

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 388 753**

51 Int. Cl.:
C12N 15/10 (2006.01)
C12N 15/01 (2006.01)
C12N 15/52 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

- 96 Número de solicitud europea: **07796357 .7**
96 Fecha de presentación: **22.06.2007**
97 Número de publicación de la solicitud: **2032698**
97 Fecha de publicación de la solicitud: **11.03.2009**

54 Título: **Evaluación sistemática de las relaciones entre secuencia y actividad usando bibliotecas de evaluación de sitios para la ingeniería de propiedades múltiples**

30 Prioridad:
23.06.2006 US 816202 P
06.06.2007 US 933312 P

45 Fecha de publicación de la mención BOPI:
18.10.2012

45 Fecha de la publicación del folleto de la patente:
18.10.2012

73 Titular/es:
Danisco US Inc.
925 Page Mill Road
Palo Alto, CA 94304, US

72 Inventor/es:
AEHLE, Wolfgang y
ESTELL, David A.

74 Agente/Representante:
Ponti Sales, Adelaida

ES 2 388 753 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

Evaluación sistemática de las relaciones entre secuencia y actividad usando bibliotecas de evaluación de sitios para la ingeniería de propiedades múltiples.

5

CAMPO DE LA INVENCION

[0001] La presente invención proporciona procedimientos para la ingeniería de proteínas. Específicamente, la invención proporciona procedimientos que utilizan bibliotecas de evaluación de sitios.

10

ANTECEDENTES DE LA INVENCION

[0002] Los expertos en la técnica conocen diversos procedimientos de ingeniería de proteínas. En general, las proteínas se modifican para obtener propiedades proteicas deseadas. En la mayoría de los procedimientos, la secuencia de nucleótidos de un gen clonado que codifica una proteína está mutada y el gen modificado se expresa para producir mutantes, que se analizan en búsqueda de actividades de interés. A menudo, las propiedades mutantes se comparan con las propiedades de la proteína natural.

15

[0003] Históricamente, el proceso de diseño de proteínas se ha enfocado como equivalente al problema de encontrar en todo el espacio de la proteína la mejor secuencia para la aplicación deseada. Este problema es extremadamente difícil y es "NP complejo". En la teoría de la complejidad, los problemas definidos como pertenecientes a la clase P se consideran algoritmos de tiempo polinomial fáciles y eficaces para los que existe solución. Los problemas NP complejos son problemas para los que actualmente no se conocen algoritmos de tiempo polinomial eficaces y si pudiera resolverse cualquier problema NP complejo, todos los problemas NP complejos podrían resolverse (Véase, p. ej., Pierce y Winfree, *Protein Engineer*, 15:779-782 [2002]). Las estrategias actuales para construir y analizar bibliotecas generalmente implican la generación de diversidad de secuencia de proteína de forma aleatoria a lo largo de la secuencia completa o de forma controlada al azar en posiciones definidas dentro de la proteína. Estas bibliotecas generalmente tienen un gran número de miembros que son "negativos" con respecto a la propiedad principal de interés, y requieren que se analicen cantidades grandes para encontrar un número relativamente pequeño de mutaciones positivas. Generalmente, las mutaciones negativas se ignoran y se obtiene solo la información de secuencia de los miembros positivos.

20

25

30

[0004] La mutagénesis de saturación (Estell y col., en *World Biotech Report* 1984, vol. 2: USA, Online Publications, Londres [1984], páginas 181-187 y Wells y col., *Gene* 34:315-323 [1985]) es una técnica que puede usarse para busca el espacio en la proteína para mutaciones que optimicen varias propiedades de una proteína. Varios grupos han desarrollado estrategias para identificar sitios que se pueden cambiar mediante mutagénesis de saturación (Reetz y col., *Angew. Chem. Int. Edn.*, 44:4192-4196 [2005]; Kato y col., *J. Mol. Biol.*, 351:683-692 [2005] y Sandberg y col., *Proc. Natl. Acad. Sci.*, 90:8367-8371 [1993]), aunque no se ha propuesto un sistema general para la identificación de sitios.

35

40

[0005] Además, debido a que la mayoría de los procedimientos de ingeniería de proteínas produce un número mayor de opciones de mutación de aminoácidos, generalmente se requiere una selección de un gran número de variantes para producir una propiedad proteica deseada. Generalmente, la selección se repite una y otra vez para obtener una variante beneficiosa. Por tanto, la mayoría de los procedimientos son laboriosos y llevan mucho tiempo. Existe una necesidad continua en la técnica de procedimientos de ingeniería de proteínas que sean eficaces y produzcan los resultados deseados.

45

RESUMEN DE LA INVENCION

[0006] Por consiguiente, la presente invención proporciona un procedimiento para variantes de ingeniería de proteínas de una proteína parental que combina mutaciones en dos o más sitios de interés; el procedimiento comprende las etapas de:

50

a) proporcionar una proteína parental y una biblioteca de evaluación de sitio de variantes de proteína de dicha proteína parental, donde la biblioteca de evaluación de sitio comprende variantes de la proteína parental modificado cada uno en uno de los dos o más sitios de interés;

55

b) comprobar dicha biblioteca de variantes de proteína y dicha proteína parental para al menos dos propiedades de interés en las respectivas pruebas de interés;

60

c) determinar un valor del índice de rendimiento para cada propiedad de interés dividiendo el valor obtenido para cada una de las variantes de proteína entre el valor obtenido para dicha proteína parental en la prueba de interés para proporcionar una diferencia de energía libre aparente ($\Delta\Delta G_{app}$) para cada una de las variantes de proteína en comparación con dicha proteína parental en la prueba de interés;

d) determinar un valor de índice de rendimiento previsto para variantes de proteína que combinan mutaciones en dos o más sitios de interés, donde el valor índice de rendimiento previsto para las variantes que combinan dos o más mutaciones se obtiene sumando los valores de $\Delta\Delta G_{app}$ de estas mutaciones;

5 e) identificar a partir del valor de índice de rendimiento previsto obtenido en la etapa (d), las variantes de proteína que combinan dos o más mutaciones y que poseen una primera propiedad mejorada con respecto a la proteína parental y una segunda propiedad que es al menos el 90% de la de la proteína parental, proporcionando de este modo una biblioteca de variantes de proteína enriquecida en miembros que tienen al menos dos de las propiedades
10 de interés.

[0007] Esto significa que la presente invención puede usarse para identificar mutaciones que proporcionan al menos una propiedad muy mejorada y al menos una propiedad adicional que no es significativamente peor que la de la proteína natural (p. ej., mejor del 110% de la proteína natural para una
15 propiedad, aún no peor del 90% de la natural para cualquier otra propiedad). Aún en otras realizaciones preferidas adicionales, las bibliotecas se construyen en función de esta información. En algunas realizaciones, las bibliotecas se construyen usando todas las mutaciones identificadas, mientras que en otras realizaciones, las bibliotecas se construyen usando un subgrupo de las mutaciones identificadas. De hecho, no se pretende que las bibliotecas estén limitadas a ningún número y/o tipo en particular de mutaciones.
20

[0008] En algunas realizaciones, la proteína es una enzima. En algunas realizaciones especialmente preferidas, la enzima se selecciona entre proteasas, transferasas, metaloproteasas, esterasas, amilasas, celulasas, oxidasas, cutinasas y lipasas. En algunas realizaciones alternativas, la proteína se selecciona a partir de anticuerpos y factores de crecimiento. Aún en realizaciones preferidas adicionales, la proteína natural es una forma madura de
25 una enzima seleccionada entre proteasas, transferasas, metaloproteasas, esterasas, amilasas, celulasas, oxidasas, cutinasas y lipasas. En algunas realizaciones preferidas, la propiedad de interés se selecciona entre carga, rendimiento de lavado, rendimiento de limpieza de superficie dura, estabilidad térmica, estabilidad de almacenamiento, estabilidad en detergente, unión al sustrato, inhibición enzimática, nivel de expresión, velocidad de reacción y degradación del sustrato. En algunas realizaciones, la proteína natural y la variante proteica son
30 componentes de al menos una composición detergente. En algunas realizaciones preferidas, el rendimiento del lavado se comprueba en una composición detergente formulada en un detergente en polvo o líquido que tiene un pH de entre 5 y 12,0.

BREVE DESCRIPCIÓN DE LOS DIBUJOS

35 [0009] En la figura 1 se proporcionan las distribuciones de los 2.851 valores de $\Delta\Delta G_{app}$ obtenidos para cada propiedad.

[0010] En la Figura 2A se proporcionan los resultados obtenidos de calcular la distribución esperada de
40 los valores de $\Delta\Delta G_{app}$ para la estabilidad en LAS y la actividad de queratina para mil combinaciones elegidas al azar de mutaciones en los cuatro sitios, en comparación con la distribución real de los valores de $\Delta\Delta G_{app}$ para 64 miembros de la biblioteca elegidos al azar.

[0011] En la figura 2B se muestra la distribución real observada para los 64 miembros de la biblioteca
45 elegidos al azar.

DESCRIPCIÓN DE LA INVENCION

[0012] La presente invención proporciona procedimientos para la ingeniería de proteínas.
50 Específicamente, la invención proporciona procedimientos que utilizan bibliotecas de evaluación de sitios.

[0013] Con fines prácticos, normalmente no es necesario encontrar la mejor secuencia en un espacio proteico para crear una proteína óptima para una aplicación en particular. En la mayoría de las aplicaciones, el problema que hay que resolver es identificar al menos una secuencia proteica que cumpla o exceda el valor mínimo
55 requerido para diversas propiedades. Esto requiere conocimiento de las mutaciones que son buenas para una propiedad en particular, así como conocer aquellas mutaciones que son malas para cualquiera de las propiedades deseadas. La presente invención proporciona medios para cumplir el objetivo identificando las posiciones en la proteína que pueden alterarse para mejorar la propiedad principal y mantener los valores de otras propiedades dentro de los límites deseados.
60

[0014] La presente invención proporciona medios para evaluar todas las posiciones en una proteína para todas las propiedades de interés construyendo "bibliotecas de evaluación de sitio" en cada sitio. En realizaciones preferidas, estas bibliotecas contienen de 9 a 19 mutaciones en cada posición y se usan para evaluar el posible uso de cada posición en la manipulación genética de la proteína y en la construcción de bibliotecas. Cada propiedad se

mide en relación con la proteína parental y se calcula una diferencia de energía libre aparente para cada mutante frente a la proteína natural. Estos valores aparentes delta delta G ("es decir, $\Delta\Delta G$ ") se usan entonces para determinar la aditividad.

- 5 **[0015]** Una forma ideal de analizar las variantes sería a través de la diferencia en la energía libre de la variante frente a la proteína parental en el proceso de interés. La energía libre de Gibbs de un proceso representa la cantidad máxima de trabajo que puede rendir un sistema. El cambio en la energía libre en relación con la enzima parental ($\Delta\Delta G$) se proporciona de la siguiente forma:

$$10 \quad \Delta\Delta G = -RT \ln (K_{\text{variante}}/K_{\text{parental}})$$

donde K_{variante} es la constante de velocidad para la enzima variante y K_{parental} es la constante de velocidad para la enzima parental, R es la constante de la ley de gas y T es la temperatura absoluta. La mayoría de los ensayos no están montados para permitir la determinación de las energías libres reales por lo que los autores utilizaron una cantidad:

$$15 \quad \Delta\Delta G_{\text{app}} = -RT \ln (P_{\text{variante}}/P_{\text{parental}})$$

donde P_{variante} es el valor de rendimiento para la variante y P_{parental} es el valor de rendimiento para la enzima parental en las mismas condiciones. Puede preverse que los valores de $\Delta\Delta G_{\text{app}}$ se comporten de forma similar a $\Delta\Delta G$ para las distribuciones de datos y la aditividad. Sin embargo, puesto que $\Delta\Delta G$ es la cantidad máxima de trabajo que la variante puede realizar en comparación con la enzima parental, la cantidad $\Delta\Delta G_{\text{app}}$ subestimarán generalmente el $\Delta\Delta G$ y llevará a resultados que parezcan sinérgicos porque las propiedades de dos posiciones aditivas pueden ser superiores al valor previsto sumando sus valores de $\Delta\Delta G_{\text{app}}$.

- 25 **[0016]** Los procedimientos de la presente invención se utilizaron para diseñar bibliotecas eficaces que se usaron para modificar por ingeniería genética múltiples propiedades en paralelo. Aunque en este documento se describe "ASP", una serina proteasa de 189 aminoácidos, los procedimientos se aplican a cualquier proteína de interés para ingeniería de proteínas. La proteasa ASP pertenece a la familia S1E (véase, p. ej., Rawlings y col., Nucleic Acids Res., 34:D270-D272 [2006]) de serina proteasas y es un homólogo de las estreptogrisinas. La enzima serina proteasa madura derivada de la cepa 69B4 de *Cellulomonas* (DSM 983316035) tiene una longitud de 189 aminoácidos (SEC ID N° 2) con una triada catalítica por His32, Asp56 y Ser137, como se muestra a continuación (con la triada catalítica indicada en negrita y subrayada):

35 **FDVIGGNAYT IGGRSRCISG FAVNGGFITA GHCGRTGATT ANPTGTFAGS
SFPGNDYAFV RTGAGVLLA QVNNYSGGRV QVAGHTAAPV GSAVCRSGST
TGWHCGTITA LNSSVTYPEG TVRGLIRTTV CAEPGDSGGG LLAGNQAQGV
TSGGSGNCRT GGTTFFQPVN PILQAYGLRM ITTDSGSSP (SEQ ID NO:2)**

- 40 **[0017]** Las bibliotecas de evaluación de sitios (BES) se construyeron como se describe en este documento introduciendo de 12 a 19 sustituciones en cada una de las 189 posiciones. Se analizaron las 2.851 mutaciones en las 189 posiciones usando tres ensayos de actividad diferentes y dos ensayos de estabilidad diferentes. Se obtuvieron de media 15 mutaciones por posición.

Evaluación de los datos de las variantes de la BES

- 45 **[0018]** En la Tabla I se proporcionan los datos para una posición en la proteína, en concreto la posición 14.

Tabla I. Datos de rendimiento para la posición 14						
Posición	Código de la variante	$\Delta\Delta G$ de la actividad caseína	$\Delta\Delta G$ de la actividad queratina	$\Delta\Delta G$ de la actividad AAPF	$\Delta\Delta G$ de estabilidad en LAS	$\Delta\Delta G$ de estabilidad térmica
14	R014T	0,35	-0,11	0,05	-0,86	-0,05
14	R014S	0,28	-0,07	0,13	-0,63	-0,05
14	R014I	0,15	-0,19	0,21	-0,53	-0,05
14	R014Q	0,10	-0,23	0,11	-0,52	-0,05
14	R014N	0,16	-0,15	0,75	-0,47	-0,05
14	R014H	0,21	-0,01	0,00	-0,23	-0,05
14	R014K	0,16	-0,10	-0,03	0,16	-0,05
14	R014R	0	0	0	0	0

14	R014W	0,30	-0,04	-0,27	0,09	0,02
14	R014G1	0,14	-0,15	0,14	-0,64	0,03
14	R014L	0,02	-0,24	-0,08	-0,22	0,09
14	R014E	-0,15	-0,15	-0,04	-1,21	0,12
14	R014D	0,17	-0,18	0,02	-1,07	0,15
14	R014P	0,33	-0,02	0,04	-0,09	0,69
14	R014M	0,06	-0,11	0,03	-0,33	0,73
14	R014A	0,14	-0,07	0,07	-0,52	1,09
14	R014C	0,18	0,13	0,74	-0,48	1,54

[0019] Se enumera el aminoácido natural como punto de referencia para cada posición. En la posición 14, R014R representa el aminoácido natural y R014X representa cada mutación medida. Para cada propiedad, se usaron 16 medidas para determinar la media y la desviación estándar de $\Delta\Delta G_{app}$ para la enzima parental. La media parental ($\mu_{parental}$) se normalizó a 0 y se determinó la desviación estándar ($\sigma_{parental}$) para $\Delta\Delta G_{app}$. Estos valores se usaron como referencia para cada propiedad en cada posición de la molécula y se enumeran en la línea R014R de la Tabla I.

[0020] En la Tabla II se proporciona un resumen de los resultados de los 2.851 mutantes. Las mutaciones se dividen en dos clases: "Arriba" y "Abajo". Un mutante está "Arriba" si el $\Delta\Delta G_{app}$ era negativo o 0 y está "Abajo" si el $\Delta\Delta G_{app}$ era positivo. La probabilidad de que una mutación esté Arriba o Abajo se determinó contando el número de mutaciones que estaban arriba o abajo y dividiendo este número entre el número total de mutaciones (es decir, 2.851 en el caso de ASP). Se encontró que la probabilidad de que una mutación estuviera Abajo (es decir "pAbajo") para una propiedad en particular oscilaba entre el 84 y el 94%. Se encontró que la probabilidad de que una mutación estuviera Abajo (es decir "pAbajo") para una propiedad en particular oscilaba entre el 6 y el 16%. Estos datos indican que las mutaciones acumuladas que son buenas para una propiedad requieren que todas las demás propiedades empeoren.

Tabla II. Resumen de los resultados para los mutantes					
	$\Delta\Delta G$ de la actividad caseína	$\Delta\Delta G$ de la actividad queratina	$\Delta\Delta G$ de la actividad AAPF	$\Delta\Delta G$ de estabilidad en LAS	$\Delta\Delta G$ de estabilidad térmica
Arriba	465	422	179	425	419
Abajo	2.386	2.429	2.672	2.426	2.432
Total	2.851	2.851	2.851	2.851	2.851
p Arriba	16%	15%	6%	15%	15%
p Abajo	84%	85%	94%	85%	85%

[0021] Las distribuciones de los 2.851 valores de $\Delta\Delta G_{app}$ obtenidos para cada propiedad se muestran en la Figura 1. En algunas realizaciones, la distribución de todas las propiedades se presentan como la suma de dos o más distribuciones Gaussianas. Esto coincide con las distribuciones de energías libres para bibliotecas publicadas en la literatura (Lancet y col., Proc. Natl. Acad. Sci. USA 90:8367-8371 [1993] y Lu y col., Proc. Natl. Acad. Sci. USA 98:1410-1415 [2001]). Por tanto, los valores medios de $\Delta\Delta G_{app}$ para cada propiedad eran todos sustancialmente peores que para la enzima parental. Para cada mutante que tenían el 1% o menos de la actividad parental ($\Delta\Delta G_{app} > 2,7$), los valores se fijaban de forma arbitraria al 1%, debido al error inherente en los sistemas de ensayo. Para cada propiedad, se obtuvo un número sustancial de mutaciones que tenían el 1% o menos de la actividad parental. Para estos datos se calcularon las medias y las desviaciones estándar, así como para los subgrupos de mutantes que mostraron más del 5% de la actividad de la enzima parental (véase la Tabla III).

[0022] Los valores medios de $\Delta\Delta G_{app}$ para cada propiedad para los 2.851 variantes variaban de 0,9 a 1,5 kcal/mol, que se correspondían con del 20 al 7% de la actividad de la enzima parental.

[0023] Es importante apreciar que estas distribuciones también representan las distribuciones de los valores de $\Delta\Delta G_{app}$ que podría esperarse en las bibliotecas aleatorias que tenían por media una mutación por miembro.

