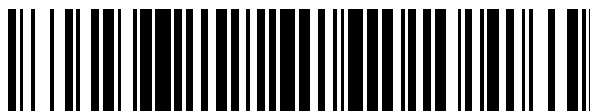


19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 789 299**

51 Int. Cl.:

**C07K 4/06** (2006.01)

**C07K 14/37** (2006.01)

**C07K 14/39** (2006.01)

**C12P 21/02** (2006.01)

**C12N 15/81** (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **29.10.2013 E 16184950 (0)**

97 Fecha y número de publicación de la concesión europea: **12.02.2020 EP 3130598**

54 Título: **Secuencias de expresión**

30 Prioridad:

**29.10.2012 EP 12190361**

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

**26.10.2020**

73 Titular/es:

**LONZA LTD (100.0%)  
Lonzastrasse  
3930 Visp, CH**

72 Inventor/es:

**GASSER, BRIGITTE;  
MATTANOVICH, DIETHARD y  
HEISS, SILVIA**

74 Agente/Representante:

**ELZABURU, S.L.P**

ES 2 789 299 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

Secuencias de expresión

La invención se refiere a elementos reguladores de un sistema de expresión de *Pichia pastoris* (*P. pastoris*), y a su uso en un método para producir una proteína de interés (PDI).

5 **Antecedentes**

La secreción eficaz de proteínas se ha llevado a cabo con hospedadores procarióticos y eucarióticos. Los ejemplos más prominentes son bacterias como *Escherichia coli*, levaduras como *Saccharomyces cerevisiae*, *Pichia pastoris* o *Hansenula polymorpha*, hongos filamentosos como *Aspergillus awamori* o *Trichoderma reesei*, o células de mamífero como, p.ej., las células CHO. Aunque se consigue fácilmente la secreción de algunas proteínas a velocidades elevadas, otras muchas proteínas se secretan solamente a niveles comparativamente bajos.

La expresión heteróloga de un gen en un organismo hospedador requiere un vector que permita la transformación estable del organismo hospedador. Este vector o casete de expresión tiene que proporcionar el gen con un promotor funcional adyacente al extremo 5' de la secuencia codificante. La transcripción se regula y se inicia así mediante esta secuencia promotora.

La ruta secretora comienza en general mediante la translocación de polipéptidos transmembrana y polipéptidos destinados a la secreción al interior del retículo endoplásmico (RE). Para ese fin, estas proteínas poseen una secuencia precursora aminoterminal, también denominada "líder", que comprende o que consiste en un péptido señal y un pro-péptido líder de secreción opcional. El péptido señal consiste en general en 13 a 36 aminoácidos más bien hidrófobos. Los péptidos señal tienen una estructura común: una región amino-terminal corta cargada positivamente (región n); una región hidrófoba central (región h); y una región carboxi-terminal más polar (región c) que contiene el sitio que se escinde mediante la peptidasa señal. En el lado interior del RE el péptido señal se escinde mediante una peptidasa señal. Tras el plegamiento adecuado del polipéptido naciente mediante chaperonas y foldasas residentes en el RE, la proteína se dirige adicionalmente para salir del RE. Este proceso se puede apoyar por la presencia de una pro-secuencia N-terminal, tal como está presente, p.ej., en el precursor del factor de conjugación alfa (MF $\alpha$ ) de *S. cerevisiae*. La proteína se transporta después al aparato de Golgi y finalmente a la membrana plasmática para la secreción al sobrenadante. El pro-péptido líder se escinde de la proteína mediante (presumiblemente) proteasas residentes en el aparato de Golgi tales como la proteasa Kex2 de *S. cerevisiae*.

Debido a que la mayoría de las levaduras no secretan grandes cantidades de proteínas endógenas, y sus proteomas extracelulares no se han caracterizados exhaustivamente hasta ahora, el número de secuencias de secreción disponibles para el uso en levaduras es limitado. Por lo tanto, se empleó la fusión de la proteína objetivo al péptido líder del factor de conjugación alfa (MF $\alpha$ ) de *S. cerevisiae* para controlar la expresión secretora en muchas especies de levaduras (que incluyen *Pichia*, *Kluyveromyces*, *Zygosaccharomyces*). Desafortunadamente, el procesamiento proteolítico del MF $\alpha$  mediante la proteasa Kex2 produce a menudo residuos de aminoácidos N-terminales heterogéneos en el producto.

El documento EP324274B1 describe la expresión y secreción mejoradas de proteínas heterólogas en levadura que emplean secuencias líderes de factor alfa de *S. cerevisiae* truncadas, y el documento EP301669B1 la secuencia líder del factor alfa de *Kluyveromyces* para la secreción de proteínas heterólogas.

De manera alternativa, se usaron los péptidos señal de fosfatasa de *S. cerevisiae* (PHO5, DK3614), sacarosa invertasa de *S. cerevisiae* (SUC, documento WO84/01153), y aspártico proteasa 3 de levadura (YAP3, documento EP792367B1) para la expresión secretora en levaduras. El documento EP0438200 (A1) describe la secuencia del péptido señal de SUC2 de *S. cerevisiae* para la expresión en *P. pastoris*.

El documento US5268273 describe una secuencia señal de fosfatasa ácida de *P. pastoris* (PHO1), en la mayoría de los casos más débil que MF $\alpha$ .

El documento US7741075 describe un péptido señal de secreción de PIR1 de *P. pastoris* para la expresión de proteínas recombinantes y péptidos del dominio de anclaje de PIR1 y PIR2 de *Pichia pastoris* para la expresión superficial recombinante.

Khasa et al. 2011 (Yeast. 28(3):213-26) describe el aislamiento de los genes PIR de *Pichia pastoris* y su utilización para la expresión en la superficie celular y la secreción de proteínas recombinantes, en particular la secreción de proteínas recombinantes en *P. pastoris*, mediante la utilización de la pre-pro señal de la proteína PpPir1p, sin una comparación con MF $\alpha$ .

El documento WO2011073367A1 y Kottmeier et al. 2011 (Applied Microbiology and Biotechnology. 91:1, 133-141) describe una secuencia señal de hidrofobina que actúa como mediador en la secreción eficaz de proteínas recombinantes en *Pichia pastoris*, en particular el uso de la pre-secuencia o pro-secuencia de hidrofobina de *Trichoderma reesei* para la secreción de eGFP en *P. pastoris*.

En el curso de la secuenciación del genoma de *P. pastoris*, se enumeraron 54 secuencias diferentes como péptidos señal predichos que incluían un sitio de escisión para posibilitar la secreción de proteínas. (De Schutter et al. Nature Biotechnology doi: 10.1038/nbt.1544 2009, y figuras complementarias).

5 El documento US2011/0021378A1 describe un grupo de 54 genes de *P. pastoris* identificados que contienen una secuencia señal, entre ellos los residuos 1-21 de SEQ ID 8 como se enumeran en el documento US2011/0021378A1, que incluyen de nuevo un sitio de escisión para posibilitar la secreción de proteínas.

10 El documento EP2258855A1 describe secuencias reguladoras de un sistema de expresión derivado de *P. pastoris*, en el que se usa una secuencia señal y líder de la proteína Epx1 de *P. pastoris* como secuencia precursora de 57 aminoácidos para facilitar la expresión y secreción de una proteína de interés PDI. Se predice que hay un sitio de escisión para una peptidasa señal tras la secuencia señal, es decir, entre la posición 20 y 21, indicado como un guion en la secuencia del sitio de escisión: VSA-AP (SEQ ID 6).

El documento WO2012/152823A1 describe secuencias señal de secreción de *P. Pastoris* y la expresión de dominios variables individuales de inmunoglobulina.

15 Es deseable proporcionar elementos reguladores alternativos adecuados para expresar una PDI en una célula hospedadora eucariótica recombinante, y métodos para producir proteínas secretadas en células eucarióticas que sean simples y eficaces, y que puedan conducir preferiblemente a un extremo N-terminal homogéneo de la PDI.

### Compendio de la invención

El objetivo se resuelve mediante la materia tal como se reivindica.

20 La invención proporciona un ácido nucleico aislado que codifica un líder, unido de forma operable a una secuencia de ácido nucleico que codifica una proteína de interés (POI), en donde el ácido nucleico que codifica el líder se condensa al extremo N-terminal del POI, en el que el líder se selecciona del grupo que consiste en

a) un péptido señal que consiste en la secuencia de 20 aminoácidos de SEQ ID 1 o una variante funcional del mismo con una o dos mutaciones puntuales, y

25 b) un péptido señal que consisten en la secuencia de 20 aminoácidos seleccionada del grupo que consiste en SEQ ID 2, 3, 4 y 5.

en donde está excluido el péptido señal caracterizado por una secuencia de 21 aminoácidos que incluye una extensión C-terminal mediante una alanina adicional, y;

en donde el ácido nucleico no codifica la secuencia de aminoácidos identificada como SEQ ID 8, y

30 en donde está excluido el péptido señal que consiste en la secuencia de aminoácidos SEQ ID 2 fusionada con un polipéptido que consiste en un dominio variable individual de inmunoglobulina.

Adicionalmente se describe en la presente memoria:

a) un péptido líder con la secuencia de aminoácidos de SEQ ID 10 o una variante funcional del mismo con una o dos mutaciones puntuales,

b) un péptido líder con la secuencia de aminoácidos seleccionada del grupo que consiste en SEQ ID 11, 12, 13 y 14.

35 El péptido señal específico se caracteriza específicamente por una secuencia señal de 20 aa, que excluye específicamente una secuencia prolongada de manera C-terminal, p.ej. una secuencia de 21 aa que incluye una prolongación C-terminal de un aminoácido adicional, tal como Alanina.

SEQ ID 1: MKXSTNLILAIAAASXVVSA, en la que

X en la posición 3 es F o L

40 X en la posición 16 es A o T

Por ejemplo, el péptido señal de 20 aminoácidos (aa), EpxL-A, tiene la secuencia de aminoácidos de SEQ ID 1, en la que X en la posición 3 es L, y X en la posición 16 es A (SEQ ID 2).

Según un ejemplo específico, el péptido señal de 20 aminoácidos (aa), EpxL-A, tiene la secuencia de aminoácidos de SEQ ID 1, en la que X en la posición 3 es F, y X en la posición 16 es A (SEQ ID 3).

45 Específicamente, la invención proporciona un péptido señal o un ácido nucleico que codifica el péptido señal que tiene una secuencia de aminoácidos seleccionada del grupo que consiste en SEQ ID 2, 3, 4 y 5.

SEQ ID 2: MKLSTNLILAIAAASAVVSA

SEQ ID 3: MKFSTNLILAIAAASAVVSA

SEQ ID 4: MKFSTNLILAIAAASTVVSA

SEQ ID 5: MKLSTNLILAIAAASTVVSA

Las sustituciones de aminoácidos en la posición 3 y 16 se indican en negrita y subrayadas.

- 5 La molécula de ácido nucleico que codifica el péptido señal también se denomina en la presente memoria "secuencia señal".

En la presente memoria se describe adicionalmente un líder que tiene la secuencia de aminoácidos de SEQ ID 10.

SEQ ID 10: MKXSTNLILAIAAASXWSAAPVAPAE~~E~~ANHLHKR, en la que

X en la posición 3 es F o L

- 10 X en la posición 16 es A o T

Específicamente, en la presente memoria se describe un líder o un ácido nucleico que codifica el líder que tiene una secuencia de aminoácidos seleccionada del grupo que consiste en SEQ ID 11, 12, 13 y 14.

SEQ ID 11: MKLSTNLILAIAAASAVVSAAPVAPAE~~E~~ANHLHKR

SEQ ID 12: MKFSTNLILAIAAASAVVSAAPVAPAE~~E~~ANHLHKR

- 15 SEQ ID 13: MKFSTNLILAIAAASTVVSAAPVAPAE~~E~~ANHLHKR

SEQ ID 14: MKLSTNLILAIAAASTVVSAAPVAPAE~~E~~ANHLHKR

Las sustituciones de aminoácidos en la posición 3 y 16 se indican en negrita y subrayadas.

Por ejemplo, el líder truncado de 36 aminoácido (aa), EpxL-KR, tiene la secuencia de aminoácidos de SEQ ID 10, en la que X en la posición 3 es L, y X en la posición 16 es A.

- 20 Según un ejemplo específico, el péptido líder de 36 aminoácidos (aa), EpxL-KR, tiene la secuencia de aminoácidos de SEQ ID 10, en la que X en la posición 3 es F, y X en la posición 16 es A (SEQ ID 12).

El líder descrito en la presente memoria con la secuencia de aminoácidos de SEQ ID 10, 11, 12, 13 o 14, o una variante funcional del mismo con una o dos mutaciones puntuales, se denomina en la presente memoria "líder truncado".

- 25 Específicamente, el ácido nucleico que codifica el líder truncado comprende o consiste en un ácido nucleico que codifica un péptido señal seleccionado del grupo que consiste en SEQ ID 2, 3, 4 y 5, preferiblemente que excluye SEQ ID 2. Específicamente, el ácido nucleico que codifica el líder truncado comprende o consiste en un ácido nucleico que codifica un péptido líder seleccionado del grupo que consiste en SEQ ID 11, 12, 13, y 14.

- 30 Específicamente, el líder truncado comprende o consiste en el péptido señal seleccionado del grupo que consiste en SEQ ID 2, 3, 4 y 5, preferiblemente que excluye SEQ ID 2.

Un ácido nucleico específicamente preferido codifica un péptido señal de la secuencia de aminoácidos de SEQ ID 3; preferiblemente, la secuencia del ácido nucleico que codifica el péptido señal es SEQ ID 16.

- 35 Por lo tanto, una realización específica de la invención se refiere a un líder o un ácido nucleico que codifica tal líder, que es un péptido señal con la secuencia de aminoácidos de SEQ ID 3, preferiblemente en el que el ácido nucleico codificante consiste en una secuencia de nucleótidos de SEQ ID 16, o una variante con optimización de codones de SEQ ID 16.

Específicamente, la secuencia de nucleótidos que codifica el líder se selecciona del grupo que consiste en SEQ ID 15, 16 y 17 o una variante optimizada codón de cualquiera de las SEQ ID 15, 16 o 17 para coincidir con el patrón de uso de codones en el organismo usado para la expresión.

- 40 para coincidir con el patrón de uso de codones en el organismo usado para la expresión

Específicamente, el líder es un péptido señal que consiste en la secuencia de aminoácidos de SEQ ID 3, preferiblemente en donde el ácido nucleico consiste en una secuencia de nucleótidos de SEQ ID 16 o una variante con optimización de codones de SEQ ID 16 para coincidir con el patrón de uso de codones en el organismo usado para la expresión

- 45 Un ácido nucleico específico descrito en la presente memoria codifica un líder, que es un péptido líder o líder

truncado con la secuencia de aminoácidos de SEQ ID 12, preferiblemente en el que el ácido nucleico consiste en una secuencia de nucleótidos de SEQ ID 19, o una variante con optimización de codones de SEQ ID 19.

SEQ ID 15: secuencia de nucleótidos obtenida de *P. pastoris*, cepa CBS7435 (Centro de Biodiversidad Fúngica CBS-KNAW, Centraalbureau voor Schimmelcultures, Utrecht, Países Bajos):

5 ATGAAGCTCTCCACCAATTTGATTCTAGCTATTGCAGCAGCTTCCGCCGTTGTCTCAGCT

SEQ ID 16: secuencia de nucleótidos tal como se usa en el Ejemplo 9, obtenida de *P. pastoris*, cepa CBS7435, mediante amplificación por PCR con el uso de los cebadores descritos en el Ejemplo 9.

ATGAAGTTCTCTACCAATTTGATTCTAGCTATTGCAGCAGCTTCCGCCGTTGTCTCAGCT

10 SEQ ID 17: secuencia de nucleótidos obtenida de *P. pastoris*, cepa DSMZ70382 (Colección Alemana de Microorganismos y Cultivos Celulares)

ATGAAGTTCTCTACCAATTTGATTCTAGCTATTGCAGCAGCATCCACTGTTGTCTCAGCT

Los nucleótidos que difieren en las secuencias anteriores están subrayados.

SEQ ID 18: secuencia de nucleótidos obtenida de *P. pastoris*, cepa CBS7435 (Centro de Biodiversidad Fúngica CBS-KNAW, Centraalbureau voor Schimmelcultures, Utrecht, Países Bajos)

15 ATGAAGCTCTCCACCAATTTGATTCTAGCTATTGCAGCAGCTTCCGCCGTTGTCTC  
AGCTGCTCCAGTTGCTCCAGCCGAAGAGGCAGCAAACCACTTGCACAAGCGT

SEQ ID 19: secuencia de nucleótidos tal como se usa en el Ejemplo 5, obtenida de *P. pastoris*, cepa CBS7435, mediante amplificación por PCR con el uso de los cebadores descritos en los Ejemplos 1 y 5.

20 ATGAAGTTCTCTACCAATTTGATTCTAGCTATTGCAGCAGCTTCCGCCGTTGTCTC  
AGCTGCTCCAGTTGCTCCAGCCGAAGAGGCAGCAAACCACTTGCACAAGCGT

SEQ ID 20: secuencia de nucleótidos obtenida de *P. pastoris*, cepa DSMZ70382 (Colección Alemana de Microorganismos y Cultivos Celulares)

ATGAAGTTCTCTACCAATTTGATTCTAGCTATTGCAGCAGCATCCACTGTTGTCTC  
AGCTGCTCCAGTTGCTCCAGCCGAAGAGGCAGCAAACCACTTGCACAAGCGT

25 Específicamente, el ácido nucleico descrito en la presente memoria es un ADN.

Específicamente, el líder descrito en la presente memoria es un polipéptido.

Según un determinado aspecto, en la presente memoria se describe un líder aislado, líder truncado, péptido señal o péptido líder, y preferiblemente el líder tiene una secuencia de aminoácidos seleccionada del grupo que consiste en SEQ ID 1, 2, 3, 4, 5.

30 Un líder preferido específicamente es un péptido que consiste en la secuencia de aminoácidos de SEQ ID 3.

Según otro aspecto, la invención proporciona además un casete de expresión que comprende el ácido nucleico descrito en la presente memoria, específicamente el ácido nucleico que codifica el líder unido de forma operable a la secuencia de ácido nucleico que codifica la PDI.

35 Específicamente se excluye la fusión del péptido señal que consiste en la secuencia de aminoácidos SEQ ID 2 con un polipéptido que consiste en un dominio variable individual de una inmunoglobulina, tal como por ejemplo un nanocuerpo.

40 Además, según los ejemplos específicos, la invención proporciona la fusión del péptido señal con la secuencia de aminoácidos seleccionada del grupo que consiste en SEQ ID 2, 3, 4 y 5 con una PDI seleccionada del grupo que consiste en factores de crecimiento, hormonas, citocinas, anticuerpos y fragmentos de anticuerpos, en particular en el que se selecciona un anticuerpo o fragmento de anticuerpo del grupo que consiste en un scFv, minicuerpo, diacuerpo, triacuerpo, tetracuerpo, Fab, proteína de fusión con Fc y un anticuerpo de longitud completa tal como, por ejemplo, IgG, IgA, IgD, IgM o Ig, preferiblemente un anticuerpo de longitud completa, un scFv o un Fab, que incluye específicamente la fusión del péptido señal que consiste en la secuencia de aminoácidos SEQ ID 2 con cualquiera de tales anticuerpos o fragmentos de anticuerpos, específicamente cualquiera de un anticuerpo de longitud completa, un scFv o un Fab.

45 Una secuencia de nucleótidos preferida específicamente que codifica la señal tal como se usa en el casete de expresión consiste en una secuencia de nucleótidos que codifica la secuencia de aminoácidos de SEQ ID 3, preferiblemente consiste en SEQ ID 16, o una variante con optimización de codones de SEQ ID 16. Tal péptido señal se fusiona preferiblemente a cualquier PDI, que incluye cualquiera de los anticuerpos o fragmentos de

anticuerpos, que incluye específicamente la fusión del péptido señal o péptido líder que consiste en la secuencia de aminoácidos SEQ ID 3 con un polipéptido que comprende o que consiste en uno o más dominios variables individuales de inmunoglobulinas, tales como por ejemplo un nanocuerpo.

5 Según otro aspecto, la invención proporciona además la casete de expresión que comprende una secuencia de nucleótidos que codifica un péptido señal, seleccionada del grupo que consiste en SEQ ID 15, 16 o 17 o una variante con optimización de codones de cualquiera de SEQ ID 15, 16 o 17 para coincidir con el patrón de uso de codón en el organismo usado para la expresión, preferiblemente en el que se excluye la construcción de fusión de la secuencia de nucleótidos de SEQ ID 15 fusionada a una secuencia de nucleótidos que codifica uno o más dominios variables individuales de inmunoglobulinas, tales como por ejemplo un nanocuerpo.

10 Incluso, según un ejemplo específico, la invención proporciona el constructo de fusión de una secuencia de nucleótidos de SEQ ID 15 fusionada a una secuencia de nucleótidos que codifica un anticuerpo o fragmento de anticuerpo seleccionado del grupo que consiste en un scFv, minicuerpo, diacuerpo, triacuerpo, tetracuerpo, proteína de fusión Fab, Fc y un anticuerpo de longitud completa tal como por ejemplo IgG, IgA, IgD, IgM o IgE, preferiblemente un anticuerpo de longitud completa, un scFv o un Fab.

