(19) OFICINA ESPAÑOLA DE PATENTES Y MARCAS

ESPAÑA

(11) Número de publicación: **2 921 137**

(51) Int. Cl.:

C12N 15/81 (2006.01)
C12P 21/02 (2006.01)
C12N 15/09 (2006.01)

(12) TRADUCCIÓN DE PATENTE EUROPEA MODIFICADA TRAS OPOSICIÓN T5

(86) Fecha de presentación y número de la solicitud internacional: **10.01.2020 PCT/EP2020/050517**

(87) Fecha y número de publicación internacional: **16.07.2020 WO20144313**

(96) Fecha de presentación y número de la solicitud europea: **10.01.2020 E 20700131 (4)**

(97) Fecha y número de publicación de la concesión europea modificada tras oposición: **25.12.2024 EP 3874047**

(54) Título: **Producción de proteínas regulada por fuente de carbono en una célula huésped recombinante**

(30) Prioridad:

11.01.2019 EP 19151376

(45) Fecha de publicación y mención en BOPI de la traducción de la patente modificada:
06.05.2025

(73) Titular/es:

LONZA LTD. (100.00%)
Lonzastrasse
3930 Visp, CH

(72) Inventor/es:

GASSER, BRIGITTE;
REBNEGGER, CORINNA;
FLORES VILLEGAS, MIRELLE CITLALI y
MATTANOVICH, DIETHARD

(74) Agente/Representante:

ELZABURU, S.L.P

ES 2 921 137 T5

DESCRIPCIÓN

Producción de proteínas regulada por fuente de carbono en una célula huésped recombinante

Campo técnico

La invención se refiere a la producción de una proteína de interés (PDI) en una célula huésped recombinante que comprende un casete de expresión heterólogo para expresar un gen de interés (GDI) que codifica la PDI, cuya célula huésped está diseñada para reducir la expresión de una proteína FLO8.

Antecedentes

Las proteínas producidas en cultivos de células huésped recombinantes se han vuelto cada vez más importantes como agentes diagnósticos y terapéuticos. Para este propósito, las células se diseñan genéticamente y/o seleccionan para producir niveles inusualmente altos de una proteína recombinante o heteróloga de interés. La optimización de las condiciones de cultivo celular es importante para la producción comercial exitosa de proteínas recombinantes o heterólogas.

La producción exitosa de proteínas de interés (PDI) se ha logrado tanto con células huésped procarióticas como eucarióticas en cultivo celular. Las células huésped eucarióticas, en particular células huésped de mamífero, levaduras u hongos filamentosos, o bacterias, se usan comúnmente como huéspedes de producción para proteínas biofarmacéuticas, así como para productos químicos a granel. Los ejemplos más destacados son levaduras como *Saccharomyces cerevisiae, Pichia pastoris* o *hansenula polymorpha,* hongos filamentosos como *Aspergillus awamori* o *Trichoderma reesei,* o células de mamífero como células CHO. La levadura metilotrófica, tal como *Pichia pastoris,* tiene buena reputación por la secreción eficaz de proteínas heterólogas. *P. pastoris* ha sido reclasificado en un nuevo género, *komagataella,* y se divide en tres especies, *k. pastoris, K. phaffii,* y *K. pseudopastoris.* Las cepas comúnmente utilizadas para aplicaciones biotecnológicas pertenecen a dos especies propuestas, *K. pastoris* y *K. phaffii.* Las cepas GS115, X-33, CBS2612 y CBS7435 son *K. phafii,* mientras que la serie SMD de cepas deficientes en proteasa (p. ej., SMD1168) se clasifica en la especie tipo, *K. pastoris,* que es la cepa de referencia para todas las cepas disponibles *P. pastoris* (Kurtzman 2009, JInd Microbiol Biotechnol. 36(11):1435-8). Mattanovich et al. (Fábricas de células microbianas 2009, 8:29 doi: 10.1186/1475-2859-8-29) describen la secuenciación del genoma de la cepa tipo DSMZ 70382 de *K. pastoris,* y analizó su secretoma y transportadores de azúcar.

El documento WO2015/158808A2 describe una célula huésped recombinante diseñada genéticamente para sobreexpresar proteínas auxiliares.

El documento WO2015/158800A1 describe la mejora de la capacidad de una célula huésped para expresar y/o secretar una PDI mediante diseño genético para subexpresar ciertas proteínas (llamadas proteínas KO) que son endógenas a la célula huésped y que han demostrado reducir el rendimiento de la producción de proteínas cuando se sobreexpresan. Tales proteínas KO, por lo tanto, se han elegido como dianas knock-out para mejorar el rendimiento. A su vez, se descubrió que la subexpresión de las proteínas KO en las líneas celulares huésped *P. pastoris* aumentan el rendimiento de las proteínas modelo entre 1,2 y 2,4 veces. Se han utilizado promotores inducibles (pAOX1) o constitutivos (pGAP). Las proteínas KO han sido identificadas como un *P. pastoris* homólogo de la proteína FLO8 de S. *cerevisiae*, una *P. pastoris* homólogo de la proteína HCH1 de S. *cerevisiae* y KO3 a *P. pastoris* homólogo de la proteína SCJ1 de S. *cerevisiae*.

Rebnegger et al. (Applied and Environmental Microbiology 2016, 82(15):4570-4583) describe cultivos de quimiostatos limitados en glucosa de un mutante de deleción *flo8* de *P. pastoris* para evitar la obstrucción del filtro.

Los promotores utilizados para la producción de proteínas en células huésped recombinantes están regulados (*p.ej.*, inducido tras la adición de metanol al medio, controlado con metanol), o constantemente activo (constitutivo). Los promotores controlados por metanol conducen a límites técnicos, tales como el calor residual en el reactor o el suministro de oxígeno.

El documento WO20137050551 A1 describe una serie de fuentes de carbono regulables promotor de *P. pastoris* (denominado pG1-pG8), e inducción de la producción de proteínas al limitar la fuente de carbono en el cultivo celular.

El documento WO2017021541 A1 describe variantes de un promotor regulable de fuente de carbono de *P. pastoris* (designadas como pG1), que están reguladas por una fuente de carbono distinta del metanol, *es decir*, no controlado con metanol, p. ej., reprimida en presencia de una fuente de carbono durante una fase de crecimiento e inducida por la limitación de la fuente de carbono en la fase de producción.

Prielhofer et al. (Microbial Cell Factories 2013; 12(5):1-10) describe promotores regulables de *P. pastoris* e inducidos sin metanol.

Prielhofer et al. (Biotechnology and Bioengineering. 2018;115:2479-2488) describe el promotor P_{GTH1} regulada por glucosa y variantes diseñadas genéticamente con propiedades de inducción muy mejoradas en comparación con las del promotor de tipo salvaje.

El documento EP2669375A1 describe una levadura con expresión de proteína mejorada por expresión de alto nivel de un homólogo de MPP1.

Hye Young Kim et al. (Biochemical and Biophysical Research Communications 2014, 449: 202-207) describen el papel de dos dominios del activador de Flo8 en la activación transcripcional de un conjunto de genes diana y el modo de acción de Flo8 al interactuar con el activador de Mss1 1.

El documento EP2952584A1 describe una producción de proteína mejorada mediante la sobreexpresión de ciertos polinucleótidos.

El documento EP2258855A1 describe ciertas secuencias líder y señal de secreción de *P. pastoris.*

El documento WO2010099195A1 describe cepas de *P. pastoris* modificados genéticamente y coexpresión de una proteína heteróloga y proteínas chaperonas.

Sumario de la invención

El objeto es mejorar la producción de proteínas en células huésped recombinantes y aumentar el rendimiento de la producción de proteínas. Otro objeto de la invención es proporcionar un método para producir una proteína heteróloga o recombinante en una célula huésped, en donde se reduce el riesgo de cambio en la morfología de la célula huésped durante los procesos de producción.

El objeto es resuelto por la materia según lo reivindicado.

De acuerdo con la invención, se proporciona una célula huésped recombinante que comprende un gen endógeno que codifica una proteína FLO8 que comprende la secuencia de aminoácidos identificada como SEQ ID NO: 1 o un homólogo de la misma, cuya célula huésped está diseñada genéticamente mediante una o más modificaciones genéticas para reducir (o suprimir) la expresión de dicho polinucleótido en comparación con la célula huésped antes de dicha una o más modificaciones genéticas, y dicha célula huésped comprende un casete de expresión heterólogo que comprende un gen de interés (GDI) para expresar dicho GDI bajo el control de un casete de expresión promotor (ECP). Específicamente, el ECP es regulable por una fuente de carbono distinta del metanol. Específicamente, el ECP es reprimible por una fuente de carbono que no es metanol.

Específicamente, el ECP es reprimible por una fuente de carbono represora, p. ej., una fuente de carbono de represión que no es metanol, tal como glucosa o glicerol, e inducible (desreprimible) al reducir la cantidad de la fuente de carbono de represión.

Específicamente, el ECP no es inducible por metanol.

Específicamente, la fuente de carbono distinta del metanol es cualquier fuente de carbono distinta del metanol que se usa adecuadamente en un cultivo de células huésped. Específicamente, la fuente de carbono distinta del metanol no es metanol.

Específicamente, la fuente de carbono distinta del metanol es otra fuente de carbono distinta del metanol. En particular, el ECP no está controlado por metanol. Aunque el cultivo celular o el medio de cultivo celular pueden comprender o libre de metanol, la ECP tal como se usa aquí no está regulada por ninguna cantidad de metanol, en particular no es inducible por metanol, por lo tanto, no está controlada por metanol. Específicamente, la ECP se puede inducir completamente en un cultivo celular libre de metanol o en un medio de cultivo celular.

Para el propósito descrito en este documento, el término "proteína FLO8" se referirá tanto a una proteína que comprende la secuencia de aminoácidos identificada como SEQ ID NO: 1, como a una secuencia de aminoácidos que tiene cierta homología con la SEQ ID NO: 1. Sin embargo, la secuencia homóloga también se denomina homólogo de FLO8.

Específicamente, el homólogo de FLO8 tiene al menos una identidad de secuencia del 25%, 30% o 35% con SEQ ID NO:1 *p.ej.*, al menos cualquiera de 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90% o 95% de identidad de secuencia, o es 100% idéntico a la SEQ ID NO:1. Específicamente, la identidad de la secuencia se determina como se describe más adelante en el presente documento, por ejemplo, cuando se compara la secuencia de longitud completa.

Específicamente, la proteína FLO8 o el homólogo respectivo comprende un dominio LisH que comprende al menos cualquiera de 50%, 60%, 65%, 70%, 75%, 80%, 85%, 90% o 95% de identidad de secuencia con SEQ ID NO:50, o es 100% idéntico a la SEQ ID NO:50.

De acuerdo con un aspecto específico, el dominio LisH de la proteína FLO8 comprende al menos una identidad de secuencia del 80%, 85%, 90% o 95% con cualquiera de los dominios LisH naturales de una proteína FLO8 de un organismo eucariótico, o un ortólogo de FLO8 respectivo, o es 100% idéntico a dicho dominio LisH natural, específicamente un dominio LisH de levadura u hongos, en particular hongos filamentosos, tal como un dominio LisH que comprende o consiste en cualquiera de:

a) SEQ ID NO:50, que es un dominio LisH de la proteína FLO8 de *K. phaffii* (SEQ ID NO: 50);

b) SEQ ID NO:51, dominio LisH de la proteína FLO8 de *K. pastoris* (BLASTp: 96% de identidad con SEQ ID NO: 50 (23/24 aa));

c) SEQ ID NO:52, que es un dominio LisH del ortólogo de FLO8 de una cepa de S. *cerevisiae* (BLASTp: 54% de identidad con SEQ ID NO: 50 (13/24 aa));

d) SEQ ID NO:53, que es un dominio LisH del ortólogo de FLO8 de otra cepa de S. *cerevisiae* (BLASTp: 54% de identidad con SEQ ID NO: 50 (13/24 aa));

e) SEQ ID NO:54, que es un dominio LisH del ortólogo de FLO8 de una cepa de *Yarrowia lipolytica* (BLASTp: 71% de identidad con SEQ ID NO: 50 (12/17 aa));

f) SEQ ID NO:55, que es un dominio LisH del ortólogo de FLO8 de *Ogataea polimorfa* (BLASTp: 62% de identidad con SEQ ID NO:50 (8/13 aa)); o

g) SEQ ID NO:56, que es un dominio LisH del ortólogo de FLO8 de una cepa de *Aspergillus niger* (BLASTp: 73% de identidad con SEQ ID NO:50 (8/11 aa)).

Típicamente, el dominio LisH de la proteína FLO8 tiene una longitud de 25-27 aminoácidos y una cierta identidad de secuencia con la proteína humana LIS1, como se describe en la base de datos Pfam, número ID PF08513, y la secuencia respectiva. En biología molecular, el dominio LisH es un dominio proteico que se encuentra en un gran número de proteínas eucariotas que tienen una amplia gama de funciones. La estructura del dominio LisH en la región N-terminal de LIS1 lo representó como un motivo de dimerización (The dimerization mechanism of LIS1 and its implication for proteins containing the LisH motif. Mateja A, Cierpicki T, Paduch M, Derewenda ZS, Otlewski J. 2006. J Mol Biol. 357(2):621-31).

Se entiende particularmente que el homólogo de FLO8 es endógeno a la célula huésped que se usa como célula huésped recombinante que produce la PDI como se describe más adelante en el presente documento. En particular, la proteína FLO8 es un ortólogo que es endógeno a la especie de la especie de célula huésped.

Específicamente, la proteína FLO8 es de origen *P. pastoris,* en particular *K. pastoris* o *K. phaffii*, si la célula huésped es *P. pastoris,* en particular *K. pastoris* y *K. phafii,* respectivamente. Alternativamente, la proteína FLO8 comprende una secuencia homóloga (u ortóloga) de tal proteína FLO8 de en *P. pastoris,* en particular *K. pastoris* o *K. phafii,* origen, cuya secuencia homóloga (ortóloga) es endógena a una célula huésped de tipo salvaje, si es de otro origen o especie. Por ejemplo, si la célula huésped es *K. phafii,* la proteína FLO8 endógena comprende o consiste en la secuencia de aminoácidos identificada como SEQ ID NO:1. Según otro ejemplo, si la célula huésped es *K. pastoris,* la proteína FLO8 endógena comprende o consiste en la secuencia de aminoácidos identificada como SEQ ID NO:3, que es 91% idéntica a SEQ ID NO:1. Sin embargo, si la célula huésped es de una especie diferente (que no sea *K. pastoris* y/o *K. phaffii*), la secuencia de la proteína FLO8 que es endógena a la célula huésped es un homólogo de la SEQ ID NO: 1 y la expresión de dicho homólogo en la célula huésped (la secuencia ortóloga de la SEQ ID NO: 1) se reduce para el propósito descrito en el presente documento.

Específicamente, cualquiera o cada una de las secuencias homólogas se caracteriza por la misma función cualitativa de la proteína FLO8 en la respectiva célula huésped de tipo salvaje que en *P. pastoris,* en particular *K. pastoris* o *K. phaffii p. ej..*, como factor de transcripción, en particular un activador transcripcional de unión al ADN implicado en la regulación de la adhesión celular, la floculación, el crecimiento invasivo o el catabolismo del almidón, aunque su actividad cuantitativa podría ser diferente en comparación con la proteína FLO8 en el tipo salvaje *K. pastoris* o *K. phaffii.*

Específicamente, la secuencia homóloga respectiva es de una especie distinta de *PAG. pastoris, en particular K. pastoris o K. phaffii, p. ej.*, otra levadura o una célula fúngica filamentosa, preferiblemente levaduras del género Komagataella o Pichia, o del género Saccharomyces o cualquier levadura metilotrófica. Sin embargo, la célula huésped puede ser una célula animal, una célula de vertebrado, una célula de mamífero, una célula humana, una célula vegetal, una célula bacteriana, una célula nematoda, una célula de invertebrado tal como una célula de insecto o una célula de molusco, o una la célula madre, y la proteína FLO8 respectiva y su homólogo descritos en el presente documento son endógenos a la célula huésped respectiva, pero su expresión se reduce o suprimió como se describe en el presente documento.

Específicamente, el homólogo de la proteína FLO8 es endógeno u originario de una especie de Pichia o endógeno u originario de cualquier otra levadura, hongo o bacteria, y tiene una identidad de secuencia del 25% SEQ ID NO: 1 o SEQ ID NO: 2, en casos específicos en al menos cualquiera de 35%, 40%, 45%, 50%, 55%, 60, 65%, 70%, 75%, 80%, 85%, 90%, 95%, o es 100% idéntico a SEQ ID NO: 1 o SEQ ID NO: 2. Específicamente, se determina que una proteína FLO8 exógena es un homólogo de la proteína FLO8 si es funcional al añadir la proteína FLO8 exógena (o el gen que codifica la proteína FLO8 exógena) a un cultivo de una cepa knockout de *flo8* de una cepa de Pichia (pastoris) o Saccharomyces (cerevisiae), cuya cepa knockout es diferente del origen de la proteína FLO8 exógena, lo que demuestra el reemplazo funcional de la proteína FLO8 endógena eliminada.

Específicamente, la proteína FLO8 es de origen *P. pastoris*, en particular codificada por un gen endógeno a la célula huésped, en donde la célula huésped es *P. pastoris.*

Específicamente, la proteína FLO8 es de origen *Komagataella phaffii*, que comprende o consiste en SEQ ID NO:1. Específicamente, tal proteína FLO8 está codificada por un gen endógeno a la célula huésped, en donde la célula huésped es *Komagataella phaffii.* Específicamente, la proteína FLO8 está codificada por la secuencia de nucleótidos identificada como SEQ ID NO:2.

Específicamente o, la proteína FLO8 es de origen *Komagataella*, que comprende al menos 90% o 91% de identidad de secuencia con SEQ ID NO:1. En concreto, la proteína FLO8 es de origen *Komagataella pastoris*, que comprende o consiste en SEQ ID NO:3. Específicamente, tal proteína FLO8 está codificada por un gen endógeno a la célula huésped, en donde la célula huésped es *Komagataella pastoris.*

Específicamente, la proteína FLO8 es de origen *Saccharomyces*, que comprende al menos un 35% de identidad de secuencia con SEQ ID NO:1. En concreto, la proteína FLO8 es de origen S. *cerevisiae*, que comprende o consiste en SEQ ID NO:5 o SEQ ID NO:6. Específicamente, dicha proteína FLO8 está codificada por un gen endógeno a la célula huésped, en donde la célula huésped es S. *cerevisiae.*

Específicamente, la proteína FLO8 es de origen *milenrama*, que comprende al menos 40% o 44% de identidad de secuencia con SEQ ID NO:1. Específicamente, la proteína FLO8 es de origen *Yarrowia lipolytica*, que comprende o consiste en SEQ ID NO:7. Específicamente, tal proteína FLO8 está codificada por un gen endógeno a la célula huésped, en donde la célula huésped es *Yarrowia lipolytica.*

Específicamente, la proteína FLO8 es de origen *Ogataea*, que comprende al menos 30% o 34% de identidad de secuencia con SEQ ID NO:1. Específicamente, la proteína FLO8 es de origen *Ogataea polimorfa*, que comprende o consiste en SEQ ID NO:8. Específicamente, tal proteína FLO8 está codificada por un gen endógeno a la célula huésped, en donde la célula huésped es *Ogataea polymorpha.*

Específicamente, la proteína FLO8 es de origen *Aspergilo*, que comprende al menos 25% o 26% de identidad de secuencia con SEQ ID NO:1. Específicamente, la proteína FLO8 es de origen *Aspergillus niger*, que comprende o consiste en SEQ ID NO:9. Específicamente, tal proteína FLO8 está codificada por un gen endógeno a la célula huésped, en donde la célula huésped es *Aspergillus niger.*

Específicamente, la célula huésped se modifica genéticamente mediante una o más modificaciones genéticas que comprenden mutaciones genómicas que reducen la transcripción y/o traducción de dicho polinucleótido que codifica dicha proteína FLO8, y/o reducen de otro modo la expresión de dicho polinucleótido y reducen la producción de dicha proteína FLO8, respectivamente.

Específicamente, dichas una o más modificaciones genéticas comprenden una interrupción, sustitución, eliminación o inactivación de (i) uno o más polinucleótidos endógenos, o una parte de los mismos; o (ii) una secuencia de control de expresión.

De acuerdo con un aspecto específico, dichas una o más modificaciones genéticas son de uno o más polinucleótidos endógenos de la célula huésped descrita en el presente documento, tales como polinucleótidos codificantes, incluyendo *p.ej.*, codificar dicho polinucleótido (o gen) la proteína FLO8, en particular la proteína de tipo salvaje (no modificada o nativa), que se produce naturalmente en la especie, tipo o cepa de la célula huésped.

De acuerdo con un aspecto específico, dichas una o más modificaciones genéticas son de una secuencia de control de expresión, incluyendo *p.ej.*, un promotor, un sitio de unión al ribosoma, secuencias de inicio y finalización transcripcionales o traduccionales, o de una secuencia potenciadora o activadora.

Se puede emplear una variedad de métodos de deiseño genético de una célula huésped para reducir la expresión de un polinucleótido endógeno, tal como un gen que codifica una proteína FLO8, incluyendo, p. ej., la interrupción del polinucleótido que codifica la proteína FLO8, la interrupción del promotor que está operativamente unido a dicho polinucleótido, reemplazando dicho promotor con otro promotor que tenga menor actividad promotora, modificando o modulando (p. ej., activando, regulando al alza, inactivando, inhibiendo o regulando a la baja) secuencias reguladoras que modulan la expresión de dicho polinucleótido, tal como usando los respectivos reguladores de la transcripción dirigido a las secuencias relevantes por una ribonucleasa guiada por ARN utilizada en un método basado en CRISPR para modificar una célula huésped, p. ej., secuencias reguladoras seleccionadas del grupo que consiste en promotor, sitios de unión ribosómicos, secuencias de inicio o finalización de la transcripción, secuencias de inicio o finalización de la traducción, secuencias potenciadoras o activadoras, secuencias represoras o inhibidoras, señales o líderes secuencias, en particular aquellas que controlan la expresión y/o secreción de una proteína.

Específicamente, dichas una o más modificaciones genéticas incluyen una o más mutaciones genómicas que incluyen la eliminación o inactivación de un gen o secuencia genómica que reduce la expresión de un gen o parte de un gen en al menos un 50%, 60%, 70%, 80%, 90%, o 95%, o incluso suprime por completo su expresión, *p.ej.*, por una knockout del gen, en comparación con el huésped respectivo sin tal modificación genética.

Específicamente, la una o más modificaciones genéticas comprenden mutaciones genómicas que deterioran constitutivamente o reducen de otro modo la expresión de uno o más polinucleótidos endógenos.

Específicamente, la una o más modificaciones genéticas comprenden mutaciones genómicas que introducen una o más secuencias reguladoras inducibles o reprimibles que alteran condicionalmente o reducen de otro modo la expresión de uno o más polinucleótidos endógenos. Tales modificaciones condicionalmente activas se dirigen particularmente a aquellos elementos reguladores y genes que están activos y/o se expresan dependiendo de las condiciones del cultivo celular.

Específicamente, la expresión de uno o más polinucleótidos endógenos se reduce, reduciendo así la expresión del polinucleótido que codifica la proteína FLO8 cuando se produce la PDI. Específicamente, tras la modificación genética, la expresión de dicha proteína FLO8 se reduce en las condiciones del cultivo de la célula huésped durante el cual se produce la PDI.

Específicamente, la célula huésped se modifica genéticamente para reducir la cantidad (p. ej., el nivel o la concentración) de dicha proteína FLO8, en al menos cualquiera de 50%, 60%, 70%, 80%, 90% o 95%, (mol/mol) en comparación con la célula huésped sin dicha modificación, o incluso en un 100%, *p.ej.* a una cantidad no detectable, suprimiendo así por completo la producción de la proteína FLO8, p. ej., mediante una desactivación del gen. De acuerdo con una realización específica, la célula huésped se modifica genéticamente para comprender una o más deleciones de (una o más) secuencias genómicas, en particular secuencias genómicas que codifican la proteína FLO8 o el respectivo homólogo de la misma. Tal célula huésped se proporciona normalmente como una cepa de deleción o knockout.

Según una realización específica, una vez que la célula huésped descrita en el presente documento se cultiva en un cultivo celular, la cantidad de proteína FLO8 total en la célula huésped o en el cultivo de la célula huésped se reduce al menos en un 50%, 60%, 70%, 80%, 90% o 95% (mol/mol), o incluso en un 100%, *p.ej.* a una cantidad no detectable, en comparación con una cantidad de referencia expresada o producida por la célula huésped antes o sin dicha modificación genética, o en comparación con una cantidad de referencia producida en un cultivo de células huésped respectivo, o en comparación con la célula huésped antes o sin dicha modificación.

Cuando se compara la célula huésped descrita en el presente documento por el efecto de dicha modificación genética para reducir la producción de dicha proteína FLO8, normalmente se compara con la célula huésped comparable antes o sin dicha modificación genética. La comparación se hace típicamente con la misma especie o tipo de célula huésped sin dicha modificación genética, que se diseña genéticamente para producir la PDI recombinante o heterólogo, en particular cuando se cultiva en condiciones para producir dicha PDI. Sin embargo, también se puede hacer una comparación con la misma especie o tipo de célula huésped que no se diseñe genéticamente más para producir la PDI recombinante o heterólogo.

Según un aspecto específico, la reducción de dicha proteína FLO8 o del respectivo homólogo de la misma está determinada por la reducción de la cantidad (*p.ej.*, el nivel o concentración) de dicha proteína FLO8 en la célula. Específicamente, la cantidad de dicha proteína FLO8 o el respectivo homólogo de la misma se determina mediante un método adecuado, tal como el empleo de transferencia Western, imágenes de inmunofluorescencia, citometría de flujo o espectrometría de masas, en particular en donde la espectrometría de masas es cromatografía líquida-espectrometría de masas (LC- MS), o espectrometría de masas en tándem de cromatografía líquida (LC-MS/MS) *p.ej.,* como se describe por Doneanu et al. (MAbs. 2012; 4(1): 24-44).

Según un aspecto específico, la célula huésped recombinante comprende solo uno o múltiples casetes de expresión heterólogos, p. ej., copias múltiples de dichos casetes de expresión, tales como al menos 2, 3, 4 o 5 copias (número de copias del gen, GCN). Por ejemplo, la célula huésped recombinante comprende hasta 2, 3, 4 o cinco copias. Cada una de las copias puede comprender o consistir en secuencias iguales o diferentes, pero incluye el ECP unido operativamente al GDI.

Según un aspecto específico, el casete de expresión heterólogo comprende o consiste en una fusión artificial de polinucleótidos, incluyendo el ECP unido operativamente al GDI, y opcionalmente otras secuencias, como una secuencia señal, líder o terminal. Específicamente, se usa un casete de expresión que es heterólogo a la célula huésped o artificial, en particular en donde el casete de expresión comprende un promotor (el ECP) y un GDI, en donde el promotor y el GDI son heterólogos entre sí, no apareciendo en tal combinación en la naturaleza *p.ej.,* en donde uno (o solo uno) del promotor y GDI es artificial o heterólogo para el otro y/o para la célula huésped descrita en el presente documento; el promotor es un promotor endógeno y el GDI es un GDI heterólogo; o el promotor es un promotor artificial o heterólogo y el GDI es un GDI endógeno; en donde tanto el promotor como el GDI son artificiales, heterólogos o de origen diferente, como de una especie o tipo (cepa) de células diferente en comparación con la célula huésped descrita en el presente documento. Específicamente, el ECP no está asociado de forma natural y/o no está ligado operativamente a dicho GDI en la célula que se utiliza como célula huésped descrita en el presente documento.

Según un aspecto específico, la ECP es inducible en presencia de una cantidad limitante del crecimiento de una fuente de carbono distinta del metanol, preferiblemente en ausencia de metanol; y reprimible en presencia de una cantidad

en exceso de una fuente de carbono distinta del metanol que es superior a la cantidad limitante del crecimiento. Específicamente, la expresión de GDI por el casete de expresión heterólogo es inducible por el ECP inducible.

Preferiblemente, la ECP es una fuente de carbono regulable, tal como reprimida en presencia de cantidades superiores a cualquiera de 1, 1,5, 2, 2,5 o 3 g/L de una fuente de carbono en el medio de cultivo celular o sobrenadante (en el presente documento denominado como una cantidad promotora-represora), e inducida o sin reprimir en presencia de una fuente de carbono no detectable o cantidades de hasta cualquiera de 0,1, 0,2, 0,3, 0,4, 0,5, 0,6, 0,7, 0,8, 0,9 o 1,0 g /L de fuente de carbono en el medio de cultivo celular o sobrenadante (en el presente documento se denomina cantidad inductora del promotor). Tales cantidades en el medio de cultivo celular o sobrenadante se entienden particularmente como la cantidad que tras la alimentación de la célula huésped y el consumo por parte de la célula huésped puede ser detectable. Por lo general, cuando se produce una PDI en condiciones que limitan el crecimiento, el cultivo celular se alimenta añadiendo una fuente de carbono suplementaria, pero en una cantidad que las células consumen inmediatamente durante la producción de la PDI, por lo que no queda nada o solo queda una cantidad baja en el mismo medio de cultivo celular o sobrenadante, *p.ej.* una cantidad de hasta 1,0 g/L.

Específicamente, la fuente de carbono que regula la ECP es cualquier otra que no sea metanol, y en el presente documento se denomina fuente de carbono distinta del metanol.

Específicamente, la fuente de carbono distinta del metanol es un carbohidrato.

Específicamente, la fuente de carbono distinta del metanol se selecciona de sacáridos, polioles, alcoholes o mezclas de cualquiera o más de los anteriores.

Específicamente, los sacáridos pueden ser cualquiera o más de los monosacáridos, tal como una hexosa, *p. ej.* glucosa, fructosa, galactosa o manosa, o disacáridos, tal como sacarosa; o un alcohol o poliol *p. ej.*, etanol o cualquier diol o triol, *p. ej.*, glicerol o una mezcla de cualquiera de los anteriores. Específicamente, cualquier fuente de carbono que no sea metanol puede usarse en el cultivo celular en una cantidad para producir dicha PDI bajo el control de la ECP.

Según un aspecto específico, la ECP comprende al menos una primera y al menos una segunda región reguladora central. Específicamente, el ECP comprende al menos dos de dichas regiones reguladoras principales primera y/o segunda. Específicamente, el ECP comprende un número limitado de dicha primera y segunda región reguladora central, en donde el número de dicha primera región reguladora central es solo uno, dos o tres, y el número de dicha segunda región reguladora central es solo uno, dos o tres. Específicamente, el ECP comprende un número igual de dicha primera y segunda región reguladora central, por ejemplo, en donde el número de dicha primera región reguladora central es uno, y el número de dicha segunda región reguladora central es uno; en donde el número de dicha primera región reguladora central es dos, y el número de dicha segunda región reguladora central es dos, o en donde el número de dicha primera región reguladora central es tres, y el número de dicha segunda región reguladora central es tres.

Específicamente, la primera región reguladora central tiene al menos un 75% de identidad de secuencia con la SEQ ID NO: 17, tal como al menos cualquiera de al menos un 80% o al menos un 90% de identidad de secuencia, y/o la segunda región reguladora central tiene al menos 75% de identidad de secuencia con la SEQ ID NO: 18, tal como al menos cualquiera de al menos 80% o al menos 90% de identidad de secuencia.

Específicamente, cada una de las regiones reguladoras centrales primera y segunda tiene una longitud de 8-16 nt.

Específicamente, la primera región reguladora central tiene una longitud de 8 a 10 nt, en particular de 9 nt. Específicamente, la primera región reguladora central comprende o consiste en SEQ ID NO: 17, o una modificación de la secuencia de nucleótidos identificada como SEQ ID NO: 17, en donde la modificación es hasta una o dos mutaciones puntuales, en particular en donde una mutación puntual es cualquiera de sustitución, inserción o eliminación de un nucleótido.

Específicamente, la segunda región reguladora central tiene una longitud de 14 a 16 nt, en particular 15 nt. Específicamente, la segunda región reguladora central comprende o consiste en SEQ ID NO: 18, o una modificación de la secuencia de nucleótidos identificada como SEQ ID NO: 18, en donde la modificación es hasta una, dos o tres mutaciones puntuales, en particular en donde uno mutación puntual es cualquiera de sustitución, inserción o eliminación de un nucleótido.

Específicamente, el ECP comprende al menos una primera y al menos una segunda región reguladora central en cualquier orden, preferiblemente en estrecha proximidad entre sí, p. ej., con hasta cualquiera de 20, 19, 18, 17, 16, 15, 14, 13, 12, 11 o 10 nt entre la primera y la segunda regiones reguladoras centrales que son más cercanas.

Específicamente, la ECP comprende una primera y una segunda región reguladora central, que están unidas a través de un espaciador, en particular separadas por una secuencia de nucleótidos (denominada en el presente documento "región central espaciadora") con una longitud de al menos cualquiera de 5, 6, 7, 8, 9, 10 nt y/o hasta cualquiera de 20, 19, 18, 17, 16, 15, 14, 13, 12, 11 o 10. En particular, la región del núcleo espaciador tiene al menos 75% de identidad de secuencia con SEQ ID NO: 36, tal como al menos cualquiera de al menos 80% o al menos 90% de identidad de secuencia, y/o comprende o consiste en SEQ ID NO: 36, o una modificación de la secuencia de

nucleótidos identificada como SEQ ID NO:36, en donde la modificación es hasta una, dos, tres o cuatro mutaciones puntuales, en particular donde una mutación puntual es cualquiera de sustitución, inserción o eliminación de un nucleótido. Específicamente, la región central del espaciador comprende o consiste en una secuencia de nucleótidos en donde la mayoría de los nucleótidos (al menos el 50% o al menos el 60%) se seleccionan de G, C o T.

Específicamente, el ECP comprende al menos una región que consiste en una secuencia de nucleótidos que, desde el extremo 5' hasta el extremo 3' consiste en los siguientes tres elementos contiguos, (i) una primera región reguladora central, (ii) una región central espaciadora, y (iii) la segunda región reguladora central, que en el presente documento se denomina "región reguladora principal".

Específicamente, el ECP comprende al menos una o dos, o solo una o dos regiones reguladoras principales, cada una de las cuales comprende o consiste en una secuencia de nucleótidos que tiene al menos una identidad de secuencia del 85%, 90% o 95% con la SEQ ID NO :35. Dicha región reguladora principal consiste preferiblemente en la primera región reguladora central, la región central espaciadora y la segunda región reguladora central.

Específicamente, la ECP comprende al menos una secuencia de polinucleótidos que tiene al menos una identidad de secuencia del 85%, 90% o 95% con la SEQ ID NO:35.

Específicamente, el ECP comprende solo una, dos o tres regiones reguladoras principales, como se describe en el presente documento, en particular en donde el número de regiones reguladoras principales es dos o tres, en donde dichas dos o tres regiones reguladoras pueden ser idénticas o diferentes entre sí. Específicamente, el ECP comprende solo dos regiones reguladoras principales, que están separadas por una secuencia de nucleótidos (en el presente documento denominada "región principal espaciadora") con una longitud de al menos cualquiera de 50, 60, 70, 80, 90, 100 nt, y/o hasta cualquiera de 500, 450, 400, 350 o 300 nt, en particular que varía entre 100 y 300 nt. Específicamente, la región principal del espaciador comprende una secuencia de nucleótidos de al menos cualquiera de 50, 60, 70, 80, 90, 100 nt de longitud que tiene al menos un 60% de identidad de secuencia con la SEQ ID NO: 37, como al menos cualquiera de al menos 70%, 75%, 80%, 85%, 90% o 95% de identidad de secuencia, y/o comprende o consiste en la SEQ ID NO: 37, o una modificación de la secuencia de nucleótidos identificada como SEQ ID NO: 37, en donde la modificación es un número de mutaciones puntuales, que es una o más, hasta cualquiera de 30, 25, 20, 15 o 10 mutaciones puntuales, en particular en donde una mutación puntual es cualquiera de sustitución, inserción, o deleción de un nucleótido.

Según un aspecto específico, la ECP comprende al menos un motivo T que consiste en una secuencia de nucleótidos en la que la mayoría de los nucleótidos es una timina (T), preferiblemente al menos cualquiera de 50%, 60%, 70%, 80%, 90% o 100% es una T, por ejemplo, que comprende o consiste en cualquiera de las SEQ ID NO: 19-34, opcionalmente sin extensión de dicho motivo T por una o más timina adicional (o adyacente) en cualquiera de los 5' o 3' final de dicho motivo T.

Específicamente, el ECP comprende solo uno o dos de dichos motivos T, o al menos dos de dichos motivos T, hasta 4 o 3 motivos T, en donde dichos motivos T son idénticos o diferentes entre sí.

Específicamente, el ECP comprende al menos uno de los motivos T aguas arriba o aguas debajo de una región reguladora principal, p. ej., uno (o solo uno o dos) motivos T aguas arriba y uno (o solo uno o dos) motivos T aguas abajo de una región reguladora principal. región. Sin embargo, según una realización específica, el ECP comprende al menos uno de dichos motivos T, en particular solo uno o dos motivos T dentro de la región principal del espaciador.

Específicamente, el ECP tiene una longitud de al menos cualquiera de 350, 400, 450, 500, 550, 600, 650, 700, 850, 900, 950 o 1000 pb, p. ej., hasta 2000 pb o hasta 1500 pb.

Específicamente, el ECP comprende una secuencia de nucleótidos en el extremo 3', p. ej., de hasta 50, 40, 30, 20, 10, 9, 8, 7, 6 o 5 nt de longitud, incluido el extremo 3', que comprende al menos parte de un sitio de iniciación de la traducción, por ejemplo, una secuencia que es al menos cualquiera de 60%, 70%, 80%, 85% o al menos 90% idéntica a cualquiera de SEQ ID NO: 38, SEQ ID NO: 39, o SEQ ID NO:40. Un sitio de iniciación de la traducción puede ser una secuencia consenso de Kozak en eucariotas y una secuencia promotora adecuada para apoyar la expresión génica.

Según un aspecto específico, la ECP comprende al menos una identidad de secuencia del 60%, 65%, 70%, 75% u 80%, en particular, al menos una identidad de secuencia del 85%, 90% o 95%., o es 100% idéntico, al menos 300 nt (consecutivos), en particular al menos cualquiera de 300, 350, 400, 450, 500, 550, 600, 650, 700, 850, 900, 950 o 1000 nt, p. ej., dentro de una región que comprende un sitio de unión del factor de transcripción (TFBS), y/o dentro de la secuencia terminal 3' que incluye el extremo 3', de cualquiera de las secuencias SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO: 12, SEQ ID NO: 13, SEQ ID NO: 14, SEQ ID NO: 15 o SEQ ID NO: 16, o cualquiera de SEQ ID NO: 41-45, hasta la longitud total de cualquiera de las secuencias de nucleótidos anteriores, en particular si la longitud total es inferior a 1000 nt.

Según un aspecto específico, la ECP comprende al menos una identidad de secuencia del 60%, 65%, 70%, 75% u 80%, en particular, al menos una identidad de secuencia del 85%, 90% o 95%, o es 100% idéntico, a cualquiera de las secuencias de longitud completa SEQ ID NO: 10, SEQ ID NO: 11, SEQ ID NO: 12, SEQ ID NO: 13, SEQ ID NO: 14, SEQ ID NO: 15, o SEQ ID NO:16, o cualquiera de SEQ ID NO:41-45.

Una realización específica se refiere a la ECP que comprende o consiste en la SEQ ID NO: 10 o SEQ ID NO: 11 o que comprende o consiste en una secuencia de nucleótidos que tiene al menos cualquiera de 60%, 65%, 70%, 75%, 80%, 85%, 90% o 95% de identidad de secuencia, o es 100% idéntica, a cualquiera de las secuencias de longitud completa de SEQ ID NO: 10, SEQ ID NO: 11, SEQ ID NO: 12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15 o SEQ ID NO:16, o cualquiera de SEQ ID NO:41-45, respectivamente.

Específicamente, el promotor ECP consta de al menos una parte o fragmento de la SEQ ID NO: 10 o SEQ ID NO: 11 con una longitud de al menos cualquiera de 300, 400, 500, 600, 700, 800, 900 o 1000 pb, en particular que incluye un TFBS y/o el terminal 3'.

Específicamente, cualquiera de las regiones reguladoras centrales primera y segunda del ECP, o la región reguladora principal del ECP, contiene uno o más TFBS. Específicamente, cada una de dichas regiones reguladoras principales primera y segunda, o ambas regiones reguladoras centrales primera y segunda juntas, o cada una de dichas regiones reguladoras principales del ECP, comprende un TFBS o al menos una parte del mismo que se considera funcional y está siendo reconocido por el respectivo factor de transcripción.

Específicamente, el TFBS se reconoce por cualquiera o más de los factores de transcripción seleccionados del grupo que consiste en Rgt1 (*p. ej.*, que comprende o consiste en la SEQ ID NO: 47), Cat8-1 (*p.ej.*, que comprende o consiste en la SEQ ID NO: 48) y Cat8-2 (*p. ej.*, que comprende o consiste en la SEQ ID NO: 49).

Un TFBS se caracteriza por ciertas secuencias de consenso, que pueden variar para el mismo factor. Los factores de transcripción específicos se identifican de la siguiente manera: Rgt1 es un activador y represor transcripcional sensible a la glucosa y regula la expresión de varios genes transportadores de glucosa (HXT). Rgt1 de *P. pastoris* comprende la secuencia de aminoácidos SEQ ID NO:47.

Cat8-1 y Cat8-2 son activadores transcripcionales de grupos de zinc que se unen a los elementos de respuesta de la fuente de carbono, necesarios para la desrepresión de una variedad de genes en condiciones de crecimiento no fermentativas. Cat8-1 y Cat8-2 de *P. pastoris* comprenden las secuencias de aminoácidos SEQ ID NO: 48 y SEQ ID NO: 49, respectivamente.

Según un aspecto específico, el ECP comprende al menos dos, tres, cuatro, cinco, seis, siete u ocho TFBS, en el que cada uno de los TFBS es reconocido individualmente por cualquiera de Rgt1, Cat8-1 o Cat8-2.

Específicamente, el ECP se caracteriza por una mayor fuerza del promotor en comparación con un promotor de referencia, en donde

- la fuerza del promotor es igual o mayor que la fuerza del promotor de referencia, en particular es al menos cualquiera de 1,1 veces, 1,2 veces, 1,3 veces, 1,4 veces, 1,5 veces, 1,6 veces, 1,7 veces, 1,8 veces, 1,9 veces, 2 veces, 2,1 veces, 2,2 veces, 2,3 veces, 2,4 veces, 2,5 veces, 2,6 veces, 2,7 veces, 2,8 veces, 2,9 veces, 3 veces, 3,3 veces, 3,5 veces, 3,8 veces, 4 veces, 4,5 veces, 5 veces, 5,5 veces o al menos 6 veces mayor cuando está en el estado inducido.

En particular, el promotor de pGAP nativo de la célula huésped, específicamente un promotor de pGAP que es endógeno y que ocurre naturalmente en la célula huésped que se usa para la producción de PDI recombinante *p. ej.*, el promotor pGAP nativo de *P. pastoris* como se usa para controlar la expresión de GAPDH en *P. pastoris,* que comprende o consiste en la SEQ ID NO: 46, puede servir como referencia en un *P. pastoris* célula huésped, para determinar la potencia mejorada del promotor ECP. Tal promotor de referencia puede usarse en experimentos de control paralelos usando la misma célula huésped y sistema de expresión, o como control interno dentro del mismo cultivo de células huésped. Tales experimentos de control para calificar la función del promotor en comparación con el promotor de referencia se llevan a cabo preferiblemente en cultivos celulares huésped *de P. pastoris*, en particular recombinantes de *P. pastoris* expresando una proteína modelo, tal como GFP o eGFP. La fuerza del promotor en comparación con la fuerza del promotor de referencia se puede determinar mediante el siguiente ensayo estándar: cepas *P. pastoris* que expresan eGFP bajo el control del promotor que se va a probar se criban en placas de 24 pocillos profundos a 25 °C con agitación a 280 rpm con 2 mL de cultivo por pocillo. Se utilizan perlas de alimentación de glucosa (6 mm, Kuhner, CH) para generar condiciones de crecimiento limitantes de la glucosa. Las células se analizan para determinar la expresión de eGFP en el estado inducido (YP + 1 perla de alimentación, durante 20-28 horas).

Según un aspecto específico, el promotor relativo o fuerza o velocidad de transcripción del ECP descrito aquí se compara con el promotor pGAP nativo de una célula de la misma especie o cepa que se usa como huésped para producir una PDI.

Específicamente, el promotor de referencia es el promotor pGAP nativo de la célula huésped. Por ejemplo, un promotor pGAP nativo de *P. pastoris* que es la secuencia promotora endógena no modificada en *P. pastoris,* como se usa para controlar la expresión de GAPDH en *P. pastoris* (GS115), *p. ej.* que comprende o consiste en la secuencia identificada como SEQ ID NO: 46 puede usarse como promotor de referencia en *P. pastoris.* Si *P. pastoris* se usa como una célula huésped recombinante para producir una PDI como se describe en este documento, la fuerza o velocidad de transcripción del ECP descrito en el presente documento se compara convenientemente con dicho promotor pGAP nativo de *P. pastoris.*

Ejemplo de secuencia de promotor pGAP nativo de *P. pastoris* (GS115) (SEQ ID NO: 46)

#	Nombre	PAS*	PIPA*	Descripción GS115
pGAP	TDH3	PAS chr2-1 0437	PIPA02510	Gliceraldehído-3-fosfato deshidrogenasa
*PAS: ORF nombre en *P. pastoris* GS115; PIPA: ORF nombre en cepa tipo *P. pastoris* DSMZ70382				

Según otro ejemplo, un promotor pGAP nativo de *S. cerevisiae* puede usarse como promotor de referencia, que es la secuencia promotora endógena no modificada en *S. cerevisiae,* como se usa para controlar la expresión de GAPDH en *S. cerevisiae.* Si *S. cerevisiae* se usa como una célula huésped recombinante para producir una PDI como se describe en el presente documento, la fuerza o velocidad de transcripción del ECP descrito en el presente documento se compara convenientemente con dicho promotor pGAP nativo de *S. cerevisiae.*

Específicamente, la fuerza del promotor está determinada por el nivel de expresión de una PDI, tal como una proteína modelo (p. ej., Green Fluorescence Protein, GFP, que incluye *p. ej.*, GFP mejorado, eGFP, Gene Bank Accession no. U57607), y/o la fuerza de transcripción, en comparación con el promotor de referencia. Preferiblemente, el análisis de transcripción es cuantitativo o semicuantitativo, empleando preferiblemente qRT-PCR, micromatrices de ADN, secuenciación de ARN y análisis de transcriptoma.

Específicamente, la ECP se caracteriza además por una proporción de inducción de promotor que se caracteriza por una alta fuerza de transcripción en el estado totalmente inducido, en comparación con un bajo nivel en el estado reprimido.

La relación de inducción del promotor se refiere específicamente a la inducción de la transcripción, incluyendo específicamente la traducción adicional y la expresión opcional de dicho PDI. La transcripción normalmente se determina como una medida de la fuerza del promotor y se refiere específicamente a la cantidad de transcritos obtenidos tras la inducción completa de dicho promotor. Dicha abundancia de transcritos puede determinarse por la fuerza de transcripción en el estado completamente inducido, que se obtiene, por ejemplo, en condiciones de cultivos de quimiostatos limitados en glucosa y se expresa en relación con la velocidad de transcripción de un promotor de referencia.

La relación de inducción es un parámetro clave para determinar la regulación de la fuente de carbono de la ECP y establece la actividad o fuerza del promotor en el estado inducido en relación con la actividad o fuerza del promotor en el estado reprimido. Por ejemplo, el nivel de expresión de una proteína indicadora (p. ej., GFP o eGFP) y/o el nivel de transcripción en el estado reprimido se determina tras la represión por exceso de glicerol, y el nivel de expresión de la proteína modelo y/o el nivel de transcripción es determinado en el estado inducido tras la inducción limitando la alimentación de glucosa.

Se considera que el promotor de ECP está desreprimido y totalmente inducido, si las condiciones de cultivo proporcionan una inducción máxima, p. ej., a concentraciones de glucosa inferiores a 0,4 g/L, preferiblemente inferiores a 0,04 g/L, concretamente inferiores a 0,02 g/L. El promotor completamente inducido muestra preferiblemente un nivel/fuerza de transcripción de al menos 20%, más preferiblemente al menos 30%, 40%, 50%, 60%, 70%, 80%, 90% y al menos 100% o incluso mayor nivel/fuerza de transcripción de al menos 150% o al menos 200% en comparación con el promotor pGAP nativo. El nivel/fuerza de la transcripción puede, por ejemplo, determinarse por la cantidad de transcritos de un gen informador, tal como eGFP al cultivar un clon en cultivo líquido. Alternativamente, la velocidad de transcripción puede determinarse por la fuerza de transcripción del gen controlado de forma nativa en un microarreglo, en donde los datos del microarreglo muestran la diferencia del nivel de expresión entre el estado reprimido y desreprimido y una alta intensidad de señal en el estado completamente inducido en comparación con un control.

Específicamente, la relación de inducción se puede determinar mediante la relación del nivel de expresión (p. ej., de una proteína modelo como GFP o eGFP) en el estado inducido frente al reprimido. La relación de inducción en comparación con un promotor de referencia se puede determinar mediante el siguiente ensayo estándar: cepas de *P. pastoris* que expresan eGFP bajo el control del promotor que se va a probar se criban en placas de 24 pocillos profundos a 25 °C con agitación a 280 rpm con 2 ml de cultivo por pocillo. Se utilizan perlas de alimentación de glucosa (6 mm, Kuhner, CH) para generar condiciones de crecimiento limitantes de la glucosa. Las células se analizan para determinar la expresión de eGFP durante la represión (YP + glicerol al 1%, fase exponencial) y la inducción (YP + 1 perla de alimentación, durante 20 a 28 horas).

Específicamente, el promotor ECP tiene una actividad o fuerza de promotor (p. ej., actividad transcripcional o fuerza de transcripción) en el estado desreprimido (inducido), que es al menos cualquiera de 1,5, 2,0, 2,5, 3, 3,5, 4, 4,5, 5, 5,5, 6, 6,5, 7, 7,5, 8, 8,5, 9, 9,5 o 10 veces mayor que en el estado reprimido. Por lo tanto, la velocidad de inducción respectiva puede ser al menos cualquiera de 1,5, 2,0, 2,5, 3, 3,5, 4, 4,5, 5, 5,5, 6, 6,5, 7, 7,5, 8, 8,5, 9, 9,5 o 10.

Sorprendentemente, resultó que la actividad transcripcional (o fuerza de transcripción) de una ECP descrita en el presente documento (cuando se induce completamente, p. ej., bajo condiciones inductoras de limitación de glucosa) en una cepa knockout de *flo8* en comparación con una cepa de tipo salvaje que comprende un locus de *flo8* y produce proteína FLO8 es mucho mayor, como al menos cualquiera de 1,5 veces, 1,6 veces, 1,7 veces, 1,8 veces, 1,9 veces, 2 veces, 3 veces, 4 veces, 5- veces, 6 veces, 7 veces, 8 veces, 9 veces o 10 veces más. Se encontró que la fuerza de transcripción en la célula huésped descrita en el presente documento aumentó significativamente, lo que permitió niveles de expresión más altos de un GDI bajo el control del ECP en un *flo8* mutante por deleción. Por el contrario, la transcripción de pGAP o pAOX no aumenta en un *flo8* mutante por deleción (como se determina en un ejemplo comparable).

Según un aspecto concreto, el casete de expresión heterólogo está comprendido en un vector o plásmido de replicación autónoma, o integrado en un cromosoma de dicha célula huésped.

El casete de expresión puede introducirse en la célula huésped e integrarse en el genoma de la célula huésped (o cualquiera de sus cromosomas) como elemento intracromosómico. *p. ej.*, en un sitio específico de integración o integrado al azar, después de lo cual se selecciona una línea de células huésped de alta producción. Alternativamente, el casete de expresión puede integrarse dentro de un elemento genético extracromosómico, tal como un plásmido o un cromosoma artificial. *p. ej.*, un cromosoma artificial de levadura (YAC). Según un ejemplo específico, el casete de expresión se introduce en la célula huésped mediante un vector, en particular un vector de expresión, que se introduce en la célula huésped mediante una técnica de transformación adecuada. Para este propósito, el GDI puede ligarse a un vector de expresión.

Un vector de expresión de levadura preferido (que se usa preferiblemente para la expresión en levadura) se selecciona del grupo que consiste en plásmidos derivados de pPICZ, pGAPZ, pPIC9, pPICZalfa, pGAPZalfa, pPIC9K, pGAPHis, pPUZZLE o Golden*Pi*CS.

Las técnicas para transfectar o transformar células huésped para introducir un vector o plásmido son bien conocidas en la técnica. Estos pueden incluir electroporación, esferoplastia, captación mediada por vesículas lipídicas, captación mediada por choque térmico, transfección mediada por fosfato de calcio (coprecipitación de fosfato de calcio/ADN), infección vírica y, en particular, el uso de virus modificados como, por ejemplo, adenovirus modificados, microinyección y electroporación.

Los transformantes como se describen en el presente documento se pueden obtener introduciendo el casete de expresión, el vector o el ADN del plásmido en un huésped y seleccionando los transformantes que expresan la proteína o el marcador de selección relevante. Las células huésped pueden tratarse para introducir ADN heterólogo o extraño mediante métodos usados convencionalmente para la transformación de células huésped, tales como el método del pulso eléctrico, el método del protoplasto, el método del acetato de litio y métodos modificados de los mismos. *P. pastoris* se transforma preferiblemente por electroporación. Los métodos preferidos de transformación para la captación del fragmento de ADN recombinante por el microorganismo incluyen transformación química, electroporación o transformación por protoplastificación.

Específicamente, el casete de expresión comprende el ECP unido operativamente al GDI que codifica el PDI, y opcionalmente comprende además secuencias señal y líder, según sea necesario para expresar y producir la PDI como una proteína secretada.

Según un aspecto específico, el casete de expresión comprende además una secuencia de nucleótidos que codifica un péptido señal que permite la secreción de la PDI preferiblemente en donde la secuencia de nucleótidos que codifica el péptido señal está fusionada adyacente o directamente al extremo 5' del GDI.

Específicamente, el péptido señal se selecciona del grupo que consiste en secuencias señal del prepropéptido del factor de apareamiento alfa de *S. cerevisiae*, los péptidos señal del gen de la fosfatasa ácida (PHO1) de *P. pastoris* y la proteína extracelular X (EPX1) (Heiss, S., V. Puxbaum, C. Gruber, F. Altmann, D. Mattanovich y B. Gasser, Microbiology 2015; 161(7): 1356-68).

Específicamente, cualquiera de las secuencias señal y/o líder como se describe en el documento WO2014067926 A1 se puede utilizar, en particular la SEQ ID NO:59 o SEQ ID NO:60.

Específicamente, las secuencias de señal como se describe en el documento WO2012152823 A1 puede utilizarse, en particular, la secuencia de señal del factor de apareamiento alfa nativo de *S. cerevisiae* identificado como SEQ ID NO:61, o mutantes de los mismos.

Según un aspecto específico, la célula huésped descrita en el presente documento puede sufrir una o más modificaciones genéticas adicionales. *p. ej.*, para mejorar la producción de proteínas.

Específicamente, la célula huésped se diseña genéticamente adicionalmente para modificar uno o más genes que influyen en la actividad proteolítica utilizada para generar cepas deficientes en proteasa, en particular una cepa deficiente en actividad de carboxipeptidasa Y. Se describen ejemplos particulares en el documento WO1992017595A1. Más ejemplos de una cepa de Pichia deficiente en proteasa con una deficiencia funcional en una

proteasa vacuolar, tal como la proteinasa A o la proteinasa B, se describen en el documento US6153424A. Otros ejemplos son las cepas Pichia que tienen una deleción *ade2* y/o deleciones de uno o ambos genes de proteasa, *PEP4* y *PRB1,* son proporcionados por *p. ej.*, ThermoFisher Scientific.

Específicamente, la célula huésped está diseñada genéticamente para modificar al menos una secuencia de ácido nucleico que codifica un producto génico funcional, en particular una proteasa, seleccionada del grupo que consiste en PEP4, PRB1, YPS1, YPS2, YMP1, YMP2, YMP1, DAP2, GRHl, PRD1, YSP3 y PRB3, como se describe en el documento WO2010099195A1.

La sobreexpresión o subexpresión de genes que codifican factores auxiliares se aplica específicamente para mejorar la expresión de un GDI, p.ej., como se describe en el documento WO2015158800A1.

Se demostró que la sobreexpresión de los siguientes genes aumenta la secreción de PDI en *P. pastoris:* PP7435_Chr3-0607, PP7435_Chr3-0933, PP7435_Chr2-0220, PP7435_Chr3-0639, PP7435_Chr4-0108, PP7435_Chr1-1232, PP7435_Chr1-1225, PP7435_Chr1-0667, y PP7435_Chr4-0448.

