | | | |
| Glu | Val | Glu | Trp | Thr | Gly | Arg | Trp | Asn | Asp | Asn | Cys | Pro | Ser | Trp | Asn | | |
| | 290 | | | | | 295 | | | | | 300 | | | | | | |
| Thr | Ile | Asp | Pro | Glu | Glu | Arg | Glu | Arg | Leu | Thr | Arg | Arg | His | Glu | Asp | | |
| 305 | | | | | 310 | | | | | 315 | | | | 320 | | | |
| Gly | Glu | Phe | Trp | Met | Ser | Phe | Ser | Asp | Phe | Leu | Arg | His | Tyr | Ser | Arg | | |
| | | | | 325 | | | | | 330 | | | | | 335 | | | |
| Leu | Glu | Ile | Cys | Asn | Leu | Thr | Pro | Asp | Thr | Leu | Thr | Ser | Asp | Thr | Tyr | | |
| | | | 340 | | | | | 345 | | | | | 350 | | | | |
| Lys | Lys | Trp | Lys | Leu | Thr | Lys | Met | Asp | Gly | Asn | Trp | Arg | Arg | Gly | Ser | | |
| | | 355 | | | | | 360 | | | | | 365 | | | | | |
| Thr | Ala | Gly | Gly | Cys | Arg | Asn | Tyr | Pro | Asn | Thr | Phe | Trp | Met | Asn | Pro | | |
| | 370 | | | | | 375 | | | | | 380 | | | | | | |
| Gln | Tyr | Leu | Ile | Lys | Leu | Glu | Glu | Glu | Asp | Glu | Asp | Glu | Glu | Asp | Gly | | |

ES 2 373 658 T3

```

385              390              395              400
Glu Ser Gly Cys Thr Phe Leu Val Gly Leu Ile Gln Lys His Arg Arg
              405              410              415
Arg Gln Arg Lys Met Gly Glu Asp Met His Thr Ile Gly Phe Gly Ile
              420              425              430
Tyr Glu Val Pro Glu Glu Leu Ser Gly Gln Thr Asn Ile His Leu Ser
              435              440              445
Lys Asn Phe Phe Leu Thr Asn Arg Ala Arg Glu Arg Ser Asp Thr Phe
              450              455              460
Ile Asn Leu Arg Glu Val Leu Asn Arg Phe Lys Leu Pro Pro Gly Glu
465              470              475              480
Tyr Ile Leu Val Pro Ser Thr Phe Glu Pro Asn Lys Asp Gly Asp Phe
              485              490              495
Cys Ile Arg Val Phe Ser Glu Lys Lys Ala Asp Tyr Gln Ala Val Asp
              500              505              510
Asp Glu Ile Glu Ala Asn Leu Glu Glu Phe Asp Ile Ser Glu Asp Asp
              515              520              525
Ile Asp Asp Gly Val Arg Arg Leu Phe Ala Gln Leu Ala Gly Glu Asp
              530              535              540
Ala Glu Ile Ser Ala Phe Glu Leu Gln Thr Ile Leu Arg Arg Val Leu
545              550              555              560
Ala Lys Arg Gln Asp Ile Lys Ser Asp Gly Phe Ser Ile Glu Thr Cys
              565              570              575
Lys Ile Met Val Asp Met Leu Asp Ser Asp Gly Ser Gly Lys Leu Gly
              580              585              590
Leu Lys Glu Phe Tyr Ile Leu Trp Thr Lys Ile Gln Lys Tyr Gln Lys
              595              600              605
Ile Tyr Arg Glu Ile Asp Val Asp Arg Ser Gly Thr Met Asn Ser Tyr
              610              615              620
Glu Met Arg Lys Ala Leu Glu Glu Ala Gly Phe Lys Met Pro Cys Gln
625              630              635              640
Leu His Gln Val Ile Val Ala Arg Phe Ala Asp Asp Gln Leu Ile Ile
              645              650              655
Asp Phe Asp Asn Phe Val Arg Cys Leu Val Arg Leu Glu Thr Leu Phe
              660              665              670
Lys Ile Phe Lys Gln Leu Asp Pro Glu Asn Thr Gly Thr Ile Glu Leu
              675              680              685
Asp Leu Ile Ser Trp Leu Cys Phe Ser Val Leu
              690              695

```

<210> 24

<211> 221

<212> PRT

<213> Virus del grabado de tabaco

5

ES 2 373 658 T3

<400> 24

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

5

<210> 25
 <211> 371
 <212> PRT
 <213> Streptococcus pyogenes

10

<400> 25

Asp Gln Asn Phe Ala Arg Asn Glu Lys Glu Ala Lys Asp Ser Ala Ile
 1 5 10 15

ES 2 373 658 T3

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

ES 2 373 658 T3

Gly Gly Gly Ala Gly Gly Phe Asn Gly Tyr Gln Ser Ala Val Val Gly
 355 360 365
 Ile Lys Pro
 370

<210> 26
 <211> 353
 5 <212> PRT
 <213> Homo sapiens
 <400> 26

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

ES 2 373 658 T3

Tyr Ile Leu His Ala Val Leu Val His Ser Gly Asp Asn His Gly Gly
 245 250 255
 His Tyr Val Val Tyr Leu Asn Pro Lys Gly Asp Gly Lys Trp Cys Lys
 260 265 270
 Phe Asp Asp Asp Val Val Ser Arg Cys Thr Lys Glu Glu Ala Ile Glu
 275 280 285
 His Asn Tyr Gly Gly His Asp Asp Asp Leu Ser Val Arg His Cys Thr
 290 295 300
 Asn Ala Tyr Met Leu Val Tyr Ile Arg Glu Ser Lys Leu Ser Glu Val
 305 310 315 320
 Leu Gln Ala Val Thr Asp His Asp Ile Pro Gln Gln Leu Val Glu Arg
 325 330 335
 Leu Gln Glu Glu Lys Arg Ile Glu Ala Gln Lys Arg Lys Glu Arg Gln
 340 345 350
 Glu

- <210> 27
- <211> 174
- 5 <212> PRT
- <213> Staphylococcus aureus
- <400> 27

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

10

ES 2 373 658 T3

Asn Gly Phe Met Thr Gln Asp Ala Lys Asn Asn Val Ile Pro Val Ser
 145 150 155 160
 Asn Gly Asp His Tyr Gln Trp Tyr Ser Ser Ile Tyr Gly Tyr
 165 170

<210> 28
 <211> 221
 5 <212> PRT
 <213> *Saccharomyces cerevisiae*
 <400> 28

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

10
 <210> 29
 <211> 166
 <212> PRT
 15 <213> *Pyrococcus horikoshii*

ES 2 373 658 T3

<400> 29

```

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

```

5 <210> 30
 <211> 316
 <212> PRT
 <213> Bacillus thermoproteolyticus

10 <400> 30

```

Ile Thr Gly Thr Ser Thr Val Gly Val Gly Arg Gly Val Leu Gly Asp
1           5           10           15
Gln Lys Asn Ile Asn Thr Thr Tyr Ser Thr Tyr Tyr Tyr Leu Gln Asp
           20           25           30
Asn Thr Arg Gly Asp Gly Ile Phe Thr Tyr Asp Ala Lys Tyr Arg Thr
           35           40           45

```

ES 2 373 658 T3

Thr Leu Pro Gly Ser Leu Trp Ala Asp Ala Asp Asn Gln Phe Phe Ala
 50 55 60
 Ser Tyr Asp Ala Pro Ala Val Asp Ala His Tyr Tyr Ala Gly Val Thr
 65 70 75 80
 Tyr Asp Tyr Tyr Lys Asn Val His Asn Arg Leu Ser Tyr Asp Gly Asn
 85 90 95
 Asn Ala Ala Ile Arg Ser Ser Val His Tyr Ser Gln Gly Tyr Asn Asn
 100 105 110
 Ala Phe Trp Asn Gly Ser Glu Met Val Tyr Gly Asp Gly Asp Gly Gln
 115 120 125
 Thr Phe Ile Pro Leu Ser Gly Gly Ile Asp Val Val Ala His Glu Leu
 130 135 140
 Thr His Ala Val Thr Asp Tyr Thr Ala Gly Leu Ile Tyr Gln Asn Glu
 145 150 155 160
 Ser Gly Ala Ile Asn Glu Ala Ile Ser Asp Ile Phe Gly Thr Leu Val
 165 170 175
 Glu Phe Tyr Ala Asn Lys Asn Pro Asp Trp Glu Ile Gly Glu Asp Val
 180 185 190
 Tyr Thr Pro Gly Ile Ser Gly Asp Ser Leu Arg Ser Met Ser Asp Pro
 195 200 205
 Ala Lys Tyr Gly Asp Pro Asp His Tyr Ser Lys Arg Tyr Thr Gly Thr
 210 215 220
 Gln Asp Asn Gly Gly Val His Ile Asn Ser Gly Ile Ile Asn Lys Ala
 225 230 235 240
 Ala Tyr Leu Ile Ser Gln Gly Gly Thr His Tyr Gly Val Ser Val Val
 245 250 255
 Gly Ile Gly Arg Asp Lys Leu Gly Lys Ile Phe Tyr Arg Ala Leu Thr
 260 265 270
 Gln Tyr Leu Thr Pro Thr Ser Asn Phe Ser Gln Leu Arg Ala Ala Ala
 275 280 285
 Val Gln Ser Ala Thr Asp Leu Tyr Gly Ser Thr Ser Gln Glu Val Ala
 290 295 300
 Ser Val Lys Gln Ala Phe Asp Ala Val Gly Val Lys
 305 310 315

<210> 31
 <211> 169
 <212> PRT
 <213> Homo sapiens

5

<400> 31

ES 2 373 658 T3

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

<210> 32
 <211> 496
 5 <212> PRT
 <213> Homo sapiens

<400> 32

Gln Tyr Ser Pro Asn Thr Gln Gln Gly Arg Thr Ser Ile Val His Leu
 1 5 10 15
 Phe Glu Trp Arg Trp Val Asp Ile Ala Leu Glu Cys Glu Arg Tyr Leu
 20 25 30
 Ala Pro Lys Gly Phe Gly Gly Val Gln Val Ser Pro Pro Asn Glu Asn
 35 40 45
 Val Ala Ile Tyr Asn Pro Phe Arg Pro Trp Trp Glu Arg Tyr Gln Pro
 50 55 60
 Val Ser Tyr Lys Leu Cys Thr Arg Ser Gly Asn Glu Asp Glu Phe Arg
 65 70 75 80
 Asn Met Val Thr Arg Cys Asn Asn Val Gly Val Arg Ile Tyr Val Asp
 85 90 95

10

ES 2 373 658 T3

Ala Val Ile Asn His Met Cys Gly Asn Ala Val Ser Ala Gly Thr Ser
 100 105 110
 Ser Thr Cys Gly Ser Tyr Phe Asn Pro Gly Ser Arg Asp Phe Pro Ala
 115 120 125
 Val Pro Tyr Ser Gly Trp Asp Phe Asn Asp Gly Lys Cys Lys Thr Gly
 130 135 140
 Ser Gly Asp Ile Glu Asn Tyr Asn Asp Ala Thr Gln Val Arg Asp Cys
 145 150 155 160
 Arg Leu Thr Gly Leu Leu Asp Leu Ala Leu Glu Lys Asp Tyr Val Arg
 165 170 175
 Ser Lys Ile Ala Glu Tyr Met Asn His Leu Ile Asp Ile Gly Val Ala
 180 185 190
 Gly Phe Arg Leu Asp Ala Ser Lys His Met Trp Pro Gly Asp Ile Lys
 195 200 205
 Ala Ile Leu Asp Lys Leu His Asn Leu Asn Ser Asn Trp Phe Pro Ala
 210 215 220
 Gly Ser Lys Pro Phe Ile Tyr Gln Glu Val Ile Asp Leu Gly Gly Glu
 225 230 235 240
 Pro Ile Lys Ser Ser Asp Tyr Phe Gly Asn Gly Arg Val Thr Glu Phe
 245 250 255
 Lys Tyr Gly Ala Lys Leu Gly Thr Val Ile Arg Lys Trp Asn Gly Glu
 260 265 270
 Lys Met Ser Tyr Leu Lys Asn Trp Gly Glu Gly Trp Gly Phe Val Pro
 275 280 285
 Ser Asp Arg Ala Leu Val Phe Val Asp Asn His Asp Asn Gln Arg Gly
 290 295 300
 His Gly Ala Gly Gly Ala Ser Ile Leu Thr Phe Trp Asp Ala Arg Leu
 305 310 315 320
 Tyr Lys Met Ala Val Gly Phe Met Leu Ala His Pro Tyr Gly Phe Thr
 325 330 335
 Arg Val Met Ser Ser Tyr Arg Trp Pro Arg Gln Phe Gln Asn Gly Asn
 340 345 350
 Asp Val Asn Asp Trp Val Gly Pro Pro Asn Asn Asn Gly Val Ile Lys
 355 360 365
 Glu Val Thr Ile Asn Pro Asp Thr Thr Cys Gly Asn Asp Trp Val Cys
 370 375 380
 Glu His Arg Trp Arg Gln Ile Arg Asn Met Val Ile Phe Arg Asn Val
 385 390 395 400
 Val Asp Gly Gln Pro Phe Thr Asn Trp Tyr Asp Asn Gly Ser Asn Gln
 405 410 415
 Val Ala Phe Gly Arg Gly Asn Arg Gly Phe Ile Val Phe Asn Asn Asp
 420 425 430

ES 2 373 658 T3

Asp Trp Ser Phe Ser Leu Thr Leu Gln Thr Gly Leu Pro Ala Gly Thr
 435 440 445
 Tyr Cys Asp Val Ile Ser Gly Asp Lys Ile Asn Gly Asn Cys Thr Gly
 450 455 460
 Ile Lys Ile Tyr Val Ser Asp Asp Gly Lys Ala His Phe Ser Ile Ser
 465 470 475 480
 Asn Ser Ala Glu Asp Pro Phe Ile Ala Ile His Ala Glu Ser Lys Leu
 485 490 495

<210> 33
 <211> 370
 5 <212> PRT
 <213> Trichoderma reesei
 <400> 33

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

10

ES 2 373 658 T3

Thr Pro His Ser Cys Thr Ala Thr Ala Cys Asp Ser Ala Gly Cys Gly
 210 215 220
 Phe Asn Pro Tyr Gly Ser Gly Tyr Lys Ser Tyr Tyr Gly Pro Gly Asp
 225 230 235 240
 Thr Val Asp Thr Ser Lys Thr Phe Thr Ile Ile Thr Gln Phe Asn Thr
 245 250 255
 Asp Asn Gly Ser Pro Ser Gly Asn Leu Val Ser Ile Thr Arg Lys Tyr
 260 265 270
 Gln Gln Asn Gly Val Asp Ile Pro Ser Ala Gln Pro Gly Gly Asp Thr
 275 280 285
 Ile Ser Ser Cys Pro Ser Ala Ser Ala Tyr Gly Gly Leu Ala Thr Met
 290 295 300
 Gly Lys Ala Leu Ser Ser Gly Met Val Leu Val Phe Ser Ile Trp Asn
 305 310 315 320
 Asp Asn Ser Gln Tyr Met Asn Trp Leu Asp Ser Gly Asn Ala Gly Pro
 325 330 335
 Cys Ser Ser Thr Glu Gly Asn Pro Ser Asn Ile Leu Ala Asn Asn Pro
 340 345 350
 Asn Thr His Val Val Phe Ser Asn Ile Arg Trp Gly Asp Ile Gly Ser
 355 360 365
 Thr Thr
 370

<210> 34
 <211> 223
 5 <212> PRT
 <213> Aspergillus niger

<400> 34

Gln Thr Met Cys Ser Gln Tyr Asp Ser Ala Ser Ser Pro Pro Tyr Ser
 1 5 10 15
 Val Asn Gln Asn Leu Trp Gly Glu Tyr Gln Gly Thr Gly Ser Gln Cys
 20 25 30
 Val Tyr Val Asp Lys Leu Ser Ser Ser Gly Ala Ser Trp His Thr Glu
 35 40 45
 Trp Thr Trp Ser Gly Gly Glu Gly Thr Val Lys Ser Tyr Ser Asn Ser
 50 55 60
 Gly Val Thr Phe Asn Lys Lys Leu Val Ser Asp Val Ser Ser Ile Pro
 65 70 75 80
 Thr Ser Val Glu Trp Lys Gln Asp Asn Thr Asn Val Asn Ala Asp Val
 85 90 95