Tabla III. Media y desviación estándar para todos los mutantes					
Media y desviación estándar para mutantes con $\leq 1\%$ de la actividad de la enzima parental					
	$\Delta\Delta G$ DE CAS	$\Delta\Delta G$ DE QUER	$\Delta\Delta G$ DE AAPF	$\Delta\Delta G$ LAS	$\Delta\Delta G$ TER
Media	1,01	0,93	1,45	1,52	1,36
Desv. Estan.	1,10	1,00	1,11	1,25	1,21
Media y desviación estándar para mutantes con $> 5\%$ de la actividad de la enzima parental					

	$\Delta\Delta G$ DE CAS	$\Delta\Delta G$ DE QUER	$\Delta\Delta G$ DE AAPF	$\Delta\Delta G$ LAS	$\Delta\Delta G$ TER
Media	0,31	0,46	0,52	0,33	0,37
Desv. Estan.	0,47	0,56	0,49	0,63	0,52

[0024] Los datos de la evaluación de sitios se comprobaron para buscar indicios de correlación entre propiedades. Los valores de $\Delta\Delta G_{app}$ para cada propiedad se representaron frente a cada una de las otras propiedades y se calcularon los coeficientes de correlación como se muestra en la Tabla IV. Las dos medidas de actividad sobre los sustratos proteicos estaban correlacionados ($r^2 = 0,77$) con sólo una débil correlación ($r^2 = 0,53$) de cada sustrato proteico con la actividad sobre el sustrato peptídico sintético AAPF. Ninguna de las dos medidas de estabilidad se correlacionaba con las medidas de actividad o entre sí.

	CAS	QUER	AAPF	LAS	TER
CAS	1				
QUER	0,77	1			
AAPF	0,53	0,53	1		
LAS	<0,01	<0,01	0,13	1	
TER	0,01	0,01	0,06	0,24	1

Evaluación de los datos de posición de BES

10

[0025] Para analizar las posiciones dentro de la secuencia de aminoácidos, se definieron dos tipos de sitios. Los sitios "Improductivos" no tienen mutantes mejores que la enzima parental mientras que los sitios "Productivos" tienen al menos una sustitución mejor que la enzima parental. En la Tabla V se proporcionan los datos de los sitios Productivos e Improductivos para cada propiedad dentro de las 189 posiciones de ASP. La probabilidad de que un sitio sea Productivo viene dada por el número de sitios Productivos dividido entre el número total de sitios (189). Aunque la probabilidad de que cualquier mutación sea mejor que la enzima parental es baja (es decir, del 6 a 28%), la probabilidad de que dicho sitio tenga al menos una mutación Arriba es bastante alta.

15

	$\Delta\Delta G$ de act caseína	$\Delta\Delta G$ de act. queratina	$\Delta\Delta$ act. AAPF	$\Delta\Delta$ estabilidad LAS	$\Delta\Delta$ estabilidad térmica
Improductivo	82	92	144	86	59
Productivo	107	97	45	103	130
p Improductivo	43%	49%	76%	46%	31%
p Productivo	57%	51%	24%	54%	69%

[0026] Fue interesante determinar cómo se distribuían los sitios Productivos e Improductivos con respecto a las características estructurales (p. ej., aminoácidos enterrados, aminoácidos que interaccionan, posiciones próximas al centro activo, etc.) en la ASP, así como los sitios de secuencia que están conservados o han cambiado durante la evolución. Para hacer esta determinación, se examinó la estructura de ASP y la secuencia se alineó con 20 homólogos no redundantes (Edgar, Nucl. Acids Res., 32:1792-1797 [2004]). Los resultados se muestran en la Tabla VI.

25

	CAS	QUER	BMI	LpH	AAPF	LAS	TER
	proporción	proporción	proporción	proporción	proporción	proporción	proporción
Próximo al centro activo							
Improductivo	1,50	1,61	1,61	1,53	1,03	1,34	0,85
Productivo	0,61	0,42	0,48	0,56	0,91	0,72	1,09
Enterrado							
Improductivo	1,65	1,66	1,71	1,62	1,31	1,67	1,39
Productivo	0,50	0,37	0,40	0,48	0,00	0,44	0,78
Sin contactos							
Improductivo	0,70	0,63	0,72	0,79	0,84	0,37	0,69
Productivo	1,18	1,30	1,18	1,12	1,40	1,48	1,13
Más de 3 contactos							
Improductivo	1,67	1,49	1,57	1,52	1,27	1,36	1,44
Productivo	0,49	0,54	0,51	0,57	0,14	0,70	0,75
Conservado							
Improductiva	2,23	2,05	2,10	2,13	1,27	1,91	1,34

Productiva	0,06	0,00	0,06	0,06	0,14	0,24	0,81
Variable							
Improductivo	0,31	0,55	0,43	0,44	0,79	0,88	0,56
Productivo	1,53	1,43	1,48	1,47	1,68	1,10	1,25
Inserción o deleción							
Improductivo	0,34	0,53	0,40	0,81	0,88	0,81	0,82
Productivo	1,50	1,52	1,51	1,22	1,56	1,22	1,04

[0027] Era destacable que para las propiedades investigadas, los sitios Productivos no se encontraban en el núcleo hidrófobo de ASP. También era interesante apreciar que ninguno de los sitios más productivos para la actividad caseína está próximo a la triada catalítica. Sólo uno de los sitios Productivos de la caseína (P118) está en contacto con el sustrato. El resto de los sitios Productivos de la caseína se distribuyen sobre lazos superficiales flexibles a lo largo de la proteína. Ninguno de los sitios Productivos para la actividad queratina se encuentra próximo al centro activo. Se encontró que estos sitios están esparcidos sobre la superficie de la molécula completa. El sitio Productivo de la queratina más próximo es R014, que siguen estando al menos a 13Å de la serina catalítica (S137, distancia Ca-Ca).

[0028] La posición de los sitios Productivos de estabilidad en LAS sigue el esquema general de estar diseminada sobre los lazos superficiales flexibles de la proteína completa. Esto también se aplica a la localización de los sitios Productivos de estabilidad térmica, con una excepción: C033 presenta interacciones de Vanderwaals con H032, que es su vecino consecutivo en la secuencia de aminoácidos.

[0029] Basándose en el alineamiento de secuencia, se identificaron sitios como "conservados" (sin diferencias en las 20 secuencias), "variables" (6 o más aminoácidos diferentes en las 20 secuencias) o "sitios de inserción o deleción" referentes a ASP. Los valores previstos se calcularon a partir de las probabilidades de que un sitio pudiera cumplir las condiciones dadas y ser Productivo o Improductivo para una propiedad determinada. Se calculó la relación entre los valores observados y los calculados; los valores por encima de 1,4 y por debajo de 0,6 se consideraron indicadores de sobrerrepresentación o subrepresentación de una clase de sitios en particular. Los valores de corte se eligieron en base a los resultados de diez grupos de datos generados aleatoriamente que coincidían con los valores de cada clase de sitios. Se ha encontrado que los restos enterrados y aquellos con varios contactos están fuertemente correlacionados con sitios Improductivos para la actividad proteasa en los dos sustratos proteicos, así como en la estabilidad a LAS. Sorprendentemente, se encontró que era más probable que las posiciones próximas al centro activo fueran Improductivas que Productivas. En el alineamiento de secuencias, era especialmente probable que los sitios conservados fuesen Improductivos para la actividad sobre los sustratos de la proteína y para la estabilidad LAS, mientras que era muy probable que los sitios altamente variables y los sitios de inserción o deleción fueran Productivos en cuanto a la actividad, con poco efecto sobre la estabilidad.

[0030] Como se indica en el Ejemplo 5, las mutaciones deletéreas para cualquier propiedad se correlacionan con mutaciones deletéreas para cualquier otra propiedad, independientemente de las correlaciones entre las propiedades. Sólo un pequeño número de posiciones (5 al 10%) tienen mutaciones que son malas para todas las propiedades. Estas posiciones definen el "plegamiento" y están conservadas en la evolución. La implicación de esto es que aunque la identificación de mutaciones beneficiosas para cualquier propiedad requiere una selección realmente predictiva para esta propiedad, puede conseguirse la identificación de mutaciones probablemente deletéreas para cualquier propiedad usando CUALQUIER selección. Una estrategia de ingeniería de proteínas simplificada es construir BES y realizar una selección usando en análisis de una única actividad y/o estabilidad. Se identifican las mutaciones deletéreas y aquellas posiciones que tienen pocas mutaciones deletéreas se usan para construir bibliotecas y mutaciones combinatorias para mejorar las propiedades múltiples. También, los sitios de selección que se encuentran en la superficie de la proteína, tienen menos interacciones y son variables en los alineamientos de secuencia proporcionando una alta proporción de sitios productivos. Los sitios que están en el interior de la molécula tienen muchos contactos y están muy conservados en la evolución tendrán una alta probabilidad de presentar mutaciones deletéreas por lo que deberían evitarse. Se contempla que cualquier procedimiento adecuado para analizar la información de secuencia y/o estructural encontrará aplicación en la presente invención, incluyendo pero sin limitaciones, procedimientos y/o programas de ordenador y/o electrónicos.

[0031] En las tablas del Ejemplo 5 se proporcionan comparaciones por parejas de los valores de variantes con más del 5% de la actividad natural y menos del 5% de actividad para cada una de las dos propiedades, junto con los coeficientes de correlación de las dos propiedades. Se muestran los resultados de tres enzimas, en concreto ASP, ACT y NPre, aunque no se pretende que la presente invención se vea limitada a estas enzimas en particular, ya que los procedimientos proporcionados en este documento encuentran su uso con cualquier proteína.

[0032] Las enzimas (ASP, ACT y NPre) y los sistemas de ensayo se describen en detalle en los documentos US WO 2005/052146, WO 2005/056782 y US 2008/0293610. Además, los procedimientos proporcionados en la solicitud de patente provisional de EE. UU. N° de serie 60/933.312 (WO 2008/002472)

encuentran su uso conjuntamente con la presente invención. Las propiedades utilizadas en este documento fueron actividad caseína (CAS), actividad queratina (QUER), actividad AAPF (AAPF), estabilidad en LAS (LAS) y estabilidad térmica para ASP, así como formación de perácido (PAF) y degradación de perácido (PAD) para ACT. En estos experimentos, las únicas propiedades que se encontró estaban correlacionadas (coeficientes de correlación > 0,5) fueron CAS, QUER y AAPF para ASP. Todas las demás no estaban correlacionadas (coeficiente de correlación <0,3). A pesar del hecho de que las propiedades no se correlacionaban, la probabilidad de que una mutación pudiera ser deletérea para las dos propiedades es mucho más alta que la esperada al azar. En la tabla se proporcionan cocientes calculados de los valores observados de las variantes previstas al azar. Los valores mayores de 1 indican correlaciones positivas y aquellos menores de 1 indican correlaciones negativas.

10

Diseño de la biblioteca

[0033] En algunas realizaciones especialmente preferidas, los datos de la biblioteca de evaluaciones de sitios se usan para diseñar una biblioteca combinatoria. La evolución dirigida tradicional permite construir bibliotecas aleatorias y analizar grandes números de bibliotecas en búsqueda de propiedades sencillas, combinar estas y repetir el proceso. Como han encontrado varios investigadores (véase Bloom y col., Curr. Opin. Struct. Biol., 15:447-452 [2005]; Bloom y col., Proc. Natl. Acad. Sci. USA 103:5869-5874 [2006] y Guo y col., Proc. Natl. Acad. Sci. USA 101:9205-9210 [2004]), la acumulación de mutaciones positivas para una propiedad normalmente lleva a la disminución de otras propiedades. Esto también se muestra fácilmente en la Tabla II, ya que la probabilidad de que cualquier mutación sea Arriba para cualquier propiedad es pequeña y la probabilidad de que cualquier mutación sea Abajo es alta (> 85%) mientras que la probabilidad de que se acumulen más de tres (3) mutaciones que aumenten la actividad y den lugar a una disminución de algunas otras propiedades es bastante alto.

[0034] Sin embargo, este problema se evita usando los datos de evaluación del sitio para construir bibliotecas que podría ser buenos para propiedades múltiples. No se incluyeron los sitios Improductivos en las bibliotecas combinatorias y los sitios Productivos se clasificaron adicionalmente por el porcentaje de mutaciones que eran Arriba. Un grupo de cuatro sitios que no interactuaban (14-24-127-159) con porcentajes altos de mutaciones Arriba tanto para la estabilidad LAS como para la actividad queratina se usaron para diseñar una biblioteca que mejorase ambas propiedades a la vez (Véase la Tabla VII).

Posición	Aminoácido natural	$\Delta\Delta$ de queratina (mejor %)	$\Delta\Delta$ de LAS (mejor %)
14	R	94	88
24	N	67	40
127	R	63	69
159	R	83	78

[0035] Asumiendo la aditividad para los sitios, se calcularon los valores previstos de $\Delta\Delta G_{app}$ para la biblioteca y se compararon con los valores determinados para la biblioteca real. En algunas realizaciones, para sitios que son aditivos para la propiedad, los resultados típicamente coinciden. Pero, en otras realizaciones, en las que los resultados no coinciden con la predicción, la forma en que disienten proporciona información sobre las interacciones de los sitios, la no aditividad de las propiedades y/o la conveniencia de los ensayos utilizados.

[0036] La distribución esperada de los valores de $\Delta\Delta G_{app}$ para la estabilidad LAS y la actividad de queratina para mil combinaciones elegidas al azar de mutaciones en los cuatro sitios se calculó y comparó con la distribución real de los valores de $\Delta\Delta G_{app}$ para 64 miembros de la biblioteca elegidos al azar. Los resultados se muestran en la Figura 2A. En la Figura 2B se muestra la distribución real observada para los 64 miembros de la biblioteca elegidos al azar. Esta biblioteca claramente tiene un gran número de miembros que son mejores que la enzima parental tanto para la estabilidad LAS como para la actividad queratina. La media observada para la actividad queratina de 0,02 Kcal coincide bien con la media prevista de -0,01 Kcal, en coincidencia con la aditividad de estos sitios. Para los resultados de estabilidad LAS, la media observada de -1,13 excedía significativamente el valor previsto de -0,28, aunque las desviaciones estándar eran similares (Véase la Tabla VIII).

Propiedad	Previsto	Observado
Actividad queratina		
Media	-0,01	0,02
Desv. estándar	0,53	0,39
Estabilidad LAS		
Media	-0,28	-1,13
Desv. estándar	1,29	1,04

[0037] En el caso de la estabilidad LAS, el ensayo original para los mutantes de BES subestimaba los valores reales de $\Delta\Delta G$. El ensayo se cambió, elevando la temperatura de incubación de 25° C a 35° C, ya que la mayoría de los miembros de la biblioteca eran estables en las condiciones del ensayo y la biblioteca se ensayó en

50

condiciones más duras. Los valores de $\Delta\Delta G_{app}$ se corrigieron para tener en cuenta esto, pero la corrección probablemente sigue subestimando los valores reales de $\Delta\Delta G$ y los sitios probablemente siguen siendo aditivos para la estabilidad LAS, dada la concordancia de la desviación estándar.

5 Definiciones

[0038] A no se que se indique otra cosa, la práctica de la presente invención implica técnicas convencionales utilizados normalmente en biología molecular, ingeniería de proteínas, microbiología y ADN recombinante, que están dentro de las habilidades de la técnica. Estas técnicas son conocidas por los expertos en la técnica y se describen en numerosos libros y trabajos de referencia (véase, p. ej., Sambrook y col., "Molecular Cloning: A Laboratory Manual", Segunda Edición (Cold Spring Harbor), [1989]) y Ausubel y col., "Current Protocols in Molecular Biology" [1987]).

[0039] Siempre que no se defina de otro modo en este documento, todos los términos técnicos y científicos utilizados en este documento tienen el mismo significado que normalmente entiende un experto en la materia a la que pertenece esta invención. Por ejemplo, en Singleton y Sainsbury, Dictionary of Microbiology and Molecular Biology, 2ª Ed., John Wiley e hijos, NY (1994) y en Hale y Marham, The Harper Collins Dictionary of Biology, Harper Perennial, NY (1991) se proporciona a los expertos en la técnica diccionarios generales de muchos términos usados en la invención. Aunque pueden usarse en la práctica de la presente invención cualesquiera procedimientos y materiales similares o equivalentes a los descritos en este documento, se describen aquí los procedimientos y materiales preferidos. Por consiguiente, los términos definidos inmediatamente a continuación se describen de forma más completa por referencia a la memoria descriptiva en su totalidad. También, como se usa en este documento, las formas singulares "uno/una" y "el/la" incluyen referencia al plural a no ser que el contexto indique claramente otra cosa. Los intervalos numéricos incluyen los números que define el intervalo. Siempre que no se indique otra cosa, los ácidos nucleicos se escribe de izquierda a derecha en la orientación 5' a 3'; las secuencias de aminoácidos se escriben de izquierda a derecha en orientación de amino a carboxilo, respectivamente. Se entenderá que esta invención no se limita a la metodología, protocolos y reactivos en particular descritos sino que pueden variar dependiente del contexto en que los expertos los usan.

[0040] La práctica de la presente invención emplea, a no ser que se indique otra cosa, técnicas convencionales de purificación de proteínas, biología molecular, microbiología, técnicas de ADN recombinante y secuenciación de proteínas, todas ellas dentro de las habilidades de la técnica.

[0041] Además, los encabezamientos proporcionados en este documento no suponen una limitación de los diversos aspectos o realizaciones que pueden hacerse en referencia a la descripción en su totalidad. Por consiguiente, los términos definidos inmediatamente a continuación se definen de forma más completa por referencia a la memoria descriptiva en su totalidad. No obstante, para facilitar la comprensión de la invención, se definen a continuación diversos términos.

[0042] Según se usa en este documento, los términos "proteasa" y "actividad proteolítica" se refieren a una proteína o péptido que muestra la capacidad de hidrolizar péptidos o sustratos que tienen enlaces peptídicos. Existen muchos procedimientos bien conocidos para medir la actividad proteolítica (Kalisz, "Microbial Proteinases," En: Fiechter (ed.), Advances in Biochemical Engineering/Biotechnology, [1988]). Por ejemplo, la actividad proteolítica puede determinarse mediante ensayos comparativos en los que se analiza la capacidad de las respectivas proteasas para hidrolizar un sustrato comercial. Entre los ejemplos de sustratos útiles en el análisis de proteasa o de actividad proteolítica se incluyen, pero sin limitaciones, dimetil caseína (Sigma C-9801), colágeno bovino (Sigma C-9879), elastina bovina (Sigma E-1625) y queratina bovina (ICN Biomedical 902111). Los ensayos colorimétricos que utilizan estos sustratos son bien conocidos en la técnica (véase, p. ej., el documento VO 99/34011 y la patente de EE. UU. N° 6.376.450). El ensayo pNA (véase p. ej., Del Mar y col., Anal. Biochem., 99:316-320 [1979]) también puede usarse para determinar la concentración de la enzima activa en las fracciones recogidas durante la elución del gradiente. Este ensayo mide la velocidad a la que se libera p-nitroanilina cuando la enzima hidroliza el sustrato sintético soluble, succinil-alanina-prolina-fenilalanina-p-nitroanilida (sAAAPF-pNA). La velocidad de producción del color amarillo procedente de la reacción de hidrólisis se mide a 410 nm en un espectrofotómetro y es proporcional a la concentración de enzima activa. Además, las medidas de absorbancia a 280 nm pueden usarse para determina la concentración de proteína total. La relación enzima activa/proteína total proporciona la pureza de la enzima.

[0043] Según se usa en este documento, los términos "proteasa ASP", "proteasa Asp" y "Asp" se refieren a las serina proteasas descritas en este documento. En algunas realizaciones preferidas, la proteasa Asp es la proteasa designada en este documento como proteasa 69B4 obtenida a partir de la cepa 69B4 de *Cellulomonas*. Por tanto, en realizaciones preferidas, el término "proteasa 69B4" se refiere a una proteasa madura presente en la naturaleza derivada de la cepa 69B4 de *Cellulomonas* (DSM 16035) que tiene secuencias de aminoácidos sustancialmente idénticas a la proporcionada en la SEC ID N° 2. En realizaciones alternativas, la presente invención proporciona porciones de la proteasa ASP.