Se demostró que la subexpresión de los siguientes genes aumenta la secreción de PDI en *P. pastoris:* PP7435_Chr1-0176, PP7435_Chr3-1062 y PP7435_Chr4-0252.

En particular, la célula huésped puede diseñarse genéticamente para sobreexpresar cualquiera o más de los factores auxiliares y aumentar la producción de las respectivas proteínas identificadas por cualquiera de las SEQ ID NO: 62-71, aumentando así aún más el rendimiento de PDI. La PDI puede ser cualquiera de péptidos, polipéptidos, proteínas o metabolitos eucarióticos, procarióticos o sintéticos de una célula huésped.

Específicamente, la PDI es heteróloga a la especie de célula huésped.

Específicamente, la PDI es un péptido, polipéptido o proteína secretado, es decir, secretado desde la célula huésped al sobrenadante del cultivo celular.

Específicamente, la PDI es una proteína eucariótica, preferiblemente una proteína derivada o relacionada de un mamífero tal como una proteína humana o una proteína que comprende una secuencia de proteína humana, o una proteína bacteriana o una proteína derivada de una bacteria.

Preferiblemente, la PDI es una proteína terapéutica que funciona en mamíferos.

En casos específicos, la PDI es una proteína multimérica, concretamente un dímero o tetrámero.

Según un aspecto específico, la PDI es un péptido o proteína seleccionada del grupo que consiste en una proteína de unión a antígeno, una proteína terapéutica, una enzima, un péptido, una proteína antibiótica, una proteína de fusión de toxina, un carbohidrato - proteína conjugada, una proteína estructural, una proteína reguladora, un antígeno de vacuna, un factor de crecimiento, una hormona, una citoquina, una enzima de proceso y una enzima metabólica.

Específicamente, la proteína de unión a antígeno se selecciona del grupo que consiste en

a) anticuerpos o fragmentos de anticuerpos, tal como cualquiera de los anticuerpos quiméricos, anticuerpos humanizados, anticuerpos biespecíficos, Fab, Fd, scFv, diacuerpos, triacuerpos, tetrámeros Fv, minicuerpos, anticuerpos de dominio único como VH, VHH, IgNAR o V -NAR;

b) miméticos de anticuerpos, tales como Adnectinas, afficuerpos, affilinas, affimeros, affitinas, alfacuerpos, anticalinos, avimeros, DARPins, fynómeros, péptidos del dominio de Kunitz, monocuerpos o NanoCLAMPS; o

c) proteínas de fusión que comprenden uno o más dominios de inmunoglobulina, dominios de anticuerpos o miméticos de anticuerpos.

Una PDI específica es una molécula de unión a antígeno tal como un anticuerpo, o un fragmento del mismo, en particular un fragmento de anticuerpo que comprende un dominio de unión a antígeno. Entre las PDI específicas se encuentran anticuerpos tal como los anticuerpos monoclonales (mAb), inmunoglobulina (Ig) o inmunoglobulina de clase G (IgG), anticuerpos de cadena pesada (HcAb) o fragmentos de los mismos, tal como la unión de fragmento-antígeno (Fab), Fd, single- fragmento de cadena variable (scFv), o variantes modificadas del mismo como, por ejemplo, dímeros Fv (diacuerpos), trímeros Fv (triacuerpos), tetrámeros Fv o minicuerpos y anticuerpos de dominio único como VH, VHH, IgNAR o V-NAR, o cualquier proteína que comprenda un dominio de pliegue de inmunoglobulina. Se pueden seleccionar moléculas de unión a antígeno adicionales de miméticos de anticuerpos o proteínas de armazón (alternativas) tales como, por ejemplo, dominios de Kunitz modificados genéticamente, Adnectinas, Afficuerpos, Affilina, Anticalinos o DARPins.

Según un aspecto específico, la PDI es, p. ej., BOTOX, Myobloc, Neurobloc, Dysport (u otros serotipos de neurotoxinas botulínicas), alglucosidasa alfa, daptomicina, YH-16, coriogonadotropina alfa, filgrastim, cetrorelix, interleucina-2, aldesleucina, teceleulina, denileucina diftitox, interferón alfa-n3 (inyección), interferón alfa-nl, DL-8234, interferón, Suntory (gamma-1a), interferón gamma, timosina alfa 1, tasonermina, DigiFab, ViperaTAb, EchiTAb, CroFab,

nesiritida, abatacept, alefacept, Rebif, eptoterminalfa, teriparatida (osteoporosis), calcitonina inyectable (enfermedad ósea), calcitonina (nasal, osteoporosis), etanercept, hemoglobina glutámera 250 (bovina), drotrecogina alfa, colagenasa, carperitida, factor de crecimiento epidérmico humano recombinante (gel tópico, cicatrización de heridas), DWP401, darbepoetina alfa, epoetina omega, epoetina beta, epoetina alfa, desirudina, lepirudina, bivalirudina, nonacog alfa, mononina, eptacog alfa (activado), factor VIII+FVW recombinante, recombinado, factor recombinante VIII, Factor VIII (recombinante), Alphnmate, octocog alfa, Factor VIII, palifermina, indiquinasa, tenecteplasa, alteplasa, pamiteplasa, reteplasa, nateplasa, monteplasa, folitropina alfa, rFSH, hpFSH, micafungina, pegfilgrastim, lenograstim, nartograstim, sermorelina, glucagón, exenatida, pramlintida, iniglucerasa, galsulfasa, leucotropina, molgramostirna, acetato de triptorelina, histrelina (implante subcutáneo, Hydron), deslorelina, histrelina, nafarelina, depósito de liberación sostenida de leuprolida (ATRIGEL), implante de leuprolida (DUROS), goserelina, Eutropina, programa KP-102, somatropina, mecasermina (falta de crecimiento), enlfavirtida, Org-33408, insulina glargina, insulina glulisina, insulina (inhalada), insulina lispro, insulina deternir, insulina (bucal, RapidMist), mecasermina rinfabate, anakinra, celmoleukin, inyección de 99 mTc-apcitide, mielopid, Betaseron, acetato de glatiramer, Gepon, sargramostim, oprelvekin, interferones alfa derivados de leucocitos humanos, Bilive, insulina (recombinante), insulina humana recombinante, insulina aspart, mecasenina, Roferon-A, interferon-alfa 2, Alfaferone, interferon alfacon-1, interferon alfa, hormona luteinizante humana recombinante Avonex', dornasa alfa, trafermin, ziconotide, taltirelin, diboterminalfa, atosiban, becaplermin, eptifibatide, Zemaira, CTC-111, Shanvac-B, vacuna contra el VPH (tetravalente), octreotida, lanreotida, ancestro, agalsidasa beta, agalsidasa alfa, laronidasa, prezatida acetato de cobre (gel tópico), rasburicasa, ranibizumab, Actimmune, PEG-Intron, Tricomin, casa recombinante inyección de desensibilización de la alergia a los ácaros del polvo, hormona paratiroidea humana recombinante (PTH) 1-84 (sc, osteoporosis), epoetina delta, antitrombina III transgénica, granditropina, Vitrase, insulina recombinante, interferón-alfa (pastilla oral), GEM-21S, vapreotida, idursulfasa, omniapatrilato, albúmina sérica recombinante, certolizumab pegol, glucarpidasa, inhibidor de la esterasa C1 recombinante humana (angioedema), lanoteplasa, hormona de crecimiento humana recombinante, enfuvirtida (inyección sin aguja, Bioject o 2000), VGV-1, interferón (alfa), lucinactant, aviptadil (inhalado, enfermedad pulmonar), icatibant, ecallantide, omiganan, Aurograb, pexigananacetate, ADI-PEG-20, LDI-200, degarelix, cintredelinbesudotox, Favld, MDX-1379, ISAtx-247, liraglutida, teriparatida (osteoporosis), tifacogina, AA4500, loción liposómica T4N5, catumaxomab, DWP413, ART-123, crisalina, desmoteplasa, amediplasa, corifolitropina alfa, TH-9507, teduglutida, Diamyd, DWP-412, hormona del crecimiento (inyección de liberación sostenida), G-CSF recombinante, insulina (inhalada, AIR), insulina (inhalada, Technosphere), insulina (inhalada, AERx), RGN-303, DiaPep277, interferón beta (infección viral por hepatitis C (VHC)), interferón alfa-n3 (oral), belatacept, parches transdérmicos de insulina, AMG-531, MBP-8298, Xerecept, opebacan, AIDSVAX, GV-1001, LymphoScan, ranpirnasa, Lipoxysan, lusupultide, MP52 (portador de fosfato tricálcico beta, regeneración de hueso), vacuna contra el melanoma, sipuleucel-T, CTP-37, Insegia, vitespen, trombina humana (congelada, sangrado quirúrgico), trombina, TransMID, alfimeprasa, Puricase, terlipresina (intravenosa, síndrome hepatorrenal), EUR-1008M, FGF-I recombinante (inyectable, enfermedad vascular), BDM-E, rotigaptida, ETC-216, P-113, MBI-594AN, duramicina (inhalada, fibrosis quística), SCV-07, OPI-45, endostatina, angiostatina, ABT-510, concentrado inhibidor de Bowman Birk, XMP-629, 99 mTc-Hynic-Annexin V, kahalalide F, CTCE-9908, teverelix (liberación prolongada), ozarelix, rornidepsin, BAY-504798, interleukin4, PRX-321, Pepscan, iboctadekin, rhlactoferrin, TRU-015, IL-21, ATN-161, cilengitide, Albuferon, Biphasix, IRX-2, omega interferon, PCK-3145, CAP-232, pasireotida, huN901-DMI, vacuna inmunoterapéutica contra el cáncer de ovario, SB-249553, Oncovax-CL, OncoVax-P, BLP-25, CerVax-16, vacuna contra el melanoma peptídico multiepítopo (MART-1, gp100, tirosinasa), nemifitida, rAAT (inhalado), rAAT (dermatológico), CGRP (inhalado, asma), pegsunercept, timosinbeta4, plitidepsina, GTP-200, ramoplanina, GRASPA, OBI-1, AC-100, calcitonina de salmón (oral,eligen), calcitonina (oral, osteoporosis), examorelina, capromorelina, Cardeva, velafermina, 1311-TM-601, KK-220, T-10, ularitide, depelestat, hematide, Chrysalin (tópica), rNAPc2, factor V111 recombinante (pegilado liposomal), bFGF, variante de estafilocinasa recombinante PEGilada, V-10153, SonoLysis Prolyse, NeuroVax, CZEN-002, terapia de neogénesis de células de los islotes, rGLP-1, BIM-51077, LY-548806, exenatida (liberación controlada, Medisorb), AVE- 0010, GA-GCB, avorelin, ACM-9604, eacetato de linaclotid, CETi-1, Hemospan, VAL (inyectable), insulina de acción rápida (inyectable, Viadel), insulina intranasal, insulina (inhalada), insulina (oral, eligen), metionil leptina humana recombinante, inyección subcutánea de pitrakinra, eccema), pitrakinra (polvo seco inhalado, asma), Multikine, RG-1068, MM-093, NBI-6024, AT-001, PI-0824, Org-39141, Cpn10 (enfermedades autoinmunes/inflamación), talactoferrina (tópica), rEV-131 (oftálmica), rEV-131 (enfermedad respiratoria), insulina humana recombinante oral (diabetes), RPI-78M, oprelvekin (o ral), CYT-99007 CTLA4-lg, DTY-001, valategrast, interferón alfa-n3 (tópico), IRX-3, RDP-58, Tauferon, lipasa estimulada por sales biliares, Merispase, alalina fosfatasa, EP-2104R, Melanotan- II, bremelanotida, ATL-104, microplasmina humana recombinante, AX-200, SEMAX, ACV-1, Xen-2174, CJC-1008, dinorfina A, SI-6603, LAB GHRH, AER-002, BGC-728, vacuna contra la malaria (virosomas, PeviPRO), ALTU-135, vacuna contra el parvovirus B19, vacuna contra la influenza (neuraminidasa recombinante), vacuna contra la malaria/VHB, vacuna contra el ántrax, Vacc-5q, Vacc-4x, vacuna contra el VIH (oral), vacuna contra el VPH, Tat Toxoid, YSPSL, CHS-13340, crema liposomal PTH(1-34) (Novasome), Ostabolin-C, análogo de PTH (tópico, psoriasis), MBRI-93.02, vacuna MTB72F (tuberculosis), vacuna MVA-Ag85A (tuberculosis), FARA04, BA-210, vacuna recombinante contra la peste FIV, AG-702, OxSODrol, rBetV1, vacuna dirigida contra alérgenos Der-p1/Der-p2/Der-p7 (alergia a los ácaros del polvo), antígeno peptídico PR1 (leucemia), vacuna ras mutante, VPH- 16 vacuna de lipopéptido E7, vacuna laberíntica (adenocarcinoma), Vacuna CML, vacuna de péptido WT1 (cáncer), IDD-5, CDX-110, Pentrys, Norelin, CytoFab, P-9808, VT-111, icrocaptida, telbermina (dermatológica, úlcera del pie diabético), rupintrivir, reticulosa, rGRF, HA, alfa-galactosidasa A, ACE-011, ALTU-140, CGX-1160, vacuna terapéutica de angiotensina, D-4F, ETC-642, APP-018, rhMBL, SCV-07 (oral, tuberculosis), DRF-7295, ABT-828, inmunotoxina específica de ErbB2 (anticancerígena), DT3SSIL-3, TST-10088, PRO-1762, Combotox, péptidos de unión al receptor

de colecistoquinina-B/gastrina, 111In-hEGF, AE-37, trasnizumab-DM1, antagonista G, IL-12 (recombinante), PM-02734, IMP-321, rhIGF-BP3, BLX-883, CUV-1647 (tópico), radioinmunoterapéuticos basados en L-19 (cáncer), Re-188-P-2045, AMG- 386, vacuna DC/1540/KLH (cáncer), VX-001, AVE-9633, AC-9301, vacuna NY-ESO-1 (péptidos), péptidos NA17.A2, vacuna contra el melanoma (antígeno pulsado terapéutico), vacuna contra el cáncer de próstata, CBP-501, lactoferrina humana recombinante (ojo seco), FX-06, AP-214, WAP-8294A (inyectable), ACP-HIP, SUN-11031, péptido YY [3-36] (obesidad, intranasal), FGLL, atacicept, BR3-Fc, BN-003, BA-058, hormona paratiroidea humana 1-34 (nasal, osteoporosis), F-18-CCR1, AT-1100 (enfermedad celíaca/diabetes), JPD-003, crema liposomal PTH(7-34) (Novasome), duramicina (oftálmica, ojo seco), CAB-2, CTCE-0214, eritropoyetina glicopegilada, EPO-Fc, CNTO-528, AMG-114, JR-013, Factor XIII, aminocandina, PN-951, 716155, SUN-E7001, TH-0318, BAY-73-7977, teverelix (liberación inmediata), EP-51216, hGH (liberación controlada, Biosfera), OGP-I, sifuvirtida, TV4710, ALG-889, Org-41259, rhCC10, F-991, timopentina (enfermedades pulmonares), r(m)CRP, insulina hepatoselectiva, subalina, proteína de fusión L19-IL-2, elafina, NMK-150, ALTU-139, EN-122004, rhTPO, agonista del receptor de trombopoyetina (trastornos trombocitopénicos), AL-108, AL-208, antagonistas del factor de crecimiento nervioso (dolor), SLV-317, CGX-1007, INNO-105, teriparatida oral (eligen), GEM-OS1, AC-162352, PRX-302, vacuna de fusión LFn-p24 (Therapore), EP-1043, vacuna pediátrica S pneumoniae, vacuna contra la malaria, Neisseria vacuna contra la meningitidis del grupo B, vacuna contra el estreptococo del grupo B neonatal, vacuna contra el ántrax, vacuna contra el VHC (gpE1+gpE2+MF-59), terapia contra la otitis media, vacuna contra el VHC (antígeno central+ISCOMATRIX), hPTH(1-34) (transdérmica, ViaDerm), 768974, SYN-101, PGN-0052, aviscumnine, BIM-23190, vacuna contra la tuberculosis, péptido de tirosinasa multiepítopo, vacuna contra el cáncer, enkastim, APC-8024, GI-5005, ACC-001, TTS-CD3, vascular dirigido TNF (tumores sólidos), desmopresina (liberación controlada bucal), onercept o TP-9201, adalimumab (HUMIRA), infliximab (REMICADE™), rituximab (RITUXAN™/MAB THERA™), etanercept (ENBREL™), bevacizumab (AVASTIN™), trastuzumab (HERCEPTIN™), pegrilgrastim (NEULASTA™), o cualquier otro punto de interés adecuado, incluidos biosimilares y biomejores.

Según un aspecto específico, la célula huésped puede ser cualquier célula animal, una célula de vertebrado, una célula de mamífero, una célula humana, una célula vegetal, una célula nematoda, una célula de invertebrado como una célula de insecto o una célula de molusco, una célula madre célula derivada de cualquiera de las anteriores, o una célula fúngica o una célula de levadura. Específicamente, la célula huésped es una célula de un género seleccionado del grupo que consiste en Pichia, Hansenula, Komagataella, Saccharomyces, Kluyveromyces, Candida, Ogataea, Yarrowia y Geotrichum, específicamente *Saccharomyces cerevisiae, Pichia pastoris, Ogataea minuta, Kluyveromces lactis, Kluyverome marxianus, Yarrowia lipolytica* o *hansenula polymorpha,* o de hongos filamentosos como *Aspergillus awamori* o *Trichoderma reesei.* Preferiblemente, la célula huésped es una levadura metilotrófica, preferiblemente *Pichia pastoris.* Aquí en *Pichia pastoris* se usa como sinónimo de todo, *Komagataella pastoris, Komagataella phaffii* y *Komagataella pseudopastoris.*

Según un aspecto específico, la célula huésped es

a) una célula de levadura de un género seleccionado del grupo que consiste en Pichia, Hansenula, Komagataella, Saccharomyces, Kluyveromyces, Candida, Ogataea, Yarrowia y Geotrichum, tal como del género Pichia (p. *Pichia pastoris, Pichia* metanolica, *Pichia kluyveri,* y *Pichia angusta*), Género Komagataella (*p. ej., Komagataella pastoris, Komagataella pseudopastoris* o *Komagataella phaffii*), género Saccharomyces (*p. ej. Saccharomyces cerevisae, Saccharomyces kluyveri, Saccharomyces uvarum),* género Kluyveromyces (p. ej., *Kluyveromyces lactis, Kluyveromyces marxianus*), el género Candida (p. ej., *Candida utilis, Candida cacaoi, Candida boidinii,*), el género Geotrichum (p. ej., *Geotrichum fermentans*), *Hansenula polymorpha, Yarrowia lipolytica,* o *Schizosaccharomyces pombe.*; o

b) una célula de hongos filamentosos, tal como *Aspergillus awamori* o *Trichoderma reesei.*

Se prefiere la especie *Pichia pastoris.* Específicamente, la célula huésped es una cepa de *Pichia pastoris* seleccionada del grupo que consiste en CBS 704, CBS 2612, CBS 7435, CBS 9173-9189, DSMZ 70877, X-33, GS115, KM71, KM71H y SMD1168.

Fuentes: CBS 704 (=NRRL Y-1603 = DSMZ 70382), CBS 2612 (=NRRL Y-7556), CBS 7435 (=NRRL Y-11430), CBS 9173-9189 (cepas CBS: CBS-KNAW Fungal Biodiversity Centre, Centraalbureau voor Schimmelculturen, Utrecht, Países Bajos) y DSMZ 70877 (German Collection of Microorganisms and Cell Cultures); cepas de Invitrogen, como X-33, GS115, KM71, KM71H y SMD1168.

Ejemplos de cepas de S preferidos *cerevisiae* incluyen W303, CEN.PK y la serie BY (colección EUROSCARF). Todas las cepas descritas anteriormente se han utilizado con éxito para producir transformantes y expresar genes heterólogos.

Según un aspecto específico, la célula huésped eucariota puede ser una célula fúngica (*p. ej.*, Aspergillus (tal como *A. niger, A. fumigatus, A. oryzae, A. nidulans*), Acremonium (tal como *A. thermophilum*), Chaetomium (tal como *C. thermophilum*), Chrysosporium (como *C. termófilo*), Cordyceps (como *C. militaris*), Corynascus, Ctenomyces, Fusarium (tal como *F. oxysporum*), Glomerella (tal como *G. graminicola*), Hipocrea (tal como *H. jecorina*), Magnaporthe (tal como *M. oryzae*), Myceliophthora (tal como *M. termófilo*), Nectria (tal como *N. hematococa*), Neurospora (tal como

n. crasa), Penicillium, Sporotrichum (tal como S. *termófilo*), Thielavia (tal como T. terrestris, *T. heterothallica*), Trichoderma (tal como T. reesei) o Verticillium (como *V dalia*)).

Según un aspecto específico, la célula de mamífero es una célula, línea celular o cepa celular humana o de roedor o bovina. Ejemplos de células de mamífero específicas adecuadas como células huésped descritas en el presente documento son líneas celulares de mieloma de ratón (NSO), líneas celulares de ovario de hámster chino (CHO), HT1080, H9, HepG2, MCF7, MDBK Jurkat, MDCK, NIH3T3, PC12, BHK (célula de riñón de hámster bebé), VERO, SP2/0, YB2/0, Y0, C127, célula L, COS, p. ej., COS1 y COS7, QC1-3, HEK-293, VERO, PER.C6, HeLA, EBl, EB2, EB3, líneas celulares oncolíticas o de hibridoma. Preferiblemente, las células de mamífero son líneas de células CHO. En una realización, la célula es una célula CHO. En una realización, la célula es una célula CHO-K1, una célula CHO-K1 SV, una célula DG44 CHO, una célula DUXB11 CHO, una célula DUKX CHO, una célula CHO-S, una célula knock-out CHO GS CHO FUT8. fuera de celda, una celda knock-out de CHO FUT8 GS, una CHOZN o una celda derivada de CHO. La célula inactivada de CHO GS (p. ej., célula GSKO) es, por ejemplo, una célula inactivada de CHO-K1 SV GS. La celda eliminatoria de CHO FUT8 es, por ejemplo, el Potelligent® CHOK1 SV (Lonza Biologics, Inc.). Las células eucariotas también incluyen células aviares, líneas celulares o cepas celulares, tales como por ejemplo, EBx® células, EB14, EB24, EB26, EB66 o EBvl3.

Según otro aspecto específico, la célula eucariota es una célula de insecto (p. ej., Sf9, Mimic™ Sf9, Sf21, High FiveTM (BT1-TN-5B1-4), o células BT1-Ea88), una célula de alga (p. ej., del género Amphora, Bacillariophyceae, Dunaliella, Chlorella, Chlamydomonas, Cyanophyta (cianobacteria), Nannochloropsis, Spirulina, u Ochromonas), o una célula vegetal (*p. ej.*, células de plantas monocotiledóneas (p. ej., maíz, arroz, trigo o setaria), o de plantas dicotiledóneas (p. ej., mandioca, patata, soja, tomate, tabaco, alfalfa, *Physcomitrella patens* o Arabidopsis).

Según un aspecto específico, la célula huésped es una célula procariótica, p. ej., una célula bacteriana. Específicamente, la célula huésped es una célula Gram-positiva como Bacillus, Streptomyces Streptococcus, Staphylococcus o Lactobacillus. El bacilo que se puede utilizar es, p. ej., *B.subtilis, B.amyloliquefaciens, B.licheniformis, B.natto o B.megaterium.* En realizaciones, la celda es *B.subtilis, tal* como *B.subtilis 3NA y B.subtilis* 168. Bacillus se puede obtener, p. ej., en Bacillus Genetic Stock Center, Biological Sciences 556, 484 West th Avenida, Colón OH 43210-1214.

En una realización, la célula procariótica es una célula Gram negativa, tal como Salmonella spp. O Escherichia *coli,* tal como por ejemplo, TG1, TG2, W3110, DH1, DHB4, DH5a, HMS 174, HMS174 (DE3), NM533, C600, HB101, JM109, MC4100, XL1-Blue y Origami, así como los derivados de cepas de E.coli B-, tal como por ejemplo BL-21 o BL21 (DE3), todas las cuales están disponibles comercialmente.

Según una realización específica, la célula procariota se selecciona del grupo que consiste en *E. coli, B. subtilis,* y Pseudomonas.

Las células huéspedes adecuadas están disponibles comercialmente, por ejemplo, de colecciones de cultivo como DSMZ (Deutsche Sammlung von Mikroorganismen and Zellkulturen GmbH, Braunschweig, Alemania) o American Type Culture Collection (ATCC).

Según un aspecto específico, la invención proporciona un método para aumentar el rendimiento de una proteína de interés (PDI) producida por una célula huésped que expresa un gen de interés (GDI) que codifica dicho PDI bajo el control de un promotor que no es inducible por metanol o reprimible por una fuente de carbono distinta del metanol (en particular, la ECP descrita en el presente documento), mediante la reducción en dicha célula huésped de la expresión de un gen que codifica una proteína FLO8 que comprende la secuencia de aminoácidos identificada como SEQ ID NO: 1 o un homólogo de la misma, en particular el gen que codifica dicha proteína FLO8, que es endógena a la célula huésped.

Específicamente, el rendimiento aumenta en al menos cualquiera de 1,2 veces, 1,3 veces, 1,4 veces, 1,5 veces, 1,6 veces, 1,7 veces, 1,8 veces, 1,9 veces, 2,0 veces, 2,1 veces, 2,2 veces, 2,3 veces, 2,4 veces, 2,5 veces, 2,6 veces, 2,7 veces, 2,8 veces, 2,9 veces, 3 veces, 3,5 veces, 4 veces, 5 veces, 5,5 veces, 6 veces, 6,5 veces, 7 veces, 7,5 veces, 8 veces, 8,5 veces, 9 veces, 9,5 veces, 10 veces, 10,5 veces, 11 veces, 11,5 veces o 12 veces, en comparación con la célula huésped comparable que expresa dicho GDI,

a) que no está diseñado para reducir la expresión de dicho gen que codifica la proteína FLO8 endógena, o en el que la expresión de dicho gen que codifica la proteína FLO8 endógena no está modificada; y opcionalmente

b) en donde el promotor que controla la expresión de dicho GDI es un promotor constitutivo, en particular un promotor GAP, o un promotor inducible por metanol, en particular un promotor AOX1.

Específicamente, el método para aumentar el rendimiento de la producción de PDI descrito en este documento emplea una célula huésped recombinante como se describe más adelante en el presente documento.

Según otro aspecto específico, la invención proporciona un método para producir una proteína de interés (PDI) codificada por un gen de interés (GDI) mediante el cultivo de la célula huésped recombinante como se describe más adelante en el presente documento en condiciones para producir dicho PDI.

Según una realización específica adicional, la invención proporciona el uso de la célula huésped descrita en este documento para la producción de una PDI.

Específicamente, la célula huésped es una línea celular cultivada en un cultivo celular, en particular una línea celular huésped de producción.

Según una realización específica, la línea celular se cultiva bajo condiciones de cultivo discontinuo, alimentación discontinua o continuo. El cultivo se puede realizar en placas de microtitulación, matraces de agitación o un biorreactor y, opcionalmente, comenzar con una fase discontinua como primera etapa, seguida de una fase discontinua alimentada o una fase de cultivo continuo como segunda etapa.

Específicamente, el método comprende las etapas:

a) cultivar la célula huésped en condiciones de crecimiento; y una etapa más

b) cultivar la célula huésped en condiciones limitantes del crecimiento en presencia de hasta 1 g/L de una segunda fuente de carbono distinta del metanol, lo que da como resultado la expresión de dicho GDI para producir dicho PDI.

Específicamente, la segunda etapa b) sigue a la primera etapa a).

Específicamente, la primera fuente de carbono es una fuente de carbono distinta del metanol denominada en el presente documento como fuente de carbono basal.

Específicamente, la célula huésped se cultiva en la primera etapa en condiciones de crecimiento en un medio de cultivo celular que comprende la primera fuente de carbono, *p. ej.* en una cantidad suficiente para permitir el crecimiento de la célula huésped en el cultivo celular, opcionalmente hasta que se consuma la cantidad de la fuente de carbono, y el cultivo adicional puede realizarse en condiciones limitantes del crecimiento.

Específicamente, la segunda fuente de carbono es una fuente de carbono distinta del metanol denominada en el presente documento como fuente de carbono suplementaria.

Específicamente, dicha primera y/o segunda fuente de carbono se selecciona de sacáridos, polioles, alcoholes o mezclas de cualquiera o más de los anteriores, como se describe más adelante en el presente documento.

Según una realización específica, la fuente de carbono basal es diferente de la fuente de carbono suplementario, *p. ej.* cuantitativa y/o cualitativamente diferentes. La diferencia cuantitativa proporciona típicamente las diferentes condiciones para reprimir o desreprimir la actividad del promotor.

De acuerdo con una realización específica adicional, las fuentes de carbono basal y suplementaria comprenden el mismo tipo de moléculas o carbohidratos, preferiblemente en diferentes concentraciones. Según otra realización específica, la fuente de carbono es una mezcla de dos o más fuentes de carbono diferentes.

Se puede usar cualquier tipo de fuente de carbono orgánico, en particular las que se usan típicamente para el cultivo de células huésped, en particular para el cultivo de células huésped eucarióticas. Según una realización específica, la fuente de carbono es una hexosa, tal como glucosa, fructosa, galactosa o manosa, un disacárido, tal como sacarosa, un alcohol, como glicerol o etanol, o una mezcla de los mismos.

Según una realización específicamente preferida, la fuente de carbono basal se selecciona del grupo que consiste en glucosa, glicerol, etanol o mezclas de los mismos. Según una realización preferida, la fuente de carbono basal es el glicerol.

Según otra realización específica, la fuente de carbono suplementaria es una hexosa tal como glucosa, fructosa, galactosa y manosa, un disacárido como sacarosa, un alcohol tal como glicerol o etanol, o una mezcla de los mismos. Según una realización preferida, la fuente de carbono suplementaria es la glucosa.

Específicamente,

a) la fuente de carbono basal se selecciona del grupo que consiste en glucosa, glicerol, etanol, una mezcla de los mismos; y

b) la fuente suplementaria de carbono es una hexosa tal como glucosa, fructosa, galactosa o manosa, un disacárido tal como sacarosa, un alcohol tal como glicerol o etanol, o una mezcla de cualquiera de los anteriores.

Ambas etapas de cultivo comprenden específicamente cultivar la línea celular en presencia de dichas fuentes de carbono. Por ejemplo, dicho cultivo de la célula huésped en condiciones de crecimiento (etapa a) se lleva a cabo utilizando una fuente de carbono basal; y dicho cultivo de la célula huésped en condiciones limitantes del crecimiento (etapa b) se lleva a cabo utilizando una fuente de carbono suplementaria, p. ej. en una cantidad limitada tal que el medio de cultivo celular comprenda hasta 1 g/L o incluso una cantidad no detectable de la fuente de carbono suplementaria en el medio de cultivo celular o sobrenadante durante el cultivo (etapa b).

Las condiciones de desrepresión (o inducción) pueden lograrse adecuadamente por medios específicos. La segunda etapa b) opcionalmente emplea un medio de alimentación que proporciona una fuente de carbono suplementaria o nula en una cantidad limitada en el medio de cultivo celular o sobrenadante. Específicamente, el medio de alimentación es químicamente definido y libre de metanol.

Específicamente, la segunda etapa b) emplea un medio de alimentación que proporciona la fuente de carbono suplementaria en una cantidad limitante de crecimiento para mantener la velocidad de crecimiento específica dentro del intervalo de 0,0001 h^{-1} a 0,2 $horas^{-1}$, preferiblemente 0,005 h^{-1} a 0,15 $horas^{-1}$.

El medio de alimentación se puede añadir al medio de cultivo en forma líquida o también en una forma alternativa, tal como un sólido, p. ej., como un comprimido u otro medio de liberación sostenida, o un gas. Sin embargo, según una realización preferida, la cantidad limitada de una fuente de carbono suplementaria añadida al medio de cultivo celular puede incluso ser cero. Preferiblemente, en condiciones de un sustrato de carbono limitado, la concentración detectable de una fuente de carbono suplementaria en el medio de cultivo es de 0-1 g/L, preferiblemente menor que cualquiera de 0,9, 0,8, 0,7, 0,6, 0,5, 0,4, 0,3, 0,2, 0,1 g/L, preferiblemente menos de cualquiera de 90, 80, 70, 60, 50, 40, 30, 20 o 10 mg/L, o incluso menos de 9, 8, 7, 6, 5, 4, 3, 2 o 1 mg/L, o específicamente 1-50 mg/L, o 1-10 mg/L, específicamente preferido 1 mg/L o incluso menos, tal como por debajo del límite de detección medido con un estándar adecuado ensayo, p. ej., determinado como una concentración residual en el medio de cultivo tras el consumo por el cultivo celular en crecimiento.

En un método preferido, la cantidad limitada de la fuente suplementaria proporciona una cantidad residual en el cultivo celular que está por debajo del límite de detección determinado en el caldo de fermentación al final de una fase de producción o en el resultado de un proceso de fermentación, preferiblemente al recolectar el producto de la fermentación.

Específicamente, dicha etapa a) cultivo se realiza en una fase discontinua; y dicha etapa b) el cultivo se realiza en una fase de cultivo de alimentación discontinua o continuo.

Específicamente, las células huésped se cultivan en un medio rico en fuentes de carbono que comprende una fuente de carbono basal durante la fase de alta velocidad de crecimiento (en condiciones de crecimiento), etapa a) (p. ej., al menos el 50%, o al menos el 60%, al menos el 70%, al menos el 80%, al menos el 90%, al menos el 95%, al menos el 98%, al menos el 99%, o hasta la velocidad de crecimiento máxima) y produciendo el PDI durante una fase de baja velocidad de crecimiento (bajo crecimiento- condiciones limitantes), etapa b) (p. ej., menos del 90%, preferiblemente menos del 80%, menos del 70%, menos del 60%, menos del 50% o menos del 40%, menos del 30%, menos del 20%, menos del 10%, menos del 5%, menos del 3%, menos del 2%, menos del 1%, menos del 0,5%, menos del 0,4%, menos del 0,3% o menos del 0,2% del velocidad máxima de crecimiento velocidad) mientras se limita la fuente de carbono, en particular mediante la alimentación de un medio mínimo definido que comprende solo la cantidad de fuente de carbono que se consume completamente cuando se mantiene el cultivo celular en la fase de producción.

Específicamente, la PDI se expresa bajo dichas condiciones limitantes del crecimiento, *p. ej.* cultivando la línea celular a una velocidad de crecimiento inferior a la velocidad máxima de crecimiento, normalmente menos del 90%, preferiblemente menos del 80%, menos del 70%, menos del 60%, menos del 50%, menos del 40%, menos del 30%, inferior al 20%, inferior al 10%, inferior al 5%, inferior al 3%, inferior al 2%, inferior al 1%, inferior al 0,5%, inferior al 0,4%, inferior al 0,3%, o menos del 0,2% de la velocidad máxima de crecimiento de las células. Normalmente, la velocidad máxima de crecimiento se determina individualmente para cada tipo de célula huésped.

Específicamente, la fase discontinua se realiza hasta que la línea celular consume una fuente de carbono basal que se añade inicialmente al cultivo celular. El método de picos de oxígeno disuelto (OD) se puede utilizar para determinar el consumo de fuente de carbono basal durante la fase discontinua.

De acuerdo con una realización específica, la fase discontinua se caracteriza por una disminución continua en la señal de presión parcial de oxígeno (pO2) y en donde el final de la fase discontinua se caracteriza por un aumento de pO2. Por lo general, mientras se consume la fuente de carbono basal durante la fase discontinua y sin añadir más fuentes de carbono como es típico en las fases discontinuas, la señal de presión parcial de oxígeno (pO2) disminuirá continuamente hasta, por ejemplo, por debajo del 65%, como por ejemplo, el 30%. Después del consumo de la fuente de carbono basal, la pO2 puede aumentar, p. ej., por encima del 30%, como por ejemplo por encima del 65%, o más, lo que indica el punto de tiempo apropiado para cambiar al sistema de alimentación discontinua utilizando medio de alimentación para añadir más fuente de carbono en condiciones limitadas de fuente de carbono.

Específicamente, la pO2 se reduce a menos del 65% o menos de saturación durante la fase discontinua seguido de un aumento de más del 65% o más de saturación al final del lote. Específicamente, la fase discontinua se realiza hasta que se produce un aumento de la señal de presión parcial de oxígeno (pO2) por encima del 65% de saturación, concretamente por encima del 70%, 75%, 80% u 85%.

Específicamente, la fase discontinua se realiza durante aproximadamente 10 a 36 h.

El término "aproximadamente" con respecto al tiempo de cultivo significará +/-5% o +/-10%.

Por ejemplo, el tiempo de ejecución del lote específico de aproximadamente de 10 a 36 h puede ser de 18 a 39,6 h, específicamente de 19 a 37,8 h.

Según una realización específica, la fase discontinua se realiza utilizando de 40 a 50 g/L de glicerol, concretamente 45 g/L de glicerol como fuente de carbono basal en medios discontinuos, y el cultivo se realiza a 25 ºC durante unas 27 a 30 h, o a 30°C durante unas 23 a 36h, o a cualquier temperatura entre 25°C y 30°C durante un tiempo de cultivo de 23 a 36h. Reducir la concentración de glicerol en el medio discontinuo disminuiría la duración de la fase discontinua, mientras que aumentar el glicerol en el medio discontinuo incluso prolongaría la fase discontinua. Como alternativa al glicerol, se puede usar glucosa, p. ej., en aproximadamente las mismas cantidades.

En un sistema típico de cultivo celular y expresión de PDI, en donde una fase discontinua es seguida por una fase de alimentación discontinua, específicamente, el cultivo en la fase de alimentación discontinua se realiza durante aproximadamente 15 a 80 h, aproximadamente 15 a 70 h, aproximadamente 15 a 60h, aproximadamente 15 a 50h, aproximadamente 15 a 45h, aproximadamente 15 a 40h, aproximadamente 15 a 35h, aproximadamente 15 a 30h, aproximadamente 15 a 35h, aproximadamente 15 a 25h, o aproximadamente 15 a 20h; preferiblemente aproximadamente 20 a 40h. Específicamente, el cultivo en fase de alimentación discontinua se realiza durante unas 80h, unas 70h, unas 60h, unas 55h, unas 50h, unas 45h, unas 40h, unas 35h, unas 33h, unas 30h, unas 25h, unas 20h, o aproximadamente las 15h.

Cualquier cultivo de alimentación discontinua de menos de 120 h o menos de 100 h o hasta 80 h, que da como resultado una producción exitosa de PDI obteniendo así un alto rendimiento se denomina en el presente documento "fermentación rápida". Específicamente, la velocidad de formación de producto específica de volumen (rP) es la cantidad de producto (mg) formado por unidad de volumen (L) y unidad de tiempo (h) (mg (L h)$^{-1}$). La velocidad de formación de productos de volumen específico también se denomina rendimiento de espacio-tiempo (STY) o productividad volumétrica.

Específicamente, el cultivo de alimentación discontinua del método descrito en el presente documento se realiza de tal manera que se obtiene un rendimiento de espacio-tiempo de aproximadamente 30 mg (L h)$^{-1}$ (es decir, 30 mg (L·h)$^{-1}$ +/-5% o +/-10%). Específicamente un rendimiento espacio-tiempo de aproximadamente 30 mg (L h)$^{-1}$ se logra dentro de aproximadamente 30 h de alimentación discontinua, específicamente al menos cualquiera de 27, 28, 29, 30, 31, 32 o 33 mg (L h)$^{-1}$ en menos de cualquiera de 33h, 32h, 31h, 30h, 29h, 28h, 27h, 26h o 25h se puede lograr el tiempo de alimentación discontinua.

Específicamente, la fase discontinua se realiza como una primera etapa a), y la fase de alimentación discontinua se realiza como una segunda etapa b).

Específicamente, la segunda paso b) emplea un medio de alimentación en una fase de alimentación discontinua que proporciona una fuente de carbono suplementaria en una cantidad limitante de crecimiento para mantener la velocidad de crecimiento específica dentro del intervalo de 0,0001 h^{-1} a 0,2 horas^{-1}, preferiblemente menos que cualquiera de 0,2, 0,15, 0,1 h^{-1} o 0,15 horas^{-1}.

Específicamente, el método de cultivo que incluye etapas de cultivo cdisontinuo y de alimentación discontinua, puede emplear particularmente una célula huésped de levadura, p. ej., una levadura de cualquiera de los géneros Saccharomyces o género Pichia o *Komagataella* género, o levadura de un género diferente a Pichia, talcomo *K. lactis, Z. rouxii, P. stipitis, H. polymorpha,* o *Y. lipolítica,* preferiblemente *Pichia pastoris* o *Komagataella pastoris.*

Según otro aspecto específico, la invención proporciona un método para producir una proteína de interés (PDI) en una célula huésped, que comprende las etapas:

a) modificar genéticamente la célula huésped para reducir la expresión de un gen endógeno que codifica una proteína FLO8 que comprende la secuencia de aminoácidos identificada como SEQ ID NO: 1 o un homólogo de la misma;

b) introducir en la célula huésped un casete de expresión heterólogo que comprende un gen de interés (GDI) que codifica o expresa dicho PDI bajo el control de un promotor de casete de expresión (ECP) que está unido operativamente al GDI, cuyo ECP es regulable o reprimible por una fuente de carbono distinta del metanol;

c) cultivar dicha célula huésped en condiciones para producir dicho PDI;

d) opcionalmente aislar dicho PDI del cultivo celular; y

e) opcionalmente purificar dicho PDI.

Específicamente, la etapa a) del método descrito en el presente documento se lleva a cabo antes, después o concomitantemente con la etapa b).

Según un aspecto específico, la célula huésped primero se modifica genéticamente para reducir la expresión de dicha proteína FLO8 o el respectivo homólogo de la misma antes de ser manipulada para producir el PDI. Según un ejemplo específico, una célula huésped de tipo salvaje se modifica genéticamente de acuerdo con la etapa a) del método

descrito en el presente documento. Específicamente, la célula huésped se proporciona al introducir dichas una o más modificaciones genéticas en una cepa de células huésped de tipo salvaje para la reducción de dicha proteína FLO8 o el respectivo homólogo de la misma.

Según otro aspecto, la célula huésped se modifica primero por ingeniería genética para producir el PDI heterólogo o recombinante, antes de modificarse genéticamente adicionalmente para reducir dicha proteína FLO8 o el respectivo homólogo de la misma. Según un ejemplo específico, una célula huésped de tipo salvaje puede diseñarse primero para que comprenda el casete de expresión para la producción de PDI. Tal célula huésped manipulada puede después modificarse adicionalmente para reducir dicha proteína FLO8 o el respectivo homólogo de la misma como se describe en el presente documento.

Según otro aspecto, la célula huésped se somete tanto a la ingeniería para la producción de PDI como a la modificación genética para la reducción de dicha proteína FLO8 o el respectivo homólogo de la misma en una etapa del método, por ejemplo, empleando el respectivo casete de expresión, reactivos y herramientas en una sola etapa. o más mezclas de reacción.

Específicamente, el método emplea etapas del método para producir la célula huésped recombinante como se describe más adelante en el presente documento.

Específicamente, el casete de expresión heteróloga comprende la ECP como se describe adicionalmente en el presente documento.

Específicamente, el PDI se puede producir cultivando la célula huésped en un medio apropiado, aislando el PDI expresado del cultivo celular, en particular del sobrenadante o medio del cultivo celular al separar las células, y purificándolo mediante un método apropiado para el producto expresado, en particular tras la separación del PDI de la célula y la purificación por medios adecuados. De ese modo, se puede producir una preparación de PDI purificada.

Figuras

Figura 1: Secuencias referidas en el presente documento

Descripción detallada del invento

Los términos específicos que se utilizan a lo largo de la memoria descriptiva tienen el siguiente significado.

El término "fuente de carbono", también denominado "sustrato de carbono", tal como se usa en el presente documento, significará un sustrato de carbono fermentable, típicamente un carbohidrato fuente, adecuado como fuente de energía para microorganismos, tal como los que son capaces de ser metabolizados por organismos huéspedes o líneas celulares de producción, en particular fuentes seleccionadas del grupo que consiste en monosacáridos, oligosacáridos, polisacáridos, alcoholes, incluido el glicerol, en forma purificada, en medios mínimos o proporcionados en materias primas, como un material nutritivo complejo. La fuente de carbono puede usarse como se describe en el presente documento como una fuente de carbono única o como una mezcla de diferentes fuentes de carbono.

Una fuente de carbono distinta del metanol se entiende aquí como una cantidad de una fuente de carbono que es libre de metanol, en particular una fuente de carbono libre de metanol.

Una "fuente de carbono basal" tal como se usa como se describe en el presente documento es típicamente una fuente de carbono adecuada para el crecimiento celular, tal como un nutriente para las células huésped, en particular para las células eucariotas. La fuente de carbono basal puede proporcionarse en un medio, tal como un medio basal o un medio complejo, pero también en un medio químicamente definido que contiene una fuente de carbono purificada. La fuente de carbono basal generalmente se proporciona en una cantidad que permita el crecimiento celular, en particular durante la fase de crecimiento en un proceso de cultivo, por ejemplo, para obtener densidades celulares de al menos 5 g/L de masa seca celular, preferiblemente al menos 10 g/L de masa seca de células, o al menos 15 g/L de masa seca de células, p. ej., presentando viabilidades de más del 90% durante los pasos de subcultivo estándar, preferiblemente más del 95%.

La fuente de carbono basal se utiliza típicamente en exceso o cantidad excedente, lo que se entiende como un exceso que proporciona energía para aumentar la biomasa, p. ej., durante el cultivo de una línea celular con una velocidad de crecimiento específica alta, tal como durante la fase de crecimiento de una línea celular en un proceso de cultivo discontinuo o de alimentación discontinua. Esta cantidad excedente es particularmente superior a la cantidad limitada de una fuente de carbono suplementaria (como se usa bajo condiciones de crecimiento limitado) para lograr una concentración residual en el caldo de fermentación que sea medible y típicamente al menos 10 veces mayor, preferiblemente al menos 50 veces o al menos 100 veces mayor que durante la alimentación con la cantidad limitada de la fuente de carbono suplementaria.

Una "fuente de carbono suplementaria" tal como se describe en el presente documento es típicamente un sustrato suplementario que facilita la producción de productos de fermentación por líneas celulares de producción, en particular en la fase de producción de un proceso de cultivo. La fase de producción sigue específicamente a una fase de

crecimiento, p. ej., en procesos de cultivo discontinuo, de alimentación discontinua y continuo. La fuente de carbono suplementaria puede estar contenida específicamente en la alimentación de un proceso de alimentación discontinua. La fuente de carbono suplementaria se emplea normalmente en un cultivo celular en condiciones limitadas de sustrato de carbono, es decir, usando la fuente de carbono en una cantidad limitada.

En el presente documento, se entiende que una "cantidad limitada" de una fuente de carbono o una "fuente de carbono limitada" se refiere específicamente al tipo y la cantidad de un sustrato de carbono que facilita la producción de productos de fermentación por líneas celulares de producción, en particular en un proceso de cultivo con control. velocidades de crecimiento inferiores a la velocidad máxima de crecimiento. La fase de producción sigue específicamente a una fase de crecimiento, p. ej., en procesos de cultivo discontinuo, de alimentación discontinua y continuo. Los procesos de cultivo celular pueden emplear cultivo discontinuo, cultivo continuo y cultivo de alimentación discontinua. El cultivo discontinuo es un proceso de cultivo mediante el cual se añade una pequeña cantidad de una solución de cultivo de semillas a un medio y las células se cultivan sin añadir un medio adicional ni descargar una solución de cultivo durante el cultivo. El cultivo continuo es un proceso de cultivo mediante el cual se añade y descarga continuamente un medio durante el cultivo. El cultivo continuo también incluye el cultivo de perfusión. El cultivo de alimentación discontinua, que es un intermedio entre el cultivo discontinuo y el cultivo continuo y también conocido como cultivo semicontinuo, es un proceso de cultivo mediante el cual se añade un medio de forma continua o secuencial durante el cultivo pero, a diferencia del cultivo continuo, la solución de cultivo no se descarga continuamente.

Específicamente preferido es un proceso de alimentación discontinua que se basa en la alimentación de un sustrato nutritivo limitante del crecimiento a un cultivo. La estrategia de alimentación discontinua, que incluye fermentación de alimentación discontinua única o alimentación discontinua repetida, se usa típicamente en procesos bioindustriales para alcanzar una alta densidad celular en el biorreactor. La adición controlada del sustrato de carbono afecta directamente la velocidad de crecimiento del cultivo y ayuda a evitar el metabolismo de desbordamiento o la formación de subproductos metabólicos no deseados. En condiciones limitadas de fuente de carbono, la fuente de carbono puede estar contenida específicamente en la alimentación de un proceso de alimentación discontinua. De este modo, el sustrato de carbono se proporciona en una cantidad limitada.

También en quimiostato o cultivo continuo como se describe en el presente documento, la velocidad de crecimiento puede controlarse estrictamente.

La cantidad limitada de una fuente de carbono se entiende en el presente documento particularmente como la cantidad de una fuente de carbono necesaria para mantener una línea celular de producción en condiciones de crecimiento limitado, p. ej., en una fase de producción o modo de producción. Tal cantidad limitada puede emplearse en un proceso de alimentación discontinua, en donde la fuente de carbono está contenida en un medio de alimentación y se suministra al cultivo a velocidades de alimentación bajas para un suministro sostenido de energía, p. ej., para producir una PDI, mientras se mantiene la biomasa a bajas velocidades específicas de crecimiento. Típicamente, se añade un medio de alimentación a un caldo de fermentación durante la fase de producción de un cultivo celular.

La cantidad limitada de una fuente de carbono puede, por ejemplo, determinarse por la cantidad residual de la fuente de carbono en el caldo de cultivo celular, que está por debajo de un umbral predeterminado o incluso por debajo del límite de detección medido en un ensayo estándar (carbohidratos). La cantidad residual normalmente se determinaría en el caldo de fermentación al recolectar un producto de fermentación.

La cantidad limitada de una fuente de carbono también se puede determinar definiendo la velocidad de alimentación promedio de la fuente de carbono al fermentador, p. ej., según lo determinado por la cantidad añadida durante todo el proceso de cultivo, p. ej., la fase de alimentación discontinua, por tiempo de cultivo, para determinar una cantidad media calculada por tiempo. Esta velocidad de alimentación promedio se mantiene baja para garantizar el uso completo de la fuente de carbono suplementaria por parte del cultivo celular, p. ej., entre 0,6 g L^{-1} h^{-1} (g fuente de carbono por L volumen de fermentación inicial y h tiempo) y 25 g L^{-1} h^{-1}, preferiblemente entre 1,6 g L^{-1} h^{-1} y 20 g L^{-1} h^{-1}.

La cantidad limitada de una fuente de carbono también se puede determinar midiendo la velocidad de crecimiento específica, velocidad de crecimiento específica que se mantiene baja, p. ej., inferior a la velocidad máxima de crecimiento específico, durante la fase de producción, p. ej., dentro de un intervalo predeterminado, tal como en el intervalo de 0.001 h^{-1} a 0,20 h^{-1}, o 0,005 horas^{-1} a 0,20 h^{-1}, preferiblemente entre 0,01 h^{-1} y 0,15h^{-1}.

Específicamente, se usa un medio de alimentación que está químicamente definido y libre de metanol.

El término "definido químicamente" con respecto al medio de cultivo celular, tal como un medio mínimo o un medio de alimentación en un proceso de alimentación discontinua, significará un medio de cultivo adecuado para el cultivo celular *in vitro* de una línea celular de producción, en donde se conocen todos los componentes químicos y (poli)péptidos. Típicamente, un medio químicamente definido está completamente libre de componentes derivados de animales y representa un entorno de cultivo celular puro y consistente.

El término "célula huésped", como se usa en el presente documento, se referirá a una sola célula, un solo clon celular o una línea celular de una célula huésped.

El término "línea celular", como se usa en el presente documento, se refiere a un clon establecido de un tipo de célula particular que ha adquirido la capacidad de proliferar durante un período de tiempo prolongado. Una línea celular se usa típicamente para expresar un gen endógeno o recombinante, o productos de una vía metabólica para producir polipéptidos o metabolitos celulares mediados por dichos polipéptidos. Una "línea de células huésped de producción" o "línea celular de producción" se entiende comúnmente como una línea celular lista para usar para el cultivo celular en un biorreactor para obtener el producto de un proceso de producción, tal como una PDI.

La célula huésped que produce la PDI como se describe en el presente documento también se denomina "célula huésped de producción" y una línea celular respectiva se denomina "línea celular de producción".

Las realizaciones específicas descritas en el presente documento se refieren a una línea de células huésped de producción que está diseñada para subexpresar un gen endógeno que codifica una proteína FLO8, y/o tiene una expresión reducida de dicho gen, y se caracteriza por un alto rendimiento de producción de PDI bajo el control de un promotor regulable de fuente de carbono (tal como un ECP descrito en el presente documento), en donde el promotor puede inducirse sin necesidad de añadir metanol al cultivo celular. Tal célula huésped resultó expresar de manera estable la PDI sin cambiar significativamente la morfología.

El término "célula huésped" se aplicará en particular a cualquier célula u organismo eucariótico o procariótico, que se utilice adecuadamente con fines de recombinación para producir una PDI o un metabolito de la célula huésped. Se entiende bien que el término "célula huésped" no incluye a los seres humanos. Específicamente, las células huésped, tal como se describen en el presente documento, son organismos artificiales y derivados de células huésped nativas (de tipo salvaje). Se entiende bien que las células huésped, los métodos y los usos descritos en el presente documento, por ejemplo, en referencia específicamente a aquellos que comprenden una o más modificaciones genéticas, dichos casetes o construcciones de expresión heteróloga, dichas células huésped transfectadas o transformadas y proteínas recombinantes, no son de origen natural, "hechas por el hombre" o sintéticas, y por lo tanto no se consideran como resultado de una "ley de la naturaleza".

El término "cultivo celular" o "que cultiva" o "cultivo" como se usa en el presente documento con respecto a una célula huésped se refiere al mantenimiento de las células en un medio artificial, *p. ej.*, un medio ambiente *in vitro*, en condiciones que favorezcan el crecimiento, la diferenciación o la viabilidad continuada, en estado activo o quiescente, de las células, específicamente en un biorreactor controlado según métodos conocidos en la industria.

Cuando se cultiva un cultivo celular usando medios de cultivo apropiados, las células se ponen en contacto con los medios en un recipiente de cultivo o con un sustrato en condiciones adecuadas para apoyar el cultivo de células en el cultivo celular. Como se describe en el presente documento, se proporciona un medio de cultivo que se puede utilizar para el crecimiento de células huésped, p. ej., células eucariotas, específicamente levaduras u hongos filamentosos. Las técnicas estándar de cultivo celular son bien conocidas en la técnica.

Los cultivos celulares como se describen en el presente documento emplean particularmente técnicas que proporcionan la producción de una PDI secretado, como para obtener el PDI en el medio de cultivo celular, que es separable de la biomasa celular, denominado en el presente documento "sobrenadante de cultivo celular", y puede purificarse para obtener el PDI con un mayor grado de pureza. Cuando una proteína (tal como, p. ej., una PDI) es producida y secretada por la célula huésped en un cultivo celular, se entiende en el presente documento que dichas proteínas son secretadas en el sobrenadante del cultivo celular, y pueden obtenerse separando el sobrenadante del cultivo celular de la biomasa de la célula huésped y, opcionalmente, purificar adicionalmente la proteína para producir una preparación de proteína purificada.

Los medios de cultivo celular proporcionan los nutrientes necesarios para mantener y hacer crecer las células de forma controlada, artificial y medioambiente *in vitro*. Las características y composiciones de los medios de cultivo celular varían según los requisitos celulares particulares. Los parámetros importantes incluyen la osmolalidad, el pH y las formulaciones de nutrientes. La alimentación de nutrientes se puede realizar de forma continua o discontinua según métodos conocidos en la técnica.

Mientras que un proceso discontinuo es un modo de cultivo celular en el que todos los nutrientes necesarios para cultivar las células están contenidos en el medio de cultivo inicial, sin suministro adicional de nutrientes adicionales durante la fermentación, en un proceso de alimentación discontinua, después de una fase discontinua, una fase de alimentación tiene lugar en la que se suministran uno o más nutrientes al cultivo mediante la alimentación. Aunque en la mayoría de los procesos el modo de alimentación es crítico e importante, la célula huésped y los métodos descritos en el presente documento no están restringidos con respecto a un determinado modo de cultivo celular.

Se puede producir una PDI recombinante usando la célula huésped y la línea celular respectiva descrita en el presente documento, cultivándola en un medio apropiado, aislando el producto o metabolito expresado del cultivo y, opcionalmente, purificándolo mediante un método adecuado.