15 Una secuencia de nucleótidos preferida específicamente es SEQ ID 16, que codifica el péptido líder de SEQ ID 3.

**Por lo tanto, según los ejemplos específicos, la invención proporciona la construcción de fusión de la secuencia de** nucleótidos SEQ ID 16 fusionada a una secuencia de nucleótidos que codifica cualquier proteína de interés, p.ej. específicamente una PDI como se describe en la presente memoria, que incluye, pero sin limitación, cualquiera de los anticuerpos o fragmentos de anticuerpos, que incluye específicamente las construcciones de fusión en las que la PDI es un polipéptido que comprende o que consiste en uno o más dominios variables individuales de inmunoglobulinas, tales como por ejemplo un nanocuerpo.

20

Mediante el casete de expresión según la invención, por primera vez fue posible proporcionar la expresión de una PDI, y específicamente una PDI madura secretada, con un residuo de aminoácido N-terminal nativo correcto, en particular sin una Alanina adicional en el extremo N-terminal.

25 Específicamente, la PDI se selecciona de proteínas terapéuticas, que incluyen anticuerpos o fragmentos de los mismos, enzimas y péptidos, antibióticos proteicos, proteínas de fusión con toxinas, conjugados carbohidrato - proteína, proteínas estructurales, proteínas reguladoras, vacunas y proteínas o partículas similares a vacunas, enzimas de procesamiento, factores de crecimiento, hormonas y citocinas, o en la que dicha PDI actúa como mediador en la producción de un metabolito de células hospedadoras, preferiblemente seleccionado del grupo que consiste en anticuerpos o fragmentos de los mismos, factores de crecimiento, hormonas y citocinas.

30

Según los ejemplos específicos, el anticuerpo o los fragmentos del mismo se seleccionan del grupo que consiste en un scFv, minicuerpo, diacuerpo, triacuerpo, tetracuerpo, Fab, proteína de fusión con Fc y un anticuerpo de longitud completa tal como por ejemplo IgG, IgA, IgD, IgM o IgE, preferiblemente un anticuerpo de longitud completa, un scFv o un Fab.

35 Específicamente, el casete de expresión según la invención es una fusión de un ácido nucleico según la invención con un ácido nucleico que codifica una PDI, y como tal es un ácido nucleico que no se da de manera natural. La invención proporciona además una proteína de fusión de un líder según la invención con una PDI, y como tal es una proteína de fusión que no se da de manera natural.

40 Por lo tanto, el líder se modifica para la producción mejorada de una PDI, p.ej. para un rendimiento de secreción incrementado, y una calidad mejorada, tal como el extremo N-terminal correcto, y emplea un líder diferente de los líderes de la técnica anterior.

El casete de expresión según la invención codifica específicamente una secuencia líder que consiste en el péptido señal. Además se describe en la presente memoria el péptido señal prolongado mediante la pro-secuencia que consiste en APVAPAEAAANHLHKKR (SEQ ID 7), cuya pro-secuencia es parte del líder de longitud completa nativo de la proteína Epx1 de *P. pastoris*, es decir, una secuencia de 16 aminoácidos idénticos a los aminoácidos 21-36 de la secuencia de 57 aminoácidos de

45

SEQ ID 8: MKFSTNLILAIAAASTVVSAAAPVAPAEAAANHLHKKRAYTDTTKTHTFTEVVTVYRT.

Ambos líderes, el péptido señal y el líder truncado descritos en la presente memoria resultaron tener sorprendentemente propiedades mejoradas inesperadas respecto de los líderes de la técnica anterior, p.ej.

50 SEQ ID 21: MKLSTNLILAIAAASAVVSAA,

también denominado en la presente memoria EpxL-AA (21 aminoácidos) según el documento US2011/0021378A1.

Además, se pudieron mostrar propiedades mejoradas en comparación con el líder de longitud completa de la proteína Epx1 de *P. pastoris* como se describió en el documento EP2258855A1, que es la secuencia de 57 aminoácidos de SEQ ID 8 o una variante de la misma con una Leucina en la posición 3.

Esto fue lo más sorprendente, porque el péptido señal de 20 aminoácidos sin una pro-secuencia o el líder truncado tiene una longitud de solamente un 35% y 63% del líder de longitud completa, respectivamente.

5 El líder de la invención tendría ventajas específicas al fusionarlo con una PDI, tal como una secreción incrementada de la PDI y/o una calidad mejorada, tal como el extremo N-terminal correcto en comparación con la fusión con el líder del factor de conjugación alfa ( $\alpha$ MF), específicamente cuando la PDI es una hormona, una citocina, un anticuerpo o fragmento de anticuerpo, p.ej. seleccionado del grupo que consiste en un scFv, minicuerpo, diacuerpo, triacuerpo, tetracuerpo, Fab, proteína de fusión con Fc y un anticuerpo de longitud completa tal como por ejemplo IgG, IgA, IgD, IgM o IgE, preferiblemente un anticuerpo de longitud completa, un scFv o un Fab.

10 El casete de expresión según la invención excluye explícitamente un ácido nucleico que codifica el líder nativo de la proteína Epx1 de *P. pastoris* como se describió en el documento EP2258855A1 (SEQ ID 8).

El casete de expresión según la invención excluye explícitamente además un ácido nucleico que codifica el líder con los aminoácidos 1-21 del líder de longitud completa de la proteína Epx1 de *P. pastoris*, es decir SEQ ID 11, como se describió en el documento US2011/0021378A1.

15 Según un aspecto específico de la invención, el casete de expresión se incorpora opcionalmente en un constructo de expresión, tal como un vector.

Según un aspecto específico adicional de la invención, el casete de expresión o la construcción de expresión comprende un promotor unido de forma operable al ácido nucleico que codifica el líder.

20 Según una realización específica, la invención proporciona una célula hospedadora de levadura recombinante que comprende un casete de expresión según la invención, específicamente una línea celular hospedadora de levadura, de manera más específica una línea celular de producción. Según un aspecto específico de la invención, el casete de expresión es un vector, o parte de un vector.

Preferiblemente, la levadura se selecciona del grupo de géneros que consiste en *Pichia*, *Candida*, *Torulopsis*, *Arxula*, *Hansenula*, *Ogatea*, *Yarrowia*, *Kluyveromyces*, *Saccharomyces*, *Komagataella*, preferiblemente una levadura metilotrófica, y se prefiere específicamente *P. pastoris*, *Komagataella pastoris*, *K. phaffii*, o *K. pseudopastoris*.

25 Según otra realización específica, la invención proporciona un método para producir una PDI en una célula hospedadora de levadura, que comprende:

- proporcionar una célula hospedadora de la invención,
- cultivar dicha célula hospedadora para expresar dicha PDI, y
- purificar la PDI para obtener una preparación de una PDI purificada.

30 Según la invención, el casete de expresión se emplea preferiblemente para facilitar la secreción de genes recombinantes en células hospedadoras de levadura, y de ese modo se incrementa el rendimiento de los productos secretados.

35 Por lo tanto, el casete de expresión según la invención proporciona la expresión y secreción eficaces de una PDI en una célula hospedadora transformada con dicho casete de expresión. A este respecto, el casete de expresión según la invención se entiende como un casete de expresión en levadura.

40 Específicamente, la PDI es una proteína recombinante, cuyo término en la presente memoria siempre se entiende que incluye polipéptidos y proteínas, tales como los producidos por la célula hospedadora recombinante. La PDI puede ser una proteína heteróloga, tal como heteróloga respecto de la levadura, p.ej., una proteína derivada de un eucariota superior tal como un ser humano o de una especie de levadura distinta de *P. pastoris*, o un polipéptido o proteína artificial. De manera alternativa, la PDI puede ser una proteína homóloga derivada de una levadura tal como p.ej. *P. pastoris*, que no se expresaría, sin embargo, o no se expresa en cantidades deseables en una levadura nativa no transformada con el vector, pero que se expresaría en cantidades deseables, p.ej. se sobreexpresaría para proporcionar cantidades significativas de la PDI o los metabolitos mediante la célula hospedadora de levadura recombinante según la invención.

45 Específicamente, la PDI comprende una secuencia de aminoácidos con una secuencia de aminoácidos N-terminal nativa. Dicha PDI comprende preferiblemente una secuencia de aminoácidos que no comprende una Alanina adicional como residuo de aminoácido N-terminal. Específicamente, la PDI no tiene residuos de aminoácidos N-terminales adicionales provenientes de la secuencia líder. Esto tiene una importancia particular para la calidad de la producción de proteínas recombinantes y la facilidad de la producción.

50 Específicamente, dicha PDI es un polipéptido o proteína secretada, que incluye moléculas extracelulares solubles o moléculas unidas a la membrana.

Específicamente, dicha PDI se selecciona de proteínas terapéuticas, que incluyen anticuerpos o fragmentos de los

5 mismos, enzimas y péptidos, antibióticos proteicos, proteínas de fusión con toxinas, conjugados carbohidrato - proteína, proteínas estructurales, proteínas reguladoras, vacunas y proteínas o partículas similares a vacunas, enzimas de procesamiento, factores de crecimiento, hormonas y citocinas, o en la que dicha PDI actúa como mediador en la producción de un metabolito de células hospedadoras. La PDI también puede ser un producto de expresión que actúa como mediador en la producción de un metabolito de células hospedadoras.

10 Específicamente, la PDI se selecciona del grupo que consiste en factores de crecimiento, hormonas, citocinas, anticuerpos y fragmentos de anticuerpos, preferiblemente seleccionados del grupo que consiste en un anticuerpo de longitud completa, tal como por ejemplo IgG, IgA, IgD, IgM o IgE, un scFv, minicuerpo, diacuerpo, triacuerpo, tetracuerpo, Fab y una proteína de fusión con Fc, preferiblemente un anticuerpo de longitud completa, un scFv o un Fab.

Específicamente, la PDI comprende un residuo de aminoácido N-terminal distinto de Alanina. Por lo tanto, no hay un sitio de escisión de peptidasa señal de VSA-AP (SEQ ID 6) como se predice en la técnica anterior (documento EP2258855A1).

15 Específicamente, la PDI puede ser una proteína secretada, tal como una proteína madura que podría ser una forma activa de una proteína o una pro-forma.

El método según la invención proporciona preferiblemente el cultivo de la célula hospedadora en un cultivo celular, y dicha PDI o metabolito se obtiene, p.ej., a medida que se secreta la PDI, que incluye proteínas o metabolitos unidos a la membrana o solubles o extracelulares, que se purifican opcionalmente del sobrenadante del cultivo celular.

20 Según un aspecto adicional, la invención proporciona el uso del ácido nucleico de la invención, en particular el ácido nucleico que codifica un péptido señal de la invención, para la secreción de una PDI de una célula hospedadora y/o para incrementar la secreción de una PDI de una célula hospedadora, preferiblemente, en el que al menos un 60, 65, 70, 75, 80, 85, 90, 95, 98, o 100% de la PDI secretada comprende una secuencia de aminoácidos N-terminal nativa.

### Figuras

25 Figura 1: secuencia promotora pG1: SEQ ID 9.

Figura 2.1: SDS-PAGE teñida con plata de los sobrenadantes reducidos de *P. pastoris* que secreta pTRP con EpxL-RT o MF $\alpha$ , respectivamente.

Figura 2.2: SDS-PAGE teñida con plata de los sobrenadantes reducidos de *P. pastoris* que secreta eGFP con EpxL-RT o MF $\alpha$ .

30 Figura 2.3: Transferencia de Western anti-HSA de los sobrenadantes reducidos de *P. pastoris* que secreta HSA con EpxL-RT o MF $\alpha$ , respectivamente.

Figura 3: Transferencia de Western teñida con Coomassie de los sobrenadantes de *P. pastoris* que secreta HSA con EpxL-RT para la secuenciación N-terminal.

35 Figura 4.1: SDS-PAGE teñida con plata de los sobrenadantes reducidos de *P. pastoris* que secreta pTRP con EpxL-RT, EpxL-KR o MF $\alpha$ , respectivamente.

Figura 4.2: SDS-PAGE teñida con plata de los sobrenadantes reducidos de *P. pastoris* que secreta eGFP con EpxL-KR o MF $\alpha$ , respectivamente.

40 Figura 4.3: Sobrenadantes de *P. pastoris* que secreta la cadena pesada de HyHEL o la cadena ligera de HyHEL con EpxL-KR, respectivamente. HC: Transferencia de Western mediante el uso del anticuerpo anti-cadena gamma de IgG; LC: SDS-PAGE y tinción de plata.

Figura 5.1: SDS-PAGE teñida con plata de los sobrenadantes reducidos de *P. pastoris* que secreta eGFP con EpxL-KR, EpxL-AA o EpxL-A, respectivamente.

Figura 5.2: SDS-PAGE teñida con plata de los sobrenadantes reducidos de *P. pastoris* que secreta LC con EpxL-KR, EpxL-AA o EpxL-A, respectivamente.

45 Figura 5.3: Transferencia de Western de los sobrenadantes no reducidos de *P. pastoris* que secreta el Fab de HyHEL bajo control de pG1 con EpxL-A o MF $\alpha$ , respectivamente.

50 Figura 6: Transferencia de Western de los sobrenadantes de *P. pastoris* que secreta la hormona del crecimiento humana (HGH), somatotropina humana, interferón alfa2a, los 3 fragmentos de anticuerpos diferentes Fab1, Fab2, Fab3, y los 2 fragmentos de anticuerpos Fv cadena simple diferentes scFv1 y scFv2, respectivamente, bajo control de pG1 con EpxL-A.



### Descripción detallada de la invención

Los términos específicos usados a lo largo de la memoria descriptiva tienen el significado siguiente.

La expresión "línea celular", tal como se usa en la presente memoria, se refiere a un clon establecido de un tipo de célula particular que ha adquirido la capacidad de proliferar a lo largo de un periodo de tiempo prolongado. La expresión "línea celular hospedadora" se refiere a una línea celular tal como se usa para expresar un gen o productos endógenos o recombinantes de una ruta metabólica para producir polipéptidos o los metabolitos celulares mediados por tales polipéptidos. Una "línea celular hospedadora de producción" o "línea celular de producción" se entiende habitualmente que es una línea celular lista para su uso para el cultivo en un biorreactor para obtener el producto de un proceso de producción, tal como una PDI. Las expresiones "levadura hospedadora" o "línea celular de levadura" o "célula hospedadora de levadura" o "célula hospedadora" u "hospedadores" significarán cualquier célula de levadura que se pueda cultivar para producir una PDI o un metabolito de célula hospedadora.

El término "expresión" o "sistema de expresión" o "casete de expresión" se refiere a las moléculas de ácido nucleico que contienen una secuencia codificante deseada de un producto de expresión tal como, p.ej., una PDI y secuencias de control tales como, p.ej., un promotor en unión operable, de forma que los hospedadores transformados o transfectados con estas secuencias son capaces de producir las proteínas codificadas o los metabolitos de las células hospedadoras. Para llevar a cabo la transformación, se puede incluir el sistema de expresión en un vector; sin embargo, el ADN relevante también se puede integrar en el cromosoma hospedador. La expresión se puede referir a los productos de expresión secretados o no secretados, que incluyen polipéptidos o metabolitos. Específicamente, el término se refiere a una célula hospedadora y a un vector compatible en condiciones adecuadas, p.ej. para la expresión de una proteína codificada por un ADN exógeno portado por el vector e introducido en la célula hospedadora.

Las "construcciones de expresión" o "vectores" usados en la presente memoria se definen como secuencias de ADN que son necesarias para la transcripción de las secuencias de nucleótidos recombinantes clonadas, es decir, de los genes recombinantes y la traducción de su mRNA en un organismo hospedador adecuado. Los vectores de expresión comprenden el casete de expresión y además comprenden normalmente un origen para la replicación autónoma en las células hospedadoras o un sitio de integración en el genoma, uno o más marcadores seleccionables (p.ej. un gen de síntesis de aminoácidos o un gen que confiere resistencia a antibióticos tales como zeocina, kanamicina, G418 o higromicina), varios sitios de escisión para enzimas de restricción, una secuencia promotora adecuada y un terminador de la transcripción, cuyos componentes se unen de forma operable entre sí. Los términos "plásmido" y "vector", tal como se usan en la presente memoria, incluyen secuencias de nucleótidos que se replican de manera autónoma, así como secuencias de nucleótidos de integración en el genoma.

Específicamente, el término se refiere a un vehículo mediante el cual se puede introducir una secuencia de ADN o ARN (p.ej. un gen exógeno) en una célula hospedadora, para transformar al hospedador y estimular la expresión (p.ej. la transcripción y traducción) de la secuencia introducida. Los plásmidos son los vectores preferidos.

Los vectores comprenden en general el ADN de un agente transmisible, en el que se inserta el ADN exógeno. Una manera habitual de insertar un segmento de ADN en otro segmento de ADN implica el uso de enzimas denominadas enzimas de restricción que escinden el ADN en sitios específicos (grupos específicos de nucleótidos) denominados sitios de restricción. Un "casete" se refiere a una secuencia codificante de ADN o segmento de ADN que codifica un producto de expresión que se puede insertar en un vector en sitios de restricción definidos. Los sitios de restricción del casete se diseñan para asegurar la inserción del casete en el marco de lectura adecuado. En general, el ADN exógeno se inserta en uno o más sitios de restricción del ADN del vector, y después es transportado por el vector a una célula hospedadora junto con el ADN del vector transmisible. Un segmento o secuencia de ADN que tiene ADN insertado o añadido, tal como un vector de expresión, también se puede denominar "construcción de ADN". Un tipo habitual de vector es un "plásmido", que en general es una molécula autocontenida de ADN bicatenario que puede aceptar fácilmente ADN adicional (exógeno) y que se puede introducir fácilmente en una célula hospedadora adecuada. Un vector plasmídico contiene a menudo ADN codificante y ADN promotor, y tiene uno o más sitios de restricción adecuados para insertar ADN exógeno. El ADN codificante es una secuencia de ADN que codifica una secuencia de aminoácidos particular para un polipéptido o proteína particular tal como, p.ej., una PDI. El ADN promotor es una secuencia de ADN que inicia, regula, o de otra manera actúa como mediador o controla la expresión del ADN codificante. El ADN promotor y el ADN codificante pueden ser del mismo gen o de genes diferentes, y pueden ser del mismo organismo o de organismos diferentes. Los vectores de clonación recombinantes incluirán a menudo uno o más sistemas de replicación para la clonación o expresión, uno o más marcadores para la selección en el hospedador, p.ej. resistencia a antibióticos, y uno o más casetes de expresión.

La expresión "variante funcional", tal como se usa en la presente memoria, p.ej. con respecto a las secuencias reguladoras según la invención, tales como el péptido señal con la secuencia de aminoácidos de SEQ ID 1, o con respecto a un líder con la secuencia de aminoácidos de SEQ ID 10, se referirá a las variantes con una o dos mutaciones puntuales en la secuencia de aminoácidos, que tienen sustancialmente la misma señal o actividad de líder en comparación con las secuencias sin modificar. Las variantes funcionales de los ácidos nucleicos descritos en la presente memoria abarcan además las secuencias con optimización de codones, también denominadas en la presente memoria "variantes con optimización de codones", que codifican cualquiera de los péptidos señal de la

presente invención. Tal optimización de codones de un ácido nucleico se entiende como la alteración sistemática de los codones en el ADN recombinante para expresarlo en un sistema heterólogo para coincidir con el patrón de uso de codones en el organismo usado para la expresión. La intención es específicamente mejorar la producción de una proteína expresada.

- 5 Se entiende que las expresiones "péptido señal", "líder" o "líder truncado", tal como se usan en la presente memoria, se refieren siempre a los aminoácidos específicos de SEQ ID 1 y SEQ ID 10, respectivamente, y también a las variantes funcionales de los mismos con una o dos mutaciones puntuales.

- 10 La expresión "sustancialmente la misma actividad de señal o líder", tal como se usa en la presente memoria, se refiere a la actividad tal como se indica hospedadora recombinante; por ejemplo, un nivel de PDI en el sobrenadante que es al menos un 50%, al menos un 60%, al menos un 70%, al menos un 80%, al menos un 90%, al menos un 95%, al menos un 98% del nivel de PDI en el sobrenadante tal como se proporciona mediante el líder de SEQ ID 1 o SEQ ID 10, respectivamente.

- 15 Una mutación puntual se entiende como la modificación de un polinucleótido que da como resultado la expresión de una secuencia de aminoácidos que difiere de la secuencia de aminoácidos sin modificar por la sustitución o el intercambio de uno o más aminoácidos individuales (no consecutivos) por un aminoácido diferente. Las variantes funcionales preferidas tienen una o más mutaciones puntuales en las posiciones 3 y 16 de SEQ ID 1 y SEQ ID 10, respectivamente.