[0044] El término “homólogos a la proteasa de *Cellulomonas*” se refiere a proteasas de origen natural que tienen secuencias de aminoácidos sustancialmente idénticas a la proteasa natural derivada de la cepa 69B4 de *Cellulomonas* o secuencias de polinucleótidos que codifican para esta proteasa de origen natural, y estas proteasas retienen las características funcionales de una serina proteasa codificada por estos ácidos nucleicos. En algunas realizaciones, estos homólogos de proteasa se denominan como “celulomonadinas”.

[0045] Según se usa en este documento, los términos “variante de proteasa”, “variante de ASP”, “variante de la proteasa ASP” y “variante de la proteasa 69B” se usan en referencia a proteasas que son similares a la ASP natural, especialmente en su función, pero tienen mutaciones en su secuencia de aminoácidos que las hacen diferentes en secuencia de la proteasa natural.

[0046] Según se usa en este documento, “*Cellulomonas* sp.” se refiere a las especies del género “*Cellulomonas*” que son bacterias Grampositivas clasificadas como miembros de la familia *Cellulomonadaceae*, Suborden *Micrococccineae*, Orden *Actinomycetales* y Clase *Actinobacteria*. Se reconoce que el género *Cellulomonas* continúa sometido a reorganización taxonómica. Por tanto, se prevé que el género incluya especie que han sido reclasificadas.

[0047] Según se usa en este documento, “el género *Bacillus*” incluye todas las especies del género “*Bacillus*” conocidas por los expertos en la técnica, incluyendo pero sin limitaciones *B. subtilis*, *B. licheniformis*, *B. lentus*, *B. brevis*, *B. stearothermophilus*, *B. alkalophilus*, *B. amyloliquefaciens*, *B. clausii*, *B. halodurans*, *B. megaterium*, *B. coagulans*, *B. circulans*, *B. lautus* y *B. thuringiensis*. Se reconoce que el género *Bacillus* sigue estando sometido a reorganización taxonómica. Por tanto, se prevé que el género incluya especies que están siendo reclasificadas incluyendo, pero sin limitaciones, organismos como *B. stearothermophilus*, que ahora se denomina “*Geobacillus stearothermophilus*”. La producción de endosporas resistentes en presencia de oxígeno se considera la característica que define al género *Bacillus*, aunque esta característica también se aplica a los recientemente denominados *Alicyclobacillus*, *Amphibacillus*, *Aneurinibacillus*, *Anoxybacillus*, *Brevibacillus*, *Filobacillus*, *Gracilibacillus*, *Halohacillus*, *Paenibacillus*, *Salibacillus*, *Thermobacillus*, *Ureibacillus* y *Virgibacillus*.

[0048] Los términos “polinucleótido” y “ácido nucleico”, usados indistintamente en este documento, se refieren a una forma polimérica de nucleótidos de cualquier longitud, de ribonucleótidos o desoxirribonucleótidos. Estos términos incluyen, pero sin limitaciones, un ADN de cadena sencilla, doble o triple, ADN genómico, ADNc, ADN, híbrido ADN-ARN y un polímero que comprende bases purínicas y pirimidínicas u otras bases nucleotídicas naturales, química o bioquímicamente modificadas, no naturales o derivatizadas. A continuación se recogen ejemplos no limitantes de polinucleótidos: genes, fragmentos de genes, fragmentos cromosómicos, EST, exones, intrones, ARNm, ARNt, ARNr, riboximas, ADNc, polinucleótidos recombinantes, polinucleótidos ramificados, plásmidos, vectores, ADN aislado de cualquier secuencia, ARN aislado de cualquier secuencia, sondas de ácidos nucleicos y cebadores. En algunas realizaciones, los polinucleótidos comprenden nucleótidos modificados, como nucleótidos metilados y análogos de nucleótidos, uracilo, otros azúcares y grupos de enlace como fluororribosa y tioato, y ramas de nucleótidos. En realizaciones alternativas, la secuencia de nucleótidos está interrumpida por componentes no nucleotídicos.

[0049] Según se usa en este documento, los términos “construcción de ADN” y “ADN transformante” se usan indistintamente para referirse al ADN usado para introducir secuencias en una célula u organismo huésped. El ADN puede generarse *in vitro* mediante PCR o cualquier otra técnica adecuada conocida por los expertos. En realizaciones especialmente preferidas, la construcción de ADN comprende una secuencia de interés (p. ej. como una secuencia entrante). En algunas realizaciones, la secuencia se une de forma operativa a elementos adicionales como elementos de control (p. ej., promotores, etc.). La construcción de ADN puede comprender además un marcador seleccionable. Puede además comprender una secuencia entrante flanqueada por cajas de homología. En una realización adicional, el ADN transformante comprende otras secuencias no homólogas, añadidas en los extremos (p. ej., secuencias de relleno y flanqueantes). En algunas realizaciones, los extremos de la secuencia entrante están cerrados de modo que el ADN transformante forma un círculo cerrado. Las secuencias transformantes pueden ser naturales, mutantes o modificadas. En algunas realizaciones, la construcción de ADN comprende secuencias homólogas al cromosoma de la célula huésped. En otras realizaciones, la construcción de ADN comprende secuencias no homólogas. Una vez que la construcción de ADN se ensambla *in vitro* puede usarse para: 1) insertar secuencias heterólogas en una secuencia diana deseada de una célula huésped y/o 2) realizar una mutagénesis de una región del cromosoma de la célula huésped (es decir, sustituir una secuencia endógena por una secuencia heteróloga), 3) delección de genes diana y/o 4) introducir un plásmido replicante en un huésped.

[0050] Según se usa en este documento, los términos “casete de expresión” y “vector de expresión” se refieren a construcciones de ácido nucleico generadas de forma recombinante o sintética, con una serie de elementos de ácido nucleico específicos que permiten la transcripción de un ácido nucleico en una célula diana. El casete de expresión recombinante puede incorporarse en un plásmido, cromosoma, ADN mitocondrial, ADN plásmido, virus o fragmento de ácido nucleico. Típicamente, la porción de casete de expresión recombinante de un vector de expresión incluye, entre otras secuencias, una secuencia de ácido nucleico que se va a transcribir y un

promotor. En realizaciones preferidas, los vectores de expresión tienen la capacidad de incorporarse y expresar fragmentos de ADN heterólogos en una célula huésped. Muchos vectores de expresión procariotas y eucariotas están disponibles en el mercado. La selección de vectores de expresión apropiados está dentro del conocimiento de los expertos en la técnica. El término "casete de expresión" se usa en este documento indistintamente con "construcción de ADN" y sus equivalentes gramaticales. La selección de vectores apropiados de expresión están dentro del conocimiento de los expertos en la técnica.

[0051] Según se usa en este documento, el término "vector" se refiere a una construcción polinucleotídica diseñada para introducir ácidos nucleicos en uno o más tipos de células. Entre los vectores se incluyen vectores de clonación, vectores de expresión, vectores lanzadera, plásmidos, casetes y similares. En algunas realizaciones, la construcción polinucleotídica comprende una secuencia de ADN que codifica la proteasa (p. ej., proteasa precursora o madura) que está unida de forma operativa a una prosequencia adecuada (p. ej. secretora, etc.) capaz de efectuar la expresión del ADN en un huésped adecuado.

[0052] Según se usa en este documento, el término "plásmido" se refiere a una construcción de ADN de doble cadena (dc) usada como vector de clonación y que forma un elemento genético autoreplicante extracromosómico en algunos eucariotas o procariotas, o integrado dentro del cromosoma huésped.

[0053] Según se usa en este documento en el contexto de la introducción de una secuencia de ácido nucleico en una célula, el término "introducido" se refiere a cualquier procedimiento adecuado para transferir la secuencia de ácido nucleico dentro de la célula. Estos procedimientos para la introducción incluyen, pero sin limitaciones, fusión del protoplasto, transfección, transformación, conjugación y transducción (véase p. ej., Ferrari y col., "Genetics" en Hardwood y col., (eds), Bacillus, Plenum Publishing Corp., páginas 57-72 [1989]).

[0054] Según se usa en este documento, los términos "transformado" y "transformado de forma estable" se refieren a una célula que tiene una secuencia de polinucleótidos no nativa (heteróloga) integrada en su genoma o como plásmido episomal que se mantiene durante al menos dos generaciones.

[0055] Un ácido nucleico está unido "de forma operativa" cuando está colocado en una relación funcional con otra secuencia de ácido nucleico. Por ejemplo, el ADN que codifica una secuencia líder secretora (es decir, un péptido señal) se une de forma operativa al ADN de un polipéptido si este se expresa como una preproteína que participa en la secreción del polipéptido; un promotor o un potenciador se une de forma operativa a una secuencia codificadora si este afecta a la transcripción de la secuencia o un sitio de unión al ribosoma se une de forma operativa a una secuencia codificadora si se coloca de modo que facilita la traducción. Generalmente, "unido de forma operativa" significa que las secuencias de ADN que se están uniendo son contiguas y, en caso de un líder secretor, contiguas y en fase de lectura. Sin embargo, no es necesario que los potenciadores estén contiguos. La unión se logra mediante ligamiento a sitios de restricción convenientes. Si estos sitios no existen se usan adaptadores o enlazadores oligonucleotídicos sintéticos de acuerdo con la práctica convencional.

[0056] Según se usa en este documento el término "gen" se refiere a un polinucleótido (p. ej., un fragmento de ADN), que codifica un polipéptido e incluye regiones que preceden y siguen a las regiones codificantes así como secuencias intermedias (intrones) entre segmentos codificantes individuales (exones).

[0057] Según se usa en este documento, "genes homólogos" se refiere a un par de genes de especies diferentes, pero normalmente relacionadas, que se corresponden entre sí y son idénticos o muy similares entre sí. El término abarca genes que están separados por especiación (es decir, el desarrollo de nuevas especies) (es decir, genes ortólogos) así como genes que se han separado por duplicación genética (es decir, genes parálogos).

[0058] Según se usa en este documento "ortólogo" y "genes ortólogos" se refieren a genes de especies diferentes que han evolucionado a partir de un gen ancestral común (es decir, un gen homólogo) mediante especiación. Típicamente, los ortólogos retiene la misma función durante el transcurso de la evolución. La identificación de ortólogos encuentra su uso en la predicción fiable de la función del gen en genomas recién secuenciados.

[0059] Según se usa en este documento, "parálogo" y "genes parálogos" se refiere a genes que están relacionados por duplicación dentro de un genoma. Mientras que los ortólogos retienen la misma función durante el transcurso de la evolución, los parálogos desarrollan nuevas funciones, a pesar de que algunas funciones están relacionadas a menudo con la original. Entre los ejemplos de genes parálogos se incluyen, pero sin limitaciones, los genes que codifican tripsina, quimotripsina, elastasa y trombina, que son todas serina proteasas y se encuentran dentro de las mismas especies.

[0060] Según se usan en este documento, las proteínas se definen como con un "plegamiento" común si tienen las mismas estructuras secundarias principales en la misma disposición y con las mismas conexiones topológicas. Proteínas diferentes con el mismo plegamiento, a menudo tienen elementos periféricos de estructura

secundaria y regiones de giro que difieren en tamaño y conformación. En algunos casos, estas regiones periféricas que difieren pueden constituir la mitad de la estructura. Las proteínas colocadas juntas en la misma categoría de plegamiento no necesariamente tienen un origen evolutivo común (p. ej., similitudes estructurales surgidas de las características físicas y químicas de proteínas que favorecen determinadas disposiciones de empaquetamiento y topologías de cadena).

[0061] Según se usa en este documento, "homología" se refiere a similitud o identidad de secuencia, prefiriéndose identidad. Esta homología se determina usando técnicas convencionales conocidas en la técnica (véase p. ej., Smith y Waterman, *Adv. Appl. Math.*, 2:482 [1981]; Needleman y Wunsch, *J. Mol. Biol.*, 48:443 [1970]; Pearson y Lipman, *Proc. Natl. Acad. Sci. USA* 85:2444 [1988]; programas como GAP, BESTFIT, FASTA y TFASTA en el paquete de software genético de Wisconsin (Genetics Computer Group, Madison, WI) y Devereux y col., *Nucl. Acid Res.*, 12:387-395 [1984]).

[0062] Según se usa en este documento, una "secuencia análoga" es aquella en la que la función del gen es esencialmente la misma que la del gen de la proteasa de la cepa 69B4 de *Cellulomonas*. Adicionalmente, genes análogos incluyen al menos una identidad de secuencia del 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 97%, 98%, 99% o 100% con la secuencia de la proteasa de la cepa 69B4 de *Cellulomonas*. Alternativamente, las secuencias análogas tienen un alineamiento de entre el 70 y el 100% de los genes encontrados en la región de la proteasa de la cepa 69B4 de *Cellulomonas* y/o tienen al menos entre 5 y 10 genes que se encuentran en la región alineados con los genes en el cromosoma de la cepa 69B4 de *Cellulomonas*. En realizaciones adicionales se aplica a la secuencia más de una de las propiedades anteriores. Las secuencias análogas se determinan mediante procedimientos conocidos de alineamiento de secuencias. Un procedimiento de alineamiento utilizado frecuentemente es BLAST, aunque como se indica anteriormente y a continuación, existen otros procedimientos que también encuentran aplicación en el alineamiento de secuencias.

[0063] Un ejemplo de un algoritmo útil es PILEUP. PILEUP crea un alineamiento de secuencia múltiple a partir de un grupo de secuencias relacionadas usando alineamientos por parejas progresivos. También puede dibujar un dendrograma que muestra las relaciones agrupadas usadas para crear el alineamiento. PILEUP usa una simplificación del procedimiento de alineamiento progresivo de Feng y Doolittle (Feng y Doolittle, *J. Mol. Evol.*, 35:351-360 [1987]). El procedimiento es similar al descrito por Higgins y Sharp (Higgins y Sharp, *CABIOS* 5:151-153 [1989]). Entre los parámetros de PILEUP útiles se incluyen un peso del hueco por defecto de 3,00, un peso de longitud del hueco por defecto de 0,10 y huecos terminales pesados.

[0064] Otro ejemplo de un algoritmo útil es el algoritmo BLAST, descrito por Altschul y col. (Altschul y col., *J. Mol. Biol.*, 215:403-410, [1990] y Karlin y col., *Proc. Natl. Acad. Sci. USA* 90:5873-5787 [1993]). Un programa BLAST especialmente útil es el programa WU-BLAST-2 (véase Altschul y col., *Meth. Enzymol.*, 266:460-480 [1996]). WU-BLAST-2 utiliza varios parámetros de búsqueda la mayoría de los cuales se establecen como valores predeterminados. Los parámetros ajustables se configuran con los siguientes valores: extensión del solapamiento =1, fracción de solapamiento =0,125, umbral de palabra (T) =11. Los parámetros HSP S y HSP S2 son valores dinámicos y son establecidos por el mismo programa dependiendo de la composición de la secuencia en particular y de la composición de la base de datos en particular frente a la que se está buscando la secuencia de interés. Sin embargo, los valores pueden ajustarse al aumento de la sensibilidad. El valor del % de identidad de secuencia de aminoácidos se determina mediante el número de restos idénticos coincidentes dividido por el número total de restos de la secuencia "más larga" en la región alineada. La secuencia "más larga" es aquella que tiene más restos reales en la región alineada (se ignoran los huecos introducidos por WU-Blast-2 para maximizar la puntuación de alineamiento).

[0065] Por tanto, "porcentaje (%) de identidad de secuencia del ácido nucleico" se define como el porcentaje de restos de nucleótidos en una secuencia candidata que son idénticos a los restos de nucleótidos de la secuencia inicial (es decir, la secuencia de interés). Un procedimiento preferido utiliza el módulo BLASTN de WU-BLAST-2 configurado con los parámetros predeterminados, con una extensión de solapamiento y una fracción de solapamiento fijado a 1 y 0,125, respectivamente.

[0066] Según se usa en este documento, "recombinante" hace referencia a una célula o vector que ha sido modificado mediante la introducción de una secuencia de ácido nucleico heterólogo y que la célula deriva de una célula modificada de este modo. Por tanto, por ejemplo, las células recombinantes expresan genes que no se encuentran de forma idéntica en la forma nativa (no recombinante) de la célula o expresan genes nativos que, por otro lado, se expresan de forma anómala, se subexpresan o no se expresan en todas como resultado de intervenciones humanas deliberadas. "Recombinación", "recombinando" y generando un ácido nucleico "recombinado" son generalmente el ensamblaje de dos o más fragmentos de ácido nucleico en el que el ensamblaje da lugar a un gen quimérico.

[0067] En una realización preferida, las secuencias de ADN mutantes se generan con mutagénesis de saturación de sitio en al menos un codón. En otra realización preferida, la mutagénesis de saturación de sitio se realiza para dos o más codones. En una realización adicional, las secuencias de ADN mutantes tienen más del 50%,

más del 55%, más del 60%, más del 65%, más del 70%, más del 75%, más del 80%, más del 85%, más del 90%, más del 95% o más del 98% de homología con la secuencia natural. En realizaciones alternativas, el ADN mutante se genera *in vivo* usando cualquier procedimiento de mutagénesis conocido como, por ejemplo, radiación, nitrosoguanidina y similares. A continuación, se aísla la secuencia de ADN deseada y se usa en los procedimientos 5 proporcionados en este documento.

[0068] Según se usa en este documento, los términos “amplificación” y “amplificación génica” se refieren a un proceso por el cual secuencias de ADN específicas se replican de forma desproporcionada de modo que el gen amplificado está presente en un número de copias mayor al que estaba inicialmente en el genoma. En algunas 10 realizaciones, la selección de células por crecimiento en presencia de un fármaco (p. ej., un inhibidor de una enzima inhibible) tiene como resultado la amplificación del gen endógeno que codifica el producto génico necesario para el crecimiento en presencia del fármaco o mediante amplificación de secuencias exógenas (es decir, de entrada) que codifican este producto génico, o ambos.

[0069] “Amplificación” es un caso especial de replicación de ácido nucleico que implica la especificidad del molde. Esto se contrastará con la replicación de molde no específica (es decir, replicación dependiente de molde pero no dependiente de un molde específico). La especificidad del molde se distingue aquí a partir de la fidelidad de la replicación (es decir, síntesis de la secuencia polinucleotídica adecuada) y la especificidad del (ribo o desoxirribo) 15 nucleótido. La especificidad del molde se describe frecuentemente en términos de especificidad de “diana”. Las secuencias diana son “dianas” en el sentido de que se busca que se separen de otros ácidos nucleicos. Las técnicas de amplificación se han diseñado principalmente para esta separación. 20

[0070] Según se usa en este documento, el término “cebador” se refiere a un oligonucleótido, de origen natural como en un digerido de restricción purificado o producido sintéticamente, que es capaz de actuar como un 25 punto de inicio de síntesis cuando se le coloca en condiciones en las que se induce la síntesis de un producto de extensión del cebador, que es complementario con una cadena de ácido nucleico (es decir, en presencia de nucleótidos y un agente inductor, tal como una ADN polimerasa y a una temperatura y pH apropiados). Preferiblemente el cebador es de cadena sencilla para una eficacia máxima de amplificación, pero alternativamente puede esta en forma de doble cadena. Si es de doble cadena, el cebador se trata primero para separar sus cadenas 30 antes de su uso para preparar los productos de extensión. Preferiblemente, el cebador es un oligodeoxirribonucleótido. El cebador debe ser suficientemente largo como para cebar la síntesis de productos de extensión en presencia del agente inductor. Las longitudes exactas de los cebadores dependerán de muchos factores como la temperatura, el origen del cebador y el uso del procedimiento.