ES 2 373 658 T3

Ala Tyr Asp Leu Phe Thr Ala Ala Asn Val Asp His Ala Thr Ser Ser
 100 105 110
 Gly Asp Tyr Glu Leu Met Ile Trp Leu Ala Arg Tyr Gly Asn Ile Gln
 115 120 125
 Pro Ile Gly Lys Gln Ile Ala Thr Ala Thr Val Gly Gly Lys Ser Trp
 130 135 140
 Glu Val Trp Tyr Gly Ser Thr Thr Gln Ala Gly Ala Glu Gln Arg Thr
 145 150 155 160
 Tyr Ser Phe Val Ser Glu Ser Pro Ile Asn Ser Tyr Ser Gly Asp Ile
 165 170 175
 Asn Ala Phe Phe Ser Tyr Leu Thr Gln Asn Gln Gly Phe Pro Ala Ser
 180 185 190
 Ser Gln Tyr Leu Ile Asn Leu Gln Phe Gly Thr Glu Ala Phe Thr Gly
 195 200 205
 Gly Pro Ala Thr Phe Thr Val Asp Asn Trp Thr Ala Ser Val Asn
 210 215 220

<210> 35
 <211> 184
 <212> PRT
 <213> Aspergillus niger
 <400> 35

5

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

10

ES 2 373 658 T3

Val Ala Asn His Phe Asn Phe Trp Ala His His Gly Phe Gly Asn Ser
 145 150 155 160
 Asp Phe Asn Tyr Gln Val Val Ala Val Glu Ala Trp Ser Gly Ala Gly
 165 170 175
 Ser Ala Ser Val Thr Ile Ser Ser
 180

<210> 36
 <211> 313
 5 <212> PRT
 <213> Streptomyces lividans
 <400> 36

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

ES 2 373 658 T3

Phe Ala Ala Leu Gly Val Asp Val Ala Ile Thr Glu Leu Asp Ile Gln
 225 230 235 240
 Gly Ala Pro Ala Ser Thr Tyr Ala Asn Val Thr Asn Asp Cys Leu Ala
 245 250 255
 Val Ser Arg Cys Leu Gly Ile Thr Val Trp Gly Val Arg Asp Ser Asp
 260 265 270
 Ser Trp Arg Ser Glu Gln Thr Pro Leu Leu Phe Asn Asn Asp Gly Ser
 275 280 285
 Lys Lys Ala Ala Tyr Thr Ala Val Leu Asp Ala Leu Asn Gly Gly Ala
 290 295 300
 Ser Ser Glu Pro Pro Ala Asp Gly Gly
 305 310

<210> 37

<211> 362

5 <212> PRT

<213> Aspergillus niger

<400> 37

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

10

ES 2 373 658 T3

His Asn Thr Asp Ala Phe Asp Val Gly Asn Ser Val Gly Val Asn Ile
 180 185 190
 Ile Lys Pro Trp Val His Asn Gln Asp Asp Cys Leu Ala Val Asn Ser
 195 200 205
 Gly Glu Asn Ile Trp Phe Thr Gly Gly Thr Cys Ile Gly Gly His Gly
 210 215 220
 Leu Ser Ile Gly Ser Val Gly Asp Arg Ser Asn Asn Val Val Lys Asn
 225 230 235 240
 Val Thr Ile Glu His Ser Thr Val Ser Asn Ser Glu Asn Ala Val Arg
 245 250 255
 Ile Lys Thr Ile Ser Gly Ala Thr Gly Ser Val Ser Glu Ile Thr Tyr
 260 265 270
 Ser Asn Ile Val Met Ser Gly Ile Ser Asp Tyr Gly Val Val Ile Gln
 275 280 285
 Gln Asp Tyr Glu Asp Gly Lys Pro Thr Gly Lys Pro Thr Asn Gly Val
 290 295 300
 Thr Ile Gln Asp Val Lys Leu Glu Ser Val Thr Gly Ser Val Asp Ser
 305 310 315 320
 Gly Ala Thr Glu Ile Tyr Leu Leu Cys Gly Ser Gly Ser Cys Ser Asp
 325 330 335
 Trp Thr Trp Asp Asp Val Lys Val Thr Gly Gly Lys Lys Ser Thr Ala
 340 345 350
 Cys Lys Asn Phe Pro Ser Val Ala Ser Cys
 355 360

- <210> 38
- <211> 383
- 5 <212> PRT
- <213> Pseudomonas cellulosa
- <400> 38

Arg Ala Asp Val Lys Pro Val Thr Val Lys Leu Val Asp Ser Gln Ala
 1 5 10 15
 Thr Met Glu Thr Arg Ser Leu Phe Ala Phe Met Gln Glu Gln Arg Arg
 20 25 30
 His Ser Ile Met Phe Gly His Gln His Glu Thr Thr Gln Gly Leu Thr
 35 40 45
 Ile Thr Arg Thr Asp Gly Thr Gln Ser Asp Thr Phe Asn Ala Val Gly
 50 55 60
 Asp Phe Ala Ala Val Tyr Gly Trp Asp Thr Leu Ser Ile Val Ala Pro
 65 70 75 80

ES 2 373 658 T3

Lys Ala Glu Gly Asp Ile Val Ala Gln Val Lys Lys Ala Tyr Ala Arg
 85 90 95
 Gly Gly Ile Ile Thr Val Ser Ser His Phe Asp Asn Pro Lys Thr Asp
 100 105 110
 Thr Gln Lys Gly Val Trp Pro Val Gly Thr Ser Trp Asp Gln Thr Pro
 115 120 125
 Ala Val Val Asp Ser Leu Pro Gly Gly Ala Tyr Asn Pro Val Leu Asn
 130 135 140
 Gly Tyr Leu Asp Gln Val Ala Glu Trp Ala Asn Asn Leu Lys Asp Glu
 145 150 155 160
 Gln Gly Arg Leu Ile Pro Val Ile Phe Arg Leu Tyr His Ala Asn Thr
 165 170 175
 Gly Ser Trp Phe Trp Trp Gly Asp Lys Gln Ser Thr Pro Glu Gln Tyr
 180 185 190
 Lys Gln Leu Phe Arg Tyr Ser Val Glu Tyr Leu Arg Asp Val Lys Gly
 195 200 205
 Val Arg Asn Phe Leu Tyr Ala Tyr Ser Pro Asn Asn Phe Trp Asp Val
 210 215 220
 Thr Glu Ala Asn Tyr Leu Glu Arg Tyr Pro Gly Asp Glu Trp Val Asp
 225 230 235 240
 Val Leu Gly Phe Asp Thr Tyr Gly Pro Val Ala Asp Asn Ala Asp Trp
 245 250 255
 Phe Arg Asn Val Val Ala Asn Ala Ala Leu Val Ala Arg Met Ala Glu
 260 265 270
 Ala Arg Gly Lys Ile Pro Val Ile Ser Glu Ile Gly Ile Arg Ala Pro
 275 280 285
 Asp Ile Glu Ala Gly Leu Tyr Asp Asn Gln Trp Tyr Arg Lys Leu Ile
 290 295 300
 Ser Gly Leu Lys Ala Asp Pro Asp Ala Arg Glu Ile Ala Phe Leu Leu
 305 310 315 320
 Val Trp Arg Asn Ala Pro Gln Gly Val Pro Gly Pro Asn Gly Thr Gln
 325 330 335
 Val Pro His Tyr Trp Val Pro Ala Asn Arg Pro Glu Asn Ile Asn Asn
 340 345 350
 Gly Thr Leu Glu Asp Phe Gln Ala Phe Tyr Ala Asp Glu Phe Thr Ala
 355 360 365
 Phe Asn Arg Asp Ile Glu Gln Val Tyr Gln Arg Pro Thr Leu Ile
 370 375 380

<210> 39
 <211> 419
 <212> PRT
 <213> Bacillus circulans

5

<400> 39

Leu Gln Pro Ala Thr Ala Glu Ala Ala Asp Ser Tyr Lys Ile Val Gly
 1 5 10 15
 Tyr Tyr Pro Ser Trp Ala Ala Tyr Gly Arg Asn Tyr Asn Val Ala Asp
 20 25 30
 Ile Asp Pro Thr Lys Val Thr His Ile Asn Tyr Ala Phe Ala Asp Ile
 35 40 45
 Cys Trp Asn Gly Ile His Gly Asn Pro Asp Pro Ser Gly Pro Asn Pro
 50 55 60
 Val Thr Trp Thr Cys Gln Asn Glu Lys Ser Gln Thr Ile Asn Val Pro
 65 70 75 80
 Asn Gly Thr Ile Val Leu Gly Asp Pro Trp Ile Asp Thr Gly Lys Thr
 85 90 95
 Phe Ala Gly Asp Thr Trp Asp Gln Pro Ile Ala Gly Asn Ile Asn Gln
 100 105 110
 Leu Asn Lys Leu Lys Gln Thr Asn Pro Asn Leu Lys Thr Ile Ile Ser
 115 120 125
 Val Gly Gly Trp Thr Trp Ser Asn Arg Phe Ser Asp Val Ala Ala Thr
 130 135 140
 Ala Ala Thr Arg Glu Val Phe Ala Asn Ser Ala Val Asp Phe Leu Arg
 145 150 155 160
 Lys Tyr Asn Phe Asp Gly Val Asp Leu Asp Trp Glu Tyr Pro Val Ser
 165 170 175
 Gly Gly Leu Asp Gly Asn Ser Lys Arg Pro Glu Asp Lys Gln Asn Tyr
 180 185 190
 Thr Leu Leu Leu Ser Lys Ile Arg Glu Lys Leu Asp Ala Ala Gly Ala
 195 200 205
 Val Asp Gly Lys Lys Tyr Leu Leu Thr Ile Ala Ser Gly Ala Ser Ala
 210 215 220
 Thr Tyr Ala Ala Asn Thr Glu Leu Ala Lys Ile Ala Ala Ile Val Asp
 225 230 235 240
 Trp Ile Asn Ile Met Thr Tyr Asp Phe Asn Gly Ala Trp Gln Lys Ile
 245 250 255
 Ser Ala His Asn Ala Pro Leu Asn Tyr Asp Pro Ala Ala Ser Ala Ala
 260 265 270
 Gly Val Pro Asp Ala Asn Thr Phe Asn Val Ala Ala Gly Ala Gln Gly
 275 280 285
 His Leu Asp Ala Gly Val Pro Ala Ala Lys Leu Val Leu Gly Val Pro
 290 295 300

ES 2 373 658 T3

Phe Tyr Gly Arg Gly Trp Asp Gly Cys Ala Gln Ala Gly Asn Gly Gln
 305 310 315 320
 Tyr Gln Thr Cys Thr Gly Gly Ser Ser Val Gly Thr Trp Glu Ala Gly
 325 330 335
 Ser Phe Asp Phe Tyr Asp Leu Glu Ala Asn Tyr Ile Asn Lys Asn Gly
 340 345 350
 Tyr Thr Arg Tyr Trp Asn Asp Thr Ala Lys Val Pro Tyr Leu Tyr Asn
 355 360 365
 Ala Ser Asn Lys Arg Phe Ile Ser Tyr Asp Asp Ala Glu Ser Val Gly
 370 375 380
 Tyr Lys Thr Ala Tyr Ile Lys Ser Lys Gly Leu Gly Gly Ala Met Phe
 385 390 395 400
 Trp Glu Leu Ser Gly Asp Arg Asn Lys Thr Leu Gln Asn Lys Leu Lys
 405 410 415
 Ala Asp Leu

- <210> 40
- <211> 317
- 5 <212> PRT
- <213> Candida antarctica
- <400> 40

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

10

ES 2 373 658 T3

```

Asp Ala Leu Ala Val Ser Ala Pro Ser Val Trp Gln Gln Thr Thr Gly
145                150                155                160
Ser Ala Leu Thr Thr Ala Leu Arg Asn Ala Gly Gly Leu Thr Gln Ile
                165                170                175
Val Pro Thr Thr Asn Leu Tyr Ser Ala Thr Asp Glu Ile Val Gln Pro
                180                185                190
Gln Val Ser Asn Ser Pro Leu Asp Ser Ser Tyr Leu Phe Asn Gly Lys
                195                200                205
Asn Val Gln Ala Gln Ala Val Cys Gly Pro Leu Phe Val Ile Asp His
                210                215                220
Ala Gly Ser Leu Thr Ser Gln Phe Ser Tyr Val Val Gly Arg Ser Ala
225                230                235                240
Leu Arg Ser Thr Thr Gly Gln Ala Arg Ser Ala Asp Tyr Gly Ile Thr
                245                250                255
Asp Cys Asn Pro Leu Pro Ala Asn Asp Leu Thr Pro Glu Gln Lys Val
                260                265                270
Ala Ala Ala Ala Leu Leu Ala Pro Ala Ala Ala Ala Ile Val Ala Gly
                275                280                285
Pro Lys Gln Asn Cys Glu Pro Asp Leu Met Pro Tyr Ala Arg Pro Phe
                290                295                300
Ala Val Gly Lys Arg Thr Cys Ser Gly Ile Val Thr Pro
305                310                315

```

<210> 41

<211> 434

5 <212> PRT

<213> secuencia artificial

<220>

<223> quimera de cobaya y homo sapiens (humano = aprox. últimos 30 aminoácidos)

10 <400> 41

```

Ala Glu Val Cys Tyr Ser His Leu Gly Cys Phe Ser Asp Glu Lys Pro
1                5                10                15
Trp Ala Gly Thr Ser Gln Arg Pro Ile Lys Ser Leu Pro Ser Asp Pro
                20                25                30
Lys Lys Ile Asn Thr Arg Phe Leu Leu Tyr Thr Asn Glu Asn Gln Asn
                35                40                45
Ser Tyr Gln Leu Ile Thr Ala Thr Asp Ile Ala Thr Ile Lys Ala Ser
                50                55                60
Asn Phe Asn Leu Asn Arg Lys Thr Arg Phe Ile Ile His Gly Phe Thr

```

ES 2 373 658 T3

| | | | | | | | | | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 65 | | | | | 70 | | | | | 75 | | | | | 80 |
| Asp | Ser | Gly | Glu | Asn | Ser | Trp | Leu | Ser | Asp | Met | Cys | Lys | Asn | Met | Phe |
| | | | | 85 | | | | | 90 | | | | | 95 | |
| Gln | Val | Glu | Lys | Val | Asn | Cys | Ile | Cys | Val | Asp | Trp | Lys | Gly | Gly | Ser |
| | | | 100 | | | | | 105 | | | | | 110 | | |
| Lys | Ala | Gln | Tyr | Ser | Gln | Ala | Ser | Gln | Asn | Ile | Arg | Val | Val | Gly | Ala |
| | | | 115 | | | | 120 | | | | | 125 | | | |
| Glu | Val | Ala | Tyr | Leu | Val | Gln | Val | Leu | Ser | Thr | Ser | Leu | Asn | Tyr | Ala |
| | | | 130 | | | 135 | | | | | 140 | | | | |
| Pro | Glu | Asn | Val | His | Ile | Ile | Gly | His | Ser | Leu | Gly | Ala | His | Thr | Ala |
| 145 | | | | | 150 | | | | | 155 | | | | | 160 |
| Gly | Glu | Ala | Gly | Lys | Arg | Leu | Asn | Gly | Leu | Val | Gly | Arg | Ile | Thr | Gly |
| | | | | 165 | | | | | 170 | | | | 175 | | |
| Leu | Asp | Pro | Ala | Glu | Pro | Tyr | Phe | Gln | Asp | Thr | Pro | Glu | Glu | Val | Arg |
| | | | 180 | | | | | 185 | | | | | 190 | | |
| Leu | Asp | Pro | Ser | Asp | Ala | Lys | Phe | Val | Asp | Val | Ile | His | Thr | Asp | Ile |
| | | | 195 | | | | 200 | | | | | 205 | | | |
| Ser | Pro | Ile | Leu | Pro | Ser | Leu | Gly | Phe | Gly | Met | Ser | Gln | Lys | Val | Gly |
| | | | 210 | | | 215 | | | | | 220 | | | | |
| His | Met | Asp | Phe | Phe | Pro | Asn | Gly | Gly | Lys | Asp | Met | Pro | Gly | Cys | Lys |
| 225 | | | | | 230 | | | | | 235 | | | | | 240 |
| Thr | Gly | Ile | Ser | Cys | Asn | His | His | Arg | Ser | Ile | Glu | Tyr | Tyr | His | Ser |
| | | | | 245 | | | | | 250 | | | | | 255 | |
| Ser | Ile | Leu | Asn | Pro | Glu | Gly | Phe | Leu | Gly | Tyr | Pro | Cys | Ala | Ser | Tyr |
| | | | 260 | | | | | 265 | | | | | 270 | | |
| Asp | Glu | Phe | Gln | Glu | Ser | Gly | Cys | Phe | Pro | Cys | Pro | Ala | Lys | Gly | Cys |
| | | | 275 | | | | 280 | | | | | 285 | | | |
| Pro | Lys | Met | Gly | His | Phe | Ala | Asp | Gln | Tyr | Pro | Gly | Lys | Thr | Asn | Ala |
| | | | 290 | | | 295 | | | | | | 300 | | | |
| Val | Glu | Gln | Thr | Phe | Phe | Leu | Asn | Thr | Gly | Ala | Ser | Asp | Asn | Phe | Thr |
| 305 | | | | | | 310 | | | | 315 | | | | | 320 |
| Arg | Trp | Arg | Tyr | Lys | Val | Thr | Val | Thr | Leu | Ser | Gly | Glu | Lys | Asp | Pro |
| | | | | 325 | | | | | 330 | | | | | 335 | |
| Ser | Gly | Asn | Ile | Asn | Val | Ala | Leu | Leu | Gly | Lys | Asn | Gly | Asn | Ser | Ala |
| | | | 340 | | | | | 345 | | | | | 350 | | |
| Gln | Tyr | Gln | Val | Phe | Lys | Gly | Thr | Leu | Lys | Pro | Asp | Ala | Ser | Tyr | Thr |
| | | | 355 | | | | 360 | | | | | 365 | | | |
| Asn | Ser | Ile | Asp | Val | Glu | Leu | Asn | Val | Gly | Thr | Ile | Gln | Lys | Val | Thr |
| | | | 370 | | | 375 | | | | | 380 | | | | |
| Phe | Leu | Trp | Lys | Arg | Ser | Gly | Ile | Ser | Val | Ser | Lys | Pro | Lys | Met | Gly |
| 385 | | | | | 390 | | | | | 395 | | | | | 400 |
| Ala | Ser | Arg | Ile | Thr | Val | Gln | Ser | Gly | Lys | Asp | Gly | Thr | Lys | Tyr | Asn |