- 20 Las mutaciones puntuales preferidas adicionalmente se refieren al intercambio de aminoácidos de la misma polaridad y/o carga. A este respecto, los aminoácidos se refieren a los veinte aminoácidos naturales codificados por sesenta y cuatro codones de tripletes. Estos 20 aminoácidos se pueden dividir en los que tienen cargas neutras, cargas positivas, y cargas negativas:

Los aminoácidos "neutros" se muestran a continuación junto con sus códigos respectivos de tres letras y de una letra, y la polaridad:

- 25 Alanina: (Ala, A) apolar, neutro;  
 Asparagina: (Asn, N) polar, neutro;  
 Cisteína: (Cys, C) apolar, neutro;  
 Glutamina: (Gln, Q) polar, neutro;  
 Glicina: (Gly, G) apolar, neutro;
- 30 Isoleucina: (Ile, I) apolar, neutro;  
 Leucina: (Leu, L) apolar, neutro;  
 Metionina: (Met, M) apolar, neutro;  
 Fenilalanina: (Phe, F) apolar, neutro;  
 Prolina: (Pro, P) apolar, neutro;
- 35 Serina: (Ser, S) polar, neutro;  
 Treonina: (Thr, T) polar, neutro;  
 Triptófano: (Trp, W) apolar, neutro;  
 Tirosina: (Tyr, Y) polar, neutro;  
 Valina: (Val, V) apolar, neutro; y
- 40 Histidina: (His, H) polar, positivo (10%) neutro (90%).  
 Los aminoácidos cargados "positivamente" son:  
 Arginina: (Arg, R) polar, positivo; y  
 Lisina: (Lys, K) polar, positivo.
- Los aminoácidos cargados "negativamente" son:
- 45 Ácido aspártico: (Asp, D) polar, negativo; y

Ácido glutámico: (Glu, E) polar, negativo.

El término "aislado" o "aislamiento", tal como se usa en la presente memoria con respecto a un ácido nucleico, una PDI u otro compuesto, se referirá a un compuesto que se ha separado lo suficientemente del medio con el que estaría asociado de manera natural, como para que exista en una forma "sustancialmente pura". "Aislado" no significa necesariamente la exclusión de mezclas artificiales o sintéticas con otros compuestos o materiales, o la presencia de impurezas que no interfieren con la actividad fundamental, y que pueden estar presentes, por ejemplo, debido a una purificación incompleta. En particular, las moléculas de ácido nucleico aisladas de la presente invención también pretenden incluir las sintetizadas químicamente. Con referencia a los ácidos nucleicos de la invención, a veces se usa la expresión "ácido nucleico aislado". Esta expresión, cuando se aplica al ADN, se refiere a una molécula de ADN que se separa de las secuencias con las que está inmediatamente contigua en el genoma natural del organismo del que se originó. Por ejemplo, un "ácido nucleico aislado" puede comprender una molécula de ADN insertada en un vector, tal como un plásmido o vector viral, o integrada en el ADN genómico de una célula procariótica o eucariótica u organismo hospedador. Específicamente, la expresión "ácido nucleico aislado" según la invención excluye que el ácido nucleico que codifica la proteína EPX1 esté unido a un ácido nucleico que codifica el líder según la presente invención. Cuando se aplica al ARN, la expresión "ácido nucleico aislado" se refiere principalmente a una molécula de ARN codificada por una molécula de ADN aislada como se definió anteriormente. De manera alternativa, la expresión se puede referir a una molécula de ARN que se ha separado suficientemente de otros ácidos nucleicos con los que estaría asociada en su estado natural (es decir, en las células o tejidos). Un "ácido nucleico aislado" (ADN o ARN) puede representar además una molécula producida directamente mediante medios biológicos o sintéticos, y separada de los otros componentes presentes durante su producción.

El término "líder", tal como se usa en la presente memoria, se entiende de la siguiente manera. Las regiones codificantes de polinucleótidos y de ácido nucleico en el casete de expresión de la invención pueden estar asociadas con regiones codificantes adicionales que codifican péptidos secretores o péptidos señal, que dirigen la secreción de una PDI. Las proteínas destinadas a la ruta secretora tienen una secuencia líder N-terminal que se escinde de la proteína madura una vez que se ha iniciado la exportación de la cadena de proteína naciente a través del retículo endoplásmico rugoso. Un líder induce que la proteína expresada se transporte hacia o fuera de la membrana plasmática, y de ese modo facilita separar y purificar la proteína expresada. En general, una proteína de membrana o una proteína secretora que se transporta al espacio periplasmático, a la membrana celular o fuera de la célula comprende tal secuencia N-terminal. Normalmente, los líderes se escinden de la proteína mediante peptidasas celulares especializadas después de transportar las proteínas.

Las proteínas secretadas por las células eucarióticas tienen en general una secuencia líder fusionada en el extremo N-terminal de la proteína, que se escinde de la proteína completa o "de longitud completa" para producir una forma secretada o "madura" de la proteína.

Los ejemplos específicos de un líder descrito en la presente memoria son un líder que consiste en un péptido señal o un líder que consiste en un péptido señal y una prosequencia, tal como el líder truncado descrito en la presente memoria. El líder que consiste en el péptido señal de SEQ ID 1 o el líder truncado de SEQ ID 10 se denominan también en la presente memoria secuencias reguladoras.

El término, tal como se usa en la presente memoria, se refiere en particular a una secuencia de control para la posible modificación de la expresión de una PDI. El líder, también denominado péptido líder, se une al extremo N-terminal de una secuencia de aminoácidos de PDI. El ácido nucleico que codifica el líder está en posición 5', y unido de forma operable al extremo 5' de la secuencia de ácido nucleico que codifica la PDI. Se puede usar cualquier secuencia líder según la invención que sea funcional en la célula hospedadora de elección.

La expresión "residuo de aminoácido N-terminal nativo" o "secuencia de aminoácidos N-terminal nativa" se entiende que se refiere a uno o más aminoácido(s) de la secuencia N-terminal, p.ej. el residuo de aminoácido N-terminal de una PDI enumerada, cuyo residuo de aminoácido se considera el correcto en comparación con la secuencia de una PDI enumerada a expresar. El residuo de aminoácido N-terminal nativo, así, proporciona un extremo N-terminal o región N-terminal nativa de una PDI, lo cual es un prerrequisito para obtener una secuencia de aminoácidos correcta, completa, para obtener un compuesto funcional sin ningún residuo(s) de aminoácido(s) N-terminal(es) adicional(es) (superfluo(s)) exógeno(s) para la PDI. En general, se entiende que cualquier proteína de tipo natural tiene un residuo de aminoácido N-terminal nativo. Además, una proteína recombinante también puede tener un residuo de aminoácido N-terminal nativo, que está predefinido y exhibe propiedades deseables de la proteína.

Específicamente, cuando el péptido señal o líder truncado descrito en la presente memoria se une directamente al residuo de aminoácido N-terminal nativo de una PDI, la proteína liberada contendrá un residuo de aminoácido N-terminal natural al menos en cierto grado, y en general no comprende una prolongación N-terminal de longitud variable. La composición preferida de una PDI se caracteriza por un residuo de aminoácido N-terminal nativo que comprende un extremo N-terminal correcto, al menos en cierto grado, que es preferiblemente la mayoría de las moléculas de PDI, o al menos un 60, 65, 70, 75, 80, 85, 90, 95, 98, o 100% (p/p) de las moléculas de PDI, sin residuos de aminoácidos adicionales en el extremo N-terminal, tal como los que se originan de un péptido señal o pro-secuencia, o un fragmento de la misma.

El término "pro-secuencia", tal como se usa en la presente memoria, se referirá a una secuencia de aminoácidos precursora unida de forma operable al extremo N-terminal de una PDI. La pro-secuencia puede tener además una secuencia señal unida de forma operable al extremo N-terminal de la pro-secuencia. En general, la pro-secuencia se escinde de la PDI para proporcionar la forma madura de la PDI.

- 5 Un ejemplo específico de una pro-secuencia descrita en la presente memoria es parte del líder truncado, y tiene la secuencia de aminoácidos de

SEQ ID 7: APVAPAEAAANHLHKR

es decir, la pro-secuencia truncada: los aminoácidos 21-36 de la secuencia líder truncada de SEQ ID 10.

10 La expresión "unido de forma operable", tal como se usa en la presente memoria, se refiere a la asociación de secuencias de nucleótidos en una única molécula de ácido nucleico, p.ej. un vector, de manera que la función de una o más secuencias de nucleótidos se ve afectada por al menos otra secuencia de nucleótidos presente en dicha molécula de ácido nucleico. Por ejemplo, un promotor está unido de forma operable con una secuencia codificante de un gen recombinante cuando es capaz de causar un efecto sobre la expresión de esa secuencia codificante. Como ejemplo adicional, un ácido nucleico que codifica un péptido señal está unido de forma operable a una  
15 secuencia de ácido nucleico que codifica una PDI cuando es capaz de expresar una proteína en la forma secretada, tal como una preforma de una proteína madura o la proteína madura. Específicamente, tales ácidos nucleicos unidos de forma operable entre sí se pueden unir inmediatamente, es decir, sin elementos adicionales o secuencias de ácido nucleico entre el ácido nucleico que codifica el péptido señal y la secuencia de ácido nucleico que codifica una PDI.

20 "Promotor", tal como se usa en la presente memoria, se refiere a una secuencia de ADN capaz de controlar la expresión de una secuencia codificante o ARN funcional. La actividad del promotor se puede estudiar por su eficacia transcripcional. Esto se puede determinar directamente midiendo la cantidad de transcripción de mRNA del promotor, p.ej. mediante transferencia de Northern, o indirectamente midiendo la cantidad de producto del gen expresado por el promotor.

25 La expresión "proteína de interés (PDI)", tal como se usa en la presente memoria, se refiere a un polipéptido o una proteína que se produce por medio de la tecnología recombinante en una célula hospedadora, también se denomina PDI recombinante o PDI producida por la célula hospedadora recombinante. De manera más específica, la PDI recombinante puede no ser natural en la célula hospedadora, es decir, una proteína heteróloga, o de otra manera puede ser nativa respecto de la célula hospedadora, es decir, una proteína homóloga respecto de la célula hospedadora, pero se produce, por ejemplo, mediante transformación con un vector auto-replicante que contiene la  
30 secuencia de ácido nucleico que codifica la PDI, o tras la integración mediante técnicas recombinantes de una o más copias de la secuencia de ácido nucleico que codifica la PDI en el genoma de la célula hospedadora, o mediante modificación recombinante de una o más secuencias reguladoras que controlan la expresión del gen que codifica la PDI, p.ej. del promotor o la secuencia señal. En casos específicos, la PDI recombinante, la PDI heteróloga u homóloga, se sobreexpresa mediante la célula hospedadora recombinante para obtener un rendimiento elevado de  
35 un producto. En ciertos casos, la PDI de expresión, tal como se usa en la presente memoria, también se refiere a cualquier producto de metabolito de la célula hospedadora mediado por la proteína expresada de manera recombinante.

40 El término "secreción", tal como se usa en la presente memoria, se refiere a la translocación de un polipéptido o proteína, específicamente una PDI, a través de la membrana plasmática y la pared celular de una célula vegetal hospedadora. La PDI secretada puede ser parte de la membrana celular en forma de una proteína unida a la membrana que está anclada dentro de la pared celular, o se puede liberar en forma de una proteína soluble en el sobrenadante celular.

45 Se entiende que el término "secreción", tal como se usa en la presente memoria con referencia a una PDI, abarca específicamente la expresión de una PDI en forma madura (que incluye las proformas de proteínas activas o las proteínas activas), en forma de una PDI unida a la membrana o en forma de una PDI extracelular.

50 El término "recombinante", tal como se usa en la presente memoria, significará "preparado mediante o el resultado de ingeniería genética". Así, un "microorganismo recombinante" comprende al menos un "ácido nucleico recombinante". Un microorganismo recombinante comprende específicamente un vector de expresión o un vector de clonación, o se ha modificado genéticamente para que contenga una secuencia de ácido nucleico recombinante. Una "proteína recombinante" se produce expresando un ácido nucleico recombinante respectivo en un hospedador.

55 Las moléculas de ácido nucleico o péptidos/polipéptidos/proteínas descritas en la presente memoria son preferiblemente recombinantes, para proporcionar fusiones de un líder con una PDI. Tal como se usa en la presente memoria, "recombinante" se refiere a una combinación artificial de dos segmentos de secuencias de otra manera separados, p.ej., mediante síntesis química o mediante la manipulación de segmentos aislados de ácidos nucleicos con técnicas de ingeniería genética. "Recombinante" también incluye la referencia a una célula o casete de expresión que se ha modificado mediante la introducción de un ácido nucleico heterólogo o una célula derivada de una célula así modificada, pero no abarca la alteración de la célula o vector mediante sucesos que se dan de manera natural (p.ej., mutación espontánea, transformación/transducción/transposición natural), tales como los que

se dan sin intervención humana deliberada.

La expresión "secuencia señal", tal como se usa en la presente memoria, se referirá a un ácido nucleico que codifica un péptido señal, que normalmente es una cadena peptídica corta (3-60 aminoácidos de longitud) que dirige el transporte de una proteína. Los péptidos señal también se pueden denominar señales de transporte, péptidos de tránsito, o señales de localización. Un péptido señal induce específicamente que una proteína expresada se transporte a lo largo de la ruta secretora. En general, una proteína de membrana o una proteína secretora que se transporta al espacio periplasmático, a la membrana celular o fuera de la célula comprende tal secuencia señal N-terminal. Normalmente, los péptidos señal se escinden de la proteína madura mediante una peptidasa señal una vez que se ha llevado a cabo la translocación de la cadena proteica naciente al retículo endoplásmico.

Un ejemplo específico de un péptido señal tal como se usa en la presente memoria es un péptido señal de SEQ ID 1, o una variante funcional del mismo con una o dos mutaciones puntuales. El péptido señal tal como se usa en la presente memoria tiene en general las características inherentes que posibilitan la escisión después del aminoácido Ala20, incluso sin la presencia del sitio de escisión de SEQ ID 6.

El péptido señal está codificado por una secuencia de ácido nucleico que va seguida en general de la secuencia de ácido nucleico que codifica una PDI, opcionalmente con una pro-secuencia en posición 3' de la secuencia señal y en posición 5' de la secuencia que codifica la PDI.

La expresión "sustancialmente pura" o "purificada", tal como se usa en la presente memoria, se referirá a una preparación que comprende al menos un 50% (p/p), preferiblemente al menos un 60%, 70%, 80%, 90% o 95% de un compuesto, tal como una molécula de ácido nucleico o una PDI. La pureza se mide mediante métodos adecuados para el compuesto (p.ej. métodos cromatográficos, electroforesis en gel de poliacrilamida, análisis mediante HPLC, y similares).

Fue sorprendente identificar y caracterizar el ácido nucleico aislado nuevo que codifica el péptido señal según la invención, cuyo péptido señal se descubrió que incorpora una característica inherente que posibilita la escisión en el extremo C-terminal independiente del siguiente residuo de aminoácido. Esto también se denomina en la presente memoria un "sitio de secreción inherente" en el extremo C-terminal. Por lo tanto, la secuencia de aminoácidos tras la secuencia del péptido señal tendría una secuencia de aminoácidos N-terminal nativa correcta, una vez que se escinde el líder.

De manera inesperada, al usar un ácido nucleico que codifica un péptido señal según la invención, se pudo omitir el sitio de escisión para la peptidasa señal que se conoce en la técnica que está localizado entre la secuencia señal y la pro-secuencia del líder de Epx1 de *P. pastoris* nativo, es decir, entre la posición 20 y 21, indicado en forma de un guion en la secuencia del sitio de escisión: VSA-AP (SEQ ID 6). Por lo tanto, se podría proporcionar un sistema de expresión mejorado para el residuo de aminoácido N-terminal correcto y nativo de una PDI, que no comprende específicamente una Alanina adicional como residuo de aminoácido N-terminal, al menos en cierto grado, p.ej. la mayoría de moléculas de PDI comprenden la secuencia N-terminal nativa, hasta el 100%, como se determina mediante LC-MS.

Por lo tanto, fue posible por primera vez usar un ácido nucleico aislado listo para su uso que codifica el péptido señal de 20 aminoácidos. Las secuencias señal de la técnica anterior (p.ej. del documento US20110021378A1) siempre incluyeron al menos un residuo de aminoácido adicional en el extremo C-terminal, que es una Alanina, que podría provocar que se expresase un extremo N-terminal erróneo de la proteína a expresar.

De forma similar, se encontró que el líder truncado descrito en la presente memoria incorporaba una característica inherente que posibilita la escisión en el extremo C-terminal independientemente del siguiente residuo de aminoácido. Esto también se denomina en la presente memoria un "sitio de escisión inherente" en el extremo C-terminal. Por lo tanto, la secuencia de aminoácidos tras el extremo C-terminal del líder truncado o tras el extremo C-terminal de la pro-secuencia del líder truncado tendría una secuencia de aminoácidos N-terminal correcta y nativa una vez que se escinde el líder.

Los ácidos nucleicos aislados y el vector de expresión descrito en la presente memoria, así, posibilitan la expresión y secreción de una PDI con el residuo de aminoácido N-terminal nativo. Esto fue lo más sorprendente, porque el rendimiento de la secreción con el péptido señal de 20 aminoácidos fue mucho mayor en comparación con los experimentos llevados a cabo con la secuencia de 21 aminoácidos, es decir, el péptido señal de 20 aminoácidos prolongado con una Alanina adicional en el extremo C-terminal como se propuso, p.ej., en el documento US20110021378A1.

Así, otra realización de la invención es el uso de un ácido nucleico o péptido líder según la invención para la secreción de una PDI y/o para incrementar la secreción de una PDI desde una célula hospedadora, preferiblemente, en el que al menos un 60, 65, 70, 75, 80, 85, 90, 95, 98, o 100% de la PDI comprende una secuencia de aminoácidos N-terminal nativa una vez que se ha escindido el líder según la invención. Preferiblemente, el incremento de la secreción es de 1,15, 1,5, 2, 5, 10, 20, 30, 40, 50, 60, 70, 80, 90 o más veces en comparación con el péptido señal conocido de SEQ ID 21. Lo más preferiblemente, la PDI es un anticuerpo o fragmento o derivado del mismo como se define adicionalmente más adelante.

Además, se descubrió de manera inesperada que se podría usar una secuencia líder truncada, que tiene una longitud de solamente un 63% del líder de longitud completa nativo de la técnica anterior, por tanto, menor que una parte significativa de la secuencia líder de longitud completa de SEQ ID 8. Fue sorprendente descubrir incluso un rendimiento de secreción incrementado con el líder truncado en comparación con los experimentos llevados a cabo con la secuencia líder de longitud completa.

Según la invención, los ácidos nucleicos aislados o el casete o vector de expresión se podrían usar con otros elementos reguladores o secuencias convencionales, empleados de manera natural o usados en general en los sistemas de expresión recombinante, p.ej. para proporcionar construcciones para la producción de proteínas recombinantes con un rendimiento elevado. Los ejemplos de secuencias reguladoras incluyen promotores, operadores, y potenciadores, sitios de unión ribosómica, y secuencias que controlan el inicio y la terminación de la transcripción y de la traducción.

Para demostrar la función de las secuencias relevantes se pueden construir casetes o vectores de expresión para controlar la expresión de una PDI, y el rendimiento expresado y/o secretado se compara respecto de las construcciones con elementos reguladores convencionales. Se puede hallar una descripción detallada del procedimiento experimental más adelante en los ejemplos.

Las secuencias identificadas se amplificaron mediante PCR de *P. pastoris* con el uso de cebadores nucleotídicos específicos, se clonaron en un vector de expresión de levadura unido de manera funcional al extremo N-terminal de la PDI, o en posición 5' de la secuencia de la PDI, y se transformaron en una línea de células hospedadoras de levadura, p.ej. *P. pastoris*, para la producción a nivel elevado de diversas PDI diferentes en forma secretada. Para estimar el efecto de las secuencias reguladoras, tales como la secuencia señal y el líder truncado según la invención, sobre el rendimiento de PDI, la línea de células hospedadoras de levadura obtenida según la invención se puede cultivar en experimentos con matraces agitados y fermentaciones semicontinuas o en quimiostato en comparación con las cepas que comprenden elementos reguladores convencionales.

Por medio de las secuencias reguladoras inventivas, en particular las secuencias que codifican el péptido señal y el vector de expresión, el método según la invención preferiblemente no solamente proporciona una producción incrementada mediante una secreción mejorada, sino que también proporciona una mayor calidad de la PDI en una célula hospedadora de levadura, y en particular en una célula hospedadora de *P. pastoris*. Un incremento de la secreción de la PDI se determina basándose en una comparación de su rendimiento de secreción en presencia de la secuencia reguladora, en particular la secuencia señal o el líder truncado, que incrementa la secreción de proteína en comparación con los elementos de la técnica anterior.

La PDI puede ser cualquier polipéptido eucariótico, procariótico o sintético. Se puede secretar como una proteína madura, como una proteína unida a la membrana o como una proteína expresada de manera extracelular. La presente invención también proporciona la producción recombinante de variantes funcionalmente equivalentes, derivados y fragmentos biológicamente activos de proteínas que se dan de manera natural. Las variantes funcionalmente equivalentes preferiblemente tienen sustancialmente las mismas características funcionales o actividad.

