

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 770 616**

51 Int. Cl.:

C12P 11/00 (2006.01)

C12N 9/10 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **21.08.2015 PCT/EP2015/069298**

87 Fecha y número de publicación internacional: **25.02.2016 WO16026976**

96 Fecha de presentación y número de la solicitud europea: **21.08.2015 E 15754200 (2)**

97 Fecha y número de publicación de la concesión europea: **06.11.2019 EP 3183355**

54 Título: **Proceso biológico para la producción de un compuesto fenólico sulfatado**

30 Prioridad:

22.08.2014 EP 14182032

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

02.07.2020

73 Titular/es:

**CYSBIO APS (100.0%)
c/o Danmarks Tekniske Universitet, Kemitorvet
220
2800 Kgs. Lyngby, DK**

72 Inventor/es:

**JENDRESEN, CHRISTIAN BILLE y
NIELSEN, ALEX TOFTGAARD**

74 Agente/Representante:

VALLEJO LÓPEZ, Juan Pedro

ES 2 770 616 T3

Aviso: En el plazo de nueve meses a contar desde la fecha de publicación en el Boletín Europeo de Patentes, de la mención de concesión de la patente europea, cualquier persona podrá oponerse ante la Oficina Europea de Patentes a la patente concedida. La oposición deberá formularse por escrito y estar motivada; sólo se considerará como formulada una vez que se haya realizado el pago de la tasa de oposición (art. 99.1 del Convenio sobre Concesión de Patentes Europeas).

DESCRIPCIÓN

Proceso biológico para la producción de un compuesto fenólico sulfatado

5 **Campo técnico de la invención**

La presente invención generalmente se refiere al campo de la biotecnología, ya que se aplica a la producción de un compuesto fenólico sulfatado usando polipéptidos o células recombinantes que comprenden dichos polipéptidos. También, la presente invención se refiere a células hospedadoras recombinantes que expresan dichos polipéptidos y composiciones que comprenden tales células hospedadoras.

10 **Antecedentes de la invención**

Una gama de compuestos fenólicos son de gran interés para la industria biotecnológica, ya que son componentes básicos para los compuestos poliméricos. Los ejemplos de tales compuestos fenólicos incluyen ácido p-cumárico (pHCA) u otros ácidos hidroxicinámicos que forman la base de muchos metabolitos secundarios, incluyendo flavonoides y estilbenos. Sin embargo, muchos de estos compuestos fenólicos son tóxicos para los organismos productores y, por tanto, limitan la productividad durante la fermentación. Por tanto, existe la necesidad de procesos de producción a gran escala, y especialmente procesos de producción biológicos a gran escala que permitan una productividad mejorada.

Además, una gama de compuestos fenólicos, y especialmente aquellos utilizados como medicamentos o aditivos alimentarios como resveratrol o vainillina, muestran poca solubilidad en agua, lo que dificulta que el cuerpo absorba estos compuestos. Por tanto, también es necesario proporcionar tales compuestos fenólicos en una forma que mejore la solubilidad y, por tanto, la biodisponibilidad, preferentemente mediante el uso de procesos biológicos de producción a gran escala.

Que las sulfotransferasas están implicadas en la sulfatación de, por ejemplo, ácidos hidroxicinámicos se desvela, por ejemplo, en *Biofactors*, vol. 39,2013, pp. 644-651 de Wong *et al.*

30

Sumario de la invención

El objetivo de la presente invención es proporcionar un método para la producción a gran escala de compuestos fenólicos sulfatados (sulfatos de arilo). Adicionalmente, Un objetivo es proporcionar un proceso biológico para la producción a gran escala de fenoles. Los inventores han desarrollado un proceso biológico con el que se logran ambos objetivos.

35

La presente invención proporciona así en un primer aspecto un proceso para la producción de un compuesto fenólico sulfatado que comprende:

40

(i') poner en contacto un medio que comprende un compuesto fenólico con una primera célula hospedadora recombinante; en donde la primera célula hospedadora recombinante comprende un polipéptido heterólogo que tiene una actividad aril sulfotransferasa; o

45

(ii') poner en contacto un medio que comprende un sustrato de carbono fermentable con una primera célula hospedadora recombinante; en donde la primera célula hospedadora recombinante comprende un polipéptido heterólogo que tiene una actividad aril sulfotransferasa; o

50

(iii') poner en contacto un medio que comprende un precursor de un compuesto fenólico con una primera célula hospedadora recombinante; en donde la primera célula hospedadora recombinante comprende un polipéptido heterólogo que tiene una actividad aril sulfotransferasa, y en donde la primera célula hospedadora recombinante se ha modificado además para tener una expresión proteica aumentada de (i) una ATP sulfurilasa, (ii) una APS quinasa y/o (iii) una PAP fosfatasa en comparación con una célula hospedadora idéntica que no lleva dicha modificación.

En un aspecto adicional la presente invención proporciona un proceso para la producción de un compuesto fenólico sulfatado, como ácido zosterico, el método comprende sulfatar un compuesto fenólico, como ácido p-cumárico, usando un polipéptido como se detalla en este documento. Particularmente, el proceso implica el uso de un polipéptido que tiene una actividad aril sulfotransferasa, tal como un polipéptido seleccionado del grupo que consiste en:

55

a) un polipéptido que comprende una secuencia de aminoácidos establecida en SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 o 13 (por ejemplo, SEQ ID NO: 1);

60

b) un polipéptido que comprende una secuencia de aminoácidos que tiene al menos el 70 %, como al menos aproximadamente el 75 %, al menos aproximadamente el 80 %, al menos aproximadamente el 85 %, al menos aproximadamente el 90 %, al menos aproximadamente el 93 %, al menos el 95 %, al menos el 96 %, al menos el 97 %, al menos el 98 %, o al menos el 99 %, de identidad de secuencia con la secuencia de aminoácidos establecida en SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 o 13 (por ejemplo, SEQ ID NO: 1); o

65

c) un polipéptido que comprende una secuencia de aminoácidos establecida en SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 o 13 (por ejemplo, SEQ ID NO: 1), en donde 1 a 50, como 1 a 40, 1 a 35, 1 a 30, 1 a 25, 1 a 20, 1 a 15, 1 a 10, 1 a 5, 1 a 3, restos de aminoácido están sustituidos, eliminados y/o insertados.

5 La presente invención proporciona en un aspecto adicional una célula hospedadora recombinante como se detalla en este documento. Particularmente, la presente invención proporciona una célula hospedadora recombinante que comprende (por ejemplo, expresa) un primer polipéptido heterólogo que tiene una actividad aril sulfotransferasa, tal como un polipéptido heterólogo seleccionado del grupo que consiste en:

- 10 a) un polipéptido que comprende una secuencia de aminoácidos establecida en SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 o 13 (por ejemplo, SEQ ID NO: 1);
b) un polipéptido que comprende una secuencia de aminoácidos que tiene al menos el 70 %, como al menos aproximadamente el 75 %, al menos aproximadamente el 80 %, al menos aproximadamente el 85 %, al menos aproximadamente el 90 %, al menos aproximadamente el 93 %, al menos el 95 %, al menos el 96 %, al menos el 97 %, al menos el 98 %, o al menos el 99 %, de identidad de secuencia con la secuencia de aminoácidos establecida en SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 o 13 (por ejemplo, SEQ ID NO: 1); o
15 c) un polipéptido que comprende una secuencia de aminoácidos establecida en SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 o 13 (por ejemplo, SEQ ID NO: 1), en donde 1 a 50, como 1 a 40, 1 a 35, 1 a 30, 1 a 25, 1 a 20, 1 a 15, 1 a 10, 1 a 5, 1 a 3, restos de aminoácido están sustituidos, eliminados y/o insertados, en donde la célula hospedadora se ha modificado adicionalmente para tener una expresión proteica aumentada de (i) una ATP sulfurilasa, (ii) una APS quinasa y/o (iii) una PAP fosfatasa en comparación con una célula hospedadora idéntica que no lleva dicha modificación.
20

25 La presente divulgación proporciona en un aspecto adicional el uso de un polipéptido como se detalla en este documento en la sulfatación de un compuesto fenólico. Particularmente, la divulgación proporciona el uso de un polipéptido que tiene una actividad aril sulfotransferasa en la sulfatación de un compuesto fenólico, tal como un polipéptido seleccionado del grupo que consiste en:

- 30 a) un polipéptido que comprende una secuencia de aminoácidos establecida en SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 o 13 (por ejemplo, SEQ ID NO: 1);
b) un polipéptido que comprende una secuencia de aminoácidos que tiene al menos el 70 %, como al menos aproximadamente el 75 %, al menos aproximadamente el 80 %, al menos aproximadamente el 85 %, al menos aproximadamente el 90 %, al menos aproximadamente el 93 %, al menos el 95 %, al menos el 96 %, al menos el 97 %, al menos el 98 %, o al menos el 99 %, de identidad de secuencia con la secuencia de aminoácidos establecida en SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 o 13 (por ejemplo, SEQ ID NO: 1); o
35 c) un polipéptido que comprende una secuencia de aminoácidos establecida en SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 o 13 (por ejemplo, SEQ ID NO: 1), en donde 1 a 50, como 1 a 40, 1 a 35, 1 a 30, 1 a 25, 1 a 20, 1 a 15, 1 a 10, 1 a 5, 1 a 3, restos de aminoácido están sustituidos, eliminados y/o insertados.

40 La presente invención proporciona en un aspecto adicional una composición que comprende una primera célula hospedadora recombinante como se detalla en este documento y una segunda célula hospedadora recombinante como se detalla en este documento. Particularmente, la presente invención proporciona una composición que comprende una primera célula hospedadora recombinante que comprende (por ejemplo, que expresa) un polipéptido heterólogo que tiene una actividad aril sulfotransferasa, tal como un polipéptido heterólogo seleccionado del grupo que consiste en:
45

- a) un polipéptido que comprende una secuencia de aminoácidos establecida en SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 o 13 (por ejemplo, SEQ ID NO: 1);
50 b) un polipéptido que comprende una secuencia de aminoácidos que tiene al menos el 70 %, como al menos aproximadamente el 75 %, al menos aproximadamente el 80 %, al menos aproximadamente el 85 %, al menos aproximadamente el 90 %, al menos aproximadamente el 93 %, al menos el 95 %, al menos el 96 %, al menos el 97 %, al menos el 98 %, o al menos el 99 %, de identidad de secuencia con la secuencia de aminoácidos establecida en SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 o 13 (por ejemplo, SEQ ID NO: 1); o
55 c) un polipéptido que comprende una secuencia de aminoácidos establecida en SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 o 13 (por ejemplo, SEQ ID NO: 1), en donde 1 a 50, como 1 a 40, 1 a 35, 1 a 30, 1 a 25, 1 a 20, 1 a 15, 1 a 10, 1 a 5, 1 a 3, restos de aminoácido están sustituidos, eliminados y/o insertados, en donde la célula hospedadora se ha modificado adicionalmente para tener una expresión proteica aumentada de (i) una ATP sulfurilasa, (ii) una APS quinasa y/o (iii) una PAP fosfatasa; y

60 una segunda célula hospedadora recombinante que comprende (por ejemplo, que expresa) un polipéptido heterólogo que tiene actividad tirosina amoniaco liasa, tal como un polipéptido seleccionado del grupo que consiste en:

- d) un polipéptido que comprende una secuencia de aminoácidos establecida en SEQ ID NO: 14, 15, 16, 17, 18, 19, 20, 21, 22 o 23 (por ejemplo, SEQ ID NO: 14);
65 e) un polipéptido que comprende una secuencia de aminoácidos que tiene al menos aproximadamente el 70 %, como al menos aproximadamente el 75 %, al menos aproximadamente el 80 %, al menos aproximadamente el

85 %, al menos aproximadamente el 90 %, al menos aproximadamente el 93 %, al menos el 95 %, al menos el 96 %, al menos el 97 %, al menos el 98 %, o al menos el 99 %, de identidad de secuencia con la secuencia de aminoácidos establecida en SEQ ID NO: 14, 15, 16, 17, 18, 19, 20, 21, 22 o 23 (por ejemplo, SEQ ID NO: 14); o f) un polipéptido que comprende una secuencia de aminoácidos establecida en SEQ ID NO: 14, 15, 16, 17, 18, 19, 20, 21, 22 o 23 (por ejemplo, SEQ ID NO: 14), en donde 1 a 50, como 1 a 40, 1 a 35, 1 a 30, 1 a 25, 1 a 20, 1 a 15, 1 a 10, 1 a 5, 1 a 3, restos de aminoácido están sustituidos, eliminados y/o insertados.

Además se desvela una composición que comprende un primer polipéptido y un segundo polipéptido. Particularmente, la presente invención proporciona una composición que comprende un primer polipéptido que tiene una actividad aril sulfotransferasa, tal como un polipéptido seleccionado del grupo que consiste en:

- a) un polipéptido que comprende una secuencia de aminoácidos establecida en SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 o 13 (por ejemplo, SEQ ID NO: 1);
- b) un polipéptido que comprende una secuencia de aminoácidos que tiene al menos el 70 %, como al menos aproximadamente el 75 %, al menos aproximadamente el 80 %, al menos aproximadamente el 85 %, al menos aproximadamente el 90 %, al menos aproximadamente el 93 %, al menos el 95 %, al menos el 96 %, al menos el 97 %, al menos el 98 %, o al menos el 99 %, de identidad de secuencia con la secuencia de aminoácidos establecida en SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 o 13 (por ejemplo, SEQ ID NO: 1); o
- c) un polipéptido que comprende una secuencia de aminoácidos establecida en SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 o 13 (por ejemplo, SEQ ID NO: 1), en donde 1 a 50, como 1 a 40, 1 a 35, 1 a 30, 1 a 25, 1 a 20, 1 a 15, 1 a 10, 1 a 5, 1 a 3, restos de aminoácido están sustituidos, eliminados y/o insertados; y

una segunda actividad del polipéptido tirosina amoniaco liasa, tal como un polipéptido seleccionado del grupo que consiste en:

- d) un polipéptido que comprende una secuencia de aminoácidos establecida en SEQ ID NO: 14, 15, 16, 17, 18, 19, 20, 21, 22 o 23 (por ejemplo, SEQ ID NO: 14);
- e) un polipéptido que comprende una secuencia de aminoácidos que tiene al menos el 70 %, como al menos aproximadamente el 75 %, al menos aproximadamente el 80 %, al menos aproximadamente el 85 %, al menos aproximadamente el 90 %, al menos aproximadamente el 93 %, al menos el 95 %, al menos el 96 %, al menos el 97 %, al menos el 98 %, o al menos el 99 %, de identidad de secuencia con la secuencia de aminoácidos establecida en SEQ ID NO: 14, 15, 16, 17, 18, 19, 20, 21, 22 o 23 (por ejemplo, SEQ ID NO: 14); o
- f) un polipéptido que comprende una secuencia de aminoácidos establecida en SEQ ID NO: 14, 15, 16, 17, 18, 19, 20, 21, 22 o 23 (por ejemplo, SEQ ID NO: 14), en donde 1 a 50, como 1 a 40, 1 a 35, 1 a 30, 1 a 25, 1 a 20, 1 a 15, 1 a 10, 1 a 5, 1 a 3, restos de aminoácido están sustituidos, eliminados y/o insertados.

Breve descripción de los dibujos

- Figura 1: Mapa del plásmido para la expresión de SULT1A1 de *Rattus norvegicus* en *Escherichia coli*
- Figura 2: Mapa del plásmido para la sobreexpresión de cysDNC en *E. coli*.
- Figura 3: Mapa del plásmido para la sobreexpresión de cysDNCQ en *E. coli*.
- Figura 4: Mapa del plásmido para la expresión de RmXAL de *Rhodotorula mucilaginosa* / *Rhodotorula rubra* en *E. coli*.
- Figura 5: Toxicidad de productos no sulfatados o sulfatados
- Figura 6: Mapa del plásmido para la expresión de tirosina amoniaco-liasa RsTAL de *Rhodobacter sphaeroides* en *E. coli*.
- Figura 7: Mapa del plásmido para la expresión de tirosina amoniaco-liasa FJTAL de *Flavobacterium johnsoniae* en *E. coli*.
- Figura 8: Mapa del plásmido para la expresión de tirosina amoniaco-liasa RcTAL de *Rhodobacter capsulatus* en *E. coli*.
- Figura 9: Mapa del plásmido para la expresión de SULT1A1 de *Rattus norvegicus* en *Saccharomyces cerevisiae* (gen nativo).
- Figura 10: Mapa del plásmido para la expresión de SULT1A1 de *Rattus norvegicus* en *Saccharomyces cerevisiae* (gen optimizado con codón).

Descripción detallada de la invención

Salvo que se defina específicamente en este documento, todos los términos técnicos y científicos utilizados tienen el mismo significado que comúnmente entiende un experto en los campos de bioquímica, genética y biología molecular.

Todos los métodos y materiales similares o equivalentes a los descritos en este documento se pueden usar en la práctica o el ensayo de la presente invención, con métodos y materiales adecuados que se describen en este documento.

La práctica de la presente invención empleará, a menos que se indique lo contrario, técnicas convencionales de biología celular, cultivo celular, biología molecular, biología transgénica, microbiología, ADN recombinante e

inmunología, que están dentro de la experiencia en la materia. Tales técnicas se explican por completo en la bibliografía. Véanse, por ejemplo, *Current Protocols in Molecular Biology* (Frederick M. AUSUBEL, 2000, Wiley and son Inc, Library of Congress, USA); *Molecular Cloning: A Laboratory Manual*, tercera edición, (Sambrook *et al*, 2001, Cold Spring Harbor, Nueva York: Cold Spring Harbor Laboratory Press); *Oligonucleotide Synthesis* (M. J. Gait ed., 1984); Mullis *et al*. Patente de Estados Unidos N.º 4.683.195; *Nucleic Acid Hybridization* (B. D. Harries y S. J. Higgins eds. 1984); *Transcription And Translation* (B. D. Hames y S. J. Higgins eds. 1984); *Culture Of Animal Cells* (R. I. Freshney, Alan R. Liss, Inc., 1987); *Immobilized Cells And Enzymes* (IRL Press, 1986); B. Perbal, *A Practical Guide To Molecular Cloning* (1984); la serie, *Methods In ENZYMOLOGY* (J. Abelson y M. Simon, eds.-in-chief, Academic Press, Inc., Nueva York), específicamente, vol. 154 y 155 (Wu *et al*. eds.) y vol. 185, "Gene Expression Technology" (D. Goeddel, ed.); *Gene Transfer Vectors For Mammalian Cells* (J. H. Miller y M. P. Calos eds., 1987, Cold Spring Harbor Laboratory); *Immunochemical Methods In Cell And Molecular Biology* (Mayer y Walker, eds., Academic Press, Londres, 1987); *Handbook Of Experimental Immunology*, Volúmenes I-IV (D. M. Weir y C. C. Blackwell, eds., 1986); y *Manipulating the Mouse Embryo*, (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1986).

15 *Polipéptidos y células hospedadoras*

Como se ha indicado anteriormente, la presente invención utiliza polipéptidos que tienen actividad aril sulfotransferasa (EC:2.8.2.1). Esto los hace particularmente adecuados para la sulfatación de compuestos fenólicos como ácido p-cumárico y sus derivados (por ejemplo, ácido cafeico, ácido ferúlico o ácido sinápico), o resveratrol.

El polipéptido que tiene actividad aril sulfotransferasa puede ser una enzima sulfotransferasa 1A1, una enzima sulfotransferasa 1A2, una enzima sulfotransferasa 1A3, una enzima sulfotransferasa 1B1, una enzima sulfotransferasa 1C1, una enzima sulfotransferasa 1C2, una enzima sulfotransferasa 1C4, o una enzima sulfotransferasa 1E1. Según determinadas realizaciones, el polipéptido que tiene actividad aril sulfotransferasa es una enzima sulfotransferasa 1A1. Según ciertas otras realizaciones, el polipéptido que tiene actividad aril sulfotransferasa es una enzima sulfotransferasa 1A2. Según determinadas realizaciones, el polipéptido que tiene actividad aril sulfotransferasa es una enzima sulfotransferasa 1B1. Según determinadas realizaciones, el polipéptido que tiene actividad aril sulfotransferasa es una enzima sulfotransferasa 1C1. Según determinadas realizaciones, el polipéptido que tiene actividad aril sulfotransferasa es una enzima sulfotransferasa 1C2. Según determinadas realizaciones, el polipéptido que tiene actividad aril sulfotransferasa es una enzima sulfotransferasa 1C4. Según otras ciertas realizaciones, el polipéptido que tiene actividad aril sulfotransferasa es una enzima sulfotransferasa 1E1 (estrógeno sulfotransferasa), como la sulfotransferasa 1E1 de *Gallus gallus domesticus*.

Según determinadas realizaciones, el polipéptido que tiene actividad aril sulfotransferasa es una aril sulfotransferasa de mamífero, como una enzima sulfotransferasa 1A1 de mamífero.

Según determinadas realizaciones, el polipéptido que tiene actividad aril sulfotransferasa es una aril sulfotransferasa de *Rattus norvegicus* o una variante de la misma. Dicha variante tiene al menos el 70 %, como al menos aproximadamente el 75 %, al menos aproximadamente el 80 %, al menos aproximadamente el 85 %, al menos aproximadamente el 90 %, al menos aproximadamente el 93 %, al menos el 95 %, al menos el 96 %, al menos el 97 %, al menos el 98 %, o al menos el 99 %, de identidad de secuencia con la secuencia de aminoácidos de la aril sulfotransferasa de *Rattus norvegicus*. Dicha variante también puede tener una secuencia de aminoácidos de la sulfotransferasa de *Rattus norvegicus*, en donde 1 a 50, como 1 a 40, 1 a 35, 1 a 30, 1 a 25, 1 a 20, 1 a 15, 1 a 10, 1 a 5, 1 a 3, restos de aminoácido están sustituidos, eliminados y/o insertados.

Se entiende que los valores anteriores definen generalmente el número total de alteraciones a la aril sulfotransferasa de referencia. Las alteraciones pueden ser únicamente sustituciones de aminoácidos, ya sean sustituciones conservadas o no conservadas, o ambas. Pueden ser únicamente deleciones de aminoácidos. Pueden ser únicamente inserciones de aminoácidos. Las alteraciones pueden ser una mezcla de estas alteraciones específicas, como sustituciones de aminoácidos e inserciones de aminoácidos.

Según determinadas realizaciones, el polipéptido que tiene actividad aril sulfotransferasa puede ser un polipéptido seleccionado del grupo que consiste en:

- a) un polipéptido que comprende una secuencia de aminoácidos establecida en SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 o 13 (por ejemplo, SEQ ID NO: 1);
- b) un polipéptido que comprende una secuencia de aminoácidos que tiene al menos el 70 %, como al menos aproximadamente el 75 %, al menos aproximadamente el 80 %, al menos aproximadamente el 85 %, al menos aproximadamente el 90 %, al menos aproximadamente el 93 %, al menos el 95 %, al menos el 96 %, al menos el 97 %, al menos el 98 %, o al menos el 99 %, de identidad de secuencia con la secuencia de aminoácidos establecida en SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 o 13 (por ejemplo, SEQ ID NO: 1); o
- c) un polipéptido que comprende una secuencia de aminoácidos establecida en SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 o 13 (por ejemplo, SEQ ID NO: 1), en donde 1 a 50, como 1 a 40, 1 a 35, 1 a 30, 1 a 25, 1 a 20, 1 a 15, 1 a 10, 1 a 5, 1 a 3, restos de aminoácido están sustituidos, eliminados y/o insertados.

Según determinadas realizaciones, un polipéptido para uso en la invención es un polipéptido según a). Por tanto, un

de secuencia con la secuencia de aminoácidos establecida en SEQ ID NO: 1.

Preferentemente, un polipéptido según b) tiene actividad aril sulfotransferasa. Más preferentemente, un polipéptido según b) tiene una actividad aril sulfotransferasa similar a la del polipéptido que comprende una secuencia de aminoácidos establecida en SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 o 13 (por ejemplo, SEQ ID NO: 1).

Según cierta realización, un polipéptido según b) tiene una actividad aril sulfotransferasa similar a la del polipéptido que comprende la secuencia de aminoácidos establecida en SEQ ID NO: 1. Según ciertas otras realizaciones, un polipéptido según b) tiene una actividad aril sulfotransferasa similar a la del polipéptido que comprende la secuencia de aminoácidos establecida en SEQ ID NO: 2. Según ciertas otras realizaciones, un polipéptido según b) tiene una actividad aril sulfotransferasa similar a la del polipéptido que comprende la secuencia de aminoácidos establecida en SEQ ID NO: 3. Según ciertas otras realizaciones, un polipéptido según b) tiene una actividad aril sulfotransferasa similar a la del polipéptido que comprende la secuencia de aminoácidos establecida en SEQ ID NO: 4. Según ciertas otras realizaciones, un polipéptido según b) tiene una actividad aril sulfotransferasa similar a la del polipéptido que comprende la secuencia de aminoácidos establecida en SEQ ID NO: 5. Según ciertas otras realizaciones, un polipéptido según b) tiene una actividad aril sulfotransferasa similar a la del polipéptido que comprende la secuencia de aminoácidos establecida en SEQ ID NO: 6. Según ciertas otras realizaciones, un polipéptido según b) tiene una actividad aril sulfotransferasa similar a la del polipéptido que comprende la secuencia de aminoácidos establecida en SEQ ID NO: 7. Según ciertas otras realizaciones, un polipéptido según b) tiene una actividad aril sulfotransferasa similar a la del polipéptido que comprende la secuencia de aminoácidos establecida en SEQ ID NO: 8. Según ciertas otras realizaciones, un polipéptido según b) tiene una actividad aril sulfotransferasa similar a la del polipéptido que comprende la secuencia de aminoácidos establecida en SEQ ID NO: 9. Según ciertas otras realizaciones, un polipéptido según b) tiene una actividad aril sulfotransferasa similar a la del polipéptido que comprende la secuencia de aminoácidos establecida en SEQ ID NO: 10. Según ciertas otras realizaciones, un polipéptido según b) tiene una actividad aril sulfotransferasa similar a la del polipéptido que comprende la secuencia de aminoácidos establecida en SEQ ID NO: 11. Según ciertas otras realizaciones, un polipéptido según b) tiene una actividad aril sulfotransferasa similar a la del polipéptido que comprende la secuencia de aminoácidos establecida en SEQ ID NO: 12. Según ciertas otras realizaciones, un polipéptido según b) tiene una actividad aril sulfotransferasa similar a la del polipéptido que comprende la secuencia de aminoácidos establecida en SEQ ID NO: 13.

Con actividad aril sulfotransferasa "similar", se entiende que el polipéptido según b) tiene al menos aproximadamente el 10 %, como al menos aproximadamente el 20 %, al menos aproximadamente el 30 %, al menos aproximadamente el 40 %, al menos aproximadamente el 50 %, al menos aproximadamente el 60 %, al menos aproximadamente el 75 %, al menos aproximadamente el 80 %, al menos aproximadamente el 90 %, al menos aproximadamente el 95 %, al menos aproximadamente el 100 %, al menos aproximadamente el 200 %, al menos aproximadamente el 300 %, al menos aproximadamente el 400 %, al menos aproximadamente el 500 %, al menos aproximadamente el 800 %, al menos aproximadamente el 1000 % o al menos aproximadamente el 2000 %, de la actividad aril sulfotransferasa del polipéptido de referencia (por ejemplo, SEQ ID NO: 1).

La actividad aril sulfotransferasa se puede determinar, por ejemplo, según el siguiente método: La actividad aril sulfotransferasa puede determinarse por la reacción de PAPS marcado radiactivamente con azufre, [³⁵S]PAPS, con el sustrato en presencia del polipéptido de interés. Esto fue descrito anteriormente, por ejemplo por Hattori *et al.* (Biosci Biotechnol Biochem. 2008; 72(2):540-7). La reacción tiene lugar en un tampón como 250 µl de fosfato sódico 50 mM, pH 6,8 con [³⁵S]PAPS 1 µM (3,7kBq) con compuesto aceptor 100 µM durante un periodo de 30 min a 30 °C. La reacción se detiene mediante la adición de 100 µl de una mezcla 1:1 de acetato de bario 0,1 M e hidróxido de bario. Se añaden 50 µl de sulfato de cinc 0,1 M, seguido de centrifugación a 1.200 x g durante 5 min. Luego se transfieren 300 µl del sobrenadante a un nuevo recipiente y se añaden 50 µl de un volumen igual de hidróxido de bario 0,1 M y sulfato de cinc 0,1 M. La mezcla se centrifuga luego a 13.000 x g durante 5 min, y se mezclan alícuotas de 300 µl del sobrenadante con 2,5 ml de Cleasol I (Nacalai Tesque, Kyoto, Japón). La radiactividad se mide luego por centelleo.

Alternativamente, la actividad de una sulfotransferasa se puede detectar mediante la medición directa del producto mediante métodos analíticos como cromatografía líquida de alto rendimiento (HPLC) o cromatografía líquida en combinación con espectrometría de masas (LC-MS).

Según otras ciertas realizaciones, un polipéptido o uso en la invención es un polipéptido según c). Por tanto, un polipéptido para uso en la invención puede ser un polipéptido que comprende una secuencia de aminoácidos establecida en SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 o 13 (por ejemplo, SEQ ID NO: 1), en donde 1 o más, tal como 2 o más, 3 o más, 4 o más, 5 o más, 6 o más, 7 o más, 8 o más, 9 o más, 10 o más, 11 o más, 12 o más, 13 o más, 14 o más, 15 o más, 16 o más, 17 o más, 18 o más, 19 o más, 20 o más, 25 o más, 30 o más, 35 o más, 40 o más, 45 o más, 50 o más, 60 o más, 70 o más, 80 o más, 90 o más, 100 o más, 110 o más, 120 o más, 130 o más, 140 o más, o 150 o más, restos de aminoácido están sustituidos, eliminados y/o insertados. Según realizaciones particulares, un polipéptido según c) comprende una secuencia de aminoácidos establecida en SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 o 13 (por ejemplo, SEQ ID NO: 1), en donde aproximadamente 1 a aproximadamente 150, como aproximadamente 1 a aproximadamente 140, aproximadamente 1 a aproximadamente 130, aproximadamente 1 a aproximadamente 120, aproximadamente 1 a aproximadamente 110, aproximadamente 1 a aproximadamente 100, aproximadamente 1 a aproximadamente 90, aproximadamente 1 a aproximadamente 80, aproximadamente 1 a

aproximadamente 70, aproximadamente 1 a aproximadamente 60, aproximadamente 1 a aproximadamente 50, aproximadamente 1 a aproximadamente 40, aproximadamente 1 a aproximadamente 35, aproximadamente 1 a aproximadamente 30, aproximadamente 1 a aproximadamente 25, aproximadamente 1 a aproximadamente 20, aproximadamente 1 a aproximadamente 15, aproximadamente 1 a aproximadamente 10, aproximadamente 1 a aproximadamente 5, o aproximadamente 1 a aproximadamente 3, restos de aminoácido están sustituidos, eliminados y/o insertados. Según realizaciones más particulares, un polipéptido según c) comprende una secuencia de aminoácidos establecida en SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 o 13 (por ejemplo, SEQ ID NO: 1), en donde aproximadamente 1 a aproximadamente 50, aproximadamente 1 a aproximadamente 40, aproximadamente 1 a aproximadamente 35, aproximadamente 1 a aproximadamente 30, aproximadamente 1 a aproximadamente 25, aproximadamente 1 a aproximadamente 20, aproximadamente 1 a aproximadamente 15, aproximadamente 1 a aproximadamente 10, aproximadamente 1 a aproximadamente 5, o aproximadamente 1 a aproximadamente 3, restos de aminoácido están sustituidos, eliminados y/o insertados. Según otras realizaciones más particulares, un polipéptido según c) comprende una secuencia de aminoácidos establecida en SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 o 13 (por ejemplo, SEQ ID NO: 1), en donde aproximadamente 1 a aproximadamente 30, como aproximadamente 1 a aproximadamente 25, aproximadamente 1 a aproximadamente 20, aproximadamente 1 a aproximadamente 15, aproximadamente 1 a aproximadamente 10, aproximadamente 1 a aproximadamente 5, o aproximadamente 1 a aproximadamente 3, restos de aminoácido están sustituidos, eliminados y/o insertados.

Según otras realizaciones más particulares, un polipéptido según c) comprende una secuencia de aminoácidos establecida en SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 o 13 (por ejemplo, SEQ ID NO: 1), en donde aproximadamente 1 a aproximadamente 25, como aproximadamente 1 a aproximadamente 20, aproximadamente 1 a aproximadamente 15, aproximadamente 1 a aproximadamente 10, aproximadamente 1 a aproximadamente 5, o aproximadamente 1 a aproximadamente 3, restos de aminoácido están sustituidos, eliminados y/o insertados.

Según realizaciones particulares, un polipéptido según c) comprende una secuencia de aminoácidos establecida en SEQ ID NO: 1, en donde aproximadamente 1 a aproximadamente 150, como aproximadamente 1 a aproximadamente 140, aproximadamente 1 a aproximadamente 130, aproximadamente 1 a aproximadamente 120, aproximadamente 1 a aproximadamente 110, aproximadamente 1 a aproximadamente 100, aproximadamente 1 a aproximadamente 90, aproximadamente 1 a aproximadamente 80, aproximadamente 1 a aproximadamente 70, aproximadamente 1 a aproximadamente 60, aproximadamente 1 a aproximadamente 50, aproximadamente 1 a aproximadamente 40, aproximadamente 1 a aproximadamente 35, aproximadamente 1 a aproximadamente 30, aproximadamente 1 a aproximadamente 25, aproximadamente 1 a aproximadamente 20, aproximadamente 1 a aproximadamente 15, aproximadamente 1 a aproximadamente 10, aproximadamente 1 a aproximadamente 5, o aproximadamente 1 a aproximadamente 3, restos de aminoácido están sustituidos, eliminados y/o insertados. Según realizaciones más particulares, un polipéptido según c) comprende una secuencia de aminoácidos establecida en SEQ ID NO: 1, en donde aproximadamente 1 a aproximadamente 50, aproximadamente 1 a aproximadamente 40, aproximadamente 1 a aproximadamente 35, aproximadamente 1 a aproximadamente 30, aproximadamente 1 a aproximadamente 25, aproximadamente 1 a aproximadamente 20, aproximadamente 1 a aproximadamente 15, aproximadamente 1 a aproximadamente 10, aproximadamente 1 a aproximadamente 5, o aproximadamente 1 a aproximadamente 3, restos de aminoácido están sustituidos, eliminados y/o insertados. Según otras realizaciones más particulares, un polipéptido según c) comprende una secuencia de aminoácidos establecida en SEQ ID NO: 1, en donde aproximadamente 1 a aproximadamente 30, como aproximadamente 1 a aproximadamente 25, aproximadamente 1 a aproximadamente 20, aproximadamente 1 a aproximadamente 15, aproximadamente 1 a aproximadamente 10, aproximadamente 1 a aproximadamente 5, o aproximadamente 1 a aproximadamente 3, restos de aminoácido están sustituidos, eliminados y/o insertados. Según otras realizaciones más particulares, un polipéptido según c) comprende una secuencia de aminoácidos establecida en SEQ ID NO: 1, en donde aproximadamente 1 a aproximadamente 25, como aproximadamente 1 a aproximadamente 20, aproximadamente 1 a aproximadamente 15, aproximadamente 1 a aproximadamente 10, aproximadamente 1 a aproximadamente 5, o aproximadamente 1 a aproximadamente 3, restos de aminoácido están sustituidos, eliminados y/o insertados.

Se entiende que los valores anteriores generalmente definen el número total de alteraciones al polipéptido de referencia (por ejemplo, SEQ ID NO: 1). Las alteraciones pueden ser únicamente sustituciones de aminoácidos, ya sean sustituciones conservadas o no conservadas, o ambas. Pueden ser únicamente deleciones de aminoácidos. Pueden ser únicamente inserciones de aminoácidos. Las alteraciones pueden ser una mezcla de estas alteraciones específicas, como sustituciones de aminoácidos e inserciones de aminoácidos.

Preferentemente, un polipéptido según c) tiene actividad aril sulfotransferasa. Más preferentemente, un polipéptido según c) tiene una actividad aril sulfotransferasa similar a la del polipéptido que comprende una secuencia de aminoácidos establecida en SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 o 13 (por ejemplo, SEQ ID NO: 1).

Según cierta realización, un polipéptido según c) tiene una actividad aril sulfotransferasa similar a la del polipéptido que comprende la secuencia de aminoácidos establecida en la SEQ ID NO: 1. Según ciertas otras realizaciones, un polipéptido según c) tiene una actividad aril sulfotransferasa similar a la del polipéptido que comprende la secuencia de aminoácidos establecida en la SEQ ID NO: 2. Según ciertas otras realizaciones, un polipéptido según c) tiene una actividad aril sulfotransferasa similar a la del polipéptido que comprende la secuencia de aminoácidos establecida en la SEQ ID NO: 3. Según ciertas otras realizaciones, un polipéptido según c) tiene una actividad aril sulfotransferasa

similar a la del polipéptido que comprende la secuencia de aminoácidos establecida en la SEQ ID NO: 4. Según ciertas otras realizaciones, un polipéptido según c) tiene una actividad aril sulfotransferasa similar a la del polipéptido que comprende la secuencia de aminoácidos establecida en la SEQ ID NO: 5. Según ciertas otras realizaciones, un polipéptido según c) tiene una actividad aril sulfotransferasa similar a la del polipéptido que comprende la secuencia de aminoácidos establecida en la SEQ ID NO: 6. Según ciertas otras realizaciones, un polipéptido según c) tiene una actividad aril sulfotransferasa similar a la del polipéptido que comprende la secuencia de aminoácidos establecida en la SEQ ID NO: 7. Según ciertas otras realizaciones, un polipéptido según c) tiene una actividad aril sulfotransferasa similar a la del polipéptido que comprende la secuencia de aminoácidos establecida en la SEQ ID NO: 8. Según ciertas otras realizaciones, un polipéptido según c) tiene una actividad aril sulfotransferasa similar a la del polipéptido que comprende la secuencia de aminoácidos establecida en la SEQ ID NO: 9. Según ciertas otras realizaciones, un polipéptido según c) tiene una actividad aril sulfotransferasa similar a la del polipéptido que comprende la secuencia de aminoácidos establecida en la SEQ ID NO: 10. Según ciertas otras realizaciones, un polipéptido según c) tiene una actividad aril sulfotransferasa similar a la del polipéptido que comprende la secuencia de aminoácidos establecida en la SEQ ID NO: 11. Según ciertas otras realizaciones, un polipéptido según c) tiene una actividad aril sulfotransferasa similar a la del polipéptido que comprende la secuencia de aminoácidos establecida en la SEQ ID NO: 12. Según ciertas otras realizaciones, un polipéptido según c) tiene una actividad aril sulfotransferasa similar a la del polipéptido que comprende la secuencia de aminoácidos establecida en la SEQ ID NO: 13.

Con actividad aril sulfotransferasa "similar" se entiende que el polipéptido según c) tiene al menos aproximadamente el 10 %, como al menos aproximadamente el 20 %, al menos aproximadamente el 30 %, al menos aproximadamente el 40 %, al menos aproximadamente el 50 %, al menos aproximadamente el 60, al menos aproximadamente el 75 %, al menos aproximadamente el 80 %, al menos aproximadamente el 90 %, al menos aproximadamente el 95 %, al menos aproximadamente el 100 %, al menos aproximadamente el 200 %, al menos aproximadamente el 300 %, al menos aproximadamente el 400 %, al menos aproximadamente el 500 %, al menos aproximadamente el 800 %, al menos aproximadamente el 1000 % o al menos aproximadamente el 2000 %, de la actividad aril sulfotransferasa del polipéptido de referencia (por ejemplo, SEQ ID NO: 1).

La actividad aril sulfotransferasa se puede determinar, por ejemplo, según el siguiente método: La actividad aril sulfotransferasa puede determinarse por la reacción de PAPS marcado radiactivamente con azufre, [³⁵S]PAPS, con el sustrato en presencia del polipéptido de interés. Esto fue descrito anteriormente, por ejemplo por Hattori *et al.* (Biosci Biotechnol Biochem. 2008; 72(2):540-7). La reacción tiene lugar en un tampón como 250 µl de fosfato sódico 50 mM, pH 6,8 con [³⁵S]PAPS 1 µM (3,7kBq) con compuesto aceptor 100 µM durante un periodo de 30 min a 30 °C. La reacción se detiene mediante la adición de 100 µl de una mezcla 1:1 de acetato de bario 0,1 M e hidróxido de bario. Se añaden 50 µl de sulfato de cinc 0,1 M, seguido de centrifugación a 1.200 x g durante 5 min. Luego se transfieren 300 µl del sobrenadante a un nuevo recipiente y se añaden 50 µl de un volumen igual de hidróxido de bario 0,1 M y sulfato de cinc 0,1 M. La mezcla se centrifuga luego a 13.000 x g durante 5 min, y se mezclan alícuotas de 300 µl del sobrenadante con 2,5 ml de Cleasol I (Nacalai Tesque, Kyoto, Japón). La radiactividad se mide luego por centelleo.

Alternativamente, la actividad de una sulfotransferasa se puede detectar mediante la medición directa del producto mediante métodos analíticos como cromatografía líquida de alto rendimiento (HPLC) o cromatografía líquida en combinación con espectrometría de masas (LC-MS).

La presente invención contempla la producción de un compuesto fenólico sulfatado a partir de un precursor del mismo, y en particular a partir de un precursor de la fórmula general (p-I) como se describe con más detalle a continuación. En este caso, puede ser adecuado emplear un polipéptido adicional (por ejemplo, segundo) que tenga actividad tirosina amoniaco liasa. Tal polipéptido puede ser un polipéptido seleccionado del grupo que consiste en:

d) un polipéptido que comprende una secuencia de aminoácidos establecida en SEQ ID NO: 14, 15, 16, 17, 18, 19, 20, 21, 22 o 23 (por ejemplo, SEQ ID NO: 14);

e) un polipéptido que comprende una secuencia de aminoácidos que tiene al menos el 70 %, como al menos aproximadamente el 75 %, al menos aproximadamente el 80 %, al menos aproximadamente el 85 %, al menos aproximadamente el 90 %, al menos aproximadamente el 93 %, al menos aproximadamente el 95 %, al menos aproximadamente el 96 %, al menos aproximadamente el 97 %, al menos aproximadamente el 98 %, o al menos aproximadamente el 99 %, de identidad de secuencia con la secuencia de aminoácidos establecida en SEQ ID NO: 14, 15, 16, 17, 18, 19, 20, 21, 22 o 23 (por ejemplo, SEQ ID NO: 14); o

f) un polipéptido que comprende una secuencia de aminoácidos establecida en SEQ ID NO: 14, 15, 16, 17, 18, 19, 20, 21, 22 o 23 (por ejemplo, SEQ ID NO: 14), en donde 1 o más, como aproximadamente 1 a aproximadamente 50, aproximadamente 1 a aproximadamente 40, aproximadamente 1 a aproximadamente 35, aproximadamente 1 a aproximadamente 30, aproximadamente 1 a aproximadamente 25, aproximadamente 1 a aproximadamente 20, aproximadamente 1 a aproximadamente 15, aproximadamente 1 a aproximadamente 10, aproximadamente 1 a aproximadamente 5, o aproximadamente 1 a aproximadamente 3, restos de aminoácido están sustituidos, eliminados y/o insertados.