ES 2 373 658 T3

```

         405               410               415
Phe Cys Ser Ser Asp Ile Val Gln Glu Asn Val Glu Gln Thr Leu Ser
        420               425               430
Pro Cys

```

- <210> 42
- <211> 471
- 5 <212> PRT
- <213> Escherichia coli
- <400> 42

```

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

```

ES 2 373 658 T3

| | | | | | | | | | | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 225 | | | | | | 230 | | | | | | 235 | | | | 240 |
| Glu | Trp | Gln | Gly | Lys | Thr | Leu | Arg | Glu | Gln | Ala | Gln | Ala | Arg | Gly | Tyr | |
| | | | | 245 | | | | | 250 | | | | | 255 | | |
| Gln | Leu | Val | Ser | Asp | Ala | Ala | Ser | Leu | Asn | Ser | Val | Thr | Glu | Ala | Asn | |
| | | | 260 | | | | | 265 | | | | | 270 | | | |
| Gln | Gln | Lys | Pro | Leu | Leu | Gly | Leu | Phe | Ala | Asp | Gly | Asn | Met | Pro | Val | |
| | | 275 | | | | | 280 | | | | | 285 | | | | |
| Arg | Trp | Leu | Gly | Pro | Lys | Ala | Thr | Tyr | His | Gly | Asn | Ile | Asp | Lys | Pro | |
| | 290 | | | | | 295 | | | | | 300 | | | | | |
| Ala | Val | Thr | Cys | Thr | Pro | Asn | Pro | Gln | Arg | Asn | Asp | Ser | Val | Pro | Thr | |
| 305 | | | | | 310 | | | | | 315 | | | | | 320 | |
| Leu | Ala | Gln | Met | Thr | Asp | Lys | Ala | Ile | Glu | Leu | Leu | Ser | Lys | Asn | Glu | |
| | | | | 325 | | | | | 330 | | | | | 335 | | |
| Lys | Gly | Phe | Phe | Leu | Gln | Val | Glu | Gly | Ala | Ser | Ile | Asp | Lys | Gln | Asp | |
| | | 340 | | | | | | 345 | | | | | 350 | | | |
| His | Ala | Ala | Asn | Pro | Cys | Gly | Gln | Ile | Gly | Glu | Thr | Val | Asp | Leu | Asp | |
| | | 355 | | | | | 360 | | | | | | 365 | | | |
| Glu | Ala | Val | Gln | Arg | Ala | Leu | Glu | Phe | Ala | Lys | Lys | Glu | Gly | Asn | Thr | |
| | | 370 | | | | 375 | | | | | | 380 | | | | |
| Leu | Val | Ile | Val | Thr | Ala | Asp | His | Ala | His | Ala | Ser | Gln | Ile | Val | Ala | |
| 385 | | | | | 390 | | | | | | 395 | | | | 400 | |
| Pro | Asp | Thr | Lys | Ala | Pro | Gly | Leu | Thr | Gln | Ala | Leu | Asn | Thr | Lys | Asp | |
| | | | 405 | | | | | | | 410 | | | | 415 | | |
| Gly | Ala | Val | Met | Val | Met | Ser | Tyr | Gly | Asn | Ser | Glu | Glu | Asp | Ser | Gln | |
| | | | 420 | | | | | 425 | | | | | 430 | | | |
| Glu | His | Thr | Gly | Ser | Gln | Leu | Arg | Ile | Ala | Ala | Tyr | Gly | Pro | His | Ala | |
| | | 435 | | | | | 440 | | | | | 445 | | | | |
| Ala | Asn | Val | Val | Gly | Leu | Thr | Asp | Gln | Thr | Asp | Leu | Phe | Tyr | Thr | Met | |
| | 450 | | | | | 455 | | | | | 460 | | | | | |
| Lys | Ala | Ala | Leu | Gly | Leu | Lys | | | | | | | | | | |
| 465 | | | | | 470 | | | | | | | | | | | |

<210> 43
 <211> 260
 <212> PRT
 <213> Bovino

5

<400> 43

ES 2 373 658 T3

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

 20 25 30
 Asp Ile Val Leu Ile Gln Glu Val Arg Asp Ser His Leu Val Ala Val
 35 40 45
 Gly Lys Leu Leu Asp Tyr Leu Asn Gln Asp Asp Pro Asn Thr Tyr His
 50 55 60
 Tyr Val Val Ser Glu Pro Leu Gly Arg Asn Ser Tyr Lys Glu Arg Tyr
 65 70 75 80
 Leu Phe Leu Phe Arg Pro Asn Lys Val Ser Val Leu Asp Thr Tyr Gln
 85 90 95
 Tyr Asp Asp Gly Cys Glu Ser Cys Gly Asn Asp Ser Phe Ser Arg Glu
 100 105 110
 Pro Ala Val Val Lys Phe Ser Ser His Ser Thr Lys Val Lys Glu Phe
 115 120 125
 Ala Ile Val Ala Leu His Ser Ala Pro Ser Asp Ala Val Ala Glu Ile
 130 135 140
 Asn Ser Leu Tyr Asp Val Tyr Leu Asp Val Gln Gln Lys Trp His Leu
 145 150 155 160
 Asn Asp Val Met Leu Met Gly Asp Phe Asn Ala Asp Cys Ser Tyr Val
 165 170 175
 Thr Ser Ser Gln Trp Ser Ser Ile Arg Leu Arg Thr Ser Ser Thr Phe
 180 185 190
 Gln Trp Leu Ile Pro Asp Ser Ala Asp Thr Thr Ala Thr Ser Thr Asn
 195 200 205
 Cys Ala Tyr Asp Arg Ile Val Val Ala Gly Ser Leu Leu Gln Ser Ser
 210 215 220
 Val Val Pro Gly Ser Ala Ala Pro Phe Asp Phe Gln Ala Ala Tyr Gly
 225 230 235 240
 Leu Ser Asn Glu Met Ala Leu Ala Ile Ser Asp His Tyr Pro Val Glu
 245 250 255
 Val Thr Leu Thr
 260

<210> 44
 <211> 686
 5 <212> PRT
 <213> Bacillus circulans

<400> 44

ES 2 373 658 T3

Ala Pro Asp Thr Ser Val Ser Asn Lys Gln Asn Phe Ser Thr Asp Val
1 5 10 15
Ile Tyr Gln Ile Phe Thr Asp Arg Phe Ser Asp Gly Asn Pro Ala Asn

ES 2 373 658 T3

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

ES 2 373 658 T3

```

          355                360                365
Gly Thr Asp Pro Asp Asn Arg Ala Arg Ile Pro Ser Phe Ser Thr Ser
   370                375                380
Thr Thr Ala Tyr Gln Val Ile Gln Lys Leu Ala Pro Leu Arg Lys Cys
385                390                395                400
Asn Pro Ala Ile Ala Tyr Gly Ser Thr Gln Glu Arg Trp Ile Asn Asn
          405                410                415
Asp Val Leu Ile Tyr Glu Arg Lys Phe Gly Ser Asn Val Ala Val Val
          420                425                430
Ala Val Asn Arg Asn Leu Asn Ala Pro Ala Ser Ile Ser Gly Leu Val
          435                440                445
Thr Ser Leu Pro Gln Gly Ser Tyr Asn Asp Val Leu Gly Gly Leu Leu
   450                455                460
Asn Gly Asn Thr Leu Ser Val Gly Ser Gly Gly Ala Ala Ser Asn Phe
465                470                475                480
Thr Leu Ala Ala Gly Gly Thr Ala Val Trp Gln Tyr Thr Ala Ala Thr
          485                490                495
Ala Thr Pro Thr Ile Gly His Val Gly Pro Met Met Ala Lys Pro Gly
          500                505                510
Val Thr Ile Thr Ile Asp Gly Arg Gly Phe Gly Ser Ser Lys Gly Thr
          515                520                525
Val Tyr Phe Gly Thr Thr Ala Val Ser Gly Ala Asp Ile Thr Ser Trp
   530                535                540
Glu Asp Thr Gln Ile Lys Val Lys Ile Pro Ala Val Ala Gly Gly Asn
545                550                555                560
Tyr Asn Ile Lys Val Ala Asn Ala Ala Gly Thr Ala Ser Asn Val Tyr
          565                570                575
Asp Asn Phe Glu Val Leu Ser Gly Asp Gln Val Ser Val Arg Phe Val
          580                585                590
Val Asn Asn Ala Thr Thr Ala Leu Gly Gln Asn Val Tyr Leu Thr Gly
          595                600                605
Ser Val Ser Glu Leu Gly Asn Trp Asp Pro Ala Lys Ala Ile Gly Pro
   610                615                620
Met Tyr Asn Gln Val Val Tyr Gln Tyr Pro Asn Trp Tyr Tyr Asp Val
625                630                635                640
Ser Val Pro Ala Gly Lys Thr Ile Glu Phe Lys Phe Leu Lys Lys Gln
          645                650                655
Gly Ser Thr Val Thr Trp Glu Gly Gly Ser Asn His Thr Phe Thr Ala
          660                665                670
Pro Ser Ser Gly Thr Ala Thr Ile Asn Val Asn Trp Gln Pro
          675                680                685

```

<210> 45

<211> 404

<212> PRT

<213> Amycolatopsis orientalis

5

ES 2 373 658 T3

<400> 45

Met Arg Val Leu Ile Thr Gly Cys Gly Ser Arg Gly Asp Thr Glu Pro
 1 5 10 15
 Leu Val Ala Leu Ala Ala Arg Leu Arg Glu Leu Gly Ala Asp Ala Arg
 20 25 30
 Met Cys Leu Pro Pro Asp Tyr Val Glu Arg Cys Ala Glu Val Gly Val
 35 40 45
 Pro Met Val Pro Val Gly Arg Ala Val Arg Ala Gly Ala Arg Glu Pro
 50 55 60
 Gly Glu Leu Pro Pro Gly Ala Ala Glu Val Val Thr Glu Val Val Ala
 65 70 75 80
 Glu Trp Phe Asp Lys Val Pro Ala Ala Ile Glu Gly Cys Asp Ala Val
 85 90 95
 Val Thr Thr Gly Leu Leu Pro Ala Ala Val Ala Val Arg Ser Met Ala
 100 105 110
 Glu Lys Leu Gly Ile Pro Tyr Arg Tyr Thr Val Leu Ser Pro Asp His
 115 120 125
 Leu Pro Ser Glu Gln Ser Gln Ala Glu Arg Asp Met Tyr Asn Gln Gly
 130 135 140
 Ala Asp Arg Leu Phe Gly Asp Ala Val Asn Ser His Arg Ala Ser Ile
 145 150 155 160
 Gly Leu Pro Pro Val Glu His Leu Tyr Asp Tyr Gly Tyr Thr Asp Gln
 165 170 175
 Pro Trp Leu Ala Ala Asp Pro Val Leu Ser Pro Leu Arg Pro Thr Asp
 180 185 190
 Leu Gly Thr Val Gln Thr Gly Ala Trp Ile Leu Pro Asp Glu Arg Pro
 195 200 205
 Leu Ser Ala Glu Leu Glu Ala Phe Leu Ala Ala Gly Ser Thr Pro Val
 210 215 220
 Tyr Val Gly Phe Gly Ser Ser Ser Arg Pro Ala Thr Ala Asp Ala Ala
 225 230 235 240
 Lys Met Ala Ile Lys Ala Val Arg Ala Ser Gly Arg Arg Ile Val Leu
 245 250 255
 Ser Arg Gly Trp Ala Asp Leu Val Leu Pro Asp Asp Gly Ala Asp Cys
 260 265 270
 Phe Val Val Gly Glu Val Asn Leu Gln Glu Leu Phe Gly Arg Val Ala

ES 2 373 658 T3

```

                275                280                285
Ala Ala Ile His His Asp Ser Ala Gly Thr Thr Leu Leu Ala Met Arg
    290                295                300
Ala Gly Ile Pro Gln Ile Val Val Arg Arg Val Val Asp Asn Val Val
    305                310                315                320
Glu Gln Ala Tyr His Ala Asp Arg Val Ala Glu Leu Gly Val Gly Val
                325                330                335
Ala Val Asp Gly Pro Val Pro Thr Ile Asp Ser Leu Ser Ala Ala Leu
                340                345                350
Asp Thr Ala Leu Ala Pro Glu Ile Arg Ala Arg Ala Thr Thr Val Ala
                355                360                365
Asp Thr Ile Arg Ala Asp Gly Thr Thr Val Ala Ala Gln Leu Leu Phe
                370                375                380
Asp Ala Val Ser Leu Glu Lys Pro Thr Val Pro Ala Leu Glu His His
    385                390                395                400
His His His His

```

<210> 46
 <211> 292
 5 <212> PRT
 <213> Pseudomonas sp.

<400> 46

```

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

```

10

ES 2 373 658 T3

```

130              135              140
His Phe Val Arg Cys Val Pro Asp Thr Ala Lys Ala Met Ala Phe Tyr
145              150              155              160
Thr Glu Val Leu Gly Phe Val Leu Ser Asp Ile Ile Asp Ile Gln Met
              165              170              175
Gly Pro Glu Thr Ser Val Pro Ala His Phe Leu His Cys Asn Gly Arg
              180              185              190
His His Thr Ile Ala Leu Ala Ala Phe Pro Ile Pro Lys Arg Ile His
              195              200              205
His Phe Met Leu Gln Ala Asn Thr Ile Asp Asp Val Gly Tyr Ala Phe
              210              215              220
Asp Arg Leu Asp Ala Ala Gly Arg Ile Thr Ser Leu Leu Gly Arg His
225              230              235              240
Thr Asn Asp Gln Thr Leu Ser Phe Tyr Ala Asp Thr Pro Ser Pro Met
              245              250              255
Ile Glu Val Glu Phe Gly Trp Gly Pro Arg Thr Val Asp Ser Ser Trp
              260              265              270
Thr Val Ala Arg His Ser Arg Thr Ala Met Trp Gly His Lys Ser Val
              275              280              285
Arg Gly Gln Arg
              290

```

<210> 47
 <211> 311
 <212> PRT
 <213> Acitenobacter sp.