Se prefieren varios enfoques diferentes para la producción de la PDI como se describe en el presente documento. Una PDI puede expresarse, procesarse y, opcionalmente, secretarse transfectando o transformando una célula huésped con un vector de expresión que contiene ADN recombinante que codifica la proteína relevante, preparando un cultivo

de la célula transfectada o transformada, haciendo crecer el cultivo, induciendo la transcripción y la producción de PDI, y recuperar la PDI.

En ciertas realizaciones, el proceso de cultivo celular es un proceso de alimentación discontinua. Específicamente, una célula huésped transformada con una construcción de ácido nucleico que codifica una PDI recombinante deseado, se cultiva en una fase de crecimiento y pasa a una fase de producción para producir una PDI recombinante deseada.

En otra realización, las células huésped descritas en el presente documento se cultivan en un modo continuo, por ejemplo, empleando un quimiostato. Un proceso de fermentación continua se caracteriza por una velocidaddefinida, constante y continua de alimentación de medio de cultivo fresco en un biorreactor, por lo que el caldo de cultivo se elimina al mismo tiempo del biorreactor a la misma velocidadveloidad de eliminación definida, constante y continua. Al mantener el medio de cultivo, la velocidad de alimentación y la velocidad de eliminación en el mismo nivel constante, los parámetros y las condiciones del cultivo celular en el biorreactor permanecen constantes.

Se entiende específicamente que un cultivo celular estable como se describe en el presente documento se refiere a un cultivo celular que mantiene las propiedades genéticas, específicamente manteniendo alto el nivel de producción de PDI, p. ej., al menos a un nivel de µg, incluso después de aproximadamente 20 generaciones de cultivo, preferiblemente al menos 30 generaciones, más preferiblemente al menos 40 generaciones, lo más preferiblemente de al menos 50 generaciones. Específicamente, se proporciona una línea celular huésped recombinante estable que se considera una gran ventaja cuando se utiliza para la producción a escala industrial.

El cultivo celular descrito en el presente documento es particularmente ventajoso para métodos a escala de fabricación industrial, p. ej., con respecto tanto al volumen como al sistema técnico, en combinación con un modo de cultivo que se basa en la alimentación de nutrientes, en particular un proceso discontinuo o de alimentación discontinua, o un proceso continuo o semicontinuo (p. ej., quimiostato).

La célula huésped descrita en el presente documento normalmente se analiza para determinar su capacidad para expresar el GDI para la producción de PDI, se analiza el rendimiento de PDI mediante cualquiera de las siguientes pruebas: ELISA, ensayo de actividad, HPLC u otras pruebas adecuadas, tales como SDS-PAGE y técnicas de transferencia de Western o espectrometría de masas.

Para determinar el efecto de una modificación genética sobre la subexpresión o reducción del gen que codifica la proteína FLO8 o su homólogo en el cultivo celular respectivo y, p. ej., sobre su efecto sobre la producción de PDI, la línea de células huésped puede cultivarse en placas de microtitulación, matraz agitador, o biorreactor usando fermentaciones de alimentación discontinua o quimiostatos en comparación con cepas sin dicha modificación genética en la célula respectiva.

El método de producción descrito en el presente documento permite específicamente la fermentación a escala piloto o industrial. La báscula de proceso industrial emplearía preferiblemente volúmenes de al menos 10 L, específicamente al menos 50 L, preferiblemente al menos 1 m^3, preferiblemente al menos 10 m^3, lo más preferiblemente al menos 100 m^3.

Se prefieren las condiciones de producción a escala industrial, que se refieren, p. ej., al cultivo de alimentación discontinua en volúmenes de reactor de 100 L a 10 m^3 o mayor, empleando tiempos de proceso típicos de varios días, o procesos continuos en volúmenes de fermentador de aproximadamente 50 - 1000 L o mayores, con velocidades de dilución de aproximadamente 0,02 - 0,15 h^{-1}.

Los dispositivos, instalaciones y métodos utilizados para el propósito descrito en el presente documento son específicamente adecuados para su uso en y con el cultivo de cualquier línea celular deseada, incluyendo las líneas celulares procarióticas y/o eucarióticas. Además, en las realizaciones, los dispositivos, las instalaciones y los métodos son adecuados para cultivar cualquier tipo de célula, incluyendo las células en suspensión o las células dependientes del anclaje (adherentes), y son adecuados para operaciones de producción configuradas para la producción de productos farmacéuticos y biofarmacéuticos, tales como productos polipeptídicos (PDI), productos de ácido nucleico (por ejemplo, ADN o ARN), o células y/o virus como los que se utilizan en terapias celulares y/o virales.

En ciertas realizaciones, las células expresan o producen un producto, tales como un producto terapéutico recombinante o de diagnóstico. Como se describe con más detalle en este documento, los ejemplos de productos producidos por las células incluyen, pero no se limitan a, PDI como los ejemplificados en este documento que incluyen moléculas de anticuerpos (p. ej., anticuerpos monoclonales, anticuerpos biespecíficos), miméticos de anticuerpos (moléculas polipeptídicas que se unen específicamente a antígenos pero que no están relacionados estructuralmente con anticuerpos, tales como p. ej., DARPins, afficuerpos, adnectinas, o IgNARs), proteínas de fusión (p. ej., proteínas de fusión Fc, citoquinas quiméricas), otras proteínas recombinantes (p. ej., proteínas glicosiladas, enzimas, hormonas), o terapéuticos virales p. ej., virus oncolíticos anticancerígenos, vectores virales para terapia génica e inmunoterapia viral), terapias celulares (p. ej., células madre pluripotentes, células madre mesenquimales y células madre adultas), vacunas o partículas encapsuladas en lípidos (p. ej., exosomas, partículas similares a virus-), ARN (tal como, p.. ej., siARN) o ADN (tales como, p. ej., ADN plasmídico), antibióticos o aminoácidos. En realizaciones, los dispositivos, instalaciones y métodos pueden usarse para producir biosimilares.

Como se mencionó, en ciertas realizaciones, los dispositivos, las instalaciones y los métodos permiten la producción de células eucariotas, p. ej., células de mamífero o células eucariotas inferiores, tales como por ejemplo células de levadura o células de hongos filamentosos, o células procariotas, tales como Gram-positivas o células Gram-negativas y/o productos de las células eucarióticas o procarióticas, p. ej., PDI que incluyen proteínas, péptidos o antibióticos, aminoácidos, ácidos nucleicos (tales como ADN o ARN), sintetizados por dichas células a gran escala. A menos que se indique lo contrario en el presente documento, los dispositivos, las instalaciones y los métodos pueden incluir cualquier volumen o capacidad de producción deseados, que incluyen, peor no se limitan a, capacidades a escala de banco, a escala piloto y a escala de producción completa.

Además, y a menos que se indique lo contrario en el presente documento, los dispositivos, instalaciones y métodos pueden incluir cualquier reactor adecuado, que incluyen, peor no se limitan a, tanque agitado, puente aéreo, fibra, microfibra, fibra hueca, matriz cerámica, lecho fluidizado, lecho fijo y /o biorreactores de lecho en chorro. Como se usa en el presente documento, "reactor" puede incluir un fermentador o una unidad de fermentación, o cualquier otro recipiente de reacción y el término "reactor" se usa indistintamente de "fermentador". Por ejemplo, en algunos aspectos, una unidad de biorreactor de ejemplo puede realizar uno o más, o todos, de los siguientes: alimentación de nutrientes y/o fuentes de carbono, inyección de gas adecuado (p. ej., oxígeno), flujo de entrada y salida de fermentación o medio de cultivo celular, separación de fases gaseosas y líquidas, mantenimiento de temperatura, mantenimiento de niveles oxígeno y CO_2, mantenimiento del nivel de pH, agitación (p. ej., agitación) y/o limpieza/esterilización. Las unidades de reactor de ejemplo, tales como una unidad de fermentación, pueden contener múltiples reactores dentro de la unidad, por ejemplo, la unidad puede tener 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90 o 100 o más biorreactores en cada unidad y/o una instalación puede contener varias unidades que tengan uno o varios reactores dentro de la instalación. En varias realizaciones, el biorreactor puede ser adecuado para procesos de fermentación discontinua, de semialimentación discontinua, de alimentación discontinua, de perfusión y/o continuos. Puede usarse cualquier diámetro de reactor adecuado. En realizaciones, el biorreactor puede tener un volumen entre aproximadamente 100 ml y aproximadamente 50 000 l. Los ejemplos no limitantes incluyen un volumen de 100 ml, 250 ml, 500 ml, 750 ml, 1 litro, 2 litros, 3 litros, 4 litros, 5 litros, 6 litros, 7 litros, 8 litros, 9 litros, 10 litros, 15 litros, 20 litros, 25 litros, 30 litros, 40 litros, 50 litros, 60 litros, 70 litros, 80 litros, 90 litros, 100 litros, 150 litros, 200 litros, 250 litros, 300 litros, 350 litros, 400 litros, 450 litros, 500 litros, 550 litros, 600 litros, 650 litros, 700 litros, 750 litros, 800 litros, 850 litros, 900 litros, 950 litros, 1000 litros, 1500 litros, 2000 litros, 2500 litros, 3000 litros, 3500 litros, 4000 litros, 4500 litros, 5000 litros, 6000 litros, 7000 litros, 8000 litros, 9000 litros, 10,000 litros, 15,000 litros, 20,000 litros, y/o 50.000 litros. Además, los reactores adecuados pueden ser de uso múltiple, de un solo uso, desechables o no desechables y pueden estar formados por cualquier material adecuado, incluyendo aleaciones metálicas tales como acero inoxidable (p. ej., 316L o cualquier otro acero inoxidable adecuado) e Inconel, plásticos y/o vidrio.

En realizaciones y a menos que se indique lo contrario en el presente documento, los dispositivos, instalaciones y métodos descritos en el presente documento también pueden incluir cualquier operación unitaria adecuada y/o equipo no mencionado, tales como operaciones y/o equipo para separación, purificación y aislamiento de dichos productos. Se puede utilizar cualquier instalación y entorno adecuado, como instalaciones tradicionales construidas con palos, instalaciones modulares, móviles y temporales, o cualquier otra construcción, instalación y/o diseño adecuado. Por ejemplo, en algunas realizaciones se pueden utilizar salas limpias modulares. Además, y a menos que se indique lo contrario, los dispositivos, sistemas y métodos descritos en el presente documento se pueden alojar y/o realizar en una sola ubicación o instalación o alternativamente se pueden alojar y/o realizar en ubicaciones y/o instalaciones separadas o múltiples.

Las técnicas adecuadas pueden abarcar el cultivo en un biorreactor comenzando con una fase discontinua, seguida de una fase de alimentación discontinua exponencial corta a una velocidad de crecimiento específica alta, seguida además por una fase de alimentación discontinua a una velocidad de crecimiento específica baja. Otra técnica de cultivo adecuada puede abarcar una fase por lotes seguida de una fase de alimentación discontinua a cualquier velocidad de crecimiento específica adecuada o combinaciones de velocidades de crecimiento específicas, como pasar de una velocidad de crecimiento alta a una baja durante el tiempo de producción del PDI, o de una velocidad de crecimiento baja a alta durante el tiempo de producción. Tiempo de producción de puntos de interés. Otra técnica de cultivo adecuada puede abarcar una fase discontinua seguida de una fase de cultivo continuo a una velocidad de dilución baja.

Una realización preferida incluye un cultivo discontinuo para proporcionar biomasa seguido de un cultivo de alimentación discontinua para la producción de PDI de alto rendimiento.

Se prefiere cultivar una célula huésped como se describe en el presente documento 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 de al menos 10 g/L de peso seco de células, preferiblemente de al menos 20 g /L de peso seco de células, preferiblemente al menos cualquiera de 30, 40, 50, 60, 70 u 80 g/L de peso seco de células. Es ventajoso prever tales rendimientos de producción de biomasa a escala piloto o industrial.

Un medio de crecimiento que permite la acumulación de biomasa, específicamente un medio de crecimiento basal, normalmente comprende una fuente de carbono, una fuente de nitrógeno, una fuente de azufre y una fuente de fosfato. Típicamente, dicho medio comprende además oligoelementos y vitaminas, y puede comprender además aminoácidos, peptona o extracto de levadura.

Las fuentes de nitrógeno preferidas incluyen $NH_4H_2PO_4$, o NHa o $(NH_4)_2SO_4$;

Las fuentes de azufre preferidas incluyen $MgSO_4$, o $(NH_4)_2SO_4$ o K_2SO_4;

Las fuentes de fosfato preferidas incluyen $NH_4H_2PO_4$, o H_3PO_4, o NaH_2PO_4, KH_2PO_4, Na_2HPO_4 o K_2HPO_4;

Otros componentes típicos del medio incluyen KCl, $CaCl_2$, y oligoelementos como: Fe, Co, Cu, Ni, Zn, Mo, Mn, I, B;

Preferiblemente, el medio se complementa con vitamina B_7;

Un medio de crecimiento típico para *P. pastoris* comprende glicerol, sorbitol o glucosa, $NH_4H_2PO_4$, $MgSO_4$, KCl, $CaCl_2$, biotina y oligoelementos.

En la fase de producción, se usa específicamente un medio de producción con solo una cantidad limitada de una fuente de carbono suplementaria.

Preferiblemente, la línea de células huésped se cultiva en un medio mineral con una fuente de carbono adecuada, lo que simplifica aún más el proceso de aislamiento de manera significativa. Un ejemplo de un medio mineral preferido es uno que contiene una fuente de carbono utilizable (*p. ej.*, glucosa, glicerol o sorbitol), sales que contienen los macroelementos (potasio, magnesio, calcio, amonio, cloruro, sulfato, fosfato) y oligoelementos (cobre, yoduro, manganeso, molibdato, cobalto, zinc y sales de hierro, y ácido bórico), y opcionalmente vitaminas o aminoácidos, por ejemplo, para complementar auxotrofias.

Específicamente, las células se cultivan en condiciones adecuadas para efectuar la expresión de la PDI deseada, que se puede purificar a partir de las células o del medio de cultivo, según la naturaleza del sistema de expresión y la proteína expresada. *p.ej.*, si la proteína está fusionada con un péptido señal y si la proteína es soluble o está unida a la membrana. Como comprenderá el experto en la materia, las condiciones de cultivo variarán según factores que incluyen el tipo de célula huésped y el vector de expresión particular empleado.

Un medio de producción típico comprende una fuente de carbono suplementaria y más $NH_4H_2PO_4$, $MgSO_4$, KCl, $CaCl_2$, biotina y oligoelementos.

Por ejemplo, la alimentación de la fuente de carbono suplementaria añadida a la fermentación puede comprender una fuente de carbono con hasta un 50% en peso de azúcares utilizables.

La fermentación se realiza preferiblemente a un pH que varía entre 3 y 8.

Los tiempos típicos de fermentación son de aproximadamente 24 a 120 horas con temperaturas en el intervalo de 20 °C a 35 °C, preferiblemente 22-30 °C.

La PDI se expresa preferiblemente empleando condiciones para producir rendimientos de al menos 1 mg/L, preferiblemente al menos 10 mg/L, preferiblemente al menos 100 mg/L, lo más preferiblemente al menos 1 g/L

El término "expresión" o "casete de expresión", como se usa en el presente documento, se refiere a moléculas de ácido nucleico que contienen una secuencia de codificación deseada y secuencias de control en enlaces operables, de modo que los huéspedes transformados o transfectados con estas secuencias sean capaces de producir las proteínas codificadas o los metabolitos de la célula huésped. Para efectuar la transformación, el sistema de expresión puede incluirse en un vector; sin embargo, el ADN relevante también puede integrarse en un cromosoma de la célula huésped. La expresión puede referirse a productos de expresión secretados o no secretados, incluyendo polipéptidos o metabolitos.

Los casetes de expresión se proporcionan convenientemente como construcciones de expresión, p. ej., de genes recombinantes y la traducción de su ARNm en un organismo huésped adecuado. Los vectores de expresión o plásmidos normalmente comprender un origen para la replicación autónoma o un locus para la integración del genoma en las células huésped, 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, nourseotricina), una serie de sitios de escisión de enzimas de restricción, una secuencia promotora adecuada y un terminador de la transcripción, cuyos componentes están operativamente unidos entre sí. Los términos "plásmido" y "vector", tal como se utilizan en el presente documento, incluyen secuencias de nucleótidos que se replican de forma autónoma, así como secuencias de nucleótidos que se integran en el genoma, tal como cromosomas artificiales, p. ej., un cromosoma artificial de levadura (YAC).

Los vectores de expresión pueden incluir, pero no se limitan a, vectores de clonación, vectores de clonación modificados y plásmidos diseñados específicamente. Los vectores de expresión preferidos descritos en el presente documento son vectores de expresión adecuados para la expresión de un gen recombinante en una célula huésped eucariota y se seleccionan dependiendo del organismo huésped. Los vectores de expresión apropiados normalmente comprenden secuencias reguladoras adecuadas para expresar ADN que codifica una PDI en una célula huésped eucariota. Los ejemplos de secuencias reguladoras incluyen promotores, operadores, potenciadores, sitios de unión

al ribosoma y secuencias que controlan el inicio y la terminación de la transcripción y la traducción. Las secuencias reguladoras normalmente están operativamente unidas a la secuencia de ADN que se va a expresar.

Para permitir la expresión de una secuencia de nucleótidos recombinantes en una célula huésped, el casete o vector de expresión descrito en el presente documento comprende un ECP, típicamente una secuencia de nucleótidos promotora que es adyacente al extremo 5' de la secuencia codificante. *p. ej.,* aguas arriba y adyacente a un gen de interés (GDI), o si se usa una secuencia señal o líder, aguas arriba y adyacente a dicha secuencia señal y líder, respectivamente, para facilitar la expresión y secreción del PDI. La secuencia promotora típicamente regula e inicia la transcripción de la secuencia de nucleótidos aguas abajo, con la que está unida operativamente, incluyendo en particular el GDI.

Las construcciones de expresión específicas descritas en el presente documento comprenden un promotor enlazado operativamente a una secuencia de nucleótidos que codifica una PDI bajo el control transcripcional de dicho promotor. Específicamente, el promotor no está asociado de forma nativa con la secuencia codificante de la PDI.

Las construcciones de expresión específicas descritas en el presente documento comprenden un polinucleótido que codifica el PDI unido a una secuencia líder que provoca la secreción del PDI de la célula huésped. La presencia de una secuencia líder de secreción de este tipo en el vector de expresión normalmente se requiere cuando el PDI destinado a la expresión y secreción recombinantes es una proteína que no se secreta naturalmente y, por lo tanto, carece de una secuencia líder de secreción natural, o su secuencia de nucleótidos se ha clonado sin su secuencia líder de secreción natural. En general, puede usarse cualquier secuencia líder de secreción eficaz para provocar la secreción del PDI de la célula huésped. La secuencia líder de secreción puede proceder de una fuente de levadura, p. del factor α de levadura como MFa de *Saccharomyces cerevisiae,* o fosfatasa de levadura, de origen mamífero o vegetal, u otras.

En realizaciones específicas, pueden usarse vectores de multiclonación, que son vectores que tienen un sitio de multiclonación. Específicamente, un gen heterólogo deseado puede integrarse o incorporarse en un sitio de multiclonación para preparar un vector de expresión. En el caso de los vectores de multiclonación, típicamente se coloca un promotor aguas arriba del sitio de multiclonación.

La célula huésped recombinante descrita en el presente documento está diseñada específicamente para reducir la cantidad de proteína FLO8 endógena de la célula huésped o el homólogo u ortólogo respectivo en la célula huésped, en particular mediante la reducción de la expresión de la secuencia del gen codificante respectivo, para subexpresar el gen.

El término "expresión génica" o "expresar un polinucleótido" como se usa en el presente documento, pretende abarcar al menos una etapa seleccionada del grupo que consiste en transcripción de ADN en ARNm, procesamiento de ARNm, maduración de ARNm, exportación de ARNm, traducción, plegamiento de proteínas y /o transporte de proteínas.

El término "reducir la expresión" generalmente se refiere a "subexpresar" y generalmente se refiere a cualquier cantidad menor que un nivel de expresión presentado por un estándar de referencia, que es la célula huésped antes de la modificación genética para reducir la expresión de un determinado polinucleótido, o que de otra manera es expresada en una célula huésped del mismo tipo o especie que no está diseñada para reducir la expresión de dicho polinucleótido. La reducción de la expresión como se describe en el presente documento se refiere específicamente a un polinucleótido o gen que codifica una proteína FLO8 definida, en particular un gen que es endógeno a la célula huésped antes de la modificación genética. En particular, el producto génico respectivo es la proteína FLO8 definida como se describe en el presente documento. Tras la modificación genéticade la célula huésped mediante modificación genética para reducir la expresión de dicho gen, la expresión de dicho producto génico o polipéptido está a un nivel que es menor que la expresión del mismo producto génico o polipéptido antes de una modificación genética de la célula huésped o en un huésped comparable que no ha sido modificado genéticamente. "Menos de" incluye, p. ej., 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80, 90% o más. La expresión "reducción de la expresión" o "subexpresión" tampoco abarca la expresión del producto génico o de un polipéptido.

Según realizaciones específicas descritas en el presente documento, la célula huésped se modifica por ingeniería genética para inactivar o eliminar (para la inactivación o eliminación de un gen o una parte del mismo) el gen de la célula huésped endógena que codifica la proteína FLO8 (como se define en el presente documento, que incluye *p.ej.* el respectivo homólogo u ortólogo), u otras secuencias de nucleótidos (codificantes o no codificantes) que confieren a la célula huésped la capacidad de expresar o producir dicha proteína FLO8.

Específicamente, se proporciona una cepa de deleción, en la que se interrumpe una secuencia de nucleótidos.

El término "interrumpir" como se usa en el presente documento se refiere a la reducción significativa hasta la eliminación completa de la expresión de una o más proteínas endógenas en una célula huésped, tal como knock-down o knockout. Esto puede medirse como presencia de una o más proteínas endógenas en un cultivo celular o medio de cultivo de la célula huésped, tal como por espectrometría de masas en la que el contenido total de una proteína endógena puede ser inferior a un umbral o no detectable.

El término "interrumpido" se refiere específicamente a un resultado de la modificación genética por al menos una etapa seleccionada del grupo que consiste en silenciamiento de genes, knock-down de genes, knockout de genes, administración de una construcción negativa dominante, knockout de genes condicional y/o por alteración del gen con respecto a un gen específico.

El término "knock-down", "reducción" o "agotamiento" en el contexto de la expresión génica como se usa en el presente documento se refiere a enfoques experimentales que conducen a una expresión reducida de un gen dado en comparación con la expresión en una célula de control. La desactivación de un gen se puede lograr por varios medios experimentales, tal como la introducción de moléculas de ácido nucleico en la célula que se hibridan con partes del ARNm del gen y conducen a su degradación (*p. ej.*, ARNsh, ARNi, ARNmi) o alterando la secuencia del gen de una manera que conduce a una transcripción reducida, estabilidad reducida del ARNm o traducción disminuida del ARNm.

Una inhibición completa de la expresión de un gen dado se denomina "knockout". El knockout de un gen significa que no se sintetizan transcritos funcionales a partir de dicho gen, lo que conduce a una pérdida de la función normalmente proporcionada por este gen. El knockout del gen se logra alterando la secuencia de ADN que conduce a la interrupción o eliminación del gen o sus secuencias reguladoras, o parte de dicho gen o secuencias reguladoras. Las tecnologías de knockout incluyen el uso de técnicas de recombinación homóloga para reemplazar, interrumpir o eliminar partes cruciales o la secuencia génica completa o el uso de enzimas que modifican el ADN, tal como el dedo de zinc o las meganucleasas, para introducir roturas de doble cadena en el ADN del gen objetivo, por ejemplo, descrito por Gaj et al. (Trends Biotechnol. 2013;31(7):397-405).

Las realizaciones específicas emplean uno o más plásmidos o casetes knockout que se transforman o transfectan en las células huésped. Mediante recombinación homóloga, el gen diana en las células huésped puede romperse. Este procedimiento generalmente se repite hasta que todos los alelos del gen objetivo se eliminen de manera estable.

Un método específico para eliminar un gen específico como se describe en el presente documento son los métodos CRISPR-Cas9 como se describe en, por ejemplo, Weninger et al. (J. Biotechnol. 2016, 235:139-49). Otro método incluye el enfoque de marcador dividido como se describe, p. ej., Heiss et al. 2013 (Appl Microbiol Biotechnol. 97(3):1241-9.)

Otra realización se refiere a la degradación del ARNm objetivo mediante el uso de ARN de interferencia pequeño (ARNip) para transfectar la célula huésped y dirigirse a un ARNm que codifica la proteína objetivo expresada endógenamente por dicha célula huésped.

La expresión de un gen se puede inhibir o reducir mediante métodos que interfieren directamente con la expresión génica, que incluyen, pero no se restringe a, la inhibición o reducción de la transcripción del ADN, *p. ej.,* por el uso de represores específicos relacionados con el promotor, por mutagénesis específica del sitio de un promotor dado, por intercambio de promotores, o inhibición o reducción de la traducción, *p. ej.,* mediante silenciamiento génico postranscripcional inducido por ARNi o ARN no codificante. La expresión de un producto génico disfuncional o inactivo con actividad reducida puede lograrse, por ejemplo, mediante mutagénesis, inserciones o deleciones aleatorias o específicas del sitio dentro del gen codificante.

La inhibición o reducción de la actividad del producto génico puede lograrse, por ejemplo, mediante la administración o la incubación con un inhibidor de la enzima respectiva, antes o simultáneamente con la expresión de la proteína. Los ejemplos de tales inhibidores incluyen, pero no se limitan a, un péptido inhibidor, un anticuerpo, un aptámero, una proteína de fusión o un anticuerpo mimético contra dicha enzima, o un ligando o receptor del mismo, o un péptido inhibidor o ácido nucleico, o una molécula pequeña con actividad de unión similar.

El silenciamiento de genes, la eliminación de genes y la eliminación de genes se refieren a técnicas mediante las cuales se reduce la expresión de un gen, ya sea mediante modificación genética o mediante tratamiento con un oligonucleótido con una secuencia complementaria a una transcripción de ARNm o un gen. Si se realiza una modificación genética del ADN, el resultado es un organismo knock-down o knockout. Si el cambio en la expresión génica es causado por un oligonucleótido que se une a un ARNm o se une temporalmente a un gen, esto da como resultado un cambio temporal en la expresión génica sin modificación del ADN cromosómico y se denomina caída transitoria.

En una desactivación transitoria, que también se abarca en el término anterior, la unión de este oligonucleótido al gen activo o sus transcritos provoca una disminución de la expresión mediante el bloqueo de la transcripción (en el caso de la unión de genes), la degradación del transcrito de ARNm (p. ej., por ARN de interferencia pequeño (siARN) o ARN antisentido) o bloqueando la traducción del ARNm.

Otros enfoques para llevar a cabo el silenciamiento génico, el knock-down o el knockout son conocidos por el experto en la literatura respectiva, y su aplicación en el contexto de la presente invención se considera rutinaria. El knockout de genes se refiere a técnicas mediante las cuales la expresión de un gen se bloquea por completo, es decir, el gen respectivo no funciona o incluso se elimina. Los enfoques metodológicos para lograr este objetivo son múltiples y conocidos por el experto en la materia. Los ejemplos son la producción de un mutante que es dominantemente negativo para el gen dado. Tal mutante se puede producir mediante mutagénesis dirigida al sitio (p. ej., deleción, deleción parcial, inserción o sustitución de ácido nucleico), mediante el uso de transposones adecuados, o mediante

otros enfoques que son conocidos por el experto en la literatura respectiva, cuya aplicación en el contexto de la presente invención se considera así como rutinario. Un ejemplo es el knockout mediante el uso de nucleasas de dedo de zinc específicas. Sigma Aldrich proporciona un kit respectivo como "CompoZR knockout ZFN". Otro enfoque abarca el uso de nucleasas efectoras similares a activadores de transcripción (TALEN).

La administración de una construcción negativa dominante implica la introducción de una secuencia que codifica un producto de expresión génica disfuncional, p. ej., por transfección. Dicha secuencia codificante está acoplada funcionalmente a un promotor fuerte, de tal manera que la expresión génica de la enzima disfuncional anula la expresión natural del producto de expresión génica, lo que, a su vez, conduce a un defecto fisiológico efectivo de la actividad respectiva de dicho producto génico de expresión.

Una desactivación genética condicional permite bloquear la expresión génica de una manera específica de tejido o de tiempo. Esto se hace, por ejemplo, introduciendo secuencias cortas llamadas sitios loxP aproximadamente el gen de interés. De nuevo, el experto en la materia conoce otros enfoques a partir de la bibliografía respectiva, y su aplicación en el contexto de la presente invención se considera rutinaria.

Otro enfoque es la alteración génica que puede conducir a un producto génico disfuncional o a un producto génico con actividad reducida. Este enfoque implica la introducción de mutaciones de cambio de marco, mutaciones sin sentido (es decir, la introducción de un codón de parada prematuro) o mutaciones que conducen a una sustitución de aminoácidos que hace que todo el producto del gen sea disfuncional o cause una actividad reducida. Tal alteración génica se puede producir, por ejemplo, mediante mutagénesis (p. ej., deleción, deleción parcial, inserción o sustitución de ácido nucleico), ya sea mutagénesis inespecífica (aleatoria) o mutagénesis dirigida al sitio. Los protocolos que describen la aplicación práctica del silenciamiento génico, knock-down génico, knockoutgénica, administración de una construcción negativa dominante, knockout génica condicional y/o alteración génica están comúnmente disponibles para el experto en la materia y están dentro de su rutina. La enseñanza técnica proporcionada en el presente documento está totalmente habilitada con respecto a todos los métodos imaginables que conducen a una inhibición o reducción de la expresión génica de un producto génico, o a la expresión de un producto génico disfuncional o inactivo, o con actividad reducida.

Las modificaciones genéticas descritas en el presente documento pueden emplear herramientas, métodos y técnicas conocidas en la técnica, tal como las descritas por J. Sambrook et al., Molecular Cloning: A Laboratory Manual (3.ª edición), Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Nueva York (2001).

El término "endógeno", como se usa en el presente documento, incluye aquellas moléculas y secuencias, en particular genes o proteínas endógenos, que están presentes en la célula huésped de tipo salvaje (nativa), antes de su modificación para reducir la expresión de los respectivos genes endógenos y/o reducir la producción de las proteínas endógenas. En particular, una molécula de ácido nucleico endógeno (p. ej., un gen) o proteína que se produce en (y se puede obtener de) una célula huésped particular tal como se encuentra en la naturaleza, se entiende que es "endógena de la célula huésped" o "endógena. a la célula huésped". Además, una célula que "expresa endógenamente" un ácido nucleico o proteína expresa ese ácido nucleico o proteína como lo hace un huésped del mismo tipo particular que se encuentra en la naturaleza. Además, una célula huésped que "produce endógenamente" o que "produce endógenamente" un ácido nucleico, proteína u otro compuesto produce ese ácido nucleico, proteína o compuesto al igual que una célula huésped del mismo tipo particular que se encuentra en la naturaleza.

Por lo tanto, incluso si una célula huésped deja de producir una proteína endógena, como en un mutante inactivado de la célula huésped, donde el gen que codifica la proteína se inactiva o elimina, la proteína todavía se denomina "endógena".

El término "heterólogo", tal como se usa en el presente documento con respecto a una secuencia de nucleótidos, una construcción tal como un casete de expresión, una secuencia de aminoácidos o una proteína, se refiere a un compuesto que es extraño a una célula huésped determinada, es decir, "exógeno", tal como no encontrado en la naturaleza en dicha célula huésped; o que se encuentra naturalmente en una célula huésped dada, p. ej., es "endógeno", sin embargo, en el contexto de una construcción heteróloga o integrada en dicha construcción heteróloga, p. ej., empleando un ácido nucleico heterólogo fusionado o junto con un ácido nucleico endógeno, lo que hace que la construcción sea heteróloga. La secuencia heteróloga de nucleótidos que se encuentra de forma endógena también se puede producir de forma no natural. *p.ej.,* mayor de lo esperado o mayor de lo que se encuentra naturalmente, cantidad en la celda. La secuencia de nucleótidos heteróloga, o un ácido nucleico que comprende la secuencia de nucleótidos heteróloga, posiblemente difiere en la secuencia de la secuencia de nucleótidos endógena pero codifica la misma proteína que se encuentra endógenamente. Específicamente, las secuencias de nucleótidos heterólogas son aquellas que no se encuentran en la misma relación con una célula huésped en la naturaleza. Cualquier secuencia de nucleótidos recombinante o artificial se entiende heteróloga. Un ejemplo de un polinucleótido heterólogo es una secuencia de nucleótidos no asociada de forma nativa con un promotor, p. ej., para obtener un promotor híbrido, o enlazado operativamente a una secuencia codificante, como se describe en el presente documento. Como resultado, se puede obtener un polinucleótido híbrido o quimérico. Otro ejemplo de un compuesto heterólogo es un polinucleótido que codifica una PDI enlazada operativamente a un elemento de control transcripcional, *p.ej.,* un promotor, al que normalmente no se une operativamente una secuencia codificante de PDI endógena de origen natural.

El término "enlazado operativamente", como se usa en el presente documento, se refiere a la asociación de secuencias de nucleótidos en una sola molécula de ácido nucleico, p. ej., un vector o un casete de expresión, de tal manera que la función de una o más secuencias de nucleótidos se ve afectada por al menos al menos otra secuencia de nucleótidos presente en dicha molécula de ácido nucleico. Mediante el enlace operativo, una secuencia de ácido nucleico se coloca en una relación funcional con otra secuencia de ácido nucleico en la misma molécula de ácido nucleico. Por ejemplo, un promotor se enlaza operativamente con una secuencia codificante de un gen recombinante, cuando es capaz de efectuar la expresión de esa secuencia codificante. Como ejemplo adicional, un ácido nucleico que codifica un péptido señal se une operativamente a una 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 enlazados operativamente entre sí pueden enlazarse 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.

Típicamente se entiende que una secuencia "promotora" está enlazada operativamente a una secuencia codificante, si el promotor controla la transcripción de la secuencia codificante. Si una secuencia promotora no está asociada de forma nativa con la secuencia codificante, su transcripción no está controlada por el promotor en células nativas (de tipo salvaje) o las secuencias se recombinan con diferentes secuencias contiguas.

El promotor que es regulable, en particular reprimible, por una fuente de carbono distinta de metanol y que se usa para el propósito descrito en el presente documento, se denomina en el presente documento "ECP". Por lo tanto, la presente divulgación con respecto al "ECP" también se referirá al "promotor que es regulable (o reprimible) por una fuente de carbono distinta del metanol", y *viceversa.*

La ECP, tal como se describe en el presente documento, en particular inicia, regula o de otro modo media o controla la expresión de un ADN que codifica una PDI. 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.

El ECP, tal como se describe en el presente documento, se entiende específicamente como un promotor regulable, en particular, un promotor regulable de fuente de carbono con diferente fuerza de promotor en el estado reprimido e inducido, en particular, un promotor regulable de fuente de carbono distinta de metanol, tal como el ECP, que es reprimible por una fuente de carbono distinta del metanol y, en particular, no inducible por el metanol. Específicamente, usando ECP que tiene actividad transcripcional en ausencia de metanol, no hay necesidad de añadir metanol al cultivo de células huésped para la producción de PDI bajo el control transcripcional de ECP.

La fuerza de la ECP se refiere específicamente a su fuerza de transcripción, representada por la eficiencia del inicio de la transcripción que ocurre en ese promotor con alta o baja frecuencia. Cuanto mayor sea la fuerza de transcripción, más frecuentemente se producirá la transcripción en ese promotor. La fuerza del promotor es una característica típica de un promotor, ya que determina la frecuencia con la que se transcribe una secuencia de ARNm dada, lo que otorga una mayor prioridad para la transcripción de algunos genes sobre otros, lo que lleva a una mayor concentración de la transcripción. Un gen que codifica una proteína que se requiere en grandes cantidades, por ejemplo, típicamente tiene un promotor relativamente fuerte. La ARN polimerasa solo puede realizar una tarea de transcripción a la vez y, por lo tanto, debe priorizar su trabajo para que sea eficiente. Las diferencias en la fuerza del promotor se seleccionan para permitir esta priorización.

La ECP utilizada en el presente documento es relativamente fuerte en el estado totalmente inducido, que típicamente se entiende como el estado de actividad máxima. La fuerza relativa se determina comúnmente con respecto a un promotor comparable, denominado en el presente documento promotor de referencia, que puede ser un promotor estándar, como el respectivo promotor pGAP de la célula utilizada como célula huésped.

La frecuencia de transcripción se entiende comúnmente como la velocidad de transcripción, p. ej., según lo determinado por la cantidad de un transcrito en un ensayo adecuado, p. ej., RT-PCR o transferencia Northern. Por ejemplo, la fuerza de transcripción de un promotor según la invención se determina en la célula huésped que es *P. pastoris* y en comparación con el promotor pGAP nativo de *P. pastoris.*

La fuerza de un promotor para expresar un gen de interés se entiende comúnmente como la fuerza de expresión o la capacidad de soportar un nivel/velocidad de expresión alto. Por ejemplo, la fuerza de expresión y/o transcripción de un promotor de la invención se determina en la célula huésped que es *P. pastoris* y en comparación con el promotor pGAP nativo de *P. pastoris.*

La fuerza de transcripción comparativa en comparación con un promotor de referencia puede determinarse mediante métodos estándar, tal como midiendo la cantidad de transcritos, p. ej., empleando un microarreglo, o también en un cultivo celular, tal como midiendo la cantidad de los respectivos productos de expresión génica en células recombinantes. En particular, la velocidad de transcripción puede determinarse por la fuerza de transcripción en una micromatriz, transferencia Northern o con PCR cuantitativa en tiempo real (qRT-PCR) o con secuenciación de ARN (RNA-seq) donde los datos muestran la diferencia del nivel de expresión entre condiciones con una velocidad de crecimiento alta y condiciones con una velocidad de crecimiento baja, o condiciones que emplean una composición de medios diferente y una intensidad de señal alta en comparación con el promotor de referencia.

La velocidad de expresión puede, por ejemplo, estar determinada por la cantidad de expresión de un gen informador, tal como eGFP.

ECP como se describe en el presente documento ejerce una fuerza de transcripción relativamente alta, p. ej., reflectada por una velocidad de transcripción o fuerza de transcripción de al menos 15% en comparación con el promotor pGAP nativo en la célula huésped, también llamado "promotor pGAP homólogo". Preferiblemente, la velocidad o fuerza de transcripción es al menos cualquiera de 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90% o 100%, o incluso mayor, como al menos cualquiera de 110%, 120%, 130%, 140%, 150%, 160%, 170%, 180%, 190% o 200% en comparación con el promotor pGAP nativo, tal como se determina en la (*p.ej.* eucariota) célula huésped seleccionada como célula huésped con fines de recombinación para producir la PDI.

El promotor pGAP nativo típicamente inicia la expresión del gen GAP que codifica la gliceraldehído-3-fosfato deshidrogenasa (GAPDH), que es un promotor constitutivo presente en la mayoría de los organismos vivos. GAPDH (EC 1\2\1\12), una enzima clave de la glucólisis y la gluconeogénesis, juega un papel crucial en el metabolismo de carbohidratos catabólicos y anabólicos.

El promotor pGAP nativo es específicamente activo en una célula eucariota recombinante de forma similar a una célula eucariota nativa de la misma especie o cepa, incluyendo la célula eucariota no modificada (no recombinante) o recombinante. Tal promotor pGAP nativo se entiende comúnmente como un promotor endógeno, por lo tanto, homólogo a la célula huésped, y puede servir como promotor estándar o de referencia con fines comparativos. La expresión relativa o fuerza de transcripción de un promotor como se describe en este documento se compara normalmente con el promotor pGAP nativo de una célula de la misma especie o cepa que se utiliza como huésped para producir una PDI.

El término "regulable" con respecto a un elemento regulador inducible o reprimible, tal como un promotor descrito en este documento, se referirá a un elemento que se reprime en una célula huésped en presencia de una cantidad excesiva de una sustancia (tal como un nutriente en el medio de cultivo celular) *p.ej.*, en la fase de crecimiento de un cultivo por lotes, y desreprimida para inducir una fuerte actividad *p.ej.,* en la fase de producción (como cuando se reduce la cantidad de un nutriente o cuando se alimenta con un sustrato suplementario), según una estrategia de alimentación por lotes. Un elemento regulador también puede diseñarse para que sea regulable, de modo que el elemento sea inactivo sin la adición de un aditivo de cultivo celular y activo en presencia de dicho aditivo. Por lo tanto, la expresión de una PDI bajo el control de dicho elemento regulador puede inducirse tras la adición de tal aditivo.

El ECP, como se describe en este documento, es un promotor regulable relativamente fuerte que normalmente se silencia o reprime en condiciones de crecimiento celular (fase de crecimiento) y se activa o desreprime en condiciones de producción (fase de producción) y, por lo tanto, es adecuado para inducir la producción de PDI en una línea celular de producción limitando la fuente de carbono.

Específicamente, el promotor como se describe en el presente documento es una fuente de carbono regulable con una fuerza de promotor diferencial determinada en una prueba que compara su fuerza en presencia de glucosa y limitación de glucosa, lo que demuestra que todavía está reprimido a concentraciones de glucosa relativamente altas, preferiblemente a concentraciones de al menos al menos 10 g/L, preferiblemente al menos 20 g/L. Específicamente, el promotor descrito en el presente documento se induce completamente a concentraciones limitadas de glucosa, considerando las concentraciones umbral de glucosa que inducen completamente al promotor, cuyo umbral es típicamente inferior a 20 g/L, preferiblemente inferior a 10 g/L, inferior o superior a 1 g/L, incluso menos de 0,1 g/L o menos de 50 mg/L, preferiblemente con una fuerza de transcripción completa de p. ej., al menos el 50% del promotor pGAP homólogo nativo, a concentraciones de glucosa de menos de 40 mg/L.

El término "represión" o "reprimido", como se usa en el presente documento dentro del contexto de la presente descripción, p. ej., para caracterizar un promotor regulable de fuente de carbono descrito en el presente documento, se refiere a la interferencia de la transcripción de un gen de interés (que codifica una proteína de interés) que está bajo el control transcripcional de un promotor que se entiende que es reprimible, lo que resulta en una disminución de la expresión de la proteína de interés por parte de la(s) célula(s).

La represión de la ECP descrita en el presente documento se produce específicamente cuando hay un agente represor en el medio de cultivo celular. Un agente de represión puede ser una determinada fuente de carbono o una cantidad de represión de una fuente de carbono, p. ej., por encima de una cierta cantidad umbral. Se dice que la expresión de un gen de interés o de una proteína de interés está "desreprimida", cuando el agente represor se elimina del medio o se reduce por debajo de una cantidad umbral que ya no es represora. Tras la desrepresión, se entiende que la ECP está completamente inducida, y la expresión de la proteína de interés es normalmente al menos 1,5 veces superior a los niveles basales de expresión de la(s) célula(s) en condiciones de represión del promotor.

Específicamente, la transcripción de un gen de interés bajo el control de la ECP descrita en el presente documento puede ser reprimida por al menos cualquiera de 30, 40, 50, 60, 70, 80, 85%, 90% o 95%, o completamente reprimida (100% reprimido) en comparación con la transcripción de dicho gen al desreprimir o inducir completamente el ECP.

La fuerza del promotor diferencial que compara la fuerza del promotor en condiciones reprimidas y desreprimidas determina las propiedades regulables de un promotor y la relación de inducción respectiva. Según con ciertas

realizaciones, la relación de inducción se entiende como una fuerza de promotor diferencial que está determinada por el inicio de la producción de PDI al cambiar a condiciones de inducción por debajo de un umbral de fuente de carbono predeterminado, y se compara con la fuerza en el estado reprimido. La fuerza de transcripción comúnmente se entiende como la fuerza en el estado totalmente inducido, es decir, que muestra actividades máximas en condiciones de desrepresión. La fuerza del promotor diferencial es, p. ej, determinado según la eficiencia o el rendimiento de la producción de PDI en una línea celular huésped recombinante en condiciones de desrepresión en comparación con las condiciones de represión, o también por la cantidad de un transcrito. El promotor regulable como se describe en el presente documento tiene una fuerza de promotor diferencial preferida (proporción de inducción), que es al menos 1,5 veces o al menos 2 veces, más preferiblemente al menos 5 veces, incluso más preferiblemente al menos 10 veces, más preferiblemente al menos 20 veces, más preferiblemente al menos 30, 40, 50 o 100 veces en el estado desreprimido (totalmente inducido) en comparación con el estado reprimido, también entendido como inducción de pliegue.

El término "mutagénesis", como se usa en el presente documento, se referirá a un método para proporcionar mutantes de una secuencia de nucleótidos, p. ej., mediante inserción, deleción y/o sustitución de uno o más nucleótidos, para obtener variantes de los mismos con al menos un cambio en la región codificante o no codificante. La mutagénesis puede ser mediante mutación aleatoria, semialeatoria o dirigida al sitio. La ECP específica descrita en el presente documento y las secuencias de nucleótidos respectivas se pueden usar para producir variantes, que también son promotores regulables que se pueden usar para el propósito que se describe en el presente documento. Tales variantes se pueden producir mediante un método de mutagénesis adecuado utilizando las secuencias de nucleótidos de ECP proporcionadas en el presente documento como secuencia original. tal método de mutagénesis abarca aquellos métodos de modificación genética del ácido nucleico o sintetizar *de novo* una secuencia de nucleótidos utilizando la información de la secuencia del promotor parental respectivo como plantilla. Los métodos de mutagénesis específicos aplican la modificación genética de promotores racionales.

El ECP ejemplar descrito en el presente documento puede, p. ej., modificarse para generar variantes de promotor con niveles de expresión alterados y propiedades reguladoras. Por ejemplo, se puede preparar una biblioteca de promotores mediante mutagénesis de secuencias de promotores seleccionadas, que se pueden usar como moléculas originales, p. ej., afinar la expresión génica en células eucariotas analizando variantes para su expresión bajo diferentes estrategias de fermentación y seleccionando variantes adecuadas. Se puede usar una biblioteca sintética de variantes, p. ej., para seleccionar un promotor que cumpla con los requisitos para producir una PDI seleccionado. Tales variantes pueden tener una eficacia de expresión aumentada en células huésped (p. ej., eucariotas) y una expresión diferencial en condiciones limitantes y ricas en fuentes de carbono. Típicamente, se producen grandes bibliotecas de genes aleatorias con una gran diversidad de genes, que pueden seleccionarse de acuerdo con un genotipo o fenotipo específicamente deseado.

Ciertas variantes de ECP pueden ser variantes de tamaño de las secuencias de nucleótidos de ECP proporcionadas en este documento y/o comprender más de uno de los elementos o regiones del promotor descrito en este documento, tales como las regiones reguladoras centrales, las regiones reguladoras principales o los motivos T, y /o comprenden uno o más (del mismo o diferentes) fragmentos de las secuencias de nucleótidos de ECP.

Los métodos de mutagénesis específicos proporcionan mutaciones puntuales de uno o más nucleótidos en una secuencia, en particular mutaciones puntuales en tándem, tal como cambiar al menos 2, 3, 4, 5, 6, 7, 8, 9, 10 o incluso nucleótidos más continuos dentro de la secuencia de nucleótidos del promotor. Una mutación puntual es típicamente al menos una deleción, inserción y/o sustitución de uno o más nucleótidos. La secuencia promotora puede estar mutada en los extremos distales, en particular dentro de la región 5' que representa hasta el 50% de la secuencia promotora de longitud completa, región 5' que puede ser muy variable sin perder sustancialmente la actividad promotora. La secuencia del promotor puede estar mutada específicamente dentro de la región reguladora principal, aunque puede preferirse que la identidad de la secuencia con la región reguladora principal ejemplar y, en particular, con la región reguladora central ejemplar sea alta, tal como p. ej., al menos cualquiera de 80%, 85%, 90% o 95%. Fuera de cualquiera de las regiones reguladoras centrales o principales, la variabilidad de la secuencia puede ser mayor y el ECP aún puede ser funcional, por ejemplo, con una identidad de secuencia de menos del 80% o menos del 85%.

Cualquier mutación dentro de las regiones reguladoras centrales o principales es típicamente conservadora, es decir, para mantener (o incluso mejorar) el reconocimiento por parte de un determinado factor de transcripción.

Específicamente, la ECP descrita en el presente documento puede comprender una secuencia de nucleótidos híbrida, p. ej., que comprende las regiones reguladoras centrales o principales descritas en el presente documento y, además, una o más regiones o secuencias promotoras alternativas (nativas o artificiales), como un sitio de iniciación de la traducción en la región 3' (específicamente el extremo 3' que comprende al menos 10 o 15 secuencias de nucleótidos del extremo 3' que incluyen el extremo 3' (p. ej., hasta 20, 25 o 30 nt) de un promotor diferente, por ejemplo, de cualquier promotor constitutivo o regulable (o inducible de otro modo), sustituyendo así la traducción sitio de iniciación del promotor ECP.

El término "secuencia de nucleótidos" o "secuencia de ácidos nucleicos" que se usa en el presente documento se refiere a ADN o ARN. "Secuencia de ácido nucleico" o "secuencia de polinucleótido" o simplemente "polinucleótido"

se refiere a un polímero de cadena simple o doble de bases de desoxirribonucleótido o ribonucleótido leídas desde el extremo 5' al 3'. Incluye casetes de expresión, plásmidos autorreplicantes, polímeros infecciosos de ADN o ARN y ADN o ARN no funcional.

El término "proteína de interés (PDI)", como se usa en el presente documento, se refiere a un polipéptido o una proteína que se produce por medio de tecnología recombinante en una célula huésped. Más específicamente, la proteína puede ser un polipéptido que no se encuentra de forma natural en la célula huésped, es decir, una proteína heteróloga, o bien puede ser nativa de la célula huésped, es decir, una proteína homóloga a la célula huésped, pero se produce, por ejemplo, mediante transformación con un vector autorreplicante que contiene la 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 el PDI en el genoma de la célula huésped, o mediante la modificación recombinante de una o más más secuencias reguladoras que controlan la expresión del gen que codifica la PDI, *p.ej.*, de la secuencia promotora. En algunos casos, el término PDI como se usa en el presente documento también se refiere a cualquier metabolito producido por la célula huésped mediado por la proteína expresada recombinantemente.

El término "andamio" como se usa en el presente documento describe un grupo multifacético de proteínas plegadas de manera estable y compacta - que difieren en tamaño, estructura y origen - que sirven como punto de partida para la generación de moléculas de unión a antígeno. Inspirado en las relaciones estructura-función de los anticuerpos (inmunoglobulinas), este andamio de proteína alternativo proporciona un marco estructural conservado y sólido que admite un sitio de interacción que se puede remodelar para el reconocimiento estricto y específico de un objetivo (bio)molecular dado.

El término "identidad de secuencia" de una variante, homólogo u ortólogo en comparación con una secuencia original de nucleótidos o aminoácidos indica el grado de identidad de dos o más secuencias. Dos o más secuencias de aminoácidos pueden tener los mismos residuos de aminoácidos o conservados en una posición correspondiente, hasta cierto punto, hasta el 100%. Dos o más secuencias de nucleótidos pueden tener pares de bases iguales o conservados en una posición correspondiente, hasta cierto punto, hasta el 100%.

La búsqueda de similitud de secuencias es una estrategia efectiva y confiable para identificar homólogos con exceso de identidad de secuencia (*p. ej.*, al menos 50%). Las herramientas de búsqueda de similitud de secuencias que se utilizan con frecuencia son *p. ej.*, BLAST, FASTA y HMMER.

Las búsquedas de similitud de secuencias pueden identificar tales proteínas o genes homólogos al detectar un exceso de similitud y una similitud estadísticamente significativa que refleja una ascendencia común. Los homólogos pueden abarcar ortólogos, que se entienden en el presente documento como la misma proteína en diferentes organismos, *p. ej.*, variantes de dicha proteína en diferentes organismos o especies diferentes.

Una secuencia ortóloga de la misma proteína en diferentes organismos o especies es típicamente homóloga a la secuencia de la proteína, específicamente de ortólogos que se originan del mismo género. Típicamente, los ortólogos tienen al menos cualquiera de 25%, 30%, 35%, 40%, 50%, 60%, 70%, 80%, 85%, 90% o 95% de identidad, hasta 100% de secuencia identidad. Específicamente, los ortólogos se pueden determinar al reemplazar la proteína FLO8 o el gen que codifica la proteína FLO8 por las secuencias ortólogas en una célula huésped, que se modifica para eliminar la proteína FLO8 endógena. Por ejemplo, si una supuesta proteína FLO8 es funcional en una célula huésped *P. pastoris* o *S. cerevisiae* reemplazando la proteína FLO8 endógena que está codificada por un gen que ha sido eliminado en tal célula huésped *P. pastoris* y *S. cerevisiae*, respectivamente, dicha supuesta proteína FLO8 puede considerarse un homólogo de la proteína FLO8 para el propósito descrito en el presente documento.

La proteína FLO8 que comprende o consiste en la secuencia de aminoácidos identificada como SEQ ID NO: 1 es de origen *K. phaffii*. Se entiende bien que hay secuencias homólogas presentes en otras células huésped eucariotas o procariotas. Por ejemplo, las células de levadura comprenden las respectivas secuencias homólogas, en particular en levaduras de *Pichia pastoris,* que ha sido reclasificado en un nuevo género, *komagataella,* y se divide en tres especies, *K. pastoris, K. phaffii,* y *K. pseudopastoris.* Las secuencias homólogas específicas son *p. ej.*, encontrado en *K. pastoris* (*p. ej.,* SEQ ID NO: 3, tal como la codificada por la secuencia de nucleótidos que comprende o consiste en SEQ ID NO: 4), *Saccharomyces cerevisiae* (*p. ej.,* SEQ ID NO: 5 o SEQ ID NO: 6), *Yarrowia lipolytica* (*p. ej.,* SEQ ID NO: 7), *Ogataea polimorfa* (*p. ej.,* SEQ ID NO: 8), o *Aspergillus niger* (*p. ej.,* SEQ ID NO: 9).

Cualquier secuencia homóloga de la proteína FLO8 con cierta identidad de secuencia descrita en el presente documento, en particular cualquier proteína FLO8 que sea un ortólogo de la proteína FLO8 de *P. pastoris* está incluida en la definición de una proteína FLO8 descrita en el presente documento.

El "porcentaje (%) de identidad de secuencia de aminoácidos" con respecto a una secuencia de aminoácidos, homólogos y ortólogos descritos en el presente documento se define como el porcentaje de residuos de aminoácidos en una secuencia candidata que son idénticos a los residuos de aminoácidos en la secuencia polipeptídica específica. después de alinear la secuencia e introducir espacios, si es necesario, para lograr el máximo porcentaje de identidad de secuencia, y no considerar ninguna sustitución conservadora como parte de la identidad de secuencia. Los expertos en la técnica pueden determinar los parámetros apropiados para medir la alineación, incluyendo los algoritmos necesarios para lograr la máxima alineación en toda la longitud de las secuencias que se comparan.

Para los fines descritos en el presente documento, la identidad de secuencia entre dos secuencias de aminoácidos se determina utilizando el programa NCBI BLAST versión BLASTP 2.8.1 con los siguientes parámetros de ejemplo: Programa: blastp, Tamaño de palabra: 6, Valor esperado: 10, Tamaño de lista de aciertos: 100, Gapcosts: 11.1, Matriz: BLOSUM62, Cadena de filtro: F, Ajuste de composición: Ajuste de matriz de puntuación de composición condicional.

El "porcentaje (%) de identidad" con respecto a una secuencia de nucleótidos, p. ej., de un promotor o un gen, se define como el porcentaje de nucleótidos en una secuencia de ADN candidata que es idéntica a los nucleótidos en la secuencia de ADN, después de alinear la secuencia e introducir huecos, si es necesario, para conseguir el máximo porcentaje de identidad de secuencia, y no considerar ninguna sustitución conservadora como parte de la identidad de secuencia. El alineamiento con el fin de determinar el porcentaje de identidad de la secuencia de nucleótidos se puede lograr de varias formas que están dentro de la experiencia en la técnica, por ejemplo, usando software disponible públicamente. Los expertos en la técnica pueden determinar los parámetros apropiados para medir la alineación, incluidos los algoritmos necesarios para lograr la máxima alineación en toda la longitud de las secuencias que se comparan.

El término "aislado" o "aislamiento", tal como se usa en el presente documento con respecto a una PDI, se referirá a tal compuesto que se ha separado lo suficiente del entorno con el que se asociaría de forma natural, en particular, un sobrenadante de cultivo celular, para existir en forma "purificada" o "sustancialmente pura". Sin embargo, "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. Los compuestos aislados se pueden formular adicionalmente para producir preparaciones de los mismos, y todavía para fines prácticos se pueden aislar; por ejemplo, una PDI se puede mezclar con vehículos o excipientes farmacéuticamente aceptables cuando se usa en diagnóstico o terapia.

El término "purificado" tal como se utiliza en el presente documento se refiere a una preparación que comprende al menos el 50% (mol/mol), preferiblemente al menos el 60%, 70%, 80%, 90% o 95% de un compuesto (*p. ej.,* una PDI). La pureza se mide por métodos apropiados para el compuesto (*p. ej.,* métodos cromatográficos, electroforesis en gel de poliacrilamida, análisis HPLC y similares). Una PDI aislada y purificada como se describe en el presente documento puede obtenerse mediante la purificación de los sobrenadantes del cultivo celular para reducir las impurezas.