Según determinadas realizaciones, un polipéptido adicional para uso en la invención es un polipéptido según d). Por tanto, un polipéptido según la invención puede ser un polipéptido que comprende una secuencia de aminoácidos establecida en SEQ ID NO: 14, 15, 16, 17, 18, 19, 20, 21, 22 o 23 (por ejemplo, SEQ ID NO: 14). Según realizaciones

Según cierta realización, un polipéptido según e) tiene una actividad tirosina amoniaco liasa similar a la del polipéptido que comprende la secuencia de aminoácidos establecida en SEQ ID NO: 14. Según ciertas otras realizaciones, un polipéptido según e) tiene una actividad tirosina amoniaco liasa similar a la del polipéptido que comprende la secuencia de aminoácidos establecida en SEQ ID NO: 15. Según ciertas otras realizaciones, un polipéptido según e) tiene una actividad tirosina amoniaco liasa similar a la del polipéptido que comprende la secuencia de aminoácidos establecida en SEQ ID NO: 16. Según ciertas otras realizaciones, un polipéptido según e) tiene una actividad tirosina amoniaco liasa similar a la del polipéptido que comprende la secuencia de aminoácidos establecida en SEQ ID NO: 17. Según ciertas otras realizaciones, un polipéptido según e) tiene una actividad tirosina amoniaco liasa similar a la del polipéptido que comprende la secuencia de aminoácidos establecida en SEQ ID NO: 18. Según ciertas otras realizaciones, un polipéptido según e) tiene una actividad tirosina amoniaco liasa similar a la del polipéptido que comprende la secuencia de aminoácidos establecida en SEQ ID NO: 19. Según ciertas otras realizaciones, un polipéptido según e) tiene una actividad tirosina amoniaco liasa similar a la del polipéptido que comprende la secuencia de aminoácidos establecida en SEQ ID NO: 20. Según ciertas otras realizaciones, un polipéptido según e) tiene una actividad tirosina amoniaco liasa similar a la del polipéptido que comprende la secuencia de aminoácidos establecida en SEQ ID NO: 21. Según ciertas otras realizaciones, un polipéptido según e) tiene una actividad tirosina amoniaco liasa similar a la del polipéptido que comprende la secuencia de aminoácidos establecida en SEQ ID NO: 22. Según ciertas otras realizaciones, un polipéptido según e) tiene una actividad tirosina amoniaco liasa similar a la del polipéptido que comprende la secuencia de aminoácidos establecida en SEQ ID NO: 23. Por actividad de tirosina amoniaco liasa "similar" se entiende que el polipéptido según e) tiene al menos aproximadamente el 10 %, como al menos aproximadamente el 20 %, al menos aproximadamente el 30 %, al menos aproximadamente el 40 %, al menos aproximadamente el 50 %, al menos aproximadamente el 60 %, al menos aproximadamente el 75 %, al menos aproximadamente el 80 %, al menos aproximadamente el 90 %, al menos aproximadamente el 95 %, al menos aproximadamente el 100 %, al menos aproximadamente el 200 %, al menos aproximadamente el 300 %, al menos aproximadamente el 400 %, al menos aproximadamente el 500 %, al menos aproximadamente el 800 %, al menos aproximadamente el 1000 % o al menos aproximadamente el 2000 %, de la actividad amoniaco liasa del polipéptido de referencia (por ejemplo, SEQ ID NO: 14).

La actividad tirosina amoniaco liasa se puede determinar, por ejemplo, según el siguiente método: Los ensayos enzimáticos se realizan en volúmenes de 200 μ l en pocillos en una placa transparente de 96 pocillos UV, siguiendo el aumento de la absorbancia a 315 nm (pHCA) usando espectrofotometría o HPLC con detección UV. Las mezclas de reacción contienen 2 μ g de proteína purificada y se inician mediante la adición de tirosina 1 mM o 6 mM después del equilibrio a 30 °C. La actividad enzimática se calcula como U/g, donde U se define como sustrato μ mol convertido por minuto. Los controles negativos no contienen proteínas purificadas. Las constantes cinéticas Km y v_{máx} se determinan a partir de ensayos que contienen tirosina de 1,56 μ M a 200 μ M.

Según otras ciertas realizaciones, un polipéptido adicional para uso en la invención es un polipéptido según f). Por tanto, un polipéptido para uso en la invención puede ser un polipéptido que comprende una secuencia de aminoácidos establecida en SEQ ID NO: 14, 15, 16, 17, 18, 19, 20, 21, 22 o 23 (por ejemplo, SEQ ID NO: 14), en donde 1 o más, tal como 2 o más, 3 o más, 4 o más, 5 o más, 6 o más, 7 o más, 8 o más, 9 o más, 10 o más, 11 o más, 12 o más, 13 o más, 14 o más, 15 o más, 16 o más, 17 o más, 18 o más, 19 o más, 20 o más, 25 o más, 30 o más, 35 o más, 40 o más, 45 o más, 50 o más, 60 o más, 70 o más, 80 o más, 90 o más, 100 o más, 110 o más, 120 o más, 130 o más, 140 o más, o 150 o más, restos de aminoácido están sustituidos, eliminados y/o insertados. Según realizaciones particulares, un polipéptido según f) comprende una secuencia de aminoácidos establecida en SEQ ID NO: 14, 15, 16, 17, 18, 19, 20, 21, 22 o 23 (por ejemplo, SEQ ID NO: 14), en donde aproximadamente 1 a unos 150, como aproximadamente 1 a aproximadamente 140, aproximadamente 1 a aproximadamente 130, aproximadamente 1 a aproximadamente 120, aproximadamente 1 a aproximadamente 110, aproximadamente 1 a aproximadamente 100, aproximadamente 1 a aproximadamente 90, aproximadamente 1 a aproximadamente 80, aproximadamente 1 a aproximadamente 70, aproximadamente 1 a aproximadamente 60, aproximadamente 1 a aproximadamente 50, aproximadamente 1 a aproximadamente 40, aproximadamente 1 a aproximadamente 35, aproximadamente 1 a aproximadamente 30, aproximadamente 1 a aproximadamente 25, aproximadamente 1 a aproximadamente 20, aproximadamente 1 a aproximadamente 15, aproximadamente 1 a aproximadamente 10, aproximadamente 1 a aproximadamente 5, o aproximadamente 1 a aproximadamente 3, restos de aminoácido están sustituidos, eliminados y/o insertados. Según realizaciones más particulares, un polipéptido según f) comprende una secuencia de aminoácidos establecida en SEQ ID NO: 14, 15, 16, 17, 18, 19, 20, 21, 22 o 23 (por ejemplo, SEQ ID NO: 14), en donde aproximadamente 1 a unos 50, aproximadamente 1 a aproximadamente 40, aproximadamente 1 a aproximadamente 35, aproximadamente 1 a aproximadamente 30, aproximadamente 1 a aproximadamente 25, aproximadamente 1 a aproximadamente 20, aproximadamente 1 a aproximadamente 15, aproximadamente 1 a aproximadamente 10, aproximadamente 1 a aproximadamente 5, o aproximadamente 1 a aproximadamente 3, restos de aminoácido están sustituidos, eliminados y/o insertados. Según otras realizaciones más particulares, un polipéptido según f) comprende una secuencia de aminoácidos establecida en SEQ ID NO: 14, 15, 16, 17, 18, 19, 20, 21, 22 o 23 (por ejemplo, SEQ ID NO: 14), en donde aproximadamente 1 a unos 30, como aproximadamente 1 a aproximadamente 25, aproximadamente 1 a aproximadamente 20, aproximadamente 1 a aproximadamente 15, aproximadamente 1 a aproximadamente 10, aproximadamente 1 a aproximadamente 5, o aproximadamente 1 a aproximadamente 3, restos de aminoácido están sustituidos, eliminados y/o insertados. Según otras realizaciones más particulares, un polipéptido según f) comprende una secuencia de aminoácidos establecida en SEQ ID NO: 14, 15, 16, 17, 18, 19, 20, 21, 22 o 23

(por ejemplo, SEQ ID NO: 14), en donde aproximadamente 1 a unos 25, como aproximadamente 1 a aproximadamente 20, aproximadamente 1 a aproximadamente 15, aproximadamente 1 a aproximadamente 10, aproximadamente 1 a aproximadamente 5, o aproximadamente 1 a aproximadamente 3, restos de aminoácido están sustituidos, eliminados y/o insertados.

5 Según realizaciones particulares, un polipéptido según f) comprende una secuencia de aminoácidos establecida en SEQ ID NO: 14, en donde aproximadamente 1 a aproximadamente 150, como aproximadamente 1 a aproximadamente 140, aproximadamente 1 a aproximadamente 130, aproximadamente 1 a aproximadamente 120, aproximadamente 1 a aproximadamente 110, aproximadamente 1 a aproximadamente 100, aproximadamente 1 a aproximadamente 90, aproximadamente 1 a aproximadamente 80, aproximadamente 1 a aproximadamente 70, aproximadamente 1 a aproximadamente 60, aproximadamente 1 a aproximadamente 50, aproximadamente 1 a aproximadamente 40, aproximadamente 1 a aproximadamente 35, aproximadamente 1 a aproximadamente 30, aproximadamente 1 a aproximadamente 25, aproximadamente 1 a aproximadamente 20, aproximadamente 1 a aproximadamente 15, aproximadamente 1 a aproximadamente 10, aproximadamente 1 a aproximadamente 5, o aproximadamente 1 a aproximadamente 3, restos de aminoácido están sustituidos, eliminados y/o insertados. Según realizaciones más particulares, un polipéptido según f) comprende una secuencia de aminoácidos establecida en SEQ ID NO: 14, en donde aproximadamente 1 a aproximadamente 50, aproximadamente 1 a aproximadamente 40, aproximadamente 1 a aproximadamente 35, aproximadamente 1 a aproximadamente 30, aproximadamente 1 a aproximadamente 25, aproximadamente 1 a aproximadamente 20, aproximadamente 1 a aproximadamente 15, aproximadamente 1 a aproximadamente 10, aproximadamente 1 a aproximadamente 5, o aproximadamente 1 a aproximadamente 3, restos de aminoácido están sustituidos, eliminados y/o insertados. Según otras realizaciones más particulares, un polipéptido según f) comprende una secuencia de aminoácidos establecida en SEQ ID NO: 14, en donde aproximadamente 1 a aproximadamente 25, como aproximadamente 1 a aproximadamente 20, aproximadamente 1 a aproximadamente 15, aproximadamente 1 a aproximadamente 10, aproximadamente 1 a aproximadamente 5, o aproximadamente 1 a aproximadamente 3, restos de aminoácido están sustituidos, eliminados y/o insertados. Según otras realizaciones más particulares, un polipéptido según f) comprende una secuencia de aminoácidos establecida en SEQ ID NO: 14, en donde aproximadamente 1 a aproximadamente 25, como aproximadamente 1 a aproximadamente 20, aproximadamente 1 a aproximadamente 15, aproximadamente 1 a aproximadamente 10, aproximadamente 1 a aproximadamente 5, o aproximadamente 1 a aproximadamente 3, restos de aminoácido están sustituidos, eliminados y/o insertados.

Se entiende que los valores anteriores generalmente definen el número total de alteraciones al polipéptido de referencia (por ejemplo, SEQ ID NO: 14). Las alteraciones pueden ser únicamente sustituciones de aminoácidos, ya sean sustituciones conservadas o no conservadas, o ambas. Pueden ser únicamente delecciones de aminoácidos. Pueden ser únicamente inserciones de aminoácidos. Las alteraciones pueden ser una mezcla de estas alteraciones específicas, como sustituciones de aminoácidos e inserciones de aminoácidos.

Preferentemente, un polipéptido según f) tiene actividad tirosina amoniaco liasa. Más preferentemente, un polipéptido según f) tiene una actividad tirosina amoniaco liasa similar a la del polipéptido que comprende una secuencia de aminoácidos establecida en SEQ ID NO: 14, 15, 16, 17, 18, 19, 20, 21, 22 o 23 (por ejemplo, SEQ ID NO: 14). Según cierta realización, un polipéptido según f) tiene una actividad tirosina amoniaco liasa similar a la del polipéptido que comprende la secuencia de aminoácidos establecida en SEQ ID NO: 14. Según ciertas otras realizaciones, un polipéptido según f) tiene una actividad tirosina amoniaco liasa similar a la del polipéptido que comprende la secuencia de aminoácidos establecida en SEQ ID NO: 15. Según ciertas otras realizaciones, un polipéptido según f) tiene una actividad tirosina amoniaco liasa similar a la del polipéptido que comprende la secuencia de aminoácidos establecida en SEQ ID NO: 16. Según ciertas otras realizaciones, un polipéptido según f) tiene una actividad tirosina amoniaco liasa similar a la del polipéptido que comprende la secuencia de aminoácidos establecida en SEQ ID NO: 17. Según ciertas otras realizaciones, un polipéptido según f) tiene una actividad tirosina amoniaco liasa similar a la del polipéptido que comprende la secuencia de aminoácidos establecida en SEQ ID NO: 18. Según ciertas otras realizaciones, un polipéptido según f) tiene una actividad tirosina amoniaco liasa similar a la del polipéptido que comprende la secuencia de aminoácidos establecida en SEQ ID NO: 19. Según ciertas otras realizaciones, un polipéptido según f) tiene una actividad tirosina amoniaco liasa similar a la del polipéptido que comprende la secuencia de aminoácidos establecida en SEQ ID NO: 20. Según ciertas otras realizaciones, un polipéptido según f) tiene una actividad tirosina amoniaco liasa similar a la del polipéptido que comprende la secuencia de aminoácidos establecida en SEQ ID NO: 21. Según ciertas otras realizaciones, un polipéptido según f) tiene una actividad tirosina amoniaco liasa similar a la del polipéptido que comprende la secuencia de aminoácidos establecida en SEQ ID NO: 22. Según ciertas otras realizaciones, un polipéptido según f) tiene una actividad tirosina amoniaco liasa similar a la del polipéptido que comprende la secuencia de aminoácidos establecida en SEQ ID NO: 23. Por actividad de tirosina amoniaco liasa "similar" se entiende que el polipéptido según f) tiene al menos aproximadamente el 10 %, como al menos aproximadamente el 20 %, al menos aproximadamente el 30 %, al menos aproximadamente el 40 %, al menos aproximadamente el 50 %, al menos aproximadamente el 60 %, al menos aproximadamente el 75 %, al menos aproximadamente el 80 %, al menos aproximadamente el 90 %, al menos aproximadamente el 95 %, al menos aproximadamente el 100 %, al menos aproximadamente el 200 %, al menos aproximadamente el 300 %, al menos aproximadamente el 400 %, al menos aproximadamente el 500 %, al menos aproximadamente el 800 %, al menos aproximadamente el 1000 % o al menos aproximadamente el 2000 %, de la actividad amoniaco liasa del polipéptido de referencia (por ejemplo, SEQ ID NO: 14).

La actividad tirosina amoniaco liasa se puede determinar, por ejemplo, según el siguiente método: Los ensayos enzimáticos se realizan en volúmenes de 200 µl en pocillos en una placa transparente de 96 pocillos UV, siguiendo el aumento de la absorbancia a 315 nm (pHCA) usando espectrofotometría o HPLC con detección UV. Las mezclas de reacción contienen 2 µg de proteína purificada y se inician mediante la adición de tirosina 1 mM o 6 mM después del equilibrio a 30 °C. La actividad enzimática se calcula como U/g, donde U se define como sustrato µmol convertido por minuto. Los controles negativos no contienen proteínas purificadas. Las constantes cinéticas Km y v_{máx} se determinan a partir de ensayos que contienen tirosina de 1,56 µM a 200 µM.

En la presente invención además se contempla el uso de un polipéptido adicional (por ejemplo, tercero) que tiene actividad de fenilalanina amoniaco liasa, como una fenilalanina amoniaco liasa (EC 4.3.1.24).

Los polipéptidos pueden emplearse según la invención en forma aislada, como en forma purificada. Los polipéptidos pueden, por ejemplo, ser expresados por una célula hospedadora recombinante, y luego purificados. Las técnicas y medios para la purificación de polipéptidos producidos por una célula hospedadora recombinante se conocen bien en la técnica. Por ejemplo, para facilitar la purificación, un motivo de aminoácido que comprende varios restos de histidina, como al menos 6, puede insertarse en el extremo C- o N-terminal del polipéptido. Un ejemplo no limitante de dicho motivo de aminoácido se proporciona en la SEQ ID NO: 24. Diversos kits de purificación para polipéptidos marcados con histidina están disponibles en fuentes comerciales como Qiagen, Hilden, Alemania; Clontech, Mountain View, CA, USA; Bio-Rad, Hercules, CA, USA y otros. Alternativamente, los polipéptidos pueden sintetizarse químicamente. Las técnicas para la síntesis química de péptidos se conocen bien e incluyen síntesis en fase líquida y síntesis en fase sólida.

Los polipéptidos también pueden emplearse según la invención como parte de una célula hospedadora recombinante. Dichas células hospedadoras recombinantes se describen con más detalles a continuación.

La presente invención también proporciona células hospedadoras recombinantes que comprenden (por ejemplo, que expresan) polipéptidos como se detalla en este documento. Generalmente, los polipéptidos según la invención serán heterólogos para las células hospedadoras, lo que significa que los polipéptidos normalmente no son encontrados ni preparados (es decir, expresados) por las células hospedadoras, pero se derivan de una especie diferente.

Por tanto, la presente invención proporciona una célula hospedadora recombinante que comprende un polipéptido heterólogo que tiene una actividad aril sulfotransferasa, y en donde la célula hospedadora se ha modificado adicionalmente para tener una expresión proteica aumentada de (i) una ATP sulfurilasa, (ii) una APS quinasa y/o (iii) una PAP fosfatasa. Según determinadas realizaciones, una célula hospedadora recombinante según la invención comprende un polipéptido heterólogo seleccionado del grupo que consiste en:

- a) un polipéptido que comprende una secuencia de aminoácidos establecida en SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 o 13 (por ejemplo, SEQ ID NO: 1);
- b) un polipéptido que comprende una secuencia de aminoácidos que tiene al menos el 70 %, como al menos aproximadamente el 75 %, al menos aproximadamente el 80 %, al menos aproximadamente el 85 %, al menos aproximadamente el 90 %, al menos aproximadamente el 93 %, al menos el 95 %, al menos el 96 %, al menos el 97 %, al menos el 98 %, o al menos el 99 %, de identidad de secuencia con la secuencia de aminoácidos establecida en SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 o 13 (por ejemplo, SEQ ID NO: 1); o
- c) un polipéptido que comprende una secuencia de aminoácidos establecida en SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 o 13 (por ejemplo, SEQ ID NO: 1), en donde 1 a 50, como 1 a 40, 1 a 35, 1 a 30, 1 a 25, 1 a 20, 1 a 15, 1 a 10, 1 a 5, 1 a 3, restos de aminoácido están sustituidos, eliminados y/o insertados, y en donde la célula hospedadora se ha modificado adicionalmente para tener una expresión proteica aumentada de (i) una ATP sulfurilasa, (ii) una APS quinasa y/o (iii) una PAP fosfatasa.

Una célula hospedadora recombinante proporcionada y utilizada según la presente invención puede comprender un polipéptido heterólogo que tiene actividad tirosina amoniaco liasa. Según determinadas realizaciones, una célula hospedadora recombinante según la invención comprende un polipéptido heterólogo seleccionado del grupo que consiste en:

- d) un polipéptido que comprende una secuencia de aminoácidos establecida en SEQ ID NO: 14, 15, 16, 17, 18, 19, 20, 21, 22 o 23 (por ejemplo, SEQ ID NO: 14);
- e) un polipéptido que comprende una secuencia de aminoácidos que tiene al menos el 70 %, como al menos aproximadamente el 75 %, al menos aproximadamente el 80 %, al menos aproximadamente el 85 %, al menos aproximadamente el 90 %, al menos aproximadamente el 93 %, al menos el 95 %, al menos el 96 %, al menos el 97 %, al menos el 98 %, o al menos el 99 %, de identidad de secuencia con la secuencia de aminoácidos establecida en SEQ ID NO: 14, 15, 16, 17, 18, 19, 20, 21, 22 o 23 (por ejemplo, SEQ ID NO: 14); o
- f) un polipéptido que comprende una secuencia de aminoácidos establecida en SEQ ID NO: 14, 15, 16, 17, 18, 19, 20, 21, 22 o 23 (por ejemplo, SEQ ID NO: 14), en donde 1 o más, como aproximadamente 1 a aproximadamente

50, aproximadamente 1 a aproximadamente 40, aproximadamente 1 a aproximadamente 35, 1 a 30, 1 a 25, 1 a 20, 1 a 15, 1 a 10, 1 a 5, 1 a 3, restos de aminoácido están sustituidos, eliminados y/o insertados.

5 Según determinadas realizaciones, una célula hospedadora recombinante según la invención reivindicada comprende un primer polipéptido heterólogo que tiene actividad aril sulfotransferasa y un segundo polipéptido heterólogo que tiene actividad tirosina amoniaco liasa. Según realizaciones particulares, una célula hospedadora recombinante según la invención reivindicada comprende un primer polipéptido heterólogo seleccionado de los polipéptidos según los puntos a) a c) como se detalla en este documento, y un segundo polipéptido heterólogo seleccionado de los polipéptidos según los puntos e) a f) como se detalla en este documento.

10 Según realizaciones más particulares, una célula hospedadora recombinante según la invención reivindicada comprende un primer polipéptido heterólogo seleccionado del grupo que consiste en:

15 a) un polipéptido que comprende una secuencia de aminoácidos establecida en SEQ ID NO: 1;

b) un polipéptido que comprende una secuencia de aminoácidos que tiene al menos el 70 %, como al menos aproximadamente el 75 %, al menos aproximadamente el 80 %, al menos aproximadamente el 85 %, al menos aproximadamente el 90 %, al menos aproximadamente el 93 %, al menos el 95 %, al menos el 96 %, al menos el 97 %, al menos el 98 %, o al menos el 99 %, de identidad de secuencia con la secuencia de aminoácidos establecida en SEQ ID NO: 1; o

20 c) un polipéptido que comprende una secuencia de aminoácidos establecida en SEQ ID NO: 1, en donde 1 a 50, como 1 a 40, 1 a 35, 1 a 30, 1 a 25, 1 a 20, 1 a 15, 1 a 10, 1 a 5, 1 a 3, restos de aminoácido están sustituidos, eliminados y/o insertados; y

25 un segundo polipéptido heterólogo seleccionado del grupo que consiste en:

d) un polipéptido que comprende una secuencia de aminoácidos establecida en SEQ ID NO: 14;

30 e) un polipéptido que comprende una secuencia de aminoácidos que tiene al menos el 70 %, como al menos aproximadamente el 75 %, al menos aproximadamente el 80 %, al menos aproximadamente el 85 %, al menos aproximadamente el 90 %, al menos aproximadamente el 93 %, al menos el 95 %, al menos el 96 %, al menos el 97 %, al menos el 98 %, o al menos el 99 %, de identidad de secuencia con la secuencia de aminoácidos establecida en SEQ ID NO: 14; o

35 f) un polipéptido que comprende una secuencia de aminoácidos establecida en SEQ ID NO: 14, en donde 1 a 50, como 1 a 40, 1 a 35, 1 a 30, 1 a 25, 1 a 20, 1 a 15, 1 a 10, 1 a 5, 1 a 3, restos de aminoácido están sustituidos, eliminados y/o insertados.

40 Según determinadas realizaciones, una célula hospedadora recombinante según la invención reivindicada comprende un primer polipéptido heterólogo que tiene actividad aril sulfotransferasa y un polipéptido heterólogo adicional (por ejemplo, tercero) que tiene actividad fenilalanina amoniaco liasa.

45 Las células hospedadoras recombinantes según la invención pueden producirse a partir de cualquier organismo hospedador adecuado, incluyendo microorganismos unicelulares o multicelulares como bacterias, levaduras, hongos, algas y plantas, y organismos eucariotas superiores, incluyendo nemátodos, insectos, reptiles, aves, anfibios y mamíferos.

50 Según determinadas realizaciones, una célula hospedadora recombinante según la invención se selecciona del grupo que consiste en bacterias, levaduras, hongos, algas y plantas.

Según ciertas otras realizaciones, una célula hospedadora recombinante según la invención se selecciona del grupo que consiste en bacterias, levaduras, hongos y algas.

55 Según ciertas otras realizaciones, una célula hospedadora recombinante según la invención se selecciona del grupo que consiste en bacterias, levadura y hongos.

Según ciertas otras realizaciones, una célula hospedadora recombinante según la invención se selecciona del grupo que consiste en bacterias y levadura.

60 Según determinadas realizaciones, una célula hospedadora recombinante según la invención no es una célula vegetal.

65 Las células hospedadoras bacterianas se seleccionan de bacterias Gram-positivas y Gram-negativas. Los ejemplos no limitantes para las células hospedadoras bacterianas Gram negativas incluyen especies del género *Escherichia*, *Erwinia*, *Klebsiella* y *Citrobacter*. Ejemplos no limitantes de células hospedadoras bacterianas Gram-positivas incluyen especies del género *Bacillus*, *Lactococcus*, *Lactobacillus*, *Clostridium*, *Corynebacterium*, *Streptomyces*, *Streptococcus*

y *Cellulomonas*.

Según determinadas realizaciones, la célula hospedadora recombinante es una bacteria, que puede ser una bacteria del género *Bacillus*, *Lactococcus*, *Lactobacillus*, *Clostridium*, *Corynebacterium*, *Geobacillus*, *Thermoanaerobacterium*,
5 *Streptococcus*, *Pseudomonas*, *Streptomyces*, *Escherichia*, *Shigella*, *Acinetobacter*, *Citrobacter*, *Salmonella*, *Klebsiella*,
Enterobacter, *Erwinia*, *Kluyvera*, *Serratia*, *Cedecea*, *Morganella*, *Hafnia*, *Edwardsiella*, *Providencia*, *Proteus* o *Yersinia*.

Según realizaciones particulares, la célula hospedadora recombinante es una bacteria del género *Bacillus*. Ejemplos
10 no limitantes de una bacteria del género *Bacillus* son *Bacillus subtilis*, *Bacillus amyloliquefaciens*, *Bacillus licheniformis*
y *Bacillus mojavensis*. Según realizaciones más particulares, la célula hospedadora recombinante es *Bacillus subtilis*.
Según otras realizaciones más particulares, la célula hospedadora recombinante es *Bacillus licheniformis*.

Según otras realizaciones particulares, la célula hospedadora recombinante es una bacteria del género *Lactococcus*.
15 Un ejemplo no limitante de una bacteria del género *Lactococcus* es *Lactococcus lactis*. Según realizaciones más
particulares, la célula hospedadora recombinante es *Lactococcus lactis*.

Según otras realizaciones particulares, La célula hospedadora recombinante es una bacteria del género
20 *Corynebacterium*. Un ejemplo no limitante de una bacteria del género *Corynebacterium* es *Corynebacterium*
glutamicum. Según realizaciones más particulares, la célula hospedadora recombinante es *Corynebacterium*
glutamicum.

Según otras realizaciones particulares, La célula hospedadora recombinante es una bacteria del género *Streptomyces*.
25 Un ejemplo no limitante de una bacteria del género *Streptomyces* son *Streptomyces lividans*, *Streptomyces coelicolor*
o *Streptomyces griseus*. Según realizaciones más particulares, la célula hospedadora recombinante es *Streptomyces*
lividans.

Según otras realizaciones más particulares, la célula hospedadora recombinante es *Streptomyces coelicolor*. Según
otras realizaciones más particulares, la célula hospedadora recombinante es *Streptomyces griseus*.

30 Según otras realizaciones particulares, la célula hospedadora recombinante es una bacteria del género *Pseudomonas*.
Un ejemplo no limitativo de una bacteria del género *Pseudomonas* es *Pseudomonas putida*. Según realizaciones más
particulares, la célula hospedadora recombinante es *Pseudomonas putida*.

Según otras realizaciones particulares, la célula hospedadora recombinante es una bacteria del género *Geobacillus*.
35 Un ejemplo no limitante de una bacteria del género *Geobacillus* es *Geobacillus thermoglucosidasius* y *Geobacillus*
stearothermophilus. Según realizaciones más particulares, la célula hospedadora recombinante es *Geobacillus*
thermoglucosidasius. Según otras realizaciones más particulares, la célula hospedadora recombinante es *Geobacillus*
stearothermophilus.

40 Según otras realizaciones particulares, la célula hospedadora recombinante es una bacteria del género
Thermoanaerobacterium. Un ejemplo no limitante de una bacteria del género *Pseudomonas* es
Thermoanaerobacterium thermosaccharolyticum. Según realizaciones más particulares, la célula hospedadora
recombinante es *Thermoanaerobacterium thermosaccharolyticum*.

45 Según otras realizaciones particulares, la célula hospedadora recombinante es una bacteria del género *Escherichia*.
Un ejemplo no limitante de una bacteria del género *Escherichia* es *Escherichia coli*. Según realizaciones más
particulares, la célula hospedadora recombinante es *Escherichia coli*.

50 Las células hospedadoras de levadura pueden derivarse de, por ejemplo, *Saccharomyces*, *Pichia*,
Schizosaccharomyces, *Zygosaccharomyces*, *Hansenula*, *Pachyosolen*, *Kluyveromyces*, *Debaryomyces*, *Yarrowia*,
Candida, *Cryptococcus*, *Komagataella*, *Lipomyces*, *Rhodospiridium*, *Rhodotorula* o *Trichosporon*.

Según determinadas realizaciones, la célula hospedadora recombinante es una levadura, que puede ser una levadura
55 del género *Saccharomyces*, *Pichia*, *Schizosaccharomyces*, *Zygosaccharomyces*, *Hansenula*, *Pachyosolen*,
Kluyveromyces, *Debaryomyces*, *Yarrowia*, *Candida*, *Cryptococcus*, *Komagataella*, *Lipomyces*, *Rhodospiridium*,
Rhodotorula o *Trichosporon*.

Según realizaciones particulares, la célula hospedadora recombinante es una levadura del género *Saccharomyces*.
60 Un ejemplo no limitante de una levadura del género *Saccharomyces* es *Saccharomyces cerevisiae*. Según
realizaciones más particulares, la célula hospedadora recombinante es *Saccharomyces cerevisiae*.

Según realizaciones particulares, la célula hospedadora recombinante es una levadura del género *Pichia*. Ejemplos
no limitantes de una levadura del género *Pichia* son *Pichia pastoris* y *Pichia kudriavzevii*. Según realizaciones más
particulares, la célula hospedadora recombinante es *Pichia pastoris*. Según otras realizaciones más particulares, la
65 célula hospedadora recombinante es *Pichia kudriavzevii*.

Las células hospedadoras de hongos pueden derivarse de, p. ej., *Aspergillus*.

5 Según determinadas realizaciones, la célula hospedadora recombinante es un hongo, como un hongo del género *Aspergillus*. Ejemplos no limitantes de un hongo del género *Aspergillus* son *Aspergillus oryzae*, *Aspergillus niger* o *Aspergillus awamarii*. Según realizaciones más particulares, la célula hospedadora recombinante es *Aspergillus oryzae*. Según otras realizaciones más particulares, la célula hospedadora recombinante es *Aspergillus niger*. Según otras realizaciones más particulares, la célula hospedadora recombinante es *Aspergillus awamarii*. Las células hospedadoras de algas pueden derivarse de, p. ej., *Chlamydomonas*, *Haematococcus*, *Phaedactylum*, *Volvox* o *Dunaliella*.

10 Según determinadas realizaciones, la célula hospedadora recombinante es un alga, que puede ser un alga del género *Chlamydomonas*, *Haematococcus*, *Phaedactylum*, *Volvox* o *Dunaliella*. Según realizaciones particulares, la célula hospedadora recombinante es una célula alga del género *Chlamydomonas*. Un ejemplo no limitante de una alga del género *Chlamydomonas* es *Chlamydomonas reinhardtii*.

15 Según realizaciones particulares, la célula hospedadora recombinante es una célula alga del género *Haematococcus*. Un ejemplo no limitante de una alga del género *Haematococcus* es *Haematococcus pluvialis*.

20 Según otras realizaciones particulares, la célula hospedadora recombinante es una célula alga del género *Phaedactylum*. Un ejemplo no limitante de una alga del género *Phaedactylum* es *Phaedactylum tricoratum*.

Una célula hospedadora de planta puede derivarse de, p. ej., soja, colza, girasol, algodón, maíz, tabaco, alfalfa, trigo, cebada, avena, sorgo, lechuga, arroz, brócoli, coliflor, repollo, chirivías, melones, zanahorias, apio, perejil, tomates, patatas, fresas, cacahuètes, uvas, cultivos de semillas de hierba, remolacha azucarera, caña de azúcar, judías, guisantes, centeno, lino, árboles de madera dura, árboles de madera blanda y hierbas forrajeras.

25 Según determinadas realizaciones, la célula hospedadora recombinante es una célula vegetal, como una célula vegetal seleccionada del grupo que consiste en soja, colza, girasol, algodón, maíz, tabaco, alfalfa, trigo, cebada, avena, sorgo, lechuga, arroz, brócoli, coliflor, repollo, chirivías, melones, zanahorias, apio, perejil, tomates, patatas, fresas, cacahuètes, uvas, cultivos de semillas de hierba, remolacha azucarera, caña de azúcar, judías, guisantes, centeno, lino, árboles de madera dura, árboles de madera blanda y hierbas forrajeras.

30 Generalmente, una célula hospedadora recombinante según la invención se ha modificado genéticamente para expresar polipéptidos como se detalla en este documento, lo que significa que en la célula hospedadora se ha introducido moléculas de ácido nucleico exógenas, como moléculas de ADN, que comprenden secuencias de nucleótidos que codifican dichos polipéptidos. Los expertos en la materia conocen técnicas para introducir moléculas de ácido nucleico exógeno, como una molécula de ADN, en las diversas células hospedadoras e incluyen transformación (por ejemplo, choque térmico o transformación natural), transfección, conjugación, electroporación, microinyección y bombardeo de micropartículas.

35 Por tanto, una célula hospedadora recombinante según la invención comprende una molécula de ácido nucleico exógena que comprende secuencias de ácido nucleico que codifican polipéptidos como se detalla en este documento.

40 Para facilitar la expresión de los polipéptidos en la célula hospedadora, las moléculas de ácido nucleico exógenas pueden comprender elementos reguladores adecuados tales como un promotor que es funcional en la célula hospedadora para causar la producción de una molécula de ARNm y que está operativamente unida a la secuencia/s de nucleótidos que codifica dicho polipéptido/s.

45 Los promotores útiles según la invención son cualquier promotor conocido que sea funcional en una célula hospedadora dada para provocar la producción de una molécula de ARNm. La persona experta conoce muchos de estos promotores. Dichos promotores incluyen promotores normalmente asociados con otros genes y/o promotores aislados de cualquier bacteria, levaduras, hongos, alga o célula vegetal. Los expertos en la técnica de biología molecular conocen generalmente el uso de promotores para expresión de proteínas, por ejemplo, véanse Sambrook et al., *Molecular cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., 1989. El promotor empleado puede ser inducible. El término "inducible" usado en el contexto de un promotor significa que el promotor solo dirige la transcripción de una secuencia de nucleótidos unida operativamente si hay un estímulo presente, como un cambio de temperatura o la presencia de una sustancia química ("inductor químico"). Como se usa en este documento, "inducción química" se refiere a la aplicación física de una sustancia exógena o endógena (incluyendo macromoléculas, p. ej., proteínas o ácidos nucleicos) a una célula hospedadora. Esto tiene el efecto de hacer que el promotor diana presente en la célula hospedadora aumente la velocidad de transcripción. Alternativamente, el promotor empleado puede ser constitutivo. El término "constitutivo" utilizado en el contexto de un promotor significa que el promotor es capaz de dirigir la transcripción de una secuencia de nucleótidos unida operativamente en ausencia de estímulo (como choque térmico, productos químicos, etc.).

50 Los ejemplos no limitantes de promotores funcionales en bacterias, como *Bacillus subtilis*, *Lactococcus lactis* o *Escherichia coli*, incluyen promotores constitutivos e inducibles como el promotor T7, los sistemas promotores de beta-lactamasa y lactosa; promotor de fosfatasa alcalina (phoA), un sistema de promotor de triptófano (trp), promotor de

tetraciclina, promotor del fago lambda, promotores de proteínas ribosómicas; y promotores híbridos como el promotor tac. También son adecuados otros promotores bacterianos y sintéticos.

5 Los ejemplos no limitantes de promotores funcionales en levadura, como *Saccharomyces cerevisiae*, incluyen promotor de xilosa, promotores GAL1 y GAL10, promotor TEF1 y promotor pgk1.

10 Los ejemplos no limitantes de promotores funcionales en hongos, como *Aspergillus Oryzae* o *Aspergillus niger*, incluyen promotores derivados del gen que codifica la amilasa TAKA de *Aspergillus oryzae*, α -amilasa neutra de *Aspergillus niger*, α -amilasa estable a ácido de *Aspergillus niger*, glucoamilasa (gluA) de *Aspergillus niger* o *Aspergillus awamsii*, acetamidasa de *Aspergillus niger*, proteasa alcalina de *Aspergillus oryzae*, triosa fosfatasa isomerasa de *Aspergillus oryzae*, proteinasa aspártica de *Rhizopus meihei*, y lipasa de *Rhizopus meihei*.

15 Los ejemplos no limitantes de promotores funcionales en alga, como *Haematococcus pluvialis*, incluir el promotor CaMV35S, el promotor SV40 y el promotor del gen RBCS2 de *Chlamydomonas reinhardtii* y el promotor del gen ARS de *Volvox carteri*. Los ejemplos no limitantes de promotores funcionales en células vegetales incluyen el promotor psbA de *Lactuca sativa*, el promotor psbA del tabaco, el promotor rrn16 PEP+NEP del tabaco, el promotor 35S de CaMV, el promotor 19S, el promotor E8 del tomate, el promotor nos, el promotor Mac y el promotor pet E o el promotor ACT1.

20 Además de un promotor, la molécula exógena de ácido nucleico puede comprender además al menos un elemento regulador seleccionado de una región no traducida 5' (5'UTR) y una región no traducida 3' (3'UTR). Los expertos conocen bien muchas de estas 5' UTR y 3' UTR derivados de procariotas y eucariotas. Dichos elementos reguladores incluyen 5' UTR y 3' UTR normalmente asociados con otros genes, y/o 5' UTR y 3' UTR aislados de cualquier bacteria, levaduras, hongos, alga o célula vegetal.

25 Si la célula hospedadora es un organismo procariota, el 5' UTR generalmente contiene un sitio de unión a ribosomas (RBS), también conocido como la secuencia Shine Dalgarno, que generalmente tiene 3-10 pares de bases cadena arriba del codón de inicio. Mientras tanto, si la célula hospedadora es un organismo eucariota, el 5' UTR generalmente contiene la secuencia consenso de Kozak. Un 5' UTR eucariota también puede contener elementos reguladores de acción cis.

30 La molécula/s de ácido nucleico exógeno puede ser un vector o parte de un vector, como un vector de expresión. Normalmente, dicho vector permanece extracromosómico dentro de la célula hospedadora, lo que significa que se encuentra fuera del núcleo o región nucleóide de la célula hospedadora.

35 Según determinadas realizaciones, una célula hospedadora recombinante según la invención no expresa una aril sulfotransferasa dependiente de PAPS endógena.

40 La presente invención también contempla que la molécula de ácido nucleico exógeno esté integrada de manera estable en el genoma de la célula hospedadora. Medios para una integración estable en el genoma de una célula hospedadora, p. ej., por recombinación homóloga, son bien conocidos por la persona experta.

45 La reacción de sulfatación depende del suministro de sulfato de 5'-fosfosulfato de 3'-fosfoadenosina (PAPS) o transferencia desde otro compuesto sulfatado. Los inventores demostraron que la reacción de sulfatación puede potenciarse mejorando el suministro de PAPS (5'-fosfosulfato de 3'-fosfoadenosina) y, además, mediante la eliminación del producto 5'-fosfato de 3'-fosfoadenosina (PAP). El suministro mejorado se obtiene por desregulación, mutación o sobreexpresión de enzimas que aumentan la concentración de PAPS o también reducen la concentración de PAP. Esto se ejemplifica en el Ejemplo 2, donde una mayor producción de ácido zosterico en *Escherichia coli* se obtiene aumentando la expresión de los genes *cysD*, *cysN*, y *cysC* que son responsables de la producción de PAPS.

50 Sin quedar unido a una teoría específica, se cree que un resto adenililo (AMP) de ATP se transfiere al sulfato para formar sulfato activado, o APS (5'-fosfosulfato de adenosina). Esta reacción extremadamente desfavorable está relacionada cinéticamente y energéticamente a la hidrólisis de GTP por la enzima ATP sulfurilasa, que se compone de dos tipos de subunidades: una adenilil transferasa (*cysD*) y una GTPasa (*cysN*). APS luego se fosforila en el 3'-hidroxilo para formar PAPS (5'-fosfosulfato de 3'-fosfoadenosina) en una reacción catalizada por APS quinasa, que está codificado por *cysC*. Adicionalmente, los inventores mejoraron aún más la producción de ácido zosterico al aumentar la expresión del gen *cysQ* que codifica una PAP fosfatasa que es responsable de la eliminación de PAP.

55 Por tanto, para mejorar aún más la producción de un compuesto fenólico sulfatado, como ácido zosterico, una célula hospedadora recombinante según la presente invención se modifica adicionalmente para tener una expresión proteica aumentada de una ATP sulfurilasa en comparación con una célula hospedadora idéntica que no lleva dicha modificación; o se modifica adicionalmente para tener una expresión proteica aumentada de una APS quinasa en comparación con una célula hospedadora idéntica que no lleva dicha modificación; y/o se modifica adicionalmente para tener una expresión proteica aumentada de una PAP fosfatasa en comparación con una célula hospedadora idéntica que no lleva dicha modificación. Por "expresión proteica aumentada" se entiende que la cantidad de la proteína respectiva producida por la célula hospedadora modificada de este modo aumenta en comparación con una célula hospedadora idéntica que no lleva dicha modificación. Más en particular, por "aumentar la expresión" se entiende que

60

65

la cantidad de proteína respectiva producida por la célula hospedadora modificada de este modo aumenta al menos el 10 %, como al menos el 20 %, al menos el 30 %, al menos el 40 %, al menos el 50 %, al menos el 60 %, al menos el 70 %, al menos el 80 %, al menos el 90 %, al menos el 100 %, al menos el 150 %, al menos el 200 %, al menos el 300 %, al menos el 400 %, al menos el 500 %, al menos el 600 %, al menos el 700 %, al menos 800 %, al menos 900 %, al menos aproximadamente el 1000 %, al menos aproximadamente el 2000 %, al menos aproximadamente el 3000 %, al menos aproximadamente el 4000 %, al menos aproximadamente el 5000 %, al menos aproximadamente el 6000 %, al menos aproximadamente el 7000 %, al menos aproximadamente el 8000 %, al menos aproximadamente el 9000 % o al menos aproximadamente el 10000 %, comparado con una célula hospedadora idéntica que no lleva dicha modificación. La cantidad de proteína en una célula dada puede determinarse mediante cualquier técnica de cuantificación adecuada conocida en la técnica, como ELISA, Inmunohistoquímica o transferencia de Western.