5

<400> 47

```

Met Glu Val Lys Ile Phe Asn Thr Gln Asp Val Gln Asp Phe Leu Arg
1              5              10              15
Val Ala Ser Gly Leu Glu Gln Glu Gly Gly Asn Pro Arg Val Lys Gln
              20              25              30
Ile Ile His Arg Val Leu Ser Asp Leu Tyr Lys Ala Ile Glu Asp Leu
              35              40              45
Asn Ile Thr Ser Asp Glu Tyr Trp Ala Gly Val Ala Tyr Leu Asn Gln
              50              55              60
Leu Gly Ala Asn Gln Glu Ala Gly Leu Leu Ser Pro Gly Leu Gly Phe
65              70              75              80
Asp His Tyr Leu Asp Met Arg Met Asp Ala Glu Asp Ala Ala Leu Gly
              85              90              95
Ile Glu Asn Ala Thr Pro Arg Thr Ile Glu Gly Pro Leu Tyr Val Ala

```

10

ES 2 373 658 T3

```

                100                105                110
Gly Ala Pro Glu Ser Val Gly Tyr Ala Arg Met Asp Asp Gly Ser Asp
                115                120                125
Pro Asn Gly His Thr Leu Ile Leu His Gly Thr Ile Phe Asp Ala Asp
                130                135                140
Gly Lys Pro Leu Pro Asn Ala Lys Val Glu Ile Trp His Ala Asn Thr
145                150                155                160
Lys Gly Phe Tyr Ser His Phe Asp Pro Thr Gly Glu Gln Gln Ala Phe
                165                170                175
Asn Met Arg Arg Ser Ile Ile Thr Asp Glu Asn Gly Gln Tyr Arg Val
                180                185                190
Arg Thr Ile Leu Pro Ala Gly Tyr Gly Cys Pro Pro Glu Gly Pro Thr
                195                200                205
Gln Gln Leu Leu Asn Gln Leu Gly Arg His Gly Asn Arg Pro Ala His
                210                215                220
Ile His Tyr Phe Val Ser Ala Asp Gly His Arg Lys Leu Thr Thr Gln
225                230                235                240
Ile Asn Val Ala Gly Asp Pro Tyr Thr Tyr Asp Asp Phe Ala Tyr Ala
                245                250                255
Thr Arg Glu Gly Leu Val Val Asp Ala Val Glu His Thr Asp Pro Glu
                260                265                270
Ala Ile Lys Ala Asn Asp Val Glu Gly Pro Phe Ala Glu Met Val Phe
                275                280                285
Asp Leu Lys Leu Thr Arg Leu Val Asp Gly Val Asp Asn Gln Val Val
                290                295                300
Asp Arg Pro Arg Leu Ala Val
205                310

```

<210> 48

<211> 414

5 <212> PRT

<213> Pseudomonas putida

<400> 48

```

Thr Thr Glu Thr Ile Gln Ser Asn Ala Asn Leu Ala Pro Leu Pro Pro
1                5                10                15
His Val Pro Glu His Leu Val Phe Asp Phe Asp Met Tyr Asn Pro Ser
                20                25                30
Asn Leu Ser Ala Gly Val Gln Glu Ala Trp Ala Val Leu Gln Glu Ser
                35                40                45
Asn Val Pro Asp Leu Val Trp Thr Arg Cys Asn Gly Gly His Trp Ile

```

10

ES 2 373 658 T3

| | | | | | | | | | | | | | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|--|--|--|--|
| 50 | | 55 | | 60 | | | | | | | | | | | | | | | |
| Ala | Thr | Arg | Gly | Gln | Leu | Ile | Arg | Glu | Ala | Tyr | Glu | Asp | Tyr | Arg | His | | | | |
| 65 | | | | | 70 | | | | | 75 | | | | | 80 | | | | |
| Phe | Ser | Ser | Glu | Cys | Pro | Phe | Ile | Pro | Arg | Glu | Ala | Gly | Glu | Ala | Tyr | | | | |
| | | | | 85 | | | | | | 90 | | | | | 95 | | | | |
| Asp | Phe | Ile | Pro | Thr | Ser | Met | Asp | Pro | Pro | Glu | Gln | Arg | Gln | Phe | Arg | | | | |
| | | | 100 | | | | | 105 | | | | | | 110 | | | | | |
| Ala | Leu | Ala | Asn | Gln | Val | Val | Gly | Met | Pro | Val | Val | Asp | Lys | Leu | Glu | | | | |
| | | | 115 | | | | 120 | | | | | | 125 | | | | | | |
| Asn | Arg | Ile | Gln | Glu | Leu | Ala | Cys | Ser | Leu | Ile | Glu | Ser | Leu | Arg | Pro | | | | |
| | | | 130 | | | | 135 | | | | | | 140 | | | | | | |
| Gln | Gly | Gln | Cys | Asn | Phe | Thr | Glu | Asp | Tyr | Ala | Glu | Pro | Phe | Pro | Ile | | | | |
| 145 | | | | | 150 | | | | | 155 | | | | | 160 | | | | |
| Arg | Ile | Phe | Met | Leu | Leu | Ala | Gly | Leu | Pro | Glu | Glu | Asp | Ile | Pro | His | | | | |
| | | | | 165 | | | | | | 170 | | | | | 175 | | | | |
| Leu | Lys | Tyr | Leu | Thr | Asp | Gln | Met | Thr | Arg | Pro | Asp | Gly | Ser | Met | Thr | | | | |
| | | | 180 | | | | | 185 | | | | | | 190 | | | | | |
| Phe | Ala | Glu | Ala | Lys | Glu | Ala | Leu | Tyr | Asp | Tyr | Leu | Ile | Pro | Ile | Ile | | | | |
| | | | 195 | | | | 200 | | | | | | | 205 | | | | | |
| Glu | Gln | Arg | Arg | Gln | Lys | Pro | Gly | Thr | Asp | Ala | Ile | Ser | Ile | Val | Ala | | | | |
| | | | 210 | | | | 215 | | | | | | 220 | | | | | | |
| Asn | Gly | Gln | Val | Asn | Gly | Arg | Pro | Ile | Thr | Ser | Asp | Glu | Ala | Lys | Arg | | | | |
| 225 | | | | | 230 | | | | | | 235 | | | | 240 | | | | |
| Met | Cys | Gly | Leu | Leu | Leu | Val | Gly | Gly | Leu | Asp | Thr | Val | Val | Asn | Phe | | | | |
| | | | | 245 | | | | | | 250 | | | | 255 | | | | | |
| Leu | Ser | Phe | Ser | Met | Glu | Phe | Leu | Ala | Lys | Ser | Pro | Glu | His | Arg | Gln | | | | |
| | | | 260 | | | | | 265 | | | | | | 270 | | | | | |
| Glu | Leu | Ile | Gln | Arg | Pro | Glu | Arg | Ile | Pro | Ala | Ala | Cys | Glu | Glu | Leu | | | | |
| | | | 275 | | | | 280 | | | | | | 285 | | | | | | |
| Leu | Arg | Arg | Phe | Ser | Leu | Val | Ala | Asp | Gly | Arg | Ile | Leu | Thr | Ser | Asp | | | | |
| | | | | | 290 | | 295 | | | | | | 300 | | | | | | |
| Tyr | Glu | Phe | His | Gly | Val | Gln | Leu | Lys | Lys | Gly | Asp | Gln | Ile | Leu | Leu | | | | |
| 305 | | | | | 310 | | | | | 315 | | | | | 320 | | | | |
| Pro | Gln | Met | Leu | Ser | Gly | Leu | Asp | Glu | Arg | Glu | Asn | Ala | Cys | Pro | Met | | | | |
| | | | | 325 | | | | | | 330 | | | | 335 | | | | | |
| His | Val | Asp | Phe | Ser | Arg | Gln | Lys | Val | Ser | His | Thr | Thr | Phe | Gly | His | | | | |
| | | | 340 | | | | | 345 | | | | | | 350 | | | | | |
| Gly | Ser | His | Leu | Cys | Leu | Gly | Gln | His | Leu | Ala | Arg | Arg | Glu | Ile | Ile | | | | |
| | | | 355 | | | | 360 | | | | | | 365 | | | | | | |
| Val | Thr | Leu | Lys | Glu | Trp | Leu | Thr | Arg | Ile | Pro | Asp | Phe | Ser | Ile | Ala | | | | |
| | | | 370 | | | | 375 | | | | | | 380 | | | | | | |
| Pro | Gly | Ala | Gln | Ile | Gln | His | Lys | Ser | Gly | Ile | Val | Ser | Gly | Val | Gln | | | | |

ES 2 373 658 T3

385 390 395 400
 Ala Leu Pro Leu Val Trp Asp Pro Ala Thr Thr Lys Ala Val
 405 410

- 5 <210> 49
 <211> 374
 <212> PRT
 <213> Equus caballus
 <400> 49

```

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

```

10

ES 2 373 658 T3

```

                245                250                255
Met Ser Asn Gly Gly Val Asp Phe Ser Phe Glu Val Ile Gly Arg Leu
                260                265                270
Asp Thr Met Val Thr Ala Leu Ser Cys Cys Gln Glu Ala Tyr Gly Val
                275                280                285
Ser Val Ile Val Gly Val Pro Pro Asp Ser Gln Asn Leu Ser Met Asn
                290                295                300
Pro Met Leu Leu Leu Ser Gly Arg Thr Trp Lys Gly Ala Ile Phe Gly
305                310                315                320
Gly Phe Lys Ser Lys Asp Ser Val Pro Lys Leu Val Ala Asp Phe Met
                325                330                335
Ala Lys Lys Phe Ala Leu Asp Pro Leu Ile Thr His Val Leu Pro Phe
                340                345                350
Glu Lys Ile Asn Glu Gly Phe Asp Leu Leu Arg Ser Gly Glu Ser Ile
                355                360                365
Arg Thr Ile Leu Thr Phe
                370

```

<210> 50
 <211> 297
 5 <212> PRT
 <213> Escherichia coli

<400> 50

```

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

```

10

ES 2 373 658 T3

```

      130              135              140
Val Lys Leu Thr Leu Asp Gln Ile Asn Thr Leu Val Thr Leu Pro Gly
145              150              155              160
Val Gly Ala Leu Lys Gln Thr Ser Gly Asp Leu Tyr Gln Met Glu Gln
      165              170              175
Ile Arg Arg Glu His Pro Asp Leu Val Leu Tyr Asn Gly Tyr Asp Glu
      180              185              190
Ile Phe Ala Ser Gly Leu Leu Ala Gly Ala Asp Gly Gly Ile Gly Ser
      195              200              205
Thr Tyr Asn Ile Met Gly Trp Arg Tyr Gln Gly Ile Val Lys Ala Leu
      210              215              220
Lys Glu Gly Asp Ile Gln Thr Ala Gln Lys Leu Gln Thr Glu Cys Asn
225              230              235              240
Lys Val Ile Asp Leu Leu Ile Lys Thr Gly Val Phe Arg Gly Leu Lys
      245              250              255
Thr Val Leu His Tyr Met Asp Val Val Ser Val Pro Leu Cys Arg Lys
      260              265              270
Pro Phe Gly Pro Val Asp Glu Lys Tyr Leu Pro Glu Leu Lys Ala Leu
      275              280              285
Ala Gln Gln Leu Met Gln Glu Arg Gly
      290              295

```

<210> 51
 <211> 268
 <212> PRT
 <213> Salmonella typhimurium

5

<400> 51

```

Met Glu Arg Tyr Glu Asn Leu Phe Ala Gln Leu Asn Asp Arg Arg Glu
1              5              10              15
Gly Ala Phe Val Pro Phe Val Thr Leu Gly Asp Pro Gly Ile Glu Gln
      20              25              30
Ser Leu Lys Ile Ile Asp Thr Leu Ile Asp Ala Gly Ala Asp Ala Leu
      35              40              45
Glu Leu Gly Val Pro Phe Ser Asp Pro Leu Ala Asp Gly Pro Thr Ile
      50              55              60
Gln Asn Ala Asn Leu Arg Ala Phe Ala Ala Gly Val Thr Pro Ala Gln
65              70              75              80
Cys Phe Glu Met Leu Ala Leu Ile Arg Glu Lys His Pro Thr Ile Pro
      85              90              95
Ile Gly Leu Leu Met Tyr Ala Asn Leu Val Phe Asn Asn Gly Ile Asp

```

10

ES 2 373 658 T3

```

                100                105                110
Ala Phe Tyr Ala Arg Cys Glu Gln Val Gly Val Asp Ser Val Leu Val
                115                120                125
Ala Asp Val Pro Val Glu Glu Ser Ala Pro Phe Arg Gln Ala Ala Leu
                130                135                140
Arg His Asn Ile Ala Pro Ile Phe Ile Cys Pro Pro Asn Ala Asp Asp
                145                150                155                160
Asp Leu Leu Arg Gln Val Ala Ser Tyr Gly Arg Gly Tyr Thr Tyr Leu
                165                170                175
Leu Ser Arg Ser Gly Val Thr Gly Ala Glu Asn Arg Gly Ala Leu Pro
                180                185                190
Leu His His Leu Ile Glu Lys Leu Lys Glu Tyr His Ala Ala Pro Ala
                195                200                205
Leu Gln Gly Phe Gly Ile Ser Ser Pro Glu Gln Val Ser Ala Ala Val
                210                215                220
Arg Ala Gly Ala Ala Gly Ala Ile Ser Gly Ser Ala Ile Val Lys Ile
                225                230                235                240
Ile Glu Lys Asn Leu Ala Ser Pro Lys Gln Met Leu Ala Glu Leu Arg
                245                250                255
Ser Phe Val Ser Ala Met Lys Ala Ala Ser Arg Ala
                260                265

```

<210> 52
 <211> 393
 <212> PRT
 <213> Actinoplanes missouriensis

5

<400> 52

```

Ser Val Gln Ala Thr Arg Glu Asp Lys Phe Ser Phe Gly Leu Trp Thr
1                5                10                15
Val Gly Trp Gln Ala Arg Asp Ala Phe Gly Asp Ala Thr Arg Thr Ala
                20                25                30
Leu Asp Pro Val Glu Ala Val His Lys Leu Ala Glu Ile Gly Ala Tyr
                35                40                45
Gly Ile Thr Phe His Asp Asp Asp Leu Val Pro Phe Gly Ser Asp Ala
                50                55                60
Gln Thr Arg Asp Gly Ile Ile Ala Gly Phe Lys Lys Ala Leu Asp Glu
                65                70                75                80
Thr Gly Leu Ile Val Pro Met Val Thr Thr Asn Leu Phe Thr His Pro
                85                90                95
Val Phe Lys Asp Gly Gly Phe Thr Ser Asn Asp Arg Ser Val Arg Arg

```

10

ES 2 373 658 T3

```

                100                105                110
Tyr Ala Ile Arg Lys Val Leu Arg Gln Met Asp Leu Gly Ala Glu Leu
                115                120                125
Gly Ala Lys Thr Leu Val Leu Trp Gly Gly Arg Glu Gly Ala Glu Tyr
                130                135                140
Asp Ser Ala Lys Asp Val Ser Ala Ala Leu Asp Arg Tyr Arg Glu Ala
145                150                155                160
Leu Asn Leu Leu Ala Gln Tyr Ser Glu Asp Arg Gly Tyr Gly Leu Arg
                165                170                175
Phe Ala Ile Glu Pro Lys Pro Asn Glu Pro Arg Gly Asp Ile Leu Leu
                180                185                190
Pro Thr Ala Gly His Ala Ile Ala Phe Val Gln Glu Leu Glu Arg Pro
                195                200                205
Glu Leu Phe Gly Ile Asn Pro Glu Thr Gly Asn Glu Gln Met Ser Asn
                210                215                220
Leu Asn Phe Thr Gln Gly Ile Ala Gln Ala Leu Trp His Lys Lys Leu
225                230                235                240
Phe His Ile Asp Leu Asn Gly Gln His Gly Pro Lys Phe Asp Gln Asp
                245                250                255
Leu Val Phe Gly His Gly Asp Leu Leu Asn Ala Phe Ser Leu Val Asp
                260                265                270
Leu Leu Glu Asn Gly Pro Asp Gly Ala Pro Ala Tyr Asp Gly Pro Arg
                275                280                285
His Phe Asp Tyr Lys Pro Ser Arg Thr Glu Asp Tyr Asp Gly Val Trp
                290                295                300
Glu Ser Ala Lys Ala Asn Ile Arg Met Tyr Leu Leu Leu Lys Glu Arg
305                310                315                320
Ala Lys Ala Phe Arg Ala Asp Pro Glu Val Gln Glu Ala Leu Ala Ala
                325                330                335
Ser Lys Val Ala Glu Leu Lys Thr Pro Thr Leu Asn Pro Gly Glu Gly
                340                345                350
Tyr Ala Glu Leu Leu Ala Asp Arg Ser Ala Phe Glu Asp Tyr Asp Ala
                355                360                365
Asp Ala Val Gly Ala Lys Gly Phe Gly Phe Val Lys Leu Asn Gln Leu
                370                375                380
Ala Ile Glu His Leu Leu Gly Ala Arg
200                205                210
                215                220
                225                230                235
                240                245                250
                255                260                265
                270                275                280
                285                290                295
                300                305                310
                315                320                325
                330                335                340
                345                350                355
                360                365                370
                375                380                385
                385                390