Una PDI mencionada en la presente memoria puede ser un producto homólogo para la célula hospedadora eucariótica o heterólogo, preferiblemente para uso terapéutico, profiláctico, diagnóstico, analítico o industrial.

La PDI es preferiblemente un polipéptido o proteína recombinante heterólogo, producido en una célula de levadura.

Específicamente, la PDI es una proteína eucariótica, preferiblemente una proteína de mamífero.

Una PDI producida según la invención puede ser una proteína multimérica, preferiblemente un dímero o tetrámero.

Según un aspecto de la invención, la PDI es una proteína recombinante o heteróloga, preferiblemente seleccionada de proteínas terapéuticas, que incluyen anticuerpos o fragmentos de los mismos, enzimas y péptidos, antibióticos proteicos, proteínas de fusión con toxinas, conjugados carbohidrato - proteína, proteínas estructurales, proteínas reguladoras, vacunas y proteínas o partículas similares a vacunas, enzimas de procesamiento, factores de crecimiento, hormonas y citocinas, o un metabolito de una PDI.

Una PDI específica es una molécula de unión a un antígeno tal como un anticuerpo, o un fragmento del mismo. Entre las PDIs específicas están los anticuerpos tales como anticuerpos monoclonales (mAbs), inmunoglobulina (Ig) o la clase de inmunoglobulina G (IgG), anticuerpos de cadena pesada (HcAbs), o fragmentos de los mismos tales como el fragmento de unión al antígeno (Fab), Fd, fragmento variable de cadena simple (scFv), o las variantes modificadas de los mismos tales como, por ejemplo, dímeros de Fv (diacuerpos), trímeros de Fv (triacuerpos), tetrámeros de Fv, o minicuerpos y anticuerpos de dominio simple similares a VH o VHH o V-NAR.

Las PDIs específicas adicionales son aprotinina, inhibidor de la ruta del factor tisular u otros inhibidores de proteasas, e insulina o precursores de insulina, análogos de insulina, hormonas del crecimiento, interleucinas, activador de plasminógeno tisular, factor de crecimiento transformante a o b, glucagón, péptido similar a glucagón 1 (GLP-1), péptido similar a glucagón 2 (GLP-2), GRPP, Factor VII, Factor VIII, Factor XIII, factor de crecimiento

- derivado de plaquetas 1, albúmina de suero, enzimas, tales como lipasas o proteasas, o un homólogo funcional, variante equivalente funcional, derivado y fragmento biológicamente activo con una función similar a la de la proteína nativa. La PDI puede ser similar estructuralmente a la proteína nativa, y se puede obtener a partir de la proteína nativa mediante la adición de uno o más aminoácidos en uno o ambos extremos C- y N-terminales o en la cadena lateral de la proteína nativa, mediante la sustitución de uno o más aminoácidos en uno o varios sitios diferentes de la secuencia de aminoácidos nativa, mediante la delección de uno o más aminoácidos en uno o ambos extremos de la proteína nativa o en uno o varios sitios de la secuencia de aminoácidos, o mediante la inserción de uno o más aminoácidos en uno o más sitios de la secuencia de aminoácidos nativa. Tales modificaciones son muy conocidas para varias de las proteínas mencionadas anteriormente.
- 5
- 10 También se puede seleccionar una PDI de sustratos, enzimas, inhibidores o cofactores que posibilitan reacciones bioquímicas en la célula hospedadora, con el objetivo de obtener el producto de dicha reacción bioquímica o una cascada de varias reacciones, p.ej. para obtener un metabolito de la célula hospedadora. Los productos ejemplares pueden ser vitaminas, tales como riboflavina, ácidos orgánicos, y alcoholes, que se pueden obtener con rendimientos incrementados tras la expresión de una proteína recombinante o una PDI descrita en la presente memoria.
- 15
- Específicamente, la célula hospedadora, que expresa un producto recombinante descrito en la presente memoria, puede ser cualquier célula de levadura adecuada para la expresión recombinante de una PDI.
- Las células hospedadoras preferidas se seleccionan del género *Pichia*, *Candida*, *Torulopsis*, *Arxula*, *Hansenula*, *Ogatea*, *Yarrowia*, *Kluyveromyces*, *Saccharomyces* o *Komagataella*, y preferiblemente una levadura metilotrófica, y se prefiere específicamente *P. pastoris*, *Komagataella pastoris*, *K. phaffii*, o *K. pseudopastoris*.
- 20
- Los ejemplos de células hospedadoras preferidas según la invención incluyen, pero sin limitación, el género *Pichia*, tal como *P. pastoris*, o *P. methanolica* o el género *Komagataella*, tal como *K. pastoris*, *K. pseudopastoris* o *K. phaffii*.
- La bibliografía más reciente divide y renombra *Pichia pastoris* en *Komagataella pastoris*, *Komagataella phaffii* y *Komagataella pseudopastoris*. En la presente memoria, se usa *Pichia pastoris* como sinónimo para todas, *Komagataella pastoris*, *Komagataella phaffii* y *Komagataella pseudopastoris*.
- 25
- Los ejemplos de las cepas de *P. pastoris* incluyen CBS 704 (=NRRL Y-1603 = DSMZ 70382), CBS 2612 (=NRRL Y-7556), CBS 7435 (=NRRL Y-11430), CBS 9173-9189 (cepas CBS: Centro de Biodiversidad Fúngica CBS-KNAW, Centraalbureau voor Schimmel-cultures, Utrecht, Países Bajos), y DSMZ 70877 (Colección Alemana de Microorganismos y Cultivos Celulares), pero también las cepas de Invitrogen, tales como X-33, GS115, KM71 y SMD1168. Todas las cepas descritas anteriormente se han usado con éxito para producir transformantes y expresar genes heterólogos.
- 30
- Según la invención, se prefiere proporcionar una célula hospedadora de levadura transformada con un vector que comprende una secuencia promotora unida de forma operable a la secuencia señal según la invención.
- Según un modo preferido de la invención, el casete de expresión según la invención comprende un promotor, tal como el promotor pGAP (gliceraldehído fosfato deshidrogenasa) de *P. pastoris*, el promotor pAOX (alcohol oxidasa) o SEQ ID 9 (Fig. 1, promotor pG1), o las variantes funcionales de los mismos. Se puede obtener una actividad de secuencia significativa con solamente parte de tales secuencias promotoras. Específicamente, una parte preferida de la secuencia promotora consiste en al menos 150 bases consecutivas o pb, más preferiblemente al menos 200 pb, al menos 300 pb, al menos 400 pb, al menos 500 pb, al menos 600 pb, al menos 700 pb, al menos 800 pb o al menos 900 pb. Preferiblemente, el extremo 3' de la región promotora está contenido en la parte preferida de la secuencia promotora.
- 35
- 40
- En un sistema de expresión preferido, el promotor es un promotor inducible o constitutivo. El promotor puede ser un promotor endógeno o heterólogo respecto de la célula hospedadora. Un ejemplo preferido de un promotor inducible es el promotor pG1, que es inducible mediante condiciones limitantes de glucosa y tiene la secuencia de nucleótidos de SEQ ID 9.
- 45
- Una célula hospedadora específica según la invención contiene secuencias promotoras heterólogas o recombinantes, que pueden derivar de una cepa diferente del hospedador de producción, tal como de otra cepa de levadura, tal como una cepa de *S. cerevisiae*. En otra realización específica, la célula hospedadora según la invención comprende un constructo de expresión recombinante según la invención que comprende el promotor que se origina del mismo género, especie o cepa que la célula hospedadora.
- 50
- El promotor puede ser cualquier secuencia de ADN que muestre actividad transcripcional en la célula hospedadora, y se puede obtener de genes que codifican proteínas homólogas o heterólogas respecto del hospedador. El promotor se obtiene preferiblemente de un gen que codifica una proteína homóloga respecto de la célula hospedadora.
- 55
- Por ejemplo, se puede obtener un promotor de una levadura, tal como una cepa de *S. cerevisiae*. Sin embargo, una realización específicamente preferida emplea un promotor que se origina de *P. pastoris* para el uso en un método

para producir una PDI recombinante en una línea celular hospedadora productora de *P. pastoris*. El origen homólogo de la secuencia de nucleótidos facilita su incorporación en la célula hospedadora del mismo género o especie, lo que permite la producción estable de una PDI, posiblemente con un rendimiento incrementado en los procesos de fabricación industriales. Además, se pueden usar variantes funcionalmente activas del promotor de otras levaduras adecuadas u otros hongos o de otros organismos tales como vertebrados o vegetales.

Las secuencias promotoras adecuadas adicionales para el uso con las células hospedadoras de levadura pueden incluir, pero sin limitación, los promotores obtenidos de genes que codifican enzimas metabólicas que se sabe que están presentes a una concentración elevada en la célula, p.ej. enzimas glucolíticas similares a triosafosfato isomerasa (TPI), fosfoglicerato quinasa (PGK), gliceraldehído-3-fosfato deshidrogenasa (GAPDH), alcohol oxidasa (AOX), lactasa (LAC) y galactosidasa (GAL).

Los ejemplos preferidos de promotores adecuados son los promotores de levaduras, que contienen una secuencia de ADN que funciona como un promotor para la transcripción génica en las células de levadura. Los ejemplos preferidos son los promotores de Mal, TPI, CUP, ADH o PGK de *S. cerevisiae*, o el promotor de glucosa-6-fosfato isomerasa de *P. pastoris* (PPGI), el promotor de 3-fosfoglicerato quinasa (PPGK) o el promotor de glicerol aldehído fosfato deshidrogenasa PGAP, el promotor de alcohol oxidasa (PAOX), el promotor de formaldehído deshidrogenasa (PFLD), el promotor de isocitrato liasa (PICL), el promotor del factor de elongación de la traducción (PTEF), y los promotores de enolasa 1 de *P. pastoris* (PENO1), triosa fosfato isomerasa (PTPI), alfa-cetoisocaproato descarboxilasa (PTH1), proteínas de las subunidades ribosómicas (PRPS2, PRPS7, PRPS31, PRPL1), miembros de la familia de proteínas de choque térmico (PSSA1, PHSP90, PKAR2), 6-fosfogluconato deshidrogenasa (PGND1), fosfoglicerato mutasa (PGPM1), transcetolasa (PTKL1), fosfatidilinositol sintasa (PPIS1), ferro-O<sub>2</sub>-oxidoreductasa (PFET3), permeasa de alta afinidad para hierro (PFTR1), fosfatasa alcalina reprimible (PPHO8), N-miristoil transferasa (PNMT1), factor de transcripción de respuesta a feromonas (PMCM1), ubiquitina (PUBI4), endonucleasa de ADN monocatenario (PRAD2) y el promotor del transportador de ADP/ATP principal de la membrana interna mitocondrial (PPET9).

Si la PDI es una proteína homóloga respecto de la célula hospedadora, es decir, una proteína que se da de manera natural en la célula hospedadora, la expresión de la PDI en la célula hospedadora se puede modular mediante el intercambio de su secuencia promotora nativa con una secuencia promotora heteróloga respecto de la célula hospedadora o con una secuencia promotora homóloga respecto de la célula hospedadora, pero diferente a la secuencia promotora nativa de dicha PDI.

Este propósito de introducir un nuevo promotor se puede conseguir, p.ej., mediante la transformación de una célula hospedadora con una molécula de ADN recombinante que comprende secuencias homólogas del gen de interés para permitir la recombinación específica de sitio, la secuencia promotora y un marcador selectivo adecuado para la célula hospedadora. La recombinación específica de sitio tendrá lugar para unir de manera operable la secuencia promotora a la secuencia de nucleótidos que codifica la PDI. Esto da como resultado la expresión de la PDI a partir de la secuencia promotora heteróloga en vez de a partir de la secuencia promotora nativa.

En una realización específicamente preferida, la secuencia promotora tiene una actividad promotora incrementada respecto de la secuencia promotora nativa de la PDI.

También es posible proporcionar un vector o casete de expresión comodín según la invención, que comprende una secuencia señal según la invención. Tal vector o casete de expresión comodín está preparado para incorporar un gen de interés que codifica una PDI. La línea celular comodín es, así, una línea celular hospedadora preformada, que se caracteriza por su capacidad de expresión. Esto sigue una estrategia de una plataforma "comodín" innovadora para la generación de líneas celulares productoras, para la producción de una PDI, p.ej., mediante el uso de la integración de un casete específico de sitio o del intercambio de casetes mediado por una recombinasa específica de sitio. Tal célula hospedadora nueva facilita la clonación de un gen de interés (GDI), p.ej. en sitios predeterminados de expresión genómica para conseguir líneas celulares de producción sumamente eficaces y reproducibles.

Según una realización preferida, el vector de expresión descrito en la presente memoria es un plásmido adecuado para la integración en el genoma de la célula hospedadora, en una única copia o en copias múltiples por célula. La secuencia de nucleótidos recombinante que codifica una PDI se puede proporcionar también en un plásmido con replicación autónoma en una única copia o en copias múltiples por célula. El plásmido preferido es un vector de expresión eucariótica, preferiblemente un vector de expresión en levadura.

Los vectores de expresión pueden incluir, pero sin limitación, vectores de clonación, vectores de clonación modificados y plásmidos diseñados específicamente. El vector de expresión preferido, tal como se usa en la invención, puede ser cualquier vector de expresión adecuado para la expresión de un gen recombinante en una célula hospedadora, y se selecciona dependiendo del organismo hospedador. El vector de expresión recombinante puede ser cualquier vector que es capaz de replicarse o integrarse en el genoma de los organismos hospedadores, también denominado vector hospedador, tal como un vector de levadura, que porta un constructo de ADN descrito en la presente memoria. Un vector de expresión en levadura preferido es adecuado para la expresión en una levadura seleccionada del grupo que consiste en levaduras metilotróficas representadas por el género *Hansenula*,



Ogatea, Pichia, Candida y Torulopsis.

En la presente invención, se prefiere usar plásmidos derivados de pPICZ, pGAPZ, pPIC9, pPICZalfa, pGAPZalfa, pPIC9K, pGAPHis o pPUZZLE como vector.

5 Según una realización preferida, se obtiene un constructo recombinante ligando los genes relevantes en un vector. Estos genes se pueden integrar de manera estable en el genoma de la célula hospedadora transformando la célula hospedadora mediante el uso de tales vectores. Los polipéptidos codificados por los genes se pueden producir mediante el uso de la línea celular hospedadora recombinante cultivando un transformante, obtenidos así en un medio adecuado, aislando la PDI expresada del cultivo, y purificándola mediante un método adecuado para el producto expresado, en particular para separar la PDI de las proteínas contaminantes.

10 La secuencia de ADN que codifica la PDI también se puede conectar de manera operable a una secuencia terminadora adecuada, por ejemplo, al terminador de AOX1 (alcohol oxidasa), terminador de CYC1 (citocromo c), terminador de TEF (factor de elongación de la traducción).

15 Los vectores de expresión pueden comprender uno o más marcadores seleccionables fenotípicos, p.ej. un gen que codifica una proteína que confiere resistencia a antibióticos o que suplementa una necesidad auxotrófica. Los vectores de levadura contienen habitualmente un origen de replicación de un plásmido de levadura, una secuencia de replicación autónoma (ARS), o, de manera alternativa, una secuencia usada para la integración en el genoma del hospedador, una región promotora, secuencias para la poliadenilación, secuencias para la terminación de la transcripción, y un marcador seleccionable.

20 Los procedimientos usados para ligar las secuencias de ADN, p.ej. que codifican la secuencia líder y/o la PDI, el promotor y el terminador, respectivamente, y para insertarlos en vectores adecuados que contienen la información necesaria para la integración o la replicación en el hospedador, son muy conocidos para las personas expertas en la técnica, y se describen, p.ej., en J. Sambrook et al., "Molecular Cloning 2nd ed.", Cold Spring Harbor Laboratory Press (1989).

25 Se entenderá que el vector se puede construir preparando primero un constructo de ADN que contiene la secuencia de ADN completa de los elementos necesarios del vector e insertando esta construcción en un vector de expresión adecuado, o insertando de manera secuencial los fragmentos de ADN que contienen la información genética para los elementos individuales, tal como la señal, el líder o la PDI, seguido de ligadura. De manera alternativa, los elementos individuales del casete de expresión también se pueden fusionar mediante PCR.

30 También se pueden usar vectores de multiclonación, que son vectores que tienen un sitio de multiclonación, en los que se puede incorporar un gen deseado en un sitio de multiclonación para proporcionar un vector de expresión. En los vectores de expresión, el promotor se coloca en posición 5' del gen de la secuencia líder y la PDI, y regula la expresión del gen. En el caso de los vectores de multiclonación, debido a que el gen de la secuencia líder y la PDI se introducen en el sitio de multiclonación, el promotor se coloca en posición 5' del sitio de multiclonación. El gen de la secuencia líder se puede fusionar al gen de la PDI durante una reacción de PCR o una preparación sintética, o el gen de la secuencia líder se puede proporcionar en el vector o casete de expresión y el gen de la PDI se puede introducir mediante procedimientos de clonación habituales.

35 El vector o casete de expresión y la construcción de ADN tal como se proporciona según la invención se pueden preparar de manera sintética mediante los métodos habituales establecidos, p.ej. el método de fosforamidita. La construcción de ADN puede ser también de origen genómico o de cADN, por ejemplo obtenido preparando una biblioteca genómica o de cADN y cribando con respecto a las secuencias de ADN que codifican todo o parte del polipéptido mediante hibridación con el uso de sondas oligonucleotídicas sintéticas de acuerdo con las técnicas habituales (Sambrook et al., Molecular Cloning: A Laboratory Manual, Cold Spring Harbor, 1989). Finalmente, la construcción de ADN puede ser de origen mixto sintético y genómico, mixto sintético y de cADN o mixto genómico y de cADN preparada renaturalizando fragmentos de origen sintético, genómico o de cADN, según sea adecuado, y los fragmentos corresponden a las diversas partes de la construcción de ADN completa, de acuerdo con las técnicas habituales.

40 Es evidente para los expertos en la técnica que las secuencias de ADN que codifican el líder y/o la PDI se pueden optimizar con respecto a la preferencia de uso de codones del organismo hospedador, según los algoritmos y los métodos del estado de la técnica (p.ej., como proporcionan los proveedores comerciales tales como, por ejemplo, GeneArt, GeneCust, GenScript o DNA2.0).

45 En otra realización preferida, el vector o casete de expresión en levadura es capaz de integrarse de manera estable en el genoma de la levadura, p.ej. mediante recombinación homóloga.

50 En un aspecto preferido, la invención se refiere a un método en el que el vector o casete de expresión comprende una secuencia líder eficaz para provocar la secreción de una PDI madura mediante la célula hospedadora transformada.

55 La secuencia señal o el líder truncado descrito en la presente memoria se puede fusionar a la secuencia de

nucleótidos que codifica una PDI destinada a la expresión recombinante mediante técnicas de clonación convencionales conocidas para una persona experta. En las realizaciones preferidas, la secuencia de nucleótidos de una PDI se fusiona a la secuencia de nucleótidos del líder de secreción, así, la secuencia señal o el líder truncado va a dirigir la proteína a la ruta secretora en la que el líder se escinde y la proteína se libera al medio.

- 5 Una célula hospedadora transformante según la invención obtenida transformando la célula con el vector o casete de expresión según la invención se puede cultivar primero preferiblemente en condiciones para hacerla crecer de manera eficaz hasta un número de células elevado sin la carga de expresar la PDI. Cuando la línea celular está preparada para la expresión de la PDI, se eligen las técnicas de cultivo para producir el producto de expresión.

- 10 Las estrategias de fermentación diferenciarían entre una fase de crecimiento y una fase de producción. El crecimiento y/o la producción pueden tener lugar de manera adecuada en modo discontinuo, modo semicontinuo o modo continuo. Se puede usar cualquier biorreactor adecuado, que incluye un reactor discontinuo, semicontinuo, continuo, de tanque agitado, o reactor con columna de burbujas.

- 15 Es ventajoso proporcionar el proceso de fermentación a escala piloto o industrial. La escala de proceso industrial emplearía preferiblemente volúmenes de al menos 10 L, específicamente al menos 50 L, preferiblemente al menos 1 m<sup>3</sup>, preferiblemente al menos 10 m<sup>3</sup>, lo más preferiblemente al menos 100 m<sup>3</sup>.

Se prefieren las condiciones de producción a escala industrial, que se refieren, p.ej., al cultivo semicontinuo en volúmenes de reactor de 100 L a 10 m<sup>3</sup> o más, que emplean tiempos de proceso típicos de varios días, o procesos continuos en volúmenes de fermentador de aprox. 50 - 1000 L o más, con velocidades de dilución de aproximadamente 0,02 - 0,4 h<sup>-1</sup>.

- 20 Las técnicas de cultivo adecuadas pueden abarcar el cultivo en un biorreactor comenzando con una fase discontinua, seguida de una fase semicontinua exponencial corta a una velocidad de crecimiento específica elevada, seguida además de una fase semicontinua a una velocidad de crecimiento específica inferior. Otra técnica de cultivo adecuada puede abarcar una fase discontinua seguida de una fase de cultivo continua a una velocidad de dilución inferior. Una realización preferida de la invención incluye un cultivo discontinuo para proporcionar la biomasa seguido de un cultivo semicontinuo para un rendimiento elevado de producción de PDI.