Según determinadas realizaciones, una célula hospedadora recombinante según la invención se modificó adicionalmente para tener una expresión proteica aumentada de una ATP sulfurilasa en comparación con una célula hospedadora idéntica que no lleva dicha modificación.

Según determinadas realizaciones, una célula hospedadora recombinante según la invención se modificó adicionalmente para tener una expresión proteica aumentada de una APS quinasa en comparación con una célula hospedadora idéntica que no lleva dicha modificación.

Según determinadas realizaciones, una célula hospedadora recombinante según la invención se modificó adicionalmente para tener una expresión proteica aumentada de una PAP fosfatasa en comparación con una célula hospedadora idéntica que no lleva dicha modificación.

Mediante cualquier medio adecuado conocido por los expertos en la materia se puede lograr un aumento de la expresión de proteínas. Por ejemplo, se puede lograr un aumento de la expresión de proteínas aumentando el número de copias del gen o genes que codifican la proteína respectiva (por ejemplo, ATP sulfurilasa, APS quinasa y/o PAP fosfatasa) en la célula hospedadora, como mediante el uso (por ejemplo, la introducción en la célula hospedadora) de vectores que comprenden el gen o genes unidos operativamente a un promotor que es funcional en la célula hospedadora para causar la producción de una molécula de ARNm. También se puede lograr un aumento de la expresión de proteínas mediante la integración de al menos una segunda copia del gen o genes que codifican la proteína respectiva en el genoma de la célula hospedadora. También se puede lograr un aumento de la expresión de proteínas aumentando la fuerza del promotor o los promotores unidos operativamente al gen o genes. También se puede lograr un aumento de la expresión de proteínas modificando el sitio de unión al ribosoma en la molécula de ARNm que codifica la proteína respectiva (por ejemplo, ATP sulfurilasa, APS quinasa y/o PAP fosfatasa). Al modificar la secuencia del sitio de unión al ribosoma se puede aumentar la velocidad de inicio de la traducción, aumentando así la eficacia de la traducción.

Los genes que codifican ATP sulfurilasa para usar según la invención pueden ser, por ejemplo, los genes *cysD* y *cysN* de *Escherichia coli* (que codifica SEQ ID NO: 25 y 26, respectivamente). Los genes codificantes alternativos de ATP sulfurilasa incluyen el gen ATP sulfurilasa ASAL de *Arabidopsis thaliana* (N.º U40715 de Registro en GenBank, Logan *et al.* (1996) *J Biol Chem* 271: 12227); el gen ATP-sulfurilasa de la cepa *Allium* (N.º AF21154 de Registro en GenBank); el gen ATP sulfurilasa de *Lotus japonicus* (N.º AW164083 de Registro en GenBank); el gen met3-1 ATP sulfurilasa de *Arabidopsis thaliana* (N.º X79210 de Registro en GenBank).

Según determinadas realizaciones, una célula hospedadora recombinante según la invención comprende una molécula de ácido nucleico exógena (como un vector) que comprende una o más secuencias de nucleótidos que codifican una ATP sulfurilasa.

Según realizaciones particulares, una célula hospedadora recombinante según la invención comprende una molécula de ácido nucleico exógena (como un vector) que comprende una secuencia de nucleótidos que codifica i) un polipéptido que comprende una secuencia de aminoácidos establecida en SEQ ID NO: 25 o ii) un polipéptido que comprende un amino secuencia ácida que tiene al menos aproximadamente el 70 %, como al menos aproximadamente el 75 %, al menos aproximadamente el 80 %, al menos aproximadamente el 85 %, al menos aproximadamente el 90 %, al menos aproximadamente el 93 %, al menos el 95 %, al menos el 96 %, al menos el 97 %, al menos el 98 %, o al menos el 99 %, de identidad de secuencia con la secuencia de aminoácidos establecida en SEQ ID NO: 25, proporcionar que la identidad de secuencia no sea del 100 %, y una secuencia de nucleótidos que codifica iii) un polipéptido que comprende una secuencia de aminoácidos establecida en la SEQ ID NO: 26 o iv) un polipéptido que comprende una secuencia de aminoácidos que tiene al menos aproximadamente el 70 %, como al menos aproximadamente el 75 %, al menos aproximadamente el 80 %, al menos aproximadamente el 85 %, al menos aproximadamente el 90 %, al menos aproximadamente el 93 %, al menos el 95 %, al menos el 96 %, al menos el 97 %, al menos el 98 %, o al menos el 99 %, de identidad de secuencia con la secuencia de aminoácidos establecida en SEQ ID NO: 26, siempre que la identidad de secuencia no sea del 100 %. Preferentemente, los polipéptidos se ensamblan para formar una proteína que tiene actividad ATP sulfurilasa.

Según realizaciones particulares, una célula hospedadora recombinante según la invención comprende una molécula

de ácido nucleico exógena (como un vector) que comprende una secuencia de nucleótidos que codifica un polipéptido que comprende una secuencia de aminoácidos establecida en SEQ ID NO: 25 y una secuencia de nucleótidos que codifica un polipéptido que comprende una amino secuencia ácida establecida en SEQ ID NO: 26.

5 Según realizaciones particulares, una célula hospedadora recombinante según la invención comprende una molécula de ácido nucleico exógena (como un vector) que comprende una secuencia de nucleótidos que codifica un polipéptido que comprende una secuencia de aminoácidos que tiene al menos aproximadamente el 70 %, como al menos aproximadamente el 75 %, al menos aproximadamente el 80 %, al menos aproximadamente el 85 %, al menos aproximadamente el 90 %, al menos aproximadamente el 93 %, al menos el 95 %, al menos el 96 %, al menos el 97 %, al menos el 98 %, o al menos el 99 %, de identidad de secuencia con la secuencia de aminoácidos establecida en SEQ ID NO: 25, proporcionar que la identidad de secuencia no sea del 100 %, y una secuencia de nucleótidos que codifica un polipéptido que comprende una secuencia de aminoácidos que tiene al menos aproximadamente el 70 %, como al menos aproximadamente el 75 %, al menos aproximadamente el 80 %, al menos aproximadamente el 85 %, al menos aproximadamente el 90 %, al menos aproximadamente el 93 %, al menos el 95 %, al menos el 96 %, al menos el 97 %, al menos el 98 %, o al menos el 99 %, de identidad de secuencia con la secuencia de aminoácidos establecida en SEQ ID NO: 26, siempre que la identidad de secuencia no sea del 100 %. Preferentemente, los polipéptidos se ensamblan para formar una proteína que tiene actividad ATP sulfúrilasa.

20 Según realizaciones particulares, una célula hospedadora recombinante según la invención comprende una molécula de ácido nucleico exógena (como un vector) que comprende una secuencia de nucleótidos que codifica un polipéptido que comprende una secuencia de aminoácidos que tiene al menos aproximadamente el 85 %, como al menos aproximadamente el 90 %, al menos aproximadamente el 93 %, al menos el 95 %, al menos el 96 %, al menos el 97 %, al menos el 98 %, o al menos el 99 %, de identidad de secuencia con la secuencia de aminoácidos establecida en SEQ ID NO: 25, proporcionar que la identidad de secuencia no sea del 100 %, y una secuencia de nucleótidos que codifica un polipéptido que comprende una secuencia de aminoácidos que tiene al menos aproximadamente el 85 %, como al menos aproximadamente el 90 %, al menos aproximadamente el 93 %, al menos el 95 %, al menos el 96 %, al menos el 97 %, al menos el 98 %, o al menos el 99 %, de identidad de secuencia con la secuencia de aminoácidos establecida en SEQ ID NO: 26, siempre que la identidad de secuencia no sea del 100 %. Preferentemente, los polipéptidos se ensamblan para formar una proteína que tiene actividad ATP sulfúrilasa.

30 Un gen codificante alternativo de ATP sulfúrilasa para usar según la invención puede ser, por ejemplo, el gen MET3 de *Saccharomyces cerevisiae* (que codifica SEQ ID NO: 68). Según realizaciones particulares, una célula hospedadora recombinante según la invención comprende una molécula de ácido nucleico exógena (como un vector) que comprende una secuencia de nucleótidos que codifica i) un polipéptido que comprende una secuencia de aminoácidos establecida en SEQ ID NO: 68 o ii) un polipéptido que comprende una amino secuencia ácida que tiene al menos aproximadamente el 70 %, como al menos aproximadamente el 75 %, al menos aproximadamente el 80 %, al menos aproximadamente el 85 %, al menos aproximadamente el 90 %, al menos aproximadamente el 93 %, al menos el 95 %, al menos el 96 %, al menos el 97 %, al menos el 98 %, o al menos el 99 %, de identidad de secuencia con la secuencia de aminoácidos establecida en SEQ ID NO: 68. Preferentemente, el polipéptido según ii) tiene actividad ATP sulfúrilasa.

45 Según realizaciones particulares, una célula hospedadora recombinante según la invención comprende una molécula de ácido nucleico exógeno (como un vector) que comprende una secuencia de nucleótidos que codifica un polipéptido que comprende una secuencia de aminoácidos establecida en SEQ ID NO: 68.

50 Según realizaciones particulares, una célula hospedadora recombinante según la invención comprende una molécula de ácido nucleico exógena (como un vector) que comprende una secuencia de nucleótidos que codifica un polipéptido que comprende una secuencia de aminoácidos que tiene al menos aproximadamente el 70 %, como al menos aproximadamente el 75 %, al menos aproximadamente el 80 %, al menos aproximadamente el 85 %, al menos aproximadamente el 90 %, al menos aproximadamente el 93 %, al menos el 95 %, al menos el 96 %, al menos el 97 %, al menos el 98 %, o al menos el 99 %, de identidad de secuencia con la secuencia de aminoácidos establecida en SEQ ID NO: 68. Preferentemente, el polipéptido tiene actividad ATP sulfúrilasa.

55 Según realizaciones particulares, una célula hospedadora recombinante según la invención comprende una molécula de ácido nucleico exógena (como un vector) que comprende una secuencia de nucleótidos que codifica un polipéptido que comprende una secuencia de aminoácidos que tiene al menos aproximadamente el 85 %, como al menos aproximadamente el 90 %, al menos aproximadamente el 93 %, al menos el 95 %, al menos el 96 %, al menos el 97 %, al menos el 98 %, o al menos el 99 %, de identidad de secuencia con la secuencia de aminoácidos establecida en SEQ ID NO: 68. Preferentemente, el polipéptido tiene actividad ATP sulfúrilasa.

60 Un gen codificante de ATP sulfúrilasa alternativo para uso según la invención puede ser, por ejemplo, el gen que codifica ATP sulfúrilasa de *Bacillus subtilis* (que codifica SEQ ID NO: 73).

65 Según realizaciones particulares, una célula hospedadora recombinante según la invención comprende una molécula de ácido nucleico exógena (como un vector) que comprende una secuencia de nucleótidos que codifica i) un polipéptido que comprende una secuencia de aminoácidos establecida en SEQ ID NO: 73 o ii) un polipéptido que

comprende un amino secuencia ácida que tiene al menos aproximadamente el 70 %, como al menos aproximadamente el 75 %, al menos aproximadamente el 80 %, al menos aproximadamente el 85 %, al menos aproximadamente el 90 %, al menos aproximadamente el 93 %, al menos el 95 %, al menos el 96 %, al menos el 97 %, al menos el 98 %, o al menos el 99 %, de identidad de secuencia con la secuencia de aminoácidos establecida en SEQ ID NO: 73. Preferentemente, el polipéptido según ii) tiene actividad ATP sulfurilasa.

Según realizaciones particulares, una célula hospedadora recombinante según la invención comprende una molécula de ácido nucleico exógeno (como un vector) que comprende una secuencia de nucleótidos que codifica un polipéptido que comprende una secuencia de aminoácidos establecida en SEQ ID NO: 73.

Según realizaciones particulares, una célula hospedadora recombinante según la invención comprende una molécula de ácido nucleico exógena (como un vector) que comprende una secuencia de nucleótidos que codifica un polipéptido que comprende una secuencia de aminoácidos que tiene al menos aproximadamente el 70 %, como al menos aproximadamente el 75 %, al menos aproximadamente el 80 %, al menos aproximadamente el 85 %, al menos aproximadamente el 90 %, al menos aproximadamente el 93 %, al menos el 95 %, al menos el 96 %, al menos el 97 %, al menos el 98 %, o al menos el 99 %, de identidad de secuencia con la secuencia de aminoácidos establecida en SEQ ID NO: 73. Preferentemente, el polipéptido tiene actividad ATP sulfurilasa.

Según realizaciones particulares, una célula hospedadora recombinante según la invención comprende una molécula de ácido nucleico exógena (como un vector) que comprende una secuencia de nucleótidos que codifica un polipéptido que comprende una secuencia de aminoácidos que tiene al menos aproximadamente el 85 %, como al menos aproximadamente el 90 %, al menos aproximadamente el 93 %, al menos el 95 %, al menos el 96 %, al menos el 97 %, al menos el 98 %, o al menos el 99 %, de identidad de secuencia con la secuencia de aminoácidos establecida en SEQ ID NO: 73. Preferentemente, el polipéptido tiene actividad ATP sulfurilasa.

Para facilitar la expresión de los polipéptidos en la célula hospedadora, la molécula de ácido nucleico exógeno puede comprender elementos reguladores adecuados, tales como un promotor que es funcional en la célula hospedadora para causar la producción de una molécula de ARNm y que está unido operativamente a las secuencias de nucleótidos que codifican dichos polipéptidos.

Un gen que codifica APS quinasa para uso según la invención puede ser, por ejemplo, el gen *cysC* de *Escherichia coli* (que codifica SEQ ID NO: 27).

En ciertos casos, se ha demostrado que un único polipéptido posee una actividad ATP sulfurilasa y una actividad 5'-adenililsulfato quinasa. Por ejemplo, se ha aislado un gen que codifica ATP sulfurilasa/APS quinasa de fuentes de ratón (N.º U34883 de Registro en GenBank, Li *et al.* (1995) J Biol Chem 70: 1945) y humano (N.º AF033026 de Registro en GenBank, Yanagisawa (1998) Biosci Biotechnol Biochem 62: 1037). Otros ejemplos de dicha enzima bifuncional incluyen enzimas 3'-fosfoadenosina 5'-fosfosulfato sintasa (PAPSS) de rata (*Rattus norvegicus*) (SEQ ID NO: 71 o 72).

Según determinadas realizaciones, una célula hospedadora recombinante según la invención comprende una molécula de ácido nucleico exógena (como un vector) que comprende una secuencia de nucleótidos que codifica una APS quinasa.

Según realizaciones particulares, una célula hospedadora recombinante según la invención comprende una molécula de ácido nucleico exógena (como un vector) que comprende una secuencia de nucleótidos que codifica i) un polipéptido que comprende una secuencia de aminoácidos establecida en SEQ ID NO: 27 o ii) un polipéptido que comprende un amino secuencia ácida que tiene al menos aproximadamente el 70 %, como al menos aproximadamente el 75 %, al menos aproximadamente el 80 %, al menos aproximadamente el 85 %, al menos aproximadamente el 90 %, al menos aproximadamente el 93 %, al menos el 95 %, al menos el 96 %, al menos el 97 %, al menos el 98 %, o al menos el 99 %, de identidad de secuencia con la secuencia de aminoácidos establecida en SEQ ID NO: 27, siempre que la identidad de secuencia no sea del 100 %. Preferentemente, dicho polipéptido según ii) tiene actividad APS quinasa.

Según realizaciones particulares, una célula hospedadora recombinante según la invención comprende una molécula de ácido nucleico exógeno (como un vector) que comprende una secuencia de nucleótidos que codifica un polipéptido que comprende una secuencia de aminoácidos establecida en SEQ ID NO: 27.

Según realizaciones particulares, una célula hospedadora recombinante según la invención comprende una molécula de ácido nucleico exógena (como un vector) que comprende una secuencia de nucleótidos que codifica un polipéptido que comprende una secuencia de aminoácidos que tiene al menos aproximadamente el 70 %, como al menos aproximadamente el 75 %, al menos aproximadamente el 80 %, al menos aproximadamente el 85 %, al menos aproximadamente el 90 %, al menos aproximadamente el 93 %, al menos el 95 %, al menos el 96 %, al menos el 97 %, al menos el 98 %, o al menos el 99 %, de identidad de secuencia con la secuencia de aminoácidos establecida en SEQ ID NO: 27, siempre que la identidad de secuencia no sea del 100 %. Preferentemente, dicho polipéptido tiene actividad de APS quinasa.

Según realizaciones particulares, una célula hospedadora recombinante según la invención comprende una molécula de ácido nucleico exógena (como un vector) que comprende una secuencia de nucleótidos que codifica un polipéptido que comprende una secuencia de aminoácidos que tiene al menos aproximadamente el 85 %, como al menos aproximadamente el 90 %, al menos aproximadamente el 93 %, al menos el 95 %, al menos el 96 %, al menos el 97 %, al menos el 98 %, o al menos el 99 %, de identidad de secuencia con la secuencia de aminoácidos establecida en SEQ ID NO: 27, siempre que la identidad de secuencia no sea del 100 %. Preferentemente, dicho polipéptido tiene actividad de APS quinasa.

Un gen codificante alternativo de APS quinasa para uso según la invención puede ser, por ejemplo, el gen MET14 de *Saccharomyces cerevisiae* (que codifica SEQ ID NO: 69).

Según realizaciones particulares, una célula hospedadora recombinante según la invención comprende una molécula de ácido nucleico exógena (como un vector) que comprende una secuencia de nucleótidos que codifica i) un polipéptido que comprende una secuencia de aminoácidos establecida en SEQ ID NO: 69 o ii) un polipéptido que comprende un amino secuencia ácida que tiene al menos aproximadamente el 70 %, como al menos aproximadamente el 75 %, al menos aproximadamente el 80 %, al menos aproximadamente el 85 %, al menos aproximadamente el 90 %, al menos aproximadamente el 93 %, al menos el 95 %, al menos el 96 %, al menos el 97 %, al menos el 98 %, o al menos el 99 %, de identidad de secuencia con la secuencia de aminoácidos establecida en SEQ ID NO: 69. Preferentemente, dicho polipéptido según ii) tiene actividad APS quinasa.

Según realizaciones particulares, una célula hospedadora recombinante según la invención comprende una molécula de ácido nucleico exógeno (como un vector) que comprende una secuencia de nucleótidos que codifica un polipéptido que comprende una secuencia de aminoácidos establecida en SEQ ID NO: 69.

Según realizaciones particulares, una célula hospedadora recombinante según la invención comprende una molécula de ácido nucleico exógena (como un vector) que comprende una secuencia de nucleótidos que codifica un polipéptido que comprende una secuencia de aminoácidos que tiene al menos aproximadamente el 70 %, como al menos aproximadamente el 75 %, al menos aproximadamente el 80 %, al menos aproximadamente el 85 %, al menos aproximadamente el 90 %, al menos aproximadamente el 93 %, al menos el 95 %, al menos el 96 %, al menos el 97 %, al menos el 98 %, o al menos el 99 %, de identidad de secuencia con la secuencia de aminoácidos establecida en SEQ ID NO: 69. Preferentemente, dicho polipéptido tiene actividad de APS quinasa.

Según realizaciones particulares, una célula hospedadora recombinante según la invención comprende una molécula de ácido nucleico exógena (como un vector) que comprende una secuencia de nucleótidos que codifica un polipéptido que comprende una secuencia de aminoácidos que tiene al menos aproximadamente el 85 %, como al menos aproximadamente el 90 %, al menos aproximadamente el 93 %, al menos el 95 %, al menos el 96 %, al menos el 97 %, al menos el 98 %, o al menos el 99 %, de identidad de secuencia con la secuencia de aminoácidos establecida en SEQ ID NO: 69. Preferentemente, dicho polipéptido tiene actividad de APS quinasa.

Un gen codificante alternativo de APS quinasa para uso según la invención puede ser, por ejemplo, el gen codificante de APS quinasa de *Bacillus subtilis* (que codifica SEQ ID NO: 74).

Según realizaciones particulares, una célula hospedadora recombinante según la invención comprende una molécula de ácido nucleico exógena (como un vector) que comprende una secuencia de nucleótidos que codifica i) un polipéptido que comprende una secuencia de aminoácidos establecida en SEQ ID NO: 74 o ii) un polipéptido que comprende un amino secuencia ácida que tiene al menos aproximadamente el 70 %, como al menos aproximadamente el 75 %, al menos aproximadamente el 80 %, al menos aproximadamente el 85 %, al menos aproximadamente el 90 %, al menos aproximadamente el 93 %, al menos el 95 %, al menos el 96 %, al menos el 97 %, al menos el 98 %, o al menos el 99 %, de identidad de secuencia con la secuencia de aminoácidos establecida en SEQ ID NO: 74. Preferentemente, dicho polipéptido según ii) tiene actividad APS quinasa.

Según realizaciones particulares, una célula hospedadora recombinante según la invención comprende una molécula de ácido nucleico exógeno (como un vector) que comprende una secuencia de nucleótidos que codifica un polipéptido que comprende una secuencia de aminoácidos establecida en SEQ ID NO: 74.

Según realizaciones particulares, una célula hospedadora recombinante según la invención comprende una molécula de ácido nucleico exógena (como un vector) que comprende una secuencia de nucleótidos que codifica un polipéptido que comprende una secuencia de aminoácidos que tiene al menos aproximadamente el 70 %, como al menos aproximadamente el 75 %, al menos aproximadamente el 80 %, al menos aproximadamente el 85 %, al menos aproximadamente el 90 %, al menos aproximadamente el 93 %, al menos el 95 %, al menos el 96 %, al menos el 97 %, al menos el 98 %, o al menos el 99 %, de identidad de secuencia con la secuencia de aminoácidos establecida en SEQ ID NO: 74. Preferentemente, dicho polipéptido tiene actividad de APS quinasa.

Según realizaciones particulares, una célula hospedadora recombinante según la invención comprende una molécula de ácido nucleico exógena (como un vector) que comprende una secuencia de nucleótidos que codifica un polipéptido

que comprende una secuencia de aminoácidos que tiene al menos aproximadamente el 85 %, como al menos aproximadamente el 90 %, al menos aproximadamente el 93 %, al menos el 95 %, al menos el 96 %, al menos el 97 %, al menos el 98 %, o al menos el 99 %, de identidad de secuencia con la secuencia de aminoácidos establecida en SEQ ID NO: 74. Preferentemente, dicho polipéptido tiene actividad de APS quinasa.

5 Alternativamente, se puede usar un polipéptido que tiene una actividad de actividad ATP sulfurilasa y APS quinasa, tal como una 3'-fosfoadenosina 5'-fosfosulfato sintasa (PAPSS).

10 Según determinadas realizaciones, una célula hospedadora recombinante según la invención comprende una molécula de ácido nucleico exógena (como un vector) que comprende una secuencia de nucleótidos que codifica una 3'-fosfoadenosina 5'-fosfosulfato sintasa.

15 Según realizaciones particulares, una célula hospedadora recombinante según la invención comprende una molécula de ácido nucleico exógena (como un vector) que comprende una secuencia de nucleótidos que codifica i) un polipéptido que comprende una secuencia de aminoácidos establecida en SEQ ID NO: 71 o ii) un polipéptido que comprende un amino secuencia ácida que tiene al menos aproximadamente el 70 %, como al menos aproximadamente el 75 %, al menos aproximadamente el 80 %, al menos aproximadamente el 85 %, al menos aproximadamente el 90 %, al menos aproximadamente el 93 %, al menos el 95 %, al menos el 96 %, al menos el 97 %, al menos el 98 %, o al menos el 99 %, de identidad de secuencia con la secuencia de aminoácidos establecida en SEQ ID NO: 71. Preferentemente, dicho polipéptido según ii) tiene actividad tanto de ATP sulfurilasa como de APS quinasa.

20 Según realizaciones particulares, una célula hospedadora recombinante según la invención comprende una molécula de ácido nucleico exógeno (como un vector) que comprende una secuencia de nucleótidos que codifica un polipéptido que comprende una secuencia de aminoácidos establecida en SEQ ID NO: 71.

30 Según realizaciones particulares, una célula hospedadora recombinante según la invención comprende una molécula de ácido nucleico exógena (como un vector) que comprende una secuencia de nucleótidos que codifica un polipéptido que comprende una secuencia de aminoácidos que tiene al menos aproximadamente el 70 %, como al menos aproximadamente el 75 %, al menos aproximadamente el 80 %, al menos aproximadamente el 85 %, al menos aproximadamente el 90 %, al menos aproximadamente el 93 %, al menos el 95 %, al menos el 96 %, al menos el 97 %, al menos el 98 %, o al menos el 99 %, de identidad de secuencia con la secuencia de aminoácidos establecida en SEQ ID NO: 71. Preferentemente, dicho polipéptido tiene actividad tanto de ATP sulfurilasa como de APS quinasa.

35 Según realizaciones particulares, una célula hospedadora recombinante según la invención comprende una molécula de ácido nucleico exógena (como un vector) que comprende una secuencia de nucleótidos que codifica un polipéptido que comprende una secuencia de aminoácidos que tiene al menos aproximadamente el 85 %, como al menos aproximadamente el 90 %, al menos aproximadamente el 93 %, al menos el 95 %, al menos el 96 %, al menos el 97 %, al menos el 98 %, o al menos el 99 %, de identidad de secuencia con la secuencia de aminoácidos establecida en SEQ ID NO: 71. Preferentemente, dicho polipéptido tiene actividad tanto de ATP sulfurilasa como de APS quinasa.

40 Según realizaciones particulares, una célula hospedadora recombinante según la invención comprende una molécula de ácido nucleico exógena (como un vector) que comprende una secuencia de nucleótidos que codifica i) un polipéptido que comprende una secuencia de aminoácidos establecida en SEQ ID NO: 72 o ii) un polipéptido que comprende un amino secuencia ácida que tiene al menos aproximadamente el 70 %, como al menos aproximadamente el 75 %, al menos aproximadamente el 80 %, al menos aproximadamente el 85 %, al menos aproximadamente el 90 %, al menos aproximadamente el 93 %, al menos el 95 %, al menos el 96 %, al menos el 97 %, al menos el 98 %, o al menos el 99 %, de identidad de secuencia con la secuencia de aminoácidos establecida en SEQ ID NO: 72. Preferentemente, dicho polipéptido según ii) tiene actividad tanto de ATP sulfurilasa como de APS quinasa.

45 Según realizaciones particulares, una célula hospedadora recombinante según la invención comprende una molécula de ácido nucleico exógeno (como un vector) que comprende una secuencia de nucleótidos que codifica un polipéptido que comprende una secuencia de aminoácidos establecida en SEQ ID NO: 72.

50 Según realizaciones particulares, una célula hospedadora recombinante según la invención comprende una molécula de ácido nucleico exógena (como un vector) que comprende una secuencia de nucleótidos que codifica un polipéptido que comprende una secuencia de aminoácidos que tiene al menos aproximadamente el 70 %, como al menos aproximadamente el 75 %, al menos aproximadamente el 80 %, al menos aproximadamente el 85 %, al menos aproximadamente el 90 %, al menos aproximadamente el 93 %, al menos el 95 %, al menos el 96 %, al menos el 97 %, al menos el 98 %, o al menos el 99 %, de identidad de secuencia con la secuencia de aminoácidos establecida en SEQ ID NO: 72. Preferentemente, dicho polipéptido tiene tanto actividad de ATP sulfurilasa como de APS quinasa.

55 Según realizaciones particulares, una célula hospedadora recombinante según la invención comprende una molécula de ácido nucleico exógena (como un vector) que comprende una secuencia de nucleótidos que codifica un polipéptido que comprende una secuencia de aminoácidos que tiene al menos aproximadamente el 85 %, como al menos

aproximadamente el 90 %, al menos aproximadamente el 93 %, al menos el 95 %, al menos el 96 %, al menos el 97 %, al menos el 98 %, o al menos el 99 %, de identidad de secuencia con la secuencia de aminoácidos establecida en SEQ ID NO: 72. Preferentemente, dicho polipéptido tiene actividad tanto de ATP sulfurilasa como de APS quinasa.

5 Para facilitar la expresión del polipéptido en la célula hospedadora, la molécula exógena de ácido nucleico puede comprender elementos reguladores adecuados, como un promotor que es funcional en la célula hospedadora para causar la producción de una molécula de ARNm y que está operativamente unida a la secuencia de nucleótidos que codifica dicho polipéptido.

10 Un gen que codifica PAP fosfatasa para uso según la invención puede ser, por ejemplo, el gen *cysQ* de *Escherichia coli* (que codifica SEQ ID NO: 28).

Según determinadas realizaciones, una célula hospedadora recombinante según la invención comprende una molécula de ácido nucleico exógena (como un vector) que comprende una secuencia de nucleótidos que codifica una PAP fosfatasa.

Según realizaciones particulares, una célula hospedadora recombinante según la invención comprende una molécula de ácido nucleico exógena (como un vector) que comprende una secuencia de nucleótidos que codifica i) un polipéptido que comprende una secuencia de aminoácidos establecida en SEQ ID NO: 28 o ii) un polipéptido que comprende un amino secuencia ácida que tiene al menos aproximadamente el 70 %, como al menos aproximadamente el 75 %, al menos aproximadamente el 80 %, al menos aproximadamente el 85 %, al menos aproximadamente el 90 %, al menos aproximadamente el 93 %, al menos el 95 %, al menos el 96 %, al menos el 97 %, al menos el 98 %, o al menos el 99 %, de identidad de secuencia con la secuencia de aminoácidos establecida en SEQ ID NO: 28, siempre que la identidad de secuencia no sea del 100 %. Preferentemente, dicho polipéptido según ii) tiene actividad de PAP fosfatasa.

Según realizaciones particulares, una célula hospedadora recombinante según la invención comprende una molécula de ácido nucleico exógena (como un vector) que comprende una secuencia de nucleótidos que codifica un polipéptido que comprende una secuencia de aminoácidos establecida en SEQ ID NO: 28.

Según realizaciones particulares, una célula hospedadora recombinante según la invención comprende una molécula de ácido nucleico exógena (como un vector) que comprende una secuencia de nucleótidos que codifica un polipéptido que comprende una secuencia de aminoácidos que tiene al menos aproximadamente el 70 %, como al menos aproximadamente el 75 %, al menos aproximadamente el 80 %, al menos aproximadamente el 85 %, al menos aproximadamente el 90 %, al menos aproximadamente el 93 %, al menos el 95 %, al menos el 96 %, al menos el 97 %, al menos el 98 %, o al menos el 99 %, de identidad de secuencia con la secuencia de aminoácidos establecida en SEQ ID NO: 28, siempre que la identidad de secuencia no sea del 100 %. Preferentemente, dicho polipéptido tiene actividad PAP fosfatasa.

Según realizaciones particulares, una célula hospedadora recombinante según la invención comprende una molécula de ácido nucleico exógena (como un vector) que comprende una secuencia de nucleótidos que codifica un polipéptido que comprende una secuencia de aminoácidos que tiene al menos aproximadamente el 85 %, como al menos aproximadamente el 90 %, al menos aproximadamente el 93 %, al menos el 95 %, al menos el 96 %, al menos el 97 %, al menos el 98 %, o al menos el 99 %, de identidad de secuencia con la secuencia de aminoácidos establecida en SEQ ID NO: 28, siempre que la identidad de secuencia no sea del 100 %. Preferentemente, dicho polipéptido tiene actividad PAP fosfatasa.

Un gen que codifica PAP fosfatasa alternativa para uso según la invención puede ser, por ejemplo, el gen MET22 de *Saccharomyces cerevisiae* (que codifica SEQ ID NO: 70).

Según realizaciones particulares, una célula hospedadora recombinante según la invención comprende una molécula de ácido nucleico exógena (como un vector) que comprende una secuencia de nucleótidos que codifica i) un polipéptido que comprende una secuencia de aminoácidos establecida en SEQ ID NO: 70 o ii) un polipéptido que comprende un amino secuencia ácida que tiene al menos aproximadamente el 70 %, como al menos aproximadamente el 75 %, al menos aproximadamente el 80 %, al menos aproximadamente el 85 %, al menos aproximadamente el 90 %, al menos aproximadamente el 93 %, al menos el 95 %, al menos el 96 %, al menos el 97 %, al menos el 98 %, o al menos el 99 %, de identidad de secuencia con la secuencia de aminoácidos establecida en SEQ ID NO: 70. Preferentemente, dicho polipéptido según ii) tiene actividad de PAP fosfatasa.

Según realizaciones particulares, una célula hospedadora recombinante según la invención comprende una molécula de ácido nucleico exógena (como un vector) que comprende una secuencia de nucleótidos que codifica un polipéptido que comprende una secuencia de aminoácidos establecida en SEQ ID NO: 70.

Según realizaciones particulares, una célula hospedadora recombinante según la invención comprende una molécula de ácido nucleico exógena (como un vector) que comprende una secuencia de nucleótidos que codifica un polipéptido que comprende una secuencia de aminoácidos que tiene al menos aproximadamente el 70 %, como al menos

aproximadamente el 75 %, al menos aproximadamente el 80 %, al menos aproximadamente el 85 %, al menos aproximadamente el 90 %, al menos aproximadamente el 93 %, al menos el 95 %, al menos el 96 %, al menos el 97 %, al menos el 98 %, o al menos el 99 %, de identidad de secuencia con la secuencia de aminoácidos establecida en SEQ ID NO: 70. Preferentemente, dicho polipéptido tiene actividad PAP fosfatasa.

5 Según realizaciones particulares, una célula hospedadora recombinante según la invención comprende una molécula de ácido nucleico exógena (como un vector) que comprende una secuencia de nucleótidos que codifica un polipéptido que comprende una secuencia de aminoácidos que tiene al menos aproximadamente el 85 %, como al menos aproximadamente el 90 %, al menos aproximadamente el 93 %, al menos el 95 %, al menos el 96 %, al menos el 10 97 %, al menos el 98 %, o al menos el 99 %, de identidad de secuencia con la secuencia de aminoácidos establecida en SEQ ID NO: 70. Preferentemente, dicho polipéptido tiene actividad PAP fosfatasa.

Un gen codificante de PAP fosfatasa alternativo para uso según la invención puede ser, por ejemplo, el gen codificante de PAP fosfatasa de *Bacillus subtilis* (que codifica SEQ ID NO: 75).

15 Según realizaciones particulares, una célula hospedadora recombinante según la invención comprende una molécula de ácido nucleico exógena (como un vector) que comprende una secuencia de nucleótidos que codifica i) un polipéptido que comprende una secuencia de aminoácidos establecida en SEQ ID NO: 75 o ii) un polipéptido que comprende un amino secuencia ácida que tiene al menos aproximadamente el 70 %, como al menos aproximadamente el 75 %, al menos aproximadamente el 80 %, al menos aproximadamente el 85 %, al menos aproximadamente el 90 %, al menos aproximadamente el 93 %, al menos el 95 %, al menos el 96 %, al menos el 20 97 %, al menos el 98 %, o al menos el 99 %, de identidad de secuencia con la secuencia de aminoácidos establecida en SEQ ID NO: 75. Preferentemente, dicho polipéptido según ii) tiene actividad de PAP fosfatasa.

25 Según realizaciones particulares, una célula hospedadora recombinante según la invención comprende una molécula de ácido nucleico exógeno (como un vector) que comprende una secuencia de nucleótidos que codifica un polipéptido que comprende una secuencia de aminoácidos establecida en SEQ ID NO: 75.

30 Según realizaciones particulares, una célula hospedadora recombinante según la invención comprende una molécula de ácido nucleico exógena (como un vector) que comprende una secuencia de nucleótidos que codifica un polipéptido que comprende una secuencia de aminoácidos que tiene al menos aproximadamente el 70 %, como al menos aproximadamente el 75 %, al menos aproximadamente el 80 %, al menos aproximadamente el 85 %, al menos aproximadamente el 90 %, al menos aproximadamente el 93 %, al menos el 95 %, al menos el 96 %, al menos el 35 97 %, al menos el 98 %, o al menos el 99 %, de identidad de secuencia con la secuencia de aminoácidos establecida en SEQ ID NO: 75. Preferentemente, dicho polipéptido tiene actividad PAP fosfatasa.

Según realizaciones particulares, una célula hospedadora recombinante según la invención comprende una molécula de ácido nucleico exógena (como un vector) que comprende una secuencia de nucleótidos que codifica un polipéptido que comprende una secuencia de aminoácidos que tiene al menos aproximadamente el 85 %, como al menos aproximadamente el 90 %, al menos aproximadamente el 93 %, al menos el 95 %, al menos el 96 %, al menos el 40 97 %, al menos el 98 %, o al menos el 99 %, de identidad de secuencia con la secuencia de aminoácidos establecida en SEQ ID NO: 75. Preferentemente, dicho polipéptido tiene actividad PAP fosfatasa.

45 Para facilitar la expresión del polipéptido en la célula hospedadora, la molécula exógena de ácido nucleico puede comprender elementos reguladores adecuados, como un promotor que es funcional en la célula hospedadora para causar la producción de una molécula de ARNm y que está operativamente unida a la secuencia de nucleótidos que codifica dicho polipéptido.

50 *Métodos y usos*

La presente invención proporciona procesos para la producción de compuestos fenólicos sulfatados. Particularmente, se proporciona un proceso para la producción de un compuesto fenólico sulfatado que comprende:

55 (i') poner en contacto un medio que comprende un compuesto fenólico con una primera célula hospedadora recombinante; en donde la primera célula hospedadora recombinante comprende un polipéptido heterólogo que tiene una actividad aril sulfotransferasa; o

(ii) poner en contacto un medio que comprende un sustrato de carbono fermentable con una primera célula hospedadora recombinante; en donde la primera célula hospedadora recombinante comprende un polipéptido heterólogo que tiene una actividad aril sulfotransferasa; o

60 (iii) poner en contacto un medio que comprende un precursor de un compuesto fenólico con una primera célula hospedadora recombinante; en donde la primera célula hospedadora recombinante comprende un polipéptido heterólogo que tiene una actividad aril sulfotransferasa, y en donde la célula hospedadora se ha modificado aún más para tener una expresión proteica aumentada de (i) una ATP sulfurilasa, (ii) una APS quinasa y/o (iii) una PAP fosfatasa.

65 Según determinadas realizaciones, el proceso para la producción de un compuesto fenólico sulfatado comprende: (i')

poner en contacto un medio que comprende un compuesto fenólico con una primera célula hospedadora recombinante; en donde la primera célula hospedadora recombinante comprende un polipéptido heterólogo que tiene actividad aril sulfotransferasa.

5 Según otras ciertas realizaciones, el proceso para la producción de un compuesto fenólico sulfatado comprende:
(i'') poner en contacto un medio que comprende un sustrato de carbono fermentable con una primera célula hospedadora recombinante; en donde la primera célula hospedadora recombinante comprende un polipéptido heterólogo que tiene actividad aril sulfotransferasa.

10 Según otras ciertas realizaciones, el proceso para la producción de un compuesto fenólico sulfatado comprende:
(i''') poner en contacto un medio que comprende un precursor de un compuesto fenólico con una primera célula hospedadora recombinante; en donde la primera célula hospedadora recombinante comprende un polipéptido heterólogo que tiene actividad aril sulfotransferasa.

15 El medio empleado puede ser cualquier medio convencional adecuado para cultivar la célula hospedadora en cuestión, y puede formarse según los principios de la técnica anterior. El medio generalmente contendrá todos los nutrientes necesarios para el crecimiento y la supervivencia de la célula hospedadora respectiva, como fuentes de carbono y nitrógeno y otras sales inorgánicas. Los medios adecuados, por ejemplo, medios mínimos o complejos, están disponibles en proveedores comerciales, o pueden prepararse según las recetas publicadas, por ejemplo, el Catálogo de cepas de la Colección Americana de Cultivos Tipo (ATCC). El medio convencional no limitante bien conocido por la persona experta incluye el caldo Luria Bertani (LB), Caldo Sabouraud Dextrosa (SD), Caldo MS, Levadura Peptona Dextrosa, BMMY, GMMY, o caldo de extracto de levadura y malta (YM), que están disponibles comercialmente. Un ejemplo no limitante de medios adecuados para cultivo de células bacterianas, como células de *B. subtilis*, *L. lactis* o *E. coli*, incluyendo medios mínimos y medios enriquecidos como caldo Luria (LB), medios M9, medios M17, medios SA, medios MOPS, Caldo Terrific, YT y otros. Los medios adecuados para cultivo de células eucariotas, las células de levadura, son RPMI 1640, MEM, DMEM, los cuales pueden complementarse con suero y/o factores de crecimiento según lo requiera la célula hospedadora particular que se cultiva. El medio para cultivar células eucariotas también puede ser cualquier tipo de medio mínimo, como medio mínimo de levadura.

30 El sustrato de carbono fermentable puede ser cualquier sustrato de carbono adecuado conocido en la técnica, y en particular cualquier sustrato de carbono usado comúnmente en el cultivo y/o fermentación de microorganismos. Los ejemplos no limitantes de sustratos de carbono fermentables adecuados incluyen carbohidratos (por ejemplo, azúcares C5 como arabinosa o xilosa, o azúcares C6 como glucosa), glicerol, glicerina, acetato, dihidroxiacetona, fuente de un carbono, metanol, metano, aceites, grasas animales, aceites vegetales, ácidos grasos, lípidos, fosfolípidos, glicerolípidos, monoglicéridos, diglicéridos, triglicéridos, fuentes de carbono renovables, polipéptidos (por ejemplo, una proteína o péptido microbiano o vegetal), extracto de levadura, componente de un extracto de levadura, peptona, casaminoácidos o cualquier combinación de dos o más de los anteriores.

40 Según determinadas realizaciones, el sustrato de carbono se selecciona del grupo que consiste en azúcares C5 (como arabinosa o xilosa), azúcares C6 (como glucosa o fructosa), lactosa, sacarosa, glicerol, glicerina, acetato, licor de maíz fermentado, extracto de levadura, componente de un extracto de levadura, peptona, casaminoácidos o combinaciones de los mismos.

45 Según determinadas realizaciones, el medio comprende glucosa.

Según ciertas otras realizaciones, el medio comprende glicerol.