[0071] Según se usa en este documento, el término “sonda” se refiere a un oligonucleótido (es decir, una secuencia de nucleótidos), ya sea de origen natural como digerido de restricción purificado o producido mediante síntesis, recombinación o amplificación por PCR, que es capaz de hibridar con otro oligonucleótido de interés. Una sonda puede ser de cadena sencilla o doble. Las sondas son útiles para la detección, identificación y aislamiento de 35 secuencias génicas en particular. Se contempla que cualquier sonda utilizada en la presente invención estará marcada con cualquier “molécula indicadora” de modo que sea detectable mediante cualquier sistema de detección incluyendo, pero sin limitaciones sistemas enzimáticos (p. ej., ELISA, así como ensayos de histoquímica basados en enzimas), de fluorescencia, radioactivos y luminiscentes. No se pretende que la presente invención se limite a un sistema de detección o marcaje en particular. 40

[0072] Según se usa en este documento, el término “diana”, cuando se usa en referencia a la reacción en cadena de la polimerasa, se refiere a la región de un ácido nucleico a la que se unen los cebadores usados para la reacción en cadena de la polimerasa. Por tanto, la “diana” se busca mediante selección a partir de otras secuencias de ácido nucleico. Un “segmento” se define como una región de ácido nucleico dentro de la secuencia diana. 45

[0073] Según se usa en este documento, el término “reacción en cadena de la polimerasa” (“PCR”) se refiere a los procedimientos de las patentes de EE. UU. N° 4.683.195, 4.683.202 y 4.965.188, que incluyen procedimientos para aumentar la concentración de un segmento de una secuencia diana en una mezcla de ADN genómico sin clonación ni purificación. Este proceso de amplificación de la secuencia diana consiste en introducir un exceso grande de dos cebadores oligonucleotídicos en la mezcla de ADN que contiene la secuencia diana deseada, 50 seguido de una secuencia precisa de termociclado en presencia de una polimerasa de ADN. Los dos cebadores son complementarios de sus respectivas cadenas de la secuencia diana de cadena doble. Para efectuar la amplificación, la mezcla se desnaturaliza y, a continuación, los cebadores hibridan con sus secuencias complementarias dentro de la molécula diana. Tras la hibridación, los cebadores se extienden con una polimerasa de modo que se forme un nuevo par de cadenas complementarias. Las etapas de desnaturalización, hibridación con el cebador y extensión de la polimerasa pueden repetirse muchas veces (es decir, desnaturalización, hibridación y extensión constituyen un 55 “ciclo”; puede haber numerosos “ciclos”) para obtener una concentración alta de un segmento amplificado de la secuencia diana deseada. La longitud del segmento amplificado de la secuencia diana deseada se determina mediante las posiciones relativas de los cebadores entre sí y, por tanto, esta longitud es un parámetro controlable. En virtud del aspecto repetitivo del proceso, el procedimiento se denomina “reacción en cadena de la polimerasa” (a 60

partir de ahora "PCR"). Puesto que los segmentos amplificados deseados de la secuencia diana se convierten en las secuencias predominantes (en términos de concentración) en la mezcla, se dice que se han "amplificado por PCR".

5 **[0074]** Según se usa en este documento, el término "reactivos de amplificación" se refiere a aquellos reactivos (desoxirribonucleótidos trifosfato, tampón, etc.) necesarios para la amplificación excepto los cebadores, el molde de ácido nucleico y la enzima de amplificación. Típicamente, los reactivos de amplificación junto con otros componentes de reacción se colocan y quedan contenidos en un recipiente de reacción (tubo de ensayo, micropocillo, etc.).

10 **[0075]** Según se usa en este documento, el término "RT-PCR" se refiere a la replicación y amplificación de secuencias de ARN. En este procedimiento, la transcripción inversa se acopla a la PCR, usando con mayor frecuencia un procedimiento enzimático en el que se emplea una polimerasa termoestable, como se describe en la patente de EE. UU. N° 5.322.770. En la RT-PCR, el molde de ARN se convierte en ADN debido a la actividad transcriptasa inversa de la polimerasa y, a continuación, se amplifica usando la actividad polimerizante de la
15 polimerasa (es decir, como en otros procedimientos de PCR).

[0076] Como se usa en este documento, los términos "endonucleasas de restricción" y "enzimas de restricción" se refieren a las enzimas bacterianas, cada una de las cuales corta el ADN de doble cadena en, o cerca de, una secuencia de nucleótidos específica.
20

[0077] Un "sitio de restricción" se refiere a una secuencia de nucleótidos reconocida y escindida por una determinada endonucleasa de restricción y es frecuentemente el sitio para la inserción de fragmentos de ADN. En determinadas realizaciones de la invención los sitios de restricción se introducen por ingeniería genética en el marcador selectivo y en los extremos 5' y 3' de la construcción de ADN.
25

[0078] "Recombinación homóloga" significa el intercambio de fragmentos de ADN entre dos moléculas de ADN o cromosomas apareados en el sitio de secuencias de nucleótidos idénticas o prácticamente idénticas. En una realización preferida, la integración cromosómica es una recombinación homóloga.

30 **[0079]** Según se usa en este documento "aminoácido" se refiere a secuencias peptídicas o proteicas, o a porciones de las mismas. Los términos "proteína", "péptido" y "polipéptido" se usan indistintamente.

[0080] Según se usa en este documento, "proteína de interés" y "polipéptido de interés" se refiere a una proteína/polipéptido que se desea evaluar y/o se está evaluando. En algunas realizaciones, la proteína de interés se
35 expresa a nivel intracelular, mientras que en otras realizaciones, es un polipéptido secretado. En realizaciones especialmente preferidas, entre estas enzimas se incluyen las serina proteasas de la presente invención. En algunas realizaciones, la proteína de interés es un polipéptido secretado que se fusiona con un péptido señal (es decir, una extensión amino terminal en una proteína que se va a secretar). Prácticamente todas las proteínas secretadas utilizan una extensión proteica aminoterminal que tiene un papel principal en el direccionamiento y la translocación
40 de proteínas precursoras a través de la membrana. Esta extensión se elimina proteolíticamente por una señal peptidasa durante la transferencia a través de la membrana o inmediatamente después.

[0081] Se dice que un polinucleótido "codifica" un ARN o un polipéptido si, en su estado nativo o cuando se manipula mediante procedimientos conocidos por los expertos en la técnica, puede transcribirse y/o traducirse
45 para producir el ARN, el polipéptido o un fragmento del mismo. También se dice que la cadena complementaria de este ácido nucleico codifica las secuencias. Según se conoce en la técnica, un ADN puede transcribirse mediante una ARN polimerasa para producir ARN, pero un ARN puede transcribirse de forma inversa mediante una transcriptasa inversa para producir un ADN. Por tanto, un ADN puede codificar un ARN y viceversa.

50 **[0082]** "Cadena huésped" o "célula huésped" se refiere a un huésped adecuado para un vector de expresión que comprende ADN según la presente invención.

[0083] Una enzima se "sobrexprende" en una célula huésped si la enzima se expresa en la célula a nivel más alto del nivel al que se expresa en una célula natural correspondiente.
55

[0084] Los términos "proteína" y "polipéptido" se usan indistintamente en este documento. A lo largo de esta descripción se usa el código de 3 letras para los aminoácidos según se define de conformidad con la Comisión de Nomenclatura Bioquímica (JCBN, por sus siglas en inglés) de la IUPAC-IUB. También se entiende que un polipéptido puede estar codificado por más de una secuencia de nucleótidos debido a la degeneración del código
60 genético.

[0085] Una "prosecuencia" es una secuencia de aminoácidos entre la secuencia señal y una proteasa madura que es necesaria para la secreción de la proteasa. La escisión de la prosecuencia da lugar a una proteasa activa madura.

[0086] El término “secuencia señal” o “péptido señal” se refiere a cualquier secuencia de nucleótidos y/o aminoácidos que puede participar en la secreción de las formas madura o precursora de la proteína. Esta definición de secuencia señal es funcional, y pretende incluir a todas aquellas secuencias de aminoácidos codificadas por la porción N-terminal del gen de la proteína que participa en la efectuación de la secreción de la proteína. A menudo, pero no en todos los casos, se une a la porción N-terminal de una proteína o a la porción N-terminal de una proteína precursora. La secuencia señal puede ser endógena o exógena. La secuencia señal puede ser tal que normalmente se asocie con la proteína (p. ej., proteasa) o puede proceder de un gen que codifica otra proteína secretada. Un ejemplo de secuencia señal exógena comprende los primeros siete restos aminoacídicos de la secuencia señal de la subtilisina de *Bacillus subtilis* fusionados con el resto de la secuencia señal de la subtilisina de *Bacillus lentus* (ATCC 21536).

[0087] El término “secuencia señal híbrida” se refiere a secuencias señal en las que parte de la secuencia se obtiene a partir del huésped de expresión fusionado con la secuencia señal del gen que se va a expresar. En algunas realizaciones, se utilizan secuencias sintéticas.

[0088] El término forma “madura” de una proteína o péptido se refiere a la forma funcional final de la proteína o el péptido. Por ejemplo, una forma madura de la proteasa de la presente invención incluye al menos la secuencia de aminoácidos idéntica a los restos de las posiciones 1-189 de la SEC ID N°2.

[0089] El término forma “precursora” de una proteína o un péptido se refiere a una forma madura de la proteína que tiene una prosequencia unida de forma operativa al extremo amino o carboxilo terminal de la proteína. El precursor también puede tener una secuencia “señal” unida de forma operativa al extremo amino terminal de la prosequencia. El precursor puede también tener polinucleótidos adicionales que estén implicados en la actividad postraduccional (p. ej., polinucleótidos escindidos de los mismos para dar lugar a la forma madura de una proteína o péptido).

[0090] “Enzima de origen natural” se refiere a aquella que tiene la secuencia de aminoácidos no modificada idéntica a la que se encuentra en la naturaleza. Entre las enzimas de origen natural se incluyen enzimas nativas, aquellas enzimas que se expresan de forma natural o que se encuentran en un microorganismo en particular.

[0091] Los términos “derivado de” y “obtenido a partir de” se refieren no sólo a una proteasa producida o producible por una cepa de un organismo en cuestión, sino a una proteasa codificada por una secuencia de ADN aislada de dicha cepa y producida en un organismo huésped que contiene dicha secuencia de ADN. Adicionalmente, el término se refiere a una proteasa codificada por una secuencia de ADN de origen sintético y/o ADNc y que tiene las características que identifican a la proteasa en cuestión. Como ejemplo, “proteasas derivadas de *Cellulomonas*” se refiere a aquellas enzimas que tienen actividad proteolítica que *Cellulomonas* producen de forma natural, así como a serina proteasas como las producidas por fuentes de *Cellulomonas* pero que a través del uso de técnicas de ingeniería genética son producidas por organismos distintos a *Cellulomonas* transformados con un ácido nucleico que codifica dichas serina proteasas.

[0092] Un “derivado” dentro del alcance de esta invención generalmente retiene la actividad proteolítica característica observada en la forma natural, nativa o parental hasta el grado en que el derivado es útil para fines similares a los de la forma natural, nativa o parental. Los derivados funcionales de la serina proteasa abarcan péptidos o fragmentos peptídicos naturales o producidos mediante síntesis o tecnología recombinante que tienen las características generales de la serina proteasa de la presente invención.

[0093] El término “derivado funcional” se refiere a un derivado de un ácido nucleico que tiene las características funcionales de un ácido nucleico que codifica la serina proteasa. Los derivados funcionales de un ácido nucleico que codifican la serina proteasa de la presente invención abarcan ácidos nucleicos o fragmento naturales, obtenidos mediante síntesis o por tecnología recombinante y codifican la serina proteasa característica de la presente invención. El ácido nucleico natural que codifica las serina proteasas según la invención incluye alelos naturales y homólogos en base a la degeneración del código genético conocida en la técnica.

[0094] El término “idéntico” en el contexto de dos ácidos nucleicos o secuencias polipeptídicas se refiere a los restos en las dos secuencias que son los mismos cuando se alienan para una correspondencia máxima medida usando uno de los siguientes algoritmos de comparación o análisis de secuencia.

[0095] El término “alineamiento óptimo” se refiere al alineamiento que ofrece el mayor porcentaje de puntuación de identidad.

[0096] “Porcentaje de identidad de secuencia”, “porcentaje de identidad de secuencia de aminoácidos”, “porcentaje de identidad de secuencia de genes” y/o “porcentaje de identidad de secuencia de ácido

nucleico/polinucleótido” con respecto a dos secuencias de aminoácidos, polinucleótidos y/o genes (según proceda) se refieren al porcentaje de residuos que son idénticos en las dos secuencias cuando las secuencias están óptimamente alineadas. Por tanto, una identidad de secuencia de aminoácidos del 80% significa que el 80% de los aminoácidos en dos secuencias polipeptídicas óptimamente alineadas son idénticos.

5

[0097] La expresión “sustancialmente idéntico” en este contexto de dos ácidos nucleicos o polipéptidos se refiere, por tanto, a un polinucleótido o polipéptido que comprende al menos una identidad de secuencia del 70%, preferiblemente al menos del 75%, preferiblemente al menos del 80%, preferiblemente al menos del 85%, preferiblemente al menos del 90%, preferiblemente al menos del 95%, preferiblemente al menos del 97%,
10 preferiblemente al menos del 98% y preferiblemente al menos del 99% en comparación con una secuencia de referencia usando los programas o algoritmos (p. ej., BLAST, ALIGN o CLUSTAL) que usan parámetros convencionales. Una indicación de que dos polipéptidos son sustancialmente idénticos es que el primer polipéptido es inmunológicamente crosreactivo con el segundo polipéptido. Típicamente, los polipéptidos que difieren en sustituciones conservadoras de aminoácidos son inmunológicamente crosreactivos. De este modo, un polipéptido es
15 sustancialmente idéntico a un segundo polipéptido, por ejemplo, cuando los dos péptidos difieren sólo en las sustituciones conservadoras. Otra indicación de que las dos secuencias de ácido nucleico son sustancialmente idénticas es que las dos moléculas hibridan entre sí en condiciones rigurosas (p. ej., dentro de un intervalo de rigurosidad media a alta).

20 **[0098]** La expresión “equivalente” en este contexto se refiere a enzimas serina proteasas que están codificadas por un polinucleótido capaz de hibridar con el polinucleótido que tiene la secuencia mostrada en la SEC ID N° 1, en condiciones de rigurosidad media a máxima. Por ejemplo, ser equivalente significa que una serina proteasa madura equivalente comprende al menos el 70%, al menos el 75%, al menos el 80%, al menos el 90%, al menos el 91%, al menos el 92%, al menos el 93%, al menos el 94%, al menos el 95%, al menos el 96%, al menos el
25 97%, al menos el 98% y/o al menos el 99% de identidad de secuencia con una serina proteasa de *Cellulomonas* madura que tiene la secuencia de aminoácidos de SEC ID N° 2.

[0099] El término “aislado” o “purificado” se refiere a un material que se extrae de su entorno original (p. ej., el entorno natural si es de origen natural). Por ejemplo, el material se sabe que está “purificado” cuando está
30 presente en una composición en particular en una concentración más alta o baja de la que existe en un organismo natural o en combinación con componentes que normalmente no están presentes tras la expresión en un organismo natural. Por ejemplo, un polinucleótido o polipéptido natural presente en un animal vivo no está aislado, pero el mismo polinucleótido o polipéptido, separado de algunos o de todos los materiales con los que coexiste en el sistema natural, está aislado. En algunas realizaciones, estos polinucleótidos son parte de un vector y/o estos
35 polinucleótidos o polipéptidos son parte de una composición y siguen estando aislados porque este vector o composición no es parte de su entorno natural. En realizaciones preferidas, se dice que un ácido nucleico o proteína está purificado, por ejemplo, si da lugar esencialmente a una banda en un gel o transferencia electroforética.

[0100] El término “aislado”, cuando se usa en referencia a una secuencia de ADN, se refiere a una
40 secuencia de ADN que se ha extraído de su entorno genético natural y, por tanto, está libre de otras secuencias codificadoras extrañas o no deseadas y está en forma adecuada para su uso dentro de sistemas de producción de proteínas genéticamente modificadas por ingeniería proteica. Estas moléculas aisladas son aquellas que se separan de su entorno natural e incluye ADNc y clones genómicos. Las moléculas de ADN aisladas de la presente invención están libres de otros genes con lo que se asocian normalmente, pero pueden incluir regiones no traducidas en los
45 extremos 5' y 3' naturales como promotores y terminadores. La identificación de regiones asociadas será evidente para un experto en la técnica (véase p. ej., Dynan y Tijan, *Nature* 316:774-78 [1985]). El término “una secuencia de ADN aislada” se denomina alternativamente como “secuencia de ADN clonada”.

[0101] El término “aislado” cuando se usa en referencia a una proteína, se refiere a una proteína que se
50 encuentra en una condición distinta de su entorno nativo. En una forma preferida, la proteína aislada está sustancialmente libre de otras proteínas, especialmente de otras proteínas homólogas. Una proteína aislada está más del 10% pura, preferiblemente más del 20% pura e incluso, más preferiblemente, más del 30% pura, según se determina mediante PAGE-SDS. Otros aspectos de la invención abarcan la proteína en una forma altamente purificada (p. ej., pura en más del 40%, pura en más del 60%, pura en más del 80%, pura en más del 90%, pura en
55 más del 95%, pura más del 97% e incluso pura en más del 99%) según se determina mediante PAGE-SDS.

[0102] Según se usa en este documento, el término “mutagénesis combinatoria” se refiere a
60 procedimientos en los que se generan bibliotecas de variantes de una secuencia inicial. En estas bibliotecas, las variantes contienen una o varias mutaciones elegidas a partir de un conjunto de mutaciones predefinidas. Además los procedimientos proporcionan medios para introducir mutaciones aleatorias que no eran miembros del conjunto de mutaciones predefinidas. En algunas realizaciones, los procedimientos incluyen aquellos mostrados en la solicitud de patente de EE. UU. N° de serie 09/699.250 (US 6.582.914). En realizaciones alternativas, los procedimientos de mutagénesis combinatoria abarcan kit disponibles en el mercado (p. ej., QuikChange® Multisite, Stratagene, San Diego, CA).