Se prefiere cultivar la línea celular hospedadora según la invención en un biorreactor en condiciones de crecimiento para obtener una densidad celular de al menos 1 g/L de peso seco de células, más preferiblemente al menos 10 g/L de peso seco de células, preferiblemente al menos 20 g/L de peso seco de células. Es ventajoso proporcionar tales rendimientos de producción de biomasa a escala piloto o industrial.

- 30 Se entiende que los métodos descritos en la presente memoria pueden incluir además cultivar dichas células hospedadoras recombinantes en condiciones que permiten la expresión de la PDI. Después se puede aislar una PDI producida de manera recombinante unida a membranas o soluble, o un metabolito de la célula hospedadora, del medio de cultivo celular y purificarlo adicionalmente mediante métodos muy conocidos para una persona experta en la técnica.

- 35 Se prefieren varias aproximaciones diferentes para la expresión y secreción de la PDI por la célula hospedadora. Las proteínas se expresan, se procesan y se secretan transformando el organismo eucariótico con un vector o casete de expresión según la invención que alberga un ADN que codifica la proteína deseada, preparando un cultivo del organismo transformado, haciendo crecer el cultivo y recuperando la proteína del medio de cultivo, p.ej. concentrando y enriqueciendo la proteína en una fracción del cultivo celular, o purificando la proteína para obtener una preparación sustancialmente pura. Además, se pueden aplicar células hospedadoras con delecciones de una o más de las proteínas contaminantes principales de las células hospedadoras (p.ej., como se describió en Heiss et al. 2012; Appl Microbiol Biotechnol. doi: 10.1007/s00253-012-4260-4) para facilitar la purificación de la PDI.

- 45 Como métodos de aislamiento y purificación para obtener un polipéptido recombinante o producto proteico, se pueden usar métodos, tales como los métodos que utilizan una diferencia de solubilidad, tales como desalación y precipitación con disolventes, métodos que utilizan una diferencia de peso molecular, tales como ultrafiltración y electroforesis en gel, métodos que utilizan una diferencia de carga eléctrica, tales como cromatografía de intercambio iónico, métodos que utilizan una afinidad específica, tales como cromatografía de afinidad, métodos que utilizan una diferencia de hidrofobicidad, tales como cromatografía líquida de alto rendimiento en fase inversa, y métodos que utilizan una diferencia del punto isoeléctrico, tales como isoelectroenfoque.

- 50 El producto altamente purificado está básicamente exento de proteínas contaminantes, y preferiblemente tiene una pureza de al menos un 90%, más preferiblemente al menos un 95%, o incluso al menos un 98%, hasta un 100%. Los productos purificados se pueden obtener mediante purificación del sobrenadante de cultivo celular o de los restos celulares.

- 55 La PDI aislada y purificada se puede identificar mediante métodos convencionales tales como transferencia de Western o ensayo de su actividad. La estructura del compuesto purificado se puede definir mediante análisis de los aminoácidos, análisis aminoterminal, análisis de la estructura primaria, y similares. Se prefiere que el compuesto sea obtenible en gran cantidad y a un nivel elevado de pureza, cumpliendo así los requerimientos necesarios para usarlo

como ingrediente activo en composiciones farmacéuticas.

5 La línea celular hospedadora preferida según la invención mantiene la integración de las secuencias reguladoras según la invención y del gen de PDI, y el nivel de expresión permanece elevado, p.ej. al menos a un nivel de  $\mu\text{g/L}$ , incluso tras alrededor de 20 generaciones de cultivo, preferiblemente al menos 30 generaciones, más preferiblemente al menos 40 generaciones, lo más preferiblemente al menos 50 generaciones. La célula hospedadora recombinante es sorprendentemente estable, lo cual es una gran ventaja al usarla para la producción de proteínas a escala industrial.

La presente invención se describe con más detalle en los ejemplos siguientes, que no pretenden limitar de ninguna manera el alcance de la invención tal como se reivindica.

10 **Ejemplos**

Ejemplo 1: Construcción de una línea de células hospedadoras de *P. pastoris* para la expresión de proteínas recombinantes mediante el uso del líder nativo de Epx1 de *P. pastoris* (SEQ ID 8) para la secreción

1a): Construcción de un vector de expresión que contiene el líder nativo de Epx1 de *P. pastoris* (secuencia señal y pro-secuencia; SEQ ID 8, EpxL-RT, la "secuencia precursora")

15 La identificación de la proteína secretada de manera nativa Epx1 en *P. pastoris* y la identificación de la supuesta secuencia líder de secreción EpxL-RT (que consiste en la secuencia señal de Epx1 y la pro-secuencia hasta el extremo N-terminal determinado experimentalmente de la proteína Epx1 madura) se describió en el documento EP2258855. Como los últimos aminoácidos que precedían al extremo N-terminal verificados de manera experimental de Epx1 madura fueron Arg-Thr (RT), la supuesta secuencia líder de secreción se denominó EpxL-RT.

20 Para generar un vector de expresión mediante el uso de la secuencia líder de secreción de la proteína Epx1 de *P. pastoris* para la secreción de una PDI, se clonó la secuencia respectiva (SEQ ID 8) en el marco de lectura con el promotor PGAP (promotor de gliceraldehído 3-fosfato deshidrogenasa) del vector de expresión pPM2\_pGAP. El vector de expresión pPM2\_pGAP es un derivado del esqueleto del vector pPuzzle\_zeoR\_AOXTT descrito en el documento WO2008/128701A2, que consiste en el origen de replicación bacteriano de pUC19, el promotor de gliceraldehído fosfato deshidrogenasa de *P. pastoris*, el terminador de la transcripción de CYC1 de *S. cerevisiae*, y el casete de resistencia al antibiótico zeocina.

La secuencia líder de secreción de Epx1 se amplificó mediante PCR a partir del ADN genómico de *P. pastoris* mediante el uso de los cebadores de oligonucleótidos (Tabla 1).

30 Tabla 1: Cebador de oligonucleótidos para la amplificación mediante PCR de la secuencia precursora de la proteína Epx1 de *P. pastoris* EpxL-RT (los sitios de restricción están subrayados)

EpxL dir.	Sbfl	SEQ ID 22 TATACCTGCAGGATGAAGTTCTCTACCAATTTGATC
EpxL inv.	Nsil	SEQ ID 23 GAAGATGCATCGTACGGTAGACAGTGACAAC

35 Posteriormente, el producto de PCR (189 pb) se digirió mediante las enzimas de restricción Sbfl y Nsil; y se ligó en un vector pPM2\_pGAP que se había linealizado mediante Sbfl y se había tratado con fosfatasa de intestino de ternero (CIP). Como Nsil y Sbfl producen extremos solapantes, la construcción resultante pPM2\_pGAPxLRT tiene un único sitio Sbfl directamente antes del codón de inicio de la secuencia EpxL-RT. Se verificó la integración correcta mediante digestión de los plásmidos resultantes con Sbfl y Ascl.

1b) Construcción de una cepa de *P. pastoris* que secreta tripsinógeno porcino recombinante mediante el uso del vector pPM2\_pGAPxLRT.

40 Para la expresión de tripsinógeno porcino recombinante (rpTRP) se sintetizó un gen artificial con optimización de codones (Geneart, Alemania). Se añadieron sitios para las enzimas de restricción Pfl23II y Sfil flanqueando el marco de lectura abierto durante la amplificación mediante PCR con el uso del plásmido administrado como molde, y los cebadores se muestran en la Tabla 2.1. El producto de PCR se digirió con Pfl23II y Sfil, y se clonó en un plásmido pPM2\_pGAPxLRT tratado con Pfl23II, Sfil y CIP (Ejemplo 1a). El plásmido ligado se transformó en *E. coli* TOP10 (Invitrogen) y se colocó en placas con LB-agar que contenía Zeocina. Se llevó a cabo un análisis con endonucleasas de restricción para confirmar la identidad correcta del plásmido pPM2\_pGAPxLRT\_rpTRP.

Tabla 2.1: Cebador de oligonucleótidos para la amplificación mediante PCR del gen de tripsinógeno porcino

EpxL-RT-pTRP dir.	Pfl23II (BsiWI)	SEQ ID 24 ATACCGTACGACTGACGACGACGACAAG
pTRP inv.	Sfil	SEQ ID 25 TTTTGGCCGAGGCGGCCCTTTCAGTTAGCAGCGATAGTTTG

1c) Construcción de una cepa de *P. pastoris* que sobreexpresa albúmina de suero humano recombinante o proteína fluorescente verde mejorada con el vector pPM2\_pGAPxLRT.

- 5 Los genes que codifican la albúmina de suero humano (HSA) o la proteína fluorescente verde mejorada (eGFP) se amplificaron mediante PCR de los vectores descritos en Stadlmayr et al. (2010, J Biotechnol. 150: 519-529) mediante el uso de los cebadores mostrados en la Tabla 2.2. Los productos de PCR se digirieron con Accl y Sfil, y se clonaron en el plásmido pPM2\_pGAPxLRT tratado con Accl y Sfil (Ejemplo 1a). El plásmido ligado se transformó en *E. coli* TOP10 (Invitrogen) y se colocó en placas con LB-agar que contenía Zeocina. Tras la verificación de la
- 10 secuencia, los vectores pPM2\_pGAPxLRT-HSA y pPM2\_pGAPxLRT-eGFP se linealizaron en la región del promotor y se transformaron en *P. pastoris*. Para HSA, se usó como referencia la construcción que usó el líder de HSA nativo descrito en Stadlmayr et al. 2010, mientras eGFP se clonó tras el líder de MFα de *S. cerevisiae* como control.

Tabla 2.2: Cebadores de oligonucleótidos para la amplificación mediante PCR de HSA y eGFP fusionadas a EpxL-RT y eGFP fusionada al líder de MFα de *S. cerevisiae*

EpxL-RT-HSA dir.	Accl	SEQ ID 26 AGGCGTCTACCGAACTGATGCACACAAGAGTGAGGTT
HSA inv.	Sfil	SEQ ID 27 GAGTGGCCGAGGCGGCCCTTATAAGCCTAAGGCAGCTTGA
EpxL-RT-eGFP dir.	Accl	SEQ ID 28 ATTTGTCTACCGAACTGTGAGCAAGGGCGAGGAGC
eGFP inv.	Sfil	SEQ ID 29 CGTTGGCCGAGGCGGCCCTTACTTGTACAGCTCGTCCATG

- 15 Ejemplo 2: Cultivo de una línea celular hospedadora de *P. pastoris* para la expresión de proteínas secretoras recombinantes y análisis del producto

20 Todos los plásmidos se linealizaron en su región promotora respectiva o en la región de integración AOX-TT antes de la electrotransformación en *P. pastoris*. Los transformantes positivos se seleccionaron en placas de agar YPD que contenían extracto de levadura (10 g/L), peptona (20 g/L), glucosa (20 g/L) y Zeocina.

2a) Cultivo de cepas de *P. pastoris* transformadas que expresan proteínas secretoras recombinantes en cultivos a pequeña escala

25 Se inocularon 5 mL de medio YP (10 g/L de extracto de levadura, 20 g/L de peptona) que contenían 10 g/L de glicerol con una única colonia de cepas de *P. pastoris* de los Ejemplos 1b, 5, 7, o 9 y se cultivaron durante la noche a 28 °C. Se transfirieron alícuotas de estos cultivos (que correspondían a una DO<sub>600</sub> final de 0,1) a 10 mL de medio de cultivo de expresión (la composición de los medios se proporciona más adelante para cada proteína recombinante) complementado con 20 g/L de glucosa y se incubó durante 48 h a 28 °C a 170 rpm en matraces Erlenmeyer de 100 mL. De manera alternativa, se usaron 2 mL de medio de cultivo de expresión para el cultivo en placas de 24 pocillos profundos. Se añadió glucosa (10 g/L) repetidamente cada 12 h, antes de recoger las células mediante

30 centrifugación a 2500xg durante 10 min a temperatura ambiente, y se prepararon para el análisis. Para la expresión controlada por pG1, se alcanzaron las condiciones de crecimiento limitante con glucosa mediante el uso de esferas de suministro de glucosa (Kühner, CH), que liberan lentamente glucosa a lo largo del tiempo según la ecuación (Glucosa)=1,63\*t<sup>0.74</sup> [mg/Disco], en vez de la complementación con glucosa. Para 10 mL de cultivo principal, se usaron 2 esferas de suministro. Se determinó la biomasa midiendo el peso de las células tras la centrifugación de 1

35 mL de suspensión celular, mientras la determinación de la proteína secretada recombinante en el sobrenadante se describe en los siguientes Ejemplos 2b-2e.

Los medios de cultivo de expresión fueron como sigue:

Para tripsinógeno porcino: por litro: 10 g de extracto de levadura, 10 g de peptona de guisante, 10,2 g de (NH<sub>4</sub>)<sub>2</sub>PO<sub>4</sub>, 1,24 g de KCl, 0,1 g de CaCl<sub>2</sub>, pH 5,0 ajustado con HCl.

Para HSA: por litro: 22 g de ácido cítrico, 3,15 g de  $(\text{NH}_4)_2\text{HPO}_4$ , 0,027 g de  $\text{CaCl}_2 \cdot 2\text{H}_2\text{O}$ , 0,9 g de KCl, 0,5 g de  $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$ , 2 ml de biotina 500 x y 1,47 mL de disolución de reserva de sales traza [por litro: 6 g de  $\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$ , 0,08 g de NaI, 3 g de  $\text{MnSO}_4 \cdot \text{H}_2\text{O}$ , 0,2 g de  $\text{Na}_2\text{MoO}_4 \cdot 2\text{H}_2\text{O}$ , 0,02 g de  $\text{H}_3\text{BO}_3$ , 0,5 g de  $\text{CoCl}_2$ , 20 g de  $\text{ZnCl}_2$ , 5 g de  $\text{FeSO}_4 \cdot 7\text{H}_2\text{O}$  y 5 mL de  $\text{H}_2\text{SO}_4$ ]; pH ajustado a 6 con KOH 5 M; esterilizado mediante filtración.

- 5 Para eGFP y para los fragmentos de anticuerpo (p.ej. Fab): por litro: 10 g de extracto de levadura, 10 g de peptona, tampón de fosfato potásico 100 mM de pH 6,0, 13,4 g de base de nitrógeno de levadura con sulfato amónico, 0,4 mg de biotina.

2b) Cuantificación de tripsina

- 10 El sobrenadante de cultivo de *P. pastoris* se desaló mediante el uso de columnas de cromatografía de exclusión por tamaño pre-empaquetadas pequeñas (Columnas de Desalación PD-10 Desechables 17-0851-01; GEHealthcare). Se aplicaron 2,5 ml de sobrenadante a la columna y se eluyó con 3,5 ml de tampón de elución (HCl 1 mM). Tras la elución, se añadieron 70  $\mu\text{l}$  de disolución de  $\text{CaCl}_2$  2 M.

- 15 Para convertir el tripsinógeno inactivo en la tripsina activa, se mezclaron 300  $\mu\text{l}$  del sobrenadante desalado (+ $\text{CaCl}_2$ ) con 690  $\mu\text{l}$  de tampón de activación (TRIS/HCl 50 mM pH 8,6;  $\text{CaCl}_2$  40 mM y 0,15 g/L de enteroquinasa, Sigma; E0632) y se incubó durante dos horas a 37 °C.

- 20 Se mezclaron 165  $\mu\text{l}$  de la mezcla de activación con 1000  $\mu\text{l}$  de disolución TAME, que contenía 446 mg/L de hidrocloreuro de éster metílico de Na-p-Tosil-L-arginina (TAME; Sigma; T4626) disuelto en tampón de dilución (TRIS/HCl 50 mM de pH 8,1;  $\text{CaCl}_2$  40 mM), y se midió la cinética de absorción a 247 nm en un espectrofotómetro a lo largo de un periodo de tiempo de 5 min a 30 °C. Si fue necesario, se diluyó la disolución de tripsina activada con tampón de dilución para alcanzar el intervalo lineal ( $\Delta A_{247}/\text{min} < 0,3$ ) de este método. Una concentración de tripsina de 1 g/L corresponde a  $\Delta A_{247}/\text{min} = 0,101$ .

2c) Cuantificación de HSA mediante ELISA

- 25 Para la cuantificación de HSA en sobrenadantes de *P. pastoris*, se usó el Kit de Cuantificación mediante ELISA de Albúmina Humana (Nº Cat. E80-129, Bethyl Laboratories, TX, EE.UU.). Se usó el patrón de HSA con una concentración de partida de 400 ng/mL. Las muestras de sobrenadante se diluyeron en consecuencia.

2d) Análisis mediante SDS-PAGE y Transferencia de Western

- 30 Para el análisis en geles de proteínas se usó el sistema NuPAGE® Novex® Bis-Tris, mediante el uso de geles de un 12 % de Bis-Tris o 4-12 % de Bis-Tris y tampón de análisis MOPS (todos de Invitrogen). Después de la electroforesis, las proteínas se visualizaron mediante tinción de plata o se transfirieron a una membrana de nitrocelulosa para el análisis mediante transferencia de Western. Por lo tanto, las proteínas se electrotransfirieron a una membrana de nitrocelulosa mediante el uso del Módulo de Transferencia XCell II™ para la transferencia húmeda (en tanque) (Invitrogen) según las instrucciones del fabricante. Después del bloqueo, las transferencias de Western se sondearon con los anticuerpos siguientes: Para HSA: conjugado anti-albúmina de suero humano-peroxidasa de rábano (HRP), Bethyl A80-129P (1:50.000); Para la cadena ligera de IgG: anticuerpo conjugado anti-cadenas ligeras kappa humanas (unidas y libres) - fosfatasa alcalina (AP), Sigma A3813 (1:5.000); Para la cadena pesada de IgG: anticuerpo anti-IgG humana (específico de cadena  $\gamma$ ) producido en cabra, Sigma I3382 (1:5.000) y conjugado anti-cabra AP (1:20.000).

- 40 La detección se llevó a cabo con el kit de detección colorimétrica de AP (BioRad) basado en el sistema NBT/BCIP para conjugados con AP, y el Sustrato Quimioluminiscente Super Signal West (Thermo Scientific) para conjugados con HRP.

2e) Cuantificación de Fab mediante ELISA

- 45 La cuantificación de Fab intacto se realizó mediante ELISA con el uso de un anticuerpo anti-IgG humana (Abcam ab7497) como anticuerpo de revestimiento (1:1.000), y un anticuerpo conjugado anti-Cadenas Ligeras Kappa Humanas (Unidas y Libres) de cabra - fosfatasa alcalina (Sigma A3813) como anticuerpo de detección (1:1.000). Se usó un fragmento de IgG, Fab/Kappa humano (Bethyl P80-115), como patrón con una concentración de partida de 50 ng/mL. Las muestras de sobrenadante se diluyeron en consecuencia. La detección se realizó con el sustrato pNPP (Sigma S0942). Los tampones de revestimiento, dilución y lavado se basaron en PBS ( $\text{KH}_2\text{PO}_4$  2 mM,  $\text{Na}_2\text{HPO}_4 \cdot 2\text{H}_2\text{O}$  10 mM, KCl 2,7 mM, NaCl 8 mM, pH 7,4) y se completaron con BSA (1% (p/v)) y/o Tween20 (0,1% (v/v)) en consecuencia.

- 50 Ejemplo 3: Expresión de proteínas recombinantes mediante una línea de células hospedadoras de *P. pastoris* mediante el uso de la secuencia precursora de Epx1 de *P. pastoris* (EpxL-RT, SEQ ID 8) para la secreción

3a) Análisis de *P. pastoris* que sobreexpresa tripsinógeno porcino recombinante mediante el uso de EpxL-RT para la secreción

Se cultivó *P. pastoris* que expresaba pTRP mediante el uso de la secuencia EpxL-RT para la secreción como se

describió en el Ejemplo 2a, y se analizó mediante el uso del ensayo de actividad de tripsina como se describió en el Ejemplo 2b y SDS-PAGE como se describió en el Ejemplo 2d. Se usó *P. pastoris* que expresaba pTRP mediante el uso de MF $\alpha$  para la secreción como referencia.

5 De manera inesperada, la expresión de pTRP mediante el uso del líder de secreción EpxL-RT condujo a una mancha de proteínas de aproximadamente 30 kDa (Figura 2.1, lado izquierdo), mientras MF $\alpha$  condujo a una pTRP secretada del tamaño correcto (25 kDa; Figura 2.1, lado derecho). Partes del pTRP secretado con EpxL-RT también condujeron a un pTRP del tamaño correcto, aunque en menor grado. Las cantidades de pTRP secretadas con EpxL-RT fueron menores del 50% de la cantidad secretada mediante el uso de MF $\alpha$  (Tabla 3, medida mediante el ensayo de actividad de tripsina como se describió en el Ejemplo 2b). La banda extendida probablemente representa una proteína de fusión EpxL-RT-pTRP, que parece debida a un procesamiento incorrecto (escisión) de la secuencia EpxL-RT.