Según ciertas otras realizaciones, el medio comprende acetato.

50 También se contempla usar almidón como sustrato de carbono. Dependiendo del microorganismo utilizado, el metabolismo del almidón puede requerir el suplemento de beta-glucosidasa, como la beta-glucosidasa de *Neurospora crassa*, al medio. Alternativamente, una célula hospedadora de recombinación según la invención además puede modificarse genéticamente para expresar una beta-glucosidasa, como la beta-glucosidasa de *Neurospora crassa*.

55 Cuando se emplea un sustrato de carbono fermentable, es posible que la célula hospedadora recombinante produzca el compuesto fenólico o un precursor del mismo directamente a partir de dicho sustrato de carbono primario.

Según determinadas realizaciones, el proceso comprende además:

60 (ii) cultivar la primera célula hospedadora recombinante en condiciones adecuadas para la producción del compuesto fenólico sulfatado correspondiente.

65 El experto en la materia conoce bien las condiciones adecuadas para cultivar la célula hospedadora respectiva. Generalmente, la célula hospedadora recombinante se cultiva a una temperatura que varía de aproximadamente 23 a aproximadamente 60 °C, tal como de aproximadamente 25 a aproximadamente 40 °C, tal como a aproximadamente 37 °C. El pH del medio puede variar de pH 1,0 a pH 14,0, tal como de aproximadamente pH 1 a aproximadamente pH 2, de aproximadamente pH 4 a aproximadamente pH 11, de aproximadamente pH 5 a aproximadamente pH 10, de

aproximadamente pH 6 a aproximadamente pH 10, o de aproximadamente pH 7 a aproximadamente pH 9,5, por ejemplo a pH 6,0, pH 7,0, pH 7,5, pH 8,0, pH 8,5, pH 9,0, pH 9,5, pH 10,0, pH 10,5 o pH 11,0.

5 El proceso puede comprender además iii) recuperar el compuesto fenólico sulfatado. El compuesto fenólico sulfatado puede recuperarse mediante un método convencional para aislar y purificar compuestos químicos de un medio. Los procedimientos de purificación bien conocidos incluyen métodos de centrifugación o filtración, precipitación y cromatográficos como, por ejemplo, cromatografía de intercambio iónico, cromatografía de filtración en gel, etc.

10 Además se proporciona un proceso para la producción de un compuesto fenólico sulfatado, como ácido zostérico, el método comprende sulfatar un compuesto fenólico, como ácido p-cumárico, usando un polipéptido que tiene actividad aril sulfotransferasa como se detalla en este documento. Dicho polipéptido puede seleccionarse del grupo que consiste en:

15 a) un polipéptido que comprende una secuencia de aminoácidos establecida en SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 o 13 (por ejemplo, SEQ ID NO: 1);

b) un polipéptido que comprende una secuencia de aminoácidos que tiene al menos el 70 %, como al menos aproximadamente el 75 %, al menos aproximadamente el 80 %, al menos aproximadamente el 85 %, al menos aproximadamente el 90 %, al menos aproximadamente el 93 %, al menos el 95 %, al menos el 96 %, al menos el 97 %, al menos el 98 %, o al menos el 99 %, de identidad de secuencia con la secuencia de aminoácidos establecida en SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 o 13 (por ejemplo, SEQ ID NO: 1); o

20 c) un polipéptido que comprende una secuencia de aminoácidos establecida en SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 o 13 (por ejemplo, SEQ ID NO: 1), en donde 1 a 50, como 1 a 40, 1 a 35, 1 a 30, 1 a 25, 1 a 20, 1 a 15, 1 a 10, 1 a 5, 1 a 3, restos de aminoácido están sustituidos, eliminados y/o insertados.

25 El experto en la materia conoce bien las condiciones adecuadas para la reacción de sulfatación. Generalmente, la reacción de sulfatación tiene lugar a una temperatura que varía de aproximadamente 23 a aproximadamente 60 °C, tal como de aproximadamente 25 a aproximadamente 40 °C, tal como a aproximadamente 37 °C. La reacción de desaminación puede tener lugar a un pH que varía de pH 1,0 a pH 14,0, tal como de aproximadamente pH 2 a aproximadamente pH 11, tal como de aproximadamente pH 5 a aproximadamente pH 10, de aproximadamente pH 6 a aproximadamente pH 10, o de aproximadamente pH 7 a aproximadamente pH 9,5, por ejemplo a pH 6,0, pH 7,0, pH 7,5, pH 8,0, pH 8,5, pH 9,0, pH 9,5, pH 10,0, pH 10,5 o pH 11,0.

30 También se desvela el uso de un polipéptido en la sulfatación de un compuesto fenólico, dicho polipéptido teniendo actividad aril sulfotransferasa como se detalla en este documento. Dicho polipéptido puede seleccionarse del grupo que consiste en:

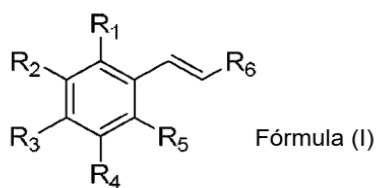
35 a) un polipéptido que comprende una secuencia de aminoácidos establecida en SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 o 13 (por ejemplo, SEQ ID NO: 1);

40 b) un polipéptido que comprende una secuencia de aminoácidos que tiene al menos el 70 %, como al menos aproximadamente el 75 %, al menos aproximadamente el 80 %, al menos aproximadamente el 85 %, al menos aproximadamente el 90 %, al menos aproximadamente el 93 %, al menos el 95 %, al menos el 96 %, al menos el 97 %, al menos el 98 %, o al menos el 99 %, de identidad de secuencia con la secuencia de aminoácidos establecida en SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 o 13 (por ejemplo, SEQ ID NO: 1); o

45 c) un polipéptido que comprende una secuencia de aminoácidos establecida en SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 o 13 (por ejemplo, SEQ ID NO: 1), en donde 1 a 50, como 1 a 40, 1 a 35, 1 a 30, 1 a 25, 1 a 20, 1 a 15, 1 a 10, 1 a 5, 1 a 3, restos de aminoácido están sustituidos, eliminados y/o insertados.

50 Para la finalidad de esta memoria descriptiva y las reivindicaciones adjuntas, debe entenderse que los compuestos fenólicos incluyen aquellos compuestos en donde un grupo hidroxilo está directamente unido a un átomo de carbono bencenoide, y qué compuestos pueden contener o no otros grupos sustituyentes.

55 Según determinadas realizaciones, el compuesto fenólico es un compuesto representado por la fórmula general (I):



en donde al menos uno de R₁, R₂, R₃, R₄, y R₅ son un grupo hidroxilo (-OH);

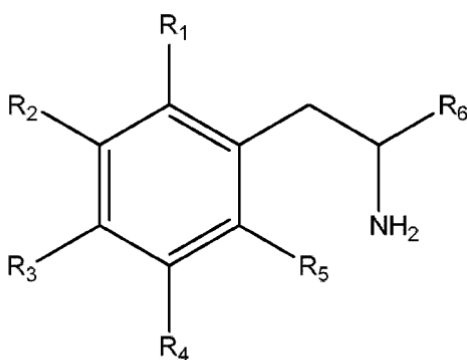
en donde R₁, R₂, R₃, R₄, R₅ y R₆ se seleccionan independientemente del grupo que consiste en haluro, hidrógeno,

hidroxilo (-OH), -OR₇, -OCOR₇, -NR₇R₈, -COR₇, -COOR₇, -SR₇, -OSO₃R₇, -OCSR₇, -POR₇R₈, alquilo, alquenilo, alquinilo, arilo y heteroarilo; en donde R₇, y R₈ se seleccionan independientemente del grupo que consiste en hidrógeno, alquilo, alquenilo, alquinilo, arilo y heteroarilo;

5 en donde R₁, R₂, R₃, R₄, R₅ y R₆, están unidos opcionalmente con un elemento puente Y_n, formando así uno o más anillos, siendo Y_n un enlace o un alquilo C₁₋₁₂ o un arilo, una estructura carbocíclica, heterocíclica o heteroaromática que tiene 1-3 anillos, 3-8 miembros anulares en cada uno y 0 a 4 heteroátomos, o un heteroalquilo que comprende 1 a 12 heteroátomos seleccionados del grupo que consiste en N, O, S, S(O)₁₋₂ y carbonilo, y en donde n es un número entero entre 1 y 12.

10 Los ejemplos específicos de compuestos de Fórmula I incluyen, pero no se limitan a, reservatrol, ácido o-, m- y p-cumárico, ácido cafeico, ácido ferúlico, ácido sinápico, curcumina, ácido rosmarínico, alcohol sináplico, alcohol coniferílico y ácido salvianólico.

15 Un precursor de un compuesto fenólico según la fórmula I puede ser un compuesto representado por la fórmula general (p-I):



Fórmula (p-I);

en donde al menos uno de R₁, R₂, R₃, R₄, y R₅ son un grupo hidroxilo (-OH);

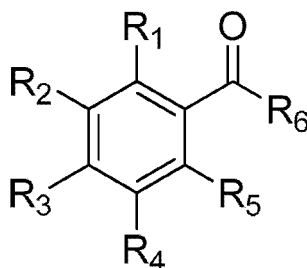
20 en donde R₁, R₂, R₃, R₄, R₅ y R₆ se seleccionan independientemente del grupo que consiste en haluro, hidrógeno, hidroxilo (-OH), -OR₇, -OCOR₇, -NR₇R₈, -COR₇, -COOR₇, -SR₇, -OSO₃R₇, -OCSR₇, -POR₇R₈, alquilo, alquenilo, alquinilo, arilo y heteroarilo; en donde R₇, y R₈ se seleccionan independientemente del grupo que consiste en hidrógeno, alquilo, alquenilo, alquinilo, arilo y heteroarilo;

25 en donde R₁, R₂, R₃, R₄, R₅ y R₆, están unidos opcionalmente con un elemento puente Y_n, formando así uno o más anillos, siendo Y_n un enlace o un alquilo C₁₋₁₂ o un arilo, una estructura carbocíclica, heterocíclica o heteroaromática que tiene 1-3 anillos, 3-8 miembros anulares en cada uno y 0 a 4 heteroátomos, o un heteroalquilo que comprende 1 a 12 heteroátomos seleccionados del grupo que consiste en N, O, S, S(O)₁₋₂ y carbonilo, y en donde n es un número entero entre 1 y 12.

30 Tal precursor puede convertirse en el compuesto fenólico por una célula hospedadora recombinante según la invención, que comprende un polipéptido que tiene actividad tirosina amoniaco liasa. Tal polipéptido eliminará el amoniaco del precursor de Fórmula (p-I) bajo la formación de la molécula correspondiente de Fórmula I. Preferentemente, el precursor p-I es el isómero L.

35 Según ciertas realizaciones, el precursor de un compuesto fenólico como se emplea en la etapa (i''') es un compuesto de la Fórmula general (p-I) como se define en este documento.

Según ciertas otras realizaciones, el compuesto fenólico es un compuesto representado por la fórmula general (II):



Fórmula (II)

40 ;

en donde al menos uno de R₁, R₂, R₃, R₄, y R₅ son un grupo hidroxilo (-OH);

en donde R₁, R₂, R₃, R₄, R₅ y R₆ se seleccionan independientemente del grupo que consiste en haluro, hidrógeno, hidroxilo (-OH), -OR₇, -OCOR₇, -NR₇R₈, -COR₇, -COOR₇, -SR₇, -OSO₃R₇, -OCSR₇, -POR₇R₈, alquilo, alqueno, alquino, arilo y heteroarilo; en donde R₇, y R₈ se seleccionan independientemente del grupo que consiste en

hidrógeno, alquilo, alqueno, alquino, arilo y heteroarilo;

en donde R₁, R₂, R₃, R₄, R₅ y R₆, están unidos opcionalmente con un elemento puente Y_n, formando así uno o más anillos, siendo Y_n un enlace o un alquilo C₁₋₁₂ o un arilo, una estructura carbocíclica, heterocíclica o heteroaromática que tiene 1-3 anillos, 3-8 miembros anulares en cada uno y 0 a 4 heteroátomos, o un heteroalquilo que comprende 1 a 12 heteroátomos seleccionados del grupo que consiste en N, O, S, S(O)₁₋₂ y carbonilo, y en donde n es un número entero entre 1 y 12.

Según determinadas realizaciones, R₆ es -COOR₇.

Según determinadas realizaciones, R₇ es hidrógeno.

Según determinadas realizaciones, R₂ es hidroxilo (-OH).

Según determinadas realizaciones, R₃ es hidroxilo (-OH).

Según determinadas realizaciones, R₄ es hidroxilo (-OH).

Según determinadas realizaciones, cada uno de R₁, R₂, R₄ y R₅ es hidrógeno.

Según ciertos modos de realización, cada uno de R₁, R₂, y R₅ es hidrógeno.

Según realizaciones particulares, el compuesto fenólico es ácido p-cumárico (Fórmula I: R₁ = H, R₂ = H, R₃ = OH, R₄ = H, R₅ = H, R₆ = COOH).

Según otras realizaciones particulares, el compuesto fenólico es el ácido cafeico (Fórmula I: R₁ = H, R₂ = H, R₃ = OH, R₄ = OH, R₅ = H, R₆ = COOH).

Según otras realizaciones particulares, el ácido fenólico es ácido ferúlico (Fórmula I: R₁ = H, R₂ = OCH₃, R₃ = OH, R₄ = H, R₅ = H, R₆ = COOH).

Según otras realizaciones particulares, el ácido fenólico es ácido sinápico (Fórmula I: R₁ = H, R₂ = OCH₃, R₃ = OH, R₄ = OCH₃, R₅ = H, R₆ = COOH).

Según otras realizaciones particulares, el compuesto fenólico es resveratrol (Fórmula I: R₁ = H, R₂ = OH, R₃ = H, R₄ = OH, R₅ = H, R₆ = p-hidroxifenilo).

Según otras realizaciones particulares, el compuesto fenólico es vainillina (Fórmula II: R₁ = H, R₂ = H, R₃ = OH, R₄ = OCH₃, R₅ = H, R₆ = H).

Según determinadas realizaciones, el compuesto fenólico es un ácido hidroxicinámico.

Según determinadas realizaciones, el compuesto fenólico es un compuesto representado por la fórmula general (I), en donde R₁ es hidrógeno; R₂, R₃ y R₄ se seleccionan independientemente del grupo que consiste en hidrógeno (H), hidroxilo (-OH), alquilo C₁₋₆ y alcoxi C₁₋₆, siempre que al menos uno de R₂, R₃ y R₄ sea hidroxilo (-OH); R₅ es hidrógeno y R₆ es COOH. Según determinadas realizaciones, el precursor de un compuesto fenólico como se emplea en la etapa (i'') es un compuesto de la fórmula general (p-I), en donde R₁ es hidrógeno; R₂, R₃ y R₄ se seleccionan independientemente del grupo que consiste en hidrógeno (H), hidroxilo (-OH), alquilo C₁₋₆ y alcoxi C₁₋₆, siempre que al menos uno de R₂, R₃ y R₄ sea hidroxilo (-OH); R₅ es hidrógeno y R₆ es COOH.

Según cierta realización, el compuesto fenólico sulfatado obtenido según la presente invención es el ácido zostérico.

Un experto en la materia conoce bien las moléculas dadoras de sulfato adecuadas metabolizadas por un polipéptido que tiene actividad aril sulfotransferasa. Los ejemplos no limitantes incluyen 5'-fosfosulfato de 3'-fosfoadenosina (PAPS), sulfato de parinitrofenilo (pNPS) y sulfato de 4-metilumbeliferilo (MUS). Dichas moléculas dadoras de sulfato pueden emplearse para facilitar la sulfatación de compuestos fenólicos según la invención.

El medio empleado para cultivar la célula hospedadora recombinante puede ser cualquier medio convencional adecuado para cultivar la célula hospedadora en cuestión, y puede formarse según los principios de la técnica anterior. El medio generalmente contendrá todos los nutrientes necesarios para el crecimiento y la supervivencia de la célula hospedadora respectiva, como fuentes de carbono y nitrógeno y otras sales inorgánicas, como sales de sulfato. Los medios adecuados, por ejemplo, medios mínimos o complejos, están disponibles en proveedores comerciales, o pueden prepararse según las recetas publicadas, por ejemplo, el Catálogo de cepas de la Colección Americana de

Cultivos Tipo (ATCC). El medio convencional no limitante bien conocido por la persona experta incluye el caldo Luria Bertani (LB), Caldo Sabouraud Dextrosa (SD), Caldo MS, Levadura Peptona Dextrosa, BMMY, GMMY, o caldo de extracto de levadura y malta (YM), que están disponibles comercialmente. Un ejemplo no limitante de medios adecuados para cultivar células bacterianas, como células de *B. subtilis*, *L. lactis* o *E. coli*, incluyendo medios mínimos y medios enriquecidos como caldo Luria (LB), medios M9, medios M17, medios SA, medios MOPS, Caldo Terrific, YT y otros. Los medios adecuados para cultivo de células eucariotas, las células de levadura, son RPMI 1640, MEM, DMEM, los cuales pueden complementarse con suero y/o factores de crecimiento según lo requiera la célula hospedadora particular que se cultiva. El medio para cultivar células eucariotas también puede ser cualquier tipo de medio mínimo, como medio mínimo de levadura.

Algunas definiciones

La "actividad aril sulfotransferasa", como se usa en este documento, se refiere a la capacidad de un polipéptido para catalizar el catalizador de la transferencia de un grupo sulfato desde una molécula dadora a una molécula aceptora de arilo.

La "actividad tirosina amoniaco liasa", como se usa en este documento, se refiere a la capacidad de un polipéptido para catalizar la conversión de L-tirosina en ácido p-cumárico.

La "actividad de fenilalanina amoniaco liasa", como se usa en este documento, se refiere a la capacidad de un polipéptido para catalizar la conversión de L-fenilalanina en ácido trans-cinámico.

"ATP sulfurilasa", como se usa en este documento, se refiere a una enzima que cataliza la reacción: $\text{ATP} + \text{sulfato} = \text{difosfato} + 5\text{'-fosfosulfato de adenosina (APS)}$.

"APS quinasa", como se usa en este documento, se refiere a una enzima que cataliza la reacción: $\text{ATP} + 5\text{'-fosfosulfato de adenosina (APS)} = \text{ADP} + 5\text{'-fosfosulfato de 3'-fosfoadenosina (PAPS)}$.

"PAP fosfatasa", como se usa en este documento, se refiere a una enzima que cataliza la reacción: $5\text{'-fosfato de 3'-fosfoadenosina (PAP)} + \text{H}_2\text{O} = \text{AMP} + \text{fosfato}$.

"Polipéptido", o "proteína" se usan indistintamente en este documento para indicar un polímero de al menos dos aminoácidos unidos covalentemente por un enlace amida, independientemente de la longitud o la modificación postraducción (por ejemplo, glucosilación, fosforilación, lipidación, miristilación, ubiquitinación, etc.). Dentro de esta definición se incluyen D- y L-aminoácidos y mezclas de D- y L-aminoácidos.

"Ácido nucleico" o "polinucleótido" se usan indistintamente en este documento para indicar un polímero de al menos dos unidades o bases de monómero de ácido nucleico (por ejemplo, adenina, citosina, guanina, timina) unidos covalentemente por un enlace fosfodiéster, independientemente de la longitud o la modificación de base.

"Recombinante" o "no natural" cuando se usa con referencia a, p. ej., una célula hospedadora, ácido nucleico o polipéptido, se refiere a un material, o un material correspondiente a la forma natural o nativa del material, que ha sido modificado de manera que de otro modo no existiría en la naturaleza, o es idéntico a los mismos pero producido o derivado de materiales sintéticos y/o mediante manipulación usando técnicas recombinantes. Los ejemplos no limitantes incluyen, entre otros, células hospedadoras recombinantes que expresan genes que no se encuentran dentro de la forma nativa (no recombinante) de la célula o que expresan genes nativos que de otro modo se expresan a un nivel diferente.

"Sustitución" o "sustituido" se refiere a la modificación del polipéptido reemplazando un resto de aminoácido con otro, por ejemplo, el reemplazo de un resto de arginina con un resto de glutamina en una secuencia de polipéptidos es una sustitución de aminoácidos.

"Sustitución conservadora" se refiere a una sustitución de un resto de aminoácido con un resto diferente que tiene una cadena lateral similar, y por tanto generalmente implica la sustitución del aminoácido en el polipéptido con aminoácidos dentro de la misma clase de aminoácidos o similar. A modo de ejemplo y no de limitación, un aminoácido con una cadena lateral alifática puede estar sustituido con otro aminoácido alifático, p. ej., alanina, valina, leucina e isoleucina; un aminoácido con cadena lateral hidroxilo está sustituido con otro aminoácido con una cadena lateral hidroxilo, p. ej., serina y treonina; un aminoácido que tiene una cadena lateral aromática se sustituye con otro aminoácido que tiene una cadena lateral aromática, p. ej., fenilalanina, tirosina, triptófano e histidina; un aminoácido con una cadena lateral básica se sustituye con otro aminoácido con una cadena lateral básica, p. ej., lisina y arginina; un aminoácido con una cadena lateral ácida se sustituye con otro aminoácido con una cadena lateral ácida, p. ej., ácido aspártico o ácido glutámico; y un aminoácido hidrófobo o hidrófilo se reemplaza con otro aminoácido hidrófobo o hidrófilo, respectivamente.

"Sustitución no conservativa" se refiere a la sustitución de un aminoácido en un polipéptido con un aminoácido con propiedades de cadena lateral significativamente diferentes. Las sustituciones no conservativas pueden usar

aminoácidos entre, en lugar de dentro de, los grupos definidos y afectan (a) la estructura del esqueleto del péptido en el área de la sustitución (por ejemplo, prolina por glicina) (b) la carga o hidrofobia, o (c) la mayor parte de la cadena lateral. A modo de ejemplo y no de limitación, una sustitución no conservativa a modo de ejemplo puede ser un aminoácido ácido sustituido con un aminoácido básico o alifático; un aminoácido aromático sustituido con un aminoácido pequeño; y un aminoácido hidrófilo sustituido con un aminoácido hidrófobo.

"Delección" o "eliminado" se refiere a la modificación del polipéptido mediante la eliminación de uno o más aminoácidos en el polipéptido de referencia. Las delecciones pueden comprender la eliminación de 1 o más aminoácidos, 2 o más aminoácidos, 5 o más aminoácidos, 10 o más aminoácidos, 15 o más aminoácidos, o 20 o más aminoácidos, hasta el 10 % del número total de aminoácidos, o hasta el 20 % del número total de aminoácidos que forman el polipéptido mientras retienen la actividad enzimática y/o retienen las propiedades mejoradas de una enzima modificada por ingeniería genética. Las delecciones pueden dirigirse a las partes internas y/o partes terminales del polipéptido, en diversas realizaciones, la eliminación puede comprender un segmento continuo o puede ser discontinuo.

"Inserción" o "insertado" se refiere a la modificación del polipéptido mediante la adición de uno o más aminoácidos al polipéptido de referencia. Las inserciones pueden comprender la adición de 1 o más aminoácidos, 2 o más aminoácidos, 5 o más aminoácidos, 10 o más aminoácidos, 15 o más aminoácidos, o 20 o más aminoácidos. Las inserciones pueden estar en las partes internas del polipéptido, o en el extremo carboxi o amino. La inserción puede ser un segmento contiguo de aminoácidos o separado por uno o más de los aminoácidos en el polipéptido de referencia.

"Célula hospedadora", como se usa en este documento, se refiere a una célula viva o microorganismo que es capaz de reproducir su material genético y junto con el material genético recombinante que se ha introducido en él, por ejemplo, mediante transformación heteróloga.

"Expresión" incluye cualquier etapa implicada en la producción de un polipéptido (por ejemplo, enzima codificada) que incluye, pero sin limitación, transcripción, modificación postranscripcional, traducción, modificación postraduccional y secreción.

Como se usa en este documento, "vector" se refiere a una molécula de ácido nucleico capaz de transportar otra molécula de ácido nucleico a la que se ha unido. Un tipo de vector es un "plásmido", que se refiere a un bucle circular de ácido nucleico bicatenario en donde se pueden ligar segmentos adicionales de ácido nucleico. Ciertos vectores son capaces de dirigir la expresión de genes a los que están unidos operativamente. En este documento dichos vectores se denominan "vectores de expresión". Ciertos otros vectores son capaces de facilitar la inserción de una molécula de ácido nucleico exógena en el genoma de una célula hospedadora. En este documento dichos vectores se denominan "vectores de transformación". En general, los vectores de utilidad en las técnicas de ácido nucleico recombinante a menudo están en forma de plásmidos. En la presente memoria descriptiva, "plásmido" y "vector" pueden usarse indistintamente ya que el plásmido es la forma más comúnmente utilizada de un vector. Los expertos en la materia conocen un gran número de vectores adecuados y están disponibles comercialmente.

Como se usa en este documento, "promotor" se refiere a una secuencia de ADN, generalmente cadena arriba (5') de la región codificante de un gen estructural, que controla la expresión de la región codificante proporcionando sitios de reconocimiento y unión para la ARN polimerasa y otros factores que pueden ser necesarios para el inicio de la transcripción. La selección del promotor dependerá de la secuencia de ácido nucleico de interés. Un "promotor funcional en una célula hospedadora" se refiere a un "promotor" que es capaz de soportar el inicio de la transcripción en dicha célula, causando la producción de una molécula de ARNm.

Como se usa en este documento, "unido operativamente" se refiere a una yuxtaposición en donde los componentes descritos están en una relación que les permite funcionar de la manera prevista. Una secuencia de control "unida operativamente" a una secuencia codificante se liga de tal manera que la expresión de la secuencia codificante se consigue en condiciones compatibles con la secuencia de control. Una secuencia promotora está "unida operativamente" a un gen cuando está suficientemente cerca del sitio de inicio de la transcripción de un gen para regular la transcripción del gen.

"Porcentaje de identidad de secuencia", "% de identidad de secuencia" y "porcentaje de identidad" se usan en este documento para referirse a comparaciones entre una secuencia de aminoácidos y una secuencia de aminoácidos de referencia. El "% de identificación de secuencia", como se usa en este documento, se calcula a partir de las dos secuencias de aminoácidos de la siguiente manera: Las secuencias se alinean usando la Versión 9 del GAP (programa de alineación global) del Genetic Computing Group, usando la matriz BLOSUM62 predeterminada (véase a continuación) con una penalización de apertura de hueco de -12 (para el primer nulo de un hueco) y una penalización de extensión de hueco de -4 (por cada nulo adicional en el hueco). Después de la alineación, el porcentaje de identidad se calcula expresando el número de coincidencias como un porcentaje del número de aminoácidos en la secuencia de aminoácidos de referencia.

Se utiliza la siguiente matriz BLOSUM62:

Ala	4																			
Arg	-1	5																		
Asn	-2	0	6																	
Asp	-2	-2	1	6																
Cys	0	-3	-3	-3	9															
Gln	-1	1	0	0	-3	5														
Glu	-1	0	0	2	-4	2	5													
Gly	0	-2	0	-1	-3	-2	-2	6												
His	-2	0	1	-1	-3	0	0	-2	8											
Ile	-1	-3	-3	-3	-1	-3	-3	-4	-3	4										
Leu	-1	-2	-3	-4	-1	-2	-3	-4	-3	2	4									
Lys	-1	2	0	-1	-3	1	1	-2	-1	-3	-2	5								
Met	-1	-1	-2	-3	-1	0	-2	-3	-2	1	2	-1	5							
Phe	-2	-3	-3	-3	-2	-3	-3	-3	-1	0	0	-3	0	6						
Pro	-1	-2	-2	-1	-3	-1	-1	-2	-2	-3	-3	-1	-2	-4	7					
Ser	1	-1	1	0	-1	0	0	0	-1	-2	-2	0	-1	-2	-1	4				
Thr	0	-1	0	-1	-1	-1	-1	-2	-2	-1	-1	-1	-1	-2	-1	1	5			
Trp	-3	-3	-4	-4	-2	-2	-3	-2	-2	-3	-2	-3	-1	1	-4	-3	-2	11		
Tyr	-2	-2	-2	-3	-2	-1	-2	-3	2	-1	-1	-2	-1	3	-3	-2	-2	2	7	
Val	0	-3	-3	-3	-1	-2	-2	-3	-3	3	1	-2	1	-1	-2	-2	0	-3	-1	4

Ala Arg Asn Asp Cys Gln Glu Gly His Ile Leu Lys Met Phe Pro Ser Thr Trp Tyr Val

"Secuencia de referencia" o "secuencia de aminoácidos de referencia" se refiere a una secuencia definida con la que se compara otra secuencia. La secuencia de aminoácidos de referencia puede ser una secuencia de aminoácidos establecida en SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22 o 23.

Los radicales /grupos alifáticos, como se menciona en este documento, están opcionalmente mono- o polisustituidos y pueden ser ramificados o no ramificados, saturados o insaturados. Los grupos alifáticos insaturados, como se define en este documento, incluyen radicales alquilo, alqueno y alquinilo. Los radicales alifáticos preferentes incluyen, pero sin limitación, metilo, etilo, vinilo (etenilo), etinilo, propilo, n-propilo, isopropilo, alilo (2-propenilo), 1-propinilo, metiletilo, butilo, n-butilo, iso-butilo, sec-butilo, *terc*-butil butenilo, butinilo, 1-metilpropilo, 2-metilpropilo, 1,1-dimeteletilo, pentilo, *n*-pentilo, 1,1-dimetilpropilo, 1,2-dimetilpropilo, 2,2-dimetilpropilo, hexilo, 1-metilpentilo, n-heptilo, n-octilo, n-nonilo y n-decilo. Los sustituyentes preferentes para los radicales alifáticos son un grupo alquilo C₁₋₄, un grupo alcoxi C₁₋₆ lineal o ramificado, F, Cl, I, Br, CF₃, CH₂F, CHF₂, CN, OH, SH, NH₂, oxo, (C=O)R', SR', SOR', SO₂R', NHR', NR'R" por lo cual R' y opcionalmente R" para cada sustituyente representa independientemente un grupo alquilo C₁₋₆ lineal o ramificado.

"Alquilo", "radical alquilo" o grupo como se usa en este documento significa hidrocarburos saturados, lineales o ramificados, que pueden ser no sustituidos o mono- o polisustituidos. Por tanto, se entiende que el alquilo insaturado incluye grupos alqueno y alquinilo, como por ejemplo -CH=CH-CH₃ o -C≡C-CH₃, mientras que el alquilo saturado incluye por ejemplo, -CH₃ y -CH₂-CH₃. "alquilo C₁₋₁₂" incluye alquilo C₁₋₂, alquilo C₁₋₃, alquilo C₁₋₄ y alquilo C₁₋₅, alquilo C₁₋₆, alquilo C₁₋₇, alquilo C₁₋₈, alquilo C₁₋₉, alquilo C₁₋₁₀ y alquilo C₁₋₁₁. En estos radicales, alquilo C₁₋₂ representa alquilo C₁₋ o C₂₋, alquilo C₁₋₃ representa alquilo C₁₋, C₂₋ o C₃₋, alquilo C₁₋₄ representa alquilo C₁₋, C₂₋, C₃₋ o C₄₋, alquilo C₁₋₅ representa alquilo C₁₋, C₂₋, C₃₋, C₄₋, o C₅₋, alquilo C₁₋₆ representa alquilo C₁₋, C₂₋, C₃₋, C₄₋, C₅₋ o C₆₋ etc. Los radicales alquilo pueden ser metilo, etilo, vinilo (etenilo), propilo, alilo (2-propenilo), 1-propinilo, metiletilo, butilo, 1-metil propilo, 2-metilpropilo, 1,1-dimeteletilo, pentilo, 1,1-dimetilpropilo, 1,2-dimetilpropilo, 2,2-dimetilpropilo, hexilo, 1-metilpentilo, si se sustituye también CHF₂, CF₃ o CH₂OH etc. Estos radicales alquilo, alqueno o alquinilo pueden estar opcionalmente mono- o polisustituidos por sustituyentes seleccionados independientemente de un grupo alquilo C₁₋₄, un grupo alcoxi C₁₋₆ lineal o ramificado, F, Cl, I, Br, CF₃, CH₂F, CHF₂, CN, OH, SH, NH₂, (C=O)R', SR', SOR', SO₂R', NHR', NR'R" por lo cual R' y opcionalmente R" para cada sustituyente representa independientemente grupo alquilo C₁₋₆ lineal o ramificado.

"Ariilo" o "radical ariilo" como se entiende en este documento significa sistemas de anillo con al menos un anillo aromático pero sin heteroátomos incluso en solo uno de los anillos. Estos radicales ariilo pueden estar opcionalmente mono- o polisustituidos por sustituyentes seleccionados independientemente de un grupo alquilo C₁₋₄, un grupo alcoxi C₁₋₆ lineal o ramificado, un grupo fenilo opcionalmente al menos mono-sustituido, F, Cl, I, Br, CF₃, CH₂F, CHF₂, CN, OH, SH, NH₂, oxo, (C=O)R', SR', SOR', SO₂R', N(C=O)-OR', NHR', NR'R" por lo cual R' y opcionalmente R" para cada sustituyente representa independientemente un grupo alquilo C₁₋₆ lineal o ramificado. Los ejemplos preferentes de radicales ariilo incluyen, pero sin limitación, radicales fenilo, naftilo, fluorantenilo, fluorenilo, tetralinilo o indanilo o antraceno, que opcionalmente pueden estar mono- o polisustituidos, si no se define lo contrario.

"Alquil-ariilo" o "radical alquil-ariilo" como se usa en este documento comprende un grupo lineal o ramificado, opcionalmente al menos una cadena de alquilo mono-sustituida que está unida a un grupo ariilo, como se ha definido anteriormente. Un radical alquil-ariilo preferente es un grupo bencilo, en donde la cadena de alquilo está opcionalmente

ramificada o sustituida. Los sustituyentes preferentes para los radicales alquil-arilo son F, Cl, Br, I, NH₂, SH, OH, SO₂, CF₃, carboxi, amido, ciano, carbamilo, nitro, fenilo, bencilo, -SO₂NH₂, alquilo C₁₋₆ y/o alcoxi C₁₋₆.

5 "Heteroarilo" o "radical heteroarilo", como se usa en este documento, se entiende que significa sistemas de anillo heterocíclico que tienen al menos un anillo aromático y pueden contener opcionalmente uno o más heteroátomos del grupo que consiste en nitrógeno, oxígeno y/o azufre y opcionalmente pueden estar mono- o polisustituidos por sustituyentes seleccionados independientemente de un grupo alquilo C₁₋₄, un grupo alcoxi C₁₋₆ lineal o ramificado, F, Cl, I, Br, CF₃, CH₂F, CHF₂, CN, OH, SH, NH₂, oxo, (C=O)R', SR', SOR', SO₂R', NHR', NR'R" por lo cual R' y 10 opcionalmente R" para cada sustituyente representa independientemente un grupo alquilo C₁₋₆ lineal o ramificado. Los ejemplos preferentes de heteroarilos incluyen, pero sin limitación, furano, benzofurano, tiofeno, benzotiofeno, pirrol, piridina, pirimidina, piridazina, pirazina, quinolina, isoquinolina, ftalazina, benzo-1,2,5-tiadiazol, benzotiazol, indol, benzotriazol, benzodioxolano, benzodioxano, bencimidazol, carbazol y quinazolina.

15 "Alcoxi", "radical alcoxi" o grupo como se usa en este documento significa un "alquilo" singular unido al oxígeno. "Alcoxi C₁₋₆" incluye alcoxi C₁₋₂, alcoxi C₁₋₃, alcoxi C₁₋₄, y alcoxi C₁₋₅, así como alcoxi C₂₋₃, alcoxi C₂₋₄, alcoxi C₂₋₅, alcoxi C₃₋₄, alcoxi C₃₋₅, y alcoxi C₄₋₅. En estos radicales, alcoxi C₁₋₂ representa alcoxi C₁₋ o C₂₋, alcoxi C₁₋₃ representa alcoxi C₁₋, C₂ o C₃₋, alquilo C₁₋₄ representa alcoxi C₁₋, C₂₋, C₃₋ o C₄₋, alcoxi C₁₋₅ representa alcoxi C₁₋, C₂₋, C₃₋, C₄₋ o C₅₋, alcoxi C₁₋₆ representa alcoxi C₁₋, C₂₋, C₃₋, C₄₋, C₅₋ o C₆₋. Los radicales alcoxi pueden ser metoxi, etoxi, propoxi, butoxi, pentiloxi o hexiloxi.

20 La expresión "precursor de un compuesto fenólico" se refiere a cualquier compuesto que pueda convertirse en un compuesto fenólico por las células hospedadoras como se describe en este documento.

25 Cuando se indique un límite numérico o un intervalo en este documento, los puntos extremos se incluyen. Además, todos los valores y sub intervalos dentro de un límite numérico o intervalo se incluyen específicamente como si estuvieran escritos explícitamente.

En los siguientes Ejemplos, los ejemplos 2 y 6 implican realizaciones de la invención:

30 Ejemplos

Ejemplo 1 - Producción de ácido zostérico en *E. coli*

35 Una gama de aril sulfotransferasas que incluye SULT1A1 de *Rattus norvegicus* (SEQ ID NO: 1), SULT1A1 de *Homo sapiens* (SEQ ID NO: 2), SULT1A1 de *Equus caballus* (SEQ ID NO: 3), SULT1A1 de *Sus scrofa domesticus* (SEQ ID NO: 4), SULT1A1 de *Canis lupus familiaris* (SEQ ID NO: 5) y SULT1E1 de *Gallus gallus domesticus* (SEQ ID NO: 6) se expresaron en *Escherichia coli*. Los genes respectivos que codifican SEQ ID NO: 1, 3, 4, 5 y 6 se clonaron amplificados a partir de ADNc de tejido hepático (Zyagen) mediante PCR usando los cebadores enumerados en la 40 Tabla 1. La secuencia de nucleótidos del gen que codifica SEQ ID NO: 2 fue un codón optimizado para la expresión en *Escherichia coli* (GeneArt, Life Technologies) y amplificado por PCR utilizando los cebadores de la Tabla 1. El plásmido pETDuet-1 se digirió con endonucleasas de restricción *NcoI* y *Sall*. Los productos de PCR se clonaron individualmente en el plásmido pETDuet-1 usando la reacción de Gibson (New England Biolabs). Los plásmidos resultantes se transformaron en BL21 (DE3) pLysS (Life Technologies). La Figura 1 muestra el mapa plasmídico del plásmido que codifica SULT1A1 de *Rattus norvegicus* (SEQ ID NO: 1).

45

Tabla 1: Descripción general de enzimas y cebadores para clonación de aril sulfotransferasas

SEQ ID NO	Nombre	Cebador directo	Cebador inverso
1	SULT1A1 <i>Rattus norvegicus</i>	CBJP472	CBJP473
2	SULT1A1 <i>Homo sapiens</i>	CBJP470	CBJP471
3	SULT1A1 <i>Equus caballus</i>	CBJP499	CBJP500
4	SULT1A1 <i>Sus scrofa domesticus</i>	CBJP505	CBJP506
5	SULT1A1 <i>Canis lupus familiaris</i>	CBJP503	CBJP504
6	SULT1E1 <i>Gallus gallus domesticus</i>	CBJP501	CBJP502

Las cepas se cultivaron en medios mínimos M9 que contenían glucosa como fuente de carbono e IPTG 0,1 mM para la inducción de la expresión génica, así como ácido p-cumárico 0,1 mM (pHCA). Después de cuatro días de 50 crecimiento, las muestras se extrajeron por filtración y se analizaron por HPLC.

La concentración de ácido p-cumárico (pHCA) y ácido zostérico en el sobrenadante se cuantificó mediante alto rendimiento (HPLC) y se comparó con los patrones químicos. La HPLC se realizó en una configuración Thermo utilizando una columna HS-F5 y fases móviles: formiato de amonio 5 mM pH 4,0 (A) y acetonitrilo (B) a 1,5 ml min⁻¹, 55 usando un gradiente de elución que comienza en B al 5 %. 0,5 min después de la inyección a 7 min, la fracción de B aumentó linealmente del 5 % al 60 %, y entre 9,5 min y 9,6 la fracción de B disminuyó nuevamente al 5 % y permaneció allí hasta los 12 min. El pHCA y el ácido zostérico se cuantificaron midiendo la absorbancia a 277 nm.

La Tabla 2 muestra el pHCA restante y el ácido zostérico producido en los medios de cultivo. El ácido zostérico se formó con una aril sulfotransferasa heteróloga expresada en un microorganismo ejemplificado por *E. coli* suministrado con el sustrato.

5 **Tabla 2: Producción de ácido zostérico en *E. coli* a partir de pHCA a través de la expresión heteróloga de sulfotransferasas.**

Enzima	pHCA restante (mM)	Ácido zostérico formado (mM)
Sin enzima	0,10	Indetectable
SULT1A1 <i>Rattus norvegicus</i>	0,02	0,10
SULT1A1 <i>Homo sapiens</i>	0,08	0,02
SULT1A1 <i>Equus caballus</i>	0,09	0,01
SULT1A1 <i>Sus scrofa domesticus</i>	0,09	0,01
SULT1A1 <i>Canis lupus familiaris</i>	0,10	0,01
SULT1E1 <i>Gallus gallus domesticus</i>	0,08	0,01

Ejemplo 2 - Aumento de la producción de ácido zostérico en *E. coli*

- 10 La adición de grupos sulfatados a las dianas depende del suministro de la molécula dadora 5'-fosfosulfato de 3'-fosfoadenosina (PAPS). Se examinó si podríamos aumentar la producción de ácido zostérico sobreexpresando enzimas que proporcionan PAPS y una enzima que elimina el producto 5'-fosfato de 3'-Fosfoadenosina(PAP).

Tabla 3: Clonación de enzimas implicadas en la activación de sulfato y eliminación de productos.

Genes	Cebador directo	Cebador inverso
<i>cysDNC</i> solo	CBJP491	CBJP492
<i>cysDNC</i> para operón artificial	CBJP491	CBJP497
<i>cysQ</i> para operón artificial	CBJP498	CBJP496

- 15 En *E. coli*, los genes *cysD* y *cysN* codifican las dos subunidades de ATP sulfurilasa (EC: 2,7.7,4), *cysC* codifica APS quinasa (EC: 2,7.1,25) y *cysQ* codifica una PAP fosfatasa.