Como métodos de aislamiento y purificación para obtener un producto de proteína o polipéptido recombinante, métodos, tales como métodos que utilizan diferencia en solubilidad, tales como precipitación con disolvente y salificación, métodos que utilizan diferencia en peso molecular, tal como ultrafiltración y electroforesis en gel, métodos que utilizan diferencia en carga eléctrica, tal como cromatografía de intercambio iónico, métodos que utilizan afinidad específica, como cromatografía de afinidad, métodos que utilizan diferencia en hidrofobicidad, tal como cromatografía líquida de alta resolución de fase inversa, y métodos que utilizan diferencia en el punto isoeléctrico, tal como enfoque isoeléctrico.

Se prefieren los siguientes métodos estándar: separación de células (desechos) y lavado por microfiltración o filtro de flujo tangencial (TFF) o centrifugación, purificación de PDI por precipitación o tratamiento térmico, activación de PDI por digestión enzimática, purificación de PDI por cromatografía, como intercambio iónico (IEX), cromatografía de interacción hidrofóbica (HIC), cromatografía de afinidad, cromatografía de exclusión por tamaño (SEC) o cromatografía HPLC, precipitación PDI de concentración y lavado por etapas de ultrafiltración.

Un producto altamente purificado está esencialmente libre de proteínas contaminantes y preferiblemente tiene una pureza de al menos 90%, más preferiblemente al menos 95%, o incluso al menos 98%, hasta 100%. Los productos purificados pueden obtenerse por purificación del sobrenadante del cultivo celular o también a partir de restos celulares.

Un punto de interés aislado y purificado se puede identificar mediante métodos convencionales como transferencia Western, HPLC, ensayo de actividad o ELISA.

El término "recombinante", tal como se usa en el presente documento, significará "preparado por o como resultado de la modificación genética". Un huésped recombinante puede modificarse genéticamente para eliminar y/o inactivar uno o más nucleótidos o secuencias de nucleótidos, y puede comprender específicamente un vector de expresión o vector de clonación que contiene una secuencia de ácido nucleico recombinante, en particular empleando una secuencia de nucleótidos ajena al huésped. Una proteína recombinante se produce expresando un ácido nucleico recombinante respectivo en un huésped. El término "recombinante" con respecto a una PDI como se usa en este documento, incluye una PDI que se prepara, expresa, crea o aísla por medios recombinantes, tal como una PDI aislado de una célula huésped transformada para expresar la PDI. De acuerdo con la presente invención técnicas convencionales de biología molecular, microbiología y ADN recombinante dentro de la experiencia del puede emplearse la técnica. Tales técnicas se explican completamente en la literatura. Véase, por ejemplo, Maniatis, Fritsch & Sambrook, "Molecular Cloning: A Laboratory Manual, Cold Spring Harbor, (1982).

Ciertas células huésped recombinantes son células huésped "modificadas genéticamente", entendidas como células huésped que han sido manipuladas mediante ingeniería genética, *es decir,* por intervención humana. Cuando se diseña una célula huésped para reducir la expresión o subexpresar un gen dado o la proteína respectiva, la célula

huésped se manipula de tal manera que la célula huésped ya no tiene la capacidad de expresar dicho gen y proteína, respectivamente, en comparación con la célula huésped bajo la misma condición antes de la manipulación, o en comparación con las células huésped que no están modificadas genéticamente de manera que dicho gen o proteína se subexprese.

Según ejemplos específicos, sorprendentemente se ha encontrado que la reducción de la expresión de un gen que codifica FLO8 en una célula huésped, en particular la deleción de dicho gen tuvo una influencia positiva en los niveles de expresión de genes controlados por promotores inducibles no controlados por metanol (ECP), lo que permite niveles de expresión más altos sin perder la regulación del promotor de la fuente de carbono. Por lo tanto, las células con y sin deleción del respectivo gen endógeno que codifica la proteína FLO8 (gen *flo8*) y se probaron la fuerza y la regulación de los genes bajo el control de un ECP. Para varios puntos de interés de ejemplo, que son proteínas modelo intracelulares o secretadas, se demostró que la expresión aumentó en las células con la eliminación del respectivo gen *flo8*, que fue rastreable a niveles de transcripción aumentados. El aumento de la expresión podría mostrarse tanto en cultivos de cribado a pequeña escala como en procesos de producción controlados en un biorreactor. Por lo tanto, se ha desarrollado un nuevo sistema de expresión que permite una formación de productos mucho mayor en medios libres de metanol. *p. ej.* en levaduras como Pichia, en comparación con sistemas de expresión previamente existentes.

Según ejemplos específicos, el efecto de la interrupción de FLO8 en la fuerza de expresión de los promotores pG1, pG3, pG4, pG6 regulados por una fuente de carbono distinta de metanol (que están regulados por una fuente de carbono distinta del metanol) se comparó con el promotor constitutivo pGAP o el promotor pAOX1 inducido por metanol. Se encontró que todos los promotores regulados por fuentes de carbono distintos de metanol tenían una transcripción mayor estadísticamente significativa en la cepa dFLO8, lo que es sorprendente e indica una mayor expresión bajo el control de dichos promotores. Por el contrario, pAOX1 inducible por metanol no mostró una fuerza de transcripción significativamente mayor en el mutante dFLO8 en comparación con el tipo salvaje, y tampoco hubo un efecto significativo en la fuerza de transcripción de pGAP. Por tanto, se concluye que la actividad del promotor pGAP o pAOX1 no se ve afectada por la infraexpresión de FLO8.

Sorprendentemente, se descubrió que la expresión de un gen heterólogo de interés para producir una proteína de interés en un cultivo celular se ha incrementado de manera efectiva tras un knockouto interrupción de FLO8, cuando se usan promotores reprimibles de fuente de carbono que no son inducibles por metanol, en comparación con el promotor pGAP estándar. Por ejemplo, se encontró un aumento en la fluorescencia de eGFP para cada uno de los promotores pG1, pG1-3, pG3, pG4, pG6, pG7 y pG8 en las cepas *dFLO8*, que varió entre 1,2 y 3,9 veces de aumento en comparación con la expresión en la célula huésped de tipo salvaje (sin interrupción de *FLO8*).

La descripción anterior se comprenderá mejor con referencia a los siguientes ejemplos. Tales ejemplos son, sin embargo, meramente representativos de métodos para practicar una o más realizaciones de la presente invención y no deben interpretarse como limitantes del alcance de la invención. El alcance de esta solicitud está siendo definido por las reivindicaciones adjuntas.

Ejemplos

Los ejemplos a continuación demostrarán que la interrupción del regulador transcripcional *FLO8* conduce a una mayor actividad transcripcional de los promotores regulados por carbono tales como pG1, pG3, pG4, pG6 y pG8, y variantes modificadas de los mismos que permiten una mayor productividad de proteínas recombinantes en condiciones de cultivo limitadas en carbono.

Ejemplo 1: Construcción de cepas d*FLO8* de *P. pastoris*

Se utilizó la cepa de tipo salvaje CBS7435 o CBS2612 de *P. pastoris* (CBS-KNAW Fungal Biodiversity Centre, Centraalbureau voor Schimmelcultures, Utrecht, Países Bajos) como cepa huésped.

Para generar las cepas mutantes dFLO8, el gen PP7435_Chr4-0252 (*FLO8*) se interrumpió con el método de casete de marcador dividido adaptado para *P. pastoris* (Gasser y *Alabama.,* 2013) y descrito en el documento WO2015158800A1. Brevemente, se amplificaron dos regiones de 1,5 kb ubicadas aproximadamente 200 pb aguas arriba y aguas abajo del inicio de la traducción del ORF utilizando los cebadores A_fw y A_bw, así como D_fw y D_bw, respectivamente **(Tabla 1).** Los fragmentos A y B resultantes se usaron para flanquear dos ca. Partes superpuestas y de 1 kb de longitud (435 bps) del casete marcador KanMX (cebadores B_fw, B_bw, C_fw y C_bw) mediante PCR de fusión, utilizando salientes en los cebadores A_bw y D_fw que eran homólogos al extremo 5' y 3' de las respectivas partes B y C del casete marcador de resistencia. Los dos fragmentos fusionados AB y CD se transformaron simultáneamente en células *P. pastoris* electrocompetentes como se describe en (Gasser *et al.,* 2013). La integración exitosa requiere tres eventos de recombinación diferentes, lo que resultó en el reemplazo de un fragmento de 0,4 kb en el extremo 5' de PP7435_Chr4-0252 y su promotor por el casete KanMX.

La selección de transformantes positivos se realizó en placas de agar-YPD selectivas (por litro: 10 g de extracto de levadura, 20 g de peptona, 20 g de glucosa, 20 g de agar-agar) que contenían Geneticina 500 µg mL^{-1}. Los mutantes de deleción correctos se verificaron mediante PCR con cebadores ubicados fuera del casete de marcador dividido (Det_fw y Det_bw, **tabla 1**) y electroforesis en gel.

Tabla 1: Cebadores para la construcción de casetes de marcadores divididos

Cebador	**Secuencia**
A_fw	CGAACATCCATCACCAAAACAC (SEQ ID NO: 72)
A bw	GTTGTCGACCTGCAGCGTACGGTGTTGCCGCGAAATG (SEQ ID NO: 73)
D_fw	TAGGTGATATCAGATCCACTGATCAATTTGCCCAAGAGACG (SEQ ID NO: 74)
D bw	GACTGTTGCGATTGCTGGTG (SEQ ID NO: 75)
B_fw	CATTTCGCGGCAACACCGTACGCTGCAGGTCGACAAC (SEQ ID NO: 76)
B bw	CGGTGAGAATGGCAAAAGCTTAT (SEQ ID NO: 77)
C_fw	AAGCCCGATGCGCCAGAGTTG (SEQ ID NO: 78)
C bw	CGTCTCTTGGGCAAATTGATCAGTGGATCTGATATCACCTA (SEQ ID NO: 79)
fw det	ATCCAGGACACGCTCATCAAG (SEQ ID NO: 80)

Ejemplo 2: Efecto de interrupción de *FLO8* en la productividad de eGFP intracelular impulsada por pG1 y pG1-3

a) Construcción de cepas dFLO8 de P. pastoris

P. pastoris CBS2612_pG1_eGFP #8 (descrito en el documento WO2013050551A1) y CBS2612_pG1-3_eGFP#1 (descrito en el documento WO2017021541A1 y Prielhofer *et al.,* 2018 como CBS2612_pGTH1-D1240) se utilizó como cepas huésped. Se ha demostrado que estas cepas han integrado una sola copia de un casete de resistencia a la zeocina junto con el casete de expresión de eGFP compuesto por el promotor pG1 regulado por glucosa (SEQ ID NO: 12) o una variante diseñada del mismo (pG1-3, SEQ ID NO :10), el GDI y el terminador de transcripción *CYC1* de *S. cerevisiae*. Las cepas mutantes d*FLO8* correspondientes se construyeron como se describe en el **Ejemplo 1.**

b) Cribado de la productividad de eGFP

Para el cribado de expresión, se inocularon colonias individuales de las cepas dFLO8, así como sus respectivas cepas parentales y una cepa de tipo salvaje no productora en 2 mL de medio YP líquido (por litro: 20 g de peptona, 10 g de extracto de levadura) que contenía Zeocina 25 µg mL^{-1} y Geneticina 500 µg mL^{-1} (si procede). Se cultivaron cultivo previocultivos previos durante aprox. 24 h a 25 °C y 280 rpm en 24-DWP y posteriormente utilizado para inocular 2 ml de medio de cribado sintético ASMv6 (la composición del medio se indica a continuación) que contiene polisacárido 50 g L^{-1} y 1,5% de enzima liberadora de glucosa (lo que permite una velocidad de liberación de glucosa de aprox. 0,8 $mg \cdot mL^{-1}$ h^{-1}; kit de desarrollo de medios m2p) a un OD_{600} inicial de 5 (condiciones inductoras). Para las condiciones de represión, se utilizó ASMv6 que contenía glicerol al 2%. A continuación, los cultivos principales se incubaron durante otras 48 h a 25 °C y 280 rpm. Para medir la expresión de eGFP, las células se diluyeron a una OD_{600} de 0,1 en disolución salina tamponada con fosfato (PBS) y se analizaron por citometría de flujo como se describe en Stadlmayr et al., 2010. Para cada muestra se analizaron 15 000 células. La auto-fluorescencia de *P. pastoris* se midió usando células de tipo salvaje *P. pastoris* y se restaron de la señal. Los niveles de expresión de eGFP normalizados (intensidad de fluorescencia relacionada con el tamaño de la célula) se proporcionan como porcentaje de la expresión controlada por pGAP **(Tabla 2).**

Medio de cribado sintético ASMv6 contenido por litro: 22,0 g de ácido cítrico monohidrato, 6,30 g $(NH_4)_2HPO_4$, 0,49 g de $MgSO_4*7H_2O$, 2,64 g KCl, 0,0535 g $CaCl_2*2H_2O$, 1,470 ml de disolución madre de sales traza de PTM0, 0,4 mg de biotina; El pH se ajustó a 6,5 con KOH (sólido).

Disolución madre de sales traza PTM0 contenida por litro:6,0 g $CuSO_4*5H_2O$, 0,08 g NaI, 3,36 g $MnSO_4*H_2O$, 0,2 g $Na_2MoO_4*2H_2O$, 0,02 g H_3BO_3, 0,82 g $CoCl_2*6H_2O$, 20,0 g ZnCl2, 65,0 g $FeSO_4*7H_2O$ y 5,0 ml H_2SO_4 (95%-98%).

Tabla 2: Impacto de d*FLO8* sobre la expresión de eGFP bajo el control de pG1 o pG1-3. Se muestran los niveles de expresión de eGFP en relación con pGAP después de 48 h de cultivo en glicerol al 2% (represión) o glucosa limitante (inducción), así como el aumento de la fluorescencia de eGFP en la cepa d*FLO8* en condiciones de inducción en comparación con pG1 o pG1-3 en el tipo salvaje.

Promotor	Célula huésped	Represión (glicerol)	Inducción (glucosa limitante)	cambio de veces pG1+dFL *O8* vs pG1 + peso	cambio de veces pG1-3+d*FLO8* vs a pG1-3+wt
pGAP	Tipo salvaje	100	100	-	-
pG1	tipo salvaje	12,31 +1,13	250,17 + 8,73	1,0	-
pG1	dFLO8	24,84 ± 3,28	950,92 ± 4,64	3,8	-
pG1	d*FLO8*_loxP	23,14 ± 0,22	983,49 ± 8,48	3,9	-
pG1-3	tipo salvaje	8,71 + 0,46	519,94 ± 53,88	2,1	1,0
pG1-3	dFLO8	17,50 ± 0,37	1448,20 ± 20,62	5,8	2,8
pG1-3	d*FLO8*_loxP*	17,86 ± 1,14	1507,22 ± 9,80	6,0	2,9
* producido por un método como se describe en el Ejemplo 2.					

La delección de *FLO8* tuvo una influencia positiva en la expresión impulsada por pG1, lo que llevó a niveles de eGFP casi 4 veces más altos en condiciones inductoras (limitantes de glucosa) **(Tabla 2).** Por lo tanto, el efecto de interrumpir *FLO8* también se estudió la variante del promotor pG1-3, que tiene una mayor expresión intrínseca *per se.* El impacto positivo también se mostró para la variante del promotor, lo que nuevamente permitió niveles de eGFP de 2,5 a 3 veces más altos en la cepa dFLO8 en comparación con la expresión del mismo promotor en el fondo de tipo salvaje **(Tabla 2).**

Ejemplo 3: Efecto de la interrupción de *FLO8* en la productividad de P_{G1-3} impulsada de proteínas recombinantes secretadas

A continuación, se evaluó el impacto de la mutación de d*FLO8* en la expresión impulsada por pG1-3 de proteínas modelo secretoras. Para ello, el casete de expresión para vHH o scR se transformó en CBS2612 y CBS2612_d*FLO8*.

a) Construcción de cepas d*FLO8* de *P. pastoris* y reciclado de marcadores de selección

Se usó la cepa de tipo salvaje CBS2612 de *P. pastoris* (CBS-KNAW Fungal Biodiversity Centre, Centraalbureau voor Schimmelcultures, Utrecht, Países Bajos) como cepa huésped.

CBS2612_KanR_d*FLO8* #2 fue construido como se describe en el **Ejemplo 1.** Para eliminar el casete del marcador de selección KanMX del genoma basado en la recombinación Cre-loxP, CBS2612_KanR_d*FLO8* #2 se transformó transitoriamente con el plásmido pTAC_Cre_hphMX4 por electroporación como se describe en Gasser *et al.,* 2013. pTAC_Cre_hphMX4 es un derivado del plásmido pYX022 (R&D Systems) y comprende un origen de replicación para *E. coli,* un casete de expresión para el gen Cre-recombinasa, un casete de resistencia a higromicina, así como un elemento ARS/CEN. Para su construcción pTAC_Cre_kanMX (Marx *et al.,* 2008) fue digerido con *Eco*R91I para eliminar el casete de resistencia a la kanamicina. El plásmido pGA26_hphMX4 se digirió con las mismas enzimas para escindir el casete de resistencia a la higromicina y los fragmentos apropiados se fusionaron mediante ligación.

La selección de transformantes positivos se realizó en YPD-agar que contenía higromicina 100 µg mL^{-1}. Posteriormente, los clones resistentes a la higromicina se volvieron a sembrar en paralelo en agar YPD (sin higromicina para promover la pérdida de pKTAC-CRE_hygR), así como en agar YPD que contenía Geneticina 500 µg mL^{-1}. Las cepas que habían perdido su resistencia a la Geneticina se controlaron más mediante PCR y electroforesis en gel empleando los cebadores Det_bw y Det_fw (**tabla 1**) para confirmar la escisión del casete de resistencia KanMX del genoma. De las cepas correspondientes, CBS2612_L_d*FLO8* #2_4 fue elegidoa para su uso posterior.

b) Construcción de P. pastoris cepas que secretan fragmentos de anticuerpos scFv (scR) y vHH bajo el control transcripcional de P_{G1-3}

Los plásmidos de expresión pPM2d_pAOX_scR y pPM2d_pAOX_vHH son derivados del esqueleto del plásmido pPUZZLE (Stadlmayr et al. 2010). Se componen de un origen de replicación para *E. coli* (pUC19), un casete de

resistencia a Zeocin, el *S. cerevisiae* líder pre-pro del factor de apareamiento alfa, los genes de interés (vHH o scR) y el *S. cerevisiae CYC1* terminador de la transcripción, así como un locus para la integración en el genoma de *P. pastoris* (región 3'AOX1). Para su construcción, los genes que codifican scFv (scR) y vHH se optimizaron por codones mediante DNA2.0 y se obtuvieron como ADN sintético (la secuencia se indica a continuación). Un marcador His6 se fusionó en el extremo C con los genes para la detección. Después de la restricción, se digirió con *Xho*I y BamHI (para scR) o EcoRV (para vHH), cada gen se ligó en pPM2d_pAOX digerido con *Xho*I y BamHI o EcoRV. El reemplazo de P_{AOX1} por P_{G1-3} se realizó mediante digestión de restricción de estos plásmidos, así como del plásmido de expresión pPM1aZ10_pG1-3_eGFP (también un derivado de pPUZZLE descrito en el documento WO2017021541A1) con *siempre*NI y *Sbf*I y fusión de los fragmentos de plásmido apropiados. El fragmento derivado de pPM1aZ10_pG1-3_eGFP también contenía la secuencia para el terminador AOX que permitía la integración dirigida a este locus.

Los plásmidos de expresión resultantes pPM1aZ30_pG1-3_scR y pPM1aZ30_pG1-3_vHH se linealizaron con *Asc*I y transformado en CBS2612 (nombre corto wt), CBS2612_KanR_d*FLO8* #2 (nombre corto d*FLO8*) o CBS2612_L_d*FLO8* #2_4 (nombre corto d*FLO8*L) por electroporación usando un protocolo estándar como se describe en Gasser et al., 2013.

La selección de transformantes positivos se realizó en agar YPD selectivo que contenía de Zeocina 50 µg mL^{-1} y Geneticina 500 µg mL^{-1} (si procede).

c) Cribado de la productividad de fragmentos de anticuerpos

Para el cribado de expresión, se inocularon colonias individuales de los respectivos transformantes en 2 mL de medio líquido YP (por litro: 20 g de peptona, 10 g de extracto de levadura) que contenía Zeocina 25 µg ml^{-1} y Geneticina 500 µg mL^{-1} (si procede) y cultivo durante aprox. 24 h a 25 °C en 24-DWP a 280 rpm. Estos cultivos se usaron para inocular 2 mL de medio de cribado sintético ASMv6 (para conocer la composición, consulte el **Ejemplo 2)** que contiene 50 g L^{-1} y 1,5% de enzima liberadora de glucosa (lo que permite una velocidad de liberación de glucosa de aprox. 0,8 mg h^{-1} ml^{-1}; kit de desarrollo de medios m2p) a un OD_{600} inicial de 8. Las condiciones de cultivo fueron similares a las condiciones previas al cultivo. Después de 48 horas, se transfirió 1 mL de la suspensión celular a un tubo de centrifugación de 1,5 mL previamente pesado y se centrifugó a 16100 g durante 5 min a temperatura ambiente. Los sobrenadantes se transfirieron con cuidado a un vial nuevo y se almacenaron a -20 °C hasta su uso posterior. Los tubos de centrifugación que contenían los gránulos se pesaron nuevamente para determinar el peso celular húmedo (WCW). La cuantificación de la proteína secretada recombinante en el sobrenadante se realizó mediante electroforesis capilar microfluídica como se describe a continuación.

d) Cuantificación por electroforesis capilar microfluídica (mCE)

Se usó el 'Sistema LabChip GX/GXII' (PerkinElmer) para el análisis cuantitativo de la titulación de proteína secretada en sobrenadantes de cultivo. Se utilizaron los consumibles 'Protein Express Lab Chip' (760499, PerkinElmer) y 'Protein Express Reagent Kit' (CLS960008, PerkinElmer). La preparación de chips y muestras se realizó según las recomendaciones del fabricante. A continuación, se da una breve descripción del procedimiento.

Preparación del chip: después de que los reactivos alcanzaron la temperatura ambiente, se transfirieron 520 y 280 µl de matriz de gel Protein Express a filtros giratorios. Se añadieron 20 µl de disolución de colorante Protein Express a los 520 µl de matriz de gel que contenía el filtro giratorio. Después de agitar brevemente en vórtice el filtro giratorio que contiene colorante en la orientación invertida, ambos filtros giratorios se centrifugaron a 9300 g durante 10 minutos. Para lavar el chip, Se añadió 120 µL de agua Milli-Q® a todos los pocillos de chips activos y el chip se sometió al programa de lavado de instrumentos. Después de otras dos etapas de enjuague con agua Milli-Q®, los fluidos restantes se aspiraron por completo y se añadieron cantidades apropiadas de las soluciones de matriz de gel filtradas, así como la solución de marcador inferior Protein Express, a los pocillos de chip apropiados.

Preparación de muestra y escalera: Para la preparación de la muestra, se mezclaron 6 µL de muestra con 21 µL de tampón de muestra en una placa de 96 microtitulaciones. Las muestras se desnaturalizaron a 100 °C durante 5 minutos y se centrifugaron a 1200 g durante 2 minutos. Posteriormente, se añadió 105 µL de agua Milli-Q® . Las disoluciones de muestra se mezclaron brevemente mediante pipeteo y se centrifugaron de nuevo a 1200 g durante 2 minutos antes de la medición. Para preparar la escalera, se desnaturalizaron 12 µL de Proteina Express Ladder a 100 °C durante 5 minutos en un tubo de PCR. Posteriormente, se añadió 120 µL de agua Milli-Q® y la disolución de escalera se agitó brevemente antes de hacer girar el tubo durante 15 segundos en una minicentrífuga.

La cuantificación se realizó empleando el software LabChip proporcionado por el fabricante y la comparación con los estándares BSA.

La **Tabla 3** muestra que las titulaciones promedio de scR son 1.8 veces más altos en el sobrenadante de d*FLO8* cepas en comparación con el tipo salvaje, mientras que la concentración de biomasa no fue diferente, lo que llevó a rendimientos de scR 1,86 veces más altos en el sobrenadante de sepas d*FLO8*. Un aumento similar en las titulaciones promedio y el rendimiento de vHH (1,7 veces) en la mutación de d*FLO8* también se observó en el caso de las cepas que expresan vHH **(tabla 4).**

Tabla 3: WCW promedio, titulaciones de productos y rendimientos de una prueba de detección de 24 DWP de CBS2612 y CB2612_L_d*FLO8* #2_4 transformado con pPM1aZ30_pG1-3_scR. Se seleccionaron 20 clones por construcción.

Cepas	WCW ± DE	Titulación ± DE	Rendimiento ± DE
	[g L^{-1}]	[mg L^{-1}]	[mg g^{-1}]
wt + pG1-3_scR #1-20	80,8 ± 3,64	29,3 ± 8,99	0,36 ± 0,107
d*FLO8*L + pG1-3_scR #1-20	79,3 ± 2,40	53,0 ± 4,53	0,67 ± 0,057

Tabla 4: WCW promedio, titulaciones de productos y rendimientos de una prueba de detección de 24 DWP de CBS2612 y CB2612_KanR_d*FLO8* #2 transformado con pPM1aZ30_pG1-3_vHH. Se seleccionaron 20 clones por construcción.

Cepas	WCW ± DE	Titulación ± DE	Rendimiento ± DE
	[g L^{-1}]	[mg L^{-1}]	[mg g^{-1}]
wt + pG1-3_vHH #1-20	92,.4 ± 3,21	56,2 ± 32,70	0,61 ± 0,351
d*FLO8* + pG1-3_vHH #1-20	90,7 ± 4,48	97,0 ± 31,70	1,04 ± 0,360

A continuación, *FLO8* se interrumpió en cuatro clones diferentes que expresan vHH seleccionados del cribado anterior (CBS2612_pG1-3_vHH #4, #5, #13 y #15) usando el enfoque de casete de marcador dividido como se describe en el **Ejemplo 1.** Posteriormente, los clones mutantes de d*FLO8*, así como los los clones parentales de *FLO8* correspondientes se cribaron en cuanto a su productividad aplicando el régimen de selección 24-DWP. La **Tabla 5** muestra que las titulaciones de vHH y el rendimiento del producto fueron de 2 a 3 veces más altos en los clones de d*FLO8.* Para verificar que los niveles de producción aumentados se basaron en una expresión transcripcional más alta, los niveles de transcripción de vHH se cuantificaron mediante qPCR en diferentes puntos de tiempo. En promedio, se observaron niveles de expresión de vHH 2 veces más altos en los clones d*FLO8*, lo que indica que las titulaciones mejoradas de vHH se basan de hecho en una mayor actividad transcripcional de pG1-3 (véase el **Ejemplo 4).**

Tabla 5: WCW promedio y titulaciones de 4 cepas diferentes que expresan vHH, así como sus correspondientes depas d*FLO8*. Se cribaron cuatro réplicas de cada cepa parental, así como seis correspondientes cepas d*FLO8*. (FC: cambio de pliegue).

Cepas	WCW ± DE	Titulación ± DE	FC
	[g L^{-1}]	[mg L^{-1}]	
pG1-3_vHH#4	81,8 ± 1,25	81,4 ± 6,42	
pG1-3_vHH #4_d*FLO8* #1-6	78,9 ± 0,94	171,9 ± 3,90	2,11
pG1-3_vHH#5	86,6 ± 1,23	37,9 ± 2,22	
pG1-3_vHH #5_d*FLO8* #1-6	79,8 ± 1,51	91,2 ± 4,17	2,41
pG1-3_vHH#13	77,6 ± 1,81	33,6 ± 1,30	
pG1-3_vHH #13_d*FLO8* #1-6	78,9 ± 1,52	101,7 ± 2,05	3,03
pG1-3_vHH#15	80,1 ± 1,23	37,7 ± 2,05	
pG1-3_vHH#15_d*FLO8* #1-6	78,7 ± 3,70	101,1 ± 4,21	2,68

Ejemplo 4: Efecto de la interrupción de *FLO8* en la transcripción de vHH controlada por pG1-3

Para probar si la interrupción de *FLO8* conduce a una mayor actividad transcripcional de los genes bajo el control de pG1-3, 6 réplicas técnicas de CBS2612_pG1-3_vHH #4 y #13, así como las cepas mutantes d*FLO8* correspondiente CBS2612_pG1-3_vHH #4_d*FLO8*_1 y #13_d*FLO8*_2 **(Tabla 5; Ejemplo 3)** se cultivaron en el formato 24-DWP como se describe en el **Ejemplo 3.** Después de 2, 19 y 26 horas, se cosechó 1 ml de cultivo de 2 réplicas y se centrifugó durante 1 minuto a 16100 g y 4 °C y se descartó el sobrenadante. Posteriormente, el sedimento celular se resuspendió en 1 ml de reactivo TRI (Sigma-Aldrich) y se almacenó a -80 °C hasta su posterior procesamiento.

El aislamiento del ARN se realizó como se describe en el **Ejemplo 6.** Para eliminar el ADN residual, las muestras de ARN se trataron con el kit ADN libre™ (Ambion) según el manual del fabricante. Posteriormente, la calidad, la pureza y la concentración del ARN se analizaron mediante electroforesis en gel y análisis espectrofotométrico utilizando un NanoDrop 2000 (Thermo Scientific).

La síntesis de ADNc se realizó con el Biozym cDNA Synthesis Kit según el manual del fabricante. Brevemente, se añadieron 500 ng de ARN total a la mezcla maestra que contenía transcriptasa inversa, dNTP, inhibidor de ARNasa y tampón de síntesis. Para cebar oligo $d(T)_{23}$ se utilizó VN (NEB). La incubación de la mezcla de reacción se realizó durante 60 minutos a 55 °C. Posteriormente, se logró la inactivación de las enzimas mediante la incubación de la mezcla de reacción a 99°C durante 5 minutos.

Para la PCR cuantitativa en tiempo real (qPCR) se utilizaron cebadores específicos de vHH **(tabla 6).** La normalización se realizó comparando con niveles de expresión **(Tabla 6).** Para el análisis, se mezcló 1 µl de ADNc, agua y cebadores con SensiMix SYBR 2x Master Mix (Bioline) y se analizó en un ciclador de PCR en tiempo real (Rotor-Gene, Qiagen).

Tabla 6: Cebadores de PCR cuantitativos en tiempo real para el análisis de transcripciones de vHH

Cebador	**Secuencia**
vHH_fw	TGTAACGTGAATGTCGGATTTG (SEQ ID NO: 81)
vHH bw	TAGTGATGGTGGTGGTGATG (SEQ ID NO: 82)
Act1_fw	CCTGAGGCTTTGTTCCACCCACT (SEQ ID NO: 83)
Act1 bw	GGAACATAGTAGCAC CGGCATAACGA (SEQ ID NO: 84)

Todas las muestras se midieron por triplicado técnico. El análisis de datos se realizó con el software Rotor-Gene empleando el método de Cuantificación comparativa (QC).

La **Tabla 7** muestra que en los niveles promedio de transcrito de vHH de las cepas d*FLO8* aumentaron en todos los puntos de tiempo de cultivo analizados.

Tabla 7: Niveles relativos promedio de transcripción de vHH en diferentes puntos de tiempo de cultivo de detección

Cepa	**Transcripción relativa**		
	2 h	**19 h**	**26 h**
pG1-3_vHH #4	1,00*	1,56	1,65
pG1-3 vHH #4 dFLO8 1	1,76	2,14	2,87
pG1-3 vHH #13	1,00*	2,97	2,90
pG1-3_vHH #13_d*FLO8*_2	2,21	3,93	5,10
	*establecido en 1		

Ejemplo 5: Impacto de interrupción de *FLO8* de la productividad de fragmentos de anticuerpos secretados impulsados por pG1-3 en cultivos por lotes alimentados por biorreactores a escala de laboratorio

Antes de que se llevaran a cabo los cultivos por lotes alimentados, el casete del marcador de resistencia Geneticin se extirpó del genoma de la cepa CBS2612_ pG1-3_vHH #4_d*FLO8*_4 (nombre corto d*FLO8*_pG1-3_vHH #4_4) **(Tabla 5, Ejemplo 3)** por recombinación mediada por Cre como se describe en el **Ejemplo 3**. Productividad de tres réplicas de la cepa resultante CBS2612_pG1-3_vHH #4_ d*FLO8*_4 #L1 (nombre corto *L*_d*FLO8* _pG1-3_vHH #4_4_1) se comparó con los niveles de producto de tres réplicas de su cepa parental en el formato de cribado 24-DPW como se describe en el Ejemplo 3 (Tabla 8). Las productividades promedio de CBS2612_pG1-3_vHH #4_d*FLO8*_4 #L1 se mantuvo similar a su cepa parental (valor p de 0,096).

Tabla 8: WCW y titulación de CBS2612_d*FLO8*_pG1-3_vHH #4_4 y CBS2612_L_d*FLO8*_pG1-3_vHH #4_4_1 en la detección de 24 DWP.

Cepas	**WCW ± DE**	**Titulación ± DE**
	[g L^{-1}]	**[mg L^{-1}]**
d*FLO8*_pG1-3_vHH #4_4	80,87 ± 1,91	306,34 ± 12,85
*L*_d*FLO8*_pG1-3_vHH #4_4_1	80,40 ± 0,57	329,92 ± 8,48

Los cultivos por lotes alimentados se realizaron con la cepa CBS2612_ pG1-3_vHH #4 y el correspondiente mutante dFLO8 loxed CBS2612_L_d*FLO8*_pG1-3_vHH #4_4_1 en biorreactores de sobremesa de 1 L (SR0700ODLS; Dasgip, Alemania). Para los cultivos previocultivos previos, se inocularon 100 mL de medio YPG que contenía 50 µg mL de Zeocin en un matraz de agitación de 1 L con 1,0 mL de criostock y se incubaron durante aprox. 24 h a 180 rpm y 25 °C. Los cultivos por lotes se operaron a un volumen de trabajo de 0,25 L y se inocularon a una OD_{600} inicial de 1,5. La composición de los medios por lotes de glicerol se proporciona a continuación. Durante todo el proceso la temperatura se controló a 30°C, el OD se mantuvo al 30% mediante el ajuste automático de la velocidad del agitador (entre 400 y 1200 rpm) y el flujo de aire (entre 9,5 y 30 sL h^{-1}) y el pH se reguló a 5,0 mediante la adición automática de 12,5% NH_4OH. Después de un pico repentino en OD que indica el final del lote, se aplicó una alimentación de glucosa incremental lineal (composición de medios que se detalla a continuación) que resultó en velocidades de crecimiento iniciales rápidas seguidas de una fase extendida de disminución gradual de µ que se optimizó específicamente para $P_{G1\text{-}3}$-basado en q_P a la cinética µ (Prielhofer et al., 2018).

Medio de lote de glicerol contenido por litro:

2 g de Ácido cítrico monohidrato ($C_6H_8O_7$*H_2O), 45 g de Glicerol, 12,6 g de $(NH_4)_2HPO_4$, 0,5 g de $MgSO^{4}$*$7H_2O$, 0,9 g de KCl, 0,022 g de $CaCl_2$*$2H_2O$, 6,6 mLde disolución madre de biotina (0,2 g L^{-1}) y 4,6 ml de disolución madre de sales traza de PTM0 (descrito en el **Ejemplo 2**). Se añadió HCl (conc.) para ajustar el pH a 5.

Medios de alimentación de glucosa contenidos por litro:

495 g de glucosa monohidrato, 5,2 g de $MgSO_4$*$7H_2O$, 8,4 g de KCl, 0,28 g de $CaCl_2$*$2H_2O$, 11,8 mL de disolución madre de biotina (0,2 g L^{-1}) y 10,1 ml de disolución madre de sales traza de PTM0 (descrito en el **Ejemplo 2)**.

La YDM y la proteína recombinante secretada se analizaron en varios puntos de tiempo a lo largo del proceso. **(Tabla 9).** Para el análisis YDM, se transfirió 1 ml de caldo de cultivo a un tubo de centrifugación presecado (a 105 °C durante al menos 24 h) y prepesado de 2 ml. Después de la centrifugación a 16100 g durante 5 minutos, el sobrenadante se transfirió con cuidado a un vial nuevo y se almacenó a -20 °C hasta su uso posterior. Los sedimentos celulares se lavaron dos veces con agua desionizada y se secaron a 105 °C durante al menos 24 h antes de medir nuevamente el peso.

Los sobrenadantes se analizaron mediante electroforesis capilar microfluídica (GXII, Perkin-Elmer) como se describe en el **Ejemplo 3.**

Tabla 9: Titulaciones de YDM y vHH durante el cultivo por lotes alimentado por biorreactor de CBS2612 _pG1-3_vHH #4 y *L*_CBS2612_d*FLO8*_pG1-3_vHH #4_4_1

	pG1-3 vHH #4		***L*_d*FLO8*_pG1-3_vHH #4_4_1**	
Tiempo	**YDM ± DE**	**Titulación ± DE**	**YDM ± DE**	**Titulación ± DE**
[h]	**[g L^{-1}]**	**[mg L^{-1}]**	**[g L^{-1}]**	**[mg L^{-1}]**
0*	22.1 ± 0.06	-	22,8 ± 0,10	-
4	34.3 ± 0.21	31,6 ± 4,85	36,7 ± 0,18	48,8 ± 1,30
8	51.9 ± 0.10	98,7 ± 26,27	57,0 ± 0,99	167,5 ± 18,17
11	63.7 ± 0.50	158,2 ± 35,89	72,7 ± 1,26	275,4 ± 27,80
24	112,2 ± 0,25	203,4 ± 38,20	123,4 ± 1,28	817,2 ± 9,64
28	122,5 ± 0,49	214,0 ± 46,04	136,7 ± 0,71	991,7 ± 76,97
32	134,6 ± 0,64	177,0 ± 32,98	141,8 ± 0,72	1.006,7 ± 25,20
* inicio de alimentación de glucosa				

Desde la **Tabla 9** se puede observar que a lo largo del proceso, las titulaciones del producto para el d*FLO8* fueron consistentemente más altos que para la cepa de fondo wt. Al final del proceso se observó un aumento de 5,7 veces en la productividad. También en el cultivo por lotes alimentados, se observaron mayores niveles de transcripción de vHH en la cepa d*FLO8* durante todo el transcurso del tiempo.

Ejemplo 6: Efecto de la interrupción de *FLO8* en la fuerza de expresión y el comportamiento regulatorio de los promotores regulados por carbono libre de metanol

En la siguiente etapa, se estudió el impacto de la interrupción de *FLO8* en la fuerza de transcripción de otros promotores regulados por carbono (descritos en el documento WO2013050551 A1 y Prielhofer et al. 2013) en condiciones de inducción. Para análisis de transcriptoma en condiciones limitantes de glucosa, se utilizaron CBS7435 de tipo salvaje y CBS7436_KanR_d*FLO8* #2. Los cultivos previos y principales se cultivaron en placas de 24 pocillos profundos (24-DWP). Para el primer cultivo previo, 2 mL de YPD (por litro: 10 g de extracto de levadura, 20 g de peptona, 20 g de glucosa) que contiene geneticina 500 µg mL^{-1} (si procede) se inocularon con una sola colonia de CBS7435 y CBS7435_KanR_d*FLO8* #2, respectivamente, y crecido durante ca. 24 h a 25°C y 280 rpm. Para el segundo cultivo previo, se utilizaron 2 ml de medio M2(D) (composición descrita a continuación) inoculado a una OD_{600} inicial de 4. Se añadieron perlas de polímero liberador de glucosa (perlas de alimentación de 12 mm, Kuhner, CH), que liberan glucosa a una velocidad no lineal de 1,63 $t^{0.74}$ mg por disco (t = tiempo [h]) y los cultivos se incubaron durante aprox. 24 h a 25 °C y 280 rpm. Para los cultivos principales se utilizaron 2 ml de medio M2 (M2(D) sin glucosa). Los cultivos se agitaron a 280 rpm y 25 °C. La liberación lenta de glucosa aseguró un crecimiento limitado de glucosa. Las muestras se tomaron después de 3 h de cultivo principal (velocidad de crecimiento específica estimada: 0,1 h^{-1}), se mezcló inmediatamente en una proporción de 2:1 con una disolución fijadora preenfriada (5% [vol/vol] de fenol en etanol [absoluto]), se alicuotó en tubos sellados y se centrifugó a 16100 g durante 1 min. Los pellets se almacenaron a -80 °C hasta su uso posterior.

M2(D) contenido por litro: 22,0 g de glucosa monohidrato, 22,0 g de ácido cítrico monohidrato, 3,15 g de $(NH_4)_2HPO_4$, 0,49 g de $MgSO_4*7H_2O$, 0,80 g de KCl, 0,0268 g de $CaCl_2*2H_2O$, 1,47 ml de disolución madre de sales traza de PTM0 (descrito en el **Ejemplo 1)** y 0,4 mg de biotina; El pH se ajustó a 5 con KOH (sólido).

Para el aislamiento del ARN, se añadieron 1 ml de reactivo TRI (Sigma-Aldrich) y 500 µl de perlas de vidrio lavadas con ácido y las células se inerrumpieron en un FastPrep-24 (mpbio) a una velocidad de 5,5 m/s durante 40 segundos. Posteriormente se añadieron 200 µL de cloroformo. Posteriormente, las muestras se agitaron vigorosamente y después se dejaron reposar durante 5 - 10 minutos a temperatura ambiente. Después de centrifugar durante 10 minutos a 16100 g y 4 °C para promover la separación de fases, la fase acuosa incolora superior que contenía el ARN se transfirió a un tubo nuevo y se añadieron 500 µl de isopropanol para precipitar el ARN. Después de 10 minutos de incubación, las muestras se centrifugaron durante 10 minutos a 16100 g y 4 °C y se descartó el sobrenadante. El sedimento de ARN se lavó una vez con etanol al 70%, se secó al aire y se volvió a suspender en agua libre de ARNasa.

Para el análisis de transcriptomas, se utilizaron. matrices de oligonucleótidos específicos de P. pastoris de diseño interno (AMAD-ID 034821, matrices personalizadas de 8 × 15K; Agilent, EE. UU.) (Graf et al.; BMC Genómica. 2008;9:390). La síntesis de ARNc, la hibridación y el escaneo se llevaron a cabo de acuerdo con el protocolo Agilent para matrices de expresión de 2 colores. Las muestras se marcaron con Cy3 y Cy5 por triplicado y se hibridaron con un grupo de referencia generado a partir de células cultivadas en diversas condiciones de cultivo. Para todas las muestras, se llevaron a cabo experimentos de cambio de tinte.

las etapas de normalización y el análisis estadístico de los datos de micromatrices incluyeron la eliminación del sesgo de color mediante el suavizado de diagrama de dispersión MA ponderado localmente (LOESS), seguido de la normalización entre matrices mediante el método "Aquantile". Para identificar genes expresados diferencialmente y calcular los valores de p, se usó un ajuste de modelo lineal con una corrección de eBayes. Los valores de p se ajustaron para pruebas múltiples con el método de descubrimiento falso (FDR) por Benjamini & Yekutieli, 2001. Se considera que los genes con valores de p ajustados <0,05 tienen una expresión diferencial estadísticamente significativa. Para identificar genes expresados diferencialmente, se aplicó además un límite de cambio de pliegue de al menos 0,58 > log2 FC < -0,58.

Todas las etapas se realizaron utilizando el software R (Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. 2010. Bioinformatics. 26:139-40.) y el paquete limma. Los cambios de pliegue para una selección de genes regulados por carbono (genes bajo el control de promotores regulados por fuentes de carbono) se muestran en la **Tabla 10.** La expresión del gen *FLO8* nativo se reduce clara y significativamente en el mutante d*FLO8* en comparación con el tipo salvaje. Para todos los genes regulados por fuentes de carbono, se puede ver el aumento de los niveles de transcripción en el mutante d*FLO8*, que alcanza una fuerza de transcripción de 3,7 a 11,4 veces mayor en las condiciones de inducción. Todos estos genes tienen una transcripción más alta estadísticamente significativa (valores p ajustados <0.05) en la cepa mutante d*FLO8*. Esto indica fuertemente que el potencial para la producción de todos los promotores regulados por carbono libre de metanol como se describe en Prielhofer et al. 2013 y Prielhofer et al. 2018 se ve potenciado por la interrupción de *FLO8*. Por el contrario, la fuerza de expresión del promotor GAP no se ve afectada por la interrupción de *FLO8*.

Tabla 10: Efecto de d*FLO8* mutante en la fuerza de transcripción del gen *FLO8* y genes controlados bajo promotores regulados por carbono en condiciones inductoras de limitación de glucosa. Se muestran los cambios de pliegue (FC) entre la cepa mutante d*FLO8* en comparación con la cepa de tipo salvaje. Los cambios de expresión con un valor de p ajustado <0,05 muestran una diferencia estadísticamente significativa.

ID de gen GS115	**ID de gen CBS7435**	**secuencia promotora**	**Promotor*/ nombre del gen**	**FC expresión d*FLO8* vs wt**	**Valor p ajustado d*FLO8* vs wt**
PAS_chr1-3_0011	PP7435_Chr1-0007	SEQ ID NO:12	G1*	11,09	7,59E-06
PAS_chr4_0550	PP7435_Chr4-0424	SEQ ID NO:13	G3*	3,66	4,96E-06
PAS_chr4_0043	PP7435_Chr4-0972	SEQ ID NO:14	G4*	11,43	4,40E-07
PAS_chr2-1_0853	PP7435_Chr2-0787	SEQ ID NO:15	G6*	8,82	1,11E-07
PAS_chr2-1_0437	PP7435_Chr2-0858	SEQ ID NO:46	BRECHA*	1,08	0,786
PAS_chr4_0711	PP7435_Chr4-0252	-	FLO8	<0,29	8,61E-05
	*Nomenclatura de promotores tal como se describe en Prielhofer et al. 2013				

Ejemplo 7: Efecto de interrupción de *FLO8* de la fuerza de transcripción del promotor pAOX1 regulado por metanol

El impacto de la interrupción de *FLO8* en la fuerza de transcripción también se determinó para los promotores estándar de P. pastoris tales como pAOX1 y pGAP en condiciones inductoras de metanol. Por lo tanto, las cepas CBS7435 que expresan un fragmento Fab recombinante bajo el control de pAOX1 y la respectiva cepa mutante dFLO8 d*FLO8*#2 se cultivaron en cultivo por lotes alimentados a base de metanol.

Los lotes alimentados se llevaron a cabo en reactores DASGIP de 1,4 L (Eppendorf, Alemania) con un volumen de trabajo máximo de 1,0 L. La temperatura de cultivo se controló a 25 °C, el pH se controló a 5,0 mediante la adición de hidróxido de amonio al 25% y la concentración oxígeno disuelto se mantuvo por encima del 20% de saturación controlando la velocidad del agitador entre 400 y 1200 rpm, y el flujo de aire entre 24 y 72 sL/h.

El inóculo para el cultivo discontinuo se cultivó en matraces de agitación que contenían 100 ml de medio YP que contenía glicerol 20 g/l y zeocina 50 µg/ml, y se incubó a 28 °C y 180 rpm durante aproximadamente 24 horas. Los cultivos se usaron para inocular el volumen inicial de 0,4 L en el biorreactor a una densidad óptica inicial (600 nm) de 1,0. El lote se terminó después de aproximadamente 24 h y se administró la primera inyección de sal (10 ml).

A continuación, se alimentó una disolución discontinua con glicerol a una velocidad constante de 5 ml/h durante 5 horas. Después, se le dio al cultivo un pulso de metanol (2 g) y una inyección de sal (10 mL). Después de que el consumo de pulsos de metanol hubiera sido indicado por un aumento en la concentración de oxígeno disuelto en el cultivo, se inició una alimentación constante con una disolución discontinua con metanol con una velocidad de alimentación de 1,0 g/h. Se administran inyecciones de sal de 10 ml cada 10 g de biomasa recién formada, lo que corresponde a ~ 43 g de medio de alimentación de metanol. Con el aumento de las concentraciones de biomasa, la velocidad de alimentación de metanol se incrementó adecuadamente cuando se pudo descartar la acumulación de metanol debido a un aumento repentino del oxígeno disuelto en el cultivo al apagar la alimentación de metanol durante un período corto de tiempo. La velocidad final de alimentación de metanol fue de 2,5 g/h.

Se tomaron muestras con frecuencia. El cultivo se recogió después de aproximadamente 100 horas cuando las densidades celulares habían alcanzado más de 100 g/l de peso seco celular.

Los medios de comunicación fueron los siguientes:

Medio por lotes (*por litro*) *contenido:* 2,0 g de ácido cítrico, 12,4 g de $(NH_4)_2HPO_4$, 0,022 g de $CaCl_2 \cdot 2H_2O$, 0,9 g de KCl, 0,5 g de $MgSO_4 \cdot 7H_2O$, 40 g de glicerol, 4,6 ml de disolución madre de sales traza de PTM1. El pH se establece en 5,0 con HCl al 25%.

Disolución discontinua con glicerol (por litro) contenía: 623 g de glicerol, 12 ml de disolución madre de sales traza de PTM0 y 40 mg de biotina. La composición de PTM0 se da en el Ejemplo 2.

Disolución discontinua con metanol (*por litro*) *de metanol puro contenido:* 12 mL de disolución madre de sales traza de PTM0 y 40 mg de biotina.

Disolución de inyección de sal (*por litro*) *contenido:* 20,8 g de $MgSO_4 \cdot 7H_2O$, 41,6 de KCl, 1,04 g de $CaCl_2 \cdot 2H_2O$

La cuantificación de Fab intacto por ELISA se realizó usando anticuerpo IgG antihumano (ab7497, Abcam) como anticuerpo de recubrimiento y un anticuerpo IgG antihumano de cabra (específico de Fab) - anticuerpo conjugado con fosfatasa alcalina (Sigma A8542) como anticuerpo de detección. Fab/Kappa humana, fragmento de IgG (Betil P80-115) se usó como estándar con una concentración inicial de 100 ng/mL, las muestras de sobrenadante se diluyeron en consecuencia. La detección se realizó con pNPP (Sigma S0942). Los tampones de recubrimiento, dilución y lavado se basaron en PBS (KH_2PO_4 2 mM, $Na_2HPO_4.2\ H_2O$ 10 mM, KCl 2,7 mM, NaCl 8 mM, pH 7,4) y se completó con BSA (1% (p/v)) y/o Tween20 (0,1% (v/v)) según corresponda.

Con respecto a la titulación del producto, el aumento en la producción de Fab bajo el control de pAOX1 en el mutante d*FLO8* fue un máximo de 1,45 veces al final del cultivo por lotes alimentado con metanol (después de 119 h, como se describe en el documento WO2015/158800A1), mientras que la formación de biomasa se redujo ligeramente (en un 10-14%).

Las muestras de micromatrices se tomaron después de 53,5 h (25 h de alimentación de metanol) y se procesaron como se describe en el **Ejemplo 6.** Los cambios de pliegues para una selección de genes entre la cepa mutante d*FLO8* en comparación con la cepa de tipo salvaje se muestran en la **Tabla 11.** Otra vez, los niveles de transcripción de *FLO8* son significativamente más bajos en el mutante d*FLO8* en comparación con el tipo salvaje. Contrariamente a los genes y promotores regulados por carbono libre de metanol que se muestran en la **Tabla 10,** el pAOX1 inducible por metanol no muestra una fuerza de transcripción significativamente mayor en el mutante d*FLO8* en comparación con el tipo salvaje en sus condiciones completamente inducidas (ya que el valor de p ajustado es mayor a 0,05). Tampoco hay un efecto significativo sobre pGAP en las células cultivadas con metanol. Por lo tanto, la transcripción de estos promotores estándar en *P. pastoris* no se ve afectada por la infraexpresión de *FLO8*.

Tabla 11: Efecto de mutante d*FLO8* en la fuerza de transcripción del gen *FLO8* y el gen controlado bajo el promotor *AOX1* inducible por metanol en condiciones de lote alimentado induciendo metanol. Se muestran los cambios de pliegue (FC) entre la cepa mutante d*FLO8* en comparación con la cepa de tipo salvaje. Los cambios de expresión con un valor de p ajustado <0,05 muestran una diferencia estadísticamente significativa.

ID de gen GS115	ID de gen CBS7435	secuencia promotora	Promotor */nombre del gen	FC expresión d*FLO8* vs wt	Valor p ajustado d*FLO8* vs wt
PAS_chr4_0821	PP7435_Chr4-0130		AOX1*	1,36	0,186
PAS_chr2-1_0437	PP7435_Chr2-0858	SEQ ID NO:46	GAP*	1,24	0,400
PAS_chr4_0711	PP7435_Chr4-0252	-	FLO8	<0,3	3,71E-08

Ejemplo 8: Efecto de la interrupción de *FLO8* en la expresión de eGFP intracelular impulsada por promotores regulados por carbono libre de metanol

Como expresión génica nativa impulsada por la mayoría de los promotores regulados por carbono libre de metanol (documento WO2013050551A1; Prielhofer et al. 2013) fue regulado significativamente al alza bajo condiciones inductoras en la interrupción de la cepa de *dFLO8* (véase **Ejemplo 6**) de *FLO8* también se probó en las respectivas cepas eGFP-informadoras descritas en el documento WO2013050551A1 y Prielhofer et al (2013). Para cada promotor no inducible por metanol, se seleccionó una cepa que albergaba el vector de expresión pPM1aZ10_pG#_eGFP (CBS2612_pG3_eGFP #1; CBS2612_pG4_eGFP #6; CBS2612_pG6_eGFP #53; X33_pG7_eGFP #1; CBS2612_pG8_eGFP #8) y *FLO8* interrumpido empleando el método de casete de marcador dividido descrito en el **Ejemplo 1.** Los cultivos de detección en condiciones inductoras (limitadoras de glucosa) se realizaron como se describe en el **Ejemplo 2** con la excepción de que el polisacárido y la enzima liberadora de glucosa se obtuvieron de un proveedor diferente (EnPump 200, Enpresso). Como diferían las propiedades de la nueva enzima liberadora de glucosa, la concentración se adaptó al 0,4% correspondiente a una velocidad constante de liberación de glucosa de aprox. 0,6 mg ml^{-1} h^{-1}. En cada caso, dos réplicas parentales respectivas, así como se examinaron al menos 7 cepas *de dFLO8* correspondientes. La productividad de eGFP- se determinó como se describe en el **Ejemplo 2** con la excepción de que los valores no se normalizaron para el tamaño de celda. La **Tabla 12** muestra que la interrupción de *FLO8* conduce a un aumento en la fluorescencia de eGFP para cada promotor que alcanza 1,7 y 1,2 veces en el caso de pG3 y pG8, respectivamente, así como 2,3 veces en el caso de pG4 y pG7 y 2,2 veces en el caso de pG6. Esto subraya aún más el potencial de *FLO8* interrupción para la mejora de otros promotores regulados por carbono libre de metanol además de los derivados de pG1 y pG1.

Estos resultados confirman el aumento de la expresión de eGFP bajo el control de pG1 o un derivado de pG1 (pG1-3) en células huésped *dFLO8* en comparación con dicha expresión en células huésped de tipo salvaje. La **Tabla 2** muestra que la interrupción de *FLO8* conduce a un aumento en la fluorescencia de eGFP que es 3,8 (o 3,9) y 2,8 (o 2,9) veces en el caso de pG1 y pG1-3, respectivamente.

Tabla 12: Impacto de d*FLO8* sobre la expresión de eGFP bajo el control de promotores regulados por carbono libre de metanol. Se muestran los niveles de expresión de eGFP como porcentajes de la expresión de pGAP, así como los cambios de pliegue de las respectivas cepas d*FLO8*_pG#_eGFP comparadas con la cepa parental de tipo salvaje después de 48 h de cultivo en condiciones limitantes de glucosa (inducción).

Promotor	Célula huésped	Expresión de eGFP relativa a pGAP ± SD	FC *dFLO8*/*peso*
pGAP	wt	100 ± 2,2%	
pG3	wt	20 ± 0,7%	
pG3	*dFLO8*	33 ± 4,5%	1,7
pG4	wt	42 ± 1,1%	
pG4	*dFLO8*	96 ± 9,0%	2,3
pG6	wt	76 ± 0,0%	

Promotor	Célula huésped	Expresión de eGFP relativa a pGAP ± SD	FC *dFLO8*/*peso*
pG6	*dFLO8*	164 ± 1,4%	2,2
pG7	wt	697 ± 1,0%	
pG7	*dFLO8*	1591 ± 2%	2,3
pG8	wt	20 ± 3,1%	
pG8	*dFLO8*	24 ± 3,4%	1,2

Referencias:

Benjamini Y & Yekutieli D (2001) The Control of the False Discovery Rate in Multiple Testing under Dependency. The Annals of Statistics 29: 1165-1188.