Tabla 3: Niveles relativos de secreción de pTRP normalizados respecto de alfaMF

Líder	Secreción de pTRP media relativa $\pm$ EEM
EpxL-RT	0,48 $\pm$ 0,02
MF $\alpha$	1,00 $\pm$ 0,05

### 3b) Análisis de *P. pastoris* que sobreexpresa eGFP recombinante mediante el uso de EpxL-RT para la secreción

15 Se cultivó *P. pastoris* que expresaba eGFP mediante el uso de la secuencia EpxL-RT para la secreción como se describió en el Ejemplo 2a y se analizó mediante el uso de SDS-PAGE y transferencia de Western (Ejemplo 2d). Aparentemente, el EpxL-RT nativo no fue capaz de secretar eGFP completamente (Figura 2.2, parte izquierda), en contraste con MF $\alpha$  (Figura 2.2, parte derecha). De manera intracelular, se pudo observar eGFP todavía unida a partes de la secuencia EpxL-RT, lo que descarta los defectos en la expresión de la proteína (no mostrado).

20 3c) Análisis de *P. pastoris* que sobreexpresa albúmina de suero humano recombinante mediante el uso de EpxL-RT para la secreción

Se cultivó *P. pastoris* que expresaba HSA mediante el uso de la secuencia EpxL-RT para la secreción como se describió en el Ejemplo 2a y se analizó mediante el uso de ELISA de HSA (Ejemplo 2c) y transferencia de Western (Ejemplo 2d). Se usó *P. pastoris* que expresaba HSA mediante el uso del líder de HSA nativo para la secreción como referencia (Kobayashi. 2006. Biologicals. 34(1):55-9.).

25 Se analizaron 12 clones individuales de cada construcción por su comportamiento de secreción después de 48 h de cultivo en matraz agitado en medio de cribado sintético. El sobrenadante se examinó de manera cualitativa mediante SDS-PAGE reductora y transferencia de Western posterior, con el uso del anticuerpo anti-HSA para la detección de HSA. Fue visible un patrón de bandas dobles diferentes para la HSA secretada mediante EpxL-RT (Figura 2.3, lado izquierdo). En comparación con la HSA secretada mediante su líder nativo (Figura 2.3, lado derecho) y la HSA purificada (no mostrada), la banda inferior fue la HSA procesada correctamente. La banda superior fue inesperada, pero debido a su peso molecular ligeramente mayor, puede representar una proteína de fusión EpxL-RT-HSA debida a un EpxL-RT procesado de manera incorrecta. Basándose en la intensidad de las bandas (tal como se analiza mediante ImageJ), hubo presente un 40-60% de la HSA secretada total en la forma de peso molecular incorrecto mayor.

Resumen: Cuando se usa la secuencia EpxL-RT para la secreción, es visible una mancha o patrón de bandas dobles para la proteína recombinante secretada, que es indicativo de un procesamiento incorrecto o de la ausencia de procesamiento del líder largo EpxL-RT. Esta característica hace que EpxL-RT que corresponde a la secuencia líder de Epx1 de longitud completa del documento EP2258855A1 sea inadecuada e inútil para la secreción de proteínas recombinantes.

Ejemplo 4: Identificación del extremo N-terminal de la banda de peso molecular mayor de cultivos de expresión de HSA

Se analizó adicionalmente el extremo N-terminal de la HSA expresada como se describió en el Ejemplo 3c anterior mediante el uso de EpxL-RT mediante secuenciación N-terminal.

45 Por lo tanto, se cargaron 500  $\mu$ l del sobrenadante respectivo en un filtro centrífugo para concentrar la proteína (filtro centrífugo de 10 kDa 0,5 ml Amicon Ultra, Millipore, UFC5010), se centrifugó durante 5 min y se recuperaron 15  $\mu$ l de muestra mediante centrifugación inversa.

50 A continuación, las muestras se separaron mediante un gel del 4-12 % de Bis-Tris NuPAGE y una transferencia de Western con tampón borato (por litro: 3,09 g de Borato [50 mM], 100 ml de MeOH a pH 9 ajustado con NaOH 1 M) llevado a cabo durante 2 h a 25 V mediante el uso de una membrana de PVDF (poli(fluoruro de vinilideno)). Antes de la transferencia, el gel se incubó en tampón borato durante 10 min, mientras la membrana se sumergió en metanol

durante 30 seg, seguido de 3 min en tampón borato.

Después de la transferencia, la membrana se tiñó durante 3 min con Coomassie (0,1 % [p/v] de R250, MeOH [40 % v/v], ácido acético [10 % v/v]), seguido de decoloración (MeOH [40 % v/v], ácido acético [10 % v/v]). La membrana se lavó con ddH<sub>2</sub>O y las bandas superiores e inferiores (Figura 3) se cortaron y se enviaron para la secuenciación de EDMAN N-terminal. Para la banda inferior, se determinó D como el aminoácido N-terminal, de manera concluyente con el primer aminoácido de HSA (extremo N-terminal de HSA, SEQ ID 30: DAHKSEV), mientras para la banda superior se determinó AYYT (SEQ ID 31) como extremo N-terminal de la proteína secretada. Estos aminoácidos son parte de la secuencia EpxL-RT (SEQ ID 8), que deja un saliente indeseado de 21 aminoácidos en HSA. La secuencia de EpxL-RT que precede a AYYT (SEQ ID 23) es KR, que podría ser parte de un motivo de escisión para peptidasa Lys-Arg dibásico. El procesamiento de los motivos Lys-Arg dibásicos mediante proteasas tales como Kex2 depende no solamente del motivo propiamente dicho, sino también de la estructura tridimensional y del medio de los aminoácidos circundantes (Bader et al. 2008, BMC Microbiol. 8: 116.). El análisis mediante BLAST y la alineación de secuencias de EpxL-RT con las secuencias de homólogos de Epx1 de otras levaduras (p.ej. *Saccharomyces cerevisiae*, género *Candida*) reveló que el motivo KR de la secuencia líder no está conservado. Por lo tanto, el procesamiento del motivo KR mediante proteasas tales como Kex2 no se pudo dar por hecho *a priori* basándose en el análisis de las secuencias.

Ejemplo 5: Construcción de una línea celular hospedadora de *P. pastoris* para la expresión de proteínas recombinantes mediante el uso de la secuencia EpxL-KR de *P. pastoris* (líder truncado, SEQ ID 10, en el que X en la posición 3 es F, y X en la posición 16 es A (SEQ ID 12)) para la secreción

Para ensayar si la secuencia KR de EpxL-RT es realmente un sitio de escisión de proteasas, se construyeron vectores para la expresión de proteínas recombinantes mediante el uso de la secuencia EpxL-KR para la secreción.

5a) Construcción de una cepa de *P. pastoris* que sobreexpresa HSA mediante el uso de la secuencia EpxL-KR para la secreción.

Se amplificó HSA mediante el uso de los cebadores de la Tabla 4.1, y se ligó en el vector pPM2\_pGAPxLRT digerido con BglII y SbfII. Tras la verificación de la secuencia, se linealizó el vector pPM2\_pGAPxLKR-HSA en la región del promotor y se transformó en *P. pastoris*.

Tabla 4.1: Cebadores de oligonucleótidos para la amplificación mediante PCR de HSA fusionada a EpxL-KR (los sitios de restricción están subrayados, la secuencia EpxL está en cursiva)

EpxL-KR_HSA dir.	BglII	SEQ ID 32 ATTC <u>GCCGAAGAGGCAGCAAACCACTTGCACAAGCG</u> TGATGCACACAAGAGTGAGGTT
HSA inv.	Sfil	SEQ ID 33 GAGT <u>GGCCGAGGCGGCC</u> TTATAAGCCTAAGGCAGCT TGA

5b) Construcción de una cepa de *P. pastoris* que sobreexpresa tripsinógeno porcino o eGFP mediante el uso de la secuencia EpxL-KR para la secreción.

Se amplificó pTRP mediante el uso de los cebadores de la Tabla 4.2, y se ligó en el vector pPM2\_pGAPxLRT digerido con PvuII y SbfII. Se generó el vector de expresión pPM2\_pGAPxLKR-eGFP de la misma manera. Tras la verificación de la secuencia, los vectores pPM2\_pGAPxLKR-pTRP y pPM2\_pGAPxLKR-eGFP se linealizaron en la región del promotor y se transformaron en *P. pastoris*.

Tabla 4.2: Cebadores de oligonucleótidos para la amplificación mediante PCR de pTRP o eGFP fusionada a EpxL-KR (los sitios de restricción están subrayados, la secuencia EpxL está en cursiva)

EpxL-KR-pTRP dir.	PvuII	SEQ ID 34 ATAC <u>CAGCTGCTCCAGTTGCTCCAGCCGAAGAGGCAGCA</u> AACCACCTTGCACAAGCGTACTGACGACGACACAAG
pTRP inv.	Sfil	SEQ ID 35 TTTT <u>GGCCGAGGCGGCC</u> TTTCAGTTAGCAGCGATAGTTT G
EpxL-KR-eGFP dir.	PvuII	SEQ ID 36 ATAC <u>CAGCTGCTCCAGTTGCTCCAGCCGAAGAGGCAGCA</u> AACCACCTTGCACAAGCGTGTGAGCAAGGGCGAGGAGC
eGFP inv.	Sfil	SEQ ID 37 CGTT <u>GGCCGAGGCGGCC</u> TTACTTGTACAGCTCGTCCATG

5c) Construcción de una cepa de *P. pastoris* que sobreexpresa la cadena ligera o la cadena pesada de un anticuerpo mediante el uso de la secuencia EpxL-KR para la secreción.

5 Se amplificó la cadena ligera (LC) y la cadena pesada (HC) del anticuerpo HyHEL mediante el uso de los cebadores de la Tabla 4.3, y se ligaron en el vector pPM2\_pGAPxLRT digerido con BglII y SbfII. Tras la verificación de la secuencia, los vectores pPM2\_pGAPxLKR-LC y pPM2\_pGAPxLKR-HC se linealizaron en la región del promotor y se transformaron en *P. pastoris*.

Tabla 4.3: Cebadores de oligonucleótidos para la amplificación mediante PCR de LC de HyHEL fusionada a EpxL-KR (los sitios de restricción están subrayados, la secuencia EpxL está en cursiva)

EpxL-KR-LC dir.	BglII	SEQ ID 38 <i>ATAAGCCGAAGAGGCAGCAAACCACTTGCACAAGCGTGAC</i> ATCGTTTTG
LC inv.	SbfII	SEQ ID 39 CTATGGCCGAGGCGGCCCTATTAACACTCACCTCTGTTG
EpxL-KR-HC dir.	BglII	SEQ ID 40 <i>ATAAGCCGAAGAGGCAGCAAACCACTTGCACAAGCGTGAC</i> GTTCAATTG
HC inv.	SbfII	SEQ ID 41 TATCGGCCGAGGCGGCCCTATTACTTACCTGGGGACAAG

10

Ejemplo 6: Expresión de proteínas recombinantes mediante una línea celular hospedadora de *P. pastoris* con el uso de EpxL-KR (líder truncado, SEQ ID 10, en el que X en la posición 3 es F, y X en la posición 16 es A (SEQ ID 12) para la secreción

15

6a) Análisis de *P. pastoris* que sobreexpresa tripsinógeno porcino recombinante mediante el uso de EpxL-KR para la secreción

Se cultivó *P. pastoris* que expresaba pTRP mediante el uso de la secuencia EpxL-KR para la secreción (Ejemplo 5b) como se describió en el Ejemplo 2a, y se analizó como en el Ejemplo 3a.

20

Sorprendentemente, se observó que el procesamiento, de hecho, mejoró mediante el uso del líder de secreción de Epx acortado EpxL-KR (Figura 4.1, parte media). Al contrario que la mancha de proteínas observada al usar EpxL-RT para la secreción (Figura 4.1, parte izquierda), la secreción de pTRP mediante el uso de EpxL-KR produjo una banda del tamaño correcto (Figura 4.1, parte media). El extremo N-terminal correcto de pTRP secretado mediante el uso de EpxL-KR se verificó mediante análisis de espectrometría de masas (LC-MS).

25

Por lo tanto, se separaron aprox. 10 µg de cada muestra con SDS-PAGE y las bandas de proteínas deseadas se cortaron y se digirieron en el gel. Las proteínas se carbamidometilaron en el gel. Las proteínas se digirieron con tripsina de grado de secuenciación (Roche) o con Glu-C (Sigma-Aldrich), LysC (Roche) y quimotripsina (Roche), o se analizaron sin digestión. Todas las reacciones proteolíticas se llevaron a cabo a 37 °C durante la noche. Después, las muestras se inyectaron directamente en el sistema de LC-MS (LC: Dionex Ultimate 3000 LC, MS: Bruker, amaZon ETD, equipado con la nano fuente en línea). Los péptidos se separaron en una columna C-18 (Dr. Maisch GmbH, columna de HPLC C-18 ReproSil-Pur 200\*0,1 mm, empaquetamiento de 3 µm, diámetro de poro de 200 Å, flujo: 0,4 µL/min) y se aplicó un gradiente lineal desde un 95% del disolvente A y 5% del disolvente B (Disolvente A: 0,1 % de FA en agua, 0,1 % de FA en ACCN) hasta un 32% de B en 40 min, seguido de un gradiente lineal de 15 min desde un 32% de B hasta un 75% de B que facilita la elución de los péptidos grandes, y se analizó mediante MS con adquisición dependiente de datos. Los datos se procesaron mediante el uso del programa informático Bruker estándar (análisis de datos) y el programa gratuito X!-Tandem combinado con GPM.

35

Al contrario que el extremo N-terminal correcto de pTRP secretado con EpxL-KR, se determinó que los aminoácidos EAEA quedaron en el extremo N-terminal del pTRP secretado al usar MFα para la secreción. Además, la cantidad de pTRP secretada mediante el uso de EpxL-KR fue más de un 20% mayor en comparación con el líder de secreción de MFα usado habitualmente (Tabla 5).

Tabla 5: Niveles relativos de secreción de pTRP normalizados respecto de MFα

Líder	Secreción de pTRP media relativa ± EEM
EpxL-KR	1,21 ± 0,02
MFα	1,00 ± 0,09



6b) Análisis de *P. pastoris* que sobreexpresa albúmina de suero humano recombinante mediante el uso de EpxL-KR para la secreción

5 Se cultivó *P. pastoris* que expresaba HSA mediante el uso de la secuencia EpxL-KR para la secreción (Ejemplo 5a) como se describió en el Ejemplo 2a y se analizó mediante el uso de transferencia de Western (Ejemplo 2d).

Al contrario que el patrón de bandas dobles observado al usar EpxL-RT para la secreción (Figura 2.3, parte izquierda), la secreción de HSA mediante el uso de EpxL-KR produjo una banda simple del tamaño correcto (no mostrado). Así, EpxL-KR fue capaz de secretar una HSA procesada correctamente.

6c) Análisis de *P. pastoris* que sobreexpresa eGFP mediante el uso de EpxL-KR para la secreción

10 Se cultivó *P. pastoris* que expresaba eGFP mediante el uso de la secuencia EpxL-KR para la secreción (Ejemplo 5b) como se describió en el Ejemplo 2a, y se analizó como en el Ejemplo 3b.

15 Solamente el líder EpxL-KR condujo a una eGFP de tamaño correcto (Figura 4.2, lado izquierdo), mientras los restos del líder de MFα condujeron a una eGFP de peso molecular mayor (Figura 4.2, lado derecho). El análisis mediante LC-MS (descrito en el Ejemplo 6a) verificó el extremo N-terminal correcto de la eGFP secretada mediante EpxL-KR, que califica el producto secretado con el contenido de moléculas con el extremo N-terminal correcto de al menos un 95%, preferiblemente al menos un 98%, aún más preferiblemente al menos un 99% o alrededor de un 100% (p/p). En contraste, el uso de MFα provocó que quedaran los aminoácidos EAEA como secuencia de aminoácidos adicional en el extremo N-terminal debido al procesamiento incorrecto del péptido líder mediante la aminopeptidasa Ste13. Según los tamaños de las bandas en la SDS-PAGE, la mayoría de eGFP secretada mediante MFα posee los salientes de aminoácidos incorrectos EAEA en el extremo N-terminal, mientras no se pueden observar bandas de tamaño incorrecto para la eGFP secretada mediante EpxL-KR. Como para pTRP (Ejemplo 6a), los niveles de secreción de eGFP mediante el uso de EpxL-KR fueron más de un 20% mayores en comparación con el uso de MFα (Tabla 6).

25 Tabla 6: Niveles relativos de secreción de eGFP normalizados respecto de MFα (cuantificados a partir de las intensidades de las bandas con el programa informático ImageJ)

Líder	Secreción de eGFP media relativa ± EEM
EpxL-KR	1,23 ± 0,05
MFα	1,00 ± 0,08

6d) Análisis de *P. pastoris* que sobreexpresa una cadena ligera de anticuerpo o una cadena pesada de anticuerpo mediante el uso de EpxL-KR para la secreción

30 Se cultivó *P. pastoris* que expresaba LC de HyHEL o HC de HyHEL mediante el uso de la secuencia EpxL-KR para la secreción (Ejemplo 5c) como se describió en el Ejemplo 2a y se analizó como en el Ejemplo 2d.

Ambas cadenas de anticuerpo se pudieron secretar mediante el uso de la secuencia EpxL-KR (Figura 4.3), lo que confirma EpxL-KR como líder de secreción valioso para la producción de anticuerpos en *P. pastoris*. El extremo N-terminal correcto del LC de HyHEL secretado mediante EpxL-KR se verificó mediante análisis de LC-MS como se describió en el Ejemplo 6a.

35 Ejemplo 7: Construcción de una línea celular hospedadora de *P. pastoris* para la expresión de proteínas recombinantes mediante el uso de la secuencia EpxL-AA (SEQ ID 21) para la secreción

40 Los documentos WO2010/135678 y US20110021378 describen la secuencia de aminoácidos de MKLSTNLILAIAAASAVVSAA (SEQ ID 21), es decir, los aminoácidos 1-21 de SEQ ID 8, que corresponden a los primeros 21 aminoácidos del líder de Epx1 de longitud completa. Sin embargo, no se proporcionan datos experimentales en los documentos WO2010/135678 y US20110021378, que muestran que esta secuencia sería adecuada realmente para la secreción de proteínas recombinantes. Para ensayar si este fragmento, denominado a continuación EpxL-AA, es suficiente para permitir la secreción de proteínas recombinantes correctamente procesadas, se construyeron vectores para la expresión de proteínas recombinantes mediante el uso de la secuencia EpxL-AA para la secreción.

45 La cadena ligera LC de anticuerpo, pTRP, y eGFP se amplificaron mediante el uso de los cebadores de la Tabla 7, y se ligaron en el vector pPM2\_pGAPxLRT digerido con BglI y SfiI. Tras la verificación de la secuencia, los vectores pPM2\_pGAPxLAA-LC, pPM2\_pGAPxLAA-pTRP y pPM2\_pGAPxLAA-eGFP se linealizaron en la región del promotor y se transformaron en *P. pastoris*, respectivamente.

Tabla 7: Cebadores de oligonucleótidos para la amplificación mediante PCR de pTRP, eGFP y LC de HyHEL fusionados a EpxL-AA (los sitios de restricción están subrayados, la secuencia EpxL está en cursiva)

EpxL-AA-LC dir.	Sbfl	SEQ ID 42 ATCACCTGCAGGATGAAGTTCTCCACCAATTTGATTCTAG CTATTGCAGCAGCTTCCGCCGTTGTCTCAGCTGCTGACAT CGTTTTGACTCAATCCCC
LC inv.	Sfil	SEQ ID 43 CTATGGCCGAGGCGGCCCTATTAACACTCACCTCTGTTG
EpxL-AA-pTRP dir.	PvuII	SEQ ID 44 ATAC <u>CAGCTG</u> CTACTGACGACGACGACAAG
pTRP inv.	Sfil	SEQ ID 45 TTTTGGCCGAGGCGGCCCTTTCAGTTAGCAGCGATAGTTTG
EpxL-AA-eGFP dir.	PvuII	SEQ ID 46 GTAC <u>CAGCTG</u> CTGTGAGCAAGGGCGAGGAGC
eGFP inv.	Sfil	SEQ ID 47 CGTTGGCCGAGGCGGCCCTTACTTGTACAGCTCGTCCATG

5 Ejemplo 8: Expresión de proteínas recombinantes mediante una línea celular hospedadora de *P. pastoris* con el uso de EpxL-AA (SEQ ID 21) para la secreción

8a) Análisis de *P. pastoris* que sobreexpresa eGFP mediante el uso de EpxL-AA para la secreción

Se cultivó *P. pastoris* que expresaba eGFP mediante el uso de la secuencia EpxL-AA para la secreción (Ejemplo 7) como se describió en el Ejemplo 2a, y se analizó como en el Ejemplo 3b.

10 En la SDS-PAGE, el tamaño de las bandas con EpxL-AA es ligeramente mayor que con EpxL-KR (Figura 5.1), lo que indica que queda al menos un residuo de Ala en el extremo N-terminal de eGFP, y de ese modo representa un extremo N-terminal no nativo de la proteína secretora recombinante.

8b) Análisis de *P. pastoris* que sobreexpresa tripsinógeno porcino recombinante mediante el uso de EpxL-AA para la secreción

15 Se cultivó *P. pastoris* que expresaba pTRP mediante el uso de la secuencia EpxL-AA para la secreción (Ejemplo 7) como se describió en el Ejemplo 2a, y se analizó como en el Ejemplo 3a.