- 5 **[0103]** Según se usa en este documento, el término “biblioteca de mutantes” se refiere a una población de células que son idénticas en la mayoría de su genoma aunque incluye homólogos diferentes de uno o más genes. Estas bibliotecas pueden usarse, por ejemplo, para identificar genes u operones con características mejoradas.
- [0104]** Según se usa en este documento, el término “gen inicial” se refiere a un gen de interés que codifica una proteína de interés que tiene que mejorarse y/o cambiarse usando la presente invención.
- 10 **[0105]** Según se usa en este documento, el término “alineamiento de secuencia múltiple” (“ASM”) se refiere a las secuencias de homólogos múltiples de un gen inicial que se alinean usando un algoritmo (p. ej., Clustal W).
- 15 **[0106]** Según se usa en este documento, los términos “secuencia consenso” y “secuencia canónica” se refieren a una secuencia de aminoácidos arquetipo frente a la cual se comparan todas las variantes de una proteína o secuencia de interés en particular. Los términos también se refieren a una secuencia en la que se muestran los nucleótidos que se presentan con mayor frecuencia en la secuencia de ADN de interés. Para cada posición de un gen, la secuencia consenso proporciona el aminoácido que es más abundante en esta posición en el ASM.
- 20 **[0107]** Según se usa en este documento, el término “mutación consenso” se refiere a una diferencia entre la secuencia de un gen inicial y una secuencia consenso. Las mutaciones consenso se identifican comparando las secuencias del gen inicial y la secuencia consenso resultante de la ASM. En algunas realizaciones, las mutaciones consenso se introducen en el gen inicial de modo que se convierten en más similares a la secuencia consenso. Las mutaciones consenso también incluyen cambios de aminoácidos en los que se cambia un aminoácido de un gen inicial por otro aminoácido que se encuentra con mayor frecuencia en una ASM en esta posición relativa a la
- 25 frecuencia de este aminoácido en el gen inicial. Por tanto, el término mutación consenso comprende todos los cambios sencillos de aminoácidos en los que se sustituye un aminoácido del gen inicial por un aminoácido que es más abundante que el aminoácido de la ASM.
- [0108]** Según se usa en este documento, el término “golpe inicial” se refiere a una variante que se identificó mediante análisis de una biblioteca de mutagénesis de consenso combinatoria. En realizaciones
- 30 preferidas, los golpes iniciales han mejorado las características funcionales mejoradas en comparación con el gen inicial.
- [0109]** Según se usa en este documento, el término “golpe mejorado” se refiere a una variante que se identificó por análisis de una biblioteca de mutagénesis de consenso combinatoria mejorada.
- 35 **[0110]** Según se usa en este documento, los términos “mutación mejorada” y “mutación de rendimiento mejorado” se refieren a una mutación que induce una mejora del rendimiento cuando se introduce en el gen inicial. En algunas realizaciones preferidas, estas mutaciones se identifican mediante golpe de secuenciación que se identificaron durante el paso de análisis del procedimiento. En la mayoría de las realizaciones, las mutaciones que
- 40 se encuentran con mayor frecuencia en los golpes son probablemente mutaciones de mejora, en comparación con una biblioteca de mutagénesis de consenso combinatoria no analizada.
- [0111]** Según se usa en este documento, el término “biblioteca de mutagénesis de consenso combinatoria potenciada” se refiere a una biblioteca MCC que se diseña y construye con base en resultados de
- 45 selección y/o secuenciación a partir de una ronda de mutagénesis MCC y selección. En algunas realizaciones, la biblioteca MCC potenciada se basa en la secuencia de un golpe inicial resultado de una ronda de MCC más temprana. En realizaciones adicionales, la MCC potenciada se diseña de modo que se favorecen las mutaciones que se observaban con frecuencia en los golpes iniciales de rondas más tempranas de mutagénesis y selección. En algunas realizaciones preferidas, esto se consigue omitiendo cebadores que codifican mutaciones que reducen el
- 50 rendimiento o aumentando la concentración de cebadores que codifican mutaciones que potencian el rendimiento en relación con otros cebadores que se usaron en bibliotecas MCC más tempranas.
- [0112]** Según se usa en este documento, el término “mutaciones que reducen el rendimiento” se refiere a mutaciones en la biblioteca de mutagénesis de consenso combinatoria que se encuentran con menos frecuencia en
- 55 los golpes resultantes del análisis en comparación con la biblioteca de mutagénesis de consenso combinatoria no analizadas. En realizaciones preferidas, el proceso de análisis elimina y/o reduce la abundancia de variantes que contienen “mutaciones que reducen el rendimiento”.
- [0113]** Según se usa en este documento, el término “ensayo funcional” se refiere a un ensayo que
- 60 proporciona una indicación de una actividad de la proteína. En realizaciones especialmente preferidas, el término se refiere a sistemas de ensayo en los que se analiza la capacidad funcional en su capacidad normal. Por ejemplo, en el caso de enzimas, un ensayo funcional implica determina la eficacia de la enzima a la hora de catalizar una reacción.

[0114] Según se usa en este documento, el término “propiedad diana” se refiere a la propiedad del gen inicial que se va a alterar. No se pretende que la presente invención se limite a cualquier propiedad diana en particular. Sin embargo, en algunas realizaciones preferidas, la propiedad objetivo es la estabilidad de un producto génico (p. ej., resistencia a la desnaturalización, proteólisis u otros factores degradativos) mientras que en otras realizaciones, se altera el nivel de producción en un huésped de producción. De hecho, se contempla que en la presente invención se usará cualquier propiedad de un gen inicial.

[0115] El término “propiedad” o los equivalentes gramaticales del mismo en el contexto de un ácido nucleico, según se usa en este documento, se refieren a cualquier característica o atributo de un ácido nucleico que puede seleccionarse o deleccionarse. Entre estas propiedades se incluyen, pero sin limitaciones, una propiedad que afecta a la unión a un polipéptido, una propiedad conferida a una célula que comprende un ácido nucleico en particular, una propiedad que afecta a la transcripción génica (p. ej., potencia del promotor, reconocimiento del promotor, regulación del promotor, función de potenciador, etc.), una propiedad que afecta al procesamiento del ARN, (p. ej., ajuste del ARN, estabilidad del ARN, conformación del ARN y modificación postranscripcional), una propiedad que afecta a la traducción (p. ej., nivel, regulación, unión del ARNm a las proteínas ribosomales, modificación postraduccional, etc.). Por ejemplo, un sitio de unión para un factor de transcripción, polimerasa, factor de regulación, etc., de un ácido nucleico puede alterarse para producir las características deseadas o para identificar características no deseadas.

[0116] El término “propiedad” o los equivalentes gramaticales del mismo en el contexto de un polipéptido (incluyendo proteínas), según se usa en este documento, se refieren a cualquier característica o atributo de un polipéptido que puede seleccionarse o deleccionarse. Entre estas propiedades se incluyen, pero sin limitaciones, la estabilidad oxidativa, especificidad de sustrato, actividad catalítica, estabilidad térmica, estabilidad alcalina, perfil de actividad de pH, resistencia a la degradación proteolítica, K_M , K_{cat} , relación K_{cat}/K_M , plegamiento de proteínas, inducción de una respuesta inmunitaria, capacidad para unirse a un ligando, capacidad para unirse a un receptor, capacidad para ser secretada, capacidad para expresarse en la superficie de una célula, capacidad para formar un oligómero, capacidad de señalización, capacidad para estimular la proliferación celular, capacidad para inhibir la proliferación celular, capacidad para inducir apoptosis, capacidad para ser modificada mediante fosforilación o glucosilación y/o capacidad para tratar una enfermedad, etc.

[0117] Según se usa en este documento, el término “selección” tiene su significado habitual en la técnica y es, en general, un proceso de múltiples etapas. En la primera etapa, se proporciona un ácido nucleico mutante o un polipéptido variante del mismo. En la segunda etapa, se determina una propiedad del ácido nucleico mutante o del polipéptido variante. En la tercera etapa, la propiedad determinada se compara con una propiedad del ácido nucleico precursor correspondiente, con la propiedad del polipéptido natural correspondiente o con la propiedad del material inicial (p. ej., la secuencia inicial) para la generación del ácido nucleico mutante.

[0118] Será aparente para el experto en la materia que el procedimiento de selección para obtener un ácido nucleico o proteína con una propiedad alterada depende de la propiedad del material inicial, cuya modificación pretende facilitar la generación del ácido nucleico mutante. Por tanto, el experto en la materia apreciará que la invención no se limita a ninguna propiedad específica seleccionada y que en la siguiente descripción de propiedades solo se enumeran ejemplos ilustrativos. Generalmente, en la técnica se describen procedimientos de selección para cualquier propiedad en particular. Por ejemplo, puede medirse la unión, pH, especificidad etc., antes y durante la mutación en la que un cambio indica una alteración. Preferiblemente, las selecciones se realizan con un alto rendimiento, incluyendo el análisis simultáneo de múltiples muestras, incluyendo, pero sin limitaciones, ensayos que utilizan chips, despliegue de fagos y sustratos múltiples y/o indicadores.

[0119] Según se usa en este documento, en algunas realizaciones, las selecciones abarcan etapas de selección en las que las variantes de interés se enriquecen a partir de una población de variantes. Entre los ejemplos de estas realizaciones se incluyen la selección de variantes que confiere una ventaja de crecimiento al organismo huésped, así como despliegue de fagos o cualquier otro procedimiento de despliegue, donde pueden capturarse variantes a partir de una población de variantes en base a sus propiedades de unión o catalíticas. En una realización preferida, una biblioteca de variantes se expone a estrés (calor, proteasa, desnaturalización, etc.) y, posteriormente, se identifican las variantes que siguen intactas en un análisis o se enriquecen mediante selección. Se pretende que el término abarque cualquier medida adecuada para la selección. De hecho, no se pretende que la presente invención se limite a cualquier procedimiento de selección en particular.

[0120] Según se usa en este documento, el término "aleatorización dirigida" se refiere a un proceso que produce varias secuencias en las que se han aleatorizado una o varias posiciones. En algunas realizaciones, la asignación aleatoria es completa (es decir, puede tener lugar en los 4 nucleótidos A, T, G y C en una posición aleatorizada). En realizaciones alternativas, la asignación aleatoria de un nucleótido se limita a un subgrupo de los cuatro nucleótidos. La aleatorización dirigida puede aplicarse a uno o varios codones de una secuencia que codifica para una o varias proteínas de interés. Cuando se expresan, las bibliotecas resultantes producen poblaciones de proteínas en las que una o más posiciones de aminoácidos pueden contener una mezcla de los 20 aminoácidos o un

subgrupo de aminoácidos, según se determine en el esquema de aleatorización del codón aleatorizado. En algunas realizaciones, los miembros individuales de una población resultante de la aleatorización diana difieren en el número de aminoácidos, debido a la inserción o delección dirigida o aleatoria de codones. En realizaciones adicionales, se incluyen aminoácidos sintéticos en las poblaciones de proteínas producidas. En algunas realizaciones preferidas, la mayoría de los miembros de una población que resulta de la asignación aleatoria dirigida muestra una homología de secuencia mayor a la secuencia consenso que el gen inicial. En algunas realizaciones, la secuencia codifica una o más proteínas de interés. En realizaciones alternativas, las proteínas tienen funciones biológicas diferentes. En algunas realizaciones preferidas, la secuencia entrante comprende al menos un marcador seleccionable. Esta secuencia puede codificar una o más proteínas de interés y puede tener otra función biológica. En muchos casos la secuencia entrante incluirá un marcador seleccionable como un gen que confiere resistencia a un antibiótico.

[0121] Los términos “secuencia modificada” y “genes modificados” se usan indistintamente en este documento para referirse a una secuencia que incluye una delección, inserción o interrupción de una secuencia de ácido nucleico natural. En algunas realizaciones preferidas, el producto de expresión de la secuencia modifica una proteína truncada (p. ej., si la modificación es una delección o interrupción de la secuencia). En algunas realizaciones especialmente preferidas, la proteína truncada retiene la actividad biológica. En realizaciones alternativas, el producto de expresión de la secuencia modificada es una proteína elongada (p. ej., modificaciones que comprenden una inserción dentro de la secuencia del ácido nucleico). En algunas realizaciones, una inserción lleva a una proteína truncada (p. ej., cuando la inserción da lugar a la formación de un codón de parada). Por tanto, una inserción puede dar lugar a una proteína truncada o una proteína elongada como producto de expresión.

[0122] Según se usa en este documento, los términos “secuencia mutante” y “gen mutante” se usan indistintamente y se refieren a una secuencia que tiene una alteración en al menos un codón que se produce en una secuencia natural en la célula huésped. El producto de expresión de la secuencia mutante es una proteína con una secuencia de aminoácidos alterada con respecto a la natural. El producto de expresión puede tener una capacidad funcional alterada (p. ej., actividad enzimática potenciada).

[0123] Se pretende que los términos “cebador mutagénico” u “oligonucleótido mutagénico” (utilizados indistintamente en este documento) se refieran a composiciones de oligonucleótidos que se corresponden con una porción de la secuencia molde y que es capaz de hibridar con la misma. Con respecto a los cebadores mutagénicos, el cebador no coincidirá de forma precisa con el ácido nucleico molde, usándose las coincidencias o no coincidencias en el cebador para introducir la mutación deseada en la biblioteca de ácidos nucleicos. Según se usa en este documento, “cebador no mutagénico” u “oligonucleótido no mutagénico” se refieren a composiciones oligonucleotídicas que coincidirán de forma precisa con el ácido nucleico molde. En una realización de la invención, solo se usan cebadores mutagénicos. En otra realización preferida de la invención, los cebadores se diseñan de modo que se haya incluido al menos una región para un cebador mutagénico; también se incluye en la mezcla de oligonucleótidos un cebador no mutagénico. Añadiendo una mezcla de cebadores mutagénicos y cebadores no mutagénicos que se corresponden con al menos uno de los cebadores mutagénicos, es posible producir una biblioteca de ácidos nucleicos resultante en la que se presenten diversos patrones de mutación combinatoria. Por ejemplo, si se desea que algunos de los miembros de la biblioteca de ácidos nucleicos mutantes retenga su secuencia precursora en determinadas posiciones mientras que otros miembros son mutantes en estos sitios, los cebadores no mutagénicos proporcionan la capacidad de obtener un nivel específico de miembros no mutantes dentro de la biblioteca de ácidos nucleicos para un resto determinado. Los procedimientos de la invención emplean oligonucleótidos mutagénicos y no mutagénicos que generalmente tienen una longitud de entre 10 y 50 bases, más preferiblemente una longitud de aproximadamente 15 a 45 bases. Sin embargo, puede ser necesario usar cebadores que son más cortos de 10 bases o más largos de 50 bases para obtener el resultado de la mutagénesis deseada. Con respecto a los correspondientes cebadores mutagénicos y no mutagénicos, no es necesario que los correspondientes oligonucleótidos sean de longitud idéntica, sino sólo que solapen en la región correspondiente a la mutación que se va a añadir. Pueden añadirse cebadores en una proporción predefinida según la presente invención. Por ejemplo, si se desea que la biblioteca resultante tenga un nivel significativo de una determinada mutación específica y una cantidad menor de una mutación diferente en el mismo sitio o en diferente sitio ajustando la cantidad de cebador añadido, es posible producir la biblioteca sesgada deseada. Alternativamente, añadiendo cantidades menores o mayores de cebadores no mutagénicos, es posible ajustar la frecuencia con que se producen las correspondientes mutaciones en la biblioteca de ácidos nucleicos mutantes.

[0124] Según se usa en este documento, la expresión “mutaciones contiguas” se refiere a mutaciones que se presentan dentro del mismo cebador oligonucleotídico. Por ejemplo, las mutaciones contiguas pueden ser adyacentes o estar próximas entre sí, sin embargo, se introducirán en los ácidos nucleicos molde mutantes resultantes para el mismo cebador.

[0125] Según se usa en este documento, la expresión “mutaciones discontinuas” se refiere a mutaciones que se presentan en cebadores oligonucleotídicos independientes. Por ejemplo, se introducirán mutaciones discontinuas en los ácidos nucleicos cebadores mutantes resultantes mediante oligonucleótidos preparados independientemente.

[0126] Los términos “secuencia natural” o “gen natural” se usan indistintamente en este documento, para referirse a una secuencia que es nativa o de origen natural en una célula huésped. En algunas realizaciones, secuencia natural se refiere a una secuencia de interés que es el punto de inicio de un proyecto de ingeniería de proteínas. La secuencia natural puede codificar una proteína homóloga o heteróloga. Una proteína homóloga es aquella que la célula huésped podría producir sin intervención. Una proteína heteróloga es aquella que la célula huésped no podría producir sino es mediante la intervención.

[0127] Según se usa en este documento, el término “anticuerpos” se refiere a inmunoglobulinas. Anticuerpos incluye, pero sin limitaciones, las inmunoglobulinas obtenidas directamente de cualquier especie a partir de la que se desea producir anticuerpos. Además, la presente invención abarca anticuerpos modificados. El término también se refiere a fragmentos de anticuerpo que retienen la capacidad de unión al epítope al que se une el anticuerpo intacto e incluye anticuerpos policlonales, anticuerpos monoclonales, anticuerpos quiméricos y anticuerpos antiidiotipo (anti-ID). Entre los fragmentos de anticuerpo se incluyen, pero sin limitaciones, las regiones determinantes de complementariedad (CDR), regiones del fragmento variable de cadena sencilla (scFv), región variable de la cadena pesada (VH) o región variable de la cadena ligera (VL). La presente invención también abarca anticuerpos policlonales y monoclonales. Preferiblemente, los anticuerpos son anticuerpos monoclonales.

[0128] El término “oxidación estable” se refiere a proteasas de la presente invención que retienen una cantidad específica de la actividad enzimática durante un periodo de tiempo determinado en condiciones predominantes durante el proceso proteolítico, de hidrolización, de limpieza o de otro tipo de la invención, por ejemplo mientras se expone o pone en contacto con agentes blanqueantes o agentes oxidantes. En algunas realizaciones, las proteasas retienen al menos el 50%, 60%, 70%, 75%, 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% o 99% de la actividad proteolítica tras el contacto con el agente blanqueante u oxidante durante un periodo de tiempo determinado, por ejemplo, al menos 1 minuto, 3 minutos, 5 minutos, 8 minutos, 12 minutos, 16 minutos, 20 minutos, etc. En algunas realizaciones, la estabilidad se mide como se describe en los ejemplos.

[0129] El término “quelante estable” se refiere a proteasas de la presente invención que retienen una cantidad específica de la actividad enzimática durante un periodo de tiempo determinado en condiciones predominantes durante el proceso proteolítico, de hidrolización, de limpieza o de otro tipo de la invención, por ejemplo mientras se expone o pone en contacto con agentes quelantes. En algunas realizaciones, las proteasas retienen al menos el 50%, 60%, 70%, 75%, 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% o 99% de la actividad proteolítica tras el contacto con el agente quelante durante un periodo de tiempo determinado, por ejemplo, al menos 10 minutos, 20 minutos, 40 minutos, 60 minutos, 100 minutos, etc. En algunas realizaciones, la estabilidad se mide como se describe en los ejemplos.

[0130] Los términos “térmicamente estable” y “termoestable” se refieren a proteasas de la presente invención que retienen una cantidad específica de la actividad enzimática tras la exposición a temperaturas identificadas durante un periodo tiempo determinado en condiciones predominantes durante el proceso proteolítico, de hidrolización, de limpieza o de otro tipo de la invención, por ejemplo mientras se expone a temperaturas alteradas. El término temperaturas alteradas incluye aumento o disminución de las temperaturas. En algunas realizaciones, las proteasas retienen al menos el 50%, 60%, 70%, 75%, 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% o 99% de la actividad proteolítica tras la exposición a temperaturas alteradas durante un periodo de tiempo determinado, por ejemplo, al menos 60 minutos, 120 minutos, 180 minutos, 240 minutos, 300 minutos, etc. En algunas realizaciones, la termoestabilidad se determina como se describe en los ejemplos.

[0131] El término “estabilidad potenciada” en el contexto de una proteasa estable a la oxidación, quelación, al calor y/o pH se refiere a una mayor actividad proteolítica retenida a lo largo del tiempo en comparación con otras serina proteasas (p. ej., proteasas subtilisinas) y/o enzimas naturales.

[0132] El término “estabilidad disminuida” en el contexto de una proteasa estable a la oxidación, quelación, al calor y/o pH se refiere a una menor actividad proteolítica retenida a lo largo del tiempo en comparación con otras serina proteasas (p. ej., proteasas subtilisinas) y/o enzimas naturales.