```

<210> 53
 <211> 348
 <212> PRT
 <213> Bacteriófago T7

5

<400> 53

ES 2 373 658 T3

Val Asn Ile Lys Thr Asn Pro Phe Lys Ala Val Ser Phe Val Glu Ser
1 5 10 15
Ala Ile Lys Lys Ala Leu Asp Asn Ala Gly Tyr Leu Ile Ala Glu Ile
20 25 30
Lys Tyr Asp Gly Val Arg Gly Asn Ile Cys Val Asp Asn Thr Ala Asn
35 40 45
Ser Tyr Trp Leu Ser Arg Val Ser Lys Thr Ile Pro Ala Leu Glu His
50 55 60
Leu Asn Gly Phe Asp Val Arg Trp Lys Arg Leu Leu Asn Asp Asp Arg
65 70 75 80
Cys Phe Tyr Lys Asp Gly Phe Met Leu Asp Gly Glu Leu Met Val Lys
85 90 95
Gly Val Asp Phe Asn Thr Gly Ser Gly Leu Leu Arg Thr Lys Trp Thr
100 105 110
Asp Thr Lys Asn Gln Glu Phe His Glu Glu Leu Phe Val Glu Pro Ile
115 120 125
Arg Lys Lys Asp Lys Val Pro Phe Lys Leu His Thr Gly His Leu His
130 135 140
Ile Lys Leu Tyr Ala Ile Leu Pro Leu His Ile Val Glu Ser Gly Glu
145 150 155 160
Asp Cys Asp Val Met Thr Leu Leu Met Gln Glu His Val Lys Asn Met
165 170 175
Leu Pro Leu Leu Gln Glu Tyr Phe Pro Glu Ile Glu Trp Gln Ala Ala
180 185 190
Glu Ser Tyr Glu Val Tyr Asp Met Val Glu Leu Gln Gln Leu Tyr Glu
195 200 205
Gln Lys Arg Ala Glu Gly His Glu Gly Leu Ile Val Lys Asp Pro Met
210 215 220
Cys Ile Tyr Lys Arg Gly Lys Lys Ser Gly Trp Trp Lys Met Lys Pro
225 230 235 240
Glu Asn Glu Ala Asp Gly Ile Ile Gln Gly Leu Val Trp Gly Thr Lys
245 250 255
Gly Leu Ala Asn Glu Gly Lys Val Ile Gly Phe Glu Val Leu Leu Glu
260 265 270
Ser Gly Arg Leu Val Asn Ala Thr Asn Ile Ser Arg Ala Leu Met Asp
275 280 285
Glu Phe Thr Glu Thr Val Lys Glu Ala Thr Leu Ser Gln Trp Gly Phe
290 295 300
Phe Ser Pro Tyr Gly Ile Gly Asp Asn Asp Ala Cys Thr Ile Asn Pro

ES 2 373 658 T3

| | | | | | | |
|---|--|-----|--|-----|--|-----|
| 305 | | 310 | | 315 | | 320 |
| Tyr Asp Gly Trp Ala Cys Gln Ile Ser Tyr Met Glu Glu Thr Pro Asp | | | | | | |
| | | 325 | | 330 | | 335 |
| Gly Ser Leu Arg His Pro Ser Phe Val Met Phe Arg | | | | | | |
| | | 340 | | 345 | | |

5 <210> 54
 <211> 42
 <212> ADN
 <213> secuencia artificial
 <220>
 <223> sitio de unión para restr1 y restr2
 <220>
 10 <221> CDS
 <222> (2)..(40)
 <223>
 <400> 54
 15
 g gtg gta tca gca ggc cac tgc tac aag tcc cgc atc cag gt 42
 Val Val Ser Ala Gly His Cys Tyr Lys Ser Arg Ile Gln
 1 5 10

20 <210> 55
 <211> 13
 <212> PRT
 <213> secuencia artificial
 <220>
 <223> sitio de unión para restr1 y restr2
 25 <400> 55
 Val Val Ser Ala Gly His Cys Tyr Lys Ser Arg Ile Gln
 1 5 10

30 <210> 56
 <211> 42
 <212> ADN
 <213> secuencia artificial
 <220>
 <223> cebador directo restr1
 35 <400> 56
 ggtggtatcc gggggccact gctacaagtc cggatccag gt 42

40 <210> 57
 <211> 42
 <212> ADN
 <213> secuencia artificial
 <220>

45 <223> cebador inverso restr2
 <400> 57
 acctggatcc gggactgta gcagtggccc gcggatacca cc 42

50 <210> 58

ES 2 373 658 T3

```

<211> 50
<212> ADN
<213> secuencia artificial
<220>
5 <223> sitio de unión para restr3 y restr4
<220>
<221> CDS
<222> (3)..(50)
<223>
10 <400> 58

    cc act ggc acg aag tgc etc atc tct ggc tgg ggc aac act gcg agc    47
      Thr Gly Thr Lys Cys Leu Ile Ser Gly Trp Gly Asn Thr Ala Ser
      1           5           10           15
    tct    50
    Ser

15 <210> 59
<211> 16
<212> PRT
<213> secuencia artificial
<220>
20 <223> sitio de unión para restr3 y restr4

<400> 59

      Thr Gly Thr Lys Cys Leu Ile Ser Gly Trp Gly Asn Thr Ala Ser Ser
      1           5           10           15

25 <210> 60
<211> 50
<212> ADN
<213> secuencia artificial
<220>
30 <223> cebador directo restr3

<400> 60
ccactggcac gaagtgcctc atctctggct ggggcaaac tgcgagctct    50

35 <210> 61
<211> 50
<212> ADN
<213> secuencia artificial
<220>
40 <223> cebador inverso restr4

<400> 61
agagctagca gtgtgcccc agccagagat gaggcactg gtaccagtgg    50

45 <210> 62
<211> 30
<212> ADN
<213> secuencia artificial
<220>
50 <223> cebador puc-directo

<400> 62
ggggtacccc accaccatga atccactct    30

```

<210> 63
 <211> 30
 <212> ADN
 <213> secuencia artificial
 5 <220>
 <223> cebador puc-inverso

 <400> 63
 cgggatccgg tatagagact gaagagatac 30
 10
 <210> 64
 <211> 39
 <212> ADN
 <213> secuencia artificial
 15 <220>
 <223> oligox-SDR1f
 <220>
 <221> misc_feature
 <222> (14)..(31)
 20 <223> cualquier nucleótido
 <220>
 <221> misc_feature
 <222> (14)..(31)
 <223> cualquier resto nucleotídico o aminoacídico
 25 <220>
 <221> CDS
 <222> (2)..(37)
 <223>
 30 <400> 64

g ggc cac tgc tac nnn nnn nnn nnn nnn nnn aag tcc cg 39
Gly His Cys Tyr Xaa Xaa Xaa Xaa Xaa Xaa Lys Ser
1 5 10

 <210> 65
 35 <211> 12
 <212> PRT
 <213> secuencia artificial
 <220>
 <221> misc_feature
 40 <222> (5)..(5)
 <223> El "Xaa" en la localización 5 representa Lys, Asn, Arg, Ser, Thr, Ile, Met, Glu, Asp, Gly, Ala, Val, Gln, His, Pro, Leu, un codón de terminación, Tyr, Trp, Cys, o Phe.

 <220>
 45 <221> misc_feature
 <222> (6)..(6)
 <223> T El "Xaa" en la localización 6 representa Lys, Asn, Arg, Ser, Thr, Ile, Met, Glu, Asp, Gly, Ala, Val, Gln, His, Pro, Leu, un codón de terminación, Tyr, Trp, Cys, o Phe.

 50 <220>
 <221> misc_feature
 <222> (7)..(7)
 <223> El "Xaa" en la localización 7 representa Lys, Asn, Arg, Ser, Thr, Ile, Met, Glu, Asp, Gly, Ala, Val, Gln, His, Pro, Leu, un codón de terminación, Tyr, Trp, Cys, o Phe.
 55
 <220>
 <221> misc_feature
 <222> (8)..(8)
 <223> El "Xaa" en la localización 8 representma Lys, Asn, Arg, Ser, Thr, Ile, Met, Glu, Asp, Gly, Ala, Val, Gln, His, Pro, Leu, un codón de terminación, Tyr, Trp, Cys, o Phe.
 60
 <220>

ES 2 373 658 T3

<221> misc_feature
 <222> (9)..(9)
 <223> El "Xaa" en la localización 96 representa Lys, Asn, Arg, Ser, Thr, Ile, Met, Glu, Asp, Gly, Ala, Val, Gln, His, Pro, Leu, un codón de terminación, Tyr, Trp, Cys, o Phe.

5

<220>
 <221> misc_feature
 <222> (10)..(10)
 <223> El "Xaa" en la localización 10 representa Lys, Asn, Arg, Ser, Thr, Ile, Met, Glu, Asp, Gly, Ala, Val, Gln, His, Pro, Leu, un codón de terminación, Tyr, Trp, Cys, o Phe.

10

<220>
 <223> oligox-SDR1f
 <220>
 <221> misc_feature
 <222> (14) .. (14)
 <223> cualquier nucleótido

15

<220>
 <221> misc_feature
 <222> (14)..(31)
 <223> cualquier resto nucleotídico o aminoacídico

20

<400> 65

25

| | | | | | | | | | | | | |
|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|
| Gly | His | Cys | Tyr | Xaa | Xaa | Xaa | Xaa | Xaa | Xaa | Xaa | Lys | Ser |
| 1 | | | 5 | | | | | | | | 10 | |

<210> 66
 <211> 45
 <212> ADN
 <213> secuencia artificial
 <220>
 <223> oligox-SDR1r
 <220>
 <221> misc_feature
 <222> (16)..(33)
 <223> cualquier aminoácido

30

<400> 66
 cgcccgggtga cgatgnnnnn nnnnnnnnnn nnnnttcaggg cctag 45

35

<210> 67
 <211> 47
 <212> ADN
 <213> secuencia artificial
 <220>
 <223> oligox-SDR2f
 <220>
 <221> CDS
 <222> (2)..(96)
 <223>
 <220>
 <221> misc_feature
 <222> (29)..(43)
 <223> cualquier resto nucleotídico o aminoacídico

40

<400> 67

c aag tgc ctc atc tct ggc tgg ggc aac nnn nnn nnn nnn nnn act g 47
Lys Cys Leu Ile Ser Gly Trp Gly Asn Xaa Xaa Xaa Xaa Xaa Thr
1 5 10 15

45

50

55

ES 2 373 658 T3

<210> 68
 <211> 15
 <212> PRT
 <213> secuencia artificial
 5 <220>
 <221> misc_feature
 <222> (10)..(10)
 <223> El "Xaa" en la localización 10 representa Lys, Asn, Arg, Ser, Thr, Ile, Met, Glu, Asp, Gly, Ala, Val, Gln, His, Pro, Leu, un codón de terminación, Tyr, Trp, Cys, o Phe.
 10 <220>
 <221> misc_feature
 <222> (11)..(11)
 15 <223> El "Xaa" en la localización 11 representa Lys, Asn, Arg, Ser, Thr, Ile, Met, Glu, Asp, Gly, Ala, Val, Gln, His, Pro, Leu, un codón de terminación, Tyr, Trp, Cys, o Phe.
 <220>
 <221> misc_feature
 20 <222> (12)..(12)
 <223> El "Xaa" en la localización 12 representa Lys, Asn, Arg, Ser, Thr, Ile, Met, Glu, Asp, Gly, Ala, Val, Gln, His, Pro, Leu, un codón de terminación, Tyr, Trp, Cys, o Phe.
 <220>
 <221> misc_feature
 25 <222> (13)..(13)
 <223> El "Xaa" en la localización 13 representa Lys, Asn, Arg, Ser, Thr, Ile, Met, Glu, Asp, Gly, Ala, Val, Gln, His, Pro, Leu, un codón de terminación, Tyr, Trp, Cys, o Phe.
 <220>
 <221> misc_feature
 30 <222> (14)..(14)
 <223> El "Xaa" en la localización 14 representa Lys, Asn, Arg, Ser, Thr, Ile, Met, Glu, Asp, Gly, Ala, Val, Gln, His, Pro, Leu, un codón de terminación, Tyr, Trp, Cys, o Phe.
 35 <220>
 <223> oligox-SDR2f
 <220>
 <221> misc_feature
 40 <222> (29)..(43)
 <223> cualquier resto nucleotídico o aminoacídico
 <400> 68

| | | | | | | | | | | | | | | |
|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|
| Lys | Cys | Leu | Ile | Ser | Gly | Trp | Gly | Asn | Xaa | Xaa | Xaa | Xaa | Xaa | Thr |
| 1 | | | | 5 | | | | | 10 | | | | | 15 |

 45 <210> 69
 <211> 55
 <212> ADN
 50 <213> secuencia artificial
 <220>
 <223> oligox-SDR2r
 <220>
 <221> misc_feature
 55 <222> (33)..(47)
 <223> cualquier base
 <220>
 <221> misc_feature
 60 <222> (33)..(47)
 <223> cualquier nucleótido
 <400> 69

ES 2 373 658 T3

catggtcac ggagtagaga cgcaccccg tgnnnnnnn nnnnnntga cgatc 55

<210> 70
 <211> 59
 5 <212> ADN
 <213> secuencia artificial
 <220>
 <223> cebador SDR1-mutnb-directo
 <220>
 10 <221> misc_feature
 <222> (24)..(40)
 <223> N=A, C, G, T; B=C, G, T; V=A, C, G

<400> 70
 15 tggatccgc gggccactgc tacnbnbn nbnnbnbn baagtcccg atccagtg 59

<210> 71
 <211> 52
 <212> ADN
 20 <213> secuencia artificial
 <220>
 <223> cebador SDR2-mutnb-inverso
 <220>
 25 <221> misc_feature
 <222> (20)..(33)
 <223> N=A, C, G, T; B=C, G, T; V=A, C, G

<400> 71
 30 ggcgccagag ctagcagtvn nvnnvnnvnn vnngtgccc cagccagaga tg 52

<210> 72
 <211> 6
 <212> PRT
 35 <213> secuencia artificial
 <220>
 <223> variante g SDR1

<400> 72
 40

Ala Phe Phe Asn Gly Asp

1 5

<210> 73
 <211> 5
 45 <212> PRT
 <213> secuencia artificial

<220>
 <223> variante g SDR2
 50

<400> 73

Arg Lys Asp Pro Trp

1 5

55 <210> 74
 <211> 234
 <212> PRT
 <213> secuencia artificial

ES 2 373 658 T3

<220>
<223> secuencia artificial

<400> 74

5

```

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

```

<210> 75
<211> 234
<212> PRT
<213> secuencia artificial

10

<220>
<223> secuencia artificial

15

<400> 75

ES 2 373 658 T3

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

<400> 75
 ggccagag ctacagctnn nnnnnnnnnn nngtgccc cagccagaga tg 52

5
 <210> 76
 <211> 12
 <212> PRT
 <213> secuencia artificial

10
 <220>
 <223> sustrato A

15
 <400> 76

Leu Leu Trp Leu Gly Arg Val Val Gly Gly Pro Val
 1 5 10

ES 2 373 658 T3

<210> 77
 <211> 12
 <212> PRT
 5 <213> secuencia artificial
 <220>
 <223> sustrato B

<400> 77
 10

Lys Lys Trp Leu Gly Arg Val Pro Gly Gly Pro Val
 1 5 10

<210> 78
 <211> 6
 15 <212> PRT
 <213> secuencia artificial
 <220>
 <223> variante 1 SDR1

<400> 78
 20

Asp Ala Val Gly Arg Asp
 1 5

<210> 79
 <211> 6
 25 <212> PRT
 <213> secuencia artificial
 <220>
 <223> variante 2 SDR1

<400> 79
 30

Asn Gly Arg Asp Leu Glu
 1 5

<210> 80
 <211> 6
 35 <212> PRT
 <213> secuencia artificial
 <220>
 40 <223> variante 3 SDR1

<400> 80

Gly Phe Val Met Phe Asn
 1 5

<210> 81
 <211> 5
 45 <212> PRT
 <213> secuencia artificial
 <220>
 50 <223> variante 1 SDR2

<400> 81

ES 2 373 658 T3

Arg Val His Pro Ser
1 5

5 <210> 82
<211> 5
<212> PRT
<213> secuencia artificial
<220>
<223> variante 2 SDR2

10 <400> 82

Val Arg Gly Thr Trp
1 5

15 <210> 83
<211> 5
<212> PRT
<213> secuencia artificial
<220>
<223> variante 3 SDR2

20 <400> 83

Arg Ser Pro Leu Thr
1 5

25 <210> 84
<211> 6
<212> PRT
<213> secuencia artificial
<220>
30 <223> variante a SDR1

<400> 84

Arg Pro Trp Asp Pro Ser
1 5

35 <210> 85
<211> 6
<212> PRT
<213> secuencia artificial
40 <220>
<223> variante b SDR1

<400> 85

Gly Phe Val Met Phe Asn
1 5

45 <210> 86
<211> 6
<212> PRT
50 <213> secuencia artificial
<220>
<223> variante c SDR1

55 <400> 86

ES 2 373 658 T3

Glu Ile Ala Asn Arg Glu
1 5

5 <210> 87
<211> 6
<212> PRT
<213> secuencia artificial
<220>
<223> variante d SDR1

10 <400> 87

Lys Ala Val Val Gly Thr
1 5

15 <210> 88
<211> 6
<212> PRT
<213> secuencia artificial
<220>
<223> variante e SDR1

20 <400> 88

Val Asn Ile Met Ala Ala
1 5

25 <210> 89
<211> 6
<212> PRT
<213> secuencia artificial
<220>
30 <223> variante f SDR1