Gasser B, Prielhofer R, Marx H, Maurer M, Nocon J, Steiger M, Puxbaum V, Sauer M & Mattanovich D (2013) Pichia pastoris: protein production host and model organism for biomedical research. Future Microbiol 8: 191-208.

Marx H, Mattanovich D & Sauer M (2008) Overexpression of the riboflavin biosynthetic pathway in Pichia pastoris. Microb Cell Fact 7: 23.

Prielhofer R, Reichinger M, Wagner N, Claes K, Kiziak C, Gasser B & Mattanovich D (2018) Superior protein titers in half the fermentation time: Promoter and process engineering for the glucose-regulated GTH1 promoter of Pichia pastoris. Biotechnol Bioeng.

Stadlmayr G, Mecklenbrauker A, Rothmuller M, Maurer M, Sauer M, Mattanovich D & Gasser B (2010) Identification and characterisation of novel Pichia pastoris promoters for heterologous protein production. J Biotechnol 150: 519-529.

Listado de secuencias

<110> Lonza S.L.

<120> PRODUCCIÓN DE PROTEÍNAS REGULADA POR FUENTE DE CARBONO EN UNA CÉLULA HUÉSPED RECOMBINANTE

<130> LO008P

<160> 86

<170> PatenteEn versión 3.5

<210> 1
<211> 806
<212> PRT
<213> Komagataella phaffii

<400> 1

```
Met Asn Lys Pro Asn Gly Ser Glu Gln Gln Pro Pro Ser Arg Gly Met
1               5                  10                  15

Lys Gln Glu Ser Gly Gly Pro Val Thr Ser Ser Thr Thr Pro Gly Thr
            20                  25                  30

Asn Thr Gly Leu Glu Asn Ser His Ser Met Gly Ala Asp Met Glu Pro
        35                  40                  45

Asp Val Gly Ala Thr Ser Pro Arg His Leu Leu Asn Gly Tyr Ile Tyr
    50                  55                  60

Asp Tyr Leu Val Lys Ser Asn Met Gln Asn Leu Ala Asp Gln Phe Ala
65                  70                  75                  80

Gln Glu Thr Glu Leu Leu Glu Thr Asp Leu Thr Val Pro Met Asp Thr
                85                  90                  95

Pro Ser Gly Tyr Leu Leu Glu Trp Trp Met Val Phe Trp Asp Leu Phe
            100                 105                 110

Asn Ala Arg Leu Lys Gln Arg Gly Ser Gln Lys Ala His Gln Tyr Ile
        115                 120                 125

Gln Leu Asn Met Leu Arg Gln Gln Gln Gln Arg Thr Met Arg Asn Thr
    130                 135                 140

Ala Arg Val Gln Lys Val Pro Leu Arg Pro His Thr Gln Ser Ser Pro
145                 150                 155                 160

Ser Met Ser Gln Thr Phe Ile Pro Gln Gln Pro Gln Gln Gln Ala Gln
                165                 170                 175
```

```
Gly Gln Gln His Ala Gln Ala Gln Ala Gln Val Gln Ala His Gln Gln
            180                 185                 190

Ala Gln His His Ala Gln Ala Gln Val Pro Val Gln Pro Gln Gln His
        195                 200                 205

Gln Leu Gly Gly Gln Thr Gln Gln Gln Gln Ser Ile Asn Thr Gly Ser
    210                 215                 220

Pro Ala Gly Pro Asn Ala Ile Asn Ser Arg Val Gln His Leu Ala Gln
225                 230                 235                 240

Gln Gln Met Asn His Leu Arg Gln Gln Ala Thr Ala Thr Thr Gln Gln
                245                 250                 255

Pro Ile Pro Gln Gln Asn Ile Pro Ser Asn Gln Gln Gly Pro Thr Gly
            260                 265                 270

Pro Tyr Pro Thr Ser Pro Ser Arg Arg Pro Arg Leu Leu Ser Asn Glu
        275                 280                 285

Ser Gly Ala Ser Ala Pro Ser Val Met Thr Lys Ser Gln Leu Gln Gly
    290                 295                 300

Val Pro Pro Ser Gln Gln Pro His Gln Gln Gln Gly Gln Gln Val Gly
305                 310                 315                 320

Pro Pro Asn Gln His Gln Gly Gln Ser Ser Ser Phe Tyr Ser Gly Met
                325                 330                 335

Pro Pro Gln Gly Val Val Val Pro His Gln Phe Asn Pro Gln Gln Tyr
            340                 345                 350

Ala Asn Met Leu Ala Arg Gln Gln His Val Gln Ala Gln Gln Gln Val
        355                 360                 365

Gln Leu Gln Gln Val Gln His Val Gln Gln Arg Gln Gln Gln Asp Gln
    370                 375                 380

Gln Gln His Arg Leu Ser Ala Gly Ser Pro Gly His Pro Ser Phe Gly
385                 390                 395                 400

Val Phe Gln Gln Pro Pro Pro Met Ser Asn His Asn Gln Val Met Ile
                405                 410                 415

Asn Gln Gln Gly Glu Thr Phe Phe Asp Pro His Ser Pro Tyr Ala Gln
            420                 425                 430
```

```
Pro Asn Gly Tyr Pro Gln Pro Gln Gln Gln Gln Gln Gln Gln Gln
        435                 440                 445

Gln Gln Gln Gln Gln Gln Pro Gln Gln Gln Gln Gln Gln Gln Gln
    450                 455                 460

Gln Lys Gln Gln Pro Pro Pro Pro Pro Arg Gln Pro Gln Arg Gln Gln
465                 470                 475                 480

Ala Met Ala Met Ala Pro Leu Pro His Ser Thr Ser Ala Ala Gly Thr
                485                 490                 495

Pro His Ser Ser Thr Thr Pro Arg Phe Ser Gln Pro Gly Pro Val Tyr
            500                 505                 510

Gln Gln Pro Leu Pro Ala Ser Gln Pro Gln His Ser Pro Pro Ser Ser
        515                 520                 525

Ile Gln Gln Pro Glu Leu Val Pro Thr Pro Gly Ser Gln His Gln Gln
    530                 535                 540

Ile Ala Gln Pro Gln Ser Gln Ser Gln His Gln Gln Ser Gln Gln Ser
545                 550                 555                 560

Gln Ser Ser Ala Ser Lys Ile Val Gly Ile Gln Glu Tyr Gln Lys Glu
                565                 570                 575

Leu Met Met Leu Glu Lys Gln Asn Lys Gln Arg His Asp Met Ala Cys
            580                 585                 590

Lys Lys Gly Ser Gly His Phe Ser Asn Phe Asp Pro Ile Pro Glu His
        595                 600                 605

Thr Pro Pro Glu Pro Lys Phe Asn Val Asn Val Met Leu Pro Pro Gln
    610                 615                 620

Asn Ser Ala Val Val Thr Lys Asn Thr Pro Gly Thr Ser Pro Gly Thr
625                 630                 635                 640

Gln Thr Gln Asn Thr Ala His Ser Thr Gly Asn Thr Ser Ala Gly Ser
                645                 650                 655

Thr Pro Asn Asn Val Ala Pro Val Arg Lys Lys Lys Glu Pro Ala Lys
            660                 665                 670

Lys Lys Ala Lys Lys Ala Thr Glu Pro Pro Thr Pro Thr Thr Pro Gln
```

```
        675                 680                 685

Thr Pro Ile Ala Ala Arg Thr His Gln Asn Ser Thr Gly Gly Ile Pro
    690                 695                 700

Gly Asn Asn Ala Ala Thr Lys Arg Arg Lys Arg Glu Pro Leu Val Asp
705                 710                 715                 720

Gln Thr Val Ser Pro Asn Leu Asn Glu Ala Ser Lys Ser Thr Lys Thr
                725                 730                 735

Gly Lys Ile Ser Ser Gln Thr Asp Phe Thr Gly Ser Asp Asn Gly Phe
            740                 745                 750

Leu Gln Asp Phe Gly Asp Gly Thr Gly Pro Pro Thr Gly Thr Asp Asp
        755                 760                 765

Met Glu Phe Asp Phe Asn Ser Phe Leu Asn Asn Glu Thr Gly Glu Pro
    770                 775                 780

Asn Ser Ser Thr Ile His Phe Asp Asn Val Phe Asn Trp Gly Glu Gly
785                 790                 795                 800

Thr Glu Ala Gly Asp Leu
                805
```

<210> 2
<211> 2421
<212> ADN
<213> Komagataella phaffii

<400> 2

atgaacaagc caaacgggtc tgaacaacaa ccaccgtcac gcggaatgaa gcaagagtca 60

ggaggcccag ttacttcatc tacgacgccg ggtaccaata ctggcctaga aaactctcat 120

tccatggggg cggatatgga gcctgatgtt ggtgctacct ctcctcgcca tcttcttaat 180

gggtacattt acgattattt agtcaaatct aacatgcaaa atttggctga tcaatttgcc 240

caagagacgg agctcttaga aacagacttg acagtaccaa tggatacgcc ttcaggctat 300

cttctagaat ggtggatggt attctgggac cttttcaatg cccgcctaaa gcaacggggt 360

tcacagaagg cccaccagta tattcagttg aacatgctac gacaacagca acagaggacc 420

atgcgaaata cagcccgtgt tcaaaaagtc ccgttgaggc cacacaccca atcatctcct 480

tcaatgtcac agacttttat tccacagcag cctcaacagc aagcacaggg acaacagcac 540

gcccaggctc aagcccaagt gcaagcacat cagcaagccc aacaccacgc gcaggcacaa 600

gtgccagtgc aaccgcaaca gcaccagcta ggaggccaaa ctcaacagca gcaatccatt 660

aacactgggt ctcctgcggg tccaaatgct atcaactcgc gtgttcaaca cttagcacaa 720

caacagatga atcaccttcg ccagcaggcg actgccacta cgcaacaacc tatcccgcaa 780

cagaatatcc catcaaacca acagggtcct acaggccctt atcctacttc cccttcaaga 840

agaccgagat tactgtctaa cgaatcgggt gcaagtgcac cctctgtaat gacaaagtca 900

cagctccaag gagtccctcc ctcacaacaa ccacaccagc agcaaggtca gcaggtaggc 960

ccccctaatc aacatcaagg tcaatcttct tccttttatt cgggcatgcc tcctcaaggg 1020

gtcgtggttc ctcatcagtt caatcctcag cagtatgcca atatgctagc aagacaacag 1080

catgtacaag ctcaacaaca ggttcagtta caacaggtcc aacatgtaca acagagacaa 1140

cagcaagacc aacaacaaca ccgcctgtcc gccggttcac cggggcaccc ttcatttggc 1200

gtttttcaac aacctcctcc gatgtcaaac cataatcagg tcatgatcaa tcagcaggga 1260

gaaacttttt ttgatccaca ttctccatat gctcaaccta acgggtaccc ccagccacag 1320

caacaacaac aacaacagca acaacaacaa caacagcagc aaccgcaaca gcagcagcag 1380

cagcagcagc aacagaagca gcaaccacca ccaccaccac gacagcctca gcgccaacaa 1440

gcgatggcca tggctcctct gcctcactct acttctgccg ccggtactcc tcactcgtcc 1500

accacaccta gattctcgca acctggtcct gtttatcagc agcctttacc tgcatctcaa 1560

ccgcaacatt ctccgccttc ttctattcag cagccggagc tagttccaac tccagggtca 1620

caacatcagc aaatagcaca accacaatca cagagccaac accagcaatc gcaacagtct 1680

caatcaagtg cttctaaaat tgtaggtata caggagtatc agaaagagct aatgatgctt 1740

gagaaacaga acaaacagcg tcatgacatg gcatgtaaga agggaagcgg gcatttttct 1800

aactttgatc caattccaga gcacacaccg cccgaaccaa aatttaatgt gaatgtaatg 1860

ctccctcccc agaactctgc agtggtcacg aagaatactc ccggaacttc acctggtaca 1920

caaactcaaa acactgcaca tagtactggt aacacttctg cggggtctac accaaataat 1980

gtcgcacctg tacgaaagaa aaaggagcca gctaaaaaga aggcaaagaa agctactgag 2040

cccccgactc ccactactcc acagactcca attgcagcta ggacacatca aaactctaca 2100

ggcggcattc ctggtaataa tgctgctact aagcgacgaa aacgggagcc gctggttgat 2160

caaactgttt cacctaacct taacgaagct tccaagtcaa caaagaccgg aaaaatttca 2220

tctcaaactg actttacagg ttctgacaat ggattcttac aggattttgg cgatggaact 2280

ggtcctccca ctggaaccga tgatatggaa tttgatttta acagttttct taataacgaa 2340

actggcgaac ctaatagttc aaccattcat tttgacaatg tattcaattg gggagaaggt 2400

accgaagccg gagatttata g 2421

<210> 3
<211> 804
<212> PRT
<213> Komagataella pastoris

<400> 3

```
Met Asn Lys Pro Asn Gly Ser Gly Gln Gln Pro Pro Ser Arg Gly Met
1               5                  10                  15

Lys Gln Asp Pro Gly Gly Pro Val Thr Ser Ser Thr Thr Pro Gly Thr
            20                  25                  30

Asn Thr Gly Phe Glu Asn Ser His Ser Met Gly Ala Asp Val Glu Pro
        35                  40                  45

Asp Val Gly Ala Ala Ser Pro Arg His Ile Leu Asn Gly Tyr Ile His
    50                  55                  60

Asp Tyr Leu Val Lys Ser Asn Met Gln Asn Leu Ala Asp Gln Phe Ala
65                  70                  75                  80

Gln Glu Ser Asp Leu Leu Glu Thr Asp Leu Thr Val Pro Met Asp Thr
                85                  90                  95

Pro Thr Gly Tyr Leu Leu Glu Trp Trp Met Val Phe Trp Asp Leu Phe
            100                 105                 110

Asn Ala Arg Leu Lys Gln Arg Gly Ser Gln Lys Ala His Gln Tyr Ile
        115                 120                 125

Gln Leu Asn Met Leu Arg Gln Gln Gln Gln Arg Thr Met Arg Asn Thr
    130                 135                 140

Ala Arg Val Gln Lys Val Pro Leu Arg Pro His Thr Gln Ser Ser Pro
145                 150                 155                 160

Ser Met Ser Gln Thr Phe Ile Pro Gln Gln Pro Gln Gln Gln Ala Gln
                165                 170                 175

Ala Gln Ala Gln Gln His Ala Gln Ala Gln Ala Gln Val Gln Ala His
            180                 185                 190

Gln Gln Ala Gln His His Ala Gln Ala Gln Val Pro Met Gln Ser Gln
        195                 200                 205

Pro His Gln Gln Gly Gly Gln Thr Gln Gln Gln Gln Pro Ile Asn Thr
    210                 215                 220

Gly Ser Pro Ala Gly Pro Asn Ala Ile Asn Ser Arg Val Gln His Leu
225                 230                 235                 240
```

```
Ala Gln Gln Gln Met Asn His Leu Arg Gln Gln Ala Thr Ala Thr Thr
                245                 250                 255

Gln Gln Pro Ile Pro Gln Gln Asn Ile Pro Ser Asn Gln Gln Gly Pro
            260                 265                 270

Ala Gly Pro Tyr Pro Thr Ser Pro Ser Arg Arg Pro Arg Leu Leu Ser
        275                 280                 285

Asn Glu Ser Gly Ala Ser Ala Pro Ser Val Met Thr Lys Ser Gln Leu
    290                 295                 300

Gln Gly Gly Pro Pro Ser Gln Gln Pro His Gln Gln Gln Ala Gln Gln
305                 310                 315                 320

Val Gly Pro Pro Asn Gln His Gln Gly Gln Ser Ser Ser Phe Tyr Ser
                325                 330                 335

Gly Met Pro Pro Gln Gly Val Val Val Pro His Gln Phe Asn Pro Gln
            340                 345                 350

Gln Tyr Ala Asn Met Leu Ala Arg Gln Gln His Val Gln Ala Gln Gln
        355                 360                 365

Gln Val Gln Leu Gln Gln Val Pro His Val Gln Gln Arg Gln Gln Gln
    370                 375                 380

Asp Gln Gln Gln His Arg Leu Ser Ala Gly Ser Pro Gly His Pro Ser
385                 390                 395                 400

Phe Gly Val Phe Gln Gln Pro Pro Pro Met Ser Asn His Asn Gln Val
                405                 410                 415

Met Ile Asn Gln Gln Gly Glu Thr Phe Phe Asp Pro His Ser Pro Tyr
            420                 425                 430

Ala Gln Pro Asn Gly Tyr Pro Gln Pro Gln Gln Gln Gln Gln Gln Gln
        435                 440                 445

Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln
    450                 455                 460

Lys Gln Gln Gln Pro Pro Pro Arg Gln Pro Gln Arg Gln Gln Ala Met
465                 470                 475                 480

Ala Thr Ala Pro Leu Pro His Ser Thr Ser Ala Ser Gly Thr Pro His
```

```
                485                 490                 495

Thr Ala Thr Thr Pro Arg Phe Ser Gln Pro Gly Pro Val Tyr Gln Gln
            500                 505                 510

Pro Leu Pro Ala Ser Gln Pro Gln His Ser Pro Pro Thr Ser Ile Gln
        515                 520                 525

Gln Gln Glu Pro Ile Pro Thr Pro Gly Ser Gln His Gln Gln Ile Ala
    530                 535                 540

Gln Pro Gln Ser Gln Asn Gln His Gln Gln Pro Gln Gln Pro Gln Ala
545                 550                 555                 560

Ser Ala Ser Lys Met Val Gly Ile Gln Glu Tyr Gln Lys Glu Leu Met
                565                 570                 575

Met Leu Glu Lys Gln Asn Lys Gln Arg His Asp Met Ala Ile Lys Lys
            580                 585                 590

Gly Ser Gly His Phe Ser Asn Phe Asp Pro Ile Pro Glu His Thr Gln
        595                 600                 605

Thr Glu Pro Lys Phe Asn Val Asn Val Met Leu Pro Pro Gln Asn Ser
    610                 615                 620

Ala Val Ala Thr Lys Asn Thr Pro Gly Thr Ser Pro Gly Thr Gln Thr
625                 630                 635                 640

Gln Asn Thr Ala His Ser Thr Gly Asn Thr Ser Ala Gly Ser Thr Pro
                645                 650                 655

Asn Asn Val Val Pro Val Arg Lys Lys Lys Glu Pro Ser Lys Lys Lys
            660                 665                 670

Ser Lys Lys Ala Thr Glu Pro Pro Thr Pro Thr Thr Pro Gln Thr Pro
        675                 680                 685

Ile Ala Ala Arg Ala His Gln Asn Ser Thr Gly Gly Ile Ser Gly Asn
    690                 695                 700

Asn Ala Ala Thr Lys Arg Arg Lys Arg Glu Pro Leu Val Asp Gln Thr
705                 710                 715                 720

Val Ser Pro Asn Leu Asn Glu Ala Ser Lys Ser Thr Lys Pro Gly Lys
                725                 730                 735
```

```
Ile Ser Ser Gln Asn Asp Phe Thr Gly Ser Asp Asn Gly Phe Leu Gln
            740                 745                 750

Asp Phe Gly Asp Gly Thr Gly Pro Pro Thr Gly Thr Asp Asp Met Glu
        755                 760                 765

Phe Asp Phe Asn Ser Phe Leu Asn Asn Glu Thr Gly Glu Pro Asn Ser
    770                 775                 780

Ser Thr Ile His Phe Asp Asn Val Phe Asn Trp Gly Glu Gly Thr Glu
785                 790                 795                 800

Ala Gly Asp Leu
```

<210> 4
<211> 2415
<212> ADN
<213> Komagataella pastoris

<400> 4

atgaacaagc caaacgggtc tggacaacaa ccaccgtcac gcggaatgaa gcaagaccca 60

ggaggcccag ttacctcatc tacgacaccc ggtaccaata ctggctttga aaactctcat 120

tccatgggag cggatgtgga gcctgatgtt ggtgccgcct ctcctcgcca tattcttaat 180

gggtacattc atgattactt agtcaaatct aatatgcaaa atttggccga tcaatttgct 240

caagagtcgg atctcttaga aacagaccta acggtaccaa tggatacacc tacaggctat 300

cttttagagt ggtggatggt attttgggac cttttcaatg cccgcctaaa gcaacgaggt 360

tcacagaagg ctcatcagta tattcagttg aacatgctac gacaacaaca gcagaggact 420

atgcgaaata cagcccgtgt tcagaaagtc ccgttgagac cacacaccca atcatctcct 480

tcaatgtcac agacctttat tccacagcag cctcaacagc aagcacaggc acaggcacag 540

cagcacgccc aagctcaagc gcaagtccag gctcatcagc aagcacagca tcatgcgcag 600

gctcaagtgc caatgcaatc gcaaccacat cagcaaggag gccaaactca acaacagcaa 660

cccattaaca ctgggtctcc tgcggggcca aatgctatca actctcgtgt gcaacactta 720

gcgcaacaac agatgaatca tcttcgccag caggcaactg ccactaccca gcaacctatc 780

ccgcaacaga atattccatc aaaccagcag ggccctgcgg gcccttatcc tacatcgcct 840

tcaagaagac cgagattact gtctaacgaa tcgggtgcaa gtgcaccctc tgtaatgacg 900

aagtcacagc tccaaggagg tcctccatca caacaaccac accaacagca agctcagcag 960

gtaggacccc ccaatcaaca tcaaggccag tcctcttcct tttattcggg catgcctcct 1020

caaggagttg tggttcctca tcagttcaat cctcagcagt atgccaatat gctagcaaga 1080

caacagcatg tgcaagctca acaacaggtt cagttacagc aggttccaca tgtgcaacaa 1140

agacaacagc aagaccaaca acaacaccgc ttgtcagccg gttcaccagg gcatccttca 1200

tttggcgttt ttcaacaacc tcctccgatg tcaaaccata atcaggtcat gatcaaccag 1260

cagggagaaa ccttttttga tcctcattca ccatatgctc aacctaatgg gtacccccag 1320

ccacagcaac aacaacaaca gcagcaacaa caacaacaac aacaacaaca gcagcagcag 1380

cagcaacaac agaagcagca acagccacca ccaagacagc ctcagcgcca acaagcgatg 1440

gctacggctc ctttgcctca ttctacttct gcctcgggta ctcctcacac ggccaccaca 1500

cctagattct cccagcctgg tcctgtttat cagcagcctt tacctgcatc tcaaccgcaa 1560

cattctccgc ctacttctat tcagcaacag gaaccaattc caactcctgg gtcacaacat 1620

cagcaaatag cacaaccgca atcacagaat caacaccagc aaccgcagca acctcaagca 1680

agtgcttcta aaatggtggg tatacaggag taccagaaag agttaatgat gcttgagaag 1740

cagaacaaac aacgtcatga catggcaatt aagaagggaa gcggacattt ctctaatttt 1800

gatccaattc cagagcacac acagactgaa ccaaaattca atgtgaacgt aatgctccct 1860

ccccagaact ctgcagtggc cacgaagaat actcctggaa cttctcctgg tacgcaaact 1920

caaaacactg cacacagtac tggcaacact tctgctgggt ctacaccgaa taatgttgta 1980

ccagttcgaa aaaagaagga gccatctaaa aagaaatcaa agaaggctac cgaacctcca 2040

actcctacta ctccacagac accaattgca gctagggcac atcaaaactc tacgggcggc 2100

atttcaggta ataatgctgc tactaagcga cgaaaaaggg agccgctggt cgaccaaaca 2160

gtttcaccta accttaacga agcttcgaaa tcaacaaagc ctgggaaaat ctcatctcaa 2220

aatgacttta caggttctga caatggattt ttacaggatt ttggcgatgg aacaggtccc 2280

cccactggaa ctgacgatat ggagtttgat tttaacagtt ttctcaacaa cgaaactggt 2340

gaacctaata gttcaaccat tcattttgac aatgtattta attgggggga gggtactgaa 2400

gccggggatc tatag 2415

<210> 5
<211> 799
<212> PRT
<213> Saccharomyces cerevisiae

<400> 5

```
Met Ser Tyr Lys Val Asn Ser Ser Tyr Pro Asp Ser Ile Pro Pro Thr
1               5                  10                  15

Glu Gln Pro Tyr Met Ala Ser Gln Tyr Lys Gln Asp Leu Gln Ser Asn
            20                  25                  30

Ile Ala Met Ala Thr Asn Ser Glu Gln Gln Arg Gln Gln Gln Gln Gln
        35                  40                  45
Gln Gln Gln Gln Gln Gln Gln Trp Ile Asn Gln Pro Thr Ala Glu Asn
    50                  55                  60

Ser Asp Leu Lys Glu Lys Met Asn Cys Lys Asn Thr Leu Asn Glu Tyr
65                  70                  75                  80

Ile Phe Asp Phe Leu Thr Lys Ser Ser Leu Lys Asn Thr Ala Ala Ala
                85                  90                  95

Phe Ala Gln Asp Ala His Leu Asp Arg Asp Lys Gly Gln Asn Pro Val
            100                 105                 110

Asp Gly Pro Lys Ser Lys Glu Asn Asn Gly Asn Gln Asn Thr Phe Ser
        115                 120                 125

Lys Val Val Asp Thr Pro Gln Gly Phe Leu Tyr Glu Trp Trp Gln Ile
    130                 135                 140

Phe Trp Asp Ile Phe Asn Thr Ser Ser Ser Arg Gly Gly Ser Glu Phe
145                 150                 155                 160

Ala Gln Gln Tyr Tyr Gln Leu Val Leu Gln Glu Gln Arg Gln Glu Gln
                165                 170                 175

Ile Tyr Arg Ser Leu Ala Val His Ala Ala Arg Leu Gln His Asp Ala
            180                 185                 190

Glu Arg Arg Gly Glu Tyr Ser Asn Glu Asp Ile Asp Pro Met His Leu
        195                 200                 205

Ala Ala Met Met Leu Gly Asn Pro Met Ala Pro Ala Val Gln Met Arg
    210                 215                 220

Asn Val Asn Met Asn Pro Ile Pro Ile Pro Met Val Gly Asn Pro Ile
225                 230                 235                 240

Val Asn Asn Phe Ser Ile Pro Pro Tyr Asn Asn Ala Asn Pro Thr Thr
                245                 250                 255

Gly Ala Thr Ala Val Ala Pro Thr Ala Pro Pro Ser Gly Asp Phe Thr
            260                 265                 270

Asn Val Gly Pro Thr Gln Asn Arg Ser Gln Asn Val Thr Gly Trp Pro
        275                 280                 285

Val Tyr Asn Tyr Pro Met Gln Pro Thr Thr Glu Asn Pro Val Gly Asn
```

```
    290                 295                 300

Pro Cys Asn Asn Asn Thr Thr Asn Asn Thr Thr Asn Asn Lys Ser Pro
305                 310                 315                 320

Val Asn Gln Pro Lys Ser Leu Lys Thr Met His Ser Thr Asp Lys Pro
                325                 330                 335

Asn Asn Val Pro Thr Ser Lys Ser Thr Arg Ser Arg Ser Ala Thr Ser
            340                 345                 350

Lys Ala Lys Gly Lys Val Lys Ala Gly Leu Val Ala Lys Arg Arg Arg
        355                 360                 365

Lys Asn Asn Thr Ala Thr Val Ser Ala Gly Ser Thr Asn Ala Cys Ser
    370                 375                 380

Pro Asn Ile Thr Thr Pro Gly Ser Thr Thr Ser Glu Pro Ala Met Val
385                 390                 395                 400

Gly Ser Arg Val Asn Lys Thr Pro Arg Ser Asp Ile Ala Thr Asn Phe
                405                 410                 415

Arg Asn Gln Ala Ile Ile Phe Gly Glu Glu Asp Ile Tyr Ser Asn Ser
            420                 425                 430

Lys Ser Ser Pro Ser Leu Asp Gly Ala Ser Pro Ser Ala Leu Ala Ser
        435                 440                 445

Lys Gln Pro Thr Lys Val Arg Lys Asn Thr Lys Lys Ala Ser Thr Ser
    450                 455                 460

Ala Phe Pro Val Glu Ser Thr Asn Lys Leu Gly Gly Asn Ser Val Val
465                 470                 475                 480

Thr Gly Lys Lys Arg Ser Pro Pro Asn Thr Arg Val Ser Arg Arg Lys
                485                 490                 495

Ser Thr Pro Ser Val Ile Leu Asn Ala Asp Ala Thr Lys Asp Glu Asn
            500                 505                 510

Asn Met Leu Arg Thr Phe Ser Asn Thr Ile Ala Pro Asn Ile His Ser
        515                 520                 525

Ala Pro Pro Thr Lys Thr Ala Asn Ser Leu Pro Phe Pro Gly Ile Asn
    530                 535                 540
```

```
Leu Gly Ser Phe Asn Lys Pro Ala Val Ser Ser Pro Leu Ser Ser Val
545                 550                 555                 560

Thr Glu Ser Cys Phe Asp Pro Glu Ser Gly Lys Ile Ala Gly Lys Asn
                565                 570                 575

Gly Pro Lys Arg Ala Val Asn Ser Lys Val Ser Ala Ser Ser Pro Leu
            580                 585                 590

Ser Ile Ala Thr Pro Arg Ser Gly Asp Ala Gln Lys Gln Arg Ser Ser
        595                 600                 605

Lys Val Pro Gly Asn Val Val Ile Lys Pro Pro His Gly Phe Ser Thr
    610                 615                 620

Thr Asn Leu Asn Ile Thr Leu Lys Asn Ser Lys Ile Ile Thr Ser Gln
625                 630                 635                 640

Asn Asn Thr Val Ser Gln Glu Leu Pro Asn Gly Gly Asn Ile Leu Glu
                645                 650                 655

Ala Gln Val Gly Asn Asp Ser Arg Ser Ser Lys Gly Asn Arg Asn Thr
            660                 665                 670

Leu Ser Thr Pro Glu Glu Lys Lys Pro Ser Ser Asn Asn Gln Gly Tyr
        675                 680                 685

Asp Phe Asp Ala Leu Lys Asn Ser Ser Ser Leu Leu Phe Pro Asn Gln
    690                 695                 700

Ala Tyr Ala Ser Asn Asn Arg Thr Pro Asn Glu Asn Ser Asn Val Ala
705                 710                 715                 720

Asp Glu Thr Ser Ala Ser Thr Asn Ser Gly Asp Asn Asp Asn Thr Leu
                725                 730                 735

Ile Gln Pro Ser Ser Asn Val Gly Thr Thr Leu Gly Pro Gln Gln Thr
            740                 745                 750

Ser Thr Asn Glu Asn Gln Asn Val His Ser Gln Asn Leu Lys Phe Gly
        755                 760                 765

Asn Ile Gly Met Val Glu Asp Gln Gly Pro Asp Tyr Asp Leu Asn Leu
    770                 775                 780

Leu Asp Thr Asn Glu Asn Asp Phe Asn Phe Ile Asn Trp Glu Gly
785                 790                 795
```

<210> 6
<211> 791
<212> PRT
<213> Saccharomyces cerevisiae

<400> 6

```
Met Ser Tyr Lys Val Asn Ser Ser Tyr Pro Asp Ser Ile Pro Pro Thr
1               5                  10                  15

Glu Gln Pro Tyr Met Ala Ser Gln Tyr Lys Gln Asp Leu Gln Ser Asn
            20                  25                  30

Ile Ala Met Ala Thr Asn Ser Glu Gln Gln Arg Gln Gln Gln Gln Trp
        35                  40                  45

Ile Asn Gln Pro Thr Ala Glu Asn Ser Asp Leu Lys Glu Lys Met Asn
    50                  55                  60

Cys Lys Asn Thr Leu Asn Glu Tyr Ile Phe Asp Phe Leu Thr Lys Ser
65                  70                  75                  80

Ser Leu Lys Asn Thr Ala Ala Ala Phe Ala Gln Asp Ala His Leu Asp
                85                  90                  95

Arg Asp Lys Gly Gln Asn Pro Ile Asp Gly Pro Lys Ser Lys Glu Asn
            100                 105                 110

Asn Gly Asn Gln Asn Thr Phe Ser Lys Val Val Asp Thr Pro Gln Gly
        115                 120                 125

Phe Leu Tyr Glu Trp Trp Gln Ile Phe Trp Asp Ile Phe Asn Thr Ser
    130                 135                 140

Ser Ser Arg Gly Gly Ser Glu Phe Ala Gln Gln Tyr Tyr Gln Leu Val
145                 150                 155                 160

Leu Gln Glu Gln Arg Gln Glu Gln Ile Tyr Arg Ser Leu Ala Ala His
                165                 170                 175

Ala Ala Arg Leu Gln His Asp Ala Glu Arg Arg Gly Glu Tyr Ser Asn
            180                 185                 190

Glu Asp Ile Asp Pro Met His Leu Ala Ala Met Met Leu Gly Asn Pro
        195                 200                 205

Met Ala Pro Ala Val Gln Met Arg Asn Val Asn Met Asn Pro Ile Pro
    210                 215                 220
```

```
Ile Pro Met Val Gly Asn Pro Ile Val Asn Asn Phe Ser Ile Pro Pro
225                 230                 235                 240

Tyr Asn Asn Ala Asn Pro Thr Thr Gly Ala Thr Ala Val Ala Pro Thr
                245                 250                 255

Ala Pro Pro Ser Gly Asp Phe Ala Asn Val Gly Pro Thr Gln Asn Arg
            260                 265                 270

Ser Gln Asn Val Thr Gly Trp Pro Val Tyr Asn Tyr Pro Met Gln Pro
        275                 280                 285

Thr Thr Glu Asn Pro Val Gly Asn Pro Cys Asn Asn Asn Thr Thr Asn
    290                 295                 300

Asn Thr Thr Asn Asn Lys Ser Pro Val Asn Gln Pro Lys Ser Leu Lys
305                 310                 315                 320

Thr Met His Ser Thr Asp Lys Pro Asn Asn Val Pro Thr Ser Lys Ser
                325                 330                 335

Thr Arg Ser Arg Ser Ala Thr Ser Lys Ala Lys Gly Lys Val Lys Ala
            340                 345                 350

Gly Leu Val Ala Lys Arg Arg Arg Lys Asn Asn Thr Ala Thr Val Ser
        355                 360                 365

Ala Gly Ser Thr Asn Ala Gly Ser Pro Asn Ile Thr Thr Pro Gly Ser
    370                 375                 380

Thr Thr Ser Glu Pro Ala Met Val Gly Ser Arg Val Asn Lys Thr Pro
385                 390                 395                 400

Arg Ser Asp Ile Ala Thr Asn Phe Arg Asn Gln Ala Ile Ile Phe Gly
                405                 410                 415

Glu Glu Asp Ile Tyr Ser Asn Ser Lys Ser Ser Pro Ser Leu Asp Gly
            420                 425                 430

Ala Ser Pro Ser Ala Leu Val Ser Lys Gln Pro Thr Lys Val Arg Lys
        435                 440                 445

Asn Thr Lys Lys Ala Ser Thr Ser Ala Phe Pro Val Glu Ser Ala Asn
    450                 455                 460

Lys Leu Gly Gly Asn Ser Val Val Thr Gly Lys Lys Arg Ser Pro Pro
465                 470                 475                 480
```

```
Asn Thr Arg Val Leu Arg Arg Lys Ser Thr Pro Ser Val Ile Leu Asn
                485                 490                 495

Ala Asp Ala Thr Lys Asp Glu Asn Asn Met Leu Arg Thr Phe Ser Asn
            500                 505                 510

Thr Thr Ala Pro Asn Ile His Ser Ala Pro Pro Thr Lys Thr Ala Asn
        515                 520                 525

Ser Leu Pro Phe Pro Gly Ile Asn Leu Gly Ser Phe Asn Lys Pro Ala
    530                 535                 540

Val Ser Ser Pro Leu Ser Ser Val Thr Glu Ser Cys Phe Asp Pro Glu
545                 550                 555                 560

Ser Gly Lys Ile Ala Gly Lys Asn Gly Pro Lys Arg Ala Val Asn Ser
                565                 570                 575

Lys Val Ser Ala Ser Ser Pro Leu Ser Ile Ala Thr Pro Pro Ser Gly
            580                 585                 590

Asp Ala Gln Glu Gln Arg Ser Ser Asn Val Pro Gly Asn Val Val Ile
        595                 600                 605

Lys Pro Pro His Gly Phe Ser Thr Thr Asn Leu Asn Ile Thr Leu Lys
    610                 615                 620

Ser Ser Lys Ile Ile Thr Ser Gln Asn Asn Thr Ala Phe Gln Glu Leu
625                 630                 635                 640

Pro Asn Gly Gly Asn Ile Leu Glu Ala Gln Val Gly Asn Asp Ser Arg
                645                 650                 655

Ser Ser Lys Gly Asn Arg Asp Thr Leu Ser Thr Pro Glu Glu Lys Lys
            660                 665                 670

Pro Ser Ser Asn Asn Gln Gly Tyr Asp Phe Asp Ala Leu Lys Asn Ser
        675                 680                 685

Ser Ser Leu Leu Phe Pro Asn Gln Ala Tyr Ala Ser Asn Asn Arg Thr
    690                 695                 700

Pro Asn Glu Asn Ser Asn Val Ala Asp Glu Thr Ser Ala Ser Thr Asn
705                 710                 715                 720

Asn Gly Asp Asn Asp Asn Thr Leu Ile Gln Pro Ser Ser Asn Val Gly
```

725 730 735

Thr Thr Leu Gly Pro Gln Gln Thr Ser Thr Asn Glu Asn Gln Asn Val
740 745 750

His Ser Gln Asn Leu Lys Phe Gly Asn Ile Gly Met Val Glu Asp Gln
755 760 765

Gly Pro Asp Tyr Asp Leu Asn Leu Leu Asp Thr Asn Glu Asn Asp Phe
770 775 780

Asn Phe Ile Asn Trp Glu Gly
785 790

<210> 7
<211> 768
<212> PRT
<213> Milenrama lipolítica

<400> 7

```
Met Pro Ala Lys Ser Asp Lys Glu Leu Leu Asn Ala Tyr Ile Tyr Asp
1               5                  10                  15

Tyr Leu Leu Lys His Asn Met His Asp Ser Ala Arg Thr Phe Gly Ala
            20                  25                  30

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

Pro Lys Pro Leu Ile Pro Ile Asp Ala Pro Gln Gly Phe Leu Tyr Glu
    50                  55                  60

Trp Trp Ala Leu Phe Trp Asp Ile Tyr Ser Ala Arg Gly Ser Lys Gly
65                  70                  75                  80

Gly Gly Ser Val Pro Ala Gln Gln Tyr Val Gln Gly Thr Met Arg Leu
                85                  90                  95

Arg Gln Glu His Ala Ala Arg Ala Gln Leu Gln Gln Gln His Gln Ala
            100                 105                 110

Gln Gln His Ala Gln Ala Gln Ala Ala Ala Gln Val Gln Gly Gln Ala
        115                 120                 125

Gln Gly Gln Gly Gln Gly Gln Asn Pro Thr Gln Gly Pro Gln Pro Gln
    130                 135                 140

Gly His Met Gly Met Pro Gly Gln Gly Pro His Gln Pro Gly Gly Pro
```

```
145                 150                 155                 160

Phe Met Asn Gly Asn Met Met Phe Pro Pro Gly Gln Met Arg Met Gly
                165                 170                 175

Gln Leu Pro Gln His Leu Gln Gln Gly Thr Gly Val Ala Gly Ala
            180                 185                 190

Asn Pro Ser Asp Asp Ser Ser Ser Pro Gly Gly Thr Ser Pro Ala Lys
        195                 200                 205

Arg Gln Arg Leu Ser Pro Asp Met Gly Gly Gln Ser His Pro Glu Gln
    210                 215                 220

Gln Gly Gln His Met Met Gly Thr Pro Asn Pro Asn Asn Pro Val Phe
225                 230                 235                 240

Asn Ser Gln Val Met Gln Gln Leu Lys Ala Ser Gln Gly Gln Met Pro
                245                 250                 255

Asn Leu Gln Gln Gln Gln Gln Ala Gln Leu Gln Gln Tyr Ser Asn Thr
            260                 265                 270

Leu Ser Leu Ala Gln Gln Arg Ala Met Met Asn Ala Lys Gly Gly Pro
        275                 280                 285

Asn Gly Gln Pro Gly Gln Pro Gly Gln Pro Gly Gln Pro Gly Gln Pro
    290                 295                 300

Gly Gln Pro Gly Gln Pro Gly Gln Pro Gly Gln Pro Leu Pro Met Gly
305                 310                 315                 320

Tyr Glu Gly Met Glu Pro Gly Phe Met Gly Gly Asn Gly Leu Ile Leu
                325                 330                 335

Asn Ala Gln Gln Met Arg Gln Gln Gln Ala Gly Ser Gly Gly Ala Ala
            340                 345                 350

Gly Ser Gly Gln Leu Asn Gly Asn Ser Asn Ala Leu His Asp Tyr Gln
        355                 360                 365

Met Gln Leu Met Leu Leu Glu Gln Gln Asn Lys Lys Arg Leu Met Val
    370                 375                 380

Ala Arg Gln Glu Gln Gln Glu Gly Gln Pro Arg Ala Glu Gly Gln Ala
385                 390                 395                 400
```

```
Ala Ala Ala Gly Ala Asn Pro Gly Ala Arg Met Ser Gly Gln Phe Lys
                405                 410                 415

Arg Pro Gly Ser Ser Pro Val Val Gly Asn Ile Gly Asp Gly Arg Arg
            420                 425                 430

Val Thr Pro Lys Leu Pro Asn Gln Pro Ser Pro Leu Ile Asp Ala Asn
        435                 440                 445

Arg Ala Ser Pro Leu Gln Ser Asn Phe Asn Gly Gln Gly Asp Phe Asn
    450                 455                 460

Val Val Val Gly Pro Ser Gly Gln Met Met Arg Met Gln Gln Pro Gln
465                 470                 475                 480

Gln Gln Gly Gly Pro Pro Gln Gln Gln Gly Gln Gln Gln Gly Gly Pro
                485                 490                 495

Gly Gln Gln Gln Gly Gly Pro Gly Gln Gln Pro Gln Gln Gly Gln Gln
            500                 505                 510

Gln Gly Gln Gln Gly Gly Pro Gln Gly Gly Gln His Arg Phe Asp Asp
        515                 520                 525

Pro Gln Gln Leu Pro Gln Ser Gln Asn Gly Gly Pro Gln Pro Gly Ser
    530                 535                 540

Ala Pro Gly Gln Leu Pro Gln Thr Gln Gly Gly Pro Gln Arg Pro Pro
545                 550                 555                 560

Ser Arg Val Ser Gln Met Pro Pro Pro Gly Val Gln Gly Gly Gln Arg
                565                 570                 575

Thr Gln Pro Ser Ser Pro Gly Gln Ala Ser Asn Pro Gly Ser Gly Gly
            580                 585                 590

Ser Ser Gly Ser Ala Pro Gly Gly Thr Thr Pro Thr Gln Ala Asn Lys
        595                 600                 605

Gln Leu Lys Gly Lys Lys Ala Glu Pro Lys Lys Arg Ala Lys Lys Gly
    610                 615                 620

Asn Gln Pro Val Thr Pro Lys Ala Val Ser Glu Ser Pro Thr Pro Thr
625                 630                 635                 640

Thr Pro Ala Thr Pro Ser Ala Ala Ser Asn Gln Ser Leu Leu Ser Lys
                645                 650                 655
```

```
Ala Thr Asn Phe Ala Ala Asn Asn Lys Gln Ala Gln Ala Ala Ala His
            660                 665                 670

Ala Gln Ala Gln Ala Gln Ala Ala Ala Gln Ala Gln Ser Gln Met His
        675                 680                 685

Met Gly Gly Met Gly Gly Ser Ser Ser Gly Leu Glu Leu Asp Val Asn
    690                 695                 700

Gly Gly Met Gly Leu Asp Asn Asp Ser Ser Ser Phe Leu Asn Asp Phe
705                 710                 715                 720

Ser Thr Gly Asn Glu Gly Asp Val Gly Asp Met Asp Phe Asp Phe Asn
                725                 730                 735

Ser Phe Leu Asn Thr Asp Asp Asn Ala Ala Gly Ala Leu Lys Phe Asp
            740                 745                 750

Ser Gly Ser Ala Phe Gly Trp Gly Glu Gly Val Glu Ala Met Asn Glu
        755                 760                 765
```

<210> 8
<211> 598
<212> PRT
<213> Ogataea polimorfa

<400> 8

```
Met Ser Gln Asn Ala Thr Pro Gln Thr Thr Glu Gly His Met Ser Met
1               5                  10                  15

Asn Asp Gln Gln Ser Asn Ser Ser Ser Gly Thr Asn Asn Thr Glu His
            20                  25                  30

Pro Ser Ser Ser Ser Val Ser Ser Val Pro Thr Val Ser Asp Ser Thr
        35                  40                  45

Ala Val Asn Ser Arg Glu Leu Leu Asn Ala Tyr Val Tyr Asp Phe Ile
    50                  55                  60

Leu Lys Ser Gly Phe Thr Ala Thr Ala Ser Ala Phe Phe Lys Glu Ala
65                  70                  75                  80

Asn Ile Pro Val Ile His Ser Glu Lys Arg Pro Thr Asn Ser Pro Ser
                85                  90                  95

Ser Ser Ala Thr Gly Thr Ser Asp Leu Pro Ala Ser Phe Met Thr Met
            100                 105                 110
```

```
Asp Ala Pro Gln Gly Phe Leu Tyr Glu Trp Trp Gln Ile Phe Trp Asp
        115                 120                 125

Val Phe Asn Ala Arg Thr Gln Arg Gly Gly Thr Thr Asn Ala Thr Gln
    130                 135                 140

Tyr Tyr His Tyr Val Asn Leu Lys Gln Lys Gln Asp His Leu Met Ser
145                 150                 155                 160

Gln Gln Ala Ala Ala Val Ala Ala Ala Ser Thr Val Met Asn Gly Asn
                165                 170                 175

Asn Thr Ser Met Ser Gly Ala Pro Val Thr Ser Ala Ala Gln Asp Ile
            180                 185                 190

Gly Val Leu Met Pro Gln Gln Gln Gln Gln Gln Gln Ala Gln Met
        195                 200                 205

Ala His Pro Pro Gln Gln Val Pro Met Ala Gln Gln Gln Arg Val Asn
    210                 215                 220

Ala Arg Met Gln Gln Pro Gln Gln Gln Gln Ser Gln Ser Met Pro Met
225                 230                 235                 240

His Pro Gln Ala Ala Ala Gln Ala Gln Met Asn Thr Leu Arg Gln Gln
                245                 250                 255

Gln Ile Ala Gln Gln Ala Gln Ala Ala Gln Val Ala Ala Gln Ala Val
            260                 265                 270

Ser Gln Arg Gly Ser Pro Ser Lys Arg Gln Arg Met Asp Ala Ala Gly
        275                 280                 285

Asn Gly Ser Ala Thr Glu Ile Asn Gln Thr Ser Ser Pro Asn Ile Ala
    290                 295                 300

Met Asn Ser Gln Gln Gln Gln Gln Gln Gln Gln His Pro Pro Met Pro
305                 310                 315                 320

Met Pro Gln Gly Met Met Ile Pro Asn Gln Phe Gly Val Pro Gln Gln
                325                 330                 335

Tyr Ala Met Phe Ala Ala Ala Gln Ala Ala Gln Gln Gly Gln Ser Gln
            340                 345                 350

Gln Lys Phe Asn Gln Tyr Met Ile Pro Asn Phe Gln Asn Gln Pro Gln
        355                 360                 365
```

```
Gln Pro Gln Gln Gln Met Leu Met His Gln Gln Met His Gln Gln Ser
    370                 375                 380

Val His Met Gln His Asp Gln Gln His Pro Pro Pro Pro Pro Pro Asn
385                 390                 395                 400

Val Pro Gln His Ser Gln Thr Phe Ser Asn Pro Asn Asp Phe Phe His
                405                 410                 415

Glu Met Pro Lys Gln Asn Asn Pro Gly Thr Ala Arg Val Asn Asp Thr
            420                 425                 430

Ala Ile Lys Asp Tyr Glu Lys Gln Leu Arg Leu Met Glu Ser Gln Asn
        435                 440                 445

Arg Arg Arg Leu Asp Val His Arg Asn Val Ser Asp Ser Lys Asp Pro
    450                 455                 460

Asn Ser Pro Gly Ser Ala Gln Phe Thr Glu Tyr Ser Ala Met Leu Asn
465                 470                 475                 480

Gln Met Pro Pro Gln His Ala Ala Ala Ser Ile Met Gln Gln Arg Ala
                485                 490                 495

Ser Pro Ala Thr Lys Ala Ser Pro Thr Thr Lys Ser Pro Ala Asn Gly
            500                 505                 510

Ala Pro Pro Ala Asn Gly Lys Gln Lys Lys Ala Pro Arg Lys Ala Arg
        515                 520                 525

Lys Asn Ser Ser Ser Val Pro Ala Thr Pro Leu Thr Pro Ala Asn Asn
    530                 535                 540

Gln Pro Thr Pro Gln Thr Asn Ile Pro Pro Thr Pro Gln Asn Thr Thr
545                 550                 555                 560

Pro Gln Gln Thr Pro Gln Ser Ser Ala Ala Ala Ala Ser Gln Val Met
                565                 570                 575

Ala Gly Lys Lys Gly Ile Lys Arg Lys Asn Arg Gly Arg Ala Ile Tyr
            580                 585                 590

Pro Ile Glu Tyr Ser Arg
        595
```

<210> 9
<211> 771
<212> PRT
<213> Aspergillus niger

<400> 9

```
Met Asn Gln Met Asn Gln Met Asn Met Ala Gly Met Asn Pro Gly Ala
1               5                  10                  15

Gly Gly Pro Val Gly Gly Val Pro Met Ile Asn Asn Gly Ser Ala Ala
            20                  25                  30

Pro Arg Asn Glu Gln Asn Ile Asn Asn Ile Pro Glu Asn Met Ile Asn
        35                  40                  45

Asn Leu Asn Thr Tyr Ile Tyr Asp Tyr Phe Leu Lys Arg Gly Tyr His
    50                  55                  60

Asp Cys Ala Arg Ala Leu Val Lys Asp Glu Ser Ile Lys Leu Asn Thr
65                  70                  75                  80

Glu Pro Pro Ile Lys Thr Ser Pro Gly His Arg Arg Asp Ala Asp Val
                85                  90                  95

Asn Gly Val Asp Gly Asp Thr Met Met Thr Asp Gly Lys Asp Gly Asp
            100                 105                 110

Lys Leu Lys Ile Pro Asp Asp Leu Pro Arg Pro Asn Leu Pro Ser Glu
        115                 120                 125

Gly Gln Gln Ser Ser Phe Leu Leu Asp Trp Phe Ser Leu Phe Trp Asp
    130                 135                 140

Phe Phe Trp Ala Gln Arg Lys Lys Gly Asn Ser Asn Asp Val Arg Ser
145                 150                 155                 160

Tyr Leu Thr His Thr Gln Asn Met Met Arg Leu Arg Glu Gln His Gln
                165                 170                 175

Asn Gln Leu Leu Arg Gln Gln Pro Leu Met Asn Gly Gln Met Gly Gln
            180                 185                 190

Met Asn Ile Arg Arg Asn Gly Met Val Pro Pro Asn Leu Gln Lys Thr
        195                 200                 205

Val Leu Gln Asn Asn Thr Thr Gly Leu Ser Gln Gln Gln Leu Ala Gln
    210                 215                 220

Met His Lys Asn Gln Gln Val Gln Met Met Gln Gln Met Gln Arg Glu
```

225 230 235 240

His Ser Asp Met Asp Met Asn Gly His Arg Pro Gln Ser Pro Ala Ser
245 250 255

Ala Glu Asn Ala Pro Ser Pro Ser Lys Arg Pro Arg Leu Glu Gly Gly
260 265 270

Pro Met Asn Gly Gln Gln Leu Ala Pro Asn Gly Arg Gly Gln Ala Gln
275 280 285

Gly Ile Pro Gly Gln Pro Thr Pro Gln Ala Leu Leu Met Gln Asn Gly
290 295 300

Leu Asn Arg Ala Met Asn Pro Asn Gln Phe Gln Ala Phe Gln Gln Ser
305 310 315 320

Gly Pro Ala Ala Gln Gln Lys Gln Met Gln Gly Met Pro Asn Gly Met
325 330 335

Met Asn Pro Ala Asn Val Met Ala Asn Pro Gln Thr Glu Met Val Ser
340 345 350

Ile Pro Glu Gly Gln Val Tyr Pro Ile Asn Gly Asp Tyr Tyr Gly Ala
355 360 365

Asn Gly Gln Met Ala Gln Val Arg Thr Gly Met Gln Thr Pro Gly Gly
370 375 380

Gln His Gly Asn His Ala Leu Gln Asp Tyr Gln Met Gln Leu Met Leu
385 390 395 400

Leu Glu Gln Gln Asn Lys Arg Arg Leu Met Met Ala Arg Gln Glu Gln
405 410 415

Asp Ser Met Ala Arg Pro Asp Gly Gln Pro Gln Met Pro Gly Gln Gln
420 425 430

Leu Pro Pro Gly Thr Ser Pro Gln Gly Ser Arg Ala Gly Thr Ser Pro
435 440 445

Asn Pro Asn Asp Gln Met Lys Arg Gly Thr Pro Lys Met Pro Gln Thr
450 455 460

Gly Leu Pro Gly Ser Pro Ser Ala Ala Asp Ala Met Ala Gln Gly Arg
465 470 475 480

```
Gly Ser Pro Ala Ser Met Asn Phe Pro Gly Gly Gln Met Pro Pro Glu
                485                 490                 495

Met Ala Gly Pro Gln Phe Phe Val Lys Asn Met Ala Asp Gly Met Ala
            500                 505                 510

Ala Pro Asn Gly Met Arg Pro Pro Ser Ser Asn Pro Ala Phe Ser Thr
        515                 520                 525

Pro Gln Met Gly Gln Pro Ile Gln Ala Gly Ala Asn Arg Met Pro Asn
    530                 535                 540

Gly Gly Trp Gln Pro Gln Gln Gly Ala Gln Gly Gln Pro Met Ala Pro
545                 550                 555                 560

Gln Gln Ser Pro Ala Thr Gln Pro Gln Ser Thr Gly Thr Pro Gln Glu
                565                 570                 575

Arg Asn Ala Met Pro Pro Pro Gln Ala Pro Pro Ala Pro Gly Ala Asn
            580                 585                 590

Val Gly Arg Thr Gln Pro Pro Ser Pro Gln Thr Ala Ala Pro Pro Thr
        595                 600                 605

Pro Gln Gln Gly Asn Lys Pro Ala Pro Lys Lys Lys Glu Thr Lys Asp
    610                 615                 620

Ser Arg Lys Arg Pro Lys Lys Gly Ala Ala Ala Ala Ala Ala Ala Ala
625                 630                 635                 640

Gln Ala Asn Thr Ala Ala Thr Pro Ser Ser Glu Ala Glu His Pro Pro
                645                 650                 655

Thr Pro Thr Pro Ser Thr Pro Ile Thr Pro Gln His Pro Asn Ser Phe
            660                 665                 670

Asn Lys Thr Gly Ala Asn Ala Thr Thr Ser Ala Pro Gln Gln Pro Thr
        675                 680                 685

Ser Ala Pro Ala Pro Pro Pro Leu Val Gln Gln Pro Pro Asp Gln Thr
    690                 695                 700

Gln Gln Pro Phe Asn Glu Leu Ser Ile Pro Asp Ala Ser Ala Phe Asn
705                 710                 715                 720

Leu Asp Phe Ser Ala Leu Glu Asn Pro Asp Ile Leu Glu Asn Phe Asp
                725                 730                 735
Phe Asp Thr Phe Leu Asn Thr Asp Ala Asp Thr Ala Gly Phe Gly Phe
            740                 745                 750

Asp Pro Asn Thr Ser Tyr Pro Thr Asp Gly Val Glu Thr Gly Ala Gly
        755                 760                 765

Asp Gly Leu
    770
```