[0133] El término “actividad de limpieza” se refiere al rendimiento de limpieza conseguido por la proteasa en condiciones prevalentes durante el proceso proteolítico, de hidrolización, de limpieza o de otro tipo de la invención. En algunas realizaciones, el rendimiento de limpieza se determina mediante la aplicación de diversos ensayos de limpieza correspondientes a manchas sensibles a enzimas, por ejemplo, proteínas de hierba, sangre, leche o huevo según se determina mediante diversas metodologías cromatográficas, espectrofotométricas u otra metodología cuantitativa tras someter a las manchas a condiciones de lavado convencionales. Entre los ejemplos de ensayos se incluyen, pero sin limitaciones, los descritos en el documento WO 99/34011 y en la patente de EE. UU. 6.605.458 así como aquellos procedimientos incluidos en los ejemplos.

[0134] El término “cantidad eficaz de limpieza” de una proteasa se refiere a la cantidad de proteasa

descrita a continuación en este documento que consigue un nivel deseado de actividad enzimática en una composición de limpieza específica. Un experto en la materia puede determinar fácilmente dichas cantidades eficaces en base a muchos factores, como la proteasa utilizada en particular, la aplicación de limpieza, la composición específica de la composición de limpieza y si se requiere una composición líquida o seca (p. ej., 5 granular o en barra), etc.

[0135] El término "materiales auxiliares de limpieza" según se usa en este documento significa cualquier material líquido, sólido o gaseoso seleccionado para el tipo en particular de composición de limpieza deseada y la forma del producto (p. ej., composición líquida, en gránulos, en polvo, en barra, en pasta, nebulizadora, en 10 comprimidos, en gel o espuma), cuyos materiales son también preferiblemente compatibles con la enzima proteasa usada en la composición. En algunas realizaciones, las composiciones granulares están en forma "compacta" mientras que, en otras realizaciones, las composiciones líquidas están en forma "concentrada".

[0136] El término "rendimiento potenciado" en el contexto de la actividad de limpieza se refiere a una 15 actividad de limpieza aumentada o mayor de determinadas manchas sensibles a enzimas, como huevo, leche, hierba o sangre, según se determina mediante evaluación normal tras un ciclo de lavado convencional y/o múltiples ciclos de lavado.

[0137] El término "rendimiento disminuido" en el contexto de la actividad de limpieza se refiere a una 20 actividad de limpieza disminuida o menor de determinadas manchas sensibles a enzimas, como huevo, leche, hierba o sangre, según se determina mediante evaluación normal tras un ciclo de lavado convencional.

[0138] El término "rendimiento comparativo" en el contexto de la actividad de limpieza se refiere al menos al 60%, al menos el 70%, al menos el 80%, al menos el 90%, al menos el 95% de la actividad de limpieza de una 25 proteasa subtilisina comparativa (p. ej., proteasas disponibles en el mercado), incluyendo pero sin limitaciones proteasas OPTIMASE™ (Genencor), productos de proteasas PURAFECT™ (Genencor), proteasas SAVINASE™ (Novozymes), variantes BPN' (véase p. ej., patente de EE. UU. N° Re 34.606), proteasa RELASE™, DURAZYME™, EVERLASE™, KANNASE™ (Novozymes), proteasas MAXACAL™, MAXAPEM™, PROPERASE™ (Genencor; véase también, la patente de EE. UU. N° Re 34.606 y patente de EE. UU. N° 5.700.676, 5.955.340, 6.312.936 y 30 6.482.628) y productos de la variante de proteasas de *B. lentus* (p. ej., los descritos en los documentos WO 92/21760, WO 95/23221 y/o WO 97/07770). Entre los ejemplos de variantes de la proteasa subtilisina se incluyen, pero sin limitaciones, aquellas que tienen sustituciones o deleciones en las posiciones de restos equivalentes a las posiciones 76, 101, 103, 104, 120, 159, 167, 170, 194, 195, 217, 232, 235, 236, 245, 248 y/o 252 de BPN'. El rendimiento de limpieza puede determinarse comparando las proteasas de la presente invención con los de las 35 proteasas subtilisinas en diversos ensayos de limpieza correspondientes a manchas sensibles a enzimas como hierba, sangre o leche según se determina mediante metodologías espectrofotométricas o analíticas tras condiciones de ciclo de lavado convencional.

[0139] Según se usa en este documento, "composiciones de limpieza de tejidos" incluye composiciones 40 detergentes para lavado a mano o a máquina incluyendo composiciones aditivas de lavandería y composiciones adecuadas para su uso en el remojo y/o pretratamiento de tejidos manchados (p. ej., prendas de ropa, ropa blanca y otros materiales textiles).

[0140] Según se usa en este documento, "composiciones de limpieza no textiles" incluye composiciones 45 de limpieza de superficies no textiles (es decir, tejidos) que incluyen, pero sin limitaciones composiciones de detergente para lavavajillas, composiciones para higiene bucal, composiciones para la limpieza de dentaduras y composiciones de higiene personal.

[0141] La forma "compacta" de las composiciones de limpieza de este documento se refleja mejor por su 50 densidad y, en términos de composición, por la cantidad de sal inorgánica de relleno. Las sales inorgánicas de relleno son componentes convencionales de composiciones detergentes en forma de polvo. En composiciones detergentes convencionales, las sales de relleno se presentan en cantidades sustanciales, típicamente del 17 al 35% en peso de la composición total. Por el contrario, en composiciones compactas, la sal de relleno está presente en cantidades que no exceden del 15% de la composición total. En algunas realizaciones, la sal de relleno se 55 presenta en cantidades que no exceden del 10%, o más preferiblemente, el 5% en peso de la composición. En algunas realizaciones, las sales inorgánicas de relleno se seleccionan entre las sales de metales alcalinos y alcalinotérreos de sulfatos y cloruros. Una sal de relleno preferida es el sulfato sódico.

PROCEDIMIENTO EXPERIMENTAL

60 **[0142]** Se pretende que las figuras adjuntas se consideren parte integral de la memoria y descripción de la invención. Los siguientes ejemplos se ofrecen para ilustrar, pero no limitar, la invención reivindicada.

[0143] En la siguiente descripción experimental se aplican las siguientes abreviaturas: IP (inhibidor de

proteasa), ppm (partes por millón); M (molar); mM (milimolar); μ M (micromolar); nM (nanomolar); mol (moles); mmol (milimoles); μ mol (micromoles); nmol (nanomoles); g (gramos); mg (miligramos); μ g (microgramos); pg (picogramos); l (litros); ml y mL (mililitros); μ l y μ L (microlitros); cm (centímetros); mm (milímetros); μ m (micrómetros); nm (nanómetros); U (unidades); V (voltios); PM (peso molecular); s (segundos); min(s) (minuto/minutos); h(s) y hr(s) (hora/horas); °C. (grados centígrados); CS (cantidad suficiente); ND (no determinado); NA (no aplicable); rpm (revoluciones por minuto); H₂O (agua); dH₂O (agua desionizada); HCl (Ácido clorhídrico); aa (aminoácido); pb (par de bases); kb (par de kilobases); kD (kilodalton); ADNc (ADN copia o complementario); ADN (ácido desoxirribonucleico); ADNcs (ADN de cadena sencilla); ADNcd (ADN de cadena doble); dNTP (desoxirribonucleótido trifosfato); ARN (ácido ribonucleico); MgCl₂ (cloruro de magnesio); NaCl (cloruro sódico); p/v (peso con respecto a volumen); v/v (volumen con respecto a volumen); g (gravedad); DO (densidad óptica); solución tamponada con fosfato de Dulbecco (DPBS); SOC (Bacto-Tripton a 2%, extracto de levadura Bacto al 0,5% Bacto, NaCl 10 mM, KCl 2,5 mM); medio Terrific Broth (TB; 12 g/l de Bacto Tripton, 24 g/l de glicerol, 2,31 de g/l KH₂PO₄ y 12,54 g/l de K₂HPO₄); DO₂₈₀ (densidad óptica a 280 nm); DO₆₀₀ (densidad óptica a 600 nm); A₄₀₅ (absorbancia a 405 nm); V_{máx} (la velocidad inicial máxima de una reacción catalizada por una enzima); PAGE (electroforesis en gel de poliacrilamida); PBS (solución salina tamponada con fosfato [NaCl 150 mM, tampón fosfato sódico 10 mM, pH 7,2]); PBST (PBS+TWEEN® 20 al 0,25%); PEG (polietilenglicol); PCR (reacción en cadena de polimerasa); RT-PCR (PCR con transcripción inversa); SDS (dodecil sulfato sódico); Tris (tris(hidroximetil)aminometano); HEPES (ácido N-[2-hidroxietil]piperacina-N-[2-etanosulfónico]); HBS (solución salina tamponada con HEPES); Tris-HCl (tris[hidroximetil]aminometano-clorhídrico); Tricina (N-[tris-(hidroximetil)-metil]-glicina); CHES (ácido 2-(N-ciclohexilamino) etanosulfónico); TAPS (ácido 3-[[tris-(hidroximetil)-metil]-amino]-propanosulfónico); CAPS (ácido 3-(ciclohexilamino)-propanosulfónico); DMSO (dimetil sulfóxido); DTT (1,4-ditio-DL-treitol); SA (ácido sinapínico (ácido s,5-dimetoxi-4-hidroxi cinámico)); TCA (ácido tricloroacético); Glut y GSH (glutación reducido); GSSG (glutación oxidado); TCEP (Tris[2-carboxietil]fosfina); Ci (Curios); mCi (miliCurios); μ Ci (microCurios); HPLC (cromatografía líquida de alta resolución); RP-HPLC (cromatografía líquida de alta resolución en fase inversa); TLC (cromatografía el capa fina); MALDI-TOF (desorción/ionización por láser asistida por matriz—tiempo de vuelo); Ts (tosilo); Bn (bencilo); Ph (fenilo); Ms (mesilo); Et (etilo), Me (metilo); *Taq* (ADN polimerasa de *Thermus aquaticus*); Klenow (fragmento grande (Klenow) de la ADN polimerasa I); EGTA (etilenglicol-bis(β -aminoetil eter) del ácido N, N, N', N'-tetraacético); EDTA (ácido etilendiaminotetraacético); bla (β -lactamasa o gen de resistencia a ampicilina); HDL (líquido alta densidad); MJ Research (MJ Research, Reno, NV); Baseclear (Baseclear BV, Inc., Leiden, Países Bajos); PerSeptive (PerSeptive Biosystems, Framingham, MA); ThermoFinnigan (ThermoFinnigan, San José, CA); Argo (Argo BioAnalytica, Morris Plains, NJ); Seitz EKS (SeitzSchenk Filtersystems GmbH, Bad Kreuznach, Alemania); Pall (Pall Corp., East Hills, NY); Spectrum (Spectrum Laboratories, Dominguez Rancho, CA); Molecular Structure (Molecular Structure Corp., Woodlands, TX); Accelrys (Accelrys, Inc., San Diego, CA); Chemical Computing (Chemical Computing Corp., Montreal, Canadá); New Brunswick (New Brunswick Scientific, Co., Edison, NJ); CFT (Center for Test Materials, Vlaardingen, Países Bajos); Procter & Gamble (Procter & Gamble, Inc., Cincinnati, OH); GE Healthcare (GE Healthcare, Chalfont St. Giles, Reino Unido); DNA2.0 (DNA2.0, Menlo Park, CA); OXOID (Oxoid, Basingstoke, Hampshire, Reino Unido); Megazyme (Megazyme International Ireland Ltd., Bray Business Park, Bray, Co., Wicklow, Irlanda); Finnzymes (Finnzymes Oy, Espoo, Finlandia); Kelco (CP Kelco, Wilmington, DE); Coming (Coming Life Sciences, Corning, NY); NEN (NEN Life Science Products, Boston, MA); Pharma AS (Pharma AS, Oslo, Noruega); Dynal (Dynal, Oslo, Noruega); Bio-Synthesis (Bio-Synthesis, Lewisville, TX); ATCC (American Type Culture Collection, Rockville, MD); Gibco/BRL (Gibco/BRL, Grand Island, NY); Sigma (Sigma Chemical Co., St. Louis, MO); Pharmacia (Pharmacia Biotech, Piscataway, NJ); NCBI (National Center for Biotechnology Information); Applied Biosystems (Applied Biosystems, Foster City, CA); BD Biosciences y/o Clontech (BD Biosciences CLONTECH Laboratories, Palo Alto, CA); Operon Technologies (Operon Technologies, Inc., Alameda, CA); MWG Biotech (MWG Biotech, High Point, NC); Oligos Etc (Oligos Etc. Inc, Wilsonville, OR); Bachem (Bachem Bioscience, Inc., King of Prussia, PA); Difco (Difco Laboratories, Detroit, MI); Mediatech (Mediatech, Herndon, VA); Santa Cruz (Santa Cruz Biotechnology, Inc., Santa Cruz, CA); Oxoid (Oxoid Inc., Ogdensburg, NY); Worthington (Worthington Biochemical Corp., Freehold, NJ); GIBCO BRL or Gibco BRL (Life Technologies, Inc., Gaithersburg, MD); Millipore (Millipore, Billerica, MA); Bio-Rad (Bio-Rad, Hercules, CA); Invitrogen (Invitrogen Corp., San Diego, CA); NEB (New England Biolabs, Beverly, MA); Sigma (Sigma Chemical Co., St. Louis, MO); Pierce (Pierce Biotechnology, Rockford, IL); Takara (Takara Bio Inc. Otsu, Japón); Roche (Hoffmann-La Roche, Basilea, Suiza); EM Science (EM Science, Gibbstown, NJ); Qiagen (Qiagen, Inc., Valencia, CA); Biodesign (Biodesign Intl., Saco, Maine); Aptagen (Aptagen, Inc., Herndon, VA); Sorvall (marca Sorvall, de Kendro Laboratory Products, Asheville, NC); Molecular Devices (Molecular Devices, Corp., Sunnyvale, CA); R&D Systems (R&D Systems, Minneapolis, MN); Stratagene (Stratagene Cloning Systems, La Jolla, CA); Marsh (Marsh Biosciences, Rochester, NY); Geneart (Geneart GmbH, Regensburg, Alemania); Bio-Tek (Bio-Tek Instruments, Winooski, VT); (Biacore (Biacore, Inc., Piscataway, NJ); PeproTech (PeproTech, Rocky Hill, NJ); SynPep (SynPep, Dublin, CA); New Objective (marca New Objective; Scientific Instrument Services, Inc., Ringoes, NJ); Waters (Waters, Inc., Milford, MA); Matrix Science (Matrix Science, Boston, MA); Dionex (Dionex, Corp., Sunnyvale, CA); Monsanto (Monsanto Co., St. Louis, MO); Wintershall (Wintershall AG, Cassel, Alemania); BASF (BASF Co., Florham Park, NJ); Huntsman (Huntsman Petrochemical Corp., Salt Lake City, UT); Enichem (Enichem Iberica, Barcelona, España); Fluka Chemie AG (Fluka Chemie AG, Buchs, Suiza); Gist-Brocades (Gist-Brocades, NV, Delft, Países Bajos); Dow Corning (Dow Corning Corp., Midland, MI) y Microsoft (Microsoft, Inc., Redmond, WA).

[0144] La serina proteasa natural usada en los siguientes ejemplos se describe en detalle en el documento WO 2005/052146 (PCT/US2004/039066).

EJEMPLO 1

5

Ensayos

[0145] En los ejemplos siguientes se usaron diversos ensayos, como determinaciones de proteínas, pruebas basadas en la aplicación y pruebas basadas en la estabilidad. Para una interpretación fácil, se muestran a continuación los siguientes ensayos y se hace referencia a los respectivos ejemplos. Todas las desviaciones de los protocolos proporcionados a continuación en cualquiera de los experimentos realizados durante el desarrollo de la presente invención se indican en los ejemplos.

A. Ensayo TCA para la determinación del contenido de proteínas en placas de microvaloración de 96 pocillos

15

[0146] Este ensayo se inició usando sobrenadantes de cultivo filtrados de placas de microvaloración crecidas durante 4 días a 33 °C con agitación a 230 rpm y aireación humidificada. Se usó para el ensayo una placa de 96 pocillos de fondo plano recién preparada. En primer lugar, se colocaron 100 µl/pocillo de HCl 0,25 N en los pocillos. A continuación, se añadieron a los pocillos 50 µl de medio de cultivo filtrado. Después se determinó la dispersión de la luz/absorbancia a 405 nm (usando el modo de mezcla de 5 s en el lector de placas), para proporcionar la lectura "blanco".

[0147] Para la prueba, se colocaron 100 µl/pocillo de TCA al 15% (p/v) en las placas y se incubaron entre 5 y 30 min a temperatura ambiente. Después se determinó la dispersión de la luz/absorbancia a 405 nm (usando el modo de mezcla de 5 s en el lector de placas).

[0148] Los cálculos se realizaron sustrayendo el blanco (es decir, sin TCA) de la lectura de prueba con TCA. Si se desea, puede generarse una curva patrón calibrando las lecturas de TCA con ensayos AAPF de clones con factores de conversión conocidos. Sin embargo, los resultados de TCA son lineales con respecto a la concentración de proteína de 50 a 500 ppm y puede, por tanto, representarse directamente frente al rendimiento de la enzima con el fin de elegir variantes con buen rendimiento.

B. Ensayo suc-AAPF-pNA de proteasas en placas de microvaloración de 96 pocillos

[0149] En este sistema de ensayo, las soluciones reactivas utilizadas fueron:

1. Tris/HCl 100 mM, pH 8,6, que contenía TWEEN®-80 al 0,005% (tampón Tris).
2. Tampón Tris 100 mM, pH 8,6 que contenía CaCl₂ 10 mM y TWEEN®-80 al 0,005% (tampón Tris).
3. suc-AAPF-pNA 160 nM en DMSO (solución madre de suc-AAPF-pNA) (Sigma: S-7388).

40

[0150] Para preparar la solución de trabajo de suc-AAPF-pNA, se añadió 1 ml de la solución patrón de AAPF a 100 ml de tampón Tris y se mezcló bien durante al menos 10 segundos.

[0151] El ensayo se realizó añadiendo 10 µl de solución de proteasa diluida a cada pocillo, seguido por la adición (rápidamente) de 190 µl de solución de trabajo de AAPF a 1 mg/ml. Las soluciones se mezclaron durante 5 s y se midió la variación de la absorbancia a 410 nm en un lector PMV a 25°C. La actividad proteasa se expresó como UA (actividad = $\Delta DO \text{ min}^{-1} \cdot \text{ml}^{-1}$).

45

C. Ensayo de hidrólisis de queratina

50

[0152] En este sistema de ensayo, las soluciones y reactivos químicos utilizados fueron:

Queratina ICN 902111

Detergente	Se disolvieron 1,6 g de detergente en 1.000 ml de agua (pH = 8,2), también se añadieron 0,6 ml de CaCl ₂ /MgCl ₂ de 10.000 gpg así como 1.190 mg de HEPES, proporcionando una dureza y potencia del tampón de 6 gpg y 5 mM, respectivamente. El pH se ajustó a 8,2 con NaOH. Ácido picrilsulfónico (TNBS) Sigma P-2297 (solución al 5% en agua).
Reactivo A	Se disolvieron conjuntamente 45,4 g de Na ₂ B ₄ O ₇ ·10 H ₂ O (Merck 6308) y 15 ml de NaOH 4N en un volumen final de 1.000 ml (calentando si es necesario).
Reactivo B	Se disolvieron conjuntamente 35,2 g de NaH ₂ PO ₄ ·1 H ₂ O (Merck 6346) y 0,6 g de Na ₂ SO ₃ (Merck 6657) en un volumen final de 1.000 ml.