<400> 89

Ala Ala Phe Asn Gly Asp
1 5

35 <210> 90
<211> 5
<212> PRT
<213> secuencia artificial
40 <220>
<223> variante a SDR2

<400> 90

Val His Pro Thr Ser
1 5

45 <210> 91
<211> 5
<212> PRT
50 <213> secuencia artificial
<220>
<223> variante b SDR2

<400> 91

Arg Ser Pro Leu Thr
1 5

5

<210> 92
 <211> 5
 <212> PRT
 <213> secuencia artificial
 <220>
 <223> variante c SDR2

10

<400> 92

Arg Gly Ala Arg Thr
1 5

15

<210> 93
 <211> 5
 <212> PRT
 <213> secuencia artificial
 <220>
 <223> variante d SDR2

20

<400> 93

25

Arg Thr Pro Ile Ser
1 5

<210> 94
 <211> 5
 <212> PRT
 <213> secuencia artificial
 <220>

30

<223> variante e SDR2

35

<400> 94

Thr Thr Ala Arg Lys
1 5

40

<210> 95
 <211> 5
 <212> PRT
 <213> secuencia artificial
 <220>

45

<223> variante f SDR2

<400> 95

Arg Lys Asp Phe Trp
1 5

50

<210> 96
 <211> 157
 <212> PRT
 <213> Homo sapiens

55

<400> 96

ES 2 373 658 T3

Val Arg Ser Ser Ser Arg Thr Pro Ser Asp Lys Pro Val Ala His Val
 1 5 10 15
 Val Ala Asn Pro Gln Ala Glu Gly Gln Leu Gln Trp Leu Asn Arg Arg
 20 25 30
 Ala Asn Ala Leu Leu Ala Asn Gly Val Glu Leu Arg Asp Asn Gln Leu
 35 40 45
 Val Val Pro Ser Glu Gly Leu Tyr Leu Ile Tyr Ser Gln Val Leu Phe

50 55 60
 Lys Gly Gln Gly Cys Pro Ser Thr His Val Leu Leu Thr His Thr Ile
 65 70 75 80
 Ser Arg Ile Ala Val Ser Tyr Gln Thr Lys Val Asn Leu Leu Ser Ala
 85 90 95
 Ile Lys Ser Pro Cys Gln Arg Glu Thr Pro Glu Gly Ala Glu Ala Lys
 100 105 110
 Pro Trp Tyr Glu Pro Ile Tyr Leu Gly Gly Val Phe Gln Leu Glu Lys
 115 120 125
 Gly Asp Arg Leu Ser Ala Glu Ile Asn Arg Pro Asp Tyr Leu Leu Phe
 130 135 140
 Ala Glu Ser Gly Gln Val Tyr Phe Gly Ile Ile Ala Leu
 145 150 155

REIVINDICACIONES

1. Un procedimiento para generar una enzima proteolítica que tenga especificidad definida no conferida por el armazón proteico para al menos un sustrato diana que comprende al menos las siguientes etapas:

- 5 (a) proporcionar un armazón proteico que tenga al menos 70 % de homología con tripsina humana I que tenga la secuencia de aminoácidos mostrada en SEC ID N°: 1, que catalice al menos una reacción química en al menos un sustrato,
- 10 (b) generar una biblioteca de enzimas proteolíticas o enzimas proteolíticas aisladas combinando un polinucleótido que codifica el armazón proteico de la etapa (a) mediante inserción o sustitución con 1 a 11 regiones determinantes de especificidad (SDR), en la que las SDR son secuencias oligonucleotídicas sintéticas completa o parcialmente aleatorias que codifican secuencias peptídicas con una longitud de menos de 50 restos aminoacídicos en una o más posiciones del grupo de posiciones dentro del polinucleótido que codifica el armazón proteico que corresponde estructuralmente o por homología de secuencia de aminoácidos a las regiones 18-25, 38-48, 54-63, 73-86, 122-130, 148-156, 165-171 y 194-204 en tripsina humana I que tiene la secuencia de aminoácidos mostrada en SEC ID N°: 1, expresar dichas enzimas y
- 15 (c) seleccionar de la biblioteca de enzimas proteolíticas generada en la etapa (b) una o más enzimas que tengan especificidades definidas no conferidas por el armazón proteico proporcionado en la etapa (a) para al menos un sustrato diana.

2. El procedimiento de acuerdo con la reivindicación 1, en el que las secuencias peptídicas insertadas o sustituidas en la etapa (b) son completa o parcialmente aleatorias y/o tienen una variación de longitud; y/o en el que la selección en la etapa (c) se consigue cribando con respecto a la actividad enzimática y/o la afinidad enzimática

- 25 (i) en concentraciones de sustrato diana bajas,
- (ii) usando el sustrato diana y al menos un sustrato más para comparación,
- (iii) añadiendo en exceso otros sustratos distintos del sustrato diana, usando este modo los sustratos diana añadidos como competidores,
- (iv) añadiendo inhibidores de enzima,
- (v) seleccionando enzimas que se unen preferentemente al sustrato diana y seleccionado de este subgrupo las enzimas que convierten el sustrato o
- (vi) cualquier combinación de las mismas.

3. El procedimiento de acuerdo con la reivindicación 1, que comprende al menos las siguientes etapas:

- 30 (a) proporcionar un primer fragmento de armazón proteico,
- (b) conectar dicho fragmento de armazón proteico mediante un engarce peptídico con una primera región determinante de especificidad, y opcionalmente
- (c) conectar el producto de la etapa (b) mediante un engarce peptídico con un péptido de región determinante de especificidad adicional o con un fragmento de armazón proteico adicional, y opcionalmente
- 35 (d) repetir la etapa (c) durante tantos ciclos como sea necesario para generar una enzima suficientemente específica y
- (e) seleccionar de la población generada en las etapas (a) - (d) una o más enzimas que tengan las especificidades deseadas para el uno o más sustratos que no se confieren por el fragmento de armazón proteico proporcionado de la etapa (a).

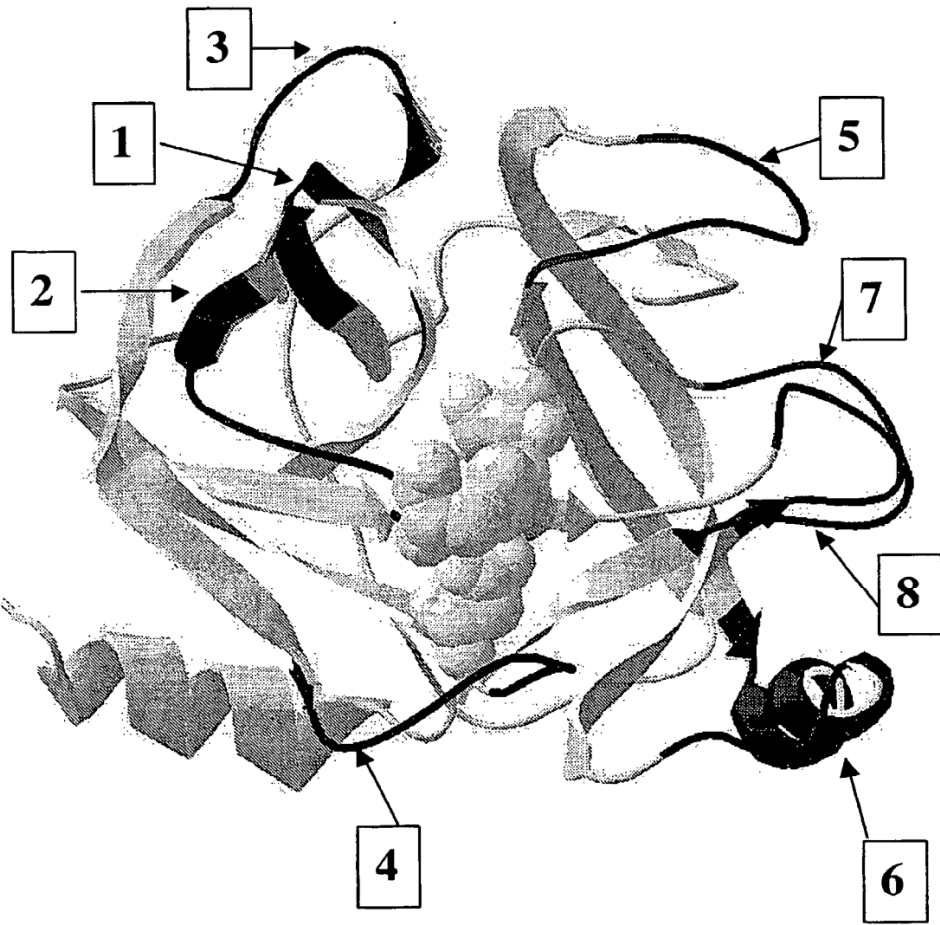


Fig. 1

ES 2 373 658 T3

| | |
|--|---|
| <p>Tripsina a-trombina enteropeptidasa</p> | <p>IVGGYNCEENSVPYQVSL---NSGYHF-CGGSLINEQWVVSAGHCY---- IVEGSDAEIGMSPWQVMLFRKSPQELL-CGASLISDRWVLTAAHCLLYPP IVGGSNAKEGAWPWVVGGL---YGGRLLCGASLVSSDWLVSAAHCVYGRN ** * * * * ** * * * * ---</p> |
| <p>Tripsina a-trombina enteropeptidasa</p> | <p>-----KSRIQVRLGEH---NIEVLEGN-EQFINAAKIIRHPQYD-RKTL WDKNFTENDLLVRIGKH---SRTRYERNIEKISMLEKIYIHPRYNWREN LE----PSKWTAILGLHMKSNTSPQTV-PRLID--EIVINPHYN-RRRK -1---- * * * * *</p> |
| <p>Tripsina a-trombina enteropeptidasa</p> | <p>NNDIMLIKLSRAVINARVSTISLPTA----PPAT-----GTKCLISGWG DRDIALMKLKKPVAFSDYIHPVCLPDR----ETAASLLQAGYKGRVTGWG DNDIAMMHLEFKVNYTDYIQPICLPEENQVFPP-----GRNC SIAGWG ** * * * * ---2-----* ***</p> |
| <p>Tripsina a-trombina enteropeptidasa</p> | <p>N-----TASSGADYPDELQCLDAPVLSQAKCEASYPG-KITSNMFCVGF NLKETWTANVGKQPSVLQVVNLPIVERPVCKDSTRI-RITDNMFCAGYK T-----VVYQGIT-ANILQEADVPLLSNERCQQQMPPEYNI TENMICAGYE --3-- * * * * * * * * *</p> |
| <p>Tripsina a-trombina enteropeptidasa</p> | <p>-EGGK--DSCQGDSSGGPVVVCNGQ----LQ-----GVVSWGDGCAQKNKP PDEGKRGDACEGDSGGPFVVMKSP----FNNRWYQMGIVSWGEGCDRDGKY -EGGI--DSCQGDSSGGPLMCQENNRWFLA-----GVTSFGYKCALPNRP * * * * * * * * * * ---4-----* * * *</p> |
| <p>Tripsina a-trombina enteropeptidasa</p> | <p>GVYTKVYNYVKWIKNTIAANS- GFYTHVFRLLKWIQKVIDQFGE GYYARVSRFTEWIQSFLH---- * * * * *</p> |

Fig. 2

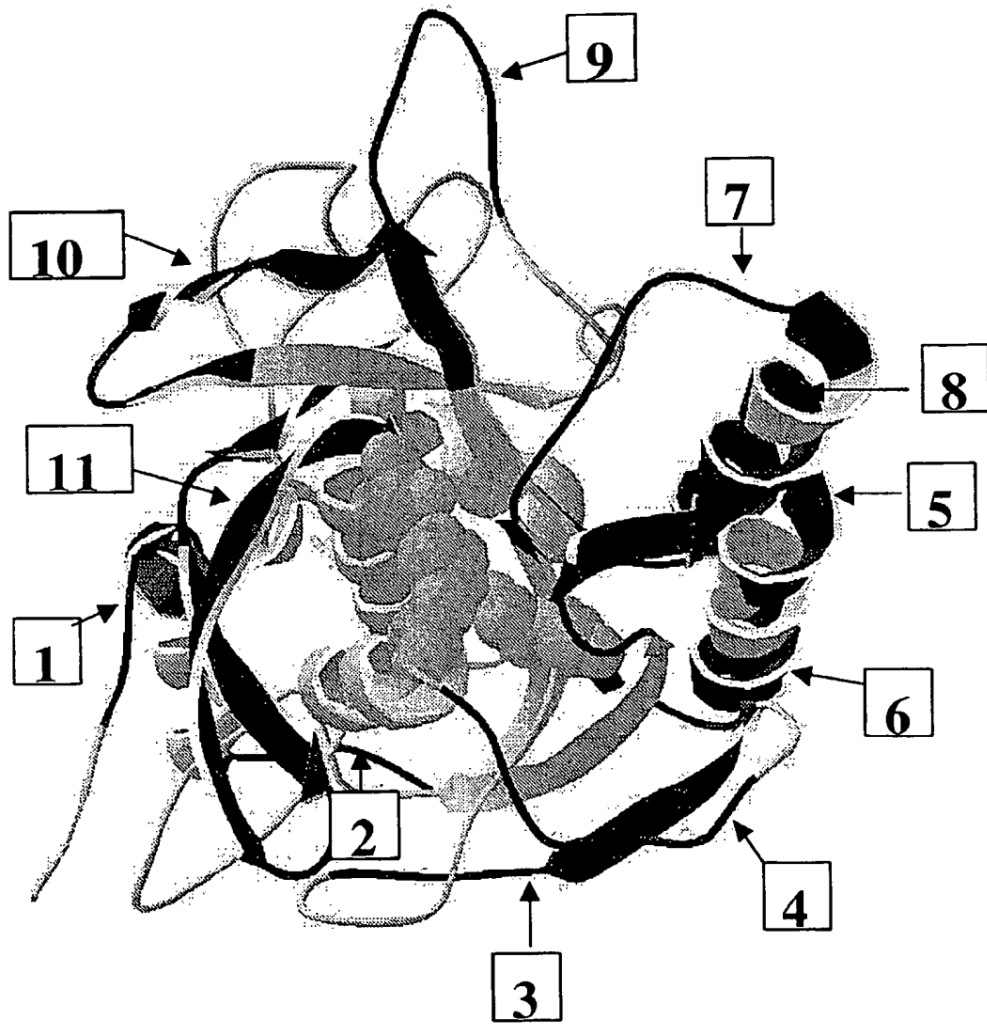


Fig. 3

ES 2 373 658 T3

```

sub      IAHEYAQSV--PY-----GISQ--IKAPALHSQGY-----
furina   VAKRRRAKRD--VYQEPTDPKFPQQWYLSGVTQRDLNVKEAWAQGF-----
PC_SK1   EKERSKRSALRDS-----ALNL--FNDPMWNQQWYLQDTRMTAALPKLDL
PC_SK5   NTHPCQ-----SD--MNIEGAWKRGY-----
          -----1-----2--

sub      -----TGSNVKVAVIDSGIDSSHPDL--NVRGGAS--FVPSETN-----P--
furina   -----TGHGIVVSILDDGIEKNHPDLAGNYDPGAS--FDVNDQD-----PDPQ
PC_SK1   HVIPVWQKGITGKGVVITVLDDGLEWNHTDIYANYDEASYDFNDNDHD-----P--
PC_SK5   -----TGKNIVVTILDDGIERTHPDL--MQNYDA--LASCDVNGNDLDPMP-----
          -----* * * *-----3-----

sub      ----YQ----DGSS--HGTHVAGTIA--AL--NNSIGVLGVSPSASLYAVKVLDS----
furina   PRYTQM----NDNR--HGTRCAGEVA--AVANNGVCGVGVAYNARIGGVRMLD----
PC_SK1   ----FPRYDPTNENK--HGTRCAGEIAMQAN--HHKCGV--GVAYNSKVGGIRMLDG----
PC_SK5   ----RY----DASNENKHGTRCAGEVA--AAANNSHCTVGIAFNAKIGGVRMLDGDVTD
          -----4-----* * * * *-----

sub      -TGSGQYSWIINGIE-WAISNMMDVINMSLG-----GPT--GSTA-----LKT--
furina   -GEVTDAVEARS-LGLNPNHIIHIYSASW-----GPEDDGKTVDGPARLAE--
PC_SK1   -IVTDAIEASSIGFN--PGHVDIYSASWGPNDDGKTVEGP--GRLA-----QKAFE
PC_SK5   MVEAKSVSFNPQHVHIYSASWGPDDDGKTVD-----GPA--PLT-----RQ--
          -5-----6-----7-----8--

sub      --VVDKAVSSG-----IVVAAAAGNEGSS-----GSTSTVGYPAKYPSTIAVGAV---
furina   --AFFRGVSQGRGGLGSIFVWASGNGGREHDSCHCDGYTNSI-YTLSISSATQFGNV---
PC_SK1   YGVKQGRQKG-----SIFVWASGNGGRQ-----GDNCDCD---GYTDSIYTISI---
PC_SK5   --AFENGVRMGRRGLGSVFVWASGNGGRSKDHCSCDGYTNSI-YTISISSTAESGKKPWY
          -----8-----*-----9-----

sub      --N-----SSNQR-----ASFSSAG-SELDVMAPGVSIQSTLPGGTYGAY
furina   --PWYSEACSSTLA-----TTYSSGNQNEKQIVTDLRQKCT-----ESH
PC_SK1   --S-----SASQGLSPWYAEKCSSTLATSYSSG-DYTDQRITSADLHNDCT----ETH
PC_SK5   LEE-----CSSTL-----ATTYSSG-ESYDKKI----ITDLRQRCTDNH
          -----10-----*-----11-----

sub      NGTSMATPHVAGAAALIL--SKHP--TWTNAQVRDRLESTATY--LG-NSFYYGKGLINV
furina   TGTSASAPLAAGIIALTLEANKNL--TWRDMQHLVVQTSKPAH--LN-ADDWATNGVGRK
PC_SK1   TGTSASAPLAAGIFALAL--EANP--NLTWRDMQHLVVWTSEYDPLA--NNPGWKKNGAGL
PC_SK5   TGTSASAPMAAGIIALAL--EANPFLIWRDVQHVIVRTSRAGH--LNANDWKINAAGFKV
          ---* * * * *-----*-----