<210> 10

<211> 1240
<212> ADN
<213> Secuencia Artificial

<220>
<223> promotor

<400> 10
caaacatttg ctccccctag tctccaggga aatgtaaaat atactgctaa tagaaaacag 60

taagacgctc agttgtcagg ataattacgt tcgactgtag taaaacagga atctgtattg 120

ttagaaagaa cgagagtttt ttacggcgcc gccatattgg gccgtgtgaa aacagcttga 180

aaccccacta ctttcaaagg ttctgttgct atacacgaac catgtttaac caacctcgct 240

tttgacttga ctgaagtcat cggttaacaa tcaagtaccc tagtctgtct gaatgctcct 300

ttccatattc agtaggtgtt tcttgcactt ttgcatgcac tgcggaagaa ttagccaata 360

gcgcgtttca tatgcgcttt tacccctct tttgtcaagc gcaaaatgcc tgtaagattt 420

ggtgggggtg tgagccgtta gctgaagtac aacaggctaa ttccctgaaa aaactgcagc 480

tcagggattc ccactatttg gtattctgat atgtttttcc tgatatgcat caaaactcta 540

atctaaaacc tgaatctccg ctatttttt ttttttttga tgaccccgtt ttcgtgacaa 600

attaatttcc aacggggtct tgtccggata agagaatttt gtttgattat ccgttcggat 660

aaatggacgc ctgctccata tttttccggt tattacccca cctggaagtg cccagaattt 720

tccggggatt acggataata cggtggtctg gattaattaa tacgagatct cagggattcc 780

cactatttgg tattctgata tgtttttcct gatatgcatc aaaactctaa tctaaaacct 840

gaatctccgc tatttttttt ttttttgat gaccccgttt tcgtgacaaa ttaatttcca 900

acggggtctt gtccggataa gagaattttg tttgattatc cgttcggata aatggacgcc 960

tgctccatat ttttccggtt attaccccac ctggaagtgc ccagaatttt ccggggatta 1020

cggataatac ggtggtctgg attaattaat acgccaagtc ttacattttg ttgcagtctc 1080

gtgcgagtat gtgcaataat aaacaagatg agccaattta ttggattagt tgcagcttga 1140

ccccgccata gctaggcata gccaagtgct atgggtgtta gatgatgcac ttggatgcag 1200

tgagttttgg agtataaaag atccttaaaa ttccaccctt 1240

<210> 11
<211> 1425
<212> ADN
<213> Secuencia Artificial

<220>
<223> promotor

<400> 11

caaacatttg ctccccctag tctccaggga aatgtaaaat atactgctaa tagaaaacag 60

taagacgctc agttgtcagg ataattacgt tcgactgtag taaaacagga atctgtattg 120

ttagaaagaa cgagagtttt ttacggcgcc gccatattgg gccgtgtgaa aacagcttga 180

aaccccacta ctttcaaagg ttctgttgct atacacgaac catgtttaac caacctcgct 240

tttgacttga ctgaagtcat cggttaacaa tcaagtaccc tagtctgtct gaatgctcct 300

ttccatattc agtaggtgtt tcttgcactt ttgcatgcac tgcggaagaa ttagccaata 360

gcgcgtttca tatgcgcttt tacccctct tttgtcaagc gcaaaatgcc tgtaagattt 420

ggtgggggtg tgagccgtta gctgaagtac aacaggctaa ttccctgaaa aaactgcagc 480

tcagggattc ccactatttg gtattctgat atgtttttcc tgatatgcat caaaactcta 540

atctaaaacc tgaatctccg ctatttttt ttttttgat gaccccgttt tcgtgacaaa 600

ttaatttcca acggggtctt gtccggataa gagaattttg tttgattatc cgttcggata 660

aatggacgcc tgctccatat ttttccggtt attaccccac ctggaagtgc ccagaatttt 720

ccggggatta cggataatac ggtggtctgg attaattaat acgccaagtc ttacattttg 780

ttgcagtctc gtgcgagtat gtgcaataat aaacaagatg agccaattta ttggattagt 840

tgcagcttga ccccgccata gctaggcata gccaagtgct atgggtgtta gatgatgcac 900

ttggatgcag tgagttttgg agtataaaag atccttaaaa ttccaccctt agatctcagg 960

gattcccact atttggtatt ctgatatgtt tttcctgata tgcatcaaaa ctctaatcta 1020

aaacctgaat ctccgctatt ttttttttt ttgatgaccc cgttttcgtg acaaattaat 1080

ttccaacggg gtcttgtccg gataagagaa ttttgtttga ttatccgttc ggataaatgg 1140

acgcctgctc catatttttc cggttattac cccacctgga agtgcccaga attttccggg 1200

gattacggat aatacggtgg tctggattaa ttaatacgcc aagtcttaca ttttgttgca 1260

gtctcgtgcg agtatgtgca ataataaaca agatgagcca atttattgga ttagttgcag 1320

cttgaccccg ccatagctag gcatagccaa gtgctatggg tgttagatga tgcacttgga 1380

tgcagtgagt tttggagtat aaaagatcct taaaattcca ccctt 1425

<210> 12
<211> 966
<212> ADN
<213> Secuencia Artificial

<220>
<223> promotor

<220>
<221> característica_misc
<222> (592)..(593)
<223> n es T o está ausente

<400> 12

caaacatttg ctccccctag tctccaggga aatgtaaaat atactgctaa tagaaaacag 60

taagacgctc agttgtcagg ataattacgt tcgactgtag taaaacagga atctgtattg 120

ttagaaagaa cgagagtttt ttacggcgcc gccatattgg gccgtgtgaa aacagcttga 180

aaccccacta ctttcaaagg ttctgttgct atacacgaac catgtttaac caacctcgct 240

tttgacttga ctgaagtcat cggttaacaa tcaagtaccc tagtctgtct gaatgctcct 300

ttccatattc agtaggtgtt tcttgcactt ttgcatgcac tgcggaagaa ttagccaata 360

gcgcgtttca tatgcgcttt tacccctct tttgtcaagc gcaaaatgcc tgtaagattt 420

ggtgggggtg tgagccgtta gctgaagtac aacaggctaa ttccctgaaa aaactgcaga 480

tagacttcaa gatctcaggg attcccacta tttggtattc tgatatgttt ttcctgatat 540

gcatcaaaac tctaatctaa aacctgaatc tccgctattt ttttttttt tnngatgacc 600

ccgttttcgt gacaaattaa tttccaacgg ggtcttgtcc ggataagaga attttgtttg 660

attatccgtt cggataaatg gacgcctgct ccatattttt ccggttatta ccccacctgg 720

aagtgcccag aattttccgg ggattacgga taatacggtg gtctggatta attaatacgc 780

caagtcttac attttgttgc agtctcgtgc gagtatgtgc aataataaac aagatgagcc 840

aatttattgg attagttgca gcttgacccc gccatagcta ggcatagcca agtgctatgg 900

gtgttagatg atgcacttgg atgcagtgag ttttggagta taaaagatcc ttaaaattcc 960

accctt 966

<210> 13
<211> 1000
<212> ADN
<213> Secuencia Artificial

<220>
<223> promotor

<400> 13

gtaaatagcg gcagcaatcc agtaaccttt tctgaatagc agagccttaa ctaaaataat 60

ggccagggta aaaaattcga aatttgacac caaaaataaa gacttgtcgt tataagtctt 120

aacaaagtcc gcaattttgg agctaacggt ggcggttgct gggatattca ataatggtag 180

aatgttgctg cgggtatatg acagagcgtg aaacacactg aacaaggtaa atggaacaac 240
agcaattgca atatggggga ggatagtcaa gaacaaagca gcaatggcaa agtactgaat 300

attctccaaa gccaaaaggt ccagtggttt caacgacaaa gtcttgttgg tatagctttg 360

gaacaaaagg acaccgaaag actcgacagc gcccacaaat acagcgttgt agaagaacga 420

attgattgct ccagagcttc taatagtcag aagatacccc aaacctccga gcaacgttag 480

cacatgacct aagaaccagg cgaagtgaag agtctggaat aacgacaccc agtcagtttt 540

tcctgagctc ctggtgggat tggtagaagc atttgatttg cttggagtgg ttttatttga 600

agatggtgtt gaagccattg ttgctaaaga gtcggagttt tgcttttagg gtttgttaag 660

caaaggagga aaaactgcgc cgtttgaagt cccaggtagt ttcgcgtgtg aggccagcca 720

gggaaagctt ccttcggtac ttttttttct tttgcaggtt ccggacggat taagcttcgg 780

gttatgaggg gggcggtagc caattccgga cacaatattg cgtcgcagct agtcacccca 840

ccataaatat acgcaggatt gaggtaataa catcgatagt cttagtaatt aatacaattc 900

agtggcgaat ttggcaacat gacgtaaggc ccactgttgt ctataaaagg ggatgaattt 960

tcatgttttt gaggcctccc ggacaattta ttgaactcaa 1000

<210> 14
<211> 1000
<212> ADN
<213> Secuencia Artificial

<220>
<223> promotor

<400> 14

tggactgttc aatttgaagt cgatgctgac gatgtcaaga gagatgctca attatatttg 60

tcatttgctg gttacactgg aaacgctact tttgttggcg gaaactctac cagtttggcc 120

gtccatgtaa acgatgtcgt tctgggccgt gaccgtttca acacgaacat aaccaatgac 180

aaatccactt acaggtctag ttcatatgga ggcaattggt accttacttc tttggatgtc 240

ccaagtgggg ctttaacgtc tggtactaac aatgtctcgt ttgtcactac aaactccgag 300

gtaaataaag gattcttgtg ggattctctc aagtttgttt ggaagttgta acaggtttat 360

aagcatatcg tgcgcttgtc cacaattgaa tcatttattg ttgcgagata catgaacaaa 420

gtgtgaactg ggacccatta ctacaattcc cacgcaaccg ttgtttcaaa gcccatattt 480

tttgacaatt gtttcgttac acccccagtt tgatgtacat cgcttgcaat gatgtgtgtc 540

ccggagtatt ttccatattc agcttgaatt cgtatactca accaatatct gggggtatac 600

ttttatgtaa cctatacaaa tcaactatac tatttcacct ttcgaccaat catctcccat 660

cttgttaagt tttgcttcct atatccctga ccctgacatc acccatgatt ccgctcaacg 720

gttctcctct acatcgtccc tcttttggag agggtgttca gtttgacatt caaattaccc 780
cccgccatca cgcgcaaccg agaccgcacc cccgaatttt cacaaattac cccacaccct 840

atactccacc actatgaggg ttattagaac tgatcacgta taaataccac cgcaagttcc 900

caagggatcg tgttcttctt ctccaattgc aatcatattt ctgactcttt ctagttcaga 960

ttaattcctt tacacttgct tttttccctt acctttatcc 1000

<210> 15
<211> 1001
<212> ADN
<213> Secuencia Artificial

<220>
<223> promotor

agaccagcag tttaactacg caaatccaca ggaatttcta catcacaata ccaatggtaa 60

taccacgacg tcaaggaatg gaaacgacga cttggaggaa gacttcgtca acctcttgcg 120

gagtacccga ggctaagaca ataagaagaa aaaaaaaaga aaagcggtgg gggagggatt 180

attaaataag gattatgtaa ccccagggta ccgttctata catatttaag gattatttag 240

gacaatcgat gaaatcggca tcaaactgga tgggagtata gtgtccggat aatcggataa 300

atcatcttgc gaggagccgc ttggttggtt ggtgagagga gtgaaatatg tgtctcctca 360

cccaagaatc gcgatatcag caccctgtgg gggacactat tggcctccct cccaaacctt 420

cgatgtggta gtgctttatt atattgatta cattgattac atagctaaac cctgcctggt 480

tgcaagttga gctccgaatt ccaatattag taaaatgcct gcaagataac ctcggtatgg 540

cgtccgaccc cgcttaatta ttttaactcc tttccaacga ggacttcgta atttttgatt 600

agggagttga gaaacggggg gtcttgatac ctcctcgatt tcagatccca ccccctctca 660

gtcccaagtg ggacccccct cggccgtgaa atgcgcgcac tttagttttt ttcgcatgta 720

aacgccggtg tccgtcaatt aaaagtcgca gactagggtg aactttacca tttttgtcgc 780

actccgtctc ctcggaatag gggtgtagta attctgcagt agtgcaattt ttaccccgcc 840

aaggggggge gaaaagagac gacctcatca cgcattctcc agtcgctctc tacgcctaca 900

gcaccgacgt agttaacttt ctcccatata taaagcaatt gccattcccc tgaaaacttt 960

<400> 15 aacctctgct ttttcttgat ttttccttgc ccaaagaaaa g 1001

<210> 16
<211> 1000
<212> ADN
<213> Secuencia Artificial

<220>
<223> promotor

<400> 16 ctgcacaacc attgccagta aggacgaaga gaaggcccca ctacccaaaa ttcaggataa 60
cgtcttcata ccatgcagcg acgcctacaa gacgctgtca agacatgcca acttcaacga 120

agtgaacttt aacacattga tcgggaaatt gaccaccaag ggaatgctgg ttgaggctgg 180

aagcgttgcc agtgtcctga gggaactgga ccgaaagttt agtaatgcat aagaggatat 240

atataggaat gcagtaataa tattagtacc cattaagtgg gctaagccat tggaaggccg 300

tctgactgat ggtggtgttc ttctcattta gatagtgcat ttgcaactac cgtctgagat 360

tgagtttgat gtgaagctcc agcgccaaaa cagtataaga accttatctc cgcattattg 420

ttcttgcgta aaagtttgtg tgaagaaaca ggggtagttg cgcagattag ttgtaatatg 480

cgcataggat gggtcattga cttctttcct cgaaagagcc acaccgttag ctaaaaaagg 540

acgcgcatct accccaaaat agaatgtggg gaaataggac gcgcaacttc ctctcaatca 600

ctggacgtca gaaaaacaaa tgcgcaatcg agtcaccctc cgtgataccc tccgtgatac 660

cccctctccg tctattctga cagcgtctcc ccatgacgtt tcaatctact tagaaaagat 720

ttcgtttttt tttccttcaa ttacacgatc tcatcttctg caagggtctg gaggacatca 780

ccaatctgcg actccataac ttagtcctga gtttatattt acgcttcatc tgatgagtag 840

gaagaaaaag tttcacgaaa ttcccccgcc aacttgccct tcggaataag cagccactct 900

ccttctgccc atagtaagct tgcgcgaggc cccaacttgg ccagaaactt taaatatgcc 960

aaacaatctc ccccaatcta agttctccct cttctaaaaa 1000

<210> 17
<211> 9
<212> ADN
<213> Secuencia Artificial

<220>
<223> región reguladora

<400> 17 ataaatgga 9

<210> 18
<211> 15
<212> ADN
<213> Secuencia Artificial

<220>
<223> región reguladora

<400> 18 catatttttc cggtt 15

<210> 19
<211> 15
<212> ADN
<213> Secuencia Artificial

<220>
<223> motivo

<400> 19 tatttttttt ttttt 15

<210> 20
<211> 15
<212> ADN
<213> Secuencia Artificial

<220>
<223> motivo

<400> 20 ttttttttt ttttt 15

<210> 21
<211> 16
<212> ADN
<213> Secuencia Artificial

<220>
<223> motivo

<400> 21 tatttttttt ttttt 16

<210> 22
<211> 16
<212> ADN
<213> Secuencia Artificial

<220>
<223> motivo

<400> 22 ttttttttt ttttt 16

<210> 23
<211> 17
<212> ADN
<213> Secuencia Artificial

<220>
<223> motivo

<400> 23 tatttttttt tttttt 17

<210> 24
<211> 17
<212> ADN
<213> Secuencia Artificial

<220>
<223> motivo

<400> 24 ttttttttt tttttt 17

<210> 25
<211> 18
<212> ADN
<213> Secuencia Artificial

<220>
<223> motivo

<400> 25 tatttttttt ttttttt 18

<210> 26
<211> 18
<212> ADN
<213> Secuencia Artificial

<220>
<223> motivo

<400> 26 ttttttttt ttttttt 18

<210> 27

<211> 19
<212> ADN
<213> Secuencia Artificial

<220>
<223> motivo

<400> 27 tatttttttt tttttttt 19

<210> 28
<211> 19
<212> ADN
<213> Secuencia Artificial

<220>
<223> motivo

<400> 28 ttttttttt tttttttt 19

<210> 29
<211> 20
<212> ADN
<213> Secuencia Artificial

<220>
<223> motivo

<400> 29 tatttttttt ttttttttt 20

<210> 30
<211> 20
<212> ADN
<213> Secuencia Artificial

<220>
<223> motivo

<400> 30 ttttttttt ttttttttt 20

<210> 31
<211> 21
<212> ADN
<213> Secuencia Artificial

<220>
<223> motivo

<400> 31 tatttttttt ttttttttt t 21

<210> 32
<211> 21
<212> ADN
<213> Secuencia Artificial

<220>
<223> motivo

<400> 32 ttttttttt ttttttttt t 21

<210> 33
<211> 22
<212> ADN
<213> Secuencia Artificial

<220>
<223> motivo

<400> 33 tatttttttt ttttttttt tt 22

<210> 34
<211> 22
<212> ADN
<213> Secuencia Artificial

<220>
<223> motivo

<400> 34 ttttttttt ttttttttt tt 22

<210> 35
<211> 33
<212> ADN
<213> Secuencia Artificial

<220>
<223> región reguladora principal

<400> 35 ataaatggac gcctgctcca tatttttccg gtt 33

<210> 36
<211> 9
<212> ADN
<213> Secuencia Artificial

<220>
<223> región central del espaciador

<400> 36 cgcctgctc 9

<210> 37
<211> 255
<212> ADN
<213> Secuencia Artificial

<220>
<223> región principal del espaciador

<400> 37
attaccccac ctggaagtgc ccagaatttt ccggggatta cggataatac ggtggtctgg 60

attaattaat acgagatctc agggattccc actatttggt attctgatat gtttttcctg 120

atatgcatca aaactctaat ctaaaacctg aatctccgct ttttttttttt tttttgatga 180

ccccgttttc gtgacaaatt aatttccaac ggggtcttgt ccggataaga gaattttgtt 240

tgattatccg ttcgg 255

<210> 38
<211> 10
<212> ADN
<213> Secuencia Artificial

<220>
<223> sitio de inicio de la traducción

<400> 38 ttccaccctt 10

<210> 39
<211> 9
<212> ADN
<213> Secuencia Artificial

<220>
<223> sitio de inicio de la traducción

<400> 39 ttcgaaacg 9

<210> 40
<211> 9
<212> ADN
<213> Secuencia Artificial

<220>
<223> sitio de inicio de la traducción

<220>
<221> característica_misc
<222> (7)..(7)
<223> R en la posición 7 es A o G, preferiblemente A

<400> 40gccgccrcc 9

<210> 41
<211> 859
<212> ADN
<213> Secuencia Artificial

<220>
<223> promotor

<400> 41

```
ggaatctgta ttgttagaaa gaacgagagt tttttacggc gccgccatat tgggccgtgt       60

gaaaacagct tgaaacccca ctactttcaa aggttctgtt gctatacacg aaccatgttt      120

aaccaacctc gcttttgact tgactgaagt catcggttaa caatcaagta ccctagtctg      180

tctgaatgct cctttccata ttcagtaggt gtttcttgca cttttgcatg cactgcggaa      240

gaattagcca atagcgcgtt tcatatgcgc ttttaccccc tcttttgtca agcgcaaaat      300

gcctgtaaga tttggtgggg gtgtgagccg ttagctgaag tacaacaggc taattccctg      360

aaaaaactgc agatagactt caagatctca gggattccca ctatttggta ttctgatatg      420

tttttcctga tatgcatcaa aactctaatc taaaacctga atctccgcta ttttttttt       480

ttttttgatg accccgtttt cgtgacaaat taatttccaa cggggtcttg tccggataag      540

agaattttgt ttgattatcc gttcggataa atggacgcct gctccatatt tttccggtta      600

ttaccccacc tggaagtgcc cagaattttc cggggattac ggataatacg gtggtctgga      660

ttaattaata cgccaagtct tacattttgt tgcagtctcg tgcgagtatg tgcaataata      720

aacaagatga gccaatttat tggattagtt gcagcttgac cccgccatag ctaggcatag      780

ccaagtgcta tgggtgttag atgatgcact tggatgcagt gagttttgga gtataaaaga      840

tccttaaaat tccacccctt                                                  859
```

<210> 42
<211> 664
<212> ADN
<213> Secuencia Artificial

<220>
<223> promotor

<400> 42

ccatattcag taggtgtttc ttgcactttt gcatgcactg cggaagaatt agccaatagc 60

gcgtttcata tgcgctttta ccccctcttt tgtcaagcgc aaaatgcctg taagatttgg 120

tgggggtgtg agccgttagc tgaagtacaa caggctaatt ccctgaaaaa actgcagata 180

gacttcaaga tctcagggat tcccactatt tggtattctg atatgttttt cctgatatgc 240

atcaaaactc taatctaaaa cctgaatctc cgctattttt ttttttttt tgatgacccc 300

gttttcgtga caaattaatt tccaacgggg tcttgtccgg ataagagaat tttgtttgat 360

tatccgttcg gataaatgga cgcctgctcc atatttttcc ggttattacc ccacctggaa 420

gtgcccagaa ttttccgggg attacggata atacggtggt ctggattaat taatacgcca 480

agtcttacat tttgttgcag tctcgtgcga gtatgtgcaa taataaacaa gatgagccaa 540

tttattggat tagttgcagc ttgaccccgc catagctagg catagccaag tgctatgggt 600

gttagatgat gcacttggat gcagtgagtt ttggagtata aaagatcctt aaaattccac 660

cctt 664

<210> 43
<211> 493
<212> ADN
<213> Secuencia Artificial

<220>
<223> promotor

<400> 43
ctgcagatag acttcaagat ctcagggatt cccactattt ggtattctga tatgtttttc 60

ctgatatgca tcaaaactct aatctaaaac ctgaatctcc gctatttttt ttttttttt 120

gatgaccccg ttttcgtgac aaattaattt ccaacggggt cttgtccgga taagagaatt 180

ttgtttgatt atccgttcgg ataaatggac gcctgctcca tatttttccg gttattaccc 240

cacctggaag tgcccagaat tttccgggga ttacggataa tacggtggtc tggattaatt 300

aatacgccaa gtcttacatt ttgttgcagt ctcgtgcgag tatgtgcaat aataaacaag 360

atgagccaat ttattggatt agttgcagct tgaccccgcc atagctaggc atagccaagt 420

gctatgggtg ttagatgatg cacttggatg cagtgagttt tggagtataa aagatcctta 480

aaattccacc ctt 493

<210> 44
<211> 370
<212> ADN
<213> Secuencia Artificial

<220>
<223> promotor

<400> 44

gaccccgttt tcgtgacaaa ttaatttcca acggggtctt gtccggataa gagaattttg 60

tttgattatc cgttcggata aatggacgcc tgctccatat ttttccggtt attaccccac 120

ctggaagtgc ccagaatttt ccggggatta cggataatac ggtggtctgg attaattaat 180

acgccaagtc ttacattttg ttgcagtctc gtgcgagtat gtgcaataat aaacaagatg 240

agccaattta ttggattagt tgcagcttga ccccgccata gctaggcata gccaagtgct 300

atgggtgtta gatgatgcac ttggatgcag tgagttttgg agtataaaag atccttaaaa 360

ttccaccctt 370

<210> 45
<211> 328
<212> ADN
<213> Secuencia Artificial

<220>
<223> promotor

<400> 45
ccggataaga gaattttgtt tgattatccg ttcggataaa tggacgcctg ctccatattt 60

ttccggttat taccccacct ggaagtgccc agaattttcc ggggattacg gataatacgg 120

tggtctggat taattaatac gccaagtctt acattttgtt gcagtctcgt gcgagtatgt 180

gcaataataa acaagatgag ccaatttatt ggattagttg cagcttgacc ccgccatagc 240

taggcatagc caagtgctat gggtgttaga tgatgcactt ggatgcagtg agttttggag 300

tataaaagat ccttaaaatt ccaccctt 328

<210> 46
<211> 1000
<212> ADN
<213> Pichia pastoris

<400> 46

```
ctgctactct ggtcccaagt gaaccacctt ttggacccta ttgaccggac cttaacttgc      60

caaacctaaa cgcttaatgc ctcagacgtt ttaatgcctc tcaacacctc caaggttgct     120

ttcttgagca tgcctactag gaactttaac gaactgtggg gttgcagaca gtttcaggcg     180

tgtcccgacc aatatggcct actagactct ctgaaaaatc acagttttcc agtagttccg     240

atcaaattac catcgaaatg gtcccataaa cggacatttg acatccgttc ctgaattata     300

gtcttccacc gtggatcatg gtgttccttt ttttcccaaa gaatatcagc atcccttaac     360

tacgttaggt cagtgatgac aatggaccaa attgttgcaa ggtttttctt tttctttcat     420
cggcacattt cagcctcaca tgcgactatt atcgatcaat gaaatccatc aagattgaaa     480

tcttaaaatt gccccttcca cttgacagga tccttttttg tagaaatgtc ttggtgtcct     540

cgtccaatca ggtagccatc tctgaaatat ctggctccgt tgcaactccg aacgacctgc     600

tggcaacgta aaattctccg gggtaaaact taaatgtgga gtaatggaac cagaaacgtc     660

tcttcccttc tctctccttc caccgcccgt taccgtccct aggaaatttt actctgctgg     720

agagcttctt ctacggcccc cttgcagcaa tgctcttccc agcattacgt tgcgggtaaa     780

acggaggtcg tgtacccgac ctagcagccc agggatggaa aagtcccggc cgtcgctggc     840

aataatagcg ggcggacgca tgtcatgaga ttattggaaa ccaccagaat cgaatataaa     900

aggcgaacac ctttcccaat tttggtttct cctgacccaa agactttaaa tttaatttat     960

ttgtccctat ttcaatcaat tgaacaacta tcaaaacaca                          1000
```

<210> 47
<211> 756
<212> PRT
<213> Pichia pastoris

<400> 47

```
Met Ile Pro Thr Ile Asp Pro Lys Asp Pro Glu Leu Val Ser Glu Asp
1               5                  10                  15

Thr Ala Gln Ser Ala Ser Ala Arg Lys Arg Ser Lys Val Ser Arg Ala
            20                  25                  30

Cys Asp Glu Cys Arg Arg Lys Lys Ile Lys Cys Asp Ala Thr Phe Leu
        35                  40                  45

Ala Asn Ser Asn Thr Leu Leu Lys Pro Cys Thr Asn Cys Tyr Lys Tyr
    50                  55                  60

Asn Cys Ser Cys Ser Phe Thr Arg Val Pro Leu Lys Arg Gly Pro Ser
65                  70                  75                  80

Lys Gly Phe Ala Arg Asp Gly Ser Gly Tyr Glu Arg Arg Arg Ser Ser
                85                  90                  95

Ser Val His Ser Val Ser Ser Ser Gln Ser Val Thr Ser Pro Val Pro
            100                 105                 110

Ser His Ala Ser Leu Pro Ile Pro Pro Ala Asn Pro Val Ser Leu Pro
        115                 120                 125

Arg Leu Asn Val Pro Gly Asp Gly Leu Leu Ser Pro Lys Ala Val Pro
    130                 135                 140
```

```
Pro Thr Asn Leu Phe Trp Lys Val Pro Tyr Glu Leu Pro Ser Phe Ser
145                 150                 155                 160

Asp Arg Arg Ser Ser Val Ala Ser Ala Asp Ser Phe Arg Arg Pro Ser
                165                 170                 175

Ile Tyr Gln Ser Asp Ser Glu Asp Asp Phe Tyr Ser Ala Thr Gly Ser
            180                 185                 190

Gln Arg Asn Ser Ile Ser Gln Ala Pro Arg Gln Arg Asn Leu Ser Pro
        195                 200                 205

Ala Leu Ser Val Ser Ser Thr Ser Ser Leu Asn Asn Arg Ile Lys Ser
    210                 215                 220

Leu Asn Met Val Ala Ser Thr Leu Glu Ser Asn Ile His Asn Tyr Tyr
225                 230                 235                 240

Ser Gln Gly Phe Asn Ser Ser Leu Pro Ile Leu Pro Leu Asp Glu Arg
                245                 250                 255

Ile Leu Ser Thr Leu Leu Ser Asn Val Ser Asn Gly Ser Ser Ser Ala
            260                 265                 270

Ser Trp Asp Ala Ile Arg Ser Pro Ile Leu Glu Leu Phe Asp Lys Ser
        275                 280                 285

Ile Leu Met Leu Leu Arg Ser Tyr Glu Ser Gln Phe Asn Phe Asn Asp
    290                 295                 300

Leu Leu Asp His Val Thr Glu Met Gln Ser Ile Tyr Pro Arg Ile Arg
305                 310                 315                 320

Ser His Leu Leu Ser Asp Glu Leu Leu Lys Leu Ile Phe Leu Met Ser
                325                 330                 335

Gly Val Leu Thr Asp Tyr Ala Leu Ile Leu Thr Gly Gln Pro Tyr Ser
            340                 345                 350

Thr Gly Leu Ser Ile Thr Val Ser Val Phe Asn Asp Trp Lys Thr Tyr
        355                 360                 365

Glu Asn Val Gln Arg Val Leu Val Ile Asn Arg Ala Gly Ser Leu Asp
    370                 375                 380

Leu Asp Tyr Asp Ser Leu Pro Phe Leu Phe Ala Arg Cys Tyr Leu Ser
385                 390                 395                 400
```

```
Leu Ala Thr Leu Asp Leu Ile Tyr Ser Leu Ser Phe Ser Ser Pro Arg
                405                 410                 415

Leu Ile Ser Ser Phe Ala Asn Leu Pro Ile Leu Asp Ile Val Gln Lys
            420                 425                 430

Cys Gly Ile Thr Lys Asp Ala Lys Leu Asp Glu Thr Pro Leu Pro Val
        435                 440                 445

Leu Asp Gln Phe Leu Asn Cys Phe Leu Pro Gly Asp Thr Tyr Pro Thr
    450                 455                 460

Ala Leu Asn Thr Leu Lys Thr Gly Leu Val Leu Leu Asp Phe Thr Asn
465                 470                 475                 480

Asn Arg Ser Thr Thr Leu Arg Phe Pro Phe Ile Asn Ile His Asp Asp
                485                 490                 495

Asn His Met Thr Gly Leu Ser His Leu Leu Ser Asn Val Ser Asp Phe
            500                 505                 510

Met Ser Gln Phe Thr Glu Val His Ser Asp Ser Lys Asp Ser Gln Leu
        515                 520                 525

Leu Phe Leu Arg Cys Ile Trp Ala Phe Trp Glu Ile Gly Ser Val Leu
    530                 535                 540

Ser Glu Leu Ile Asp His Phe Ile Ser Ser Ser Ala Asn Ser Gln Val
545                 550                 555                 560

Gly Asp Lys Asp Ala Ser Phe Phe Tyr Glu His Gln Leu Lys Val Thr
                565                 570                 575

Thr Leu Leu Gly Thr Phe Ser Asn Ile Ala Ser Ala Phe Leu Thr Ser
            580                 585                 590

Ser Thr Thr Ala Ala Ser His Pro Pro Pro Ser Ile Ser Pro Phe His
        595                 600                 605

Ile Ile Ser Met Val Glu Ser Phe Lys Met Val Gln Phe Leu Asn Lys
    610                 615                 620

Leu Ile Ala Ser Phe Ile Ser Leu Asn Glu Lys Leu Glu Lys Arg Glu
625                 630                 635                 640

Leu Glu Asp Glu Leu Ser Lys Cys Lys Glu Glu Leu Asn Asn Leu Asn
```

645 650 655

Glu Arg Phe Gln Ala Val Ser Ser Val Gln Thr Leu Pro Val Val His
660 665 670

Val Leu Phe Arg Asp Leu Val Phe Ser Ser Asn Arg Leu Asp Thr Gln
675 680 685

Arg Asp Arg Ala Ser Ser Val Val Ser Ala Thr Thr Thr Thr Ser Thr
690 695 700

Ala Thr Thr Thr Ala Thr Thr Lys Lys Ser Ser Phe Gly Asn Leu Leu
705 710 715 720

His Ser Asp Glu Glu Asn Ile Leu Pro Thr Val Ile Asp Trp Cys Lys
725 730 735

Glu Gln Lys His Ser Ala Glu Met Phe Leu Asn Lys Asn Asp Leu Asn
740 745 750

Gly Trp Leu Tyr
755

<210> 48
<211> 1036
<212> PRT
<213> Pichia pastoris

<400> 48

```
Met Met Pro Glu Glu Gln Val Thr Ser Pro Gln Arg Lys His Gln Lys
1               5                   10                  15

Ser Lys Ala Lys Thr Ile Arg Ala Pro Gly Ser Ser Ile Glu Arg Val
            20                  25                  30

Ala Gln Ala Cys Asp Arg Cys Arg Ser Lys Lys Thr Arg Cys Asp Gly
        35                  40                  45

Lys Arg Pro Gln Cys Ser Gln Cys Ala Ala Val Gly Phe Glu Cys Lys
    50                  55                  60

Ile Ser Asp Lys Leu Ser Arg Arg Ala Phe Pro Arg Gly Tyr Thr Glu
65                  70                  75                  80

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

His Lys Leu Ile Asp Leu Lys Asn Glu Gln Val Glu Ile Lys Asn Arg
```

```
            100                 105                 110

Ile Asp Gln Glu Ser Thr Leu Thr Asn Glu Asn Leu Thr Leu Leu Asn
        115                 120                 125

Lys Glu Gln Glu Val Ser His Ser Gly Asn Ile His His His Ala Asp
    130                 135                 140

Gly Glu Pro Cys Asn Cys Ala Asn Ser Val Ser Ala Arg Pro Val Ser
145                 150                 155                 160

Ile Ala Gly Ser Val Asp Ile Asp Thr Thr Asp Leu Ser Asp Glu Asp
                165                 170                 175

Asp Ser Leu Tyr Ser Ala Ala Ser Tyr Asn Ala Lys Tyr His Gln Thr
            180                 185                 190

Gly Thr Ser Gly Pro Glu Met Val Arg Leu Ser Gln Arg Tyr Ser Ser
        195                 200                 205

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

Ala Ala Ala Ala Ile Ser Ile Gln Asn Lys Met Arg Thr Gln Thr Phe
225                 230                 235                 240

Val Asn Leu Ala Asn Leu Val Ala Met Ser Ile Pro Arg Thr Thr Glu
                245                 250                 255

Glu Thr Leu Phe Ile Ala Ser Leu Leu Ala Lys Ile Cys Asn Val His
            260                 265                 270

Gly Phe Gln Ser Lys Ala Pro Ile Leu Thr Ala Lys Ser Ile Ala Leu
        275                 280                 285

Leu Lys Asp Lys Tyr Asn Tyr Gly Asn Asp Glu Val Phe Ala Asn Ile
    290                 295                 300

Thr Leu Lys Asn Val Asn Phe Asn Lys Leu Thr Ser Gln Gln Ser Gln
305                 310                 315                 320

Gln Phe Phe Gln Ser Leu Asn Leu Pro Asn Gln Val Asn Leu Asp Leu
                325                 330                 335

Phe Ile Thr Thr Phe Phe Asn Thr Trp Asn Asn Phe Ile Pro Ile Ile
            340                 345                 350
```

```
Asn Arg His Ile Phe Met Ser Ser Tyr Ile Lys Phe Asn Lys Ser Arg
        355                 360                 365

Glu Thr Met Phe Thr Asp Asn Ser Met Phe Gly Asn Glu Lys Phe Gly
    370                 375                 380

Glu Ile Leu Leu Leu Ile Thr Thr Met Val Met Leu Ser Gln Glu Arg
385                 390                 395                 400

Asn Asn Asn Arg Glu Ala Val Pro Ser Ser Ser Tyr Lys Lys Asp Ser
                405                 410                 415

Thr Pro His Pro His Arg Pro Asp Ala Ser Ser Gln Ser Asn Val Glu
            420                 425                 430

Ile Leu Gln Tyr Tyr Asp His Leu Ile His Glu Phe Ile Lys Ser Asn
        435                 440                 445

Ile Ser Asp Asp Cys Ser Leu Pro Thr Leu Glu Ser Leu Ser Leu Gln
    450                 455                 460

Leu Leu Tyr Cys Leu Ala Ile Gly Asp Leu Thr Thr Ser Tyr Glu Leu
465                 470                 475                 480

Arg Gly Lys Ile Ile Thr Met Gly Gln Gln Leu Arg Leu His Arg Cys
                485                 490                 495

Pro Ser Ala Val Leu Gly Thr Asn Gly Ser Lys Val Ser Gln Met Gln
            500                 505                 510

Gln Gly Glu Arg Arg Ile Leu Phe Trp Cys Ile Tyr Ile Leu Asp Thr
        515                 520                 525

Phe Ser Ala Leu Ile Leu Gly Val Pro Arg Leu Leu Lys Asp Tyr Glu
    530                 535                 540

Ile Glu Cys Ala Leu Pro Phe Ser Asn Glu Ser Asn Asn Ala Asn Val
545                 550                 555                 560

Lys Gly Ser Ile Glu Asn Thr Thr Asn Thr Val Ile Ile Asn Asn Ile
                565                 570                 575

Lys Leu Ser Leu Ala Gly Lys Val Ser Asp Cys Ala Leu Ala Val Met
            580                 585                 590

Arg Tyr Ser Lys Val Leu Gly Asn Ile Leu Asp Ser Ile Phe Gln Arg
        595                 600                 605
```

```
Ser Ser Ile Asn Asn Pro Ser Val Val Ser Lys Ser Thr Asn Ile Thr
    610                 615                 620

Glu Glu Thr Cys Leu Leu His Glu His Ala Leu Asp Leu Trp Arg Arg
625                 630                 635                 640

Glu Leu Ser Pro His Ile Asn Val Asp Leu Asp Lys Ser Pro Gly Gly
                645                 650                 655

Val Glu Tyr Glu Arg Leu Ser Asp Asn Gln Leu Thr Ile Leu Phe Leu
            660                 665                 670

Tyr Tyr His Ala Lys Ile Leu Ile Tyr Leu Pro Leu Met Ala Asn Glu
        675                 680                 685

Ser Ser Gln Ser Arg Ser Ser Ala Ser Tyr Ile Asn Ile Gln Gln Ser
    690                 695                 700

Thr Thr Ser Ile Leu Ala Ile Ala Asn Thr Leu Ala Thr Lys Glu Arg
705                 710                 715                 720

Asn Phe Tyr Phe Leu Pro Leu Pro Val Asn Leu Ser Arg Glu Lys Val
                725                 730                 735

Arg Leu Ala Phe Leu Ser Ala Lys Gly Ser Leu Glu Tyr Ala Arg Gly
            740                 745                 750

Gly Ala Leu Phe Gln Glu Ser Lys Ile Leu Leu Ala Ser Val Ile Asn
        755                 760                 765

Glu Leu Lys Ile Glu Thr Ser Ile Gly Met Leu Gly Cys Leu Ser Val
    770                 775                 780

Pro Cys Met Glu Ala Val Asp Asn Ala Met Glu Gln Ile Met Ala Leu
785                 790                 795                 800

Pro Gly Lys Val Ser Ser Val Asn Gly Ser Asn Ser Glu Met Lys Arg
                805                 810                 815

Ser Ser Ser Lys Arg Lys Ser Ser Pro Leu Arg Gln Asp Ile Ser Gly
            820                 825                 830

Asp Glu Arg Lys Ser His Asn Ile Glu Val Ser Asp Ser Arg Thr Pro
        835                 840                 845

Ser Val Gln Ser Ser Leu Tyr Pro Gln Pro Asn Gln Met His His Pro
    850                 855                 860
```

```
Asn Ile Ile Lys Ser Glu Asn Asn Glu Gln Met Ile Pro Glu Asn Asp
865                 870                 875                 880

Thr Pro Gly Ala Ile Asn Asp Ile Phe Thr Ser His Ser Pro Pro Gly
                885                 890                 895

Thr Val Thr Ser Met Lys Glu Glu Asp Leu Pro Ile Lys Val Pro Ile
            900                 905                 910

Leu Leu Gln Thr Gln Gln Arg Gln Ile Tyr Asn Asn Asn Pro Asn His
        915                 920                 925

Ser Leu Phe Ser Gln Gln Pro Gly Thr Gln Val Leu Ser Gly Gln Gln
    930                 935                 940

Met Pro Gly Pro Ser Ser Thr Asp Gln Gln Phe Lys Arg Ile Thr Thr
945                 950                 955                 960

Pro Asp Gly Leu Asp Ser Leu Met Met Gln Asp Phe Gly Val Asp Ala
                965                 970                 975

Ser Leu Gly Leu Pro Met Leu Asp Phe Asp Phe Asn Phe Asp Phe Glu
            980                 985                 990

Asn Val Gln Asn Asn Tyr Ser Gln  Ser Asn Val Ser Pro  Pro Asn Ser
        995                 1000                1005

Glu Ser  Val Pro Ser Ser Ile  Gln Gly Thr His Ser  Asn Asp Pro
    1010                1015                1020

Lys Asp  Ser Gln Val Ser Ala  Gly Ser Leu Phe Gly  Leu
    1025                1030                1035
```

<210> 49
<211> 887
<212> PRT
<213> Pichia pastoris

<400> 49

```
Met Lys Glu Asn Gln Ala Ser Asn Lys Phe Asn Leu Ile Lys Asn Pro
1               5                  10                  15

Ile Thr Gly Lys Pro Arg Ile Ser Gln Ala Cys Asp Arg Cys Arg Ile
            20                  25                  30

Lys Lys Ile Lys Cys Asp Gly Thr Leu Pro Ser Cys Thr Asn Cys Ser
        35                  40                  45
Lys Ile Gly Phe Val Cys Lys Ile Ser Asp Arg Leu Thr Arg Ser Ser
    50                  55                  60

Phe Pro Lys Gly Tyr Thr Lys Asn Leu Glu Gln Lys Leu Ile Asp Met
65                  70                  75                  80

Glu Leu Asp Arg Asn Arg Leu Met Leu Glu Leu Asn Arg Ile Lys Lys
                85                  90                  95

Glu Gly Phe Asp Gly Thr Asn Asn Asn Ile Ala Met Ala Ser Ser Val
            100                 105                 110

Ser Ser Ser Glu Asn Leu Lys Ser Asp Asp Ser Ser Glu Cys Gln Ser
        115                 120                 125

Val Thr Val Ser Leu Ser Ser Thr Ser Gly Pro Ser Leu Ser Pro Glu
    130                 135                 140

Pro Lys Gln Asp Asp Phe Arg Phe Arg Val Gly Met Asp Gly Ser Phe
145                 150                 155                 160

Val Leu Asn Gln Phe Leu Gln Ser Pro Leu Met Asp Tyr Ile Lys Ser
                165                 170                 175

Leu Asn Val Leu Gln Phe Asn Gly Cys Ala Asn Phe Asp Gln Ser Phe
            180                 185                 190

Asn Asp Asp Pro Leu Val Leu Asn Lys Tyr His Met Asn Leu Asn Arg
        195                 200                 205

Phe Leu Asn Leu Ile Phe Tyr Lys Leu Leu Leu Pro Leu Ile His Arg
    210                 215                 220

Asn Ser Asn Thr Leu Asn Glu Lys Phe Ala Glu Asp Asn Asn Ser Leu
225                 230                 235                 240

Asp Ser Leu Ile Trp Lys Phe Phe Thr Asn Tyr Asn Lys Leu Ile Pro
                245                 250                 255

Ile Leu Glu Phe Asp Ser Phe Tyr Lys Asp Tyr Leu Gln Phe Ile His
            260                 265                 270

Lys Tyr Tyr Ser Asn Asn Gln Val Phe Val Asp Gly Phe Arg Lys Tyr
        275                 280                 285

Phe Glu Phe Ser Glu Phe Glu Gln Cys Phe Ile Val Lys Leu Ile Leu
```

```
    290                 295                 300

Ile Leu Lys Phe Thr Leu Pro Val Ile His Asp Thr Ser Val Pro Ser
305                 310                 315                 320

Glu Ile Tyr Arg Leu Ile Ser Met Asp Ser Leu Gln Arg Leu Phe Gly
                325                 330                 335

Asn Ile Asp Phe Leu Lys Pro Ser Thr Asp Lys Val Ser Ile Leu Leu
            340                 345                 350

Leu Val Leu His Tyr Met Val Leu Tyr Glu Ser Pro Lys Ser Leu Leu
        355                 360                 365

Asp Thr Gln Asp Glu Ala Gln Lys Tyr Asp Glu Phe Ile Gly Asn Leu
    370                 375                 380

Leu Ser Thr Ala Val His His Ile Thr Ser Leu Arg Leu His Ile Asp
385                 390                 395                 400

Pro Arg Lys Leu Gln Phe Pro Arg Pro Leu Pro Ser Asn Gly Asn Arg
                405                 410                 415

Leu Arg Ile Lys Leu Ser Trp Cys Tyr Lys Leu Ile Ser Lys Leu Phe
            420                 425                 430

Arg Val Ile Tyr Asn Ile Asp Asn Asp Ser Leu Tyr Ser Leu Asp Asp
        435                 440                 445

Ser His Leu Pro Glu Leu Gln Ser Ile Ser Ile Leu His Glu Glu Leu
    450                 455                 460

Asp Val Thr Ile Gln Phe Asn Asn Leu Leu Asn Leu Ile Pro Asn Asn
465                 470                 475                 480

Phe His Ser Leu Arg Asp Lys Gln Ser Leu Ser Lys Ile Lys Thr Gln
                485                 490                 495

Leu Leu Glu Trp His Lys Asn Phe Asn Thr Glu Phe Val Glu His Phe
            500                 505                 510

Asn Leu Asn Asp Thr Asp Ser Asp Glu Leu Ser Ala Glu Lys Ile Asn
        515                 520                 525

Val Leu Arg Ser Lys Leu Ile Ser Leu Asn Arg Leu Asn Cys Tyr Asn
    530                 535                 540
```

```
Ser Tyr Phe Gln Leu Val Ile Glu Leu Gln Leu Lys Glu Asn Leu Asp
545                 550                 555                 560

Ser Val Val Ser Gly Ile Phe Gly Leu Ser Asn Glu Met Leu Ile Asp
                565                 570                 575

Asn Lys Ser Ser Thr Glu Leu Leu Asn Thr Leu Gln Gln Thr Pro Ile
            580                 585                 590

Ile His Gln Ser Ser Ile Leu Val Ser Leu Cys Tyr Arg Ile Gln Thr
        595                 600                 605

Gly Asn Leu Gln Asp Glu Ile Cys Ser Ile Leu Val Asn Asn Tyr Glu
    610                 615                 620

Lys Leu Leu Gln Cys Asn Asp Ala Gly Leu Pro Ile Lys Ile Leu Pro
625                 630                 635                 640

Gln Leu Val His Tyr Phe Lys Gly Lys Ile Ser Thr Asn Leu Ser Asn
                645                 650                 655

Ser Ala Ala His Glu Asp Leu Met Asn Met Phe Thr Leu Asn Asp Asn
            660                 665                 670

Leu Ser Thr Thr Thr Thr Asp Leu Asp Ser Phe Ile Ile Pro Pro Lys
        675                 680                 685

Arg Lys Gln Asp Gln Thr Leu Pro Ile Gly Thr Lys Arg Ser Lys Ser
    690                 695                 700

Ala Ser Thr Ser Ser Val Ile Ser Ser Asp Asp Cys Ser Leu Phe Ser
705                 710                 715                 720

Asn Ser Leu Ser Val Pro Thr Thr Phe Ser Gly Ser Ser Ile Ser Val
                725                 730                 735

Gly Met Asp Asn Pro Pro Ser Ser Leu Phe Gly Ser Tyr Lys Arg Pro
            740                 745                 750

Ser Ser Ile Val Lys Gln Glu Pro Thr Ile Asn Pro Arg Ser Asn Gly
        755                 760                 765

Thr Asn Thr Asp Ser Asn Leu Phe Asp Thr Phe Asn Asp Ser Ile Lys
    770                 775                 780

Gly Ser Leu Asn Asn Gly Leu Lys Lys Leu Lys Asp Ile Arg Cys Asn
785                 790                 795                 800
```

Ser Val Val Glu Arg Ser His Ser Ser Gln Arg Asn Asp Phe Leu Met
805 810 815

Asp Gln Glu Asp Ser Ile Thr Lys Glu Thr Ile Asn Phe Ser Glu Leu
820 825 830

Phe Thr Cys Gly Thr Pro Thr Ala Ser Gln Ser Ile Asp Arg Ser Pro
835 840 845

Lys Ser Leu Leu Leu Asn Asp Leu Ala Ile Ala Pro Asp Thr Leu Val
850 855 860

Ile Lys Pro Asp Ala Glu Asp Leu Asp Arg Leu Lys Asn Lys Ile Arg
865 870 875 880

Ser Val Lys Ser Thr Val His
885

<210> 50
<211> 26
<212> PRT
<213> Komagataella phaffii

<400> 50
Leu Leu Asn Gly Tyr Ile Tyr Asp Tyr Leu Val Lys Ser Asn Met Gln
1 5 10 15

Asn Leu Ala Asp Gln Phe Ala Gln Glu Thr
20 25

<210> 51
<211> 25
<212> PRT
<213> Komagatella pastoris

<400> 51
Leu Asn Gly Tyr Ile His Asp Tyr Leu Val Lys Ser Asn Met Gln Asn
1 5 10 15

Leu Ala Asp Gln Phe Ala Gln Glu Ser
20 25

<210> 52
<211> 25
<212> PRT
<213> Saccharomyces cerevisiae

<400> 52
Leu Asn Glu Tyr Ile Phe Asp Phe Leu Thr Lys Ser Ser Leu Lys Asn
1 5 10 15

Thr Ala Ala Ala Phe Ala Gln Asp Ala
20 25

<210> 53
<211> 25
<212> PRT
<213> Saccharomyces cerevisiae

<400> 53

Leu Asn Glu Tyr Ile Phe Asp Phe Leu Thr Lys Ser Ser Leu Lys Asn
1 5 10 15

Thr Ala Ala Ala Phe Ala Gln Asp Ala
20 25

<210> 54
<211> 27
<212> PRT
<213> Milenrama lipolítica

<400> 54
Glu Leu Leu Asn Ala Tyr Ile Tyr Asp Tyr Leu Leu Lys His Asn Met
1 5 10 15

His Asp Ser Ala Arg Thr Phe Gly Ala Glu Ala
20 25

<210> 55
<211> 27
<212> PRT
<213> Ogataea polimorfa

<400> 55
Glu Leu Leu Asn Ala Tyr Val Tyr Asp Phe Ile Leu Lys Ser Gly Phe
1 5 10 15

Thr Ala Thr Ala Ser Ala Phe Phe Lys Glu Ala
20 25

<210> 56
<211> 26
<212> PRT
<213> Aspergillus niger

<400> 56
Asn Asn Leu Asn Thr Tyr Ile Tyr Asp Tyr Phe Leu Lys Arg Gly Tyr
1 5 10 15
His Asp Cys Ala Arg Ala Leu Val Lys Asp
20 25

<210> 57
<211> 774
<212> ADN
<213> Secuencia Artificial

<220>
<223> secuencia scR

<400> 57

caggaacaac taatggagtc tgggggtggt ttggttaccc tgggtggttc tcttaagctt 60

tcatgtaagg cctctggtat tgatttttcg cactacggta tctcctgggt tagacaagct 120

cctggaaaag gtctggaatg gatcgcttac atttacccaa attacggttc tgttgactat 180

gcctcctggg tcaatggtag gttcactatt tcccttgaca acgctcagaa cacggtattc 240

ctacagatga tctccctaac cgctgctgat actgcaacct acttctgtgc tcgtgacaga 300

ggttactact ctggctctcg tggaactaga cttgacttat ggggacaagg tactctcgtt 360

accatctcta gtggtggagg tggttctgga ggaggaggtt ccggcggagg tggtagcgag 420

ctggtcatga ctcaaacccc tccatcccta tctgcatcag tcggtgaaac cgttagaatt 480

agatgccttg catctgagtt cttgttcaac ggtgtgtcct ggtatcaaca aaagcctggt 540

aagcctccaa agtttctcat ttctggtgcc tcaaacctcg aatctggagt gccaccaaga 600

ttttccggat ctggctctgg tactgactac actctgacaa ttggtggtgt tcaagctgag 660

gatgttgcta cctactattg tctcggtggt tactcaggat cttccggcct aactttcggt 720

gccggtacaa acgtcgagat caaaggtgga catcaccacc accatcacta atag 774

<210> 58
<211> 837
<212> ADN
<213> Secuencia Artificial

<220>
<223> secuencia vHH

<400> 58
caggttcagc tgcaggagtc cggtggtggt ctggttcaag ccggtggttc attaagattg 60

tcctgtgctg cctctggtag aactttcact tctttcgcaa tgggttggtt tagacaagca 120

cctggaaaag agagagagtt tgttgcttct atctccagat ccggtacttt aactagatac 180

gctgactctg ccaagggtag attcactatt tctgttgaca acgccaagaa cactgtttct 240

ttgcaaatgg acaaccttaa cccagatgac accgcagtct attactgtgc cgctgacttg 300

cacagaccat acggtccagg aacccaaaga tccgatgagt acgattcttg gggtcaggga 360

actcaagtca ctgtctcttc aggtggtgga tctggtggtg gaggttcagg tggtggagga 420
tccggtggtg gtggttctgg tggtggtgga tctggtggag gtgaagttca acttgtcgaa 480

tccggtggtg cacttgtcca acctggtgga tctcttagac tttcttgtgc cgcctccggt 540

tttcctgtta accgttactc tatgcgttgg tacagacaag cccctggaaa agaacgtgaa 600

tgggttgccg gaatgtcctc agctggtgac agatcctcct acgaagattc tgtgaaggga 660

cgtttcacca tctccagaga tgacgcccgt aacaccgttt accttcaaat gaactccctt 720

aagcctgagg atactgccgt ctactattgt aacgtgaatg tcggatttga atactgggga 780

cagggaaccc aagttactgt ctcttccggt ggacatcacc accaccatca ctaatag 837

<210> 59
<211> 20
<212> PRT
<213> Secuencia Artificial

<220>
<223> secuencia de señales

<220>
<221> CARACTERÍSTICAS_MISC
<222> (3)..(3)
<223> X en la posición 3 es F o L

<220>
<221> CARACTERÍSTICAS_MISC
<222> (16)..(16)
<223> X en la posición 16 es A o T

<400> 59

```
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
```

<210> 60
<211> 36
<212> PRT
<213> Secuencia Artificial

<220>
<223> secuencia líder

<220>
<221> CARACTERÍSTICAS_MISC
<222> (3)..(3)
<223> X en la posición 3 es F o L

<220>
<221> CARACTERÍSTICAS_MISC
<222> (16)..(16)
<223> X en la posición 16 es A o T

<400> 60

```
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
```

<210> 61
<211> 85
<212> PRT
<213> Saccharomyces cerevisiae

<400> 61

```
Met Arg Phe Pro Ser Ile Phe Thr Ala Val Leu Phe Ala Ala Ser Ser
1               5                   10                  15

Ala Leu Ala Ala Pro Val Asn Thr Thr Thr Glu Asp Glu Thr Ala Gln
            20                  25                  30

Ile Pro Ala Glu Ala Val Ile Gly Tyr Ser Asp Leu Glu Gly Asp Phe
        35                  40                  45

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

Phe Ile Asn Thr Thr Ile Ala Ser Ile Ala Ala Lys Glu Glu Gly Val
65                  70                  75                  80

Ser Leu Glu Lys Arg
                85
```

<210> 62
<211> 274
<212> PRT
<213> Komagataella pastoris

<400> 62

```
Met Ser Ser Asp Ala Val Glu Gln Leu Glu Asn Phe Gln Leu Ile Lys
1               5                  10                  15

Phe Asp Arg Phe Asp Pro Ser Thr Gln Ser Thr Ile Arg Ile Ala Arg
            20                  25                  30

Ser Pro Lys Pro Ile Pro Val Lys Val Val Ile Val Gly Asp Gly Gly
        35                  40                  45
Cys Gly Lys Thr Cys Leu Leu Asn Val Phe Ala Thr Gly Thr Phe Pro
    50                  55                  60

Glu Ala Tyr Val Pro Thr Ile Ile Glu Asn Val Val Ile Thr Leu Val
65                  70                  75                  80

Thr Pro Thr Gly Gln Ile Ala Ala Val Thr Leu Trp Asp Thr Ala Gly
                85                  90                  95

Gln Glu Glu Tyr Asp Arg Leu Arg Pro Leu Ser Tyr Ser Asp Val Asp
            100                 105                 110

Val Val Leu Leu Cys Tyr Ser Ile Asp Asn Leu Ser Thr Phe His Asn
        115                 120                 125

Val Ala Asp Lys Trp Tyr Pro Glu Val Ala His Phe Cys Pro Asn Thr
    130                 135                 140

Pro Ile Ile Leu Val Gly Thr Lys Ser Asp Met Arg Arg His Gln Lys
145                 150                 155                 160

Ser Gln Pro His Phe Val Ser Pro Gln Asp Ser Ser Gln Leu Ala Arg
                165                 170                 175

Gln Met Gly Ala Val Met Asn Ile Glu Cys Ser Ala Lys Glu Val Ser
            180                 185                 190

Asn Val Asn Ile Val Phe Asp Ala Ala Val Ser Tyr Cys Leu Ser Asn
        195                 200                 205

Ser Arg Pro Lys Thr Arg Gly Asp Asn Asp Asn Asn Arg Ser Asn Arg
    210                 215                 220

Arg Leu Ser Arg Ala Lys Arg Ala Ser Met Phe Ile Arg Gly Lys Asp
225                 230                 235                 240

Val Ser Ser Thr Ser Gly Asn Ser Arg Glu Glu Leu Val Glu Tyr Asp
                245                 250                 255

Gln Asp Gly Leu Ala Ile Ile Pro Asp Arg Lys Lys Arg Lys Cys Ser
            260                 265                 270

Ile Ile
```