Procedimiento:

- 5 **[0153]** Antes de las incubaciones, la queratina se tamizó a través de un cedazo de 100 µm en pequeñas porciones cada vez. A continuación, se agitaron 10 g de la queratina < 100 µm en solución detergente durante al menos 20 minutos a temperatura ambiente con ajuste regular del pH a 8,2. Finalmente, la suspensión se centrifugó durante 20 minutos a temperatura ambiente (Sorvall, rotor GSA, 13.000 rpm). A continuación se repitió este procedimiento. Finalmente, el sedimento húmedo se resuspendió en detergente en un volumen total de 200 ml y la suspensión se mantuvo en agitación durante el pipeteo. Antes de la incubación, las placas de microvaloración (PMV) se llenaron con 200 µl de sustrato por pocillo con una pipeta multicanal Biohit y punta de 1.200 µl (6 dispensaciones de 200 µl y dispensados lo más rápido posible para evitar la agregación de la queratina en las puntas). A continuación se añadieron 10 µl de cultivo filtrado a las PMV que contenían el sustrato. Las placas se cubrieron con cinta adhesiva se colocaron en un incubador y se incubaron a 20°C durante 3 horas a 350 rpm (Innova 4330 [New Brunswick]).
- 10 Tras la incubación, las placas se centrifugaron durante 3 minutos a 3.000 rpm (Centrífuga de Sigma 6K 15). Aproximadamente a los 15 minutos de retirar la primera placa del incubador se preparó el reactivo TNBS mezclando 1 ml de solución de TNBS por cada 50 ml de reactivo A.

- 15 **[0154]** Las PMV se llenaron con 60 µl de TNBS con reactivo A por pocillo. A partir de las placas incubadas se transfirieron 10 µl a las PMV con TNBS con reactivo A. Las placas se cubrieron con cinta adhesiva y se agitaron durante 20 minutos en un agitador de mesa (BMG Thremostar) a temperatura ambiente y a 500 rpm. Finalmente, se añadieron a los pocillos 200 µl de reactivo B, se mezcló durante 1 minuto en un agitador y se midió la absorbancia a 405 nm con el lector de PMV.

25 Cálculo de la actividad hidrolizante de queratina

- [0155]** El valor de absorbancia obtenido se corrigió con respecto al valor blanco (sustrato sin enzima). La absorbancia resultante proporciona una medida de la actividad hidrolítica. Se calculó el índice de rendimiento para cada muestra (variante). El índice de rendimiento compara el rendimiento de la variante (valor actual) y la enzima estándar (valor teórico) a la misma concentración de proteína. Además, pueden calcularse los valores teóricos usando los parámetros de la ecuación de Langmuir de la enzima estándar. Un índice de rendimiento (IR) que sea mayor de 1 (IR>1) identifica una variante mejor (en comparación con el estándar [p. ej., natural], mientras que un IR de 1 (IR=1) identifica una variante que tiene el mismo rendimiento que el estándar y un IR menor de 1 (IR<1) identifica una variante que tiene peor rendimiento que el estándar. Por tanto, el IR identifica ganadores así como
- 30 variantes cuyo uso es menos deseable en determinadas circunstancias.

D. Ensayo de hidrólisis de dimetilcaseína (96 pocillos)

- 40 **[0156]** En este sistema de ensayo, las soluciones y reactivos químicos utilizados fueron:

Dimetilcaseína (DMC):	Sigma C-9801
TWEEN®-80:	Sigma P-8074
Tampón PIPES (ácido libre)	Sigma P-1851; se disuelven 15,1 g en aproximadamente 960 ml de agua, se ajusta el pH a 7,0 con NaOH 4N, se añade 1 ml de TWEEN®- 80 al 5% y el volumen se lleva a 1.000 ml. La concentración final de PIPES y TWEEN®-80 es de 50 mM y 0.005%, respectivamente.
Ácido picrilsulfónico (TNBS):	Sigma P-2297 (solución al 5% en agua)
Reactivo A:	Se disuelven conjuntamente 45,4 g de Na ₂ B ₄ O ₇ ·10 H ₂ O (Merck 6308) y 15 ml de NaOH 4 N en un volumen final de 1.000 ml (calentando si es necesario)
Reactivo B:	Se disuelven conjuntamente 35,2 g de NaH ₂ PO ₄ ·1 H ₂ O (Merck 6346) y 0,6 g de Na ₂ SO ₃ (Merck 6657) en un volumen final de 1.000 ml.

Procedimiento:

5 [0157] Para preparar el sustrato se disolvieron 4 g de DMC en 400 ml de tampón PIPES. Los sobrenadantes de cultivo filtrados se diluyeron con tampón PIPES; la concentración final de los controles en la placa de crecimiento era de 20 ppm. A continuación, se añadieron 10 µl de cada sobrenadante diluido a 200 µl de sustrato en los pocillos de una PMV. La placa PMV se cubrió con cinta adhesiva, se agitó durante algunos segundos y se colocó en una estufa a 37°C durante 2 horas sin agitación.

10 [0158] Aproximadamente a los 15 minutos de retirar la 1ª placa de la estufa, se preparó el reactivo mezclando 1 ml de solución TNBS con 50 ml de reactivo A. Las PMV se llenaron con 60 µl de TNBS más reactivo A por pocillo. Las placas incubadas se agitaron durante algunos segundos, tras lo cual se transfirieron 10 µl a las PMV con TNBS más reactivo A. Las placas se cubrieron con cinta adhesiva y se agitaron durante 20 minutos en un agitador de mesa (BMG Thremostar) a temperatura ambiente y 500 rpm. Finalmente, se añadieron a los pocillos 200 µl de reactivo B, se mezclaron durante 1 minuto en un agitador y se determinó la absorbancia a 405 nm usando un lector de PMV.

Cálculo de la actividad hidrolizante de dimetilcaseína:

20 [0159] El valor de absorbancia obtenido se corrigió con respecto al valor blanco (sustrato sin enzima). La absorbancia resultante es una medida de la actividad hidrolítica. La actividad específica (arbitraria) de una muestra se calculó dividiendo la absorbancia entre la concentración de proteína determinada.

E. Ensayo de termoestabilidad

25 [0160] Este ensayo se basa en la hidrólisis de dimetilcaseína, antes y después de calentar el sobrenadante de cultivo tamponado. Se usan los mismos reactivos químicos y soluciones reactivas descritas en el ensayo de hidrólisis de dimetilcaseína.

Procedimiento:

30 [0161] Los sobrenadantes de cultivo filtrados se diluyeron hasta 20 ppm en tampón PIPES (en base a la concentración de los controles en las placas de cultivo). A continuación, se colocaron 50 µl de sobrenadante diluido en los pocillos vacíos de una PMV. La placa PMV se incubó en un incubador iEMS/agitador HT (Thermo Labsystems) durante 90 minutos a 60°C y 400 rpm. Las placas se enfriaron en hielo durante 5 minutos. A 35 continuación se añadieron 10 µl de la solución a una PMV recién preparada que contenía 200 µl de sustrato dimetilcaseína/pocillo. Esta PMV se cubrió con cinta adhesiva, se agitó durante algunos segundos y se colocó en una estufa a 37°C durante 2 horas sin agitación. Se usó el mismo procedimiento de detección utilizado para el ensayo de hidrólisis de DMC.

40 **Cálculo de la termoestabilidad:**

[0162] La actividad residual de una muestra se expresó como la relación entre la absorbancia final y la absorbancia inicial ambas corregidas con los blancos.

45 **F. Ensayo de estabilidad en LAS**

[0163] La estabilidad en LAS se midió tras la incubación de la proteasa del ensayo en presencia de LAS (dodecilsulfonato sódico) al 0,06% y la actividad residual se determinó usando el ensayo de AAPF.

50 **Reactivos:**

[0164]

Dodecilsulfonato, sal sódico (=LAS): Sigma D-2525

55 TWEEN®-80: Sigma P-8074

Tampón TRIS (ácido libre) (Sigma T-1378); se disuelven 6,35 g en aproximadamente 960 ml de agua; el pH se ajusta a 8,2 con HCl 4N. La concentración final de TRIS es de 52,5 mM.

Solución madre de LAS: preparar una solución de LAS al 10,5% en agua MQ (=10,5 g por 100 ml de MQ)

Tampón TRIS- 100 mM / pH 8,6 (Tris 100 mM/Tween80 al 0,005%)

60 Tampón TRIS-Ca pH 8,6 (Tris 100 mM/CaCl₂ 10 mM/Tween80 al 0,005%)

Equipo:

[0165]

PMV de fondo redondo: Costar (N° 9017)

Biomek FX

Multipipeta ASYS

Lector de PMV Spectramax

5 Incubador/agitador iEMS

Incubador/agitador Innova 4339

Pipeta multicanal Biohit

Agitador BMG ThermoStar

10 Procedimiento:

[0166] Se preparó una solución de LAS al 0,063% en tampón Tris 52,5 mM, pH 8,2. La solución de trabajo de AAPF se preparó añadiendo 1 ml de solución madre de AAPF a 100 mg/l (en DMSO) a 100 ml de tampón TRIS (100 mM), pH 8,6. Para diluir los sobrenadantes, se llenaron placas de fondo plano con tampón de dilución, se añadió una alícuota del sobrenadante y se mezcló bien. La relación de dilución depende de la concentración de los controles de ASP en las placas de crecimiento (actividad AAPF). La concentración de proteína deseada era de 80 ppm.

[0167] Se añadieron 10 µl del sobrenadante diluido a 190 µl de tampón LAS al 0,063%/pocillo. Las PMV se cubrieron con cinta adhesiva, se agitaron durante algunos segundos y se colocaron en un incubador (Innova 4230) a 25° o 35°C durante 60 minutos con una agitación de 200 rpm. La actividad inicial ($t=10$ minutos) se determinó tras 10 minutos de incubación transfiriendo 10 µl de la mezcla de cada pocillo a una PMV recién preparada que contenía 190 µl de solución de trabajo AAPF. Estas soluciones se mezclaron bien y se determinó la actividad AAPF usando un lector de PMV (20 lecturas en 5 minutos y 25°C).

25

[0168] La actividad final ($t=60$ minutos) se determinó retirando otros 10 µl de solución de la placa en incubación después de 60 minutos de incubación. La actividad AAPF se determinó después como se describió anteriormente. Los cálculos se realizaron como sigue: el % de actividad residual era $[\text{valor } t=60] * 100 / [\text{valor } t=10]$.

30 EJEMPLO 2

Producción de proteasa 69B4 a partir de la bacteria alcalifílica grampositiva 69B4

[0169] En este ejemplo se proporciona una descripción de la cepa 68B4 de *Cellulomonas* utilizada para aislar inicialmente la nueva proteasa 69B4 proporcionada por la presente invención. El microorganismo alcalifílico cepa 69B.4 de *Cellulomonas* (DSM 16035) se aisló a 37°C en un medio que contenía caseína alcalina (g.l^{-1}) (véase, p. ej., Duckworth y col., FEMS Microbiol. Ecol., 19:181-191 [1996]).

Glucosa (Merck 1.08342)	10
Peptona (Difco 0118)	5
Extracto de levadura (Difco 0127)	5
K ₂ HPO ₄	1
MgSO ₄ .7H ₂ O	0,2
NaCl	40
Na ₂ CO ₃	10
Caseína	20
Agar	20

40 **[0170]** También se usó un medio de cultivo alcalino adicional (medio alcalófilo de Grant) para cultivar la cepa 69B.4 de *Cellulomonas*, como se proporciona a continuación:

Solución A del medio alcalófilo de Grant ("GAM") (g.l^{-1})

Glucosa (Merck 1.08342)	10
Peptona (Difco 0118)	5
Extracto de levadura (Difco 0127)	5
K ₂ HPO ₄	1
MgSO ₄ .7H ₂ O	0,2

45

Disueltos en 800 ml de agua destilada y esterilizado mediante autoclave.

Solución B de GAM (g.L⁻¹)

NaCl	40
Na ₂ CO ₃	10

[0171] Disueltos en 200 ml de agua destilada y esterilizado mediante autoclave.

5

[0172] El medio GAM completo se preparó mezclando la solución A (800 ml) con la solución B (200 ml). El medio sólido se preparó mediante la adición de agar (2% p/v).

Condiciones de crecimiento

10

[0173] A partir de un vial de cultivo de glicerol recién descongelado (conservado como glicerol congelado (20% v/v, solución madre conservada a -80°C), los microorganismos se inocularon usando un asa de inoculación en el medio alcalófilo de Grant (GAM) descrito anteriormente en placas de agar y se crecieron durante al menos 2 días a 37°C. A continuación, se usó una colonia para inocular un matraz en agitación de 500 ml que contenía 100 ml de GAM a pH 10. Este matraz se incubó a continuación a 37°C en un agitador rotatorio a 280 rpm durante 1-2 días hasta que se obtuvo un buen crecimiento (según observación visual). A continuación se usaron posteriormente 100 ml de medio de cultivo para inocular un fermentador de 7 l que contenía 5 litros de GAM. Las fermentaciones se desarrollaron a 37°C durante 2-3 días para obtener la producción máxima de proteasa. Las condiciones completamente aeróbicas se mantuvieron durante todo el tiempo inyectando aire a un flujo de 5 l/min, en la región de la bomba que estaba rotando a aproximadamente 500 rpm. El pH se estableció a un valor de 10 al inicio pero no se controló durante la fermentación.

15

20

Preparación de muestras de enzima sin procesar de 69B4

[0174] El medio de cultivo se recogió del fermentador y las células se retiraron mediante centrifugación durante 30 min a 5.000 x g a 10°C. El sobrenadante resultante se aclaró mediante filtración en profundidad sobre Seitz EKS (SeitzSchenk Filtersystems). El sobrenadante de cultivo estéril resultante se concentró además aproximadamente 10 veces mediante ultrafiltración usando un casete de ultrafiltración con un punto de corte de 10 kDa (Minisette de 10 kDa Pall Omega; Pall). Las muestras de 69B4 sin procesar concentradas resultantes se congelaron y conservaron a -20°C hasta su uso posterior.

25

30

Purificación

[0175] El medio de cultivo separado de las células se dializó frente a ácido (2-(4-morfolino)-etano sulfónico ("MES") 20 mM, pH 5,4, CaCl₂ 1mM usando una tripa de diálisis Spectra-Pro7 (Spectrum) con un peso molecular de exclusión (PME) de 8 K. La diálisis se realizó durante la noche o hasta que la conductividad de la muestra fuese menor o igual a la conductividad del tampón MES. La muestra de enzima dializada se purificó usando BioCad VISION (Applied Biosystems) con una columna de intercambio catiónico POROS de sulfo-propilo de alta densidad (HS) 20 (20 micras) de 10 x 100 mm (7,845 ml) (PerSeptive Biosystems). Tras cargar la enzima en la columna previamente equilibrada a 5 ml/min, la columna se lavó a 40 ml/min con un gradiente de pH de MES 25 mM, pH 6,2, CaCl₂ 1 mM a ácido (N-[2-hidroxietil]-piperazina-N'-[2-etano] sulfónico [C₈H₁₈N₂O₄S, N° CAS 7365-45-9] ("HEPES") 25 mM, pH 8,0, CaCl₂ 0,1 mM en 25 volúmenes de columna. Se recogieron fracciones (8 ml) durante el desarrollo de la cromatografía. El paso de lavado a pH 8,0 se mantuvo durante 5 volúmenes de columna y, a continuación, la enzima se eluyó usando un gradiente (0-100 mM de NaCl en el mismo tampón en 35 volúmenes de columna). La actividad proteasa en las fracciones se controla usando el ensayo pNA (ensayo sAAPF-pNA; DelMar, y col., *supra*). La actividad proteasa que eluye a NaCl 40 mM se concentró y se cambió el tampón (usando concentrador VIVA Science de PME a 5K de 20 ml) a MES 20 mM, pH 5,8, CaCl₂ 1 mM. Este material se usó para la caracterización adicional de la enzima.

35

40

45

50 EJEMPLO 3**Producción de proteasas ASP en *B. subtilis***

[0176] Los experimentos realizados para producir la proteasa 69B4 (también denominada en este documento como "ASP", "Asp", "proteasa ASP" y "proteasa Asp") en *B. subtilis* se describen en la solicitud de patente de EE. UU. N° de serie 10/576.331 (WO 2005/052146).

55

[0177] La secuencia de ADN (secuencia de ADN de ASP sintética) se proporciona a continuación con el uso de codones adaptado para las especies de *Bacillus*, y codifica la proteína precursora de ASP natural.

60

ATGACACCACGAACTGTCACAAGAGCTCTGGCTGTGGCAACAGCAGCTGCTACACT
CTTGGCTGGGGGTATGGCAGCACAAGCTAACGAACCGGCTCCTCCAGGATCTGCAT
CAGCCCCCTCCACGATTAGCTGAAAACTTGACCCTGACTTACTTGAAGCAATGGAA
CGCGATCTGGGGTTAGATGCAGAGGAAGCAGCTGCAACGTTAGCTTTTCAGCATGA
CGCAGCTGAAACGGGAGAGGCTCTTGCTGAGGAACTCGACGAAGATTTTCGCGGGC
ACGTGGGTTGAAGATGATGTGCTGTATGTTGCAACCACTGATGAAGATGCTGTTGA
AGAAGTCGAAGGCGAAGGAGCAACTGCTGTGACTGTTGAGCATTCTCTTGCTGATT
TAGAGGCGTGGAAGACGGTTTTGGATGCTGCGCTGGAGGGTCATGATGATGTGCCT
ACGTGGTACGTGACGTGCCTACGAATTCGGTAGTCGTTGCTGTAAAGGCAGGAGC
GCAGGATGTAGCTGCAGGACTTGTGGAAGGCGCTGATGTGCCATCAGATGCGGTCA
CTTTGTAGAAACGGACGAAACGCCTAGAACGATGTTTCGACGTAATTGGAGGCA
ACGCATATACTATTGGCGGCCGGTCTAGATGTTCTATCGGATTCGCAGTAAAC
GGTGGCTTCATTACTGCCGGTCACTGCGGAAGAACAGGAGCCACTACTGCCA
ATCCGACTGGCACATTTGCAGGTAGCTCGTTTCCGGGAAATGATTATGCATTC
GTCCGAACAGGGGCAGGAGTAAATTTGCTTGCCCAAGTCAATAACTACTCGGG
CGGCAGAGTCCAAGTAGCAGGACATACGGCCGCACCAGTTGGATCTGCTGTA
TGCCGCTCAGGTAGCACTACAGGTTGGCATTGCGGAACTATCACGGCGCTGA
ATTCGTCTGTACGTATCCAGAGGGAACAGTCCGAGGACTTATCCGCACGACG
GTTTGTGCCGAACCAGGTGATAGCGGAGGTAGCCTTTTAGCGGGAAATCAAG
CCCAAGGTGTCACGTCAGGTGTTCTGGAAATTGTCGGACGGGGGGAACAAC
ATTCTTTCAACCAGTCAACCCGATTTTGCAGGCTTACGGCCTGAGAATGATTA
CGACTGACTCTGGAAGTTCCCTGCTCCAGCACCTACATCATGTACAGGCTACGC
AAGAACGTTACAGGAACCCTCGCAGCAGGAAGAGCAGCAGCTCAACCGAACGGT
AGCTATGTTCAAGGTCAACCGGAGCGGTACACATTCCGTCTGTCTCAATGGACCTAG
CGGTGCGGACTTTGATTTGTATGTGCAGCGATGGAATGGCAGTAGCTGGGTAACCG
TCGCTCAATCGACATCGCCGGGAAGCAATGAAACCATTACGTACCGCGGAAATGCT
GGATATTATCGCTACGTGGTTAACGCTGCGTCAGGATCAGGAGCTTACACAATGGG
ACTCACCTCCCCTGA (SEQ ID NO:1)

[0178] En la secuencia anterior, el texto en negrita indica el ADN que codifica la proteasa madura, la fuente convencional indica la secuencia líder y el texto subrayado indica las prosequencias N-terminal y C-terminal.