```

Fig. 4

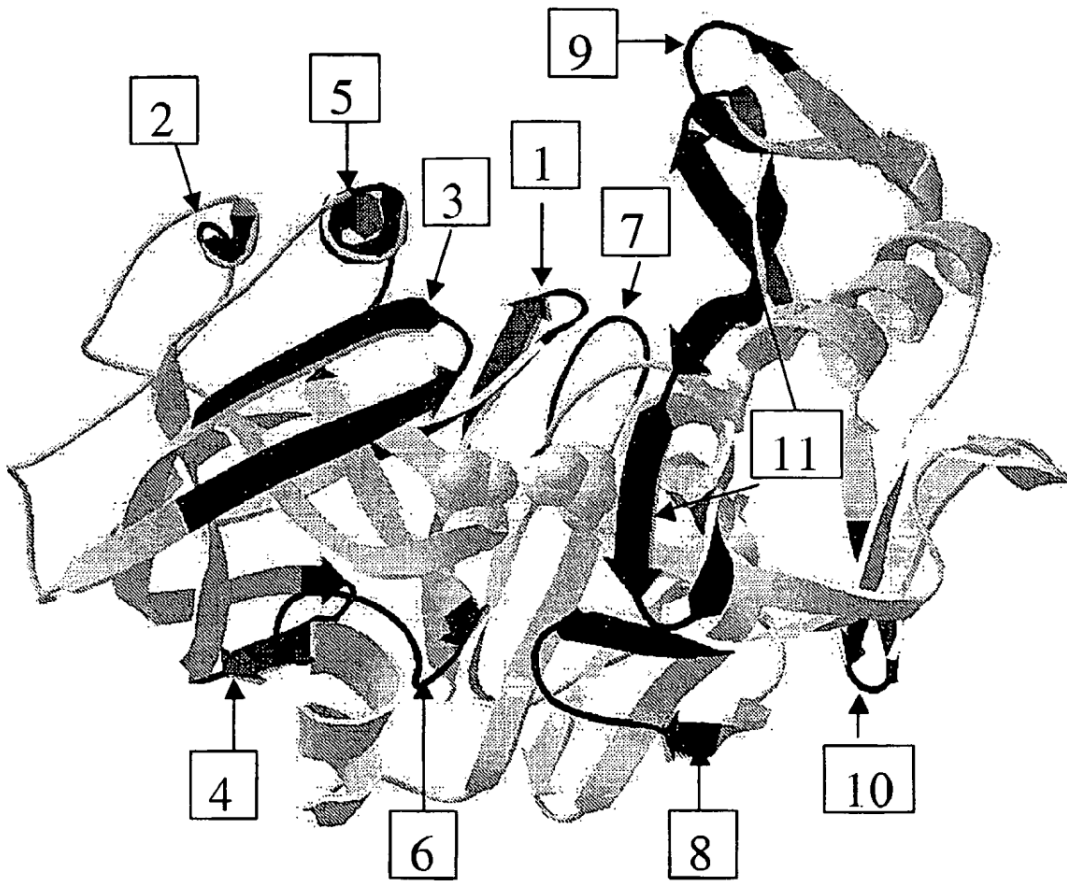


Fig.5

ES 2 373 658 T3

Peps. TLVDEQP----LENYLDMEYFGTIGIGTPAQDFTVVFDTGSSNLWVPSVYCSSL--ACTN
 Secr. EMVDN-----LRGKSGQGYVEMTVGSPPTLNILVDTGSSNFVGAAPHFPL-----
 Cath. PAVTEGPIPEVLKNYMDAQYYGEIGIGTPPQCFTVVFDTGSSNLWVPSIHCKLLDIACWI
 * ----1----- * * * * * ***** * -----2--

Peps. HNRFPEDSSTYQSTSETVSITYGTGSMTGILGYDTVQV-----G--GISDTN
 Secr. HRYRQQLSSTYRDLRKGVYVPTYQKWEGLGTDLVSI-----PHGPNVTVRA
 Cath. HHKYNSDKSSTYVKNGTSFDIHYGSGSLSGYLSQDTSVPCQSASSASALG--GVKVER
 - **** -----3----* * * -----4-----

Peps.: QIFGLSETEPGSFLYYAPFDGILGLAYPSIS--SSGATPVFDNIWNQGLVSQDLFSVYLS
 Secr. NIAAITESDK-FFINGSNWEGILGLAYAEIARPDSSLEPFDSLVKQTHVP-NLFLSLQLC
 Cath. QVFGAATKQPGITFIAAKFDGILGMAYPRIS--VNNVLPVFDNLMQQLVDQNIFSFYLS
 -----5-----***** -----6----- ** * * * *

Peps. ADD-----KS--GSVVIFGGIDSSYYTGS LNWVPVIVEGYWQITVDSITMNGETI
 Secr. GAGFPLNQSEVLASV--GGSMIIGGIDHSLYTGSLWYTPIRREWYVEVI IVRVEINGQDL
 Cath. RDP-----DAQPGELMLGGTDSKYYKGSLSYLNVTBKAYWQVHLDQVEVASGLT
 -----7----- ** * * * * * -----8-----

Peps. A--CAEGC--QAIVDGTGSLLTGPTSPIANIQSDIGASENSD-----GDMVVSCSAI
 Secr. KMDCKEYNYDKSIVDSGTTNLRRLPKKVFEEAVKSIKAASSTEKFPDGFWLGEQLV-CWQA
 Cath. L--CKEGC--EAIVDGTGSLMVGPVDEVRELQKAI GAVPLIQ-----GEYMIPCEKV
 * * **** * * * * * -----9----- *

Peps. SSLPDIVFTI-----NGVQYPVPPSAYILQSEGS----CISGFQGMNVP-TESG
 Secr. GTPWNIFPVISLYLMGEVTNQSFRTILPQQYLRPVEDV----ATSQDDCYKFAISQSS
 Cath. STLPAITLKL-----GGKGYKLSPEYTLKVSQAGKTLCLSGFMGMMDIP-PPSG
 * -----10----- * * -----11-----

Peps. ELWILGDVFI RQYFTVFD RANNQVGLAPVA
 Secr. TGTVMGAVIMEGFYVVFDRARKRIGFAVSA
 Cath. PLWILGDVFI RYTYVFD RDNMRVGF AEA
 -- * * **** * * *

Fig. 6

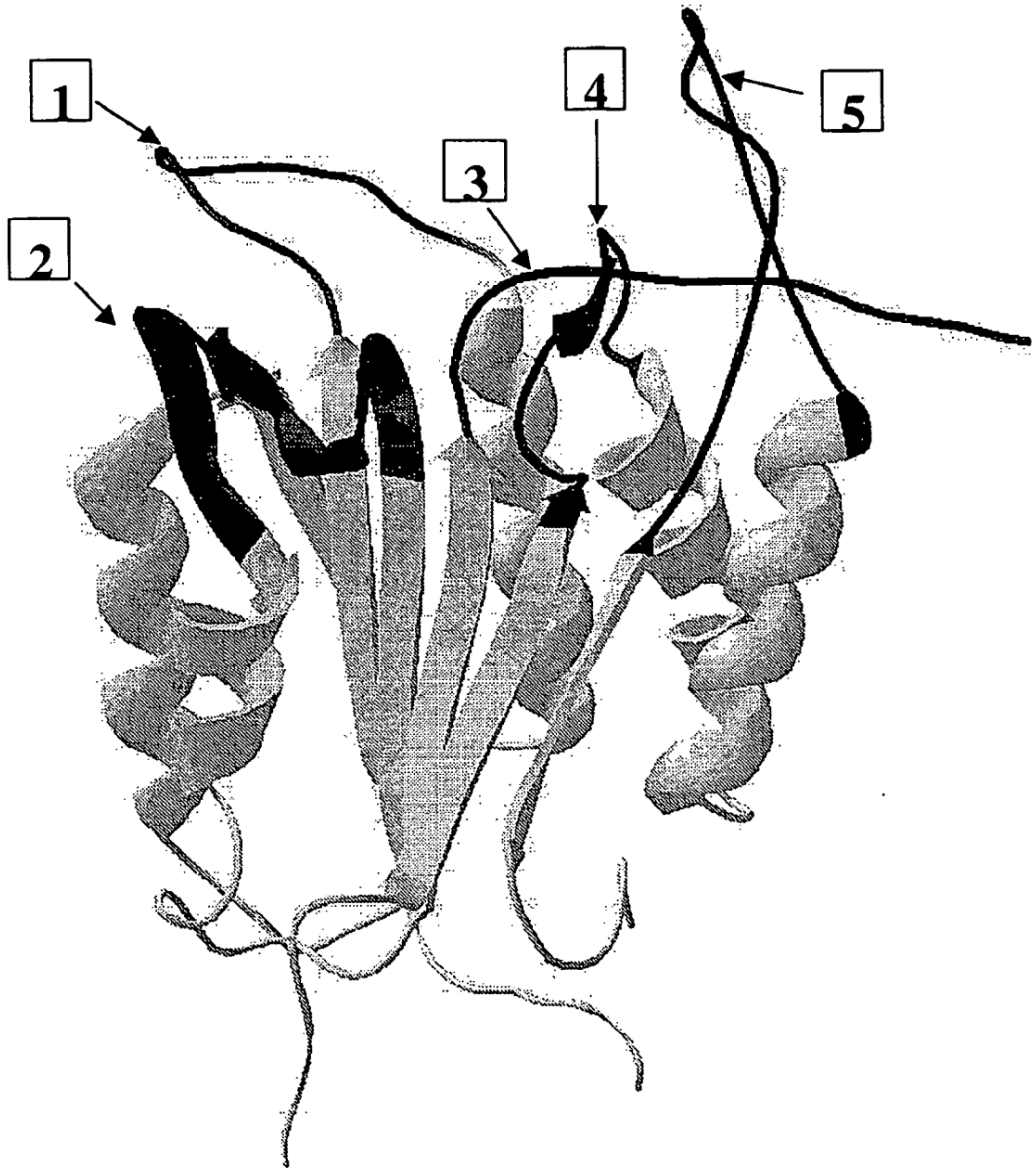


Fig. 7

ES 2 373 658 T3

01 MLEADDQGCI EEQGVEDSAN EDSVDAKPDR SSFVPSLFSK KKNVTMRSI KTTRDRVPTY
61 QYNMFEKLG KCIIINNKNF DKVIGMGVRN GTDKDAEALF KCFRSLGFDV IVYNDSCAK
-----1-----
121 MQDLLKASE EDHTNAACFA CILLSHGEEN VIYGKDGVTP IKDLTAHFRG DRSKILLEKP
-----2-----
181 KLFFIQACRG TELDDGIQAD SGPINDTDAN PRYKIPVEAD FLFAYSTVPG YYSWRSPGRG
-----3----- -----4-----
241 SWFVQALCSI LEEHGKDLEI MQILTRVNR VARHFESQSD DPHFHEKKQI PCVVSMLTKE
--- -----5-----
301 LYFSQ