- 20 El grupo *cysDNC* se amplificó por PCR con ADN cromosómico MG1655 de *E. coli* utilizando los cebadores que se muestran en la tabla 3. El plásmido pRSFDuet-1 (Life Technologies) fue digerido por las endonucleasas de restricción NdeI y BglIII. El grupo génico se insertó en el plásmido digerido usando la reacción de Gibson (New England Biolabs). La figura 2 muestra el plásmido resultante. Para la expresión combinada de *cysDNC* y *cysQ* en un operón artificial, *cysDNCQ*, las dos partes fueron amplificadas por PCR con ADN cromosómico MG1655 de *E. coli* usando los cebadores que se muestran en la Tabla 3. De nuevo, las partes se insertaron en el plásmido digerido. La Figura 3 muestra los plásmidos resultantes. El plásmido que expresa SULT1A1 de *Homo sapiens* (SEQ ID NO: 2) del ejemplo 1 se co-transformó en células BL21(DE3)pLysS de *E. coli* (DE3) (Life Technologies) con el plásmido que expresa *cysDNC* o *cysDNCQ*.

- 30 Las células se cultivaron como en el Ejemplo 1 y los sobrenadantes se analizaron para la formación del producto como en el ejemplo 1. La cepa que expresa SULT1A1 en combinación con *cysDNCQ* también se cultivó sin adición de IPTG para inducción. La Tabla 4 muestra las concentraciones de pHCA y ácido zostérico.

Tabla 4: Concentraciones de pHCA y ácido zostérico en medios de cultivo con *E. coli* que expresa una aril sulfotransferasa en combinación con *cysDNC* y *cysQ*.

Enzimas	Inducción	pHCA restante (mM)	Ácido zostérico formado (mM)
SULT1A1 <i>Homo sapiens</i>	IPTG 0,1 mM	0,08	0,02
SULT1A1 <i>Homo sapiens</i> , <i>CysDNC</i>	IPTG 0,1 mM	0,06	0,06
SULT1A1 <i>Homo sapiens</i> , <i>CysDNCQ</i>	IPTG 0,1 mM	0,04	0,09
SULT1A1 <i>Homo sapiens</i> , <i>CysDNCQ</i>	Ninguno	0,10	Indetectable

- 35 Esto muestra que una mayor cantidad de pHCA se transforma en ácido zostérico cuando aumenta la expresión proteica de *cysDNC*. Aún se forma más ácido zostérico cuando la expresión de proteína *cysQ* aumenta adicionalmente.

40 **Ejemplo 3 - Se puede formar un producto sulfatado *in vivo* por coexpresión de una ruta heteróloga y una aril sulfotransferasa**

La producción de un producto sulfatado se puede lograr biológicamente mediante la expresión de aril sulfotransferasa como se muestra en el ejemplo 1. El sustrato para sulfatación también puede estar formado por un organismo biológico, y aquí se mostrará para un organismo que expresa una ruta heteróloga que conduce a un compuesto

fenólico y que expresa una sulfotransferasa que actúa sobre el compuesto fenólico.

La enzima RmXAL de *Rhodotorula mucilaginosa/Rhodotorula rubra* (SEQ ID NO: 20) tiene actividad tirosina amoniaco liasa, catalizando así la desaminación no oxidativa del aminoácido tirosina, liberando ácido p-cumárico (pHCA) y amoniaco. El gen que codifica RmXAL fue optimizado por codón usando algoritmos estándar para la expresión en *E. coli* disponible en GeneArt (Life Technologies) y amplificado por PCR utilizando los cebadores que se muestran en la tabla 5 e insertados en el vector pCDFDuet-1 (Novagen/Life Technologies), que había sido digerido por las enzimas de restricción NdeI y BglII, utilizando la reacción de Gibson (New England Biolabs). La Figura 4 proporciona una imagen del plásmido que expresa RmXAL.

Tabla 5: Cebadores utilizados para la clonación de tirosina amoniocoliasa

Genes	Cebador directo	Cebador inverso
RmXAL	CBJP487	CBJP488

El plásmido resultante se co-transformó en células BL21(DE3)pLysS de *E. coli* (Life Technologies) solas o junto con el plásmido que expresa SULT1A1 de *Homo sapiens* (ejemplo 1). Las cepas resultantes se cultivaron en medios M9 con glucosa como fuente de carbono, con IPTG 0,1 mM para inducción de la expresión génica. Para análisis de formación del producto se tomaron muestras como se describió anteriormente (ejemplo 1). La Tabla 6 muestra las concentraciones resultantes de pHCA y ácido zosterico. RmXAL permitió la producción de pHCA sin adición de ningún sustrato, proporcionando así una vía heteróloga desde el metabolismo normal de las células a un producto heterólogo. La expresión adicional de una aril sulfotransferasa, ejemplificada por SULT1A1 de *Homo sapiens*, permitió la conversión *in vivo* de pHCA a ácido zosterico. Por tanto, una aril sulfotransferasa puede actuar sobre un compuesto producido *in vivo* y las células pueden liberar el producto sulfatado resultante al medio.

Tabla 6: Concentraciones de pHCA y ácido zosterico en medios de cultivo con *E. coli* que expresa una aril sulfotransferasa en combinación con una tirosina amoniaco liasa.

Enzimas	pHCA (mM)	Ácido zosterico formado (mM)
RmXAL	0,04	Indetectable
SULT1A1 <i>Homo sapiens</i> , RmXAL	0,02	0,01

Ejemplo 4 - Disminución de la toxicidad del producto sulfatado

MG1655 de *E. coli* se cultivó en medios mínimos M9 definidos químicamente con glucosa al 0,2 % como fuente de carbono sin adición adicional o con adiciones de 10 mM, 20 mM, 25 mM, 30 mM, ácido p-cumárico 35 mM o 40 mM (pHCA), o con 20 mM o 40 mM del éster sulfato de pHCA (ácido zosterico). Todas las preparaciones de medios se habían ajustado a pH 7. Las células se cultivaron a 37 °C con agitación a 250 rpm en un agitador orbital. Las tasas de crecimiento se examinaron siguiendo la densidad óptica a 600 nm. Las tasas de crecimiento resultantes en la fase de crecimiento exponencial se muestran en la Figura 5. Los cuadrados rellenos representan tasas de crecimiento en medios con pHCA. Los cuadrados abiertos representan tasas de crecimiento en medios con ácido zosterico. Y el círculo representa la tasa de crecimiento en los medios sin ninguna de estas adiciones. Es evidente que la presencia de pHCA es tóxica para las células, mientras que el éster sulfato, ácido zosterico es mucho menor.

Ejemplo 5 - Suministro *in vivo* de precursor de producto sulfatado

El sustrato que es sujeto de sulfatación puede suministrarse al medio vacío de dichos precursores o puede ser proporcionado por microorganismos en el medio. Aquí se muestra que el ácido p-cumárico que se sulfata para generar ácido zosterico, puede ser producido *in vivo* por la expresión de una tirosina amoniaco-liasa.

Los genes que codifican las tirosina amoniaco-liasas RcTAL (de *Rhodobacter capsulatus*; SEQ ID NO: 48), RsTAL (de *Rhodobacter sphaeroides*; SEQ ID NO: 17) y FjTAL (de *Flavobacterium johnsoniae*; SEQ ID NO: 14) se clonaron en vectores de expresión como sigue. Los genes (SEQ ID NO: 49, 50 y 51, respectivamente) fueron optimizados para *E. coli* y sintetizados por GeneArt, se amplificaron por PCR usando los oligonucleótidos que se muestran en la tabla a continuación, y se clonaron en pCDFDuet-1 (Novagen): El plásmido se digirió con *NdeI* y *BglII* y se purificó en gel. Los genes se insertaron mediante ensamblaje isotérmico utilizando Gibson Assembly Master Mix (New England Biolabs) y se transformaron en DH5 α químicamente competente (cepa de laboratorio) o NEB5 α (New England Biolabs), seleccionando resistencia a 50 $\mu\text{g ml}^{-1}$ de espectinomicina en medio LB. Los plásmidos resultantes pCBJ215 (RsTAL), pCBJ228 (FjTAL) y pCBJ297 (RcTAL) (Figuras 6 a 8, respectivamente) se co-transformaron por electroporación en la cepa de expresión BL21 de *E. coli* (DE3) (Invitrogen/life Technologies) junto con un plásmido basado en pETDuet-1 que expresa SULT1A1 de rata (Ejemplo 1). Los cultivos de transformación se sembraron en placas sobre LB que contenían 50 $\mu\text{g ml}^{-1}$ de espectinomicina y 100 $\mu\text{g ml}^{-1}$ de ampicilina. También se hizo una cepa de control que portaba pCDFDuet-1.

Cebadores:

Oligonucleótido	Gen	Dirección	Secuencia
CBJP483	RsTAL	Directo	CATCTTAGTATATTAGTTAAGTATAAGAAGGAG ATATACATATGCTGGCAATGAGCCCT
CBJP484	RsTAL	Inverso	TGGCCGGCCGATATCCAATTGATTAACCGGAC TCTGTTGC
CBJP555	FjTAL	Directo	CATCTTAGTATATTAGTTAAGTATAAGAAGGAG ATATACATATGAACACCATCAACGAATATCTG
CBJP556	FjTAL	Inverso	TGGCCGGCCGATATCCAATTGATTAATTGTTAA TCAGGTGGTCTTTTACTTTCTG
CBJP745	RcTAL	Directo	CATCTTAGTATATTAGTTAAGTATAAGAAGGAG ATATACATATGCTGGATGCAACCATTGG
CBJP746	RcTAL	Inverso	TGGCCGGCCGATATCCAATTGATTATGCCGGA GGATCCGCT

Las cepas que albergan plásmidos recombinantes se cultivaron previamente en medio líquido 2xYT con 100 µg ml⁻¹ de ampicilina y 50 µg ml⁻¹ de espectinomicina y se incubaron a 37 °C y 250 rpm durante la noche. Al día siguiente, cada cultivo previo se transfirió a 5 ml de medio mínimo M9 con glucosa al 0,2 %, tirosina 2 mM e IPTG 1 mM para inducción de expresión. Los cultivos se colocaron en una incubadora a 37 °C con agitación a 250 rpm durante la noche. Los sobrenadantes se recogieron luego por centrifugación dos veces y se aplicaron al análisis de HPLC como se describe en el ejemplo 1, y los títulos de ácido p-cumárico (pHCA) y ácido zostérico (ZA) se cuantificaron utilizando patrones químicos y se presentan en la tabla a continuación.

Sulfotransferasa	Tirosina amoniaco-liasa	pHCA µM	ZA µM
SULT1A1 rata	Ninguno	0	0
SULT1A1 rata	RsTAL	78	<1
SULT1A1 rata	RcTAL	20	<1
SULT1A1 rata	FjTAL	398	16

Aquí, es evidente que el ácido zostérico se forma cuando hay un suministro de ácido p-cumárico exógeno o si las células son capaces de producir ácido p-cumárico. En conclusión, se puede formar un producto sulfatado a partir de una molécula precursora no sulfatada, cuando este se produce *in vivo*.

Adicionalmente, los datos muestran sorprendentemente que el empleo de la tirosina amoniaco-liasa de *Flavobacterium johnsoniae* (FjTAL; SEQ ID NO: 14) da como resultado un mayor suministro de molécula precursora no sulfatada (aquí: ácido p-cumárico), que a su vez conduce a un mayor rendimiento de producto sulfatado (aquí: ácido zostérico) en comparación con otras tirosina amoniaco-liasas.

20 Ejemplo 6 - Producción de productos sulfatados en otros hospedadores

Se demostró que se puede producir ácido zostérico *in vivo* en *Escherichia coli* por expresión de una aril sulfotransferasa. Para demostrar que la reacción es posible en otros microorganismos, aquí se muestra que la levadura *Saccharomyces cerevisiae* también se puede utilizar como hospedador para la producción.

El gen que codifica la aril sulfotransferasa SULT1A (Ejemplo 1) se clonó después de un promotor TEF1 en un plásmido episómico con un origen de replicación de 2 micrómetros como sigue. El gen se amplificó por PCR usando los cebadores CBJP633 y CBJP634. Alternativamente, el gen fue optimizado por codón para *E. coli* y sintetizado por GeneArt y amplificado por los cebadores CBJP635 y CBJP636. El promotor TEF1 (Jensen *et al.*, 2014, FEMS Yeast Res 14: 238-248) se amplificó por PCR utilizando los cebadores PTEF1_fw y PTEF1_rv. El plásmido pCfB132 (Jensen *et al.*, mencionado anteriormente) se digirió mediante las enzimas de restricción AsiSI y Nt.BsmI. Los tres fragmentos - plásmido, promotor TEF1 y gen que codifica SULT1A1 - se ensamblaron usando un procedimiento de clonación por escisión de uracilo, dando como resultado los plásmidos pCBJ283 y pCBJ284 (Figuras 9 y 10, respectivamente, que posteriormente se transformó en la cepa CEN.PK102-5B de *Saccharomyces cerevisiae* seleccionando para crecimiento en placas de medios sintéticos que carecen de uracilo. También se realizó una cepa de control por transformación de pCfB132 en CEN.PK102-5B.

Cebadores:

Oligonucleótido	Gen/promotor	Dirección	Secuencia
CBJP633	SULT1A1 rata	Directo	AGTGCAGGUAAAACAATGgagttctcccgtcca
CBJP634	SULT1A1 rata	Inverso	CGTGCGAUTCAtagttcacaacgaaactg
CBJP635	SULT1A1 rata (<i>E. coli</i>)	Directo	ATCTGTCAUAAAACAATGgaatttcacgtccgc
CBJP636	SULT1A1 rata (<i>E. coli</i>)	Inverso	CACGCGAUTCAcagttcacaacgaaattgaa
PTEF1_fw	PTEF1	Directo	Cacgcgauccacacacccatagcttc
PTEF1_rv	PTEF1	Inverso	Cgtgcgauggaagtacctcaaga

Las cepas se cultivaron en medio Delft modificado (Jensen *et al.*, mencionado anteriormente) con 20 mg/ml de histidina y 60 mg/ml de leucina y ácido p-cumárico 10 mM durante la noche a 30 °C con aireación. El sobrenadante se aisló y examinó por HPLC como se describe en el Ejemplo 1. La siguiente tabla muestra que el ácido zostérico (ZA) fue producido por la cepa que expresa SULT1A1 y no por la cepa de control que carece de una sulfotransferasa.

Sulfotransferasa	μMZA (promedios y desviaciones estándar de experimentos repetidos)
Ninguno	0 \pm 0
SULT1A1 rata (nativo)	37,8 \pm 5,7
SULT1A1 rata (codón optimizado para <i>E. coli</i>)	46,2 \pm 3,5

Es evidente que el ácido zostérico se forma solo cuando se expresa una sulfotransferasa en la levadura, y que el gen que lo codifica puede ser natural o codificado por un gen sintético con un codón específico de optimización. En conclusión, las reacciones de sulfatación se muestran catalizadas por sulfotransferasas en *E. coli* también se catalizan cuando las sulfotransferasas se expresan en otros organismos, como se demuestra aquí para la levadura *S. cerevisiae*. La eficacia de la producción puede verse afectada por medios tales como el uso de codón de los genes que codifican la sulfotransferasa. Por tanto las levaduras que expresan sulfotransferasas pueden ser capaces de desintoxicar compuestos aromáticos como ácido p-cumárico y formar productos sulfatados como ácido zostérico.

Ejemplo 7 - Una gama de compuestos son sustratos para sulfatación *in vivo*

Aquí se muestra que la expresión de una aril sulfotransferasa puede ser capaz de convertir varios sustratos. Algunos de estos son inhibidores que se pueden encontrar en el hidrolizado de biomasa utilizado como sustrato para crecimiento celular y producción en biotecnología. Los compuestos también incluyen algunos que son de interés biotecnológico como productos de un cultivo celular o son algunos cuyo éster sulfato es de interés económico.

Se examinaron diferentes sulfotransferasas para determinar sus especificidades de sustrato contra tres sustratos. Se probaron las sulfotransferasas mencionadas en el ejemplo 1, así como otras adicionales. Los genes que las codifican se clonaron como se describe en el ejemplo 1 usando los cebadores que se muestran en la tabla a continuación de las bibliotecas de ADNc de los organismos respectivos, excepto para SULT1A1 de rata (*Rattus norvegicus*) optimizado por codón para *E. coli* (descrito anteriormente). Los vectores resultantes se transformaron en BL21(DE3)pLysS.

Cebadores:

Oligonucleótido	Gen	Dirección	Secuencia
CBJP517	SULT1C1 <i>Gallus gallus domesticus</i>	Directo	TAGAAATAATTTTGTTTAACTTTA AGAAGGAGATATACCatggcctgg ataaaatgg
CBJP518	SULT1C1 <i>Gallus gallus domesticus</i>	Inverso	TAAGCATTATGCGGCCGCAAGCT TGtcacaattccatgcgaaaaactag
CBJP533	SULT1A1 <i>Rattus norvegicus</i> (Codón optimizado para <i>E. coli</i>)	Directo	TAGAAATAATTTTGTTTAACTTTA AGAAGGAGATATACCatggaatttc acgtcc
CBJP534	SULT1A1 <i>Rattus norvegicus</i> (Codón optimizado para <i>E. coli</i>)	Inverso	TAAGCATTATGCGGCCGCAAGCT TGttacagttcacaacgaaattg

Las cepas resultantes se cultivaron en medio M9 que contenía pHCA 100 μM , resveratrol 95 μM o kaempferol 87 μM . Los cultivos se cultivaron durante la noche a 37 °C, 300 rpm. Al día siguiente, los sobrenadantes se aislaron y se examinaron por HPLC como se describe en el ejemplo 1. BL21(DE3)pLysS se usaron como cepa de control y no convirtieron los sustratos.

Enzima	pHCA	resveratrol	kaempferol
	100 μ M	95 μ M	87 μ M
SULT1A1 <i>Rattus norvegicus</i>	93 %	93 %	95 %
SULT1C1 <i>Gallus gallus domesticus</i>	26 %	100 %	80 %
SULT1A1 <i>Rattus norvegicus</i> (Codón optimizado para <i>E. coli</i>)	73 %	58 %	38 %
SULT1A1 humano	39 %	36 %	97 %
SULT1A1 <i>Equus caballus</i>	21 %	100 %	96 %
SULT1E1 <i>Gallus gallus domesticus</i>	17 %	100 %	47 %
SULT1A1 <i>Canis lupus familiaris</i>	34 %	61 %	60 %
SULT1A1 <i>Sus scrofa domesticus</i>	8 %	88 %	45 %

La tabla muestra el porcentaje de conversión de los diferentes sustratos por las células que expresan las diferentes sulfotransferasas. Los resultados muestran que varias sulfotransferasas, y especialmente la aril sulfotransferasa de rata (*Rattus norvegicus*), puede emplearse en la sulfatación de compuestos fenólicos.

Para probar aún más la gama de sustratos que se pueden sulfatar, se usaron cepas que transportan plásmidos que expresan SULT1A1 de rata (*Rattus norvegicus*) y SULT1E1 de pollo (*Gallus gallus domesticus*) (Ejemplo 1) clonados en el vector de expresión pETDuet-1, y cysDNCQ de *E. coli* clonado en el vector de expresión pRSFDuet-1 (Ejemplo 2). Los plásmidos se introdujeron en la cepa de expresión BL21(DE3)pLysS de *E. coli* como se describió anteriormente, seleccionando transformantes con antibióticos apropiados, es decir, 34 μ g ml⁻¹ de cloranfenicol para pLysS, 100 μ g ml⁻¹ de ampicilina para vectores basados en pETDuet-1 y 100 μ g ml⁻¹ de kanamicina para vectores basados en pRSFDuet-1. La siguiente tabla muestra la combinación de genes sobreexpresados en plásmidos. También se examinó una cepa de control sin un gen de sulfotransferasa o un operón cysDNCQ.

Cepas de <i>E. coli</i>	Sulfotransferasa	Genes Cys
Cepa de control	-	-
SULT1A1 rata	SULT1A1 rata	-
SULT1E1 pollo	SULT1E1 pollo	-
SULT1A1 rata + CysDNCQ	SULT1A1 rata	CysDNCQ

Las cepas fueron precultivadas en medio 2xYT con antibióticos apropiados. Se usaron 10 μ l de estos precultivos para inocular medios M9 con IPTG 1 mM y ninguno o un sustrato único para sulfatación. Después del crecimiento durante la noche a 37 °C, 300 rpm, los sobrenadantes se retiraron y se examinaron por HPLC como se describe en el Ejemplo 1. Los compuestos se detectaron por absorbancia UV. La siguiente tabla muestra el porcentaje de reducción de la concentración en las cepas que expresan sulfotransferasas solas o en combinación con genes cysDNCQ en comparación con la cepa de control.

Compuesto	Comenzar concentración en μ M	SULT1A1	SULT1E1	SULT1A1 + CysDNCQ
Ácido ferúlico	110	72 %	67 %	100 %
Quercetina	85	75 %	74 %	81 %
Ácido 4-hidroxibenzoico	287	5 %	4 %	6 %
4-acetamidofenol	114	24 %	10 %	30 %
Ácido 3-hidroxi-4-metoxicinámico	132	51 %	24 %	62 %
Ácido 4-hidroxifenilpirúvico	255	47 %	100 %	64 %
Ácido 3-(4-hidroxifenil)propiónico	241	3 %	1 %	7 %
Ácido vanílico	173	33 %	0 %	39 %
Luteolina	61	27 %	0 %	37 %
Apigenina	77	41 %	98 %	99 %
fisetina	81	98 %	98 %	100 %

En conclusión, Una amplia gama de compuestos fenólicos son sustratos para sulfotransferasas. En los ejemplos mostrados, la conversión se ve reforzada por la sobreexpresión de genes cysDNCQ. Algunos de estos compuestos y sus ésteres de sulfato son de interés en biotecnología. Además, algunos de estos compuestos son inhibidores del crecimiento y la función celular y, por tanto, la conversión por sulfatación es de interés para uso en sistemas biológicos.

LISTADO DE SECUENCIAS

<110> Danmarks Tekniske Universitet

<120> Procesos biológicos para la producción de sulfatos de arilo

5

<130> P81404775PCT00

<150> EP14182032

<151> 22-08-2014

10

<160> 75

<170> PatentIn versión 3.5

<210> 1

<211> 291

15

<212> PRT

<213> *Rattus norvegicus*

<400> 1

ES 2 770 616 T3

Met Glu Phe Ser Arg Pro Pro Leu Val His Val Lys Gly Ile Pro Leu
 1 5 10 15

Ile Lys Tyr Phe Ala Glu Thr Ile Gly Pro Leu Gln Asn Phe Thr Ala
 20 25 30

Trp Pro Asp Asp Leu Leu Ile Ser Thr Tyr Pro Lys Ser Gly Thr Thr
 35 40 45

Trp Met Ser Glu Ile Leu Asp Met Ile Tyr Gln Gly Gly Lys Leu Glu
 50 55 60

Lys Cys Gly Arg Ala Pro Ile Tyr Ala Arg Val Pro Phe Leu Glu Phe
 65 70 75 80

Lys Cys Pro Gly Val Pro Ser Gly Leu Glu Thr Leu Glu Glu Thr Pro
 85 90 95

Ala Pro Arg Leu Leu Lys Thr His Leu Pro Leu Ser Leu Leu Pro Gln
 100 105 110

Ser Leu Leu Asp Gln Lys Val Lys Val Ile Tyr Ile Ala Arg Asn Ala
 115 120 125

Lys Asp Val Val Val Ser Tyr Tyr Asn Phe Tyr Asn Met Ala Lys Leu
 130 135 140

His Pro Asp Pro Gly Thr Trp Asp Ser Phe Leu Glu Asn Phe Met Asp
 145 150 155 160

Gly Glu Val Ser Tyr Gly Ser Trp Tyr Gln His Val Lys Glu Trp Trp

ES 2 770 616 T3

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

<210> 2

<211> 295

5 <212> PRT

<213> Homo sapiens

<400> 2

ES 2 770 616 T3

Met Glu Leu Ile Gln Asp Thr Ser Arg Pro Pro Leu Glu Tyr Val Lys
1 5 10 15

Gly Val Pro Leu Ile Lys Tyr Phe Ala Glu Ala Leu Gly Pro Leu Gln
20 25 30

Ser Phe Gln Ala Arg Pro Asp Asp Leu Leu Ile Ser Thr Tyr Pro Lys
35 40 45

Ser Gly Thr Thr Trp Val Ser Gln Ile Leu Asp Met Ile Tyr Gln Gly
50 55 60

Gly Asp Leu Glu Lys Cys His Arg Ala Pro Ile Phe Met Arg Val Pro
65 70 75 80

Phe Leu Glu Phe Lys Ala Pro Gly Ile Pro Ser Gly Met Glu Thr Leu

ES 2 770 616 T3

85 90 95

Lys Asp Thr Pro Ala Pro Arg Leu Leu Lys Thr His Leu Pro Leu Ala
100 105 110

Leu Leu Pro Gln Thr Leu Leu Asp Gln Lys Val Lys Val Val Tyr Val
115 120 125

Ala Arg Asn Ala Lys Asp Val Ala Val Ser Tyr Tyr His Phe Tyr His
130 135 140

Met Ala Lys Val His Pro Glu Pro Gly Thr Trp Asp Ser Phe Leu Glu
145 150 155 160

Lys Phe Met Val Gly Glu Val Ser Tyr Gly Ser Trp Tyr Gln His Val
165 170 175

Gln Glu Trp Trp Glu Leu Ser Arg Thr His Pro Val Leu Tyr Leu Phe
180 185 190

Tyr Glu Asp Met Lys Glu Asn Pro Lys Arg Glu Ile Gln Lys Ile Leu
195 200 205

Glu Phe Val Gly Arg Ser Leu Pro Glu Glu Thr Val Asp Phe Met Val
210 215 220

Gln His Thr Ser Phe Lys Glu Met Lys Lys Asn Pro Met Thr Asn Tyr
225 230 235 240

Thr Thr Val Pro Gln Glu Phe Met Asp His Ser Ile Ser Pro Phe Met
245 250 255

Arg Lys Gly Met Ala Gly Asp Trp Lys Thr Thr Phe Thr Val Ala Gln
260 265 270

Asn Glu Arg Phe Asp Ala Asp Tyr Ala Glu Lys Met Ala Gly Cys Ser
275 280 285

Leu Ser Phe Arg Ser Glu Leu
290 295

<210> 3

<211> 295

5 <212> PRT

<213> Equus caballus

<400> 3

Met Glu Leu Ile Gln Asp Thr Ser Arg Pro Pro Leu Lys Tyr Val Lys

10

ES 2 770 616 T3

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

ES 2 770 616 T3

Arg Lys Gly Ile Ala Gly Asp Trp Lys Asn Thr Phe Thr Val Ala Gln
 260 265 270

Asn Glu His Phe Asp Thr Asp Tyr Ala Glu Lys Met Ala Gly Cys Lys
 275 280 285

Leu Ser Phe Arg Ser Glu Val
 290 295

<210> 4

<211> 295

5 <212> PRT

<213> Sus scrofa

<400> 4

Met Glu Pro Val Gln Asp Thr Tyr Arg Pro Pro Leu Glu Tyr Val Lys
 1 5 10 15

Gly Val Pro Leu Ile Lys Tyr Phe Ala Glu Ala Leu Gly Pro Leu Glu
 20 25 30

Ser Phe Gln Ala Trp Pro Asp Asp Val Leu Ile Ser Thr Tyr Pro Lys
 35 40 45

Ser Gly Thr Thr Trp Val Ser Glu Ile Leu Asp Leu Ile Tyr Gln Gly
 50 55 60

Gly Asp Leu Gln Lys Cys Gln Arg Ala Pro Ile Phe Val Arg Val Pro
 65 70 75 80

Phe Leu Glu Phe Lys Ile Pro Gly Cys Pro Thr Gly Phe Glu Leu Leu
 85 90 95

Lys Asp Thr Pro Ala Pro Arg Leu Leu Lys Thr His Leu Pro Leu Ala
 100 105 110

Leu Leu Pro Gln Thr Leu Leu Asp Gln Lys Val Lys Val Val Tyr Val
 115 120 125

Ala Arg Asn Ala Lys Asp Val Ala Val Ser Tyr Tyr His Phe Tyr Arg
 130 135 140

Met Ala Lys Val His Pro Asn Pro Gly Thr Trp Asp Ser Phe Leu Glu
 145 150 155 160

Asp Phe Met Ala Gly Glu Val Ser Tyr Gly Ser Trp Tyr Gln His Val
 165 170 175

10

ES 2 770 616 T3

Gln Glu Trp Trp Glu Leu Arg His Thr His Pro Val Leu Tyr Leu Phe
 180 185 190

Tyr Glu Asp Met Lys Glu Asn Pro Lys Arg Glu Ile Gln Lys Ile Leu
 195 200 205

Glu Phe Val Gly Arg Ser Leu Pro Glu Glu Thr Val Glu Asp Ile Val
 210 215 220

Gln His Thr Ser Phe Gln Glu Met Lys Asn Asn Ala Met Thr Asn Tyr
 225 230 235 240

Arg Thr Leu Pro Ser Asp Leu Leu Asp His Ser Ile Ser Ala Phe Met
 245 250 255

Arg Lys Gly Ile Thr Gly Asp Trp Lys Ser Thr Phe Thr Val Ala Gln
 260 265 270

Asn Glu Arg Phe Glu Ala Asp Tyr Ala Glu Lys Met Ala Gly Cys Asn
 275 280 285

Leu Arg Phe Arg Ser Glu Leu
 290 295

<210> 5

<211> 295

5 <212> PRT

<213> Canis lupus

<400> 5

Met Glu Asp Ile Pro Asp Thr Ser Arg Pro Pro Leu Lys Tyr Val Lys
 1 5 10 15

Gly Ile Pro Leu Ile Lys Tyr Phe Ala Glu Ala Leu Glu Ser Leu Gln
 20 25 30

Asp Phe Gln Ala Gln Pro Asp Asp Leu Leu Ile Ser Thr Tyr Pro Lys
 35 40 45

Ser Gly Thr Thr Trp Val Ser Glu Ile Leu Asp Met Ile Tyr Gln Asp
 50 55 60

Gly Asp Val Glu Lys Cys Arg Arg Ala Pro Val Phe Ile Arg Val Pro
 65 70 75 80

Phe Leu Glu Phe Lys Ala Pro Gly Ile Pro Thr Gly Leu Glu Val Leu
 85 90 95

10

ES 2 770 616 T3

Lys Asp Thr Pro Ala Pro Arg Leu Ile Lys Thr His Leu Pro Leu Ala
 100 105 110

Leu Leu Pro Gln Thr Leu Leu Asp Gln Lys Val Lys Val Val Tyr Val
 115 120 125

Ala Arg Asn Ala Lys Asp Val Ala Val Ser Tyr Tyr His Phe Tyr Arg
 130 135 140

Met Ala Lys Val His Pro Asp Pro Asp Thr Trp Asp Ser Phe Leu Glu
 145 150 155 160

Lys Phe Met Ala Gly Glu Val Ser Tyr Gly Ser Trp Tyr Gln His Val
 165 170 175

Gln Glu Trp Trp Glu Leu Ser His Thr His Pro Val Leu Tyr Leu Phe
 180 185 190

Tyr Glu Asp Met Lys Glu Asn Pro Lys Arg Glu Ile Gln Lys Ile Leu
 195 200 205

Lys Phe Val Gly Arg Ser Leu Pro Glu Glu Thr Val Asp Leu Ile Val
 210 215 220

Gln His Thr Ser Phe Lys Glu Met Lys Asn Asn Ser Met Ala Asn Tyr
 225 230 235 240

Thr Thr Leu Ser Pro Asp Ile Met Asp His Ser Ile Ser Ala Phe Met
 245 250 255

Arg Lys Gly Ile Ser Gly Asp Trp Lys Thr Thr Phe Thr Val Ala Gln
 260 265 270

Asn Glu Arg Phe Asp Ala Asp Tyr Ala Lys Lys Met Glu Gly Cys Gly
 275 280 285

Leu Ser Phe Arg Thr Gln Leu
 290 295

<210> 6

<211> 294

5 <212> PRT

<213> Gallus gallus

<400> 6

Met Gly Asn Asp Glu Val Ile Arg Gln Asp Leu Gly Cys Leu Tyr Asp
 1 5 10 15

10

ES 2 770 616 T3

Ile Pro Leu Tyr Lys Cys Phe Val Ala Gly Trp Pro Gln Val Glu Ala
 20 25 30

Phe Gln Ala Arg Pro Asp Asp Leu Leu Ile Ala Thr Tyr Pro Lys Ser
 35 40 45

Gly Thr Thr Trp Leu Ser Glu Ile Leu Asp Ala Ile Tyr His Asp Gly
 50 55 60

Asp Leu Glu Lys Cys Arg Arg Asp Ala Ile Tyr Asn Arg Val Pro Phe
 65 70 75 80

Leu Glu Met Lys Ala Pro Gly Ile Leu Ser Gly Val Glu Gln Leu Glu
 85 90 95

Lys Ile Pro Ser Pro Arg Leu Val Lys Thr His Leu Pro Val His Leu
 100 105 110

Leu Pro Ala Ser Phe Gln Glu Lys Asp Cys Lys Val Ile Tyr Met Ala
 115 120 125

Arg Asn Ala Lys Asp Val Val Ile Ser Tyr Tyr Tyr Phe Tyr Gln Met
 130 135 140

Ala Lys Ile His Pro Asp Pro Gly Thr Leu Ser Glu Phe Leu Gln Ala
 145 150 155 160

Phe Met Asp Gly Lys Val Ala Tyr Gly Ser Trp Tyr Lys His Val Lys
 165 170 175

Gly Trp Trp Glu Lys Arg His Glu Lys Arg Leu Leu Tyr Leu Phe Tyr
 180 185 190

Glu Asp Met Lys Lys Asp Pro Arg Arg Glu Ile Gln Lys Ile Leu Gln
 195 200 205

Phe Leu Gly Lys Glu Val Ala Glu Glu Thr Val Ala Arg Ile Leu His
 210 215 220

His Thr Ser Phe Gln Glu Met Lys Lys Asn Pro Ala Thr Asn Tyr Glu
 225 230 235 240

Thr Met Pro Thr Glu Leu Met Asp His Ser Leu Ser Pro Phe Met Arg
 245 250 255

Lys Gly Ile Ser Gly Asp Trp Ala Asn His Phe Thr Val Ala Gln Asn
 260 265 270

ES 2 770 616 T3

Glu Arg Phe Asp Gln His Tyr Gln Gln Gln Met Ala Gly Ser Asp Leu
 275 280 285

Cys Phe Gln Met Glu Ala
 290

<210> 7

<211> 307

5 <212> PRT
 <213> Gallus gallus

<400> 7

Met Ala Leu Asp Lys Met Glu Asn Leu Ser Leu Glu Glu Asn Met Leu
 1 5 10 15

Arg Ser Glu Met Gly Glu Val Gln Gly Ile Pro Val Thr Lys Pro Thr
 20 25 30

Cys Asp Ile Trp Asp Gln Val Trp Asn Phe Lys Ala Arg Pro Asp Asp
 35 40 45

Leu Leu Val Ala Thr Tyr Ala Lys Ala Gly Thr Thr Trp Thr Gln Glu
 50 55 60

Ile Val Asp Met Ile Gln Gln Asn Gly Asp Ile Glu Lys Cys Arg Arg
 65 70 75 80

Ala Ser Thr Tyr Lys Arg His Pro Phe Leu Glu Trp Tyr Ile Pro Asp
 85 90 95

Ser Ser Pro Leu Gly Tyr Ser Gly Leu Lys Leu Ala Glu Ala Met Pro
 100 105 110

Ser Pro Arg Thr Met Lys Thr His Leu Pro Val Gln Leu Val Pro Pro
 115 120 125

Ser Phe Trp Glu Gln Asn Cys Lys Ile Ile Tyr Val Ala Arg Asn Ala
 130 135 140

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

Leu Pro Asp Pro Gly Thr Ile Glu Glu Phe Thr Glu Lys Phe Met Asn
 165 170 175

Gly Glu Val Leu Trp Gly Ser Trp Tyr Asp His Val Lys Gly Trp Trp
 180 185 190

10

ES 2 770 616 T3

Lys Ala Lys Asp Lys His Arg Ile Leu Tyr Leu Phe Tyr Glu Asp Met
 195 200 205

Lys Glu Asn Pro Lys Arg Glu Ile Gln Lys Ile Met Lys Phe Leu Glu
 210 215 220

Lys Asp Leu Asp Glu Glu Val Leu Asn Lys Ile Ile Tyr Asn Thr Ser
 225 230 235 240

Phe Glu Ile Met Lys Asp Asn Pro Met Thr Asn Tyr Thr Lys Asp Phe
 245 250 255

Val Gly Val Met Asp His Ser Val Ser Pro Phe Met Arg Lys Gly Ser
 260 265 270

Val Gly Asp Trp Lys Asn Tyr Phe Thr Val Ala Leu Asn Lys Lys Phe
 275 280 285

Asp Gln Asp Tyr Lys Lys Lys Met Ala Asp Thr Ser Leu Val Phe Arg
 290 295 300

Met Glu Leu
 305

<210> 8

<211> 301

5 <212> PRT

<213> Danio rerio

<400> 8

Met Asp Leu Pro Asp Ile Ser Ser Ile Lys Leu Pro Ser Arg Pro Lys
 1 5 10 15

Ile Phe Glu Phe Glu Gly Ile Ser Met Ile Ser Tyr Phe Thr Asp Asn
 20 25 30

Trp Glu Lys Leu Lys Asn Phe Gln Ala Arg Pro Asp Asp Ile Leu Ile
 35 40 45

Ala Thr Tyr Pro Lys Ala Gly Thr Thr Trp Val Ser Tyr Ile Leu Asp
 50 55 60

Leu Leu Tyr Phe Gly Lys Val Glu Pro Asn Gly Gln Ser Ser Leu Pro
 65 70 75 80

Ile Tyr Met Arg Val Pro Phe Leu Glu Ser Cys Phe Pro Gly Met Pro
 85 90 95

10

ES 2 770 616 T3

Ser Gly Thr Glu Leu Ala Asp Asn Leu Pro Asn Ser Pro Arg Leu Ile
 100 105 110

Lys Thr His Leu Pro Val Gln Leu Val Pro Lys Ser Phe Trp Gly Gln
 115 120 125

Asn Ser Lys Val Val Tyr Val Ala Arg Asn Ala Lys Asp Asn Val Val
 130 135 140

Ser Phe Phe His Phe Asp Arg Met Asn His Gly Gln Pro Glu Pro Gly
 145 150 155 160

Asp Trp Asp Thr Phe Leu Gln Ala Phe Ile Lys Gly Glu Arg Val Phe
 165 170 175

Gly Ser Trp Phe Asp His Val Cys Gly Trp Trp Glu Lys Lys Lys Thr
 180 185 190

Tyr Pro Asn Leu His Tyr Met Phe Tyr Glu Asp Ile Ala Lys Asp Ile
 195 200 205

Asn Gly Glu Val Glu Ser Leu Cys Thr Phe Leu Lys Leu Ser Arg Ser
 210 215 220

Asp Glu Glu Lys Glu Lys Ile Ile Asn Gly Val Gln Phe Asp Ala Met
 225 230 235 240

Lys Gln Asn Val Met Thr Asn Tyr Ser Thr Ile Pro Thr Met Asp Phe
 245 250 255

Thr Ile Ser Pro Phe Met Arg Lys Gly Lys Val Gly Asp Trp Lys Asn
 260 265 270

His Phe Thr Val Ala Gln Asn Glu Gln Phe Asp Glu Asp Tyr Lys Glu
 275 280 285

Lys Met Lys Asn Thr Thr Leu Asn Phe Arg Thr Lys Ile
 290 295 300

<210> 9

<211> 300

5 <212> PRT
 <213> Danio rerio

<400> 9

Met Glu Ile Gln Gly Lys Ser Ser Thr Asp Leu Pro Asp Arg Pro Glu
 1 5 10 15

10

ES 2 770 616 T3

Ile Phe Glu Phe Glu Gly Ile Ser Met Val Glu His Phe Thr Lys Asn
 20 25 30

Trp Glu Asn Val Lys Asn Phe Gln Ala Arg Pro Asp Asp Ile Leu Ile
 35 40 45

Ala Thr Tyr Pro Lys Ala Gly Thr Thr Trp Val Ser Asn Ile Leu Asp
 50 55 60

Leu Leu Tyr Phe Gly Lys Glu Asp Pro Lys Arg Gln Thr Thr Lys Pro
 65 70 75 80

Ile Tyr Lys Arg Val Pro Phe Leu Glu Ser Cys Phe Pro Glu Met Gln
 85 90 95

Ser Gly Thr Glu Leu Ala Asn Asn Leu Pro Thr Ser Pro Arg Leu Ile
 100 105 110

Lys Thr His Leu Pro Val Gln Leu Val Pro Gln Ser Phe Trp Glu Lys
 115 120 125

Asn Ser Arg Val Ala Tyr Val Ala Arg Asn Ala Lys Asp Asn Ala Val
 130 135 140

Ser Tyr Phe His Phe Asn Arg Met Asn Lys Ala Gln Pro Glu Pro Gly
 145 150 155 160

Asp Trp Asn Thr Phe Leu Glu Glu Phe Met Lys Gly Lys Met Val Phe
 165 170 175

Gly Ser Trp Phe Asp His Val Cys Gly Trp Trp Glu Lys Lys Lys Thr
 180 185 190

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

Lys Gly Glu Val Glu Ser Leu Cys Thr Phe Leu Lys Leu Ser Arg Ser
 210 215 220

Asp Glu Glu Lys Glu Lys Ile Ile Asn Gly Ile Gln Phe Asp Ala Met
 225 230 235 240

Lys Gln Asn Lys Met Thr Asn Tyr Ser Thr Val Leu Val Met Asp Phe
 245 250 255

Thr Ile Ser Pro Phe Met Arg Lys Gly Lys Val Gly Asp Trp Lys Asn
 260 265 270

ES 2 770 616 T3

His Phe Thr Val Ala Gln Asn Glu Gln Phe Asn Glu Asp Tyr Lys Gln
 275 280 285

Lys Met Lys Asn Ser Thr Leu Lys Phe Pro Thr Glu
 290 295 300

<210> 10

<211> 338

5 <212> PRT

<213> Drosophila melanogaster

<400> 10

Met Pro Gln Ser Ser Phe Phe Ala Lys Ser Val Pro Phe Glu Gln Ile
 1 5 10 15

Asp Lys Leu Ala Ile Ser Gly Gly Tyr Ser Ser Ile Phe Ala Ser Ser
 20 25 30

Lys Pro Ser Val Pro Val Val Gly Asn Trp Glu Gln Arg Phe Cys Arg
 35 40 45

Leu Ala Asp Thr Phe Gln Pro Val Leu Asp Arg Val Tyr Asp Phe Glu
 50 55 60

Val Arg Asp Asp Asp Val Trp Ile Val Thr Leu Pro Lys Cys Gly Thr
 65 70 75 80

Thr Trp Met Gln Glu Leu Ala Trp Leu Val Ile Asn Glu Cys Asp Phe
 85 90 95

Glu Thr Ala Lys Ser Val Asp Leu Thr His Arg Ser Pro Phe Leu Glu
 100 105 110

Phe Asn Gly Val Val Pro Asn Val Pro His Asp Thr Ile Ala Ala Ala
 115 120 125

Asn Ala Leu Pro Ser Pro Arg Leu Ile Lys Ser His Leu Pro Ala Trp
 130 135 140

Met Leu Pro Arg Gln Ile Trp Ser Lys Arg Pro Lys Ile Ile Tyr Val
 145 150 155 160

Tyr Arg Asn Pro Lys Asp Ala Ala Ile Ser Tyr Phe His His Trp Arg
 165 170 175

Gly Met Val Gly Tyr Gln Gly Thr Lys Ser Asp Phe Met His Ser Phe
 180 185 190

10

ES 2 770 616 T3

Ile Asp Gly Tyr Val Asn Phe Thr Pro Cys Trp Pro His Ile Leu Asp
 195 200 205

Phe Trp Gln Leu Arg His Glu Pro Asn Ile Phe Phe Thr Ser Tyr Glu
 210 215 220

Arg Met Lys Gly Gln Leu Gly Gln Val Ile Ser Glu Val Ala Gln Phe
 225 230 235 240

Leu Glu Arg Ser Val Ser Gln Glu Gln Met Gln Gln Met Gln Arg His
 245 250 255

Leu Ser Phe Glu Ser Met Arg Asp Asn Pro Ala Cys Asn His Val Lys
 260 265 270

Glu Phe Glu Ser Met Lys Ala Ala Ala Gly Arg Glu Val Glu Glu Phe
 275 280 285

Arg Phe Val Arg Arg Gly Val Val Gly Ser His Lys Asp Glu Leu Thr
 290 295 300

Ala Asp Ile Ile Arg Glu Phe Asp Leu Trp Ser Asp Ser Asn Leu Arg
 305 310 315 320

Asp Phe Lys Leu Asn Met Asp Asp Phe Ala Asn Tyr Ser Lys Phe Ala
 325 330 335

Ser Thr

<210> 11

<211> 313

5 <212> PRT

<213> Drosophila melanogaster

<400> 11

Met Asn Arg Val Gln Val Thr Pro Arg Ser Tyr Pro Thr Asn Leu Ile
 1 5 10 15

Asp Lys Asp Trp Gly Asn Arg Lys Leu Phe Tyr Thr Lys Asp Ser Glu
 20 25 30

Asn Phe Leu Arg Leu Val His Asp Met Lys Leu Arg Asp Asp Asp Val
 35 40 45

Trp Ile Val Thr Leu Pro Lys Cys Gly Thr Thr Trp Met Gln Glu Leu
 50 55 60

10

ES 2 770 616 T3

Leu Trp Leu Leu Leu Asn Asn Cys Asp Phe Glu Gly Ala Leu Ala Lys
65 70 75 80

Asp Gln Glu Leu Arg Thr Pro Phe Leu Glu Phe Gly Tyr Ser Val Phe
85 90 95

His Asp Pro Asn Arg Ser Phe Gly Pro Ile Glu Asp Leu Lys Ser Pro
100 105 110

Arg Leu Ile Lys Ser His Leu Ser Leu Ala Leu Leu Pro Ser Lys Leu
115 120 125

Trp Glu Gly Lys Asn Lys Val Ile Tyr Val Ser Arg Asn Pro Leu Asp
130 135 140

Ser Tyr Val Ser Arg Tyr Tyr His Gly Val Ser Phe Gly Phe Asn Tyr
145 150 155 160

Gly Lys Ser Leu His Gln Tyr Phe Asp Glu Val Leu Ala Ser Asp Asp
165 170 175

Phe Pro Thr Glu Phe Ile Glu His Ala His Glu Phe Tyr Gln Leu Arg
180 185 190

Asn Glu Pro Trp Val Phe Tyr Thr Ser Phe Glu Met Met Lys Lys Asp
195 200 205

Leu Arg Gly Val Ile Asn Asp Val Ser Arg Phe Leu Asn Lys Pro Ile
210 215 220

Asn Asp Gln Gln Met Glu Lys Leu Leu Lys His Leu Ser Phe Ala Glu
225 230 235 240

Met Lys Lys Asn Pro Thr Thr Asn His Leu Trp Glu Leu Ala Gln Val
245 250 255

Gln His Glu Asn Ala Gly Lys Glu Met His Pro Phe Val Arg Arg Gly
260 265 270

Asp Val Asn Gly Tyr Lys Asp Glu Leu Lys Pro Glu Gln Ile Glu Lys
275 280 285

Ala Asn Val Arg Ile Gln Glu Val Leu Ala Lys Asn Gly Val Thr Leu
290 295 300

Asp Glu Leu Leu Leu Leu Lys Asp Gln

305

310

ES 2 770 616 T3

<211> 346

<212> PRT

<213> Drosophila melanogaster

5 <400> 12

```

Met Glu Asn Thr Pro Leu Lys Phe Pro His Glu Ile Arg Asp Val Glu
 1                               5                               10                               15

Glu Ser Thr Asn Ala Glu Leu Leu Asp His Phe His Gly Glu Arg Thr
                20                               25                               30

Gly Phe Val Gln Val Gly Ser Glu Gly Tyr Phe Phe Pro His Lys Tyr
                35                               40                               45

Lys Asp Glu Ala Glu Arg Tyr Tyr Asn Phe Glu Ala Arg Pro Asp Asp
 50                               55                               60

Val Trp Ile Ala Thr Val Pro Arg Ser Gly Thr Thr Trp Thr Gln Glu
65                               70                               75                               80

Leu Ile Trp Leu Val Ala Asn Gly Leu Asp Phe Glu His Ala Gln Glu
                85                               90                               95

Arg Pro Leu Thr Glu Arg Phe Pro Phe Phe Glu Phe Pro Leu Phe Val
                100                               105                               110

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

Glu Ala Leu Glu Phe Ile Glu Lys Ile Ala Arg Pro Gly Tyr Glu Ala
                130                               135                               140

Leu Ser Glu Ile Pro Arg Ser Gln Arg Arg Phe Ile Lys Thr His Phe
145                               150                               155                               160

Pro Phe Ser Leu Met Pro Pro Ser Val Leu Glu Lys Lys Cys Lys Val
                165                               170                               175

Ile Tyr Val Val Arg Asp Pro Lys Asp Val Ala Val Ser Tyr Tyr His
                180                               185                               190

Leu Asn Arg Leu Phe Arg Thr Gln Gly Tyr Val Gly Asp Phe Glu Arg
                195                               200                               205

Tyr Trp His Tyr Phe Gln Asn Gly Leu Asn Pro Trp Leu Pro Tyr Tyr