Se secretó pTRP mediante el uso de la secuencia EpxL-AA, sin embargo, los niveles de secreción fueron inferiores que con la secuencia EpxL-KR (Tabla 8). En la SDS-PAGE, el tamaño de las bandas con EpxL-AA es ligeramente mayor que con EpxL-KR (Figura 5.1), lo que indica que queda al menos un residuo de Ala en el extremo N-terminal de pTRP.

20 De hecho, la secuenciación N-terminal del pTRP secretado mediante EpxL-AA reveló que el extremo N-terminal de la proteína secretora recombinante contuvo un aminoácido adicional, Ala, que quedó de la secuencia EpxL-AA, y de ese modo representa un extremo N-terminal no nativo de la proteína secretora recombinante.

Tabla 8: Niveles relativos de secreción de pTRP normalizados respecto de EpxL-KR

Líder	Secreción de pTRP media relativa $\pm$ EEM
EpxL-KR	1,00 $\pm$ 0,12
EpxL-AA	0,79 $\pm$ 0,04

25 8c) Análisis de *P. pastoris* que sobreexpresa LC mediante el uso de EpxL-AA para la secreción

30 Se cultivó *P. pastoris* que expresaba LC de HyHEL mediante el uso de la secuencia EpxL-AA para la secreción (Ejemplo 7) como se describió en el Ejemplo 2a, y se analizó como en el Ejemplo 2d. Para LC, los niveles de secreción con EpxL-AA fueron inferiores en comparación con EpxL-KR o MF $\alpha$  (Figura 5.2). Como para pTRP, la secuenciación N-terminal de la LC secretada mediante EpxL-AA mostró que el extremo N-terminal de la proteína secretora recombinante contuvo un aminoácido adicional, Ala, que quedó de la secuencia EpxL-AA.

Este es de nuevo un residuo indeseado derivado de la secuencia señal, que hace que EpxL-AA sea inadecuada

para la producción de proteínas recombinantes secretadas.

Ejemplo 9: Construcción de una línea celular hospedadora de *P. pastoris* para la expresión de proteínas recombinantes mediante el uso de la secuencia EpxL-A (secuencia del péptido señal, SEQ ID 1, en el que X en la posición 3 es F, y X en la posición 16 es A, SEQ ID 3) para la secreción

- 5 Por lo tanto, se construyeron cepas de *P. pastoris* que secretaban proteínas recombinantes mediante una secuencia señal que consiste justo en los primeros 20 aminoácidos del líder de Epx1 nativo, denominado EpxL-A. La Ala representa el último aminoácido antes (en posición C-terminal) del sitio de escisión predicho de la peptidasa señal, mientras en el lado N-terminal del sitio de escisión SP la proteína recombinante comienza inmediatamente. Para verificar que la naturaleza de este aminoácido N-terminal no influye en la escisión de SP, se ensayó la secreción de tres proteínas recombinantes diferentes: eGFP (que comienza con el aminoácido hidrófobo Val), y el anticuerpo LC y HC (que comienza con el aminoácido cargado negativamente Asp).

La cadena ligera LC de anticuerpo, Fab-HC, y eGFP se amplificaron mediante el uso de los cebadores de la Tabla 9, y se ligaron en el vector pPM2\_pGAPxLRT digerido con BglI y SfiI. Tras la verificación de la secuencia, los vectores se linealizaron en la región del promotor y se transformaron en *P. pastoris*.

- 15 Para LC y Fab-HC, el promotor pGAP también se intercambi6 por el promotor inducible pG1 (SEQ ID 9) mediante el uso de Apal y SbfI, y despu6 los casetes de expresi6n para ambas cadenas se combinaron en un vector mediante el uso de las enzimas de restricci6n compatibles MreI y AgeI. Tras la verifi6n de la secuencia, el vector se linealiz6 en la regi6n AOX-TT y se transform6 en *P. pastoris*.

- 20 Tabla 9: Cebadores de oligonucle6tidos para la amplificaci6n mediante PCR de eGFP y LC de HyHEL y Fab-HC fusionados a EpxL-A (los sitios de restricci6n est6n subrayados, la secuencia EpxL est6 en cursiva)

EpxL-A-LC dir.	SbfI	SEQ ID 48 TACTCCTGCAGGATGAAGTTCTCCACCAATTTGATTCTAG CTATTGCAGCAGCTTCCGCCGTTGTCTCAGCTGACGTTCA ATTGCAAGAATCTGG
LC inv.	SfiI	SEQ ID 49 CTATGGCCGAGGCGGCCCTATTAACACTCACCTCTGTTG
EpxL-A-eGFP dir.	PvuII	SEQ ID 50 GTACCAGCTGTGAGCAAGGGCGAGGAGC
eGFP inv.	SfiI	SEQ ID 51 CGTTGGCCGAGGCGGCCCTTACTTGTACAGCTCGTCCATG
EpxL-A-Fab HC dir.	SbfI	SEQ ID 52 TACTCCTGCAGGATGAAGTTCTCCACCAATTTGATTCTAG CTATTGCAGCAGCTTCCGCCGTTGTCTCAGCTGACGTTCA ATTGCAAGAATCTGG
Fab-HC inv.	SfiI	SEQ ID 53 TCATGGCCGAGGCGGCCCTTACTTGTACAGGACTTTG GCTC

Ejemplo 10: Expresi6n de prote6nas recombinantes mediante una l6nea celular hospedadora de *P. pastoris* con el uso de EpxL-A (secuencia del p6ptido se6al, SEQ ID 1, en el que X en la posici6n 3 es F, y X en la posici6n 16 es A (SEQ ID 3) para la secreci6n

- 25 10a) An6lisis de *P. pastoris* que sobreexpresa eGFP mediante el uso de EpxL-A para la secreci6n

Se cultiv6 *P. pastoris* que expresaba eGFP mediante el uso de la secuencia EpxL-A para la secreci6n (Ejemplo 9) como se describi6 en el Ejemplo 2a, y se analiz6 como en el Ejemplo 3b.

Como se puede observar en la Figura 5.1, los niveles de secreci6n de eGFP mediante el uso de EpxL-A fueron comparables a EpxL-KR, y mayores que con EpxL-AA.

- 30 Al contrario que las secuencias se6al de hidrofobina y las secuencias l6der descritas en Kottmaier et al. (2011 Appl Microbiol Biotechnol. 91: 133-141) y el prepro-l6der de MF $\alpha$ , la secuencia EpxL-A y la secuencia EpxL-KR no dan como resultado un transporte vacuolar de eGFP (como se demuestra mediante microscop6a de barrido l6ser confocal - datos no mostrados). Ambas cepas exhiben una distribuci6n de eGFP intracelular representativa de las prote6nas presentes en la ruta secretora (tinci6n principalmente del ret6culo endopl6smico); por lo tanto, el p6ptido se6al EpxL-A y el l6der truncado EpxL-KR parecen ser muy adecuados para la secreci6n de prote6nas recombinantes, ya que exhiben un "fenotipo secretor".

10b) Análisis de *P. pastoris* que sobreexpresa LC mediante el uso de EpxL-A para la secreción

Se cultivó *P. pastoris* que expresaba LC de HyHEL mediante el uso de la secuencia EpxL-A para la secreción (Ejemplo 9) como se describió en el Ejemplo 2a, y se analizó como en el Ejemplo 6d.

5 El uso de EpxL-A para la secreción de la cadena ligera del anticuerpo condujo a una proteína de tamaño correcto (Figura 5.2), por lo cual la secreción fue tan eficaz como con EpxL-KR y más de 8 veces mayor que con EpxL-AA (Tabla 10). Los niveles de secreción se compararon mediante la cuantificación de las intensidades de las bandas del gel mostrado en la Figura 5.2 con el uso del programa informático ImageJ.

Tabla 10: Niveles relativos de secreción de LC de HyHEL normalizados respecto de EpxL-AA.

Líder	Secreción de LC media relativa ± EEM
EpxL-KR	8,82 ± 0,53
EPxL-AA	1,00 ± 0,55
EPxL-A	9,53 ± 0,87

10 Se verificó que el extremo N-terminal de LC secretado con EpxL-A mediante LC-MS fue el extremo N-terminal correcto DIVLTQSP (Asp-Ile-Val-Leu-Thr-Gln-Ser-Pro, SEQ ID 54).

Por lo tanto, el péptido señal de Epx1, EpxL-A, es suficiente y adecuado para secretar proteínas recombinantes con el extremo N-terminal correcto, al contrario que la secuencia más larga EpxL-AA.

15 Además de la expresión bajo control del promotor constitutivo pGAP, también se ensayó la secreción de LC con EpxL-A bajo control del promotor inducible pG1 (SEQ ID 9). La construcción de la cepa se describió en el Ejemplo 9. Se generaron condiciones limitantes de glucosa en los cultivos de cribado mediante el uso de esferas de suministro de glucosa como se describió en el Ejemplo 2a. En tales condiciones de inducción, se observó la secreción de la cadena ligera.

20 10c) Análisis de *P. pastoris* que sobreexpresa un fragmento Fab de anticuerpo mediante el uso de EpxL-A para la secreción

Se cultivó *P. pastoris* que expresaba el fragmento Fab del anticuerpo HyHEL (que consiste en el fragmento de la cadena ligera ( $v_L$  y  $c_L$ ) y de la cadena pesada ( $v_H$  y  $c_{H1}$ )) mediante el uso de la secuencia EpxL-A para la secreción de ambas cadenas (cepa descrita en el Ejemplo 9) como se describió en el Ejemplo 2a, y se analizó como en los Ejemplos 2d y 2e.

25 Sorprendentemente, mediante el uso de la secuencia EpxL-A para la secreción de LC y HC del Fab de HyHEL, los niveles del Fab secretado intacto fueron hasta diez veces mayores que con el prepro-líder de MF $\alpha$  (determinado mediante ELISA como se describió en el Ejemplo 2e). El rendimiento de Fab por biomasa fue 0,15-0,35 mg de Fab/DO al usar EpxL-A, en comparación con 0,03-0,13 mg de Fab/DO al usar MF $\alpha$  para la secreción. Además, los sobrenadantes de *P. pastoris* que secretaba Fab de HyHEL bajo control de pG1 con EpxL-A no contienen niveles elevados de LC libre o agregados de mayor peso molecular de la proteína recombinante (Figura 5.3).

Ejemplo 11: Expresión de proteínas recombinantes mediante una línea celular hospedadora de *P. pastoris* con el uso de EpxL-A (secuencia del péptido señal, SEQ ID 1, en el que X en la posición 3 es F, y X en la posición 16 es A, SEQ ID 3) para la secreción

35 Se ensayó la secreción de ocho proteínas a partir de *P. pastoris* mediante el uso de EpxL-A (SEQ ID 3): Hormona del Crecimiento Humana (HGH), Somatotropina, Interferón alfa2a (IFN- $\alpha$  2a), los dos scFvs diferentes marcados con his (scFvs1 y scFvs2) y los 3 Fabs diferentes Fab1, Fab2 y Fab3.

40 Los genes se sometieron a optimización de codones para *P. pastoris* y se sintetizaron mediante GeneArt (Alemania). Los vectores obtenidos se digirieron con SbfI y SfiI, y se ligaron los genes en el vector pPM2aZ30\_pG1. En el caso de los Fabs, los casetes de expresión para ambas cadenas se combinaron en un vector mediante el uso de las enzimas de restricción compatibles MreI y AgeI. Tras la verificación de la secuencia, los vectores se linealizaron en la región del terminador AOX y se transformaron en *P. pastoris*.

45 Se cultivaron cepas de *P. pastoris* que expresaban las proteínas recombinantes Hormona del Crecimiento Humana, Somatotropina, Interferón alfa2a, los dos scFvs marcados con his scFv1 y scFv2 y los tres Fabs Fab1, Fab2 y Fab3 (Fabs que consisten en la cadena ligera ( $v_L$  y  $c_L$ ) y el fragmento de la cadena pesada ( $v_H$  y  $c_{H1}$ )) mediante el uso de la secuencia EpxL-A (SEQ ID 3) para la secreción en un medio que contenía por litro: 10 g de extracto de levadura, 10 g de peptona, tampón de fosfato potásico 100 mM de pH 6,0, 13,4 g de base de nitrógeno de levadura con sulfato amónico y 0,4 mg de biotina. Se llevaron a cabo cribados a pequeña escala en placas de 24 pocillos con dos esferas de suministro (Kuhner, diámetro de 6 mm).

Se analizaron los sobrenadantes como en el Ejemplo 2d. Se detectaron PDIs en las transferencias de Western mediante el uso de anticuerpos específicos: Para Somatotropina y Hormona del Crecimiento Humana: Anticuerpo policlonal GH1; Proteintech 17867-1AP (1:5.000) y anticuerpo Anti-IgG de Conejo (molécula completa)-Fosfatasa Alcalina, Sigma A3687 (1:12.000). Para Interferón-alfa 2a: Interferón, anticuerpo alfa 2a; Antibodies-Online ABIN573795 (1:1.500) y anticuerpo Anti-IgG de Conejo (molécula completa)-Fosfatasa Alcalina, Sigma A3687 (1:12.000). Para los scFvs marcados con His: Conjugado Penta-His HRP; QIAGEN 10149928 (1:1.500).

Todas las proteínas ensayadas se secretaron con éxito en el medio de cultivo al usar la secuencia EpxL-A (SEQ ID 3) (Figura 6).

**Listado de secuencias**

10 <110> Lonza Ltd  
 <120> Secuencias de expresión

15 <130> LO002P  
 <160> 53  
 <170> PatentIn version 3.5

20 <210> 1  
 <211> 20  
 <212> PRT  
 <213> Pichia pastoris

25 <220>  
 <221> PÉPTIDO  
 <222> (3)..(3)  
 <223> X es bien F o L

30 <220>  
 <221> PÉPTIDO  
 <222> (16)..(16)  
 <223> X es bien A o T

35 <400> 1  
 Met Lys Xaa Ser Thr Asn Leu Ile Leu Ala Ile Ala Ala Ala Ser Xaa  
 1 5 10 15  
 Val Val Ser Ala  
 20

40 <210> 2  
 <211> 20  
 <212> PRT  
 <213> Pichia pastoris

<400> 2  
 Met Lys Leu Ser Thr Asn Leu Ile Leu Ala Ile Ala Ala Ala Ser Ala  
 1 5 10 15  
 Val Val Ser Ala  
 20

45 <210> 3  
 <211> 20  
 <212> PRT

50 <213> Pichia pastoris

<400> 3  
 Met Lys Phe Ser Thr Asn Leu Ile Leu Ala Ile Ala Ala Ala Ser Ala  
 1 5 10 15

ES 2 789 299 T3

Val Val Ser Ala  
                   20  
 <210> 4  
 <211> 20  
 <212> PRT  
 5 <213> Pichia pastoris  
  
 <400> 4  
 Met Lys Phe Ser Thr Asn Leu Ile Leu Ala Ile Ala Ala Ala Ser Thr  
   1                  5                  10                  15  
  
 Val Val Ser Ala  
                   20  
 10 <210> 5  
 <211> 20  
 <212> PRT  
 <213> Pichia pastoris  
  
 <400> 5  
 Met Lys Leu Ser Thr Asn Leu Ile Leu Ala Ile Ala Ala Ala Ser Thr  
   1                  5                  10                  15  
  
 Val Val Ser Ala  
                   20  
 15 <210> 6  
 <211> 5  
 <212> PRT  
 <213> Pichia pastoris  
 20 <400> 6  
 Val Ser Ala Ala Pro  
   1                  5  
 <210> 7  
 <211> 16  
 25 <212> PRT  
 <213> Pichia pastoris  
  
 <400> 7  
 Ala Pro Val Ala Pro Ala Glu Glu Ala Ala Asn His Leu His Lys Arg  
   1                  5                  10                  15  
 30 <210> 8  
 <211> 57  
 <212> PRT  
 <213> Pichia pastoris  
 35 <400> 8  
 Met Lys Phe Ser Thr Asn Leu Ile Leu Ala Ile Ala Ala Ala Ser Thr  
   1                  5                  10                  15  
  
 Val Val Ser Ala Ala Pro Val Ala Pro Ala Glu Glu Ala Ala Asn His  
                   20                  25                  30  
  
 Leu His Lys Arg Ala Tyr Tyr Thr Asp Thr Thr Lys Thr His Thr Phe  
           35                  40                  45  
  
 Thr Glu Val Val Thr Val Tyr Arg Thr  
   50                  55  
 <210> 9  
 <211> 1001  
 <212> ADN

ES 2 789 299 T3

<213> Pichia pastoris

<400> 9

	atttccaccc ccatcccagt agaatgtagg gtccccaac atttgctccc cctagtctcc	60
	agggaaatgt aaaatatact gctaatagaa aacagtaaga cgctcagttg tcaggataat	120
	tacgttcgac tgtagtaaaa caggaatctg tattgttaga aagaacgaga gttttttacg	180
	gcgccgcat attgggccgt gtgaaaacag cttgaaaccc cactactttc aaaggttctg	240
	ttgctataca cgaacatgt ttaaccaacc tcgcttttga cttgactgaa gtcacggtt	300
	aacaatcaag taccctagtc tgtctgaatg ctcctttcca tattcagtag gtgtttcttg	360
	cacttttgca tgcactgcyg aagaattagc caatagcgcyg tttcatatgc gcttttacc	420
	cctcttttgt caagcgcaaa atgcctgtaa gatttggtyg gggtygtgagc cgttagctga	480
	agtacaacag gctaattccc tgaaaaact gcagatagac ttcaagatct cagggattcc	540
	cactatttgg tattctgata tgtttttcct gatatgcatc aaaactctaa tctaaaacct	600
	gaatctccgc tatttttttt ttttttttga tgaccccgty ttcgtgacaa attaatttcc	660
	aacggggtct tgtccggata agagaatttt gtttgattat ccgtytcggat aaatggacyc	720
	ctgctccata tttttccggt tattaccca cctggaagty cccagaattt tccggggatt	780
	acggataata cggtygtctg gattaattaa tacgccaagt cttacatttt gtygcagtct	840
	cgtgcygagta tgytgaataa taacaagat gagccaattt attggattag tgycagctty	900
	accccgccat agctaggcat agccaagtyc tatggtygtt agatgatgca ctyggatgca	960
5	gtgagtyttg gagtataaaa gatccttaa attccacct t	1001

<210> 10

<211> 36

<212> PRT

10 <213> Pichia pastoris

<220>

<221> PÉPTIDO

<222> (3)..(3)

15 <223> X es bien F o L

<220>

<221> PÉPTIDO

<222> (16)..(16)