Fig. 8

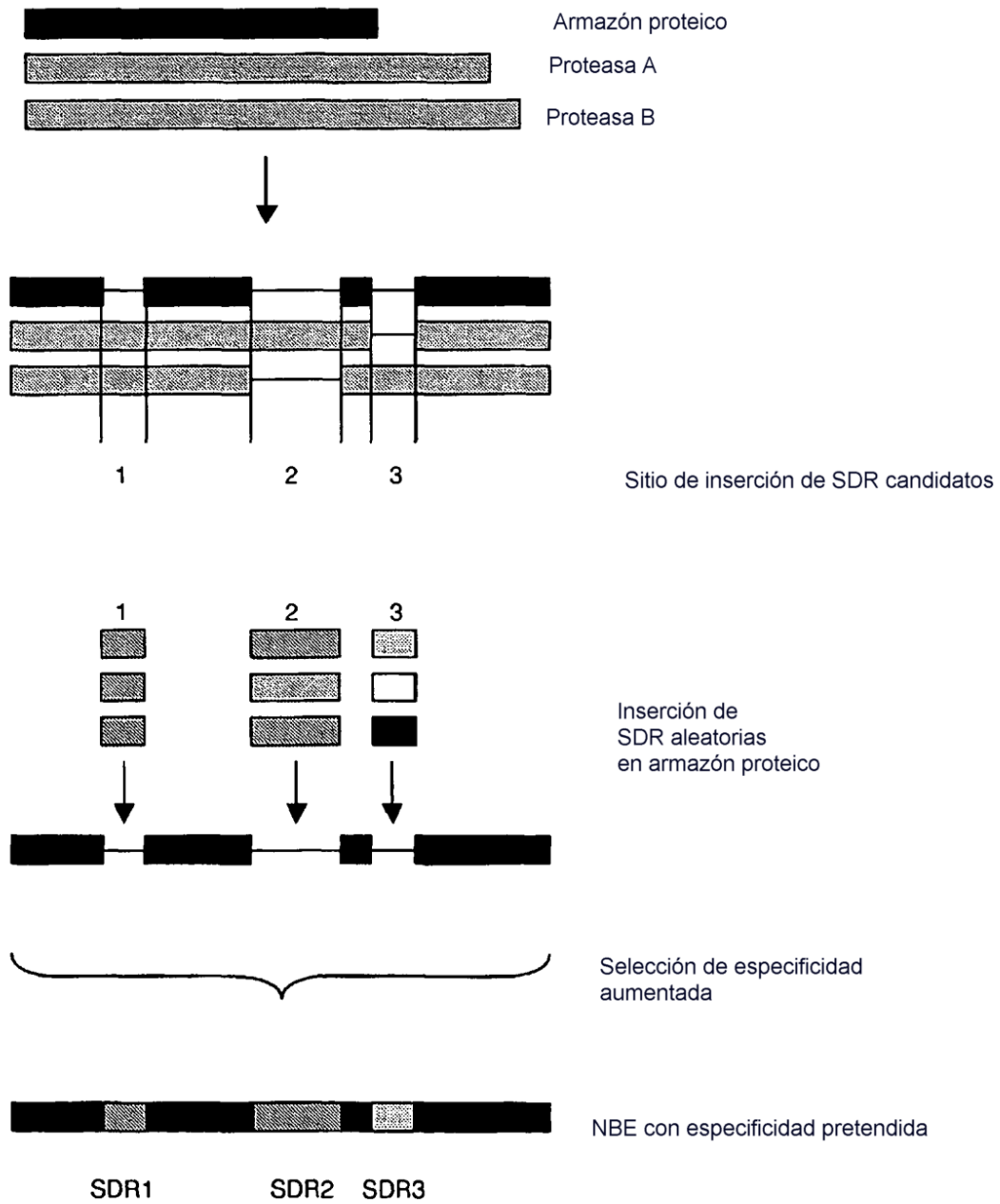


Fig. 9

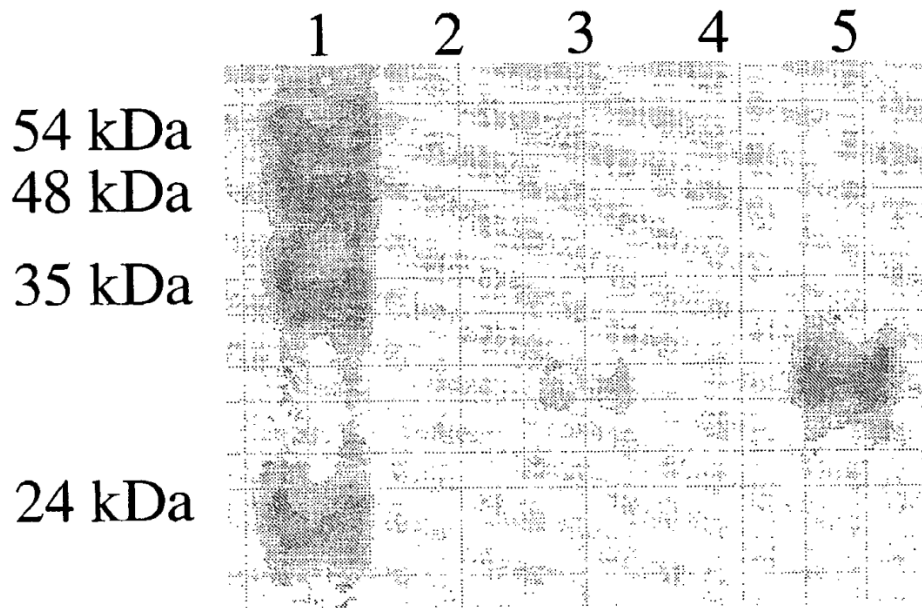


Fig. 10

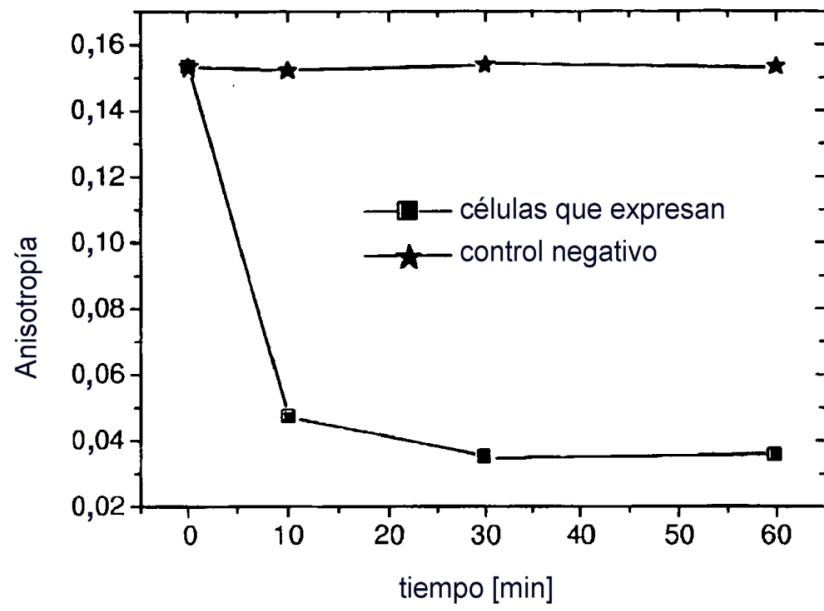


Fig. 11

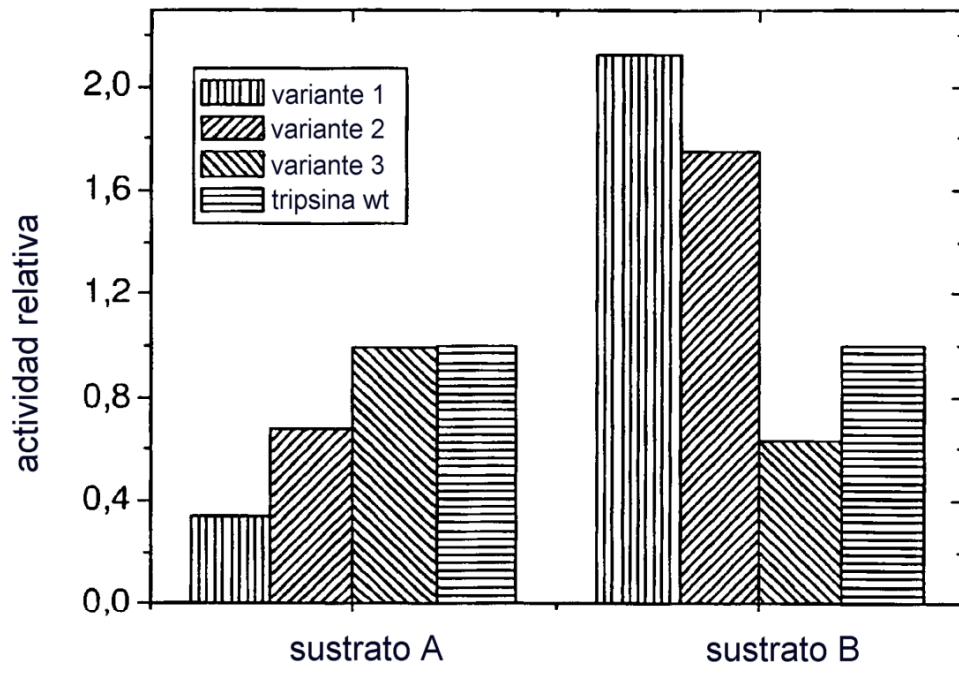


Fig. 12

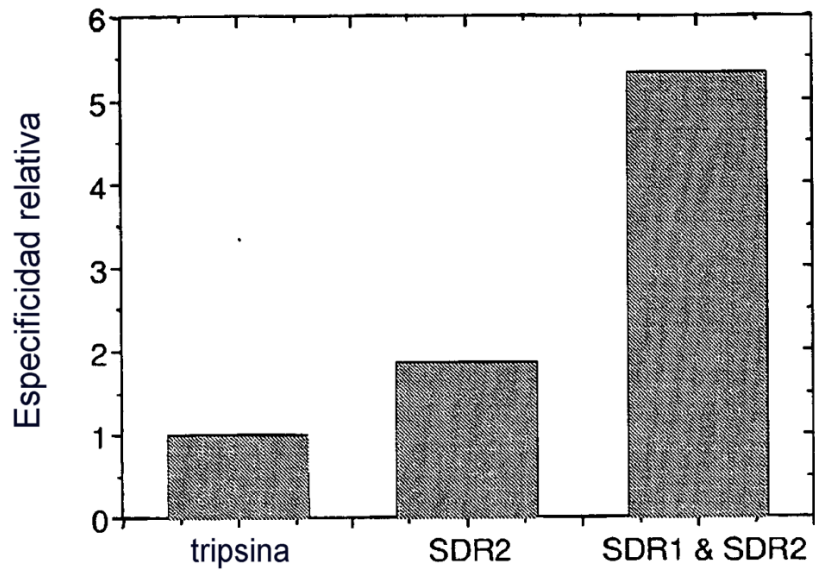


Fig. 13

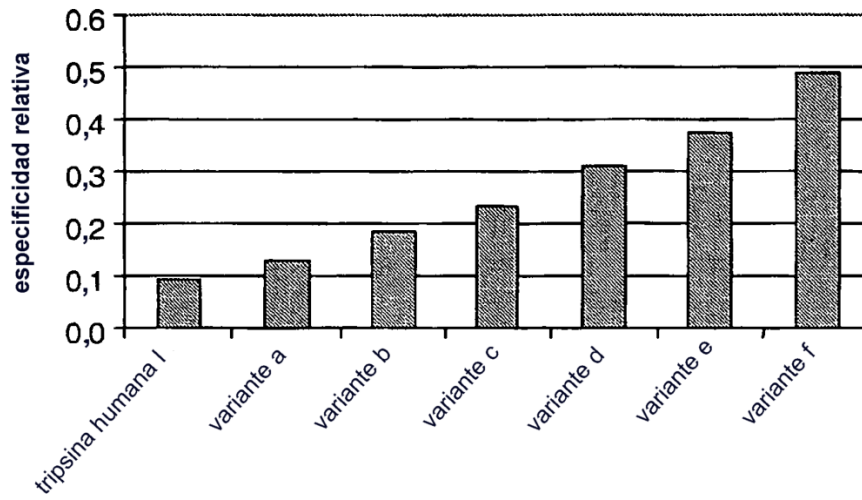


Fig. 14

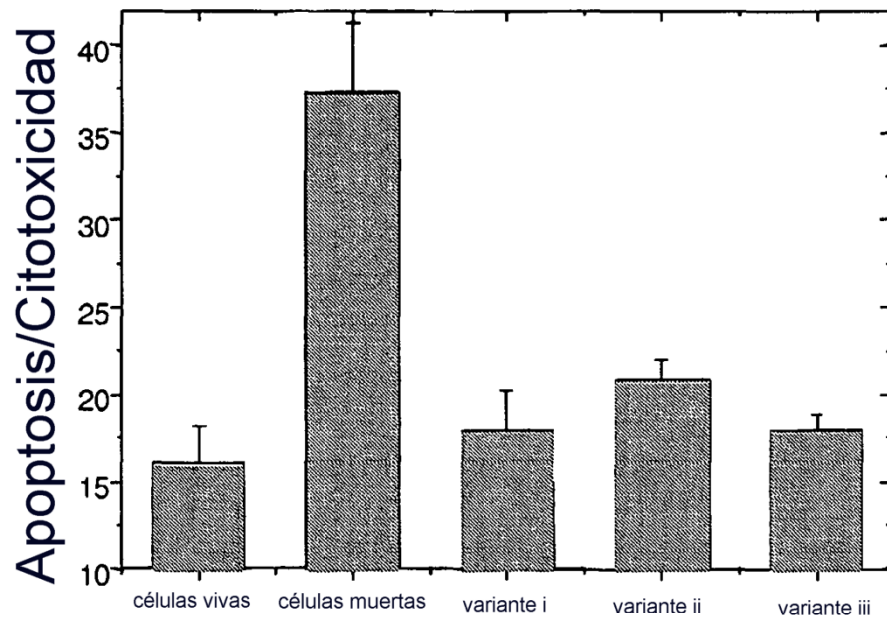


Fig. 15

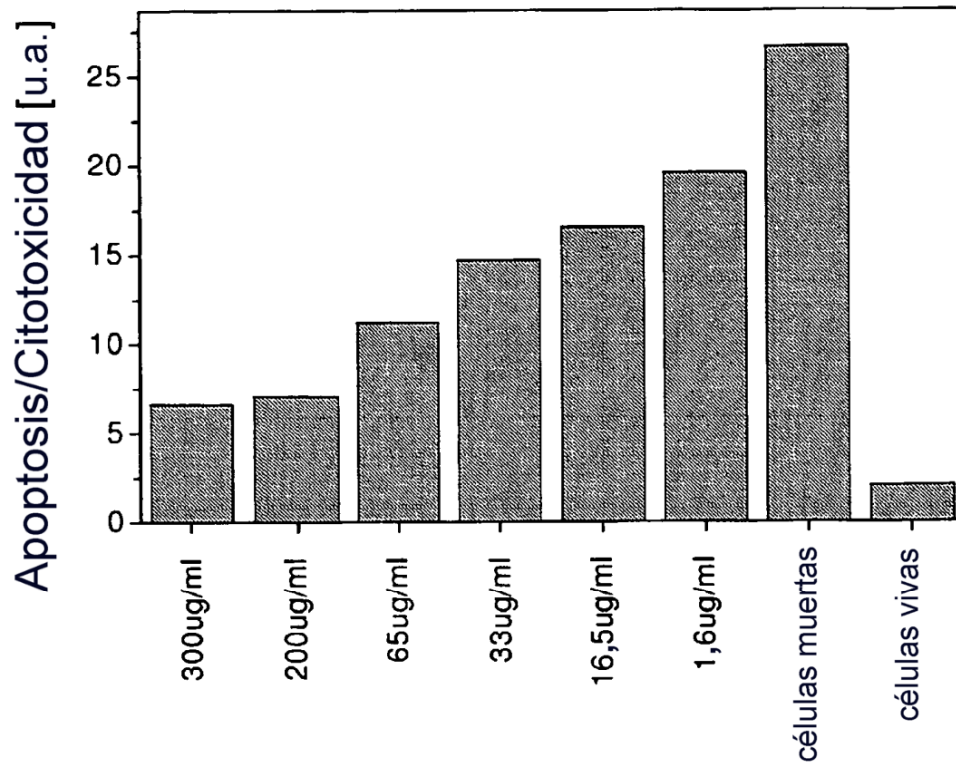


Fig. 16

digestión proteolítica de TNF

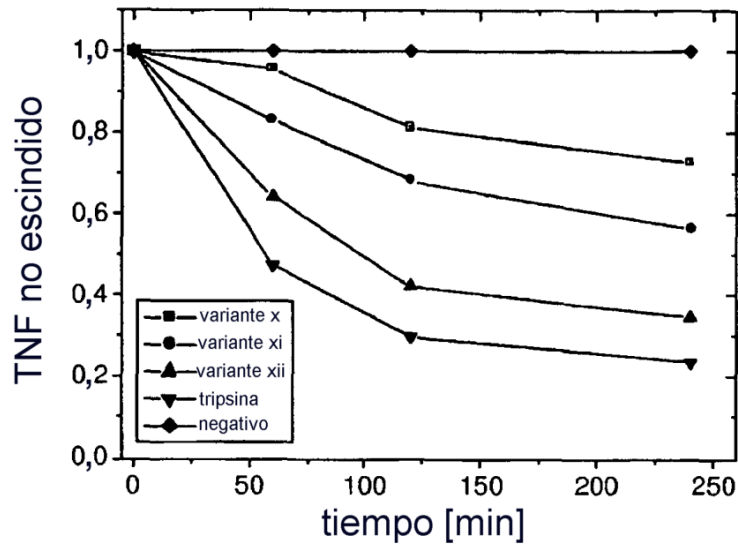


Fig. 17a

digestión proteolítica de proteína de suero

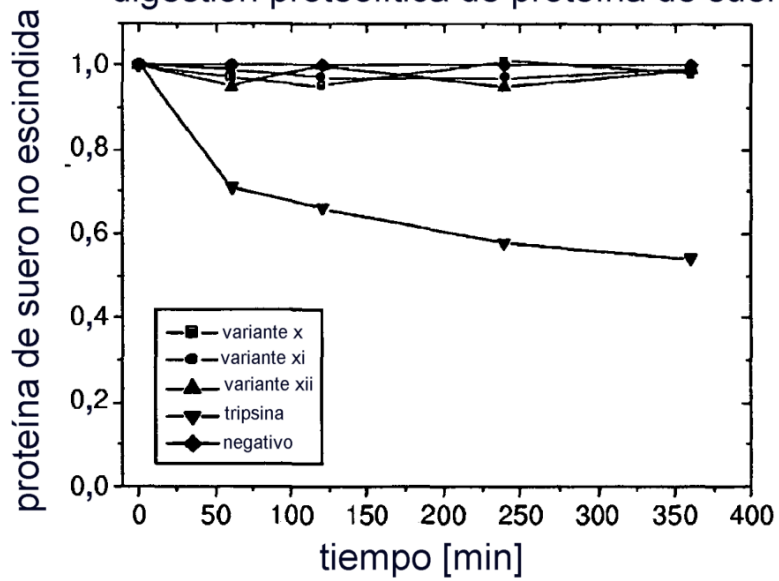


Fig. 17b

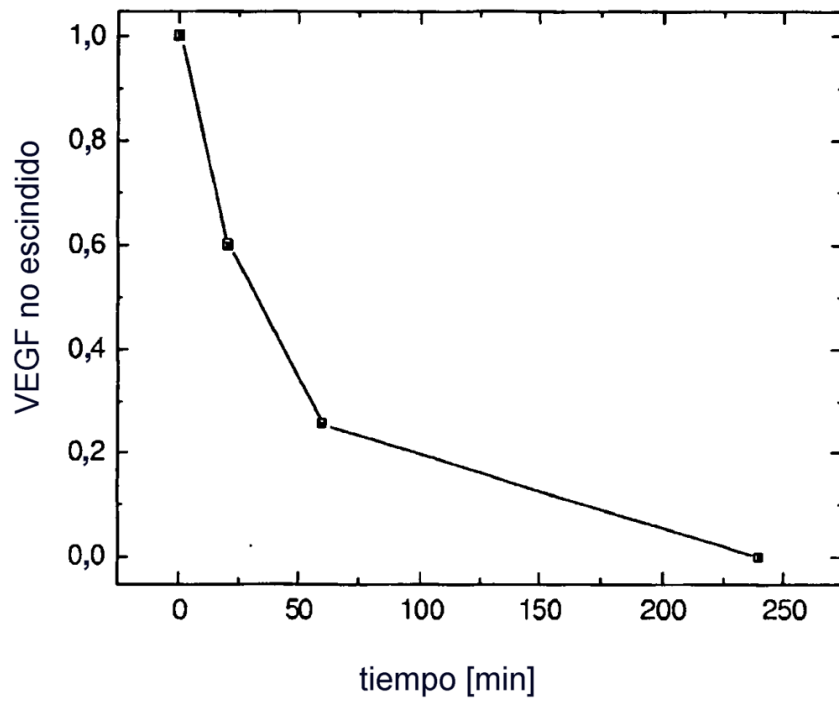


Fig. 18