<210> 63
<211> 95
<212> PRT
<213> Komagataella pastoris

<400> 63

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

Ile Gly Glu Thr Thr Thr Ser Ser Thr Thr Ala Ser Leu Ser Glu Ser
            20                  25                  30

Pro Thr Leu Val Trp Val Thr Gly Thr Asp Ala Ser Gly Arg Leu Ala
        35                  40                  45

Thr Thr Gln Ser Ala Tyr Thr Gln Gln Phe Ser Gln Leu Tyr Ser Ser
    50                  55                  60

Ile Ala Ser Pro Ser Ser Gly Ser Ile Gly Leu Gly Thr Ile Gln Gly
65                  70                  75                  80

Thr Val Gly Ile Val Arg Thr Tyr Glu Thr Ile Thr Leu Ala Ser
                85                  90                  95
```

<210> 64
<211> 86
<212> PRT
<213> Komagataella pastoris

<400> 64

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

Ala Ala Asn Leu Asp Lys Lys Gln Asp Glu Pro Thr Ser Ala Arg Ser
            20                  25                  30

Ala Gly Ala Gly Gly Ser Ser Ser Thr Met Leu Lys Leu Tyr Thr Asp
        35                  40                  45

Glu Ala Gln Gly Leu Lys Val Asp Pro Leu Ile Val Leu Val Leu Ala
    50                  55                  60

Val Gly Phe Ile Phe Ser Val Ile Gly Leu His Val Val Ala Lys Leu
65                  70                  75                  80

Thr Gly Lys Leu Ile Asn
                85
```

<210> 65
<211> 361
<212> PRT
<213> Komagataella pastoris

<400> 65

```
Met Thr Pro Arg Ser His Ile Phe Phe Asp Ile Ser Ile Asn Asn Gln
1               5                   10                  15

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

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

Lys Ser Gly Lys Pro Leu His Tyr Lys Gly Ser Thr Phe His Arg Ile
    50                  55                  60

Ile Lys Asp Phe Met Val Gln Gly Gly Asp Phe Thr Asn Gly Asn Gly
65                  70                  75                  80

Thr Gly Gly Glu Ser Ile Tyr Gly Glu Lys Phe Glu Asp Glu Asn Phe
                85                  90                  95

Gln Leu Thr His Asp Lys Pro Phe Leu Leu Ser Met Ala Asn Ala Gly
            100                 105                 110

Pro Gly Thr Asn Gly Ser Gln Phe Phe Ile Thr Thr Val Pro Thr Pro
        115                 120                 125

His Leu Asp Asn Lys His Val Val Phe Gly Lys Val Ile Ala Gly Lys
    130                 135                 140

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

Glu Pro Val Val Ile Glu Asp Cys Gly Glu Leu Pro Glu Asp Ala Asp
                165                 170                 175

Leu Thr Ile Ser Asp Glu Thr Gly Asp Lys Tyr Glu Glu Val Leu Lys
            180                 185                 190

Asp Asn Glu Asn Ile Asp Ile Asp Asp Phe Glu Gln Val Tyr Gln Ala
        195                 200                 205

Ile Thr Glu Ile Lys Glu Leu Gly Thr Lys Tyr Phe Lys Asn Gly Asp
    210                 215                 220

Thr Lys Ile Ala Phe Glu Lys Tyr Gln Lys Ala Ala Asn Tyr Leu Leu
225                 230                 235                 240
```

Glu Tyr Ile Pro Ser Asp Leu Ser Glu Glu Gln Ser Ser Lys Leu Glu
245 250 255

Leu Leu Lys Thr Ser Val Phe Ser Asn Val Ala Leu Ala Gly Leu Lys
260 265 270

Val Ser Lys Phe Lys Asp Thr Ile Lys Tyr Ala Thr Leu Val Ile Glu
275 280 285

Asp Glu Ser Ala Asp Ala Lys Ala Lys Ser Lys Gly Tyr Tyr Arg Arg
290 295 300

Gly Ser Ala Tyr Ser Ser Leu Lys Asp Glu Asp Ser Ala Ile Ser Asp
305 310 315 320

Phe Gln Lys Ala Leu Glu Leu Ser Pro Gly Asp Pro Ala Ile Ser Gln
325 330 335

Ser Leu Gln Arg Thr Thr Lys Ala Arg Lys Asp Arg Leu Ala Lys Glu
340 345 350

Lys Ala Ala Leu Ser Lys Phe Phe Glu
355 360

<210> 66
<211> 127
<212> PRT
<213> Komagataella pastoris

<400> 66
Met Thr Asn Trp Lys Ala Ile Leu Thr Pro Ala Gln Tyr Gln Val Leu
1 5 10 15

Arg Leu Gly Gly Thr Glu Arg Pro Tyr Thr Gly Gln Tyr Val Asn Phe
20 25 30

Lys Lys Asn Gly Thr Tyr Leu Cys Ser Gly Cys Gln Thr Pro Leu Tyr
35 40 45

Lys Ser Gly Thr Lys Phe Asp Ser Ser Cys Gly Trp Pro Ala Phe Tyr
50 55 60

Glu Ala Leu Pro Gly Ala Val Lys Arg Ile Glu Asp Asn Ser Leu Gly
65 70 75 80

Met Arg Arg Ile Glu Ile Arg Cys Ser Lys Cys Asp Gly His Leu Gly
85 90 95
His Val Phe Glu Gly Glu Gly Phe Asp Thr Pro Thr Asp Ser Arg His
100 105 110

Cys Val Asn Ser Ile Ser Leu Lys Phe Gln Gly Glu Glu Glu Asn
115 120 125

<210> 67
<211> 546
<212> PRT
<213> Komagataella pastoris

<400> 67

```
Met Ser His Leu Leu Leu Arg Asp Ser Phe Trp Gly Arg Thr Ile Tyr
1               5                  10                  15

His Leu Ser Lys His Arg Tyr Phe Ser Phe Pro Glu Glu Lys Asp Gly
            20                  25                  30

Phe Ile Ala Pro Glu Lys Tyr Tyr Leu Asn Met Asp Gln Val Ser Ile
        35                  40                  45

His Ala Glu Ser Glu Lys Asn Ile Val Glu Gly Leu Val Asp Thr Ser
    50                  55                  60

Asn Ser Ser Leu Glu Glu Val Lys Thr Thr Arg Val Ile Val Asp Trp
65                  70                  75                  80

Asp Glu Tyr Asp Gln Lys Glu Asn Pro Gln Asn Trp Ser Ser Leu Leu
                85                  90                  95

Lys Cys Phe Val Val Phe Glu Val Gly Ile Leu Thr Val Ala Val Tyr
            100                 105                 110

Met Gly Ser Ala Ile Tyr Thr Pro Gly Ile Glu Asp Ile Met Arg Asp
        115                 120                 125

Leu Asn Val Ser Arg Thr Val Ala Thr Leu Pro Leu Thr Leu Phe Val
    130                 135                 140

Ile Gly Tyr Ala Val Gly Pro Met Ile Phe Ser Pro Met Ser Glu His
145                 150                 155                 160

Pro Ala Ile Gly Arg Thr Thr Ile Tyr Val Trp Thr Leu Phe Ile Phe
                165                 170                 175

Ala Ile Leu Gln Ile Pro Thr Ala Leu Thr Thr Asn Ile Ala Gly Phe
            180                 185                 190
```

```
Cys Ile Leu Arg Phe Ile Gly Gly Phe Phe Ala Ser Pro Ala Leu Ala
        195                 200                 205

Thr Gly Pro Ala Ser Val Gly Asp Val Ile Ala Ile Pro His Leu Pro
    210                 215                 220

Val Gly Leu Gly Leu Trp Ser Ile Cys Ala Val Cys Gly Pro Ser Leu
225                 230                 235                 240

Gly Pro Leu Phe Gly Ala Ile Phe Ser Gln Leu Val Ser Trp Arg Trp
                245                 250                 255

Cys Phe Trp Phe Leu Leu Ile Thr Ser Gly Thr Leu Phe Ile Val Leu
            260                 265                 270

Gly Phe Thr Leu Pro Glu Thr Tyr Val Pro Thr Leu Leu Tyr Arg Lys
        275                 280                 285

Ala Arg Arg Leu Arg Ala Leu Thr Lys Asn Glu Leu Ile Ile Ser Lys
    290                 295                 300

Gly Glu Leu Asp Ile Gln Asp Arg Thr Ala Lys Glu Val Leu Ile Glu
305                 310                 315                 320

Cys Leu Trp Arg Pro Val Asp Ile Ser Phe Arg Asp Pro Val Val Leu
                325                 330                 335

Met Ile Asn Leu Tyr Ile Ser Met Val Tyr Ser Ile Trp Tyr Ile Trp
            340                 345                 350

Phe Glu Ala Phe Pro Ile Val Phe Leu Glu Ile Tyr Gly Phe Ser Leu
        355                 360                 365

Ile Gly Met Gly Ala Ser Phe Ala Gly Ile Leu Ile Gly Val Leu Ile
    370                 375                 380

Cys Ser Ala Cys Tyr Cys Tyr Ala Cys His Val Thr Phe Ala Arg Arg
385                 390                 395                 400

Ile Ile Ala Asn Glu Thr Ile His Pro Glu Phe Phe Val Pro Gly Ala
                405                 410                 415

Ile Ile Gly Gly Cys Ile Met Pro Thr Gly Ile Phe Ile Leu Gly Trp
            420                 425                 430

Thr Ala Thr Lys Ser Val His Trp Ile Val Pro Ile Ile Gly Ser Gly
        435                 440                 445
```

```
Leu Phe Ala Ala Gly Gly Tyr Leu Ile Phe Gln Thr Leu Phe Asn Tyr
    450                 455                 460

Leu Ala Met Ser Phe Pro Arg Tyr Met Ala Ser Ala Phe Ala Gly Asn
465                 470                 475                 480

Asp Leu Phe Arg Ser Phe Ser Ala Ser Val Phe Pro Leu Phe Gly His
                485                 490                 495

Ala Leu Tyr Ala Asn Leu Gly Ser Glu Lys Phe Pro Val Gly Trp Gly
            500                 505                 510

Ser Ser Val Leu Gly Phe Ile Thr Val Ala Met Ile Ala Ile Pro Val
        515                 520                 525

Thr Phe Met Arg Tyr Gly Pro Arg Leu Arg Ala Asn Ser Arg Tyr Ala
    530                 535                 540

Gly Pro
545
```

<210> 68
<211> 355
<212> PRT
<213> Komagataella pastoris

<400> 68

```
Met Thr Asp Tyr Val Thr Ser Lys Arg Pro Asp Asn Val Leu Asn Trp
1               5                   10                  15

Thr Ser Ile His Val Ser Ser Trp Ile Gly Glu Thr Ile Pro Glu Ile
            20                  25                  30

Asp Pro Ser Leu Leu Gln Asn Phe Leu Glu His Asp Ile Ala Gly Asp
        35                  40                  45

Val Leu Pro Tyr Leu Lys Ser Glu Asp Leu Lys Glu Ile Gly Ile Asn
    50                  55                  60

Glu Leu Lys His Arg Ile Ser Ile Lys Lys Asn Ile His Glu Leu Leu
65                  70                  75                  80

Val Ser Asn Glu Lys His Ile Asp Thr Ser Ile Leu Ser Asp Thr Ala
                85                  90                  95

Thr Glu Leu Gly Thr Leu Ile Leu Thr Asn Lys Phe Ile Thr Gln Met
            100                 105                 110
```

```
Ala Asn Arg Lys Asn Val Val Asp Asp Ser Thr His His Ser Asn Asn
        115                 120                 125

Arg Arg Leu Thr Glu Gln Phe Asn Lys Leu Arg Lys Asp Leu Leu Pro
    130                 135                 140

Ile Phe Lys Trp Ile Lys Glu Thr Gln Pro Leu Pro Thr Pro Glu Asn
145                 150                 155                 160

Thr His Phe Ala Asn Met Gly Ser Val Pro Ala Ser Pro Val Glu His
                165                 170                 175

Thr Ser Gly Glu Ser Thr Leu Ser Asn Pro Ser Leu Ser Thr Ile Asn
            180                 185                 190

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

Lys Pro Thr Leu Ser Ser Arg Arg Gln Ser His Ala Leu Ser Pro Thr
    210                 215                 220

Gly Glu His Leu Asn Val Ser Ser Ser Ser Pro Ser Thr Gly Asn Phe
225                 230                 235                 240

Glu Thr Leu Asn Gly Glu Arg Pro Asn Leu Arg Ser Ala Ser Ser Gly
                245                 250                 255

Ser Gln Glu His Thr Glu Asn Glu Leu Leu Lys Pro Leu Arg Val Lys
            260                 265                 270

Ala Asp Glu Pro Cys Tyr Lys Val Ile Gln Asn Ala Met Lys Arg His
        275                 280                 285

Gly Leu Ser Val Asp Asp Trp Arg Lys Tyr Ala Leu Val Ile Cys Tyr
    290                 295                 300

Gly Asp Glu Glu Arg Val Leu Gly Leu His Glu Lys Pro Gly Ser Ile
305                 310                 315                 320

Phe Lys Glu Leu Lys Asp Gln Lys Gln Asn Pro Ala Ile Met Leu Arg
                325                 330                 335

Gln Ile Asp Thr Asn Asn Asp Asp Gln Asn His Ile Glu Thr Pro Gly
            340                 345                 350

Gly Arg Leu                                                 355
```

<210> 69
<211> 1027
<212> PRT
<213> Komagataella pastoris

<400> 69

```
Met Thr Thr Asn Gly Gln Lys Arg Gln Lys Thr Arg Lys Pro Leu Leu
1               5                  10                  15

Ile Asn Ala Phe Val Met Gly Cys Ala Gly Leu Gln Asn Pro Gly Leu
            20                  25                  30

Trp Lys His Pro Lys Asp Ser Ser His Arg Phe Asn Gln Ile Asp His
        35                  40                  45

Trp Thr Tyr Leu Ala Lys Leu Ala Glu Lys Gly Lys Phe Asn Ala Leu
    50                  55                  60

Phe Ile Ala Asp Val Leu Gly Gly Tyr Asp Val Tyr Lys Gly Pro Glu
65                  70                  75                  80

Asn Leu Ala Thr Pro Ala Val Ala Gly Ala Gln Trp Pro Val Thr Glu
                85                  90                  95

Pro Ser Ala Val Val Ser Ala Met Ala Ala Val Thr Thr Asn Leu Ala
            100                 105                 110

Phe Gly Val Thr Phe Ser Thr Ile Ser Glu Ala Pro Tyr His Phe Ala
        115                 120                 125

Arg Arg Leu Ser Thr Leu Asp His Leu Thr Lys Gly Arg Ile Gly Trp
    130                 135                 140

Asn Val Val Ser Ser Tyr Leu Glu Ser Ala Ala Arg Asn Leu Leu Asn
145                 150                 155                 160

Gly Glu Lys Leu Asp Glu His Asp Gln Arg Tyr Leu Lys Ala Glu Glu
                165                 170                 175

Tyr Ile Gln Ile Val Tyr Glu Leu Leu Leu Ser Ser Trp Arg Asp Asp
            180                 185                 190

Ala Val Val Leu Asp Lys Lys Ala Gly Val Tyr Thr Asp Pro Thr Arg
        195                 200                 205

Phe Arg Lys Ile Asn Phe Glu Gly Lys Phe Phe Lys Val Pro Gly Pro
```

```
    210                 215                 220

His Ile Val Asp Pro Thr Pro Gln Arg Leu Pro Val Ile Leu Gln Ala
225                 230                 235                 240

Gly Thr Ser Lys Val Gly Lys Glu Phe Ala Ala Lys His Ala Glu Ile
                245                 250                 255

Val Phe Val Ile Ser Phe Ser Pro Asp Asp Leu Lys Pro Lys Ile Ala
            260                 265                 270

Glu Val Arg Gln Leu Ala Lys Glu Lys Phe Gly Arg Asn His Asp Asp
        275                 280                 285

Ile Lys Phe Val Ala Leu Ala Thr Pro Val Ile Gly Ala Thr His Glu
    290                 295                 300

Leu Ala Glu Glu Lys Tyr Gln Glu Leu Leu Ser Tyr Gly Asp Ile Glu
305                 310                 315                 320

Gly Ala Gln Ala Leu Phe Gly Gly Trp Thr Gly Ile Asp Leu Ser Gln
                325                 330                 335

Tyr Gly Glu Asp Glu Glu Leu Gly Asn Val Ser Ser Asn Ala Met Arg
            340                 345                 350

Gly Ala Val Gln Asn Trp Thr Lys Ala Ile Pro Asn Glu Lys Arg Trp
        355                 360                 365

Thr Arg Lys Val Ile Ala Lys Gln Ile Thr Val Gly Gly Leu Gly Pro
    370                 375                 380

Ala Phe Val Gly Thr Pro Glu Glu Ile Ala Asp Glu Leu Glu His Trp
385                 390                 395                 400

Ser Asp His Ala Gly Leu Asp Gly Phe Asn Phe Thr Tyr Ala Val Asn
                405                 410                 415

Pro Leu Ser Phe Glu Glu Ile Val Glu Asp Leu Ile Pro Val Leu Gln
            420                 425                 430

Arg Arg Gly Leu Ala Gln Lys Glu Tyr Pro Asn Pro Glu Thr Gly Ser
        435                 440                 445

Thr Phe Arg Lys Asn Leu Phe Gly Thr Asp Phe Val Pro Ser Thr His
    450                 455                 460
```

```
Pro Ala Tyr Asn Leu Arg Trp Arg Ala Gly Val Ser Lys Glu Glu Phe
465                 470                 475                 480

Glu Lys Ser Leu Asn Ala Thr Thr Asn Trp Tyr Ser Ser Phe Ala Arg
                485                 490                 495

Ser Gly Ala Leu Gly Glu Leu His Asn Thr Cys Arg Ile Leu Tyr Leu
            500                 505                 510

Gln Ile Val Lys Tyr Lys Tyr Arg Leu Arg Val Arg Ser Glu Gly Asn
        515                 520                 525

Ser Ile Pro Phe Ala Lys Met Thr Lys Glu Asn Glu Ala Lys Arg Gln
    530                 535                 540

Lys Thr Ser Gln Pro Lys Ala Lys Lys Gln Leu Ile Ile Asn Ala Phe
545                 550                 555                 560

Met Ser Gly Ser Ser Gly Asn Gln Ser Pro Gly Leu Trp Ser Tyr Pro
                565                 570                 575

Gly Asp Lys Ser Thr Glu Tyr Thr Thr Leu Asp Tyr Trp Val Glu Leu
            580                 585                 590

Ala Gln Lys Leu Glu Lys Ala Lys Phe His Ser Ile Phe Ile Ala Asp
        595                 600                 605

Val Leu Gly Gly Tyr Asp Val Tyr Asn Gly Pro Gly Asn Tyr Ser Ala
    610                 615                 620

Ala Ala Lys Ser Gly Ala Gln Phe Pro Met Ile Glu Pro Ser Ala Ala
625                 630                 635                 640

Val Thr Ala Met Ala Ala Ala Thr Lys Ser Ile Thr Phe Gly Val Thr
                645                 650                 655

Phe Ser Thr Ile Ser Glu Ala Pro Tyr His Phe Ala Arg Arg Leu Gly
            660                 665                 670

Thr Leu Asp Leu Leu Thr Asn Gly Arg Val Gly Trp Asn Ile Val Ser
        675                 680                 685

Ser Tyr Leu Asp Ser Ala Ala Arg Asn Leu Leu Asn Gly Glu Pro Leu
    690                 695                 700

Pro Leu His Ala Asp Arg Tyr Lys Arg Ala Glu Glu Phe Leu Gln Val
705                 710                 715                 720
```

```
Val Tyr Arg Leu Phe Leu Ser Ser Trp Arg Asp Asp Ala Tyr Lys Leu
                725                 730                 735

Asp Lys Lys Thr Arg Thr Phe Ala Asp Pro Lys Leu Ile Arg Thr Ile
            740                 745                 750

Asp His Val Gly Glu Phe Phe Asn Val Pro Gly Pro Gln Phe Leu Pro
        755                 760                 765

Pro Thr Pro Gln Arg Leu Pro Leu Ile Leu Gln Ala Gly Thr Ser Lys
    770                 775                 780

Val Gly Met Asp Tyr Ala Ala Lys His Ala Glu Val Val Phe Leu Ala
785                 790                 795                 800

Ser Phe Asp Pro Glu Ser Leu Gln Glu Lys Ile Lys Thr Val Arg Asp
                805                 810                 815

Ile Ala Glu Thr Lys Tyr Asn Arg Pro Arg Asp Ser Ile Lys Phe Leu
            820                 825                 830

Ile Leu Ile Thr Val Val Ile Ala Asp Thr His Glu Asp Ala Val Lys
        835                 840                 845

Arg Tyr Glu Asp Leu Ala Ser Tyr Ala Asp Leu Glu Gly Ala Gln Ala
    850                 855                 860

Leu Phe Ser Gly Trp Thr Gly Ile Asp Ile Gly Lys Tyr Gly Glu Asp
865                 870                 875                 880

Glu Pro Leu Glu His Val Glu Ser Asn Ala Ile Lys Ser His Val Lys
                885                 890                 895

Asn Trp Thr Lys Phe Lys Asp Asn Lys Pro Arg Ala Arg Lys Asp Ile
            900                 905                 910

Ala Lys Gln Ile Gly Val Gly Gly Ser Gly Pro Leu Leu Val Gly Ser
        915                 920                 925

Val Gln Glu Ile Ala Asp Glu Leu Glu Arg Trp Ala Glu Val Ser Asp
    930                 935                 940

Leu Asp Gly Phe Asn Phe Ala Tyr Ala Asp Tyr Pro Gln Thr Phe Asp
945                 950                 955                 960

Asp Ile Ile Glu Lys Leu Leu Pro Glu Leu Asn Lys Arg Gly Val Phe
                965                 970                 975
```

```
Trp Asp Asp Tyr Lys Ile Pro Gly Gly Thr Phe Arg Glu Ser Val Phe
            980                 985                 990

Gly Arg Lys Phe Val Asp Lys Asp  His Pro Ala Tyr Asp  Leu Arg Trp
        995                 1000                1005

Arg Ser  Asp Gln Thr Arg Glu  Glu Phe Glu Lys Lys  Leu Ala Glu
    1010                1015                1020

Leu Glu  Lys Lys
    1025
```

<210> 70
<211> 194
<212> PRT
<213> Komagataella pastoris

<400> 70

```
Met Arg Phe Ser Asn Val Val Leu Thr Ala Ile Ala Ala Ala Gly Val
1               5                   10                  15

Gln Ala Asp Glu Ala Leu Tyr Thr Val Phe Tyr Asn Asp Val Thr Glu
            20                  25                  30

Asn Ala Gln Glu Tyr Leu Ser Tyr Ile Gln Ala Asn Thr Ala Ala Gly
        35                  40                  45

Phe Thr Asp Leu Leu Ser Leu Tyr Thr Glu Leu Ala Thr Tyr Thr Asp
    50                  55                  60

Asp Ser Tyr Thr Ser Ile Phe Thr Glu Glu Asp Phe Pro Ala Ser Glu
65                  70                  75                  80

Leu Ser Ser Phe Val Val Asn Leu Pro Trp Tyr Ser Ser Arg Ile Glu
                85                  90                  95

Pro Gln Val Ala Ala Ala Glu Thr Gly Glu Ser Glu Glu Glu Ser Glu
            100                 105                 110

Thr Gly Glu Ser Glu Glu Glu Ser Glu Thr Gly Glu Glu Thr Glu Thr
        115                 120                 125

Glu Thr Gly Ser Glu Ser Glu Ser Glu Ser Glu Ser Glu Thr Ser Ala
    130                 135                 140

Thr Gly Thr Gly Thr Gly Thr Ser Ala Ser Glu Ser Ala Glu Thr Glu
145                 150                 155                 160
Thr Ser Thr Asp Ala Ala Val Ser Ile Asp His Pro Lys Ser Thr Leu
                165                 170                 175

Leu Met Gly Leu Thr Ala Ala Val Val Ser Ile Thr Phe Gly Val Phe
            180                 185                 190

Ala Leu
```

<210> 71
<211> 478

<212> PRT
<213> Komagataella pastoris

<400> 71

```
Met Ser Ser Phe Arg Val Leu Asp Leu Val Lys Pro Phe Thr Pro Phe
1               5                  10                  15

Leu Pro Glu Val Ile Ser Pro Glu Arg Lys Val Pro Phe Gln Gln Lys
            20                  25                  30

Leu Met Trp Thr Gly Val Thr Leu Leu Ile Phe Leu Val Met Ser Glu
        35                  40                  45

Ile Pro Leu Tyr Gly Ile Thr Ser Ser Asp Ser Ser Asp Pro Leu Phe
    50                  55                  60

Trp Leu Arg Met Met Leu Ala Ser Asn Arg Gly Thr Leu Met Glu Leu
65                  70                  75                  80

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

Gly Ile Gln Ile Leu Asp Val Asn Met Glu Asn Lys Ala Asp Arg Glu
            100                 105                 110

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

Gln Ala Thr Val Tyr Val Leu Thr Gly Met Tyr Gly Pro Pro Gly Glu
    130                 135                 140

Leu Gly Val Gly Val Cys Leu Leu Leu Val Leu Gln Leu Val Phe Ala
145                 150                 155                 160

Gly Ile Val Val Ile Leu Leu Asp Glu Leu Leu Gln Lys Gly Tyr Gly
                165                 170                 175
```

```
Leu Gly Ser Gly Ile Ser Leu Phe Met Ala Thr Asn Ile Cys Glu Gln
            180                 185                 190

Ile Phe Trp Lys Thr Phe Ala Pro Thr Thr Val Asn Arg Gly Arg Gly
        195                 200                 205

Lys Glu Phe Glu Gly Ala Phe Ile Ser Phe Phe His Leu Ile Leu Thr
    210                 215                 220

Lys Lys Asp Lys Lys Arg Ala Leu Leu Glu Ser Phe Tyr Arg Asp Asn
225                 230                 235                 240

Ala Pro Asn Met Phe Gln Val Ile Ala Thr Leu Val Val Phe Phe Thr
                245                 250                 255

Val Val Tyr Leu Gln Gly Phe Arg Leu Glu Ile Pro Val Lys Ser Thr
            260                 265                 270

Arg Gln Arg Gly Pro Tyr Gly Thr Tyr Pro Ile Arg Leu Phe Tyr Thr
        275                 280                 285

Ser Asn Met Pro Ile Met Leu Gln Ser Ala Leu Thr Ser Asn Ile Phe
    290                 295                 300

Ile Ile Ser Gln Met Leu Tyr Ser His Phe Pro Asp Asn Ala Phe Val
305                 310                 315                 320

Lys Leu Ile Gly Thr Trp Glu Ala Gln Pro Gly Ser Ala Gln Leu Phe
                325                 330                 335

Ala Ala Ser Gly Leu Ala Tyr Tyr Met Gln Pro Pro Met Ser Leu Ser
            340                 345                 350

Gln Ala Leu Leu Asp Pro Ile Lys Thr Val Val Tyr Val Val Phe Val
        355                 360                 365

Leu Thr Thr Cys Ala Ile Phe Ser Lys Thr Trp Ile Glu Ile Ser Gly
    370                 375                 380

Ser Ser Pro Arg Asp Val Ala Lys Gln Phe Lys Asp Gln Gly Leu Val
385                 390                 395                 400

Ile Ala Gly His Arg Asp Ala Thr Val Tyr Lys Glu Leu Lys Lys Ile
                405                 410                 415

Ile Pro Thr Ala Ala Ala Phe Gly Gly Ala Thr Ile Gly Ala Leu Ser
```

420 425 430

Val Val Ser Asp Leu Leu Gly Thr Leu Gly Ser Gly Thr Ser Ile Leu
435 440 445

Leu Ala Val Thr Thr Ile Tyr Gly Tyr Tyr Glu Leu Ala Val Lys Glu
450 455 460

Gly Gly Phe Ser Lys Gly Gly Pro Ser Gly Phe Val Asp Leu
465 470 475

<210> 72
<211> 22
<212> ADN
<213> Secuencia Artificial

<220>
<223> cebador

<400> 72 cgaacatcca tcaccaaaac ac 22

<210> 73
<211> 37
<212> ADN
<213> Secuencia Artificial

<220>
<223> cebador

<400> 73 gttgtcgacc tgcagcgtac ggtgttgccg cgaaatg 37

<210> 74
<211> 41
<212> ADN
<213> Secuencia Artificial

<220>
<223> cebador

<400> 74 taggtgatat cagatccact gatcaatttg cccaagagac g 41

<210> 75
<211> 20
<212> ADN
<213> Secuencia Artificial

<220>
<223> cebador

<400> 75 gactgttgcg attgctggtg 20

<210> 76
<211> 37
<212> ADN
<213> Secuencia Artificial

<220>
<223> cebador

<400> 76 catttcgcgg caacaccgta cgctgcaggt cgacaac 37

<210> 77
<211> 23
<212> ADN

<213> Secuencia Artificial

<220>
<223> cebador

<400> 77 cggtgagaat ggcaaaagct tat 23

<210> 78
<211> 21
<212> ADN
<213> Secuencia Artificial

<220>
<223> cebador

<400> 78 aagcccgatg cgccagagtt g 21

<210> 79
<211> 41
<212> ADN
<213> Secuencia Artificial

<220>
<223> cebador

<400> 79 cgtctcttgg gcaaattgat cagtggatct gatatcacct a 41

<210> 80
<211> 21
<212> ADN
<213> Secuencia Artificial

<220>
<223> cebador

<400> 80 atccaggaca cgctcatcaa g 21

<210> 81
<211> 22
<212> ADN
<213> Secuencia Artificial

<220>
<223> cebador

<400> 81 tgtaacgtga atgtcggatt tg 22

<210> 82
<211> 20
<212> ADN
<213> Secuencia Artificial

<220>
<223> cebador

<400> 82 tagtgatggt ggtggtgatg 20

<210> 83
<211> 23
<212> ADN
<213> Secuencia Artificial

<220>
<223> cebador

<400> 83 cctgaggctt tgttccaccc act 23

<210> 84
<211> 26
<212> ADN
<213> Secuencia Artificial

<220>
<223> cebador

<400> 84 ggaacatagt agcaccggca taacga 26

<210> 85
<211> 10
<212> ADN
<213> Secuencia Artificial

<220>
<223> sitio de inicio de la traducción

<400> 85 cctgcaggcc 10

<210> 86
<211> 9
<212> ADN
<213> Secuencia Artificial

<220>
<223> sitio de inicio de la traducción

<220>
<221> característica_misc
<222> (1)..(6)
<223> n en las posiciones 1-6 es cualquier ácido nucleico, preferiblemente A

<220>
<221> característica_misc
<222> (8)..(9)
<223> n en las posiciones 8-9 es cualquier ácido nucleico, preferiblemente A

<400> 86 nnnnnnann 9

REIVINDICACIONES

1. Una célula huésped recombinante que comprende un gen endógeno que codifica una proteína FLO8 que comprende la secuencia de aminoácidos identificada como SEQ ID NO: 1 o un homólogo de la misma, cuya célula huésped está modificada genéticamente mediante una o más modificaciones genéticas para reducir la expresión de dicho gen en comparación con el huésped. antes de dichas una o más modificaciones genéticas, y cuya célula huésped comprende un casete de expresión heterólogo que comprende un gen de interés (GDI) bajo el control de un promotor del casete de expresión (ECP) cuyo ECP es reprimible por una fuente de carbono distinta del metanol y comprende:

(i) cualquiera de SEQ ID NO: 10-16, o

(ii) cualquiera de SEQ ID NO: 41-45, o

(iii) al menos 85% de identidad de secuencia con una región de al menos 300 nt que incluye el extremo 3' de cualquiera de SEQ ID NO: 10-16 o SEQ ID NO: 41-45.

2. La célula huésped de la reivindicación 1, en donde dicho polinucleótido endógeno es un gen que codifica dicha proteína FLO8, en donde dichas una o más modificaciones genéticas comprenden una interrupción, sustitución, deleción o knockoutde (i) uno o más polinucleótidos endógenos, o una parte de los mismos; o (ii) una secuencia de control de la expresión, preferiblemente donde dicha secuencia de control de la expresión se selecciona del grupo que consiste en un promotor, un sitio de unión ribosómico, secuencias de inicio y finalización transcripcionales o traduccionales, una secuencia potenciadora y activadora.

3. La célula huésped de la reivindicación 1 o 2, en donde dicho gen endógeno que codifica dicha proteína FLO8 o dicho homólogo es eliminado por dichas una o más modificaciones genéticas.

4. La célula huésped de una cualquiera de las reivindicaciones 1 a 3, en donde la ECP es inducible en presencia de una cantidad limitante del crecimiento de una fuente de carbono distinta del metanol, preferiblemente en ausencia de metanol; y reprimible en presencia de una cantidad en exceso de una fuente de carbono distinta del metanol que es superior a la cantidad limitante del crecimiento.

5. La célula huésped de la reivindicación 4, en donde la ECP comprende o consiste en la SEQ ID NO:10 o SEQ ID NO:11.

6. La célula huésped de cualquiera de las reivindicaciones 1 a 5, en donde el casete de expresión comprende además una secuencia de nucleótidos que codifica un péptido señal que permite la secreción de una proteína de interés (PDI) que está codificada por el GDI, preferiblemente en donde la secuencia de nucleótidos que codifica el péptido señal se fusiona junto al extremo 5' del GDI.

7. La célula huésped de cualquiera de las reivindicaciones 1 a 6, en donde el GDI codifica una proteína de interés (PDI) que es un péptido o proteína seleccionada del grupo que consiste en una proteína de unión a antígeno, una proteína terapéutica, una enzima, un péptido, un antibiótico de proteína, una proteína de fusión de toxina, un carbohidrato - proteína conjugada, una proteína estructural, una proteína reguladora, un antígeno de vacuna, un factor de crecimiento, una hormona, una citocina, una enzima de proceso y una enzima metabólica.

8. La célula huésped de cualquiera de las reivindicaciones 1 a 7, que es

a) una célula de levadura de un género seleccionado del grupo que consiste en Pichia, Hansenula, Komagataella, Saccharomyces, Kluyveromyces, Candida, Ogataea, Yarrowia y Geotrichum, tal como *Pichia pastoris, Komagataella phaffii, Komagataella pastoris, Komagataella pseudopastoris, Saccharomyces cerevisiae, Ogataea minuta, Kluyveromces lactis, Kluyveromes marxianus, Yarrowia lipolytica* o *Hansenula polimorfa*; o

b) una célula de hongos filamentosos, tal como *Aspergillus awamori* o *Trichoderma reesei.*

9. Un método para aumentar el rendimiento de una proteína de interés (PDI) producida por una célula huésped que expresa un gen de interés (GDI) que codifica dicha PDI bajo el control de un promotor que no es inducible por metanol y es reprimible por una fuente de carbono que no es metanol, reduciendo en dicha expresión en la célula huésped de un gen que codifica una proteína FLO8 que comprende la secuencia de aminoácidos identificada como SEQ ID NO: 1 o un homólogo de la misma.

10. Un método para producir una proteína de interés (PDI) codificada por un gen de interés (GDI) mediante el cultivo de la célula huésped de cualquiera de las reivindicaciones 1 a 8 en condiciones para producir dicha PDI.

11. El método de la reivindicación 9 o 10, que comprende las etapas:

a) cultivar la célula huésped bajo condiciones de crecimiento; y

b) cultivar la célula huésped en condiciones limitantes del crecimiento en presencia de hasta 1 g/L de una segunda fuente de carbono distinta del metanol, lo que da como resultado la expresión de dicho GDI para producir dicha PDI,

preferiblemente en donde se realiza dicha etapa a) de cultivo en una fase discontinua; y dicha etapa b) el cultivo se realiza en una fase de cultivo de alimentación discontinua o continuo.

12. El método de la reivindicación 11, en donde dicha primera o segunda fuente de carbono se selecciona de sacáridos, polioles, alcoholes o mezclas de cualquiera o más de los anteriores.

13. Un método para producir una proteína de interés (PDI) en una célula huésped, que comprende las etapas:

a) diseñar genéticamente la célula huésped para reducir la expresión de un gen endógeno que codifica una proteína FLO8 que comprende la secuencia de aminoácidos identificada como SEQ ID NO: 1 o un homólogo de la misma;

introducir en la célula huésped un casete de expresión heterólogo que comprende un gen de interés (GDI) que expresa dicho PDI bajo el control de un promotor de casete de expresión (ECP) que está unido operativamente al GDI, cuyo ECP es reprimible por una fuente de carbono distinta de metanol y comprende:

(i) cualquiera de SEQ ID NO: 10-16, o

(ii) cualquiera de SEQ ID NO: 41-45, o

(iii) al menos un 85% de identidad de secuencia con una región de al menos 300 nt que incluye el extremo 3' de cualquiera de SEQ ID NO: 10-16 o SEQ ID NO: 41-45;

b) cultivar dicha célula huésped en condiciones para producir dicho PDI;

c) opcionalmente aislar dicho PDI del cultivo celular; y

d) opcionalmente purificar dicho PDI.

Fig. 1:

Secuencia de aminoácido, proteína FLO8 *Komagataella phaffii*, SEQ ID NO:1

>XP_002494151/PP7435_Chr4-0252/PAS_chr4_0711 (Flo8_Komagataella phaffii CBS7435/GS115)
dominio LisH subrayado

MNKPNGSEQQPPSRGMKQESGGPVTSSTTPGTNTGLENSHSMGADMEPDVGATSPRHLLNGYIYDYLVKSNMQNLA
DQFAQETELLETDLTVPMDTPSGYLLEWWMVFWDLFNARLKQRGSQKAHQYIQLNMLRQQQQRTMRNTARVQKVPL
RPHTQSSPSMSQTFIPQQPQQQAQGQQHAQAQAQVQAHQQAQHHAQAQVPVQPQQFQLGGQTQQQQSINTGSPAGP
NAINSPVQHLAQQQMNHLRQQATATTQQPIPQQNIPSNQQGPTGPYPTSPSRRPRLLSNESGASAPSVMTKSQLQG
VPPSQQPHQQQGQQVGPPNQHQGQSSSFYSGMPPQGVVVPHQFNPQQYANMLARQQHVQAQQQVQLQQVQHVQQRQ
QQDQQQHRLSAGSPGHPSFGVFQQPPPMSNHNQVMINQQGETFFDPHSPYAQPNGYPQPQQQQQQQQQQQQQQQPQ
QQQQQQQQCKQQPPPPPPRQPQPQQAMAMAPLPHSTSAAGTPHSSTTPRFSQPGPVYQQPLPASQPQHSPPSSIQQP
ELVPTPGSQHQQIAQPQSQSQHQQSQQSQSSASKIVGIQEYQKELMMLEKQNKQRHDMACKKGSGHFSNFDPIPEH
TPPEPKFNVNVMLPPQNSAVVTKNTPGTSPGTQTQNTAHSTGNTSAGSTPNNVAPVRKKKEPAKKKAKKATEPPTP
TTPQTPIAARTHQNSTGGIPGNNAATKRRKREPLVDQTVSPNLNEASKSTKTGKISSQTDFTGSDNGFLQDFGDGT
GPPTGTDDMEFDFNSFLNNETGEPNSSTIHFDNVFNWGEGTEAGDL

Secuencia de nucleótido, gen flo8 *Komagataella phaffii*, SEQ ID NO:1

>XM_002494106.1 Proteína hipotética GS115 Komagataella phaffii (PAS_chr4_0711), ARNm

ATGAACAAGCCAAACGGGTCTGAACAACAACCACCGTCACGCGGAATGAAGCAAGAGTCAGGAGGCCCAG
TTACTTCATCTACGACGCCGGGTACCAATACTGGCCTAGAAAACTCTCATTCCATGGGGGCGGATATGGA
GCCTGATGTTGGTGCTACCTCTCCTCGCCATCTTCTTAATGGGTACATTTACGATTATTTAGTCAAATCT
AACATGCAAAATTTGGCTGATCAATTTGCCCAAGAGACGGAGCTCTTAGAAACAGACTTGACAGTACCAA
TGGATACGCCTTCAGGCTATCTTCTAGAATGGTGGATGGTATTCTGGGACCTTTTCAATGCCCGCCTAAA
GCAACGGGGTTCACAGAAGGCCCACCAGTATATTCAGTTGAACATGCTACGACAACAGCAACAGAGGACC
ATGCGAAATACAGCCCGTGTTCAAAAAGTCCCGTTGAGGCCACACACCCAATCATCTCCTTCAATGTCAC
AGACTTTTATTCCACAGCAGCCTCAACAGCAAGCACAGGGACAACAGCACGCCCAGGCTCAAGCCCAAGT
GCAAGCACATCAGCAAGCCCAACACCACGCGCAGGCACAAGTGCCAGTGCAACCGCAACAGCACCAGCTA
GGAGGCCAAACTCAACAGCAGCAATCCATTAACACTGGGTCTCCTGCGGGTCCAAATGCTATCAACTCGC
GTGTTCAACACTTAGCACAACAACAGATGAATCACCTTCGCCAGCAGGCGACTGCCACTACGCAACAACC
TATCCCGCAACAGAATATCCCATCAAACCAACAGGGTCCTACAGGCCCTTATCCTACTTCCCCTTCAAGA
AGACCGAGATTACTGTCTAACGAATCGGGTGCAAGTGCACCCTCTGTAATGACAAAGTCACAGCTCCAAG
GAGTCCCTCCCTCACAACAACCACACCAGCAGCAAGGTCAGCAGGTAGGCCCCCCTAATCAACATCAAGG
TCAATCTTCTTCCTTTTATTCGGGCATGCCTCCTCAAGGGGTCGTGGTTCCTCATCAGTTCAATCCTCAG
CAGTATGCCAATATGCTAGCAAGACAACAGCATGTACAAGCTCAACAACAGGTTCAGTTACAACAGGTCC
AACATGTACAACAGAGACAACAGCAAGACCAACAACAACACCGCCTGTCCGCCGGTTCACCGGGGCACCC
TTCATTTGGCGTTTTTCAACAACCTCCTCCGATGTCAAACCATAATCAGGTCATGATCAATCAGCAGGGA
GAAACTTTTTTTGATCCACATTCTCCATATGCTCAACCTAACGGGTACCCCCAGCCACAGCAACAACAAC
AACAACAGCAACAACAACAACAACAGCAGCAACCGCAACAGCAGCAGCAGCAGCAGCAGCAACAGAAGCA
GCAACCACCACCACCACCACGACAGCCTCAGCGCCAACAAGCGATGGCCATGGCTCCTCTGCCTCACTCT
ACTTCTGCCGCCGGTACTCCTCACTCGTCCACCACACCTAGATTCTCGCAACCTGGTCCTGTTTATCAGC
AGCCTTTACCTGCATCTCAACCGCAACATTCTCCGCCTTCTTCTATTCAGCAGCCGGAGCTAGTTCCAAC
TCCAGGGTCACAACATCAGCAAATAGCACAACCACAATCACAGAGCCAACACCAGCAATCGCAACAGTCT
CAATCAAGTGCTTCTAAAATTGTAGGTATACAGGAGTATCAGAAAGAGCTAATGATGCTTGAGAAACAGA
ACAAACAGCGTCATGACATGGCATGTAAGAAGGGAAGCGGGCATTTTTCTAACTTTGATCCAATTCCAGA
GCACACACCGCCCGAACCAAAATTTAATGTGAATGTAATGCTCCCTCCCCAGAACTCTGCAGTGGTCACG
AAGAATACTCCCGGAACTTCACCTGGTACACAAACTCAAAACACTGCACATAGTACTGGTAACACTTCTG
CGGGGTCTACACCAAATAATGTCGCACCTGTACGAAAGAAAAAGGAGCCAGCTAAAAAGAAGGCAAAGAA
AGCTACTGAGCCCCCGACTCCCACTACTCCACAGACTCCAATTGCAGCTAGGACACATCAAAACTCTACA
GGCGGCATTCCTGGTAATAATGCTGCTACTAAGCGACGAAAACGGGAGCCGCTGGTTGATCAAACTGTTT
CACCTAACCTTAACGAAGCTTCCAAGTCAACAAAGACCGGAAAAATTTCATCTCAAACTGACTTTACAGG
TTCTGACAATGGATTCTTACAGGATTTTGGCGATGGAACTGGTCCTCCCACTGGAACCGATGATATGGAA
TTTGATTTTAACAGTTTTCTTAATAACGAAACTGGCGAACCTAATAGTTCAACCATTCATTTTGACAATG
TATTCAATTGGGGAGAAGGTACCGAAGCCGGAGATTTATAG

Fig. 1 (continuación)

Secuencia de aminoácido, proteína FLO8 *Komagataella pastoris*, SEQ ID NO:3

>ANZ77695.1 BA75_04606T0 [Komagataella pastoris] dominio LisH subrayado

MNKPNGSGQQPPSRGMKQDPGGPVTSSTTPGTNTGFENSHSMGADVEPDVGAASPRHILNGYIHDYLVKSNMQNLA
DQFAQESDLLETDLTVPMDTPTGYLLEWWMVFWDLFNARLKQRGSQKAHQYIQLNMLRQQQQRTMRNTARVQKVPL
RPHTQSSPSMSQTFIPQQPQQQAQAQAQQHAQAQAQVQAHQQAQHHAQAQVPMQSQPHQQGGQTQQQQPINTGSPA
GPNAINSRVQHLAQQQMNHLRQQATATTQQPIPQQNIPSNQQGPAGPYPTSPSRRPRLLSNESGASAPSVMTKSQL
QGGPPSQQPHQQQAQQVGPPNQHQGQSSSFYSGMPPQGVVVPHQFNPQQYANMLARQQHVQAQQQVQLQQVPHVQQ
RQQQDQQQHRLSAGSPGHPSFGVFQQPPPMSNHNQVMINQQGETFFDPHSPYAQPNGYPQPQQQQQQQQQQQQQQQ
QQQQQQQQKQQQPPPRQPQRQQAMATAPLPHSTSASGTPHTATTPRFSQPGPVYQQPLPASQPQHSPPTSIQQQEP
IPTPGSQHQQIAQPQSQNQHQQPQQPQASASKMVGIQEYQKELMMLEKQNKQRHDMAIKKGSGHFSNFDPIPEHTQ
TEPKFNVNVMLPPQNSAVATKNTPGTSPGTQTQNTAHSTGNTSAGSTPNNVVPVRKKKEPSKKKSKKATEPPTPTT
PQTPIAARAHQNSTGGISGNNAATKRRKREPLVDQTVSPNLNEASKSTKPGKISSQNDFTGSDNGFLQDFGDGTGP
PTGTDDMEFDFNSFLNNETGEPNSSTIHFDNVFNWGEGTEAGDL

Secuencia de nucleótido, proteína FLO8 *Komagataella pastoris*, SEQ ID NO:4

>CP014587.1:1153190-1155604 cepa ATCC Komagataella pastoris cromosoma 28485 secuencia 4

ATGAACAAGCCAAACGGGTCTGGACAACAACCACCGTCACGCGGAATGAAGCAAGACCCAGGAGGCCCAG
TTACCTCATCTACGACACCCGGTACCAATACTGGCTTTGAAAACTCTCATTCCATGGGAGCGGATGTGGA
GCCTGATGTTGGTGCCGCCTCTCCTCGCCATATTCTTAATGGGTACATTCATGATTACTTAGTCAAATCT
AATATGCAAAATTTGGCCGATCAATTTGCTCAAGAGTCGGATCTCTTAGAAACAGACCTAACGGTACCAA
TGGATACACCTACAGGCTATCTTTTAGAGTGGTGGATGGTATTTTGGGACCTTTTCAATGCCCGCCTAAA
GCAACGAGGTTCACAGAAGGCTCATCAGTATATTCAGTTGAACATGCTACGACAACAACAGCAGAGGACT
ATGCGAAATACAGCCCGTGTTCAGAAAGTCCCGTTGAGACCACACACCCAATCATCTCCTTCAATGTCAC
AGACCTTTATTCCACAGCAGCCTCAACAGCAAGCACAGGCACAGGCACAGCAGCACGCCCAAGCTCAAGC
GCAAGTCCAGGCTCATCAGCAAGCACAGCATCATGCGCAGGCTCAAGTGCCAATGCAATCGCAACCACAT
CAGCAAGGAGGCCAAACTCAACAACAGCAACCCATTAACACTGGGTCTCCTGCGGGGCCAAATGCTATCA
ACTCTCGTGTGCAACACTTAGCGCAACAACAGATGAATCATCTTCGCCAGCAGGCAACTGCCACTACCCA
GCAACCTATCCCGCAACAGAATATTCCATCAAACCAGCAGGGCCCTGCGGGCCCTTATCCTACATCGCCT
TCAAGAAGACCGAGATTACTGTCTAACGAATCGGGTGCAAGTGCACCCTCTGTAATGACGAAGTCACAGC
TCCAAGGAGGTCCTCCATCACAACAACCACACCAACAGCAAGCTCAGCAGGTAGGACCCCCCAATCAACA
TCAAGGCCAGTCCTCTTCCTTTTATTCGGGCATGCCTCCTCAAGGAGTTGTGGTTCCTCATCAGTTCAAT
CCTCAGCAGTATGCCAATATGCTAGCAAGACAACAGCATGTGCAAGCTCAACAACAGGTTCAGTTACAGC
AGGTTCCACATGTGCAACAAAGACAACAGCAAGACCAACAACAACACCGCTTGTCAGCCGGTTCACCAGG
GCATCCTTCATTTGGCGTTTTTCAACAACCTCCTCCGATGTCAAACCATAATCAGGTCATGATCAACCAG
CAGGGAGAAACCTTTTTTGATCCTCATTCACCATATGCTCAACCTAATGGGTACCCCCAGCCACAGCAAC
AACAACAACAGCAGCAACAACAACAACAACAACAACAACAGCAGCAGCAGCAGCAACAACAGAAGCAGCA
ACAGCCACCACCAAGACAGCCTCAGCGCCAACAAGCGATGGCTACGGCTCCTTTGCCTCATTCTACTTCT
GCCTCGGGTACTCCTCACACGGCCACCACACCTAGATTCTCCCAGCCTGGTCCTGTTTATCAGCAGCCTT
TACCTGCATCTCAACCGCAACATTCTCCGCCTACTTCTATTCAGCAACAGGAACCAATTCCAACTCCTGG
GTCACAACATCAGCAAATAGCACAACCGCAATCACAGAATCAACACCAGCAACCGCAGCAACCTCAAGCA
AGTGCTTCTAAAATGGTGGGTATACAGGAGTACCAGAAAGAGTTAATGATGCTTGAGAAGCAGAACAAAC
AACGTCATGACATGGCAATTAAGAAGGGAAGCGGACATTTCTCTAATTTTGATCCAATTCCAGAGCACAC
ACAGACTGAACCAAAATTCAATGTGAACGTAATGCTCCCTCCCCAGAACTCTGCAGTGGCCACGAAGAAT
ACTCCTGGAACTTCTCCTGGTACGCAAACTCAAAACACTGCACACAGTACTGGCAACACTTCTGCTGGGT
CTACACCGAATAATGTTGTACCAGTTCGAAAAAAGAAGGAGCCATCTAAAAAGAAATCAAAGAAGGCTAC
CGAACCTCCAACTCCTACTACTCCACAGACACCAATTGCAGCTAGGGCACATCAAAACTCTACGGGCGGC
ATTTCAGGTAATAATGCTGCTACTAAGCGACGAAAAAGGGAGCCGCTGGTCGACCAAACAGTTTCACCTA
ACCTTAACGAAGCTTCGAAATCAACAAAGCCTGGGAAAATCTCATCTCAAAATGACTTTACAGGTTCTGA
CAATGGATTTTTACAGGATTTTGGCGATGGAACAGGTCCCCCCACTGGAACTGACGATATGGAGTTTGAT
TTTAACAGTTTTCTCAACAACGAAACTGGTGAACCTAATAGTTCAACCATTCATTTTGACAATGTATTTA
ATTGGGGGGAGGGTACTGAAGCCGGGGATCTATAG

Fig. 1 (continuación)

Ortólogo de FLO8 de S. cerevisiae (cepa ATCC 204508/S288c), SEQ ID NO:3

>sp|P40068|FLO8_YEAST activador transcripcional de FLO8 OS=Saccharomyces cerevisiae (cepa TCC 204508/S288c) OX=559292 GN=FLO8 PE=5 SV=2 dominio LisH subrayado

MSYKVNSSYPDSIPPTEQPYMASQYKQDLQSNIAMATNSEQQRQQQQQQQQQQQQWINQPTAENSDLKEKMNCKNT
LNEYIFDFLTKSSLKNTAAAFAQDAHLDRDKGQNPVDGPKSKENNGNQNTFSKVVDTPQGFLYEWWQIFWDIFNTS
SSRGGSEFAQQYYQLVLQEQRQEQIYRSLAVHAARLQHDAERRGEYSNEDIDPMHLAAMMLGNPMAPAVQMRNVNM
NPIPIPMVGNPIVNNFSIPPYNNANPTTGATAVAPTAPPSGDFTNVGPTQNRSQNVTGWPVYNYPMQPTTENPVGN
PCNNNTTNNTTNNKSPVNQPKSLKTMHSTDKPNNVPTSKSTRSRSATSKAKGKVKAGLVAKRRRKNNTATVSAGST
NACSPNITTPGSTTSEPAMVGSRVNKTPRSDIATNFRNQAIIFGEEDIYSNSKSSPSLDGASPSALASKQPTKVRK
NTKKASTSAFPVESTNKLGGNSVVTGKKRSPPNTRVSRRKSTPSVILNADATKDENNMLRTFSNTIAPNIHSAPPT
KTANSLPFPGINLGSFNKPAVSSPLSSVTESCFDPESGKIAGKNGPKRAVNSKVSASSPLSIATPRSGDAQKQRSS
KVPGNVVIKPPHGFSTTNLNITLKNSKIITSQNNTVSQELPNGGNILEAQVGNDSRSSKGNRNTLSTFEEKKPSSN
NQGYDFDALKNSSSLLFPNQAYASNNRTPNENSNVADETSASTNSGDNDNTLIQPSSNVGTTLGPQQTSTNENQNV
HSQNLKFGNIGMVEDQGPDYDLNLLDTNENDFNFINWEG

Ortólogo de FLO8 de S. cerevisiae (CEN.PK113-7D), SEQ ID NO:6

>EIW10903.1 Flo8p [Saccharomyces cerevisiae CEN.PK113-7D] dominio LisH subrayado

MSYKVNSSYPDSIPPTEQPYMASQYKQDLQSNIAMATNSEQQRQQQQWINQPTAENSDLKEKMNCKNTLNEYIFDF
LTKSSLKNTAAAFAQDAHLDRDKGQNPIDGPKSKENNGNQNTFSKVVDTPQGFLYEWWQIFWDIFNTSSSRGGSEF
AQQYYQLVLQEQRQEQIYRSLAAHAARLQHDAERRGEYSNEDIDPMHLAAMMLGNPMAPAVQMRNVNMNPIPIPMV
GNPIVNNFSIPPYNNANPTTGATAVAPTAPPSGDFANVGPTQNRSQNVTGWPVYNYPMQPTTENPVGNPCNNNTTN
NTTNNKSPVNQPKSLKTMHSTDKPNNVPTSKSTRSRSATSKAKGKVKAGLVAKRRRKNNTATVSAGSTNAGSPNIT
TPGSTTSEPAMVGSRVNKTPRSDIATNFPNQAIIFGEEDIYSNSKSSPSLDGASPSALVSKQPTKVRKNTKKASTS
AFPVESANKLGGNSVVTGKKRSPPNTRVLRRKSTPSVILNADATKDENNMLRTFSNTTAPNIHSAPPTKTANSLPF
PGINLGSFNKPAVSSPLSSVTESCFDPESGKIAGKNGPKRAVNSKVSASSPLSIATPPSGDAQEQRSSNVPGNVVI
KPPHGFSTTNLNITLKSSKIITSQNNTAFQELPNGGNILEAQVGNDSRSSKGNRDTLSTPEEKKPSSNNQGYDFDA
LKNSSSLLFPNQAYASNNRTPNENSNVADETSASTNNGDNDNTLIQPSSNVGTTLGPQQTSTNENQNVHSQNLKFG
NIGMVEDQGPDYDLNLLDTNENDFNFINWEG