5

Expresión del gen ASP sintético

[0179] La expresión del gen de ASP sintético se describe en la solicitud de patente de EE. UU. Nº de serie 0/576.331 (WO 2005/052146).

10

EJEMPLO 4

Producción de mutantes combinatorios y bibliotecas de mutación múltiples.

[0180] En este ejemplo, se describen los procedimientos utilizados para construir mutantes combinatorios y bibliotecas de mutación múltiple.

Construcción de mutantes combinatorios

5

[0181] La construcción de mutantes combinatorios de ASP se describe en la solicitud de patente de EE. UU. N° de serie 10/576.331 (WO 2005/052146).

Construcción de una biblioteca de mutación múltiple

10

[0182] La biblioteca de mutación múltiple se construyó como se señala en el kit QCMS de Stratagene con la excepción de la concentración del cebador usando en las reacciones. Específicamente, 1 µl del plásmido pUC18-ASP purificado y metilado (aproximadamente 70 ng) se mezcló con 15 µl de agua destilada estéril, 1,5 µl de dNTP, 2,5 µl de tampón 10x, 1 µl de mezcla de la enzima y 1,0 µl de mezcla del cebador mutante (para un total de 100 pmoles de cebadores). La mezcla del cebador se preparó usando 10 µl de cada uno de los dieciocho cebadores mutantes (100 pmoles/µl); añadiendo 50 ng de cada cebador para la biblioteca como recomienda Stratagene, lo que resultaba en pocas mutaciones en una ronda previa de mutagénesis. Por tanto, el protocolo se modificó en la ronda presente de mutagénesis para incluir un total de 100 pmoles de cebadores en cada reacción. Las condiciones de ciclado fueron 95°C durante 1 min, seguido de 20 ciclos de 95°C durante 1 min, 55°C durante 1 min y 65°C durante 12 min en un termociclador PTC2-200 de MJ Research usando tubos de PCR de pared delgada de 0,2 ml. El producto de reacción se digirió con 1 µl de *DpnI* del kit QCMS incubando a 37°C durante toda la noche. Se añadieron 0,5 µl más de *DpnI* y la reacción se incubó durante 1 hora.

[0183] Posteriormente, el ADN de la biblioteca (producto pUC 18-ASP de cadena sencilla sometido a mutagénesis) se electroporó en células de *E. coli* electrocompetentes (Invitrogen, N° de cat., C4040-52, Ont Shot® TOP10 Electrocomp™ *E. coli, dam+*) y se realizó un crecimiento selectivo en placas de agar que contenían 100 mg/l de ampicilina que tuvo como resultado la biblioteca de mutación múltiple de ASP en células de *E. coli*. Se recogieron las colonias (decenas de miles) y se usó el kit de ADN miniprep spin de Qiagen (N° de cat. 27106) para preparar el ADN plásmido mediante las etapas señaladas en el manual del kit miniprep de Qiagen. El ADN miniprep se eluyó con 50 µl de tampón EB de Qiagen proporcionado en el kit.

[0184] El ADN miniprep se digirió usando las enzimas de restricción de ADN *PstI* e *HindIII*. La mezcla de fragmentos de la biblioteca de ASP (*PstI* x *HindIII*) se purificó en gel y se clonó en el fragmento vector pHPLT *HindIII* x *PstI* de 4.154 pares de bases mediante una reacción ligasa usando la ADN ligasa de T4 de Invitrogen (N° de Cat. 15224-025), utilizando el protocolo de Invitrogen según se recomienda para la clonación general de extremos cohesivos. En otra estrategia, los fragmentos de la biblioteca sintética de ASP se produjeron mediante GeneArt. Estos fragmentos de la biblioteca de ASP también se digirieron con *PstI* e *HindIII*, se purificaron y clonaron en el fragmento vector pHPLT *HindIII* x *PstI* de 4.154 pares de bases mediante una reacción ligasa.

[0185] Para transformar la mezcla de reacción de ligamiento directamente en las células de *Bacillus*, el ADN de la biblioteca (mezcla de fragmento de la biblioteca de ASP clonado en pHPLT) se amplificó usando el kit TempliPhi (Amersham, N° de cat. 25-6400). Con este objetivo, se mezcló 1 µl de la mezcla de reacción de ligamiento con 5 µl de tampón de muestra del kit TempliPhi y se calentó durante 3 minutos a 95°C para desnaturalizar el ADN. La reacción se colocó en hielo para su enfriamiento durante 2 minutos y, a continuación, se centrifugó brevemente. Después se añadieron 5 µl de tampón de reacción y 0,2 µl de polimerasa phi29 del kit TempliPhi y las reacciones se incubaron a 30°C en un aparato de PCR MJ Research durante 4 horas. La enzima phi29 se inactivó por calor en las reacciones mediante incubación a 65°C durante 10 min en el aparato de PCR.

[0186] Para la transformación de las bibliotecas en *Bacillus* se mezclaron 0,1 µl del producto de reacción de amplificación TempliPhi con 500 µl de las células de *B. subtilis* competentes ($\Delta aprE$, $\Delta nprE$, *oppA*, $\Delta spoII E$, *degUHy32*, $\Delta amyE::(xylR,pxyA-comK)$) seguido de agitación vigorosa a 37°C durante 1 hora y se dispusieron 100 y 500 µl en placas de HI-agar que contenía sulfato de neomicina a 200 ppm (Sigma, N° de Cat, N-1876; que contiene 732 µg de neomicina por mg) y leche desnatada al 0,5%. Se picaron para su análisis noventa y cinco clones de la biblioteca.

55

[0187] La mutagénesis funcionó bien, ya que sólo el 14% de los clones eran iguales a la secuencia estructural (ASP con R014I-A064K-T086K-T116E-R123F) y aproximadamente el 3% de los clones presentaba mutaciones extras. El resto de los clones secuenciados (72%) eran todos mutantes y, de estos, aproximadamente el 94% eran mutantes exclusivos. Los resultados de secuenciación de la biblioteca se proporcionan a continuación en la Tabla 4-1.

60

G54D					
N24A					
N24Q					
N24T					
N67S					
R127K					
R159F					
R159K					
R159K					
R159N					
R159N					
G78D	R159F				
N24Q	R35E				
N67S	R159E				
R127K	R159E				
R127K	R159K				
R127K	R159N				
R127Q	R159K				
R35D	R159E				
R35D	R159K				
R35E	R159K				
G54D	R127K	R159K			
G78D	R127K	R159K			
G78D	R127K	R159E			
G78D	R127Q	R159K			
N24A	N67A	R159K			
N24A	N67S	R159K			
N24E	R35D	G78D			
N24T	N67S	R159E			
N67L	G78D	R159K			
R35D	G78D	R159K			
N24A	R35E	G78D	R159N		
N24D	R35D	G78D	R159F		
N24E	G54D	G78D	R159K		
N24E	R35D	G78D	R127K	R159N	
N24Q	G54D	G78D	R159N		
N24Q	N67L	G78D	R159E		
N24Q	R35D	R127K	R159K		
N24T	R35D	G78D	R159K		
N24T	R35D	G78D	R159K		
N67S	G78D	R127K	R159K		
R35D	G78D	R127K	R159E		
R35D	G78D	R127K	R159N		
R35D	G78D	R127Q	R159K		
R35E	G54D	N67A	R159F		
R35E	N67S	G78D	R127Q		
N24A	G54D	N67S	G78D	R159F	
N24A	R35D	N67A	G78D	R159F	
N24Q	R35D	N67L	G78D	R159K	
N24Q	R35D	N67L	G78D	R159N	
N24Q	R35D	N67S	R127K	R159E	
N24Q	R35E	N67A	R127K	R159E	
N24Q	R35E	N67A	G78D	R159E	
N24T	N67A	G78D	R127Q	R159N	
N24T	R35E	N67A	G78D	R127Q	
R35E	G54D	N67S	G78D	R159K	
N24A	G54D	N67S	G78D	R127K	R159K
N24A	R35E	N67S	G78D	R127K	R159K
N24E	R35E	G54D	N67S	R127K	R159N

N24Q	R35D	N67S	G78D	R127K	R159F	
N24T	G54D	N67S	G78D	R127Y	R159E	
N24E	R35E	G54D	N67S	G78D	R127K	R159K

EJEMPLO 5

Correlación de mutaciones deletéreas para propiedades múltiples

- 5 **[0188]** En este ejemplo, se ilustra el principio de que las mutaciones deletéreas para cualquier propiedad se correlacionan con mutaciones deletéreas para cualquier otra propiedad, independientemente de las correlaciones entre las propiedades. Como se indica en este documento sólo un número pequeño de posiciones (5-10%) tienen mutaciones que son malas para todas las propiedades. Estas posiciones definen el plegamiento y están conservadas en la evolución. La implicación de esto es que aunque la identificación de mutaciones beneficiosas para cualquier propiedad requiere una selección realmente predictiva para esta propiedad, puede conseguirse la identificación de mutaciones probablemente deletéreas para cualquier propiedad usando cualquier selección, incluyendo pero sin limitaciones los procedimientos proporcionados en este documento.
- 10
- 15 **[0189]** Las enzimas variantes (ASP, ACT y NPR_e) se produjeron como se describe en este documento y dentro de las solicitudes de patente de EE. UU. N° de serie 0/576.331, 10/581.014, 11/581.102 y 11/583.334 (documentos WO 2005/052146, WO 2005/056782 y US 2008/0293610). Las tablas siguientes proporcionan comparaciones por parejas de los valores de variantes con más del 5% de la actividad natural y menos del 5% de actividad para cada una de las dos propiedades, junto con los coeficientes de correlación de las dos propiedades.
- 20 Los sistemas de ensayo utilizados en este ejemplo también se proporcionan en estas solicitudes. Las propiedades utilizadas en este documento fueron actividad caseína (CAS), actividad queratina (QUER), actividad AAPF (AAPF), estabilidad en LAS (LAS) y estabilidad térmica para ASP, así como formación de perácido (PAF) y degradación de perácido (PAD) para ACT.
- 25 **[0190]** Como se indica en las tablas siguientes, las únicas propiedades que se descubrió que estaban correlacionadas (coeficientes de correlación > 0,5) fueron CAS, QUER y AAPF para ASP. Todas las demás no estaban correlacionadas (coeficiente de correlación < 0,3). A pesar del hecho de que las propiedades no se correlacionaban, la probabilidad de que una mutación pudiera ser deletérea para las dos propiedades es mucho más alta que la esperada por probabilidad. En la tabla se proporcionan los cocientes calculados de los valores observados y esperados de las variantes elegidas al azar. Los valores mayores de 1 indican correlaciones positivas y aquellos menores de 1 indican correlaciones negativas.
- 30

Valor	Observado		Esperado		Observado/esperado
	CAS	QUER	CAS	QUER	
<=5%	892	674	31%	24%	
<=5%	1.959	2.177	69%	76%	
ambos >5%	1.877	66%	1.496	52%	1,25
uno >5%	382	13%	1.144	40%	0,33
ambos <=5%	592	21%	211	7%	2,81
al menos uno >5%	2.259	79%	2.640	93%	0,86

Tabla 5-2. Comparación de los resultados de CAS y AAPF para ASP

Valor	Observado		Esperado		Observado/esperado
	CAS	AAPF	CAS	AAPF	
<=5%	892	1.263	31%	44%	
>5%	1.959	1.588	69%	56%	
ambos >5%	1.576	55%	1.091	38%	1,44
uno >5%	395	14%	1.365	48%	0,29
ambos <=5%	880	31%	395	14%	2,23
	Observado		Esperado		
al menos uno >5%	1.971	69%	2.456	86%	0,80

Tabla 5-3. Comparación de los resultados de CAS y LAS para ASP

Valor	Observado		Esperado		Observado/esperado
	CAS	LAS	CAS	LAS	
<=5%	892	1.450	31%	51%	
>5%	1.959	1.401	69%	49%	
ambos >5%	1.393	49%	963	34%	1,45
	574	20%	1.435	50%	0,40
uno >5%	884	31%	45	16%	1,95
ambos <=5%					
al menos uno >5%	196	69%	2.397	84%	0,82

Tabla 5-4. Comparación de los resultados de CAS y estabilidad térmica para ASP

Valor	Observado		Esperado		Observado/esperado
	CAS	TER	CAS	TER	
<=5%	8.921	1.198	31%	42%	
>5%	1.959	1.653	69%	58%	
ambos >5%	1.508	53%	1.136	40%	1,33
uno >5%	596	21%	1.340	47%	0,44
ambos <=5%	747	26%	375	13%	1,99
al menos uno >5%	2.104	74%	2.476	87%	0,85

Tabla 5-5. Comparación de los resultados de QUER y AAPF para ASP

Valor	Observado		Esperado		Observado/esperado
	QUER	AAPF	QUER	AAPF	
<=5%	674	1.263	24%	44%	
>5%	2.177	1.588	76%	56%	
ambos >5%	1.566	55%	1.213	43%	1,29
uno >5%	633	22%	1.340	47%	0,47
ambos <=5%	652	23%	299	10%	2,18
al menos uno >5%	2.199	77%	2.552	90%	0,86

5

Tabla 5-6. Comparación de los resultados de PAF y PAD para ACT

Valor	Observado		Esperado		Observado/esperado
	PAF	PAD	PAF	PAD	

ES 2 388 753 T3

<=5% >5%	541	751		19%	26%	
	2.536	2.326		89%	82%	
ambos>5%	2.187	77%		2.069	73%	1,06
uno >5%	488	17%		639	22%	0,76
ambos <=5%	402	14%		143	5%	282
al menos uno >5%	2.675	94%		2.708	95%	0,99

REIVINDICACIONES

1. Un procedimiento para variantes de ingeniería de proteínas de una proteína parental que combina mutaciones en dos o más sitios de interés; el procedimiento comprende las etapas de:
- 5 a) proporcionar una proteína parental y una biblioteca de evaluación de sitio de variantes de proteína de dicha proteína parental, donde la biblioteca de evaluación de sitio comprende variantes de la proteína parental modificada cada una en uno de los dos o más sitios de interés;
- 10 b) comprobar dicha biblioteca de variantes de proteína y dicha proteína parental para al menos dos propiedades de interés en las respectivas pruebas de interés;
- c) determinar un valor índice de rendimiento para cada propiedad de interés dividiendo el valor obtenido para cada una de las variantes de proteína entre el valor obtenido para dicha proteína parental en la prueba de interés para
- 15 proporcionar una diferencia de energía libre aparente ($\Delta\Delta G_{app}$) para cada una de las variantes de proteína comparado con dicha proteína parental en la prueba de interés.
- d) determinar un valor índice de rendimiento previsto para variantes de proteína que combinan mutaciones a dos o más sitios de interés, en el que el valor índice de rendimiento previsto para las variantes que combinan dos o más
- 20 mutaciones se obtienen añadiendo los valores de $\Delta\Delta G_{app}$ de estas mutaciones.
- e) identificar a partir del valor índice de rendimiento previsto obtenido en la etapa (d) las variantes de proteína que combinan dos o más mutaciones y que poseen una primera propiedad mejorada con respecto a la proteína parental y una segunda propiedad que es al menos el 90% de la de la proteína parental, proporcionando de este modo una
- 25 biblioteca de variantes de proteína enriquecida en miembros que tienen al menos dos de las propiedades de interés.
2. El procedimiento de la reivindicación 1, en el que las propiedades de interés se seleccionan entre carga, rendimiento de lavado, rendimiento de limpieza de superficie dura, estabilidad térmica, estabilidad de almacenamiento, estabilidad en detergente, unión al sustrato, inhibición enzimática, nivel de expresión, velocidad de
- 30 reacción y degradación del sustrato.
3. El procedimiento de la reivindicación 1 o de la reivindicación 2, en el que dicha proteína es una enzima.
- 35 4. El procedimiento de la reivindicación 3, en el que dicha enzima se selecciona entre proteasas, transferasas, metaloproteasas, esterasas, amilasas, celulasas, oxidasas, cutinasas y lipasas.

FIGURA 1.

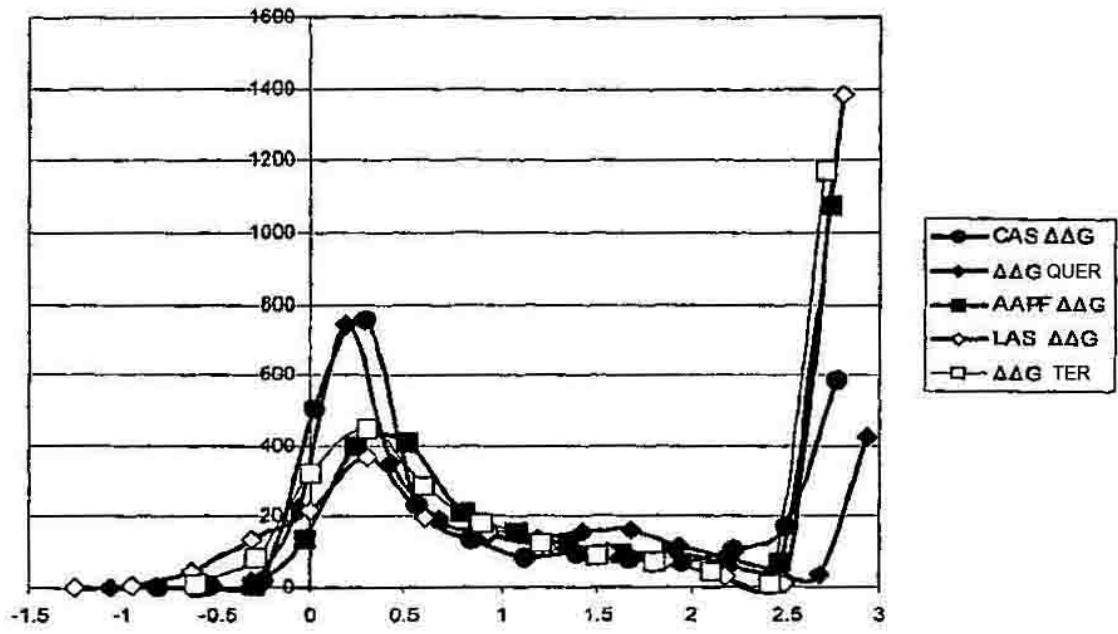


FIGURA 2A.

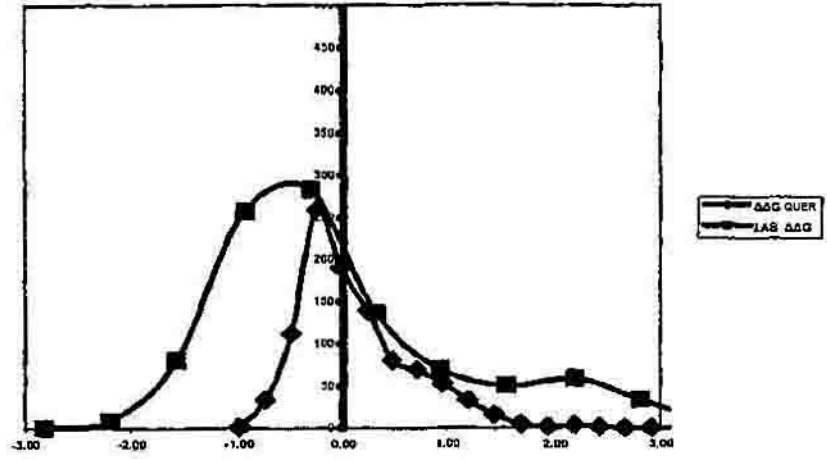


FIGURA 2B.