20 <223> X es bien A o T

<400> 10

Met Lys Xaa Ser Thr Asn Leu Ile Leu Ala Ile Ala Ala Ala Ser Xaa  
 1 5 10 15

Val Val Ser Ala Ala Pro Val Ala Pro Ala Glu Glu Ala Ala Asn His  
 20 25 30

Leu His Lys Arg  
 35

25 <210> 11  
 <211> 36  
 <212> PRT  
 <213> Pichia pastoris

30 <400> 11

ES 2 789 299 T3

Met Lys Leu Ser Thr Asn Leu Ile Leu Ala Ile Ala Ala Ala Ser Ala  
1 5 10 15

Val Val Ser Ala Ala Pro Val Ala Pro Ala Glu Glu Ala Ala Asn His  
20 25 30

Leu His Lys Arg  
35

5 <210> 12  
<211> 36  
<212> PRT  
<213> Pichia pastoris

10 <400> 12

Met Lys Phe Ser Thr Asn Leu Ile Leu Ala Ile Ala Ala Ala Ser Ala  
1 5 10 15

Val Val Ser Ala Ala Pro Val Ala Pro Ala Glu Glu Ala Ala Asn His  
20 25 30

Leu His Lys Arg  
35

15 <210> 13  
<211> 36  
<212> PRT  
<213> Pichia pastoris

<400> 13

Met Lys Phe Ser Thr Asn Leu Ile Leu Ala Ile Ala Ala Ala Ser Thr  
1 5 10 15

Val Val Ser Ala Ala Pro Val Ala Pro Ala Glu Glu Ala Ala Asn His  
20 25 30

Leu His Lys Arg  
35

20 <210> 14  
<211> 36  
<212> PRT  
<213> Pichia pastoris

<400> 14

Met Lys Leu Ser Thr Asn Leu Ile Leu Ala Ile Ala Ala Ala Ser Thr  
1 5 10 15

Val Val Ser Ala Ala Pro Val Ala Pro Ala Glu Glu Ala Ala Asn His  
20 25 30

Leu His Lys Arg  
35

30 <210> 15  
<211> 60  
<212> ADN  
<213> Pichia pastoris



ES 2 789 299 T3

<400> 15

5 atgaagctct ccaccaattt gattctagct attgcagcag cttccgccgt tgtctcagct 60

<210> 16  
<211> 60  
<212> ADN  
<213> Pichia pastoris

10 <400> 16

atgaagtct ctaccaattt gattctagct attgcagcag cttccgccgt tgtctcagct 60

15 <210> 17  
<211> 60  
<212> ADN  
<213> Pichia pastoris

20 <400> 17

atgaagtct ctaccaattt gatcttagct attgcagcag catccactgt tgtctcagct 60

25 <210> 18  
<211> 108  
<212> ADN  
<213> Pichia pastoris

30 <400> 18

atgaagctct ccaccaattt gattctagct attgcagcag cttccgccgt tgtctcagct 60

gctccagttg ctccagccga agaggcagca aaccacttgc acaagcgt 108

35 <210> 19  
<211> 108  
<212> ADN  
<213> Pichia pastoris

<400> 19

atgaagttct ctaccaattt gattctagct attgcagcag cttccgccgt tgtctcagct 60

40 gctccagttg ctccagccga agaggcagca aaccacttgc acaagcgt 108

<210> 20  
<211> 108  
<212> ADN  
<213> Pichia pastoris

45 <400> 20

atgaagttct ctaccaattt gatcttagct attgcagcag catccactgt tgtctcagct 60

50 gctccagttg ctccagccga agaggcagca aaccacttgc acaagcgt 108

<210> 21  
<211> 21  
<212> PRT  
<213> Pichia pastoris

55 <400> 21

ES 2 789 299 T3

Met Lys Leu Ser Thr Asn Leu Ile Leu Ala Ile Ala Ala Ala Ser Ala  
 1 5 10 15

Val Val Ser Ala Ala  
 20

<210> 22  
 <211> 36  
 <212> ADN  
 <213> artificial

5 <220>  
 <223> cebador

10 <400> 22  
 tatacctgca g gatgaagt cctaccaat ttgatc 36

<210> 23  
 <211> 31  
 <212> ADN  
 <213> artificial

<220>  
 <223> cebador

20 <400> 23  
 gaagatgcat cgtacggtag acagtgacaa c 31

<210> 24  
 <211> 28  
 <212> ADN  
 <213> artificial

<220>  
 <223> cebador

30 <400> 24  
 ataccgtacg actgacgacg acgacaag 28

35 <210> 25  
 <211> 40  
 <212> ADN  
 <213> artificial

40 <220>  
 <223> cebador

45 <400> 25  
 ttttgccga ggcggcctt cagttagcag cgatagttg 40

<210> 26  
 <211> 37  
 <212> ADN  
 <213> artificial

50 <220>  
 <223> cebador

55 <400> 26  
 aggcgtctac cgaactgatg cacacaagag tgaggtt 37

<210> 27  
 <211> 39  
 <212> ADN  
 <213> artificial

60

<220>  
 <223> cebador

5 <400> 27  
 gagtggccga ggcggcctta taagcctaag gcagcttga 39

10 <210> 28  
 <211> 35  
 <212> ADN  
 <213> artificial

15 <220>  
 <223> cebador

15 <400> 28  
 attgtctac cgaactgtga gcaagggcga ggagc 35

20 <210> 29  
 <211> 39  
 <212> ADN  
 <213> artificial

25 <220>  
 <223> cebador

25 <400> 29  
 cgttggccga ggcggcctta ctgtacagc tcgtccatg 39

30 <210> 30  
 <211> 7  
 <212> PRT  
 <213> artificial

35 <220>  
 <223> péptido

35 <400> 30  
 Asp Ala His Lys Ser Glu Val  
 1 5

40 <210> 31  
 <211> 4  
 <212> PRT  
 <213> artificial

45 <220>  
 <223> péptido

45 <400> 31  
 Ala Tyr Tyr Thr  
 1

50 <210> 32  
 <211> 58  
 <212> ADN  
 <213> artificial

55 <220>  
 <223> cebador

60 <400> 32  
 attcgccgaa gaggcagcaa accacttgca caagcgtgat gcacacaaga gtgaggtt 58

60 <210> 33  
 <211> 39  
 <212> ADN  
 <213> artificial

ES 2 789 299 T3

<220>  
 <223> cebador

5 <400> 33  
 gagtggccga ggcggccta taagcctaag gcagcttga 39

10 <210> 34  
 <211> 75  
 <212> DNA  
 <213> artificial

15 <220>  
 <223> cebador

<400> 34  
 ataccagctg ctccagttgc tccagccgaa gaggcagcaa accacttgca caagcgtact 60  
 gacgacgacg acaag 75

20 <210> 35  
 <211> 40  
 <212> ADN  
 <213> artificial

25 <220>  
 <223> cebador

30 <400> 35  
 ttttggccga ggcggcctt cagtagcag cgatagttg 40

35 <210> 36  
 <211> 76  
 <212> ADN  
 <213> artificial

40 <220>  
 <223> cebador

<400> 36  
 ataccagctg ctccagttgc tccagccgaa gaggcagcaa accacttgca caagcgtgtg 60  
 agcaagggcg aggagc 76

45 <210> 37  
 <211> 39  
 <212> ADN  
 <213> artificial

50 <220>  
 <223> cebador

<400> 37  
 cgttggccga ggcggccta ctgtacagc tcgtccatg 39

55 <210> 38  
 <211> 49  
 <212> ADN  
 <213> artificial

60 <220>  
 <223> cebador

ES 2 789 299 T3

<400> 38  
 ataagccgaa gaggcagcaa accactgca caagcgtgac atcgtttg 49

5 <210> 39  
 <211> 39  
 <212> ADN  
 <213> artificial

10 <220>  
 <223> cebador

<400> 39  
 ctatggccga ggcggcccta ttaactca cctctgtg 39

15 <210> 40  
 <211> 49  
 <212> ADN  
 <213> artificial

20 <220>  
 <223> cebador

<400> 40  
 ataagccgaa gaggcagcaa accactgca caagcgtgac gttcaattg 49

25 <210> 41  
 <211> 39  
 <212> ADN  
 <213> artificial

30 <220>  
 <223> cebador

<400> 41  
 tatcggccga ggcggcccta ttactacct ggggacaag 39

35 <210> 42  
 <211> 98  
 <212> ADN  
 <213> artificial

40 <220>  
 <223> cebador

45 <400> 42  
 atcacctgca ggatgaagtt ctccaccaat ttgattctag ctattgcagc agcttccgcc 60  
 gttgtctcag ctgctgacat cgttttgact caatcccc 98

50 <210> 43  
 <211> 39  
 <212> ADN  
 <213> artificial

55 <220>  
 <223> cebador

<400> 43  
 ctatggccga ggcggcccta ttaactca cctctgtg 39

60 <210> 44  
 <211> 30  
 <212> ADN  
 <213> artificial

<220>  
 <223> cebador

5 <400> 44  
 ataccagctg ctactgacga cgacgacaag 30

<210> 45  
 <211> 40  
 10 <212> ADN  
 <213> artificial

<220>  
 <223> cebador

15 <400> 45  
 tttggccga ggcggcctt cagtagcag cgatagttg 40

<210> 46  
 <211> 31  
 20 <212> ADN  
 <213> artificial

<220>  
 <223> cebador

25 <400> 46  
 gtaccagctg ctgtgagcaa gggcgaggag c 31

<210> 47  
 <211> 39  
 30 <212> ADN  
 <213> artificial

<220>  
 <223> cebador

<400> 47  
 40 cgttggccga ggcggccta ctgtacagc tcgtccatg 39

<210> 48  
 <211> 95  
 <212> ADN  
 <213> artificial

45 <220>  
 <223> cebador

<400> 48  
 50 tactcctgca ggatgaagtt ctccaccaat ttgattctag ctattgcagc agcttccgcc 60  
 gttgtctcag ctgacgttca attgcaagaa tctgg 95

<210> 49  
 <211> 39  
 55 <212> ADN  
 <213> artificial

<220>  
 <223> cebador

60 <400> 49  
 ctatggccga ggcggcccta ttaactca cctctgtg 39

<210> 50

ES 2 789 299 T3

<211> 28  
<212> ADN  
<213> artificial

5 <220>  
<223> cebador

<400> 50  
gtaccagctg tgagcaaggg cgaggagc 28

10 <210> 51  
<211> 39  
<212> ADN  
<213> artificial

15 <220>  
<223> cebador

<400> 51  
20 cgttgccga ggcggccta ctgtacagc tcgtccatg 39

<210> 52  
<211> 95  
<212> ADN  
25 <213> artificial

<220>  
<223> cebador

30 <400> 52  
tactcctgca ggatgaagt cccaccaat ttgattctag ctattgcagc agcttccgcc 60  
gttgtctcag ctgacgttca attgcaagaa tctgg 95

<210> 53  
35 <211> 44  
<212> ADN  
<213> artificial

<220>  
40 <223> cebador

<400> 53  
tcatggcga ggcggccta ttactgtca caggacttg gctc 44

**REIVINDICACIONES**

1. Un ácido nucleico aislado que codifica un líder unido de forma operable a una secuencia de ácidos nucleicos que codifica una proteína de interés (PDI), en donde el ácido nucleico que codifica el líder está fusionado a un extremo N-terminal, cuyo líder se selecciona del grupo que consiste en
  - 5 a) un péptido señal que consiste en la secuencia de 20 aminoácidos de SEQ ID 1 o una variante funcional del mismo con una o dos mutaciones puntuales, y
  - b) un péptido señal que consiste en la secuencia de 20 aminoácidos seleccionada del grupo que consiste en SEQ ID 2, 3, 4 y 5.
- en donde se excluye el péptido señal se caracteriza por una secuencia de 21 aminoácidos que incluye una extensión N-terminal mediante una alanina adicional, y;
  - 10 en donde el ácido nucleico no codifica la secuencia de aminoácidos identificada como SEQ ID 8, y
  - en donde está excluido el péptido señal que consiste en la secuencia de aminoácidos SEQ ID 2 fusionado con un polipéptido que consiste en un dominio variable único de inmunoglobulina.
2. El ácido nucleico de la reivindicación 1, en donde la secuencia de nucleótidos que codifica el líder se selecciona del grupo que consiste en SEQ ID 15, 16 y 17 o una variante con optimización de codones de cualquiera de SEQ ID 15, 16 o 17 para coincidir con el patrón de uso de codones del organismo usado para la expresión.
  - 15
3. El ácido nucleico de la reivindicación 2, en donde el líder es un péptido señal que consiste en la secuencia de aminoácidos de SEQ ID 3, preferiblemente en el que el ácido nucleico consiste en una secuencia de nucleótidos de SEQ ID 16, o una variante con optimización de codones de SEQ ID 16 para coincidir con el patrón de uso de
  - 20 codones del organismo usado para la expresión.
4. Casete de expresión que comprende el ácido nucleico de cualquiera de las reivindicaciones 1 a 3.
5. El casete de expresión según de la reivindicación 4, que comprende una secuencia de nucleótidos que codifica un péptido señal seleccionado del grupo que consiste en SEQ ID 15, 16 y 17 o una variante con optimización de
  - 25 codones de cualquiera de SEQ ID 15, 16, o 17 para coincidir con el patrón de uso de codones del organismo usado para la expresión.
6. El casete de expresión según la reivindicación 4 o 5, en el que dicha PDI comprende una secuencia de aminoácidos con una secuencia de aminoácidos N-terminal nativa.
7. El casete de expresión según cualquiera de las reivindicaciones 4 a 7, en el que la PDI se selecciona de proteínas terapéuticas, que incluyen anticuerpos o fragmentos de los mismos, preferiblemente seleccionados del
  - 30 grupo que consiste en un anticuerpo de longitud completa, un scFv, minicuerpo, diacuerpo, triacuerpo, tetracuerpo, Fab y una proteína de fusión con Fc, enzimas y péptidos, antibióticos proteicos, proteínas de fusión con toxinas, conjugados carbohidrato - proteína, proteínas estructurales, proteínas reguladoras, vacunas y proteínas o partículas similares a vacunas, enzimas de procesamiento, factores de crecimiento, hormonas y citocinas, preferiblemente seleccionada del grupo que consiste en anticuerpos o fragmentos de los mismos, o en donde dicha PDI media la
    - 35 producción de un metabolito de células hospedadoras.
8. El casete de expresión según cualquiera de las reivindicaciones 4 a 7, que comprende además un promotor unido de forma operable al ácido nucleico que codifica el líder.
9. Una célula hospedadora de levadura recombinante que comprende el casete de expresión según cualquiera de las reivindicaciones 4 a 8, preferiblemente en el que la levadura se selecciona del grupo de géneros que consiste en
  - 40 Pichia, Candida, Torulopsis, Arxula, Hansenula, Ogatea, Yarrowia, Kluyveromyces, Saccharomyces y Komagataella.
10. Un método para producir una PDI en una célula hospedadora de levadura, que comprende:
  - proporcionar la célula hospedadora según la reivindicación 9,
  - cultivar dicha célula hospedadora para expresar dicha PDI, y
  - purificar la PDI para obtener una preparación de una PDI purificada.
11. El método según la reivindicación 10, en el que dicha PDI comprende una secuencia de aminoácidos con una
  - 45 secuencia de aminoácidos N-terminal nativa, preferiblemente en el que dicha PDI comprende una secuencia de aminoácidos que no comprende una Alanina adicional como residuo de aminoácido N-terminal.
12. El método según las reivindicaciones 10 u 11, en el que la PDI se selecciona de proteínas terapéuticas, que incluyen anticuerpos o fragmentos de los mismos, preferiblemente seleccionados del grupo que consiste en un
  - 50 anticuerpo de longitud completa, un scFv, minicuerpo, diacuerpo, triacuerpo, tetracuerpo, Fab y una proteína de



fusión con Fc, enzimas y péptidos, antibióticos proteicos, proteínas de fusión con toxinas, conjugados carbohidrato - proteína, proteínas estructurales, proteínas reguladoras, vacunas y proteínas o partículas similares a vacunas, enzimas de procesamiento, factores de crecimiento, hormonas y citocinas, o en el que dicha PDI actúa como mediador en la producción de un metabolito de células hospedadoras.

- 5 13. El uso del ácido nucleico según cualquiera de las reivindicaciones 1 a 3, para la secreción de una PDI desde una célula hospedadora y/o para incrementar la secreción de una PDI desde una célula hospedadora, preferiblemente en el que al menos un 60, 65, 70, 75, 80, 85, 90, 95, 98, o 100% de la PDI secretada comprende una secuencia de aminoácidos N-terminal nativa.

Fig. 1: Promotor pG1 (SEQ ID 9)

ATTTCCACCCCATCCCAGTAGAATGTAGGGTCCCCAACATTTGCTCCCCCTAG  
TCTCCAGGGAAATGTAAAATATACTGCTAATAGAAAACAGTAAGACGCTCAGTTGT  
CAGGATAATTACGTTTCGACTGTAGTAAAACAGGAATCTGTATTGTTAGAAAGAACG  
AGAGTTTTTTACGGCGCCGCCATATTGGGCCGTGTGAAAACAGCTTGAAACCCCA  
CTACTTTCAAAGGTTCTGTTGCTATACACGAACCATGTTTAACCAACCTCGCTTTT  
GACTTGACTGAAGTCATCGGTTAACAATCAAGTACCCTAGTCTGTCTGAATGCTCC  
TTTCCATATTCAGTAGGTGTTTCTTGCACTTTTGCATGCACTGCGGAAGAATTAGC  
CAATAGCGCGTTTCATATGCGCTTTTACCCCTCTTTTGTCAAGCGCAAATGCCT  
GTAAGATTTGGTGGGGGTGTGAGCCGTTAGCTGAAGTACAACAGGCTAATTCCT  
GAAAAAAGTGCAGATAGACTTCAAGATCTCAGGGATTCCCCTATTTGGTATTCTG  
ATATGTTTTTCTGATATGCATCAAACTCTAATCTAAACCTGAATCTCCGCTATT  
TTTTTTTTTTTTTTGATGACCCCGTTTTCTGACAAATTAATTTCCAACGGGGTCTT  
GTCCGGATAAGAGAATTTTGTGGATTATCCGTTCCGGATAAATGGACGCCTGCTCC  
ATATTTTTCCGGTTATTACCCACCTGGAAGTGCCCAGAATTTTCCGGGGATTACG  
GATAATACGGTGGTCTGGATTAATTAATACGCCAAGTCTTACATTTTGTGTCAGTC  
TCGTGCGAGTATGTGCAATAATAACAAGATGAGCCAATTTATTGGATTAGTTGCA  
GCTTGACCCCGCCATAGCTAGGCATAGCCAAGTGCTATGGGTGTTAGATGATGCA  
CTTGGATGCAGTGAGTTTTGGAGTATAAAAGATCCTTAAATTCACCCCTT

Fig. 2.1

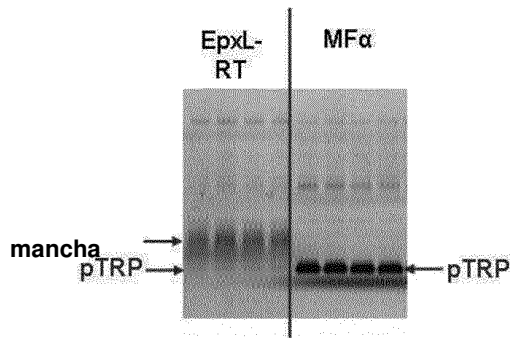


Fig. 2.2

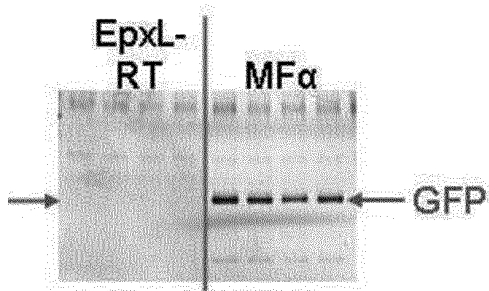


Fig. 2.3

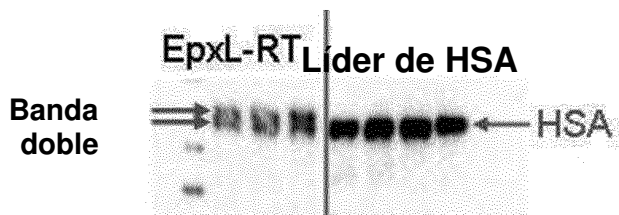


Fig. 3

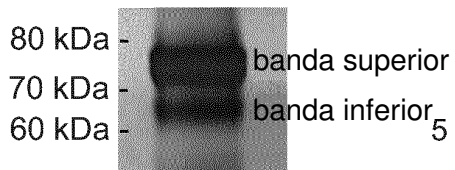


Fig. 4.1

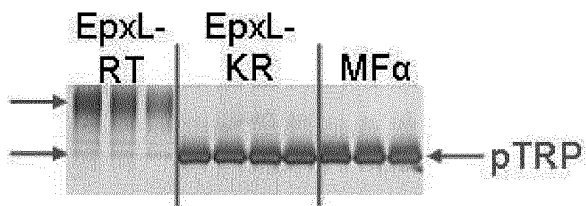


Fig. 4.2

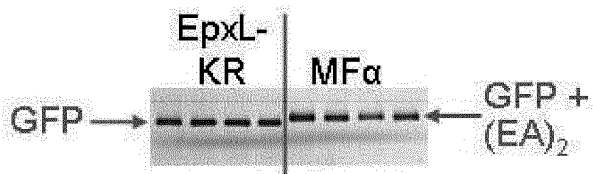


Fig. 4.3

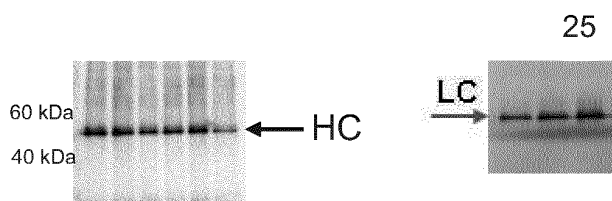


Fig. 5.1

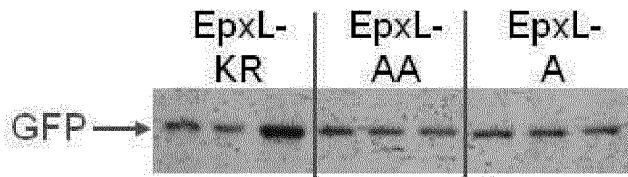


Fig. 5.2

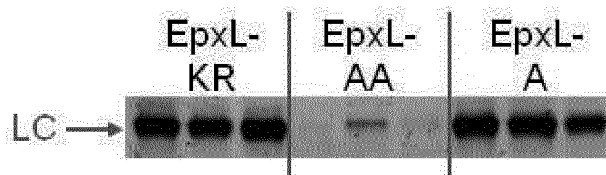


Fig. 5.3

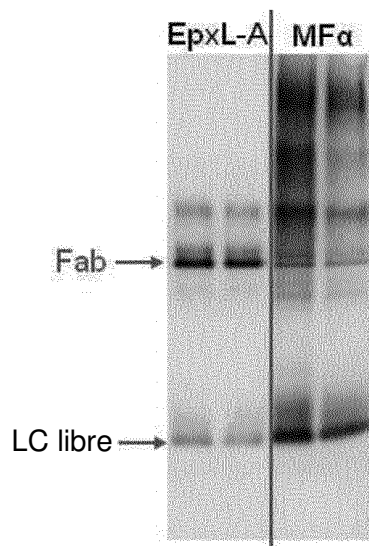


Fig. 6