Ortólogo de FLO8 de Yarrowia lipolytica (CLIB122), SEQ ID NO:7

>XP_503484.1 YALi0E03102p [*Yarrowia lipolytica* CLIB122] dominio LisH subrayado

MPAKSDKELLNAYIYDYLLKHNMHDSAPTFGAEAKVVPNVKKEDDKDLPKPLIPIDAPQGFLYEWWALFWDIYSAR
GSKGGGSVPAQQYVQGTMPLRQEHAARAQLQQQHQAQQHAQAQAAAQVQGQAQGQGQGQNPTQGPQPQGHMGMPGQ
GPHQPGGPFMNGNMMFPPGQMRMGQLPQHLQQQGTGVAGANPSDDSSSPGGTSPAKRQRLSPDMGGQSHPEQQGQH
MMGTPNPNNPVFNSQVMQQLKASQGQMPNLQQQQQAQLQQYSNTLSLAQQRAMMNAKGGPNGQPGQPGQPGQPGQP
GQPGQPGQPGQPLPMGYEGMEPGFMGGNGLILNAQQMRQQQAGSGGAAGSGQLNGNSNALHDYQMQLMLLEQQNKK
RLMVARQEQQEGQPRAEGQAAAAGANPGARMSGQFKRPGSSPVVGNIGDGRRVTPKLPNQPSPLIDANRASPLQSN
FNGQGDFNVVVGPSGQMMRMQQPQQQGGPPQQQGQQQGGPGQQQGGPGQQPQQGQQQGQQGGPQGGQHPFDDPQQL
PQSQNGGPQPGSAPGQLPQTQGGPQRPPSRVSQMPPPGVQGGQRTQPSSPGQASNPGSGGSSGSAPGGTTPTQANK
QLKGKKAEPKKRAKKGNQPVTPKAVSESPTPTTPATPSAASNQSLLSKATNFAANNKQAQAAAHAQAQAQAAAQAQ
SQMHMGGMGGSSSGLELDVNGGMGLDNDSSSFLNDFSTGNEGDVGDMDFDFNSFLNTDDNAAGALKFDSGSAFGWG
EGVEAMNE

Fig. 1 (continuación)

Ortólogo de FLO8 de Ogataea polymorpha, SEQ ID NO:8

>XP_018209309.1 proteína hipotética OGAPODRAFT_77715 [Ogataea polymorpha]
dominio LisH subrayado

MSQNATPQTTEGHMSMNDQQSNSSSGTNNTEHPSSSSVSSVPTVSDSTAVNSRELLNAYVYDFILKSGFTATASAF
FKEANIPVIHSEKRPTNSPSSSATGTSDLPASFMTMDAPQGFLYEWWQIFWDVFNARTQRGGTTNATQYYHYVNLK
QKQDHLMSQQAAAVAAASTVMNGNNTSMSGAPVTSAAQDIGVLMPQQQQQQQQAQMAHPPQQVPMAQQQRVNARMQ
QPQQQQSQSMPMHPQAAAQAQMNTLRQQQIAQQAQAAQVAAQAVSQRGSPSKRQRMDAAGNGSATEINQTSSPNIA
MNSQQQQQQQQHPPMPMPQGMMIPNQFGVPQQYAMFAAAQAAQQGQSQQKFNQYMIPNFQNQPQQPQQQMLMHQQM
HQQSVHMQHDQQHPPPPPPPNVPQHSQTFSNPNDFFHEMPKQNNPGTARVNDTAIKDYEKQLRLMESQNRRRLDVHR
NVSDSKDPNSPGSAQFTEYSAMLNQMPPQHAAASIMQQRASPATKASPTTKSPANGAPPANGKQKKAPRKARKNSS
SVPATPLTPANNQPTPQTNIPPTPQNTTPQQTPQSSAAAASQVMAGKKGIKRKNRGRAIYPIEYSR

Ortólogo de FLO8 de Aspergillus niger (CBS 513.88), SEQ ID NO:9
Aspergillus niger

>XP_001395127.2 proteína de la ruta de proteína quinasa dependiente de AMPc (Som1)
[Aspergillus niger CBS 513.88]
dominio LisH subrayado

MNQMNQMNMAGMNPGAGGPVGGVPMINNGSAAPRNEQNINNIPENMINNLNTYIYDYFLKRGYHDCARALVKDESI
KLNTEPPIKTSPGHRRDADVNGVDGDTMMTDGKDGDKLKIPDDLPRPNLPSEGQQSSFLLDWFSLFWDFFWAQRKK
GNSNDVRSYLTHTQNMMRLREQHQNQLLRQQPLMNGQMGQMNIRRNGMVPPNLQKTVLQNNTTGLSQQQLAQMHKN
QQVQMMQQMQREHSDMDMNGHRPQSPASAENAPSPSKRPRLEGGPMNGQQLAPNGRGQAQGIPGQPTPQALLMQNG
LNRAMNPNQFQAFQQSGPAAQQKQMQGMPNGMMNPANVMANPQTEMVSIPEGQVYPINGDYYGANGQMAQVRTGMQ
TPGGQHGNHALQDYQMQLMLLEQQNKRRLMMARQEQDSMARPDGQPQMPGQQLPPGTSPQGSRAGTSPNPNDQMKR
GTPKMPQTGLPGSPSAADAMAQGRGSPASMNFPGGQMPPEMAGPQFFVKNMADGMAAPNGMRPPSSNPAFSTPQMG
QPIQAGANRMPNGGWQPQQGAQGQPMAPQQSPATQPQSTGTPQERNAMPPPQAPPAPGANVGRTQPPSPQTAAPPT
PQQGNKPAPKKKETKDSRKRPKKGAAAAAAAAQANTAATPSSEAEHPPTPTPSTPITPQHPNSFNKTGANATTSAP
QQPTSAPAPPPLVQQPPDQTQQPFNELSIPDASAFNLDFSALENPDILENFDFDTFLNTDADTAGFGFDPNTSYPT
DGVETGAGDGL

Fig. 1 (continuación)

SEQ ID NO:10 [pG1-3. pGTH1-D1240]

Motivos T: cursiva y subrayado; primera y segunda regiones regulatorias centrales; cursiva, doble subrayado, región central del espaciador ente primera y segunda región central; negrita; sitio de iniciación de la traducción; línea punteada; negrita: región regulatoria repetida

CAAACATTTGCTCCCCCTAGTCTCCAGGGAAATGTAAAATATACTGCTAATAGAAAACAGTAAGACGCTCAGTTGT
CAGGATAATTACGTTCGACTGTAGTAAAACAGGAATCTGTATTGTTAGAAAGAACGAGAGTTTTTTACGGCGCCGC
CATATTGGGCCGTGTGAAAACAGCTTGAAACCCCACTACTTTCAAAGGTTCTGTTGCTATACACGAACCATGTTTA
ACCAACCTCGCTTTTGACTTGACTGAAGTCATCGGTTAACAATCAAGTACCCTAGTCTGTCTGAATGCTCCTTTCC
ATATTCAGTAGGTGTTTCTTGCACTTTTGCATGCACTGCGGAAGAATTAGCCAATAGCGCGTTTCATATGCGCTTT
TACCCCCTCTTTTGTCAAGCGCAAAATGCCTGTAAGATTTGGTGGGGGTGTGAGCCGTTAGCTGAAGTACAACAGG
CTAATTCCCTGAAAAAACTGCAGCTCAGGGATTCCCACTATTTGGTATTCTGATATGTTTTTCCTGATATGCATCA
AAACTCTAATCTAAAACCTGAATCTCCGC*TATTTTTTTTTTTTTTTT*GATGACCCCGTTTTCGTGACAAATTAATTT
CCAACGGGGTCTTGT**CCGGATAAGAGAATTTTGTTTGATTATCCGTTCGG*ATAAATGGACGCCTGCTCCATATTTT***
***TCCGGTT*ATTACCCCACCTGGAAGTGCCCAGAATTTTCCGGGGATTACGGATAATA**CGGTGGTCTGGATTAATTAA
TACGAGATCTCAGGGATTCCCACTATTTGGTATTCTGATATGTTTTTCCTGATATGCATCAAAACTCTAATCTAAA
ACCTGAATCTCCGC*TATTTTTTTTTTTTTTTT*GATGACCCCGTTTTCGTGACAAATTAATTTCCAACGGGGTCTTGT
CCGGATAAGAGAATTTTGTTTGATTATCCGTTCGG*ATAAATGGACGCCTGCTCCATATTTTTCCGGTT*ATTACCCC
ACCTGGAAGTGCCCAGAATTTTCCGGGGATTACGGATAATACGGTGGTCTGGATTAATTAATACGCCAAGTCTTAC
ATTTTGTTGCAGTCTCGTGCGAGTATGTGCAATAATAAACAAGATGAGCCAATTTATTGGATTAGTTGCAGCTTGA
CCCCGCCATAGCTAGGCATAGCCAAGTGCTATGGGTGTTAGATGATGCACTTGGATGCAGTGAGTTTTGGAGTATA
AAAGATCCTTAAAATTCCACCCTT

SEQ ID NO:11 [pG1-D1427. pG1-4 (PG1-D1427-2xTA(T)14]:

Ejemplo que comprende dos motivos T que son (T)n (n=14), cada uno extendido precediendo TA (extendiendo el motivo T y su extremo 5' para convertirse en motivo TA(T)n]

Motivos T: cursiva y subrayado; primera y segunda regiones regulatorias centrales; cursiva, doble subrayado, región central del espaciador ente primera y segunda región central; negrita; sitio de iniciación de la traducción; línea punteada; negrita: región regulatoria repetida

CAAACATTTGCTCCCCCTAGTCTCCAGGGAAATGTAAAATATACTGCTAATAGAAAACAGTAAGACGCTCAGTTGT
CAGGATAATTACGTTCGACTGTAGTAAAACAGGAATCTGTATTGTTAGAAAGAACGAGAGTTTTTTACGGCGCCGC
CATATTGGGCCGTGTGAAAACAGCTTGAAACCCCACTACTTTCAAAGGTTCTGTTGCTATACACGAACCATGTTTA
ACCAACCTCGCTTTTGACTTGACTGAAGTCATCGGTTAACAATCAAGTACCCTAGTCTGTCTGAATGCTCCTTTCC
ATATTCAGTAGGTGTTTCTTGCACTTTTGCATGCACTGCGGAAGAATTAGCCAATAGCGCGTTTCATATGCGCTTT
TACCCCCTCTTTTGTCAAGCGCAAAATGCCTGTAAGATTTGGTGGGGGTGTGAGCCGTTAGCTGAAGTACAACAGG
CTAATTCCCTGAAAAAACTGCAGCTCAGGGATTCCCACTATTTGGTATTCTGATATGTTTTTCCTGATATGCATCA
AAACTCTAATCTAAAACCTGAATCTCCGC*TATTTTTTTTTTTTTTTT*GATGACCCCGTTTTCGTGACAAATTAATTTC
CAACGGGGTCTTGT**CCGGATAAGAGAATTTTGTTTGATTATCCGTTCGG*ATAAATGGACGCCTGCTCCATATTTTT***
***CCGGTT*ATTACCCCACCTGGAAGTGCCCAGAATTTTCCGGGGATTACGGATAATA**CGGTGGTCTGGATTAATTAAT
ACGCCAAGTCTTACATTTTGTTGCAGTCTCGTGCGAGTATGTGCAATAATAAACAAGATGAGCCAATTTATTGGAT
TAGTTGCAGCTTGACCCCGCCATAGCTAGGCATAGCCAAGTGCTATGGGTGTTAGATGATGCACTTGGATGCAGTG
AGTTTTGGAGTATAAAAGATCCTTAAAATTCCACCCTTAGATCTCAGGGATTCCCACTATTTGGTATTCTGATATG
TTTTTCCTGATATGCATCAAAACTCTAATCTAAAACCTGAATCTCCGC*TATTTTTTTTTTTTTTTT*GATGACCCCGTT
TTCGTGACAAATTAATTTCCAACGGGGTCTTGT**CCGGATAAGAGAATTTTGTTTGATTATCCGTTCGG*ATAAATGG***
***ACGCCTGCTCCATATTTTTCCGGTT*ATTACCCCACCTGGAAGTGCCCAGAATTTTCCGGGGATTACGGATAATA**CG
GTGGTCTGGATTAATTAATACGCCAAGTCTTACATTTTGTTGCAGTCTCGTGCGAGTATGTGCAATAATAAACAAG
ATGAGCCAATTTATTGGATTAGTTGCAGCTTGACCCCGCCATAGCTAGGCATAGCCAAGTGCTATGGGTGTTAGAT
GATGCACTTGGATGCAGTGAGTTTTGGAGTATAAAAGATCCTTAAAATTCCACCCTT

Fig. 1 (continuación)

pG1, SEQ ID NO:12

CAAACATTTGCTCCCCCTAGTCTCCAGGGAAATGTAAAATATACTGCTAATAGAAAACAGTAAGACGCTCAGTTGT
CAGGATAATTACGTTCGACTGTAGTAAAACAGGAATCTGTATTGTTAGAAAGAACGAGAGTTTTTTACGGCGCCGC
CATATTGGGCCGTGTGAAAACAGCTTGAAACCCCACTACTTTCAAAGGTTCTGTTGCTATACACGAACCATGTTTA
ACCAACCTCGCTTTTGACTTGACTGAAGTCATCGGTTAACAATCAAGTACCCTAGTCTGTCTGAATGCTCCTTTCC
ATATTCAGTAGGTGTTTCTTGCACTTTTGCATGCACTGCGGAAGAATTAGCCAATAGCGCGTTTCATATGCGCTTT
TACCCCCTCTTTTGTCAAGCGCAAAATGCCTGTAAGATTTGGTGGGGGTGTGAGCCGTTAGCTGAAGTACAACAGG
CTAATTCCCTGAAAAAACTGCAGATAGACTTCAAGATCTCAGGGATTCCCACTATTTGGTATTCTGATATGTTTTT
CCTGATATGCATCAAAACTCTAATCTAAAACCTGAATCTCCGCTATTTTTTTTTTTTTTnnGATGACCCCGTTTTC
GTGACAAATTAATTTCCAACGGGGTCTTGTCCGGATAAGAGAATTTTGTTTGATTATCCGTTCGGATAAATGGACG
CCTGCTCCATATTTTTCCGGTTATTACCCCACCTGGAAGTGCCCAGAATTTTCCGGGGATTACGGATAATACGGTG
GTCTGGATTAATTAATACGCCAAGTCTTACATTTTGTTGCAGTCTCGTGCGAGTATGTGCAATAATAAACAAGATG
AGCCAATTTATTGGATTAGTTGCAGCTTGACCCCGCCATAGCTAGGCATAGCCAAGTGCTATGGGTGTTAGATGAT
GCACTTGGATGCAGTGAGTTTTGGAGTATAAAAGATCCTTAAAATTCCACCCTT

en donde en posiciones 592 y 593 es T o está ausente

pG3, SEQ ID NO:13

GTAAATAGCGGCAGCAATCCAGTAACCTTTTCTGAATAGCAGAGCCTTAACTAAAATAATGGCCAGGGTAAAAAAT
TCGAAATTTGACACCAAAAATAAAGACTTGTCGTTATAAGTCTTAACAAAGTCCGCAATTTTGGAGCTAACGGTGG
CGGTTGCTGGGATATTCAATAATGGTAGAATGTTGCTGCGGGTATATGACAGAGCGTGAAACACACTGAACAAGGT
AAATGGAACAACAGCAATTGCAATATGGGGGAGGATAGTCAAGAACAAAGCAGCAATGGCAAAGTACTGAATATTC
TCCAAAGCCAAAAGGTCCAGTGGTTTCAACGACAAAGTCTTGTTGGTATAGCTTTGGAACAAAAGGACACCGAAAG
ACTCGACAGCGCCCACAAATACAGCGTTGTAGAAGAACGAATTGATTGCTCCAGAGCTTCTAATAGTCAGAAGATA
CCCCAAACCTCCGAGCAACGTTAGCACATGACCTAAGAACCAGGCGAAGTGAAGAGTCTGGAATAACGACACCCAG
TCAGTTTTTCCTGAGCTCCTGGTGGGATTGGTAGAAGCATTTGATTTGCTTGGAGTGGTTTTATTTGAAGATGGTG
TTGAAGCCATTGTTGCTAAAGAGTCGGAGTTTTGCTTTTAGGGTTTGTTAAGCAAAGGAGGAAAAACTGCGCCGTT
TGAAGTCCCAGGTAGTTTCGCGTGTGAGGCCAGCCAGGGAAAGCTTCCTTCGGTACTTTTTTTTCTTTTGCAGGTT
CCGGACGGATTAAGCTTCGGGTTATGAGGGGGGCGGTAGCCAATTCCGGACACAATATTGCGTCGCAGCTAGTCAC
CCCGCCATAAATATACGCAGGATTGAGGTAATAACATCGATAGTCTTAGTAATTAATACAATTCAGTGGCGAATTT
GGCAACATGACGTAAGGCCCACTGTTGTCTATAAAAGGGGATGAATTTTCATGTTTTTGAGGCCTCCCGGACAATT
TATTGAACTCAA

pG4, SEQ ID NO:14

TGGACTGTTCAATTTGAAGTCGATGCTGACGATGTCAAGAGAGATGCTCAATTATATTTGTCATTTGCTGGTTACA
CTGGAAACGCTACTTTTGTTGGCGGAAACTCTACCAGTTTGGCCGTCCATGTAAACGATGTCGTTCTGGGCCGTGA
CCGTTTCAACACGAACATAACCAATGACAAATCCACTTACAGGTCTAGTTCATATGGAGGCAATTGGTACCTTACT
TCTTTGGATGTCCCAAGTGGGGCTTTAACGTCTGGTACTAACAATGTCTCGTTTGTCACTACAAACTCCGAGGTAA
ATAAAGGATTCTTGTGGGATTCTCTCAAGTTTGTTTGGAAGTTGTAACAGGTTTATAAGCATATCGTGCGCTTGTC
CACAATTGAATCATTTATTGTTGCGAGATACATGAACAAAGTGTGAACTGGGACCCATTACTACAATTCCCACGCA
ACCGTTGTTTCAAAGCCCATATTTTTTGACAATTGTTTCGTTACACCCCCAGTTTGATGTACATCGCTTGCAATGA
TGTGTGTCCCGGAGTATTTTCCATATTCAGCTTGAATTCGTATACTCAACCAATATCTGGGGGTATACTTTTATGT
AACCTATACAAATCAACTATACTATTTCACCTTTCGACCAATCATCTCCCATCTTGTTAAGTTTTGCTTCCTATAT
CCCTGACCCTGACATCACCCATGATTCCGCTCAACGGTTCTCCTCTACATCGTCCCTCTTTTGGAGAGGGTGTTCA
GTTTGACATTCAAATTACCCCCCGCCATCACGCGCAACCGAGACCGCACCCCCGAATTTTCACAAATTACCCCACA
CCCTATACTCCACCACTATGAGGGTTATTAGAACTGATCACGTATAAATACCACCGCAAGTTCCCAAGGGATCGTG
TTCTTCTTCTCCAATTGCAATCATATTTCTGACTCTTTCTAGTTCAGATTAATTCCTTTACACTTGCTTTTTTCCC
TTACCTTTATCC

Fig. 1 (continuación)

pG6, SEQ ID NO:15

AGACCAGCAGTTTAACTACGCAAATCCACAGGAATTTCTACATCACAATACCAATGGTAATACCACGACGTCAAGG
AATGGAAACGACGACTTGGAGGAAGACTTCGTCAACCTCTTGCGGAGTACCCGAGGCTAAGACAATAAGAAGAAAA
AAAAAAGAAAAGCGGTGGGGGAGGGATTATTAAATAAGGATTATGTAACCCCAGGGTACCGTTCTATACATATTTA
AGGATTATTTAGGACAATCGATGAAATCGGCATCAAACTGGATGGGAGTATAGTGTCCGGATAATCGGATAAATCA
TCTTGCGAGGAGCCGCTTGGTTGGTTGGTGAGAGGAGTGAAATATGTGTCTCCTCACCCAAGAATCGCGATATCAG
CACCCTGTGGGGGACACTATTGGCCTCCCTCCCAAACCTTCGATGTGGTAGTGCTTTATTATATTGATTACATTGA
TTACATAGCTAAACCCTGCCTGGTTGCAAGTTGAGCTCCGAATTCCAATATTAGTAAAATGCCTGCAAGATAACCT
CGGTATGGCGTCCGACCCCGCTTAATTATTTTAACTCCTTTCCAACGAGGACTTCGTAATTTTTGATTAGGGAGTT
GAGAAACGGGGGGTCTTGATACCTCCTCGATTTCAGATCCCACCCCCTCTCAGTCCCAAGTGGGACCCCCCTCGGC
CGTGAAATGCGCGCACTTTAGTTTTTTTCGCATGTAAACGCCGGTGTCCGTCAATTAAAAGTCGCAGACTAGGGTG
AACTTTACCATTTTTGTCGCACTCCGTCTCCTCGGAATAGGGGTGTAGTAATTCTGCAGTAGTGCAATTTTTACCC
CGCCAAGGGGGGGCGAAAAGAGACGACCTCATCACGCATTCTCCAGTCGCTCTCTACGCCTACAGCACCGACGTAG
TTAACTTTCTCCCATATATAAAGCAATTGCCATTCCCCTGAAAACTTTAACCTCTGCTTTTTCTTGATTTTTCCTT
GCCCAAAGAAAAG

pG8, SEQ ID NO:16

CTGCACAACCATTGCCAGTAAGGACGAAGAGAAGGCCCCACTACCCAAAATTCAGGATAACGTCTTCATACCATGC
AGCGACGCCTACAAGACGCTGTCAAGACATGCCAACTTCAACGAAGTGAACTTTAACACATTGATCGGGAAATTGA
CCACCAAGGGAATGCTGGTTGAGGCTGGAAGCGTTGCCAGTGTCCTGAGGGAACTGGACCGAAAGTTTAGTAATGC
ATAAGAGGATATATATAGGAATGCAGTAATAATATTAGTACCCATTAAGTGGGCTAAGCCATTGGAAGGCCGTCTG
ACTGATGGTGGTGTTCTTCTCATTTAGATAGTGCATTTGCAACTACCGTCTGAGATTGAGTTTGATGTGAAGCTCC
AGCGCCAAAACAGTATAAGAACCTTATCTCCGCATTATTGTTCTTGCGTAAAAGTTTGTGTGAAGAAACAGGGGTA
GTTGCGCAGATTAGTTGTAATATGCGCATAGGATGGGTCATTGACTTCTTTCCTCGAAAGAGCCACACCGTTAGCT
AAAAAAGGACGCGCATCTACCCCAAAATAGAATGTGGGGAAATAGGACGCGCAACTTCCTCTCAATCACTGGACGT
CAGAAAAACAAATGCGCAATCGAGTCACCCTCCGTGATACCCTCCGTGATACCCCCTCTCCGTCTATTCTGACAGC
GTCTCCCCATGACGTTTCAATCTACTTAGAAAAGATTTCGTTTTTTTTTCCTTCAATTACACGATCTCATCTTCTG
CAAGGGTCTGGAGGACATCACCAATCTGCGACTCCATAACTTAGTCCTGAGTTTATATTTACGCTTCATCTGATGA
GTAGGAAGAAAAAGTTTCACGAAATTCCCCCGCCAACTTGCCCTTCGGAATAAGCAGCCACTCTCCTTCTGCCCAT
AGTAAGCTTGCGCGAGGCCCCAACTTGGCCAGAAACTTTAAATATGCCAAACAATCTCCCCCAATCTAAGTTCTCC
CTCTTCTAAAAA

Primera región regulatoria central, SEQ ID NO: 17

ATAAATGGA

Segunda región regulatoria central, SEQ ID NO: 18

CATATTTTTCCGGTT

Fig. 1 (continuación)

Motivo T identificado por cualquiera de las SEQ:19-34

- $TA(T)_{13}$: SEQ ID NO:19: *TATTTTTTTTTTTT*
- $(T)_{15}$: SEQ ID NO:20 *TTTTTTTTTTTTTT*
- $TA(T)_{14}$: SEQ ID NO:21: *TATTTTTTTTTTTTT*
- $(T)_{16}$: SEQ ID NO:22: *TTTTTTTTTTTTTTT*
- $TA(T)_{15}$: SEQ ID NO:23: *TATTTTTTTTTTTTTT*
- $(T)_{17}$: SEQ ID NO:24: *TTTTTTTTTTTTTTTT*
- $TA(T)_{16}$: SEQ ID NO:25: *TATTTTTTTTTTTTTTT*
- $(T)_{18}$: SEQ ID NO:26: *TTTTTTTTTTTTTTTTT*
- $TA(T)_{17}$: SEQ ID NO:27: *TATTTTTTTTTTTTTTTT*
- $(T)_{19}$: SEQ ID NO:28: *TTTTTTTTTTTTTTTTT*
- $TA(T)_{18}$: SEQ ID NO:29: *TATTTTTTTTTTTTTTTTT*
- $(T)_{20}$: SEQ ID NO:30: *TTTTTTTTTTTTTTTTTT*
- $TA(T)_{19}$: SEQ ID NO:31: *TATTTTTTTTTTTTTTTTTT*
- $(T)_{21}$: SEQ ID NO:32: *TTTTTTTTTTTTTTTTTTT*
- $TA(T)_{20}$: SEQ ID NO:33: *TATTTTTTTTTTTTTTTTTTT*
- $(T)_{22}$: SEQ ID NO:34: *TTTTTTTTTTTTTTTTTTTT*

Región regulatoria principal: (SEQ ID NO:35): ATAAATGGACGCCTGCTCCATATTTTTCCGGTT

(incluyendo la primera y segunda región central, SEQ ID NO: 17 y 18, subrayada)

Región central espaciadora: (SEQ ID NO:36): CGCCTGCTC

Región principal espaciadora: (SEQ ID NO: 37) [incl. un motivo T16, cursiva y subrayado]

ATTACCCCACCTGGAAGTGCCCAGAATTTTCCGGGGATTACGGATAATACGGTGGTCTGGATTAATTAATACGA
GATCTCAGGGATTCCCACTATTTGGTATTCTGATATGTTTTTCCTGATATGCATCAAAACTCTAATCTAAAACCTGA
ATCTCCGC*TTTTTTTTTTTTTTTT*GATGACCCCGTTTTCGTGACAAATTAATTTCCAACGGGGTCTTGT**CCGGATAA**
GAGAATTTTGTTTGATTATCCGTTCGG

Sitio de iniciación de traducción ilustrativo (de pG1): (SEQ ID NO:38): TTCCACCCTT

Sitio de iniciación de traducción ilustrativo (de pAOX1): (SEQ ID NO:39): TTCGAAACG

Fig. 1 (continuación)

Sitio de iniciación de traducción ilustrativo (secuencia condenso de eucariotas):

(SEQ ID NO:40):

gccgccRcc

en donde R en la posición 7 es A o G, preferiblemente A

[letras mayúsculas indican bases altamente conservadas; R=purina 8ª o G, con A siendo más frecuente aquí), véase ta,bien Kozak M (1987). "An analysisi dof 5'-noncoding sequences from 699 vertebrate Messenger RNAs". Nucleic Acids Res. 15 (20): 8125-8148.]

Sitio de iniciación de traducción ilustrativo (p. ej., utilizando en vectores pPUZZLE):

(SEQ ID NO:85): CCTGCAGGCC

Sitio de iniciación de traducción ilustrativo (secuencia condenso de *P. pastoris*):

(SEQ ID NO:86):

nnnnnnAnn

en donde

n en posiciones 1-6 es cualquier ácido nucleico, preferiblemente A,

n en posiciones 8-9 es cualquier ácido nucleico, preferiblemente A

pG1a (SEQ ID NO:41)

GGAATCTGTATTGTTAGAAAGAACGAGAGTTTTTTACGGCGCCGCCATATTGGGCCGTGTGAAAACAGCTTGAAAC
CCCACTACTTTCAAAGGTTCTGTTGCTATACACGAACCATGTTTAACCAACCTCGCTTTTGACTTGACTGAAGTCA
TCGGTTAACAATCAAGTACCCTAGTCTGTCTGAATGCTCCTTTCCATATTCAGTAGGTGTTTCTTGCACTTTTGCA
TGCACTGCGGAAGAATTAGCCAATAGCGCGTTTCATATGCGCTTTTACCCCCTCTTTTGTCAAGCGCAAAATGCCT
GTAAGATTTGGTGGGGGTGTGAGCCGTTAGCTGAAGTACAACAGGCTAATTCCCTGAAAAAACTGCAGATAGACTT
CAAGATCTCAGGGATTCCCACTATTTGGTATTCTGATATGTTTTTCCTGATATGCATCAAAACTCTAATCTAAAAC
CTGAATCTCCGCTATTTTTTTTTTTTTTTTGATGACCCCGTTTTCGTGACAAATTAATTTCCAACGGGGTCTTGTC
CGGATAAGAGAATTTTGTTTGATTATCCGTTCGGATAAATGGACGCCTGCTCCATATTTTTCCGGTTATTACCCCA
CCTGGAAGTGCCCAGAATTTTCCGGGGATTACGGATAATACGGTGGTCTGGATTAATTAATACGCCAAGTCTTACA
TTTTGTTGCAGTCTCGTGCGAGTATGTGCAATAATAAACAAGATGAGCCAAATTATTGGATTAGTTGCAGCTTGAC
CCCGCCATAGCTAGGCATAGCCAAGTGCTATGGGTGTTAGATGATGCACTTGGATGCAGTGAGTTTTGGAGTATAA
AAGATCCTTAAAATTCCACCCTT

pG1b (SEQ ID NO:42)

CCATATTCAGTAGGTGTTTCTTGCACTTTTGCATGCACTGCGGAAGAATTAGCCAATAGCGCGTTTCATATGCGCT
TTTACCCCCTCTTTTGTCAAGCGCAAAATGCCTGTAAGATTTGGTGGGGGTGTGAGCCGTTAGCTGAAGTACAACA
GGCTAATTCCCTGAAAAAACTGCAGATAGACTTCAAGATCTCAGGGATTCCCACTATTTGGTATTCTGATATGTTT
TTCCTGATATGCATCAAAACTCTAATCTAAAACCTGAATCTCCGCTATTTTTTTTTTTTTTTGATGACCCCGTTT
TCGTGACAAATTAATTTCCAACGGGGTCTTGTCCGGATAAGAGAATTTTGTTTGATTATCCGTTCGGATAAATGGA
CGCCTGCTCCATATTTTTCCGGTTATTACCCCACCTGGAAGTGCCCAGAATTTTCCGGGGATTACGGATAATACGG
TGGTCTGGATTAATTAATACGCCAAGTCTTACATTTTGTTGCAGTCTCGTGCGAGTATGTGCAATAATAAACAAGA
TGAGCCAATTTATTGGATTAGTTGCAGCTTGACCCCGCCATAGCTAGGCATAGCCAAGTGCTATGGGTGTTAGATG
ATGCACTTGGATGCAGTGAGTTTTGGAGTATAAAAGATCCTTAAAATTCCACCCTT

Fig. 1 (continuación)

pG1c (SEQ ID NO:43)

CTGCAGATAGACTTCAAGATCTCAGGGATTCCCACTATTTGGTATTCTGATATGTTTTTCCTGATATGCATCAAAA
CTCTAATCTAAAACCTGAATCTCCGCTATTTTTTTTTTTTTTTGATGACCCCGTTTTCGTGACAAATTAATTTCC
AACGGGGTCTTGTCCGGATAAGAGAATTTTGTTTGATTATCCGTTCGGATAAATGGACGCCTGCTCCATATTTTTC
CGGTTATTACCCCACCTGGAAGTGCCCAGAATTTTCCGGGGATTACGGATAATACGGTGGTCTGGATTAATTAATA
CGCCAAGTCTTACATTTTGTTGCAGTCTCGTGCGAGTATGTGCAATAATAAACAAGATGAGCCAATTTATTGGATT
AGTTGCAGCTTGACCCCGCCATAGCTAGGCATAGCCAAGTGCTATGGGTGTTAGATGATGCACTTGGATGCAGTGA
GTTTTGGAGTATAAAAGATCCTTAAAATTCCACCCTT

pG1d (SEQ ID NO:44)

GACCCCGTTTTCGTGACAAATTAATTTCCAACGGGGTCTTGTCCGGATAAGAGAATTTTGTTTGATTATCCGTTCG
GATAAATGGACGCCTGCTCCATATTTTTCCGGTTATTACCCCACCTGGAAGTGCCCAGAATTTTCCGGGGATTACG
GATAATACGGTGGTCTGGATTAATTAATACGCCAAGTCTTACATTTTGTTGCAGTCTCGTGCGAGTATGTGCAATA
ATAAACAAGATGAGCCAATTTATTGGATTAGTTGCAGCTTGACCCCGCCATAGCTAGGCATAGCCAAGTGCTATGG
GTGTTAGATGATGCACTTGGATGCAGTGAGTTTTGGAGTATAAAAGATCCTTAAAATTCCACCCTT

pG1e (SEQ ID NO:45)

CCGGATAAGAGAATTTTGTTTGATTATCCGTTCGGATAAATGGACGCCTGCTCCATATTTTTCCGGTTATTACCCC
ACCTGGAAGTGCCCAGAATTTTCCGGGGATTACGGATAATACGGTGGTCTGGATTAATTAATACGCCAAGTCTTAC
ATTTTGTTGCAGTCTCGTGCGAGTATGTGCAATAATAAACAAGATGAGCCAATTTATTGGATTAGTTGCAGCTTGA
CCCCGCCATAGCTAGGCATAGCCAAGTGCTATGGGTGTTAGATGATGCACTTGGATGCAGTGAGTTTTGGAGTATA
AAAGATCCTTAAAATTCCACCCTT

Secuencia de promotor pGAP nativo de p. pastoris (GS115): SEQ ID NO: 46

CTGCTACTCTGGTCCCAAGTGAACCACCTTTTGGACCCTATTGACCGGACCTTAACTTGCCAAACCTAAACGCTTA
ATGCCTCAGACGTTTTAATGCCTCTCAACACCTCCAAGGTTGCTTTCTTGAGCATGCCTACTAGGAACTTTAACGA
ACTGTGGGGTTGCAGACAGTTTCAGGCGTGTCCCGACCAATATGGCCTACTAGACTCTCTGAAAAATCACAGTTTT
CCAGTAGTTCCGATCAAATTACCATCGAAATGGTCCCATAAACGGACATTTGACATCCGTTCCTGAATTATAGTCT
TCCACCGTGGATCATGGTGTTCCTTTTTTTCCCAAAGAATATCAGCATCCCTTAACTACGTTAGGTCAGTGATGAC
AATGGACCAAATTGTTGCAAGGTTTTTCTTTTTCTTTCATCGGCACATTTCAGCCTCACATGCGACTATTATCGAT
CAATGAAATCCATCAAGATTGAAATCTTAAAATTGCCCCTTTCACTTGACAGGATCCTTTTTTGTAGAAATGTCTT
GGTGTCCTCGTCCAATCAGGTAGCCATCTCTGAAATATCTGGCTCCGTTGCAACTCCGAACGACCTGCTGGCAACG
TAAAATTCTCCGGGGTAAAACTTAAATGTGGAGTAATGGAACCAGAAACGTCTCTTCCCTTCTCTCTCCTTCCACC
GCCCGTTACCGTCCCTAGGAAATTTTACTCTGCTGGAGAGCTTCTTCTACGGCCCCCTTGCAGCAATGCTCTTCCC
AGCATTACGTTGCGGGTAAAACGGAGGTCGTGTACCCGACCTAGCAGCCCAGGGATGGAAAAGTCCCGGCCGTCGC
TGGCAATAATAGCGGGCGGACGCATGTCATGAGATTATTGGAAACCACCAGAATCGAATATAAAAGGCGAACACCT
TTCCCAATTTTGGTTTCTCCTGACCCAAAGACTTTAAATTTAATTTATTTGTCCCTATTTCAATCAATTGAACAAC
TATCAAAACACA

Fig. 1 (continuación)

Rgt1 de *P. Pastoris* (PAS_chr1-3_0233): SEQ ID NO:47

MIPTIDPKDPELVSEDTAQSASARKRSKVSRACDECRRKKIKCDATFLANSNTLLKPCTNCYKYNCSCSFTRVPLK
RGPSKGFARDGSGYERRRSSSVHSVSSSQSVTSPVPSHASLPIPPANPVSLPRLNVPGDGLLSPKAVPPTNLFWKV
PYELPSFSDRRSSVASADSFRRPSIYQSDSEDDFYSATGSQRNSISQAPRQRNLSPALSVSSTSSLNNRIKSLNMV
ASTLESNIHNYYSQGFNSSLPILPLDERILSTLLSNVSNGSSSASWDAIRSPILELFDKSILMLLRSYESQFNFND
LLDHVTEMQSIYPRIRSHLLSDELLKLIFLMSGVLTDYALILTGQPYSTGLSITVSVFNDWKTYENVQRVLVINRA
GSLDLDYDSLPFLFARCYLSLATLDLIYSLSFSSPRLISSFANLPILDIVQKCGITKDAKLDETPLPVLDQFLNCF
LPGDTYPTALNTLKTGLVLLDFTNNRSTTLRFPFINIHDDNHMTGLSHLLSNVSDFMSQFTEVHSDSKDSQLLFLR
CIWAFWEIGSVLSELIDHFISSSANSQVGDKDASFFYEHQLKVTTLLGTFSNIASAFLTSSTTAASHPPPSISPFH
IISMVESFKMVQFLNKLIASFISLNEKLEKRELEDELSKCKEELNNLNERFQAVSSVQTLPVVHVLFRDLVFSSNR
LDTQRDRASSVVSATTTTSTATTTATTKKSSFGNLLHSDEENILPTVIDWCKEQKHSAEMFLNKNDLNGWLY

Cat8-1 de *P. Pastoris* (PAS_chr2-1_0757): SEQ ID NO:48

MMPEEQVTSPQRKHQKSKAKTIRAPGSSIERVAQACDRCRSKKTRCDGKRPQCSQCAAVGFECKISDKLSRRAFPR
GYTETLEERIRELEFENKKLHKLIDLKNEQVEIKNRIDQESTLTNENLTLLNKEQEVSHSGNIHHHADGEPCNCAN
SVSARPVSIAGSVDIDTTDLSDEDDSLYSAASYNAKYHQTGTSGPEMVRLSQRYSSGNFNDPLSFEQSNAPGAAAA
ISIQNKMRTQTFVNLANLVAMSIPRTTEETLFIASLLAKICNVHGFQSKAPILTAKSIALLKDKYNYGNDEVFANI
TLKNVNFNKLTSQQSQQFFQSLNLPNQVNLDLFITTFFNTWNNFIPIINRHIFMSSYIKFNKSRETMFTDNSMFGN
EKFGEILLLITTMVMLSQERNNNREAVPSSSYKKDSTPHPHRPDASSQSNVEILQYYDHLIHEFIKSNISDDCSLP
TLESLSLQLLYCLAIGDLTTSYELRGKIITMGQQLRLHRCPSAVLGTNGSKVSQMQQGERRILFWCIYILDTFSAL
ILGVPRLLKDYEIECALPFSNESNNANVKGSIENTTNTVIINNIKLSLAGKVSDCALAVMRYSKVLGNILDSIFQR
SSINNPSVVSKSTNITEETCLLHEHALDLWRRELSPHINVDLDKSPGGVEYERLSDNQLTILFLYYHAKILIYLPL
MANESSQSRSSASYINIQQSTTSILAIANTLATKERNFYFLPLPVNLSREKVRLAFLSAKGSLEYARGGALFQESK
ILLASVINELKIETSIGMLGCLSVPCMEAVDNAMEQIMALPGKVSSVNGSNSEMKRSSSKRKSSPLRQDISGDERK
SHNIEVSDSRTPSVQSSLYPQPNQMHHPNIIKSENNEQMIPENDTPGAINDIFTSHSPPGTVTSMKEEDLPIKVPI
LLQTQQRQIYNNNPNHSLFSQQPGTQVLSGQQMPGPSSTDQQFKRITTPDGLDSLMMQDFGVDASLGLPMLDFDFN
FDFENVQNNYSQSNVSPPNSESVPSSIQGTHSNDPKDSQVSAGSLFGL

Cat8-2 de *P. Pastoris* (PAS_chr4_0540): SEQ ID NO:49

MKENQASNKFNLIKNPITGKPRISQACDRCRIKKIKCDGTLPSCTNCSKIGFVCKISDRLTRSSFPKGYTKNLEQK
LIDMELDRNRLMLELNRIKKEGFDGTNNNIAMASSVSSSENLKSDDSSECQSVTVSLSSTSGPSLSPEPKQDDFRF
RVGMDGSFVLNQFLQSPLMDYIKSLNVLQFNGCANFDQSFNDDPLVLNKYHMNLNRFLNLIFYKLLLPLIHRNSNT
LNEKFAEDNNSLDSLIWKFFTNYNKLIPILEFDSFYKDYLQFIHKYYSNNQVFVDGFRKYFEFSEFEQCFIVKLIL
ILKFTLPVIHDTSVPSEIYRLISMDSLQRLFGNIDFLKPSTDKVSILLLVLHYMVLYESPKSLLDTQDEAQKYDEF
IGNLLSTAVHHITSLRLHIDPRKLQFPRPLPSNGNRLRIKLSWCYKLISKLFRVIYNIDNDSLYSLDDSHLPELQS
ISILHEELDVTIQFNNLLNLIPNNFHSLRDKQSLSKIKTQLLEWHKNFNTEFVEHFNLNDTDSDELSAEKINVLRS
KLISLNRLNCYNSYFQLVIELQLKENLDSVVSGIFGLSNEMLIDNKSSTELLNTLQQTPIIHQSSILVSLCYRIQT
GNLQDEICSILVNNYEKLLQCNDAGLPIKILPQLVHYFKGKISTNLSNSAAHEDLMNMFTLNDNLSTTTTDLDSFI
IPPKRKQDQTLPIGTKRSKSASTSSVISSDDCSLFSNSLSVPTTFSGSSISVGMDNPPSSLFGSYKRPSSIVKQEP
TINPRSNGTNTDSNLFDTFNDSIKGSLNNGLKKLKDIRCNSVVERSHSSQRNDFLMDQEDSITKETINFSELFTCG
TPTASQSIDRSPKSLLLNDLAIAPDTLVIKPDAEDLDRLKNKIRSVKSTVH

Dominio LisH de la proteína FLO8 de *K. phaffii* (SEQ ID NO: 50)

LLNGYIYDYLVKSNMQNLADQFAQET

Dominio LisH de la proteína FLO8 de *K. pastoris* (SEQ ID NO: 51)

LNGYIHDYLVKSNMQNLADQFAQES

BLASTp: 96% de identidad de SEQ ID NO: 50 (23/24 aa)

Fig. 1 (continuación)

Dominio LisH del ortólogo de FLO8 de *S. cerevisae* (cepa ATCC 204508 / S288c) (SEQ ID NO: 52)

LNEYIFDFLTKSSLKNTAAAFAQDA

BLASTp: 54% de identidad de SEQ ID NO: 50 (13/24 aa)

LNEYIFDFLTKSSLKNTAAAFAQDA

Dominio LisH del ortólogo de FLO8 de *S. cerevisae* (CEN.PK113-7D) (SEQ ID NO: 53)

LNEYIFDFLTKSSLKNTAAAFAQDA

BLASTp: 54% de identidad de SEQ ID NO: 50 (13/24 aa)

Dominio LisH del ortólogo de FLO8 de *Yarrowia lipolytica* (CLIB112) (SEQ ID NO: 54)

ELLNAYIYDYLLKHNMHDSARTFGAEA

BLASTp: 71% de identidad de SEQ ID NO: 50 (12/17 aa)

Dominio LisH del ortólogo de FLO8 de *Ogataea polymorpha,* (SEQ ID NO: 55)

ELLNAYVYDFILKSGFTATASAFFKEA

BLASTp: 62% de identidad de SEQ ID NO: 50 (8/13 aa)

Dominio LisH del ortólogo de FLO8 de *Aspergillus niger* (CBC 513.88) (SEQ ID NO: 56)

NNLNTYIYDYFLKRGYHDCARALVKD

BLASTp: 73% de identidad de SEQ ID NO: 50 (8/11 aa)

Secuencia csR (SEQ ID NO: 57)

CAGGAACAACTAATGGAGTCTGGGGGTGGTTTGGTTACCCTGGGTGGTTCTCTTAAGCTTTCATGTAAGGCCTCTG
GTATTGATTTTTCGCACTACGGTATCTCCTGGGTTAGACAAGCTCCTGGAAAAGGTCTGGAATGGATCGCTTACAT
TTACCCAAATTACGGTTCTGTTGACTATGCCTCCTGGGTCAATGGTAGGTTCACTATTTCCCTTGACAACGCTCAG
AACACGGTATTCCTACAGATGATCTCCCTAACCGCTGCTGATACTGCAACCTACTTCTGTGCTCGTGACAGAGGTT
ACTACTCTGGCTCTCGTGGAACTAGACTTGACTTATGGGGACAAGGTACTCTCGTTACCATCTCTAGTGGTGGAGG
TGGTTCTGGAGGAGGAGGTTCCGGCGGAGGTGGTAGCGAGCTGGTCATGACTCAAACCCCTCCATCCCTATCTGCA
TCAGTCGGTGAAACCGTTAGAATTAGATGCCTTGCATCTGAGTTCTTGTTCAACGGTGTGTCCTGGTATCAACAAA
AGCCTGGTAAGCCTCCAAAGTTTCTCATTTCTGGTGCCTCAAACCTCGAATCTGGAGTGCCACCAAGATTTTCCGG
ATCTGGCTCTGGTACTGACTACACTCTGACAATTGGTGGTGTTCAAGCTGAGGATGTTGCTACCTACTATTGTCTC
GGTGGTTACTCAGGATCTTCCGGCCTAACTTTCGGTGCCGGTACAAACGTCGAGATCAAAGGTGGACATCACCACC
ACCATCACTAATAG

Secuencia vHH (SEQ ID NO: 58)

CAGGTTCAGCTGCAGGAGTCCGGTGGTGGTCTGGTTCAAGCCGGTGGTTCATTAAGATTGTCCTGTGCTGCCTCTG
GTAGAACTTTCACTTCTTTCGCAATGGGTTGGTTTAGACAAGCACCTGGAAAAGAGAGAGAGTTTGTTGCTTCTAT
CTCCAGATCCGGTACTTTAACTAGATACGCTGACTCTGCCAAGGGTAGATTCACTATTTCTGTTGACAACGCCAAG
AACACTGTTTCTTTGCAAATGGACAACCTTAACCCAGATGACACCGCAGTCTATTACTGTGCCGCTGACTTGCACA
GACCATACGGTCCAGGAACCCAAAGATCCGATGAGTACGATTCTTGGGGTCAGGGAACTCAAGTCACTGTCTCTTC
AGCTGGTGGATCTGGTGGTGGAGGTTCAGGTGGTGGAGGATCCGGTGGTGGTGGTTCTGGTGGTGGTGGATCTGGT
GGAGGTGAAGTTCAACTTGTCGAATCCGGTGGTGCACTTGTCCAACCTGGTGGATCTCTTAGACTTTCTTGTGCCG
CCTCCGGTTTTCCTGTTAACCGTTACTCTATGCGTTGGTACAGACAAGCCCCTGGAAAAGAACGTGAATGGGTTGC
CGGAATGTCCTCAGCTGGTGACAGATCCTCCTACGAAGATTCTGTGAAGGGACGTTTCACCATCTCCAGAGATGAC
GCCCGTAACACCGTTTACCTTCAAATGAACTCCCTTAAGCCTGAGGATACTGCCGTCTACTATTGTAACGTGAATG
TCGGATTTGAATACTGGGGACAGGGAACCCAAGTTACTGTCTCTTCCGGTGGACATCACCACCACCATCACTAATA
G

Fig. 1 (continuación)

Secuencia señal (SEQ ID NO: 59)

MK**X**STNLILAIAAAS**X**VVSA, en donde

X en posición 3 es tanto F como L; y

X en posición 16 es tanto A como T

Secuencia líder (SEQ ID NO: 60)

MK**X**STNLILAIAAAS**X**VVSAAPVAPAEEAANHLHKR, en donde

X en posición 3 es tanto F como L

X en posición 16 es tanto A como T

Secuencia señal, aMF, *S. cerevisae* (SEQ ID NO: 61):

MRFPSIFTAVLFAASSALAAPVNTTTEDETAQIPAEAVIGYSDLEGDFDVAVLPFSNSTNNGLLFINTTIASIAAK
EEGVSLEKR.

Factor auxiliar 1, *Komagataella pastoris* (SEQ ID NO: 62):

MSSDAVEQLE NFQLIKFDRF DPSTQSTIRI ARSPKPIPVK VVIVGDGGCG
KTCLLNVFAT GTFPEAYVPT IIENVVITLV TPTGQIAAVT LWDTAGQEEY
DRLRPLSYSD VDVVLLCYSI DNLSTFHNVA DKWYPEVAHF CPNTPIILVG
TKSDMRRHQK SQPHFVSPQD SSQLARQMGA VMNIECSAKE VSNVNIVFDA
AVSYCLSNSR PKTRGDNDNN RSNRRLSRAK RASMFIRGKD VSSTSGNSRE
ELVEYDQDGL AIIPDRKKRK CSII

Factor auxiliar 2, *Komagataella pastoris* (SEQ ID NO: 63):

MLNKLFIAIL IVITAVIGET TTSSTTASLS ESPTLVWVTG TDASGRLATT
QSAYTQQFSQ LYSSIASPSS GSIGLGTIQG TVGIVRTYET ITLAS

Factor auxiliar 3, *Komagataella pastoris* (SEQ ID NO: 64):

MSTAIPGGQR TLAKRRAANL DKKQDEPTSA RSAGAGGSSS TMLKLYTDEA
QGLKVDPLIV LVLAVGFIFS VIGLHVVAKL TGKLIN

Factor auxiliar 4, *Komagataella pastoris* (SEQ ID NO: 65):

MTPRSHIFFD ISINNQPAGR IIFELFNDIV PKTAENFRAL STGEKGIGKS
GKPLHYKGST FHRIIKDFMV QGGDFTNGNG TGGESIYGEK FEDENFQLTH
DKPFLLSMAN AGPGTNGSQF FITTVPTPHL DNKHVVFGKV IAGKATVRKI
ERNSEGEAPI EPVVIEDCGE LPEDADLTIS DETGDKYEEV LKDNENIDID
DFEQVYQAIT EIKELGTKYF KNGDTKIAFE KYQKAANYLL EYIPSDLSEE
QSSKLELLKT SVFSNVALAG LKVSKFKDTI KYATLVIEDE SADAKAKSKG
YYRRGSAYSS LKDEDSAISD FQKALELSPG DPAISQSLQR TTKARKDRLA
KEKAALSKFF E

Factor auxiliar 5, *Komagataella pastoris* (SEQ ID NO: 66):

MTNWKAILTP AQYQVLRLGG TERPYTGQYV NFKKNGTYLC SGCQTPLYKS
GTKFDSSCGW PAFYEALPGA VKRIEDNSLG MRRIEIRCSK CDGHLGHVFE
GEGFDTPTDS RHCVNSISLK FQGEEEN

Fig. 1 (continuación)

Factor auxiliar 6, *Komagataella pastoris* (SEQ ID NO: 67):

MSHLLLRDSF WGRTIYHLSK HRYFSFPEEK DGFIAPEKYY LNMDQVSIHA
ESEKNIVEGL VDTSNSSLEE VKTTRVIVDW DEYDQKENPQ NWSSLLKCFV
VFEVGILTVA VYMGSAIYTP GIEDIMRDLN VSRTVATLPL TLFVIGYAVG
PMIFSPMSEH PAIGRTTIYV WTLFIFAILQ IPTALTTNIA GFCILRFIGG
FFASPALATG PASVGDVIAI PHLPVGLGLW SICAVCGPSL GPLFGAIFSQ
LVSWRWCFWF LLITSGTLFI VLGFTLPETY VPTLLYRKAR RLRALTKNEL
IISKGELDIQ DRTAKEVLIE CLWRPVDISF RDPVVLMINL YISMVYSIWY
IWFEAFPIVF LEIYGFSLIG MGASFAGILI GVLICSACYC YACHVTFARR
IIANETIHPE FFVPGAIIGG CIMPTGIFIL GWTATKSVHW IVPIIGSGLF
AAGGYLIFQT LFNYLAMSFP RYMASAFAGN DLFRSFSASV FPLFGHALYA
NLGSEKFPVG WGSSVLGFIT VAMIAIPVTF MRYGPRLRAN SRYAGP

Factor auxiliar 7, *Komagataella pastoris* (SEQ ID NO: 68):

MTDYVTSKRP DNVLNWTSIH VSSWIGETIP EIDPSLLQNF LEHDIAGDVL
PYLKSEDLKE IGINELKHRI SIKKNIHELL VSNEKHIDTS ILSDTATELG
TLILTNKFIT QMANRKNVVD DSTHHSNNRR LTEQFNKLRK DLLPIFKWIK
ETQPLPTPEN THFANMGSVP ASPVEHTSGE STLSNPSLST INAGEGVNSA
VAGQSLGRKP TLSSRRQSHA LSPTGEHLNV SSSSPSTGNF ETLNGERPNL
RSASSGSQEH TENELLKPLR VKADEPCYKV IQNAMKRHGL SVDDWRKYAL
VICYGDEERV LGLHEKPGSI FKELKDQKQN PAIMLRQIDT NNDDQNHIET
PGGRL

Factor auxiliar 8, *Komagataella pastoris* (SEQ ID NO: 69):

MTTNGQKRQK TRKPLLLINAF VMGCAGLQNP GLWKHPKDSS HRFNQIDHWT
YLAKLAEKGK FNALFIADVL GGYDVYKGPE NLATPAVAGA QWPVTEPSAV
VSAMAAVTTN LAFGVTFSTI SEAPYHFARR LSTLDHLTKG RIGWNVVSSY
LESAARNLLN GEKLDEHDQR YLKAEEYIQI VYELLLSSWR DDAVVLDKKA
GVYTDPTRFR KINFEGKFFK VPGPHIVDPT PQRLPVILQA GTSKVGKEFA
AKHAEIVFVI SFSPDDLKPK IAEVRQLAKE KFGRNHDDIK FVALATPVIG
ATHELAEEKY QELLSYGDIE GAQALFGGWT GIDLSQYGED EELGNVSSNA
MRGAVQNWTK AIPNEKRWTR KVIAKQITVG GLGPAFVGTP EEIADELEHW
SDHAGLDGFN FTYAVNPLSF EEIVEDLIPV LQRRGLAQKE YPNPETGSTF
RKNLFGTDFV PSTHPAYNLR WRAGVSKEEF EKSLNATTNW YSSFARSGAL
GELHNTCRIL YLQIVKYKYR LRVRSEGNSI PFAKMTKENE AKRQKTSQPK
AKKQLIINAF MSGSSGNQSP GLWSYPGDKS TEYTTLDYWV ELAQKLEKAK
FHSIFIADVL GGYDVYNGPG NYSAAAKSGA QFPMIEPSAA VTAMAAATKS
ITFGVTFSTI SEAPYHFARR LGTLDLLTNG RVGWNIVSSY LDSAARNLLN
GEPLPLHADR YKRAEEFLQV VYRLFLSSWR DDAYKLDKKT RTFADPKLIR
TIDHVGEFFN VPGPQFLPPT PQRLPLILQA GTSKVGMDYA AKHAEVVFLA
SFDPESLQEK IKTVRDIAET KYNRPRDSIK FLILITVVIA DTHEDAVKRY
EDLASYADLE GAQALFSGWT GIDIGKYGED EPLEHVESNA IKSHVKNWTK
FKDNKPRARK DIAKQIGVGG SGPLLVGSVQ EIADELERWA EVSDLDGFNF
AYADYPQTFD DIIEKLLPEL NKRGVFWDDY KIPGGTFRES VFGRKFVDKD
HPAYDLRWRS DQTREEFEKK LAELEKK

Factor auxiliar 9, *Komagataella pastoris* (SEQ ID NO: 70):

MRFSNVVLTA IAAAGVQADE ALYTVFYNDV TENAQEYLSY IQANTAAGFT
DLLSLYTELA TYTDDSYTSI FTEEDFPASE LSSFVVNLPW YSSRIEPQVA
AAETGESEEE SETGESEEES ETGEETETET GSESESESES ETSATGTGTG
TSASESAETE TSTDAAVSID HPKSTLLMGL TAAVVSITFG VFAL

Fig. 1 (continuación)

Factor auxiliar 10, *Komagataella pastoris* (SEQ ID NO: 71):

MSSFRVLDLV KPFTPFLPEV ISPERKVPFQ QKLMWTGVTL LIFLVMSEIP
LYGITSSDSS DPLFWLRMML ASNRGTLMEL GISPIVTSGM VFQLLQGIQI
LDVNMENKAD RELFQTAQKV FAILLSIGQA TVYVLTGMYG PPGELGVGVC
LLLVLQLVFA GIVVILLDEL LQKGYGLGSG ISLFMATNIC EQIFWKTFAP
TTVNRGRGKE FEGAFISFFH LILTKKDKKR ALLESFYRDN APNMFQVIAT
LVVFFTVVYL QGFRLEIPVK STRQRGPYGT YPIRLFYTSN MPIMLQSALT
SNIFIISQML YSHFPDNAFV KLIGTWEAQP GSAQLFAASG LAYYMQPPMS
LSQALLDPIK TVVYVVFVLT TCAIFSKTWI EISGSSPRDV AKQFKDQGLV
IAGHRDATVY KELKKIIPTA AAFGGATIGA LSVVSDLLGT LGSGTSILLA
VTTIYGYYEL AVKEGGFSKG GPSGFVDL