```

ES 2 770 616 T3

210 215 220

Ser His Val Lys Glu Ala Arg Glu His Ala His Leu Ser Asn Val Leu
 225 230 235 240

Phe Leu Arg Tyr Glu Asp Met Leu Ala Asp Leu Pro Gly Ala Ile Asn
 245 250 255

Ser Ile Ala Ser Phe Leu Glu Cys Pro Pro Lys Pro Glu Asp Met Asp
 260 265 270

Arg Leu Leu Asp His Leu Ser Ile Arg Ser Phe Arg Glu Asn Lys Ser
 275 280 285

Val Asn Met His Glu Met Ala Ser Val Gly Val Leu Asn Lys Gly Glu
 290 295 300

Ala Gly Phe Val Arg Ser Gly Ala Lys Thr Ala Tyr Gln Pro Gln Gln
 305 310 315 320

Glu Phe Val Glu Asn Pro Lys Leu Leu Lys Ser Ala Asn Glu Trp Val
 325 330 335

Glu Gln Asn Ile Lys Ser Phe Lys Thr Ile
 340 345

<210> 13

<211> 323

5 <212> PRT

<213> Arabidopsis thaliana

<400> 13

Met Glu Met Asn Leu Arg Ile Glu Asp Leu Asn Glu Glu Thr Lys Thr
 1 5 10 15

Leu Ile Ser Ser Leu Pro Ser Asp Lys Asp Phe Thr Gly Lys Thr Ile
 20 25 30

Cys Lys Tyr Gln Gly Cys Trp Tyr Thr His Asn Val Leu Gln Ala Val
 35 40 45

Leu Asn Phe Gln Lys Ser Phe Lys Pro Gln Asp Thr Asp Ile Ile Val
 50 55 60

Ala Ser Phe Pro Lys Cys Gly Thr Thr Trp Leu Lys Ala Leu Thr Phe
 65 70 75 80

Ala Leu Leu His Arg Ser Lys Gln Pro Ser His Asp Asp Asp His Pro

10

ES 2 770 616 T3

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

ES 2 770 616 T3

<211> 506

<212> PRT

<213> *Flavobacterium johnsoniae*

5 <400> 14

```

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

Ile Phe Gly Asn Gln Lys Val Thr Ile Ser Asp Val Val Val Asn Arg
 20           25           30

Val Asn Glu Ser Phe Asn Phe Leu Lys Glu Phe Ser Gly Asn Lys Val
 35           40           45

Ile Tyr Gly Val Asn Thr Gly Phe Gly Pro Met Ala Gln Tyr Arg Ile
 50           55           60

Lys Glu Ser Asp Gln Ile Gln Leu Gln Tyr Asn Leu Ile Arg Ser His
 65           70           75           80

Ser Ser Gly Thr Gly Lys Pro Leu Ser Pro Val Cys Ala Lys Ala Ala
 85           90           95

Ile Leu Ala Arg Leu Asn Thr Leu Ser Leu Gly Asn Ser Gly Val His
 100          105          110

Pro Ser Val Ile Asn Leu Met Ser Glu Leu Ile Asn Lys Asp Ile Thr
 115          120          125

Pro Leu Ile Phe Glu His Gly Gly Val Gly Ala Ser Gly Asp Leu Val
 130          135          140

Gln Leu Ser His Leu Ala Leu Val Leu Ile Gly Glu Gly Glu Val Phe
 145          150          155          160

Tyr Lys Gly Glu Arg Arg Pro Thr Pro Glu Val Phe Glu Ile Glu Gly
 165          170          175

Leu Lys Pro Ile Gln Val Glu Ile Arg Glu Gly Leu Ala Leu Ile Asn
 180          185          190

Gly Thr Ser Val Met Thr Gly Ile Gly Val Val Asn Val Tyr His Ala
 195          200          205

Lys Lys Leu Leu Asp Trp Ser Leu Lys Ser Ser Cys Ala Ile Asn Glu
 210          215          220
    
```

ES 2 770 616 T3

Leu Val Gln Ala Tyr Asp Asp His Phe Ser Ala Glu Leu Asn Gln Thr
 225 230 235 240

Lys Arg His Lys Gly Gln Gln Glu Ile Ala Leu Lys Met Arg Gln Asn
 245 250 255

Leu Ser Asp Ser Thr Leu Ile Arg Lys Arg Glu Asp His Leu Tyr Ser
 260 265 270

Gly Glu Asn Thr Glu Glu Ile Phe Lys Glu Lys Val Gln Glu Tyr Tyr
 275 280 285

Ser Leu Arg Cys Val Pro Gln Ile Leu Gly Pro Val Leu Glu Thr Ile
 290 295 300

Asn Asn Val Ala Ser Ile Leu Glu Asp Glu Phe Asn Ser Ala Asn Asp
 305 310 315 320

Asn Pro Ile Ile Asp Val Lys Asn Gln His Val Tyr His Gly Gly Asn
 325 330 335

Phe His Gly Asp Tyr Ile Ser Leu Glu Met Asp Lys Leu Lys Ile Val
 340 345 350

Ile Thr Lys Leu Thr Met Leu Ala Glu Arg Gln Leu Asn Tyr Leu Leu
 355 360 365

Asn Ser Lys Ile Asn Glu Leu Leu Pro Pro Phe Val Asn Leu Gly Thr
 370 375 380

Leu Gly Phe Asn Phe Gly Met Gln Gly Val Gln Phe Thr Ala Thr Ser
 385 390 395 400

Thr Thr Ala Glu Ser Gln Met Leu Ser Asn Pro Met Tyr Val His Ser
 405 410 415

Ile Pro Asn Asn Asn Asp Asn Gln Asp Ile Val Ser Met Gly Thr Asn
 420 425 430

Ser Ala Val Ile Thr Ser Lys Val Ile Glu Asn Ala Phe Glu Val Leu
 435 440 445

Ala Ile Glu Met Ile Thr Ile Val Gln Ala Ile Asp Tyr Leu Gly Gln
 450 455 460

Lys Asp Lys Ile Ser Ser Val Ser Lys Lys Trp Tyr Asp Glu Ile Arg
 465 470 475 480

ES 2 770 616 T3

Asn Ile Ile Pro Thr Phe Lys Glu Asp Gln Val Met Tyr Pro Phe Val
 485 490 495

Gln Lys Val Lys Asp His Leu Ile Asn Asn
 500 505

<210> 15

<211> 552

5 <212> PRT

<213> Herpetosiphon aurantiacus

<400> 15

Met Ser Thr Thr Leu Ile Leu Thr Gly Glu Gly Leu Gly Ile Asp Asp
 1 5 10 15

Val Val Arg Val Ala Arg His Gln Asp Arg Val Glu Leu Thr Thr Asp
 20 25 30

Pro Ala Ile Leu Ala Gln Ile Glu Ala Ser Cys Ala Tyr Ile Asn Gln
 35 40 45

Ala Val Lys Glu His Gln Pro Val Tyr Gly Val Thr Thr Gly Phe Gly
 50 55 60

Gly Met Ala Asn Val Ile Ile Ser Pro Glu Glu Ala Ala Glu Leu Gln
 65 70 75 80

Asn Asn Ala Ile Trp Tyr His Lys Thr Gly Ala Gly Lys Leu Leu Pro
 85 90 95

Phe Thr Asp Val Arg Ala Ala Met Leu Leu Arg Ala Asn Ser His Met
 100 105 110

Arg Gly Ala Ser Gly Ile Arg Leu Glu Ile Ile Gln Arg Met Val Thr
 115 120 125

Phe Leu Asn Ala Asn Val Thr Pro His Val Arg Glu Phe Gly Ser Ile
 130 135 140

Gly Ala Ser Gly Asp Leu Val Pro Leu Ile Ser Ile Thr Gly Ala Leu
 145 150 155 160

Leu Gly Thr Asp Gln Ala Phe Met Val Asp Phe Asn Gly Glu Thr Leu
 165 170 175

Asp Cys Ile Ser Ala Leu Glu Arg Leu Gly Leu Pro Arg Leu Arg Leu
 180 185 190

ES 2 770 616 T3

Gln Pro Lys Glu Gly Leu Ala Met Met Asn Gly Thr Ser Val Met Thr
 195 200 205

Gly Ile Ala Ala Asn Cys Val His Asp Ala Arg Ile Leu Leu Ala Leu
 210 215 220

Ala Leu Glu Ala His Ala Leu Met Ile Gln Gly Leu Gln Gly Thr Asn
 225 230 240

Gln Ser Phe His Pro Phe Ile His Arg His Lys Pro His Thr Gly Gln
 245 250 255

Val Trp Ala Ala Asp His Met Leu Glu Leu Leu Gln Gly Ser Gln Leu
 260 265 270

Ser Arg Asn Glu Leu Asp Gly Ser His Asp Tyr Arg Asp Gly Asp Leu
 275 280 285

Ile Gln Asp Arg Tyr Ser Leu Arg Cys Leu Pro Gln Phe Leu Gly Pro
 290 295 300

Ile Ile Asp Gly Met Ala Phe Ile Ser His His Leu Arg Val Glu Ile
 305 310 315 320

Asn Ser Ala Asn Asp Asn Pro Leu Ile Asp Thr Ala Ser Ala Ala Ser
 325 330 335

Tyr His Gly Gly Asn Phe Leu Gly Gln Tyr Ile Gly Val Gly Met Asp
 340 345 350

Gln Leu Arg Tyr Tyr Met Gly Leu Met Ala Lys His Leu Asp Val Gln
 355 360 365

Ile Ala Leu Leu Val Ser Pro Gln Phe Asn Asn Gly Leu Pro Ala Ser
 370 375 380

Leu Val Gly Asn Ile Gln Arg Lys Val Asn Met Gly Leu Lys Gly Leu
 385 390 395 400

Gln Leu Thr Ala Asn Ser Ile Met Pro Ile Leu Thr Phe Leu Gly Asn
 405 410 415

Ser Leu Ala Asp Arg Phe Pro Thr His Ala Glu Gln Phe Asn Gln Asn
 420 425 430

Ile Asn Ser Gln Gly Phe Gly Ser Ala Asn Leu Ala Arg Gln Thr Ile
 435 440 445

ES 2 770 616 T3

Gln Thr Leu Gln Gln Tyr Ile Ala Ile Thr Leu Met Phe Gly Val Gln
 450 455 460

Ala Val Asp Leu Arg Thr His Lys Leu Ala Gly His Tyr Asn Ala Ala
 465 470 475 480

Glu Leu Leu Ser Pro Leu Thr Ala Lys Ile Tyr His Ala Val Arg Ser
 485 490 495

Ile Val Lys His Pro Pro Ser Pro Glu Arg Pro Tyr Ile Trp Asn Asp
 500 505 510

Asp Glu Gln Val Leu Glu Ala His Ile Ser Ala Leu Ala His Asp Ile
 515 520 525

Ala Asn Asp Gly Ser Leu Val Ser Ala Val Glu Gln Thr Leu Ser Gly
 530 535 540

Leu Arg Ser Ile Ile Leu Phe Arg
 545 550

<210> 16

<211> 552

5 <212> PRT

<213> Herpetosiphon aurantiacus

<400> 16

Met Arg His Gln Val Thr Leu Thr Gly Ala Gly Leu Thr Ile Glu Asp
 1 5 10 15

Val Val Arg Val Ala Arg His His Gln Pro Val Gly Leu Thr Asp Asn
 20 25 30

Pro Glu Ile Leu Gln Arg Ile Glu Asp Ser Cys Ala Tyr Ile Asn Asp
 35 40 45

Ala Val Lys Ala Ser Lys Pro Val Tyr Gly Val Thr Thr Gly Phe Gly
 50 55 60

Gly Met Ala Asp Val Val Ile Ser Ser Glu Glu Ala Ala Asp Leu Gln
 65 70 75 80

Asn Asn Ala Ile Trp Tyr His Lys Thr Gly Ala Gly Lys Leu Leu Pro
 85 90 95

Leu Ala Asp Val Arg Ala Ala Met Leu Leu Arg Ala Asn Ser His Met
 100 105 110

10

ES 2 770 616 T3

Arg Gly Val Ser Gly Ile Arg Leu Glu Ile Ile Gln Arg Met Met Thr
 115 120 125

Phe Leu Asn Ala Asn Val Thr Pro His Val Arg Glu Phe Gly Ser Ile
 130 135 140

Gly Ala Ser Gly Asp Leu Val Pro Leu Ile Ser Ile Thr Gly Ala Leu
 145 150 155 160

Leu Gly Thr Asp Pro Ala Phe Arg Val Asp Phe Asp Gly Glu Asn Ile
 165 170 175

Asp Cys Leu Glu Ala Leu Glu Arg Leu Asn Leu Pro Arg Leu Glu Leu
 180 185 190

Leu Pro Lys Glu Gly Leu Ala Met Met Asn Gly Thr Ser Val Met Thr
 195 200 205

Gly Ile Ala Ser Asn Val Leu His Asp Ala Arg Ile Leu Leu Gly Leu
 210 215 220

Ala Leu Asn Ile His Gly Leu Met Ile Gln Gly Leu Gln Gly Thr Asn
 225 230 235 240

Gln Ser Phe His Pro Phe Ile His Gln His Lys Ala His Thr Gly Gln
 245 250 255

Val Trp Ala Ala Asp His Met Leu Gln Ile Leu Glu Gly Ser Ala Leu
 260 265 270

Ser Arg Asp Glu Leu Asp Gly Arg His Glu Tyr Arg Glu Gly Asp Leu
 275 280 285

Ile Gln Asp Arg Tyr Ser Leu Arg Cys Leu Pro Gln Phe Leu Gly Pro
 290 295 300

Ile Ile Asp Gly Met Ala Tyr Ile Thr His His Leu Arg Val Glu Ile
 305 310 315 320

Asn Ser Ala Asn Asp Asn Pro Leu Ile Asn Thr Glu Ala Gly Ala Ser
 325 330 335

Tyr His Gly Gly Asn Phe Leu Gly Gln Tyr Ile Gly Val Gly Met Asp
 340 345 350

Gln Leu Arg Tyr Tyr Met Gly Leu Met Ala Lys His Leu Asp Val Gln

ES 2 770 616 T3

355	360	365																				
Ile	Ala	Leu	Leu	Val	Ser	Pro	Gln	Phe	Asn	Asn	Gly	Leu	Ser	Ala	Ser							
370						375					380											
Leu	Val	Gly	Asn	Thr	Asp	Arg	Lys	Val	Asn	Met	Gly	Leu	Lys	Gly	Leu							
385					390					395					400							
Gln	Ile	Ser	Gly	Asn	Ser	Ile	Met	Pro	Ile	Leu	Gly	Phe	Leu	Gly	Asn							
				405					410					415								
Ser	Leu	Ala	Asp	Arg	Phe	Pro	Thr	His	Ala	Glu	Gln	Phe	Asn	Gln	Asn							
			420					425					430									
Ile	Asn	Ser	Gln	Gly	Phe	Gly	Ser	Ala	Asn	Leu	Ala	Arg	Gln	Thr	Ile							
		435					440					445										
Glu	Thr	Leu	Gln	Gln	Tyr	Ile	Ala	Ile	Ala	Leu	Ile	Phe	Gly	Val	Gln							
	450					455					460											
Ala	Val	Asp	Leu	Arg	Thr	Phe	Lys	Arg	Thr	Gly	His	Tyr	Asn	Ala	Val							
465					470					475					480							
Glu	Thr	Leu	Ser	Pro	Met	Thr	Ala	Lys	Leu	Tyr	Ser	Ala	Met	Arg	Glu							
				485					490					495								
Val	Val	Gly	Lys	Pro	Ile	Ser	His	Glu	Arg	Pro	Tyr	Ile	Trp	Asn	Asp							
			500					505					510									
Asn	Glu	Gln	Ala	Leu	Glu	Gln	His	Ile	Ser	Ala	Ile	Val	Ser	Asp	Ile							
		515					520					525										
Thr	Asn	Asp	Gly	Ile	Ile	Pro	Gln	Ala	Ile	Gln	Glu	Thr	Leu	Asp	Ser							
	530					535					540											
Leu	Arg	Ser	Ile	Ile	Leu	Phe	Ala															
545					550																	

<210> 17

<211> 523

5 <212> PRT

<213> Rhodobacter sphaeroides

<400> 17

Met	Leu	Ala	Met	Ser	Pro	Pro	Lys	Pro	Ala	Val	Glu	Leu	Asp	Arg	His
1			5						10					15	

Ile Asp Leu Asp Glu Ala His Ser Val Ala Ser Gly Gly Ala Arg Ile

10

ES 2 770 616 T3

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

ES 2 770 616 T3

Arg Val Val Arg His Val Ile Ala Glu Arg Arg Leu Gly Ala Ser Asp
 275 280 285

Ile Gly Thr Glu Pro Glu Ala Gly Gln Asp Ala Tyr Ser Leu Arg Cys
 290 295 300

Ala Pro Gln Val Leu Gly Ala Gly Phe Asp Thr Leu Ala Trp His Asp
 305 310 315 320

Arg Val Leu Thr Ile Glu Leu Asn Ala Val Thr Asp Asn Pro Val Phe
 325 330 335

Pro Pro Asp Gly Ser Val Pro Ala Leu His Gly Gly Asn Phe Met Gly
 340 345 350

Gln His Val Ala Leu Thr Ser Asp Ala Leu Ala Thr Ala Val Thr Val
 355 360 365

Leu Ala Gly Leu Ala Glu Arg Gln Ile Ala Arg Leu Thr Asp Glu Arg
 370 375 380

Leu Asn Arg Gly Leu Pro Pro Phe Leu His Arg Gly Pro Ala Gly Leu
 385 390 395 400

Asn Ser Gly Phe Met Gly Ala Gln Val Thr Ala Thr Ala Leu Leu Ala
 405 410 415

Glu Met Arg Ala Thr Gly Pro Ala Ser Ile His Ser Ile Ser Thr Asn
 420 425 430

Ala Ala Asn Gln Asp Val Val Ser Leu Gly Thr Ile Ala Ala Arg Leu
 435 440 445

Cys Arg Glu Lys Ile Asp Arg Trp Ala Glu Ile Leu Ala Ile Leu Ala
 450 455 460

Leu Cys Leu Ala Gln Ala Ala Glu Leu Arg Cys Gly Ser Gly Leu Asp
 465 470 475 480

Gly Val Ser Pro Ala Gly Lys Lys Leu Val Gln Ala Leu Arg Glu Gln
 485 490 495

Phe Pro Pro Leu Glu Thr Asp Arg Pro Leu Gly Gln Glu Ile Ala Ala
 500 505 510

Leu Ala Thr His Leu Leu Gln Gln Ser Pro Val
 515 520

ES 2 770 616 T3

<211> 510

<212> PRT

<213> Saccharothrix espanaensis

5 <400> 18

Met Thr Gln Val Val Glu Arg Gln Ala Asp Arg Leu Ser Ser Arg Glu
 1 5 10 15

Tyr Leu Ala Arg Val Val Arg Ser Ala Gly Trp Asp Ala Gly Leu Thr
 20 25 30

Ser Cys Thr Asp Glu Glu Ile Val Arg Met Gly Ala Ser Ala Arg Thr
 35 40 45

Ile Glu Glu Tyr Leu Lys Ser Asp Lys Pro Ile Tyr Gly Leu Thr Gln
 50 55 60

Gly Phe Gly Pro Leu Val Leu Phe Asp Ala Asp Ser Glu Leu Glu Gln
 65 70 75 80

Gly Gly Ser Leu Ile Ser His Leu Gly Thr Gly Gln Gly Ala Pro Leu
 85 90 95

Ala Pro Glu Val Ser Arg Leu Ile Leu Trp Leu Arg Ile Gln Asn Met
 100 105 110

Arg Lys Gly Tyr Ser Ala Val Ser Pro Val Phe Trp Gln Lys Leu Ala
 115 120 125

Asp Leu Trp Asn Lys Gly Phe Thr Pro Ala Ile Pro Arg His Gly Thr
 130 135 140

Val Ser Ala Ser Gly Asp Leu Gln Pro Leu Ala His Ala Ala Leu Ala
 145 150 155 160

Phe Thr Gly Val Gly Glu Ala Trp Thr Arg Asp Ala Asp Gly Arg Trp
 165 170 175

Ser Thr Val Pro Ala Val Asp Ala Leu Ala Ala Leu Gly Ala Glu Pro
 180 185 190

Phe Asp Trp Pro Val Arg Glu Ala Leu Ala Phe Val Asn Gly Thr Gly
 195 200 205

Ala Ser Leu Ala Val Ala Val Leu Asn His Arg Ser Ala Leu Arg Leu
 210 215 220

ES 2 770 616 T3

Val Arg Ala Cys Ala Val Leu Ser Ala Arg Leu Ala Thr Leu Leu Gly
 225 230 235 240

Ala Asn Pro Glu His Tyr Asp Val Gly His Gly Val Ala Arg Gly Gln
 245 250 255

Val Gly Gln Leu Thr Ala Ala Glu Trp Ile Arg Gln Gly Leu Pro Arg
 260 265 270

Gly Met Val Arg Asp Gly Ser Arg Pro Leu Gln Glu Pro Tyr Ser Leu
 275 280 285

Arg Cys Ala Pro Gln Val Leu Gly Ala Val Leu Asp Gln Leu Asp Gly
 290 295 300

Ala Gly Asp Val Leu Ala Arg Glu Val Asp Gly Cys Gln Asp Asn Pro
 305 310 315 320

Ile Thr Tyr Glu Gly Glu Leu Leu His Gly Gly Asn Phe His Ala Met
 325 330 335

Pro Val Gly Phe Ala Ser Asp Gln Ile Gly Leu Ala Met His Met Ala
 340 345 350

Ala Tyr Leu Ala Glu Arg Gln Leu Gly Leu Leu Val Ser Pro Val Thr
 355 360 365

Asn Gly Asp Leu Pro Pro Met Leu Thr Pro Arg Ala Gly Arg Gly Ala
 370 375 380

Gly Leu Ala Gly Val Gln Ile Ser Ala Thr Ser Phe Val Ser Arg Ile
 385 390 395 400

Arg Gln Leu Val Phe Pro Ala Ser Leu Thr Thr Leu Pro Thr Asn Gly
 405 410 415

Trp Asn Gln Asp His Val Pro Met Ala Leu Asn Gly Ala Asn Ser Val
 420 425 430

Phe Glu Ala Leu Glu Leu Gly Trp Leu Thr Val Gly Ser Leu Ala Val
 435 440 445

Gly Val Ala Gln Leu Ala Ala Met Thr Gly His Ala Ala Glu Gly Val
 450 455 460

Trp Ala Glu Leu Ala Gly Ile Cys Pro Pro Leu Asp Ala Asp Arg Pro
 465 470 475 480

ES 2 770 616 T3

Leu Gly Ala Glu Val Arg Ala Ala Arg Asp Leu Leu Ser Ala His Ala
 485 490 495

Asp Gln Leu Leu Val Asp Glu Ala Asp Gly Lys Asp Phe Gly
 500 505 510

<210> 19

<211> 519

5 <212> PRT

<213> Rheinheimera sp. A13L

<400> 19

Met Arg Ser Glu Gln Leu Thr Leu Glu Asp Val Glu Ala Ile Ala Leu
 1 5 10 15

Gly Arg Gln Thr Leu Val Val Thr Glu Lys Gln Met His Ala Val Glu
 20 25 30

Asn Ala His Lys Phe Leu Cys Arg Ala Ile Ser Asp Arg Lys Arg Ile
 35 40 45

Tyr Gly Val Thr Thr Gly Tyr Gly Pro Leu Ala Thr Thr Glu Val Asp
 50 55 60

Pro Arg Gln Ser Ala Leu Leu Gln Gln Asn Leu Val His His Leu Cys
 65 70 75 80

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

Val Ala Arg Leu Ile Ser Leu Leu Ser Gly His Ser Gly Ala Asn Pro
 100 105 110

Leu Leu Ile Lys Arg Met Gln Glu Trp Leu Asp Ala Asp Ile Val Pro
 115 120 125

Phe Ile Pro Cys Arg Gly Thr Val Gly Ala Ser Gly Asp Leu Thr Pro
 130 135 140

Leu Ala His Leu Ala Arg Ala Leu Ser Gly Gly Gly Lys Val Ser Ile
 145 150 155 160

Lys Gly Gly Leu Trp Ile Glu Ser Arg Asp Ala His Gln Gln Leu Gly
 165 170 175

Trp Gln Pro Leu Val Leu Lys Gly Lys Asp Ala Ile Ser Leu Val Asn
 180 185 190

10

ES 2 770 616 T3

Gly Thr Ser Ala Thr Val Gly Ile Ala Ala Leu Asn Ala Thr Ala Ala
195 200 205

Glu Arg Ala Leu Lys Leu Ser Thr Leu Leu Val Leu Leu Tyr Ala Glu
210 215 220

Leu Leu Asn Gly His Arg Glu Ala Phe His Pro Ala Ile Gly Gln Leu
225 230 235 240

Arg Pro His Pro Gly Gln Gln Lys Leu His Ser Trp Leu Trp Ser Leu
245 250 255

Ser Ala Ser Ser Asp Ala Leu Val Pro Trp Cys Ala Glu Ser Arg Asn
260 265 270

Leu Asn Leu Met Gly Glu Asp Ile Gln Gln Asn Gln Pro Leu Leu Gln
275 280 285

Asp Ala Tyr Thr Leu Arg Cys Ala Pro Gln Ala Leu Gly Ala Ala Leu
290 295 300

Asp Val Ile Ser Gln His Ala Thr Thr Val Lys Ile Glu Leu Ser Ala
305 310 315 320

Val Thr Asp Asn Pro Leu Leu Phe Ala Glu Asp Glu Leu Ile Leu His
325 330 335

Gly Gly Asn Phe Phe Gly Gln His Leu Ala Phe Ala Ser Asp His Leu
340 345 350

Asn Asn Ala Leu Ile Gln Met Ala Leu Tyr Ser Glu Arg Arg Ile Ala
355 360 365

Arg Ile Thr Asp Pro Leu Arg Asn Lys Gly Leu Pro Ala Phe Met Gln
370 375 380

Pro Leu Asp Thr Gly Leu His Ser Gly Phe Met Gly Ala Gln Val Cys
385 390 395 400

Ala Thr Ser Leu Val Ala Glu Leu Arg Ser Gln Ala Met Pro Ala Ser
405 410 415

Ile Gln Ser Ile Pro Thr Asn Ala Asp Asn Gln Asp Ile Val Pro Leu
420 425 430

Gly Thr Ile Ala Ala Arg Arg Ala Ser Thr Ser Leu Thr Gln Leu Tyr

ES 2 770 616 T3

435

440

445

Gln Ile Leu Ala Ile Glu Ala Leu Val Leu Val Gln Gly Ala Glu Leu
450 455 460

Lys Asn Thr His Ser Phe Ser His Ser Ser Gln Val Val Cys Ala Trp
465 470 475 480

Leu Arg Gln Tyr Ala Leu Pro Leu Lys Glu Asp Arg Ala Leu Ser Glu
485 490 495

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

Ser Leu Ile Glu Leu Leu Ala
515

<210> 20

<211> 713

5 <212> PRT

<213> Rhodotorula mucilaginosa

<400> 20

ES 2 770 616 T3

Met Ala Pro Ser Val Asp Ser Ile Ala Thr Ser Val Ala Asn Ser Leu
 1 5 10 15

Ser Asn Gly Leu His Ala Ala Ala Ala Asn Gly Gly Asp Val His
 20 25 30

Lys Lys Thr Ala Gly Ala Gly Ser Leu Leu Pro Thr Thr Glu Thr Thr
 35 40 45

Gln Leu Asp Ile Val Glu Arg Ile Leu Ala Asp Ala Gly Ala Thr Asp
 50 55 60

Gln Ile Lys Leu Asp Gly Tyr Thr Leu Thr Leu Gly Asp Val Val Gly
 65 70 75 80

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

Arg Glu Lys Ile Asp Ala Ser Val Glu Phe Leu Arg Thr Gln Leu Asp
 100 105 110

Asn Ser Val Tyr Gly Val Thr Thr Gly Phe Gly Gly Ser Ala Asp Thr
 115 120 125

Arg Thr Glu Asp Ala Ile Ser Leu Gln Lys Ala Leu Leu Glu His Gln

ES 2 770 616 T3

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

ES 2 770 616 T3

His Ala Val Leu Ser Leu Glu Ala Gly Gln Ser Thr Thr Asp Asn Pro
 385 390 395 400

Leu Ile Asp Leu Glu Asn Lys Met Thr His His Gly Gly Ala Phe Met
 405 410 415

Ala Ser Ser Val Gly Asn Thr Met Glu Lys Thr Arg Leu Ala Val Ala
 420 425 430

Leu Met Gly Lys Val Ser Phe Thr Gln Leu Thr Glu Met Leu Asn Ala
 435 440 445

Gly Met Asn Arg Ala Leu Pro Ser Cys Leu Ala Ala Glu Asp Pro Ser
 450 455 460

Leu Ser Tyr His Cys Lys Gly Leu Asp Ile Ala Ala Ala Ala Tyr Thr
 465 470 475 480

Ser Glu Leu Gly His Leu Ala Asn Pro Val Ser Thr His Val Gln Pro
 485 490 495

Ala Glu Met Gly Asn Gln Ala Ile Asn Ser Leu Ala Leu Ile Ser Ala
 500 505 510

Arg Arg Thr Ala Glu Ala Asn Asp Val Leu Ser Leu Leu Leu Ala Thr
 515 520 525

His Leu Tyr Cys Val Leu Gln Ala Val Asp Leu Arg Ala Met Glu Phe
 530 535 540

Glu His Thr Lys Ala Phe Glu Pro Met Val Thr Glu Leu Leu Lys Gln
 545 550 555 560

His Phe Gly Ala Leu Ala Thr Ala Glu Val Glu Asp Lys Val Arg Lys
 565 570 575

Ser Ile Tyr Lys Arg Leu Gln Gln Asn Asn Ser Tyr Asp Leu Glu Gln
 580 585 590

Arg Trp His Asp Thr Phe Ser Val Ala Thr Gly Ala Val Val Glu Ala
 595 600 605

Leu Ala Gly Gln Glu Val Ser Leu Ala Ser Leu Asn Ala Trp Lys Val
 610 615 620

Ala Cys Ala Glu Lys Ala Ile Ala Leu Thr Arg Ser Val Arg Asp Ser
 625 630 635 640

ES 2 770 616 T3

Phe Trp Ala Ala Pro Ser Ser Ser Ser Pro Ala Leu Lys Tyr Leu Ser
 645 650 655

Pro Arg Thr Arg Val Leu Tyr Ser Phe Val Arg Glu Glu Val Gly Val
 660 665 670

Lys Ala Arg Arg Gly Asp Val Tyr Leu Gly Lys Gln Glu Val Thr Ile
 675 680 685

Gly Thr Asn Val Ser Arg Ile Tyr Glu Ala Ile Lys Ser Gly Cys Ile
 690 695 700

Ala Pro Val Leu Val Lys Met Met Ala
 705 710

<210> 21

<211> 689

5 <212> PRT

<213> Trichosporon cutaneum

<400> 21

Met Phe Ile Glu Thr Asn Val Ala Lys Pro Ala Ser Thr Lys Ala Met
 1 5 10 15

Asn Ala Gly Ser Ala Lys Ala Ala Pro Val Glu Pro Phe Ala Thr Tyr
 20 25 30

Ala His Ser Gln Ala Thr Lys Thr Val Ser Ile Asp Gly His Thr Met
 35 40 45

Lys Val Gly Asp Val Val Ala Val Ala Arg His Gly Ala Lys Val Glu
 50 55 60

Leu Ala Ala Ser Val Ala Gly Pro Val Arg Ala Ser Val Asp Phe Lys
 65 70 75 80

Glu Ser Lys Lys His Thr Ser Ile Tyr Gly Val Thr Thr Gly Phe Gly
 85 90 95

Gly Ser Ala Asp Thr Arg Thr Ser Asp Thr Glu Ala Leu Gln Ile Ser
 100 105 110

Leu Leu Glu His Gln Leu Cys Gly Phe Leu Pro Thr Asp Ala Thr Tyr
 115 120 125

Glu Gly Met Leu Leu Ala Ala Met Pro Ile Pro Ile Val Arg Gly Ala
 130 135 140

10

ES 2 770 616 T3

Met Ala Val Arg Val Asn Ser Cys Val Arg Gly His Ser Gly Val Arg
145 150 155 160

Leu Glu Val Leu Gln Ser Phe Ala Asp Phe Ile Asn Arg Gly Leu Val
165 170 175

Pro Cys Val Pro Leu Arg Gly Thr Ile Ser Ala Ser Gly Asp Leu Ser
180 185 190

Pro Leu Ser Tyr Ile Ala Gly Ala Ile Cys Gly His Pro Asp Val Lys
195 200 205

Val Phe Asp Thr Ala Ala Ser Pro Pro Thr Val Leu Thr Ser Pro Glu
210 215 220

Ala Ile Ala Lys Tyr Gly Leu Lys Thr Val Lys Leu Ala Ser Lys Glu
225 230 235 240

Gly Leu Gly Leu Val Asn Gly Thr Ala Val Ser Ala Ala Ala Gly Ala
245 250 255

Leu Ala Leu Tyr Asp Ala Glu Cys Leu Ala Ile Met Ser Gln Thr Asn
260 265 270

Thr Val Leu Thr Val Glu Ala Leu Asp Gly His Val Gly Ser Phe Ala
275 280 285

Pro Phe Ile Gln Glu Ile Arg Pro His Ala Gly Gln Ile Glu Ala Ala
290 295 300

Arg Asn Ile Arg His Met Leu Gly Gly Ser Lys Leu Ala Val His Glu
305 310 315 320

Glu Ser Glu Leu Leu Ala Asp Gln Asp Ala Gly Ile Leu Arg Gln Asp
325 330 335

Arg Tyr Ala Leu Arg Thr Ser Ala Gln Trp Ile Gly Pro Gln Leu Glu
340 345 350

Ala Leu Gly Leu Ala Arg Gln Gln Ile Glu Thr Glu Leu Asn Ser Thr
355 360 365

Thr Asp Asn Pro Leu Ile Asp Val Glu Gly Gly Met Phe His His Gly
370 375 380

Gly Asn Phe Gln Ala Met Ala Val Thr Ser Ala Met Asp Ser Ala Arg
385 390 395 400

ES 2 770 616 T3

Ile Val Leu Gln Asn Leu Gly Lys Leu Ser Phe Ala Gln Val Thr Glu
405 410 415

Leu Ile Asn Cys Glu Met Asn His Gly Leu Pro Ser Asn Leu Ala Gly
420 425 430

Ser Glu Pro Ser Thr Asn Tyr His Cys Lys Gly Leu Asp Ile His Cys
435 440 445

Gly Ala Tyr Cys Ala Glu Leu Gly Phe Leu Ala Asn Pro Met Ser Asn
450 455 460

His Val Gln Ser Thr Glu Met His Asn Gln Ser Val Asn Ser Met Ala
465 470 475 480

Phe Ala Ser Ala Arg Arg Thr Met Glu Ala Asn Glu Val Leu Ser Leu
485 490 495

Leu Leu Gly Ser Gln Met Tyr Cys Ala Thr Gln Ala Leu Asp Leu Arg
500 505 510

Val Met Glu Val Lys Phe Lys Met Ala Ile Val Lys Leu Leu Asn Glu
515 520 525

Thr Leu Thr Lys His Phe Ala Ala Phe Leu Thr Pro Glu Gln Leu Ala
530 535 540

Lys Leu Asn Thr His Ala Ala Ile Thr Leu Tyr Lys Arg Leu Asn Gln
545 550 555 560

Thr Pro Ser Trp Asp Ser Ala Pro Arg Phe Glu Asp Ala Ala Lys His
565 570 575

Leu Val Gly Val Ile Met Asp Ala Leu Met Val Asn Asp Asp Ile Thr
580 585 590

Asp Leu Thr Asn Leu Pro Lys Trp Lys Lys Glu Phe Ala Lys Glu Ala
595 600 605

Gly Asn Leu Tyr Arg Ser Ile Leu Val Ala Thr Thr Ala Asp Gly Arg
610 615 620

Asn Asp Leu Glu Pro Ala Glu Tyr Leu Gly Gln Thr Arg Ala Val Tyr
625 630 635 640

Glu Ala Val Arg Ser Glu Leu Gly Val Lys Val Arg Arg Gly Asp Val

ES 2 770 616 T3

645

650

655

Ala Glu Gly Lys Ser Gly Lys Ser Ile Gly Ser Ser Val Ala Lys Ile
660 665 670

Val Glu Ala Met Arg Asp Gly Arg Leu Met Gly Ala Val Gly Lys Met
675 680 685

Phe

<210> 22

<211> 716

5 <212> PRT

<213> Rhodosporidium toruloides

<400> 22

Met Ala Pro Ser Leu Asp Ser Ile Ser His Ser Phe Ala Asn Gly Val
1 5 10 15

Ala Ser Ala Lys Gln Ala Val Asn Gly Ala Ser Thr Asn Leu Ala Val
20 25 30

Ala Gly Ser His Leu Pro Thr Thr Gln Val Thr Gln Val Asp Ile Val
35 40 45

Glu Lys Met Leu Ala Ala Pro Thr Asp Ser Thr Leu Glu Leu Asp Gly
50 55 60

Tyr Ser Leu Asn Leu Gly Asp Val Val Ser Ala Ala Arg Lys Gly Arg
65 70 75 80

Pro Val Arg Val Lys Asp Ser Asp Glu Ile Arg Ser Lys Ile Asp Lys
85 90 95

Ser Val Glu Phe Leu Arg Ser Gln Leu Ser Met Ser Val Tyr Gly Val
100 105 110

Thr Thr Gly Phe Gly Gly Ser Ala Asp Thr Arg Thr Glu Asp Ala Ile
115 120 125

Ser Leu Gln Lys Ala Leu Leu Glu His Gln Leu Cys Gly Val Leu Pro
130 135 140

Ser Ser Phe Asp Ser Phe Arg Leu Gly Arg Gly Leu Glu Asn Ser Leu
145 150 155 160

Pro Leu Glu Val Val Arg Gly Ala Met Thr Ile Arg Val Asn Ser Leu

10

ES 2 770 616 T3

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

ES 2 770 616 T3

Ala Val Ala Asn Thr Met Glu Lys Thr Arg Leu Gly Leu Ala Gln Ile
 420 425 430

Gly Lys Leu Asn Phe Thr Gln Leu Thr Glu Met Leu Asn Ala Gly Met
 435 440 445

Asn Arg Gly Leu Pro Ser Cys Leu Ala Ala Glu Asp Pro Ser Leu Ser
 450 455 460

Tyr His Cys Lys Gly Leu Asp Ile Ala Ala Ala Ala Tyr Thr Ser Glu
 465 470 475 480

Leu Gly His Leu Ala Asn Pro Val Thr Thr His Val Gln Pro Ala Glu
 485 490 495

Met Ala Asn Gln Ala Val Asn Ser Leu Ala Leu Ile Ser Ala Arg Arg
 500 505 510

Thr Thr Glu Ser Asn Asp Val Leu Ser Leu Leu Leu Ala Thr His Leu
 515 520 525

Tyr Cys Val Leu Gln Ala Ile Asp Leu Arg Ala Ile Glu Phe Glu Phe
 530 535 540

Lys Lys Gln Phe Gly Pro Ala Ile Val Ser Leu Ile Asp Gln His Phe
 545 550 555 560

Gly Ser Ala Met Thr Gly Ser Asn Leu Arg Asp Glu Leu Val Glu Lys
 565 570 575

Val Asn Lys Thr Leu Ala Lys Arg Leu Glu Gln Thr Asn Ser Tyr Asp
 580 585 590

Leu Val Pro Arg Trp His Asp Ala Phe Ser Phe Ala Ala Gly Thr Val
 595 600 605

Val Glu Val Leu Ser Ser Thr Ser Leu Ser Leu Ala Ala Val Asn Ala
 610 615 620

Trp Lys Val Ala Ala Ala Glu Ser Ala Ile Ser Leu Thr Arg Gln Val
 625 630 635 640

Arg Glu Thr Phe Trp Ser Ala Ala Ser Thr Ser Ser Pro Ala Leu Ser
 645 650 655

Tyr Leu Ser Pro Arg Thr Gln Ile Leu Tyr Ala Phe Val Arg Glu Glu
 660 665 670

ES 2 770 616 T3

Leu Gly Val Lys Ala Arg Arg Gly Asp Val Phe Leu Gly Lys Gln Glu
675 680 685

Val Thr Ile Gly Ser Asn Val Ser Lys Ile Tyr Glu Ala Ile Lys Ser
690 695 700

Gly Arg Ile Asn Asn Val Leu Leu Lys Met Leu Ala
705 710 715

<210> 23

<211> 737

5 <212> PRT

<213> Phanerochaete chrysosporium

<400> 23

Met Pro Ser Arg Ile Asp Tyr Tyr Thr Ser Ser Gly Asn Gly Tyr Ala
1 5 10 15

Gln Ser Arg Lys Ser Ser Ala Ile Tyr Pro Ala Ser Ala Ser Thr Gly
20 25 30

His Ala Ala Pro Ser Thr Glu Arg Lys Pro Glu Leu Leu Asp Lys Phe
35 40 45

Val Glu Ala Tyr Asp Glu Leu Gln Ser Tyr Arg Glu Gly Lys Pro Val
50 55 60

Ile Val Asp Gly His Asn Leu Ser Ile Pro Ala Val Ala Ala Thr Ala
65 70 75 80

Arg Phe Gly Ala Ala Val Val Leu Asp Glu Asn Pro Glu Thr His Glu
85 90 95

Arg Val Leu Gln Ser Arg Arg Val Ile Val Asp Lys Val Ser Thr Gln
100 105 110

Arg Ser Val Tyr Gly Val Ser Thr Gly Phe Gly Gly Ser Ala Asp Thr
115 120 125

Arg Thr Ser Asp Pro Leu Gln Leu Gly His Ala Leu Leu Gln His Gln
130 135 140

His Val Gly Val Leu Pro Thr Gln Thr Glu Ser Pro Leu Pro Ala Leu
145 150 155 160

Pro Leu Gly Asp Pro Leu Ala Thr Thr Ser Met Pro Glu Ala Trp Val
165 170 175

10

ES 2 770 616 T3

Arg Gly Ala Ile Leu Ile Arg Met Asn Ser Leu Ile Arg Gly His Ser
 180 185 190

Gly Val Arg Trp Glu Leu Ile Glu Lys Met Gly Glu Leu Leu Arg Glu
 195 200 205

Asn Ile Thr Pro Leu Val Pro Leu Arg Gly Ser Ile Ser Ala Ser Gly
 210 215 220

Asp Leu Ser Pro Leu Ser Tyr Ile Ala Gly Thr Leu Ile Gly Ser Pro
 225 230 235 240

Ala Ile Arg Val Phe Asp Gly Pro Ala Ser Tyr Gly Ala Arg Arg Ile
 245 250 255

Leu Pro Ser Asn Ile Ala Leu Ala Asn His Gly Val Ala Pro Ile Pro
 260 265 270

Leu Ser Ser Lys Glu His Leu Gly Ile Leu Asn Gly Thr Ala Phe Ser
 275 280 285

Ala Ser Val Gly Ala Leu Ala Leu Asn Glu Ala Val His Leu Ser Leu
 290 295 300

Leu Ala Gln Val Cys Thr Ala Met Gly Thr Glu Ala Met Ile Gly Ala
 305 310 315 320

Val Gly Ser Phe Asp Ala Phe Ile His Asp Thr Ala Arg Pro His Pro
 325 330 335

Gly Gln Val Glu Val Ala Arg Asn Val Arg Thr Leu Leu Glu Asp Ser
 340 345 350

Gln Met Ala Val Lys Ala Glu Asp Glu Val His Ile Ala Glu Asp Glu
 355 360 365

Gly Glu Leu Arg Gln Asp Arg Tyr Pro Leu Arg Thr Ala Ala Gln Phe
 370 375 380

Leu Gly Pro Gln Ile Glu Asp Ile Leu Ser Ala His Glu Thr Val Thr
 385 390 395 400

Leu Glu Cys Asn Ser Thr Thr Asp Asn Pro Leu Ile Asp Gly Glu Thr
 405 410 415

Gly Thr Val His His Gly Gly Asn Phe Gln Ala Met Ala Val Thr Asn
 420 425 430

ES 2 770 616 T3

Ala Met Glu Lys Thr Arg Leu Ala Ile His His Ile Gly Lys Leu Leu
435 440 445

Phe Ala Gln Ala Thr Glu Leu Ile Asn Pro Met Met Asn Arg Gly Leu
450 455 460

Pro Pro Asn Leu Ala Ala Thr Asp Pro Ser His Asn Tyr Phe Ala Lys
465 470 475 480

Gly Val Asp Ile His Leu Ala Ala Tyr Val Gly Glu Leu Gly Phe Leu
485 490 495

Ala Ser Pro Val Ser Ser His Ile Gln Ser Ala Glu Met His Asn Gln
500 505 510

Ala Val Asn Ser Leu Ala Leu Val Ser Ala Arg Tyr Thr Ile Ser Ala
515 520 525

Leu Asp Val Leu Ser Leu Leu Thr Ala Ala Tyr Leu Tyr Val Leu Cys
530 535 540

Gln Ala Leu Asp Leu Arg Ala Met His Asn Asp Leu Gln Ser Ser Leu
545 550 555 560

Ser Ala Ile Val Arg Glu Leu Leu Pro Lys His Phe Pro Ser Ala Ala
565 570 575

Lys Arg Ala Asp Ala Leu Leu Pro Ile Leu Glu Arg Thr Ile Phe Arg
580 585 590

Ala Leu Asn Ser Ser Ser Ser Ala Asp Cys Lys Ala Arg Met Val Ser
595 600 605

Val Ala Ala Ser Thr Thr Thr Pro Leu Val Asp Phe Leu Ser Ala Asp
610 615 620

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

Leu Ala Thr Arg Ala Ala Asp Ala Leu Thr Thr Leu Arg Thr Gln Tyr
645 650 655

Leu Glu Gly Ala Arg Gly Ala Ala Pro Ala Ser Lys Tyr Leu Gly Lys
660 665 670

Thr Arg Pro Val Tyr Glu Phe Val Arg Val Thr Leu Asn Val Pro Met

ES 2 770 616 T3

675

680

685

His Gly Arg Glu Asn Leu His Asn Phe Glu Met Gly Pro Gly Val Glu
690 695 700

Asp Gly Ile Ile Gly Asn Asn Ile Ser Thr Ile Tyr Glu Ala Ile Arg
705 710 715 720

Asp Gly Lys Met Gln Asn Val Val Met Gln Leu Val Lys Ser Ile Lys
725 730 735

Ala

<210> 24

<211> 14

5 <212> PRT

<213> Secuencia artificial

<220>

<223> Etiqueta His

10

<400> 24

Met Ala His His His His His His Glu Asn Leu Tyr Phe Gln
1 5 10

15 <210> 25

<211> 302

<212> PRT

<213> *Escherichia coli*

20 <400> 25

ES 2 770 616 T3

Met Asp Gln Ile Arg Leu Thr His Leu Arg Gln Leu Glu Ala Glu Ser
1 5 10 15

Ile His Ile Ile Arg Glu Val Ala Ala Glu Phe Ser Asn Pro Val Met
 20 25 30

Leu Tyr Ser Ile Gly Lys Asp Ser Ser Val Met Leu His Leu Ala Arg
 35 40 45

Lys Ala Phe Tyr Pro Gly Thr Leu Pro Phe Pro Leu Leu His Val Asp
 50 55 60

Thr Gly Trp Lys Phe Arg Glu Met Tyr Glu Phe Arg Asp Arg Thr Ala
65 70 75 80

Lys Ala Tyr Gly Cys Glu Leu Leu Val His Lys Asn Pro Glu Gly Val
 85 90 95

ES 2 770 616 T3

Ala Met Gly Ile Asn Pro Phe Val His Gly Ser Ala Lys His Thr Asp
 100 105 110

Ile Met Lys Thr Glu Gly Leu Lys Gln Ala Leu Asn Lys Tyr Gly Phe
 115 120 125

Asp Ala Ala Phe Gly Gly Ala Arg Arg Asp Glu Glu Lys Ser Arg Ala
 130 135 140

Lys Glu Arg Ile Tyr Ser Phe Arg Asp Arg Phe His Arg Trp Asp Pro
 145 150 155 160

Lys Asn Gln Arg Pro Glu Leu Trp His Asn Tyr Asn Gly Gln Ile Asn
 165 170 175

Lys Gly Glu Ser Ile Arg Val Phe Pro Leu Ser Asn Trp Thr Glu Gln
 180 185 190

Asp Ile Trp Gln Tyr Ile Trp Leu Glu Asn Ile Asp Ile Val Pro Leu
 195 200 205

Tyr Leu Ala Ala Glu Arg Pro Val Leu Glu Arg Asp Gly Met Leu Met
 210 215 220

Met Ile Asp Asp Asn Arg Ile Asp Leu Gln Pro Gly Glu Val Ile Lys
 225 230 235 240

Lys Arg Met Val Arg Phe Arg Thr Leu Gly Cys Trp Pro Leu Thr Gly
 245 250 255

Ala Val Glu Ser Asn Ala Gln Thr Leu Pro Glu Ile Ile Glu Glu Met
 260 265 270

Leu Val Ser Thr Thr Ser Glu Arg Gln Gly Arg Val Ile Asp Arg Asp
 275 280 285

Gln Ala Gly Ser Met Glu Leu Lys Lys Arg Gln Gly Tyr Phe
 290 295 300

<210> 26

<211> 475

5 <212> PRT

<213> *Escherichia coli*

<400> 26

Met Asn Thr Ala Leu Ala Gln Gln Ile Ala Asn Glu Gly Gly Val Glu
 1 5 10 15

10

ES 2 770 616 T3

Ala Trp Met Ile Ala Gln Gln His Lys Ser Leu Leu Arg Phe Leu Thr
 20 25 30

Cys Gly Ser Val Asp Asp Gly Lys Ser Thr Leu Ile Gly Arg Leu Leu
 35 40 45

His Asp Thr Arg Gln Ile Tyr Glu Asp Gln Leu Ser Ser Leu His Asn
 50 55 60

Asp Ser Lys Arg His Gly Thr Gln Gly Glu Lys Leu Asp Leu Ala Leu
 65 70 75 80

Leu Val Asp Gly Leu Gln Ala Glu Arg Glu Gln Gly Ile Thr Ile Asp
 85 90 95

Val Ala Tyr Arg Tyr Phe Ser Thr Glu Lys Arg Lys Phe Ile Ile Ala
 100 105 110

Asp Thr Pro Gly His Glu Gln Tyr Thr Arg Asn Met Ala Thr Gly Ala
 115 120 125

Ser Thr Cys Glu Leu Ala Ile Leu Leu Ile Asp Ala Arg Lys Gly Val
 130 135 140

Leu Asp Gln Thr Arg Arg His Ser Phe Ile Ser Thr Leu Leu Gly Ile
 145 150 155 160

Lys His Leu Val Val Ala Ile Asn Lys Met Asp Leu Val Asp Tyr Ser
 165 170 175

Glu Glu Thr Phe Thr Arg Ile Arg Glu Asp Tyr Leu Thr Phe Ala Gly
 180 185 190

Gln Leu Pro Gly Asn Leu Asp Ile Arg Phe Val Pro Leu Ser Ala Leu
 195 200 205

Glu Gly Asp Asn Val Ala Ser Gln Ser Glu Ser Met Pro Trp Tyr Ser
 210 215 220

Gly Pro Thr Leu Leu Glu Val Leu Glu Thr Val Glu Ile Gln Arg Val
 225 230 235 240

Val Asp Ala Gln Pro Met Arg Phe Pro Val Gln Tyr Val Asn Arg Pro
 245 250 255

Asn Leu Asp Phe Arg Gly Tyr Ala Gly Thr Leu Ala Ser Gly Arg Val
 260 265 270

ES 2 770 616 T3

Glu Val Gly Gln Arg Val Lys Val Leu Pro Ser Gly Val Glu Ser Asn
 275 280 285
 Val Ala Arg Ile Val Thr Phe Asp Gly Asp Arg Glu Glu Ala Phe Ala
 290 295 300
 Gly Glu Ala Ile Thr Leu Val Leu Thr Asp Glu Ile Asp Ile Ser Arg
 305 310 315 320
 Gly Asp Leu Leu Leu Ala Ala Asp Glu Ala Leu Pro Ala Val Gln Ser
 325 330 335
 Ala Ser Val Asp Val Val Trp Met Ala Glu Gln Pro Leu Ser Pro Gly
 340 345 350
 Gln Ser Tyr Asp Ile Lys Ile Ala Gly Lys Lys Thr Arg Ala Arg Val
 355 360 365
 Asp Gly Ile Arg Tyr Gln Val Asp Ile Asn Asn Leu Thr Gln Arg Glu
 370 375 380
 Val Glu Asn Leu Pro Leu Asn Gly Ile Gly Leu Val Asp Leu Thr Phe
 385 390 395 400
 Asp Glu Pro Leu Val Leu Asp Arg Tyr Gln Gln Asn Pro Val Thr Gly
 405 410 415
 Gly Leu Ile Phe Ile Asp Arg Leu Ser Asn Val Thr Val Gly Ala Gly
 420 425 430
 Met Val His Glu Pro Val Ser Gln Ala Thr Ala Ala Pro Ser Glu Phe
 435 440 445
 Ser Ala Phe Glu Leu Glu Leu Asn Ala Leu Val Arg Arg His Phe Pro
 450 455 460
 His Trp Gly Ala Arg Asp Leu Leu Gly Asp Lys
 465 470 475

<210> 27

<211> 201

5 <212> PRT

<213> *Escherichia coli*

<400> 27

ES 2 770 616 T3

Met Ala Leu His Asp Glu Asn Val Val Trp His Ser His Pro Val Thr
 1 5 10 15

Val Gln Gln Arg Glu Leu His His Gly His Arg Gly Val Val Leu Trp
 20 25 30

Phe Thr Gly Leu Ser Gly Ser Gly Lys Ser Thr Val Ala Gly Ala Leu
 35 40 45

Glu Glu Ala Leu His Lys Leu Gly Val Ser Thr Tyr Leu Leu Asp Gly
 50 55 60

Asp Asn Val Arg His Gly Leu Cys Ser Asp Leu Gly Phe Ser Asp Ala
 65 70 75 80

Asp Arg Lys Glu Asn Ile Arg Arg Val Gly Glu Val Ala Asn Leu Met
 85 90 95

Val Glu Ala Gly Leu Val Val Leu Thr Ala Phe Ile Ser Pro His Arg
 100 105 110

Ala Glu Arg Gln Met Val Arg Glu Arg Val Gly Glu Gly Arg Phe Ile
 115 120 125

Glu Val Phe Val Asp Thr Pro Leu Ala Ile Cys Glu Ala Arg Asp Pro
 130 135 140

Lys Gly Leu Tyr Lys Lys Ala Arg Ala Gly Glu Leu Arg Asn Phe Thr
 145 150 155 160

Gly Ile Asp Ser Val Tyr Glu Ala Pro Glu Ser Ala Glu Ile His Leu
 165 170 175

Asn Gly Glu Gln Leu Val Thr Asn Leu Val Gln Gln Leu Leu Asp Leu
 180 185 190

Leu Arg Gln Asn Asp Ile Ile Arg Ser
 195 200

5 <210> 28
 <211> 246
 <212> PRT
 <213> *Escherichia coli*

10 <400> 28

ES 2 770 616 T3

Met Leu Asp Gln Val Cys Gln Leu Ala Arg Asn Ala Gly Asp Ala Ile
 1 5 10 15

Met Gln Val Tyr Asp Gly Thr Lys Pro Met Asp Val Val Ser Lys Ala
 20 25 30

Asp Asn Ser Pro Val Thr Ala Ala Asp Ile Ala Ala His Thr Val Ile
 35 40 45

Met Asp Gly Leu Arg Thr Leu Thr Pro Asp Val Pro Val Leu Ser Glu
 50 55 60

Glu Asp Pro Pro Gly Trp Glu Val Arg Gln His Trp Gln Arg Tyr Trp
 65 70 75 80

Leu Val Asp Pro Leu Asp Gly Thr Lys Glu Phe Ile Lys Arg Asn Gly
 85 90 95

Glu Phe Thr Val Asn Ile Ala Leu Ile Asp His Gly Lys Pro Ile Leu
 100 105 110

Gly Val Val Tyr Ala Pro Val Met Asn Val Met Tyr Ser Ala Ala Glu
 115 120 125

Gly Lys Ala Trp Lys Glu Glu Cys Gly Val Arg Lys Gln Ile Gln Val
 130 135 140

Arg Asp Ala Arg Pro Pro Leu Val Val Ile Ser Arg Ser His Ala Asp
 145 150 155 160

Ala Glu Leu Lys Glu Tyr Leu Gln Gln Leu Gly Glu His Gln Thr Thr
 165 170 175

Ser Ile Gly Ser Ser Leu Lys Phe Cys Leu Val Ala Glu Gly Gln Ala
 180 185 190

Gln Leu Tyr Pro Arg Phe Gly Pro Thr Asn Ile Trp Asp Thr Ala Ala
 195 200 205

Gly His Ala Val Ala Ala Ala Ala Gly Ala His Val His Asp Trp Gln
 210 215 220

Gly Lys Pro Leu Asp Tyr Thr Pro Arg Glu Ser Phe Leu Asn Pro Gly
 225 230 235 240

Phe Arg Val Ser Ile Tyr
 245

ES 2 770 616 T3

<210> 29
<211> 58
<212> ADN
<213> Secuencia artificial

5

<220>
<223> Cebador CBJP472

10 <400> 29
tagaaataat tttgtttaac ttaagaagg agatatacca tggagtctc ccgccac 58

<210> 30
<211> 49
<212> ADN
<213> Secuencia artificial

15

<220>
<223> Cebador CBJP473

20 <400> 30
taagcattat gcggccgcaa gctgtcata gtcacaacg aaactgaa 49

<210> 31
<211> 62
<212> ADN
<213> Secuencia artificial

25

<220>
<223> Cebador CBJP470

30 <400> 31
tagaaataat tttgtttaac ttaagaagg agatatacca tggaactgat tcaggatacc 60
ag 62

35 <210> 32
<211> 47
<212> ADN
<213> Secuencia artificial

40 <220>
<223> Cebador CBJP471

45 <400> 32
taagcattat gcggccgcaa gctgttaca gttcgctacg aaagctc 47

ES 2 770 616 T3

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

5

<220>
<223> Cebador CBJP499

<400> 33
10 tagaaataat tttgttaac ttaagaagg agatatacca tggagctgat ccaggacacc 60

<210> 34
<211> 45
<212> ADN
15 <213> Secuencia artificial

<220>
<223> Cebador CBJP500

<400> 34
20 taagcattat gcggccgcaa gctgtcaca cctctgagcg gaagc 45

<210> 35
<211> 57
25 <212> ADN
<213> Secuencia artificial

<220>
<223> Cebador CBJP505

<400> 35
30 tagaaataat tttgttaac ttaagaagg agatatacca tggagccggt ccaggac 57

<210> 36
35 <211> 45
<212> ADN
<213> Secuencia artificial

<220>
40 <223> Cebador CBJP506

<400> 36
taagcattat gcggccgcaa gctgtcaca gctcagagcg gaagc 45

<210> 37
45 <211> 57

ES 2 770 616 T3

<212> ADN
 <213> Secuencia artificial

5 <220>
 <223> Cebador CBJP503

<400> 37
 tagaaataat tttgtttaac ttaagaagg agatatacca tggaggacat tcccgac 57

10 <210> 38
 <211> 45
 <212> ADN
 <213> Secuencia artificial

15 <220>
 <223> Cebador CBJP504

20 <400> 38
 taagcattat gcggccgcaa gctgtcaca gctgtgtgcg gaagc 45

<210> 39
 <211> 62
 <212> ADN
 <213> Secuencia artificial

25 <220>
 <223> Cebador CBJP501

30 <400> 39
 tagaaataat tttgtttaac ttaagaagg agatatacca tggggaatga tgaggtgatc 60
 ag 62

<210> 40
 <211> 54

35 <212> ADN
 <213> Secuencia artificial

<220>
 <223> Cebador CBJP502

40 <400> 40
 taagcattat gcggccgcaa gctgttact ctgtctattg caattatta cagg 54

<210> 41

45 <211> 65

ES 2 770 616 T3

<212> ADN
<213> Secuencia artificial

5 <220>
<223> Cebador CBJP491

<400> 41

catcttagta tattagttaa gtataagaag gagatataca tatggatcaa atacgactta 60

10 ctcac 65

<210> 42

<211> 46

<212> ADN
<213> Secuencia artificial

15

<220>
<223> Cebador CBJP492

20 <400> 42
tgcccgccg atatccaatt gatcaggatc tgataatc gttctg 46

<210> 43

<211> 24

25 <212> ADN
<213> Secuencia artificial

<220>
<223> Cebador CBJP497

30 <400> 43
tcaggatctg ataatatcgt tctg 24

<210> 44

<211> 59

35 <212> ADN
<213> Secuencia artificial

<220>
<223> Cebador CBJP498

40 <400> 44
cagaacgata ttatcagatc ctgataagtt aacaccgctc acagagacga ggtggagaa 59

<210> 45

45 <211> 46

ES 2 770 616 T3

<212> ADN
 <213> Secuencia artificial

5 <220>
 <223> Cebador CBJP496

<400> 45
 tggccggccg atatccaatt gattagtaaa tagacactct gaacc 46

10 <210> 46
 <211> 62
 <212> ADN
 <213> Secuencia artificial

15 <220>
 <223> Cebador CBJP487

<400> 46
 catcttagta tattagttaa gtataagaag gagatataca tatggcaccg agcggtgata 60
 gc 62

20 <210> 47
 <211> 47
 <212> ADN
 <213> Secuencia artificial

25 <220>
 <223> Cebador CBJP488

30 <400> 47
 tggccggccg atatccaatt gattaggcca tcattttaac cagaacc 47

<210> 48
 <211> 540

35 <212> PRT
 <213> Rhodobacter capsulatus

<400> 48

ES 2 770 616 T3

Met Leu Asp Ala Thr Ile Gly Arg Lys Arg Met Thr Leu Gln Ser Gln
1 5 10 15

Thr Ala Lys Asp Cys Leu Ala Leu Asp Gly Ala Leu Thr Leu Val Gln
20 25 30

Cys Glu Ala Ile Ala Thr His Arg Ser Arg Ile Ser Val Thr Pro Ala
35 40 45

Leu Arg Glu Arg Cys Ala Arg Ala His Ala Arg Leu Glu His Ala Ile
50 55 60

ES 2 770 616 T3

Ala Glu Gln Arg His Ile Tyr Gly Ile Thr Thr Gly Phe Gly Pro Leu
65 70 75 80

Ala Asn Arg Leu Ile Gly Ala Asp Gln Gly Ala Glu Leu Gln Gln Asn
85 90 95

Leu Ile Tyr His Leu Ala Thr Gly Val Gly Pro Lys Leu Ser Trp Ala
100 105 110

Glu Ala Arg Ala Leu Met Leu Ala Arg Leu Asn Ser Ile Leu Gln Gly
115 120 125

Ala Ser Gly Ala Ser Pro Glu Thr Ile Asp Arg Ile Val Ala Val Leu
130 135 140

Asn Ala Gly Phe Ala Pro Glu Val Pro Ala Gln Gly Thr Val Gly Ala
145 150 155 160

Ser Gly Asp Leu Thr Pro Leu Ala His Met Val Leu Ala Leu Gln Gly
165 170 175

Arg Gly Arg Met Ile Asp Pro Ser Gly Arg Val Gln Glu Ala Gly Ala
180 185 190

Val Met Asp Arg Leu Cys Gly Gly Pro Leu Thr Leu Ala Ala Arg Asp
195 200 205

Gly Leu Ala Leu Val Asn Gly Thr Ser Ala Met Thr Ala Ile Ala Ala
210 215 220

Leu Thr Gly Val Glu Ala Ala Arg Ala Ile Asp Ala Ala Leu Arg His
225 230 235 240

Ser Ala Val Leu Met Glu Val Leu Ser Gly His Ala Glu Ala Trp His
245 250 255

Pro Ala Phe Ala Glu Leu Arg Pro His Pro Gly Gln Leu Arg Ala Thr
260 265 270

Glu Arg Leu Ala Gln Ala Leu Asp Gly Ala Gly Arg Val Cys Arg Thr
275 280 285

Leu Thr Ala Ala Arg Arg Leu Thr Ala Ala Asp Leu Arg Pro Glu Asp
290 295 300

His Pro Ala Gln Asp Ala Tyr Ser Leu Arg Val Val Pro Gln Leu Val

ES 2 770 616 T3

305 310 315 320

Gly Ala Val Trp Asp Thr Leu Asp Trp His Asp Arg Val Val Thr Cys
 325 330 335

Glu Leu Asn Ser Val Thr Asp Asn Pro Ile Phe Pro Glu Gly Cys Ala
 340 345 350

Val Pro Ala Leu His Gly Gly Asn Phe Met Gly Val His Val Ala Leu
 355 360 365

Ala Ser Asp Ala Leu Asn Ala Ala Leu Val Thr Leu Ala Gly Leu Val
 370 375 380

Glu Arg Gln Ile Ala Arg Leu Thr Asp Glu Lys Leu Asn Lys Gly Leu
 385 390 395 400

Pro Ala Phe Leu His Gly Gly Gln Ala Gly Leu Gln Ser Gly Phe Met
 405 410 415

Gly Ala Gln Val Thr Ala Thr Ala Leu Leu Ala Glu Met Arg Ala Asn
 420 425 430

Ala Thr Pro Val Ser Val Gln Ser Leu Ser Thr Asn Gly Ala Asn Gln
 435 440 445

Asp Val Val Ser Met Gly Thr Ile Ala Ala Arg Arg Ala Arg Ala Gln
 450 455 460

Leu Leu Pro Leu Ser Gln Ile Gln Ala Ile Leu Ala Leu Ala Leu Ala
 465 470 475 480

Gln Ala Met Asp Leu Leu Asp Asp Pro Glu Gly Gln Ala Gly Trp Ser
 485 490 495

Leu Thr Ala Arg Asp Leu Arg Asp Arg Ile Arg Ala Val Ser Pro Gly
 500 505 510

Leu Arg Ala Asp Arg Pro Leu Ala Gly Asp Ile Glu Ala Val Ala Gln
 515 520 525

Gly Leu Arg His Pro Ser Ala Ala Asp Pro Pro Ala
 530 535 540

<210> 49

<211> 1620

5 <212> ADN

<213> Rhodobacter capsulatus

ES 2 770 616 T3

<400> 49

```

atgctggatg caaccattgg tcgtaaactg atgaccctgc agagccagac cgaaaagat      60
tgtctggcac tggatggtgc actgaccctg gttcagtgag aagcaattgc aaccatcgt      120
agccgtatta gcgttacacc ggcaactgct gaacggtgtg cacgtgccca tgcacgtctg      180
gaacatgcaa ttgccgaaca gcgtcatatt tatggtatta ccaccggtt  tggtcgctg      240
gcaaactgct tgattggtgc agatcagggt gcagaactgc agcagaatct gatttatcat      300
ctggcaaccg gtgtgggtcc gaaactgagc tgggctgaag cccgtgcaact gatgctggca      360
cgtctgaata gcatcctgca ggggtcaagc ggtgcaagtc cggaaaccat tgatgcatt      420
gttgccgttc tgaatgcagg ttttgaccgc gaagttccgg cacagggcac cgttgggtgc      480
agcggtgatc tgactccgct ggcccatatg gttctggccc tgcaaggtcg tggtcgtatg      540
attgatccga gcggctcgtg tcaagaggca ggcgcagtta tggatcgtct gtgtggcgtg      600
ccgctgacac tggcagcacg tgatggtctg gcgctggtta atggcaccag cgcaatgacc      660
gcaattgcag cactgaccgg tgttgaagcc gcacgtgcaa ttgatgcagc cctgcgtcat      720
agcgcagttc tgatggaagt tctgagcggg catgcagaag catggcatcc ggcatctgcg      780
gaactgcgct cgcacccggg tcagctcgtg gcaaccgaac gtctggcaca gccctggat      840
ggcgcaggtc gtgtttgtcg taccctgacc gcagcacgtc gtctgacagc agccgatctg      900
cgtccggaag atcatcctgc acaggatgca tatagcctgc gtgttgttcc gcagctggtt      960
gggtgcagttt gggataccct ggattggcat gatcgtgttg ttacctgtga actgaatagc     1020
gttaccgata atccgatttt tccggaaggt tgtgcagttc ctgccctgca tggtaggcaat     1080
tttatgggtg ttcattgttc actggcaagt gatgcactga atgccgcaact ggttaccctg     1140
gcaggtctgg ttgaactgca gattgccctg ctgaccgatg aaaaactgaa taaaggctctg     1200
cctgcctttc tgcatggcgg tcaggctggt ctgcagagcg gttttatggg agcacaggtt     1260
accgcaaccg cactgctggc agaaatcgtg gcaaatgcga caccggttag cgttcagagc     1320
ctgagcacca atggtgcgaa tcaggatggt gttagcatgg gtacaattgc cgcacgtcgt     1380
gcgctgcgac agctgctgcc gctgagccag attcaggcaa tcctggccct ggctctggcc     1440
caggcaatgg atctgctgga tgatccgga ggcaggcag gttggagtct gaccgcacgt     1500
gatctgcgtg atcgtattcg tgcagttagt ccgggtctgc gtgcagatcg tcctctggca     1560
ggcgatattg aagcagttgc acagggactg cgtcatccga gcgcagcga tcctccggca     1620

```

5

<210> 50

<211> 1569

<212> ADN

<213> Rhodobacter sphaeroides

10

<400> 50

ES 2 770 616 T3

atgctggcaa tgagccctcc gaaaccggca gttgaactgg atcgtcatat tgatctggat 60
 gaagcacata gcgttgcaag cgggtggtgca cgtattgttc tggcaccgcc tgcacgtgat 120
 cgttgtcgtg caagcgaagc acgtctgggt gcagttattc gtgaagcccg tcatgtttat 180
 ggtctgacca ccggttttgg tccgctggca aatcgtctgg ttagcgggtga aaatgttcgt 240
 accctgcagg caaatctggt tcatcatctg gccagcgggtg tgggtccgggt tctggattgg 300
 accacgcgac gtgcaatggt gctggcacgc ctggttgcaa ttgccagggtg tgcgagcgggt 360
 gcaagtgaag gtacaattgc acgtctgatt gatctgctga atagcgaact ggcaccggca 420
 gtgccgatgc gtggcaccgt tgggtcatca ggtgatctga ctccgctggc ccatatggtt 480
 ctgtgtctgc agggctcgtg tgattttctg gatcgtgatg gcaccctctt ggatggtgcc 540
 gaaggtctgc gtcgtggtcg tctgcagccg ctggatctga gccatcgtga tgcactggca 600
 ctggttaatg gcaccagcgc aatgacaggt attgactggg tgaatgcaca tgcctgtcgt 660
 catctgggta attgggcagt tgcactgacc gactgctggg ccgaatgtct ggggtggtcgt 720
 accgaagcat gggcagcagc actgagcgat ctgctccgc atccgggtca gaaagatgca 780
 gcagcccgtc tgcgtgcacg tgttgatggt agcgcacgtg tggttcgtca tgttattgca 840
 gaacgtcgcc tgggtgccag cgatattggc accgaaccgg aagcagggtca ggatgcatat 900
 agcctgcggt gtgcaccgca ggttctgggt gccggttttg ataccctggc atggcatgat 960
 cgtgttctga ccatgaact gaatgcagtt accgataatc cggtttttcc tccggatggt 1020
 agtgttccgg cactgcatgg tggcaathtt atgggtcagc atgttgccct gacctcagat 1080
 gccctggcaa ccgcagtgac cgttctggca ggtctggccg aacgtcagat tgcccgtctg 1140
 accgatgaac gtctgaatcg tggctgcct ccgttctgc accgtggtcc ggcaggcctg 1200
 aatagtggct ttatgggtgc acaggttacc gcaacagccc tgctggcaga aatgcgtgca 1260
 accggtccgg caagcattca tagcattagc accaatgcag caaatcagga tgttgttagc 1320
 ctgggtacga ttgccgcacg tctgtgtcgt gaaaaaattg atcgttgggc agaaattctg 1380
 gccattctg cactgtgtct ggcacaggca gcagaactgc gttgcggtag tggcctggat 1440
 ggcgtttcac cggcaggtaa aaaactggtt caggcactgc gcgaacagtt tccgcctctg 1500
 gaaaccgatc gtccgctggg tcaagaaatt gcagcactgg caaccatct gctgcaacag 1560
 agtccggtt 1569

<210> 51

<211> 1518

5

<212> ADN

<213> *Flavobacterium johnsoniae*

<400> 51

10

atgaacacca tcaacgaata tctgagcctg gaagaatttg aagccattat ctttggcaat 60

ES 2 770 616 T3

cagaaagtga ccattagtga tgttgttg aatcgcgta acgagagctt taactttctg 120
aaagaattta gcggcaacaa agtgatctat ggtgtgaata ccggttttgg tccgatggca 180
cagtatcgta ttaaagaaag cgatcagatt cagctgcagt ataactctgat tcgtagccat 240
agcagcggca ccggtaaacc gctgagtcgg gtttgtgcaa aagcagcaat tctggcacgt 300
ctgaataccc tgagtctggg taatagcggg gttcatccga gcgttattaa tctgatgagc 360
gaactgatca acaaagatat cacaccgctg atttttgaac atggtggtgt tgggtgcaagc 420
ggtgatctgg ttacgctgag ccatctggca ctggttctga ttggtgaagg tgaagttttc 480
tataaagggtg aacgtcgtcc gacaccgaa gtttttgaaa ttgaaggtct gaaaccgatc 540
caggtgaaa ttccggaagg tctggccctg attaatggca ccagcgttat gaccggtatt 600
ggtgttgta atgtgtacca tgcaaaaaa ctgctggatt ggagcctgaa aagcagctgt 660
gcaattaatg aactggttca ggcatatgat gatcacttta gcgcagaact gaatcagacc 720
aaacgtcata aaggtcagca agaaattgca ctgaaaatgc gtcagaatct gagcagatagc 780
accctgattc gcaaacgtga agatcatctg tatagcgggtg aaaacaccga agaaatcttc 840
aaagaaaag tgcaagagta ttatagcctg cgttgtgttc cgcagattct gggcccggtt 900
ctggaacca ttaacaatgt tgcaagcatt ctggaagatg aatttaacag cgcaaacgat 960
aaccgatca tcgatgtaa aaaccagcat gtttatcacg gtggcaattt tcatggtgat 1020
tatatcagcc tggaaatgga taaactgaaa atcgtgatta ccaaactgac catgctggca 1080
gaacgtcagc tgaattatct gctgaatagc aaaattaacg aactgctgcc tccgtttggt 1140
aatctgggca ccctggggtt taactttggt atgcagggtg ttcagtttac cgcaaccagc 1200
accaccgag aaagccagat gctgagcaat ccgatgatg ttcatagcat tccgaacaat 1260
aatgataacc aggatattgt tagcatgggc accaatagcg cagttattac cagcaaagtt 1320
atcgaaaatg ctttgaagt tctggccatt gaaatgatta ccattgttca ggcgattgat 1380
tatctgggcc agaaagataa aatcagcagc gtttagcaaaa aatggtatga tgaaatccgc 1440
aacatcatcc cgaccttaa agaagatcag gtgatgtatc cgttcgtgca gaaagtaaaa 1500
gaccacctga ttaacaat 1518

<210> 52

<211> 59

5 <212> ADN
<213> Secuencia artificial

<220>

<223> Cebador CBJP483

10

<400> 52

catcttagta tattagttaa gtataagaag gagatataca tatgctggca atgagccct 59

<210> 53

15

<211> 41

ES 2 770 616 T3

	<212> ADN		
	<213> Secuencia artificial		
5	<220>		
	<223> Cebador CBJP484		
	<400> 53		
	tgcccgccg atatcaatt gattaaaccg gactctgtg c	41	
10	<210> 54		
	<211> 65		
	<212> ADN		
	<213> Secuencia artificial		
15	<220>		
	<223> Cebador CBJP555		
	<400> 54		
	catcttagta tattagttaa gtataagaag gagatataca tatgaacacc atcaacgaat	60	
20	atctg		65
	<210> 55		
	<211> 55		
25	<212> ADN		
	<213> Secuencia artificial		
	<220>		
	<223> Cebador CBJP556		
30	<400> 55		
	tgcccgccg atatcaatt gattaattgt taatcagggtg gtctttact ttctg	55	
	<210> 56		
	<211> 61		
35	<212> ADN		
	<213> Secuencia artificial		
	<220>		
	<223> Cebador CBJP745		
40	<400> 56		
	catcttagta tattagttaa gtataagaag gagatataca tatgctggat gcaaccattg	60	
	g		61

ES 2 770 616 T3

	<210> 57	
	<211> 41	
	<212> ADN	
	<213> Secuencia artificial	
5		
	<220>	
	<223> Cebador CBJP746	
10	<400> 57 tggccggccg atatccaatt gattatgccg gaggatccgc t	41
	<210> 58	
	<211> 33	
	<212> ADN	
15	<213> Secuencia artificial	
	<220>	
	<223> Cebador CBJP633	
20	<400> 58 agtcaggua aaacaatgga gttctcccg cca	33
	<210> 59	
	<211> 30	
25	<212> ADN	
	<213> Secuencia artificial	
	<220>	
	<223> Cebador CBJP634	
30	<400> 59 cgtgcgautc atagttcaca acgaaactg	30
	<210> 60	
35	<211> 34	
	<212> ADN	
	<213> Secuencia artificial	
	<220>	
40	<223> Cebador CBJP635	
	<400> 60 atctgtcaua aaacaatgga atttcacgt ccgc	34
45	<210> 61	
	<211> 32	

ES 2 770 616 T3

	<212> ADN <213> Secuencia artificial	
5	<220> <223> Cebador CBJP636	
	<400> 61 cacgcgautc acagttcaca acgaaatttg aa	32
10	<210> 62 <211> 25 <212> ADN <213> Secuencia artificial	
15	<220> <223> Cebador PTEF1_fw	
20	<400> 62 cacgcgaugc acacaccata gcttc	25
	<210> 63 <211> 25 <212> ADN <213> Secuencia artificial	
25	<220> <223> Cebador PTEF1_rv	
30	<400> 63 cgtgcgaugg aagtaccttc aaaga	25
	<210> 64 <211> 58 <212> ADN <213> Secuencia artificial	
35	<220> <223> Cebador CBJP517	
40	<400> 64 tagaaataat tttgtttaac ttaagaagg agatatacca tggccctgga taaaatgg	58
	<210> 65 <211> 49 <212> ADN <213> Secuencia artificial	
45	<212> ADN <213> Secuencia artificial	

ES 2 770 616 T3

<220>
<223> Cebador CBJP518

5 <400> 65
taagcattat gcggccgcaa gcttgcaca attccatgcg aaaaactag 49

<210> 66
<211> 56
<212> ADN
10 <213> Secuencia artificial

<220>
<223> Cebador CBJP533

15 <400> 66
tagaaataat tttgttaac ttaagaagg agatatacca tggaaatttc acgtcc 56

<210> 67
<211> 47
20 <212> ADN
<213> Secuencia artificial

<220>
<223> Cebador CBJP534

25 <400> 67
taagcattat gcggccgcaa gctgttaca gttcacaacg aaatttg 47

<210> 68
30 <211> 511
<212> PRT
<213> *Saccharomyces cerevisiae*

<400> 68
35

ES 2 770 616 T3

Met Pro Ala Pro His Gly Gly Ile Leu Gln Asp Leu Ile Ala Arg Asp
1 5 10 15

Ala Leu Lys Lys Asn Glu Leu Leu Ser Glu Ala Gln Ser Ser Asp Ile
20 25 30

Leu Val Trp Asn Leu Thr Pro Arg Gln Leu Cys Asp Ile Glu Leu Ile
35 40 45

Leu Asn Gly Gly Phe Ser Pro Leu Thr Gly Phe Leu Asn Glu Asn Asp
50 55 60

Tyr Ser Ser Val Val Thr Asp Ser Arg Leu Ala Asp Gly Thr Leu Trp
65 70 75 80

Thr Ile Pro Ile Thr Leu Asp Val Asp Glu Ala Phe Ala Asn Gln Ile
85 90 95

Lys Pro Asp Thr Arg Ile Ala Leu Phe Gln Asp Asp Glu Ile Pro Ile
100 105 110

Ala Ile Leu Thr Val Gln Asp Val Tyr Lys Pro Asn Lys Thr Ile Glu
115 120 125

Ala Glu Lys Val Phe Arg Gly Asp Pro Glu His Pro Ala Ile Ser Tyr
130 135 140

Leu Phe Asn Val Ala Gly Asp Tyr Tyr Val Gly Gly Ser Leu Glu Ala
145 150 155 160

Ile Gln Leu Pro Gln His Tyr Asp Tyr Pro Gly Leu Arg Lys Thr Pro
165 170 175

Ala Gln Leu Arg Leu Glu Phe Gln Ser Arg Gln Trp Asp Arg Val Val
180 185 190

Ala Phe Gln Thr Arg Asn Pro Met His Arg Ala His Arg Glu Leu Thr
195 200 205

Val Arg Ala Ala Arg Glu Ala Asn Ala Lys Val Leu Ile His Pro Val
210 215 220

Val Gly Leu Thr Lys Pro Gly Asp Ile Asp His His Thr Arg Val Arg
225 230 235 240

Val Tyr Gln Glu Ile Ile Lys Arg Tyr Pro Asn Gly Ile Ala Phe Leu

ES 2 770 616 T3

	245		250		255										
Ser	Leu	Leu	Pro 260	Leu	Ala	Met	Arg	Met 265	Ser	Gly	Asp	Arg	Glu	Ala	Val 270
Trp	His	Ala 275	Ile	Ile	Arg	Lys	Asn 280	Tyr	Gly	Ala	Ser	His 285	Phe	Ile	Val
Gly	Arg 290	Asp	His	Ala	Gly	Pro 295	Gly	Lys	Asn	Ser	Lys 300	Gly	Val	Asp	Phe
Tyr 305	Gly	Pro	Tyr	Asp	Ala 310	Gln	Glu	Leu	Val	Glu 315	Ser	Tyr	Lys	His	Glu 320
Leu	Asp	Ile	Glu	Val 325	Val	Pro	Phe	Arg	Met 330	Val	Thr	Tyr	Leu	Pro 335	Asp
Glu	Asp	Arg	Tyr 340	Ala	Pro	Ile	Asp	Gln 345	Ile	Asp	Thr	Thr	Lys 350	Thr	Arg
Thr	Leu	Asn 355	Ile	Ser	Gly	Thr	Glu 360	Leu	Arg	Arg	Arg	Leu	Arg 365	Val	Gly
Gly	Glu 370	Ile	Pro	Glu	Trp	Phe 375	Ser	Tyr	Pro	Glu	Val 380	Val	Lys	Ile	Leu
Arg 385	Glu	Ser	Asn	Pro	Pro 390	Arg	Pro	Lys	Gln	Gly 395	Phe	Ser	Ile	Val	Leu 400
Gly	Asn	Ser	Leu 405	Thr	Val	Ser	Arg	Glu	Gln 410	Leu	Ser	Ile	Ala	Leu	Leu 415
Ser	Thr	Phe	Leu 420	Gln	Phe	Gly	Gly	Gly 425	Arg	Tyr	Tyr	Lys	Ile	Phe	Glu 430
His	Asn 435	Asn	Lys	Thr	Glu	Leu	Leu	Ser 440	Leu	Ile	Gln	Asp 445	Phe	Ile	Gly
Ser	Gly 450	Ser	Gly	Leu	Ile	Ile	Pro	Asn 455	Gln	Trp	Glu 460	Asp	Asp	Lys	Asp
Ser 465	Val	Val	Gly	Lys	Gln 470	Asn	Val	Tyr	Leu	Leu 475	Asp	Thr	Ser	Ser	Ser 480
Ala	Asp	Ile	Gln	Leu 485	Glu	Ser	Ala	Asp	Glu 490	Pro	Ile	Ser	His	Ile	Val 495
Gln	Lys	Val	Val	Leu 500	Phe	Leu	Glu	Asp 505	Asn	Gly	Phe	Phe	Val	Phe	

ES 2 770 616 T3

<210> 69

<211> 202

<212> PRT

5 <213> *Saccharomyces cerevisiae*

<400> 69

```

Met Ala Thr Asn Ile Thr Trp His Pro Asn Leu Thr Tyr Asp Glu Arg
1           5           10           15

Lys Ala Leu Arg Lys Gln Asp Gly Cys Thr Ile Trp Leu Thr Gly Leu
          20           25           30

Ser Ala Ser Gly Lys Ser Thr Ile Ala Cys Ala Leu Glu Gln Leu Leu
          35           40           45

Leu Gln Lys Asn Leu Ser Ala Tyr Arg Leu Asp Gly Asp Asn Ile Arg
          50           55           60

Phe Gly Leu Asn Lys Asp Leu Gly Phe Ser Glu Lys Asp Arg Asn Glu
65           70           75           80

Asn Ile Arg Arg Ile Ser Glu Val Ser Lys Leu Phe Ala Asp Ser Cys
          85           90           95

Ala Ile Ser Ile Thr Ser Phe Ile Ser Pro Tyr Arg Val Asp Arg Asp
          100          105          110

Arg Ala Arg Glu Leu His Lys Glu Ala Gly Leu Lys Phe Ile Glu Ile
          115          120          125

Phe Val Asp Val Pro Leu Glu Val Ala Glu Gln Arg Asp Pro Lys Gly
          130          135          140

Leu Tyr Lys Lys Ala Arg Glu Gly Val Ile Lys Glu Phe Thr Gly Ile
145           150           155           160

Ser Ala Pro Tyr Glu Ala Pro Lys Ala Pro Glu Leu His Leu Arg Thr
          165          170          175

Asp Gln Lys Thr Val Glu Glu Cys Ala Thr Ile Ile Tyr Glu Tyr Leu
          180          185          190

Ile Ser Glu Lys Ile Ile Arg Lys His Leu
          195          200

```

10

<210> 70

ES 2 770 616 T3

<211> 357

<212> PRT

<213> *Saccharomyces cerevisiae*

5 <400> 70

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

ES 2 770 616 T3

Leu Lys Asp Thr Lys Asp Met Ile Thr Leu Glu Gly Val Glu Lys Gly
 225 230 235 240

His Ser Ser His Asp Glu Gln Thr Ala Ile Lys Asn Lys Leu Asn Ile
 245 250 255

Ser Lys Ser Leu His Leu Asp Ser Gln Ala Lys Tyr Cys Leu Leu Ala
 260 265 270

Leu Gly Leu Ala Asp Val Tyr Leu Arg Leu Pro Ile Lys Leu Ser Tyr
 275 280 285

Gln Glu Lys Ile Trp Asp His Ala Ala Gly Asn Val Ile Val His Glu
 290 295 300

Ala Gly Gly Ile His Thr Asp Ala Met Glu Asp Val Pro Leu Asp Phe
 305 310 315 320

Gly Asn Gly Arg Thr Leu Ala Thr Lys Gly Val Ile Ala Ser Ser Gly
 325 330 335

Pro Arg Glu Leu His Asp Leu Val Val Ser Thr Ser Cys Asp Val Ile
 340 345 350

Gln Ser Arg Asn Ala
 355

<210> 71

<211> 581

5 <212> PRT

<213> *Rattus norvegicus*

<400> 71

Met Glu Ile Pro Gly Ser Leu Cys Lys Lys Val Lys Leu Ser Asn Asn
 1 5 10 15

Ala Gln Asn Trp Gly Met Gln Arg Ala Thr Asn Val Thr Tyr Gln Ala
 20 25 30

His His Val Ser Arg Asn Lys Arg Gly Gln Val Val Gly Thr Arg Gly
 35 40 45

Gly Phe Arg Gly Cys Thr Val Trp Leu Thr Gly Leu Ser Gly Ala Gly
 50 55 60

Lys Thr Thr Val Ser Met Ala Leu Glu Glu Tyr Leu Val Cys His Gly
 65 70 75 80

10

ES 2 770 616 T3

Ile Pro Cys Tyr Thr Leu Asp Gly Asp Asn Ile Arg Gln Gly Leu Asn
85 90 95

Lys Asn Leu Gly Phe Ser Pro Glu Asp Arg Glu Glu Asn Val Arg Arg
100 105 110

Ile Ala Glu Val Ala Lys Leu Phe Ala Asp Ala Gly Leu Val Cys Ile
115 120 125

Thr Ser Phe Ile Ser Pro Tyr Thr Gln Val Arg Gln Gly Phe Thr Gly
130 135 140

Ile Asp Ser Glu Tyr Glu Lys Pro Glu Ala Pro Glu Leu Val Leu Lys
145 150 155 160

Thr Asp Ser Cys Asp Val Asn Asp Cys Val Gln Gln Val Val Glu Leu
165 170 175

Leu Gln Glu Arg Asp Ile Val Pro Val Asp Ala Ser Tyr Glu Val Lys
180 185 190

Glu Leu Tyr Val Pro Glu Asn Lys Leu His Leu Ala Lys Thr Asp Ala
195 200 205

Glu Ala Leu Pro Ala Leu Lys Ile Asn Lys Val Asp Met Gln Trp Val
210 215 220

Gln Val Leu Ala Glu Gly Trp Ala Thr Pro Leu Asn Gly Phe Met Arg
225 230 235 240

Glu Arg Glu Tyr Leu Gln Cys Leu His Phe Asp Cys Leu Leu Asp Gly
245 250 255

Gly Val Ile Asn Leu Ser Val Pro Ile Val Leu Thr Ala Thr Gln Glu
260 265 270

Asp Lys Glu Arg Leu Asp Gly Cys Thr Ala Phe Ala Leu Val Tyr Glu
275 280 285

Gly Arg Arg Val Ala Ile Leu Arg Asn Pro Glu Phe Phe Glu His Arg
290 295 300

Lys Glu Glu Arg Cys Ala Arg Gln Trp Gly Thr Thr Cys Arg Ser His
305 310 315 320

Pro Tyr Ile Lys Met Ile Leu Glu Gln Gly Asp Trp Leu Ile Gly Gly
325 330 335

ES 2 770 616 T3

Asp Leu Gln Val Leu Asp Arg Ile Tyr Trp Asn Asp Gly Leu Asp Gln
 340 345 350

Tyr Arg Leu Thr Pro Ala Glu Leu Lys Gln Lys Phe Lys Asp Met Asn
 355 360 365

Ala Asp Ala Val Phe Ala Phe Gln Leu Arg Asn Pro Val His Asn Gly
 370 375 380

His Ala Leu Leu Met Gln Asp Thr His Lys Gln Leu Leu Glu Arg Gly
 385 390 395 400

Tyr Arg Arg Pro Val Leu Leu Leu His Pro Leu Gly Gly Trp Thr Lys
 405 410 415

Asp Asp Asp Val Pro Leu Met Trp Arg Met Lys Gln His Ala Ala Val
 420 425 430

Leu Glu Glu Gly Ile Leu Asn Pro Glu Thr Thr Val Val Ala Ile Phe
 435 440 445

Pro Ser Pro Met Met Tyr Ala Gly Pro Thr Glu Val Gln Trp His Cys
 450 455 460

Arg Ala Arg Met Val Ala Gly Ala Asn Phe Tyr Ile Val Gly Arg Asp
 465 470 475 480

Pro Ala Gly Met Pro His Pro Glu Thr Gly Lys Asp Leu Tyr Glu Pro
 485 490 495

Thr His Gly Ala Lys Val Leu Thr Met Ala Pro Gly Leu Ile Thr Leu
 500 505 510

Glu Ile Val Pro Phe Arg Val Ala Ala Tyr Asn Lys Lys Lys Lys Arg
 515 520 525

Met Asp Tyr Tyr Asp Ser Asp His His Glu Asp Phe Glu Phe Ile Ser
 530 535 540

Gly Thr Arg Met Arg Lys Leu Ala Arg Glu Gly Gln Lys Pro Pro Glu
 545 550 555 560

Gly Phe Met Ala Pro Lys Ala Trp Thr Val Leu Val Glu Tyr Tyr Lys
 565 570 575

Ser Leu Glu Lys Ala
 580

<210> 72

<211> 614

5 <212> PRT

<213> *Rattus norvegicus*

ES 2 770 616 T3

<400> 72

Met Ser Glu Ile Lys Lys Gln Lys Thr Asp Gln Gln Lys Ser Thr Asn
 1 5 10 15

Val Val Tyr Gln Ala His His Val Ser Arg Asn Lys Arg Gly Gln Val
 20 25 30

Val Gly Thr Arg Gly Gly Phe Arg Gly Cys Thr Val Trp Leu Thr Gly
 35 40 45

Leu Ser Gly Ala Gly Lys Thr Thr Ile Ser Phe Ala Leu Glu Glu Tyr
 50 55 60

Leu Val Ser His Ala Ile Pro Cys Tyr Ser Leu Asp Gly Asp Asn Val
 65 70 75 80

Arg His Gly Leu Asn Lys Asn Leu Gly Phe Ser Ala Gly Asp Arg Glu
 85 90 95

Glu Asn Ile Arg Arg Ile Ala Glu Val Ala Lys Leu Phe Ala Asp Ala
 100 105 110

Gly Leu Val Cys Ile Thr Ser Phe Ile Ser Pro Phe Ala Lys Asp Arg
 115 120 125

Glu Asn Ala Arg Lys Ile His Glu Ser Ala Gly Leu Pro Phe Phe Glu
 130 135 140

Ile Phe Val Asp Ala Pro Leu Asn Ile Cys Glu Ser Arg Asp Val Lys
 145 150 155 160

Gly Leu Tyr Lys Arg Ala Arg Ala Gly Glu Ile Lys Gly Phe Thr Gly
 165 170 175

Ile Asp Ser Asn Tyr Glu Lys Pro Glu Thr Pro Glu Cys Val Leu Lys
 180 185 190

Thr Asn Leu Ser Ser Val Ser Asp Cys Val Gln Gln Val Val Glu Leu
 195 200 205

Leu Gln Glu Gln Ser Ile Val Pro His Thr Thr Ile Lys Gly Ile His
 210 215 220

ES 2 770 616 T3

Glu Leu Phe Val Pro Glu Asn Lys Ile Asp Gln Ile Arg Ala Glu Leu
 225 230 235 240

Glu Thr Leu Pro Ser Leu Pro Ile Thr Lys Leu Asp Leu Gln Trp Val
 245 250 255

Gln Ile Leu Ser Glu Gly Trp Ala Thr Pro Leu Lys Gly Phe Met Arg
 260 265 270

Glu Lys Glu Tyr Leu Gln Thr Leu His Phe Asp Thr Leu Leu Asp Asp
 275 280 285

Gly Val Ile Asn Met Ser Ile Pro Ile Val Leu Pro Val Ser Gly Asp
 290 295 300

Asp Lys Ala Arg Leu Glu Gly Cys Ser Lys Phe Ala Leu Met Tyr Glu
 305 310 315 320

Gly Arg Arg Val Ala Leu Leu Gln Asp Pro Glu Phe Tyr Glu His Arg
 325 330 335

Lys Glu Glu Arg Cys Ser Arg Val Trp Gly Thr Ala Ser Ala Lys His
 340 345 350

Pro His Ile Lys Met Val Met Glu Gly Gly Asp Trp Leu Val Gly Gly
 355 360 365

Asp Leu Gln Val Leu Glu Arg Ile Arg Trp Asn Asp Gly Leu Asp Gln
 370 375 380

Tyr Arg Leu Thr Pro Leu Glu Leu Lys Gln Lys Cys Lys Asp Met Asp
 385 390 395 400

Ala Asp Ala Val Phe Ala Phe Gln Leu Arg Asn Pro Val His Asn Gly
 405 410 415

His Ala Leu Leu Met Gln Asp Thr Arg Arg Arg Leu Leu Glu Arg Gly
 420 425 430

Tyr Lys His Pro Val Leu Leu Leu His Pro Leu Gly Gly Trp Thr Lys
 435 440 445

Asp Asp Asp Val Pro Leu Asp Trp Arg Met Lys Gln His Ala Ala Val
 450 455 460

Leu Glu Glu Gly Ile Leu Asp Pro Lys Ser Thr Ile Val Ala Ile Phe

ES 2 770 616 T3

Arg Lys Cys Gly Asn Met Gly Thr Ser Lys Thr Cys Pro His Ser Pro
 325 330 335

Arg Asp His Ile His Leu Ser Gly Thr Lys Val Arg Glu Leu Leu Arg
 340 345 350

Gln Gly Lys Lys Pro Pro Lys Glu Phe Ser Arg Pro Glu Val Ala Ala
 355 360 365

Val Leu Ile Lys Gly Leu His Gln Gln Pro Val Ala Ile Lys Gln Asn
 370 375 380

Ser Gly Glu Leu Gln
 385

<210> 74

<211> 199

5 <212> PRT

<213> Bacillus subtilis

<400> 74

Met Thr His Asn Pro Asn Ile Ile Trp His Pro Ala Ala Ile Ser Lys
 1 5 10 15

Ser Asp Arg Gln Ser Leu Asn Gly His Lys Ser Cys Val Leu Trp Phe
 20 25 30

Thr Gly Leu Ser Gly Ser Gly Lys Ser Val Leu Ala Asn Ala Val Asp
 35 40 45

Glu Lys Leu Tyr Arg Lys Gly Ile Gln Ser Tyr Val Leu Asp Gly Asp
 50 55 60

Asn Ile Arg His Gly Leu Asn Lys Asp Leu Gly Phe Gln Thr Gly Asp
 65 70 75 80

Arg Ile Glu Asn Ile Arg Arg Ile Gly Glu Val Ala Lys Leu Phe Val
 85 90 95

Asp Ser Gly Gln Met Ile Leu Thr Ala Phe Ile Ser Pro Phe Arg Glu
 100 105 110

Asp Arg Asp Met Val Arg Ala Leu Phe Pro Lys Gly Glu Phe Phe Glu
 115 120 125

Ile Tyr Val Lys Cys Pro Leu His Val Cys Glu Gln Arg Asp Pro Lys
 130 135 140

ES 2 770 616 T3

Gly Leu Tyr Lys Lys Ala Arg Asn Gly Glu Ile Lys His Phe Thr Gly
145 150 155 160

Ile Asp Ser Pro Tyr Glu Ala Pro Leu Ser Pro Asp Phe Ile Ile Glu
165 170 175

Ser Asp Gln Thr Ser Ile Ser Asp Gly Ala Asp Leu Ile Ile Asn Ala
180 185 190

Leu Gln Asn Arg Gly Ile Ile
195

<210> 75

5 <211> 313

<212> PRT

<213> Bacillus subtilis

10 <400> 75

ES 2 770 616 T3

Leu Phe Pro Asn Thr Thr Glu Lys Thr Leu Lys Tyr Ala Gly Glu Leu
 165 170 175

Ile Gln Tyr Pro Phe Ser Ser Ser Glu Leu Phe Asn Gln Leu Tyr Glu
 180 185 190

Thr Lys Leu Asn Val Val Lys Leu Asn Gly Phe Ile Phe Gln Asn Val
 195 200 205

Ser Leu Ser Glu Asn Gly Ala Ala Ser Val Phe Ile Lys Lys Asp Thr
 210 215 220

Leu Glu Lys Phe Gly Thr Thr Ala Ser Glu Ala Ser Gln Leu Val Gly
 225 230 235 240

Thr Leu Gly Asn Ile Ser Gly Ile Arg Ala Trp Val Phe Phe Val Glu
 245 250 255

Glu Asp Asp Gln Ile Arg Val Arg Phe Arg Ser Lys Gly Pro Val Ile
 260 265 270

Asn Gly Leu Ala Arg Lys Tyr Asn Gly Gly Gly His Pro Leu Ala Ser
 275 280 285

Gly Ala Ser Ile Tyr Ser Trp Asp Glu Ala Asp Arg Ile Leu Ala Asp
 290 295 300

Leu Glu Thr Leu Cys Lys Glu His Glu
 305 310

REIVINDICACIONES

1. Un proceso para la producción de un compuesto fenólico sulfatado, que comprende:

- 5 (i') poner en contacto un medio que comprende un compuesto fenólico con una primera célula hospedadora recombinante; o
 (i'') poner en contacto un medio que comprende un sustrato de carbono fermentable con una primera célula hospedadora recombinante; o
 10 (i''') poner en contacto un medio que comprende un precursor de un compuesto fenólico con una primera célula hospedadora recombinante;

en donde la primera célula hospedadora recombinante comprende un polipéptido heterólogo que tiene una actividad aril sulfotransferasa, y

- 15 en donde la primera célula hospedadora recombinante ha sido modificada adicionalmente para tener una expresión proteica aumentada de (i) una ATP sulfurilasa, (ii) una APS quinasa y/o (iii) una PAP fosfatasa en comparación con una célula hospedadora idéntica que no lleva dicha modificación.

2. El proceso según la reivindicación 1, que comprende adicionalmente:

- 20 (ii) cultivar la primera célula hospedadora recombinante en condiciones adecuadas para la producción del compuesto fenólico sulfatado correspondiente.

3. El proceso según las reivindicaciones 1 o 2, en donde el polipéptido heterólogo que tiene una actividad aril sulfotransferasa es un polipéptido que comprende una secuencia de aminoácidos que tiene al menos el 70 % de identidad de secuencia con la secuencia de aminoácidos establecida en SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 o 13, y tiene actividad aril sulfotransferasa.

- 25

4. El proceso según cualquiera de las reivindicaciones 1-3, en donde dicha primera célula hospedadora recombinante comprende además un polipéptido heterólogo que tiene una actividad tirosina amoniaco liasa o un polipéptido heterólogo que tiene actividad fenilalanina amoniaco liasa.

- 30

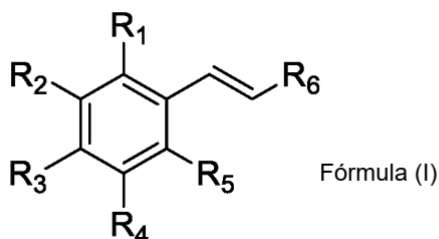
5. El proceso según una cualquiera de las reivindicaciones 1-4, en donde en la etapa (i'), (i'') o (i''') el medio se pone en contacto adicionalmente con una segunda célula hospedadora recombinante que comprende un polipéptido heterólogo que tiene una actividad tirosina amoniaco liasa o una tercera célula hospedadora recombinante que comprende un polipéptido heterólogo que tiene actividad fenilalanina amoniaco liasa.

- 35

6. El proceso según las reivindicaciones 4 o 5, en donde el polipéptido heterólogo que tiene una actividad tirosina amoniaco liasa es un polipéptido que comprende una secuencia de aminoácidos que tiene al menos el 70 % de identidad de secuencia con la secuencia de aminoácidos establecida en SEQ ID NO: 14, 15, 16, 17, 18, 19, 20, 21, 22 o 23, y tiene actividad tirosina amoniaco liasa.

- 40

7. El proceso según una cualquiera de las reivindicaciones 1-6, en donde el compuesto fenólico está representado por la fórmula general (I):



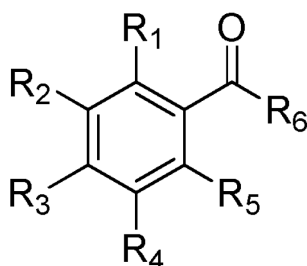
45 en donde al menos uno de R₁, R₂, R₃, R₄ y R₅ es un grupo hidroxilo (-OH);

en donde R₁, R₂, R₃, R₄, R₅ y R₆ se seleccionan independientemente del grupo que consiste en haluro, hidrógeno, hidroxilo (-OH), -OR₇, -OCOR₇, -NR₇R₈, -COR₇, -COOR₇, -SR₇, -OSO₃R₇, -OCSR₇, -POR₇R₈, alquilo, alquenilo, alquinilo, arilo y heteroarilo; en donde R₇, y R₈ se seleccionan independientemente del grupo que consiste en hidrógeno, alquilo, alquenilo, alquinilo, arilo y heteroarilo;

- 50

en donde R₁, R₂, R₃, R₄, R₅ y R₆, están unidos opcionalmente a un elemento puente Y_n, formando así uno o más anillos, siendo Y_n un enlace o un alquilo C₁₋₁₂ o un arilo, una estructura carbocíclica, heterocíclica o heteroaromática que tiene 1-3 anillos, 3-8 miembros anulares en cada uno y 0 a 4 heteroátomos, o un heteroalquilo que comprende 1 a 12 heteroátomos seleccionados del grupo que consiste en N, O, S, S(O)₁₋₂ y carbonilo, y en donde n es un número

- 55 entero entre 1 y 12; o en donde el compuesto fenólico está representado por la fórmula general (II):



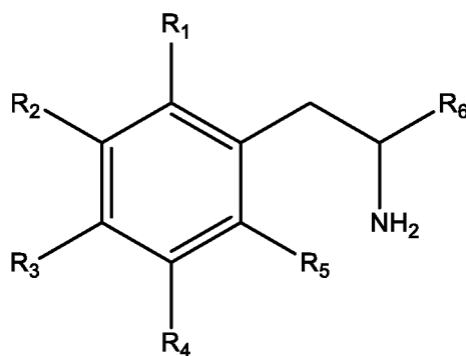
Fórmula (II)

en donde al menos uno de R_1 , R_2 , R_3 , R_4 y R_5 es un grupo hidroxilo (-OH);

5 en donde R_1 , R_2 , R_3 , R_4 , R_5 y R_6 se seleccionan independientemente del grupo que consiste en haluro, hidrógeno, hidroxilo (-OH), $-OR_7$, $-OCOR_7$, $-NR_7R_8$, $-COR_7$, $-COOR_7$, $-SR_7$, $-OSO_3R_7$, $-OCSR_7$, $-POR_7R_8$, alquilo, alquenilo, alquinilo, arilo y heteroarilo; en donde R_7 y R_8 se seleccionan independientemente del grupo que consiste en hidrógeno, alquilo, alquenilo, alquinilo, arilo y heteroarilo;

10 en donde R_1 , R_2 , R_3 , R_4 , R_5 y R_6 , están unidos opcionalmente a un elemento puente Y_n , formando así uno o más anillos, siendo Y_n un enlace o un alquilo C_{1-12} o un arilo, una estructura carbocíclica, heterocíclica o heteroaromática que tiene 1-3 anillos, 3-8 miembros anulares en cada uno y 0 a 4 heteroátomos, o un heteroalquilo que comprende 1 a 12 heteroátomos seleccionados del grupo que consiste en N, O, S, $S(O)_{1-2}$ y carbonilo, y en donde n es un número entero entre 1 y 12.

15 8. El proceso según una cualquiera de las reivindicaciones 1-6, en donde el precursor de un compuesto fenólico en la etapa (i''') es un compuesto de la fórmula general (p-I):



Fórmula (p-I);

en donde al menos uno de R_1 , R_2 , R_3 , R_4 y R_5 es un grupo hidroxilo (-OH);

20 en donde R_1 , R_2 , R_3 , R_4 , R_5 y R_6 se seleccionan independientemente del grupo que consiste en haluro, hidrógeno, hidroxilo (-OH), $-OR_7$, $-OCOR_7$, $-NR_7R_8$, $-COR_7$, $-COOR_7$, $-SR_7$, $-OSO_3R_7$, $-OCSR_7$, $-POR_7R_8$, alquilo, alquenilo, alquinilo, arilo y heteroarilo; en donde R_7 y R_8 se seleccionan independientemente del grupo que consiste en hidrógeno, alquilo, alquenilo, alquinilo, arilo y heteroarilo;

25 en donde R_1 , R_2 , R_3 , R_4 , R_5 y R_6 , están unidos opcionalmente a un elemento puente Y_n , formando así uno o más anillos, siendo Y_n un enlace o un alquilo C_{1-12} o un arilo, una estructura carbocíclica, heterocíclica o heteroaromática que tiene 1-3 anillos, 3-8 miembros anulares en cada uno y 0 a 4 heteroátomos, o un heteroalquilo que comprende 1 a 12 heteroátomos seleccionados del grupo que consiste en N, O, S, $S(O)_{1-2}$ y carbonilo, y en donde n es un número entero entre 1 y 12.

30 9. El proceso según las reivindicaciones 7 u 8, en donde cada uno de R_1 , R_2 , R_4 y R_5 es hidrógeno, R_3 es hidroxilo (-OH) y R_6 es $-COOH$.

35 10. Una célula hospedadora recombinante que comprende un primer polipéptido heterólogo que tiene actividad aril sulfotransferasa y ha sido modificado adicionalmente para tener una expresión proteica aumentada de (i) una ATP sulfurilasa, (ii) una APS quinasa y/o (iii) una PAP fosfatasa en comparación con una célula hospedadora idéntica que no lleva dicha modificación.

40 11. La célula hospedadora recombinante según la reivindicación 10, en donde el polipéptido heterólogo que tiene actividad aril sulfotransferasa es un polipéptido que comprende una secuencia de aminoácidos que tiene al menos el 70 % de identidad de secuencia con la secuencia de aminoácidos establecida en SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9,

10, 11, 12 o 13, y tiene actividad aril sulfotransferasa.

5 12. La célula hospedadora recombinante según las reivindicaciones 10 u 11, que comprende además un segundo polipéptido heterólogo que tiene actividad tirosina amoniaco liasa y/o un tercer polipéptido heterólogo que tiene actividad fenilalanina amoniaco liasa.

10 13. La célula hospedadora recombinante según la reivindicación 12, en donde el heterólogo que tiene actividad tirosina amoniaco liasa es un polipéptido que comprende una secuencia de aminoácidos que tiene al menos el 70 % de identidad de secuencia con la secuencia de aminoácidos establecida en SEQ ID NO: 14, 15, 16, 17, 18, 19, 20, 21, 22 o 23, y tiene actividad tirosina amoniaco liasa.

15 14. Una composición que comprende una primera célula hospedadora recombinante que comprende un polipéptido heterólogo que tiene actividad aril sulfotransferasa y ha sido modificado adicionalmente para tener una expresión proteica aumentada de (i) una ATP sulfurilasa, (ii) una APS quinasa y/o (iii) una PAP fosfatasa en comparación con una célula hospedadora idéntica que no lleva dicha modificación; y una segunda célula hospedadora recombinante que comprende un polipéptido heterólogo que tiene actividad tirosina amoniaco liasa o una tercera célula hospedadora recombinante que comprende un polipéptido heterólogo que tiene actividad fenilalanina amoniaco liasa.

20 15. La composición según la reivindicación 14, en donde el polipéptido heterólogo que tiene una actividad aril sulfotransferasa es un polipéptido que comprende una secuencia de aminoácidos que tiene al menos aproximadamente 70 % de identidad de secuencia con la secuencia de aminoácidos establecida en SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 o 13, y tiene actividad aril sulfotransferasa; y/o en donde el polipéptido heterólogo que tiene actividad tirosina amoniaco liasa es un polipéptido que comprende una secuencia de aminoácidos que tiene al menos un 70 % de identidad de secuencia con la secuencia de aminoácidos establecida en SEQ ID NO: 14, 15, 16,
25 17, 18, 19, 20, 21, 22 o 23, y tiene actividad tirosina amoniaco liasa.

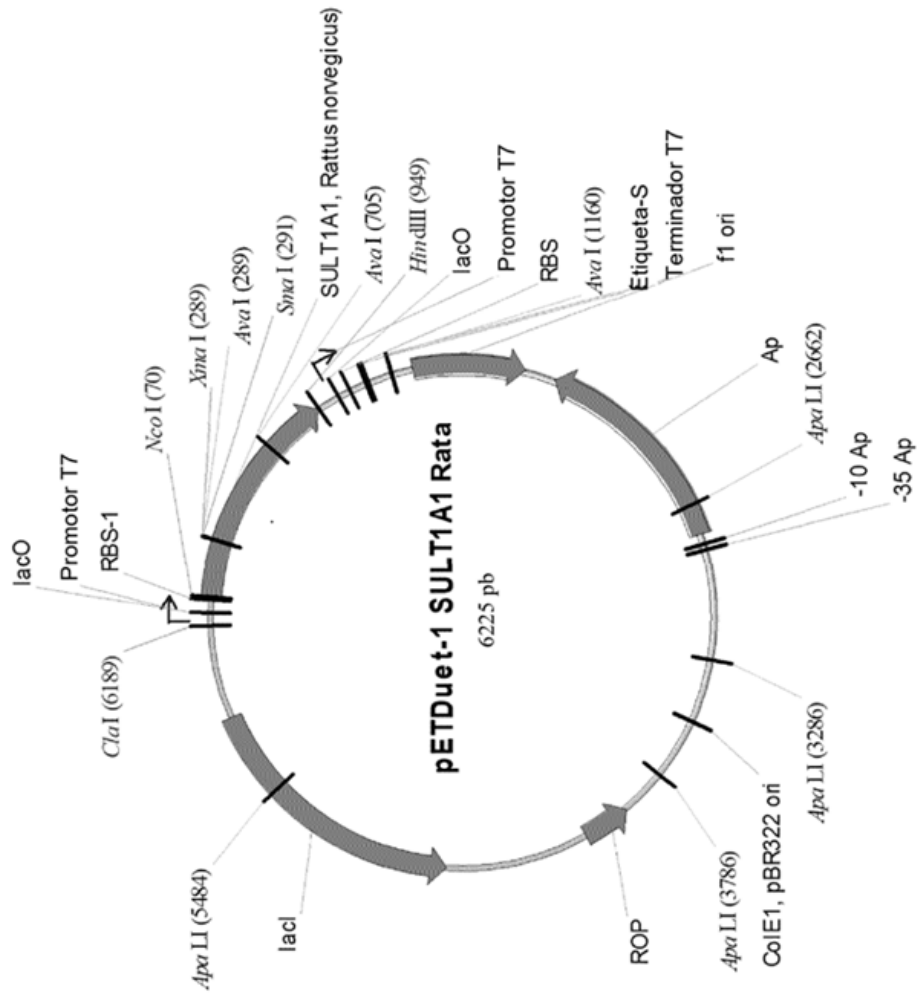


Figura 1

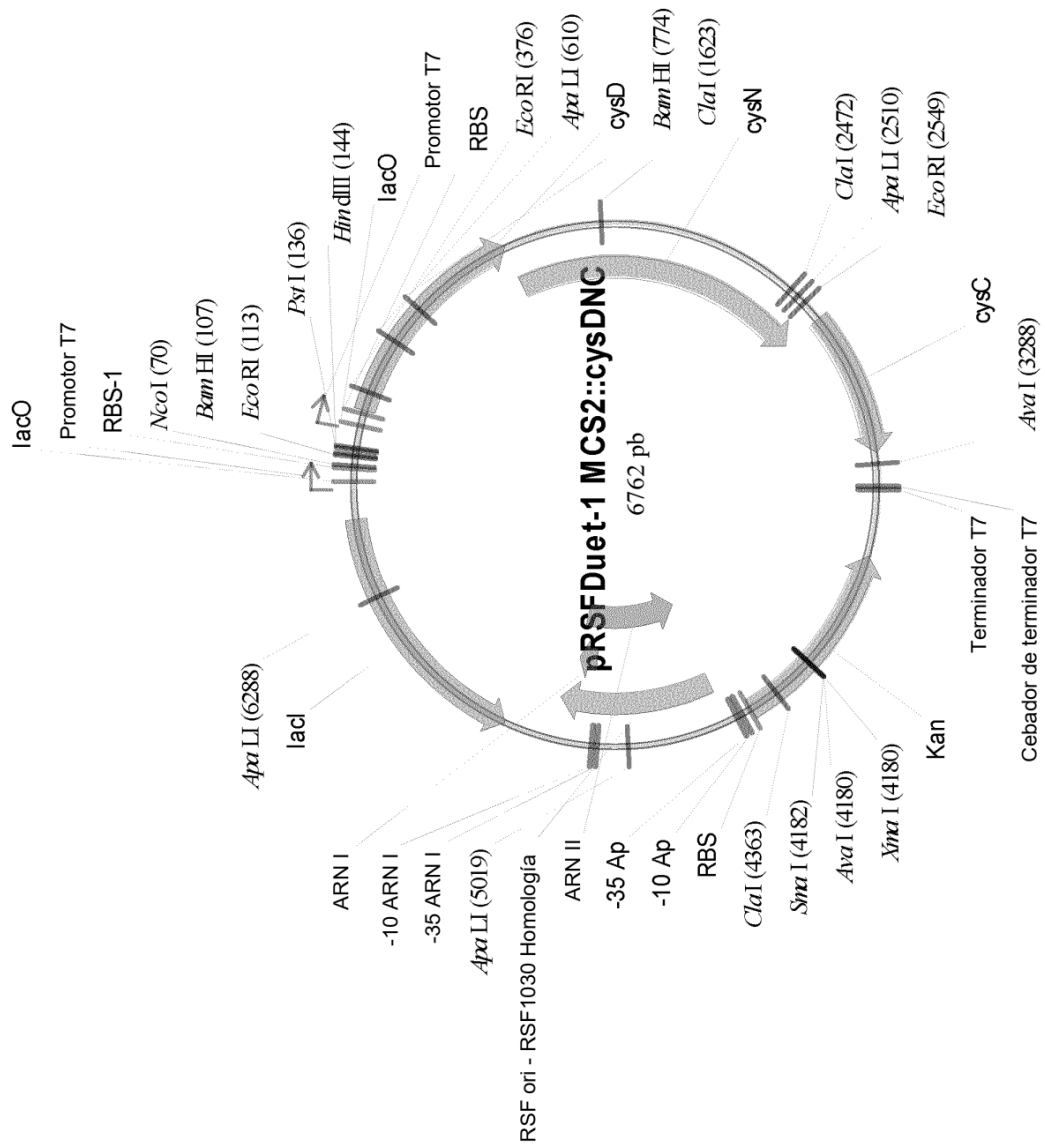


Figura 2

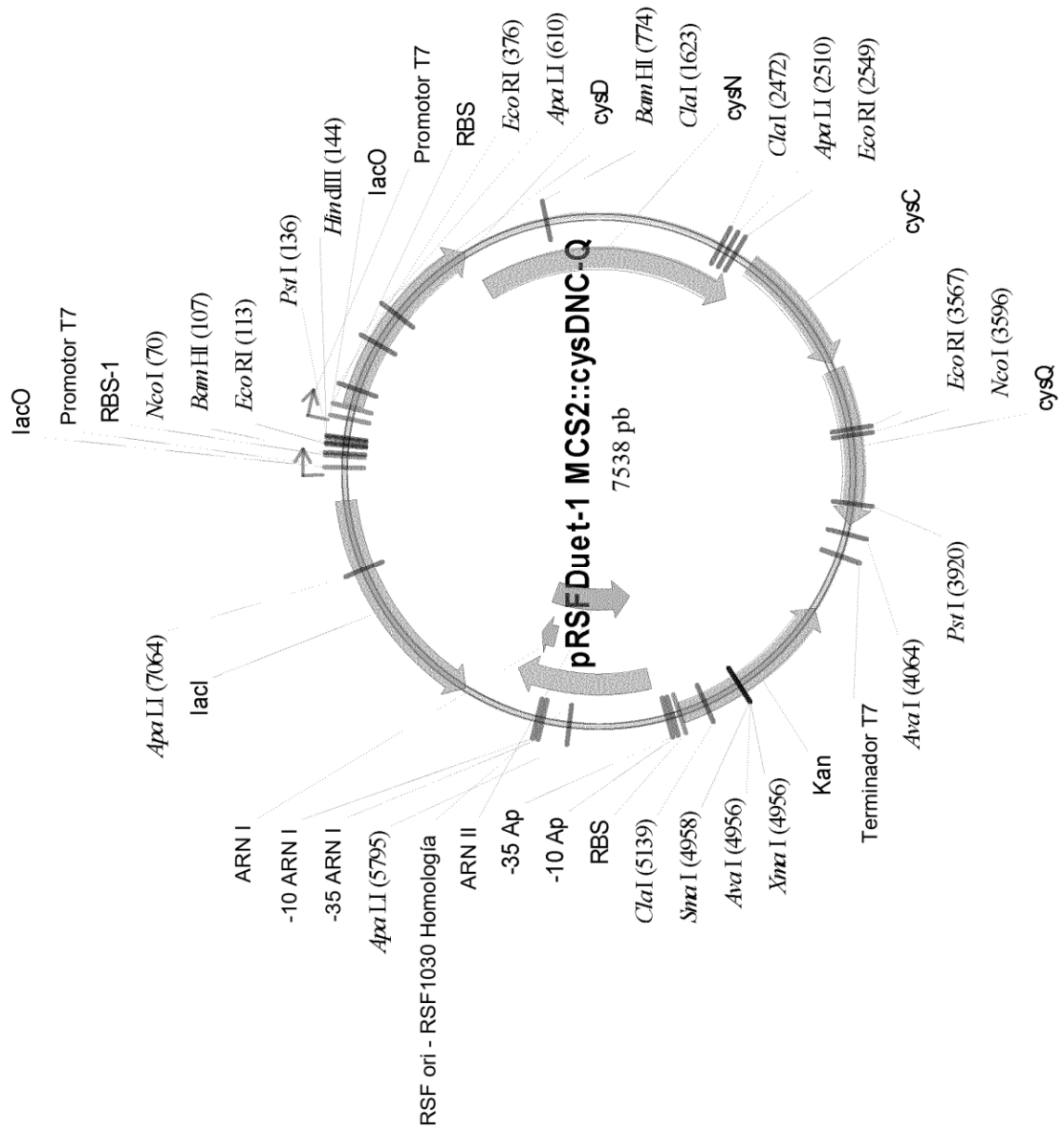


Figura 3



Figura 4

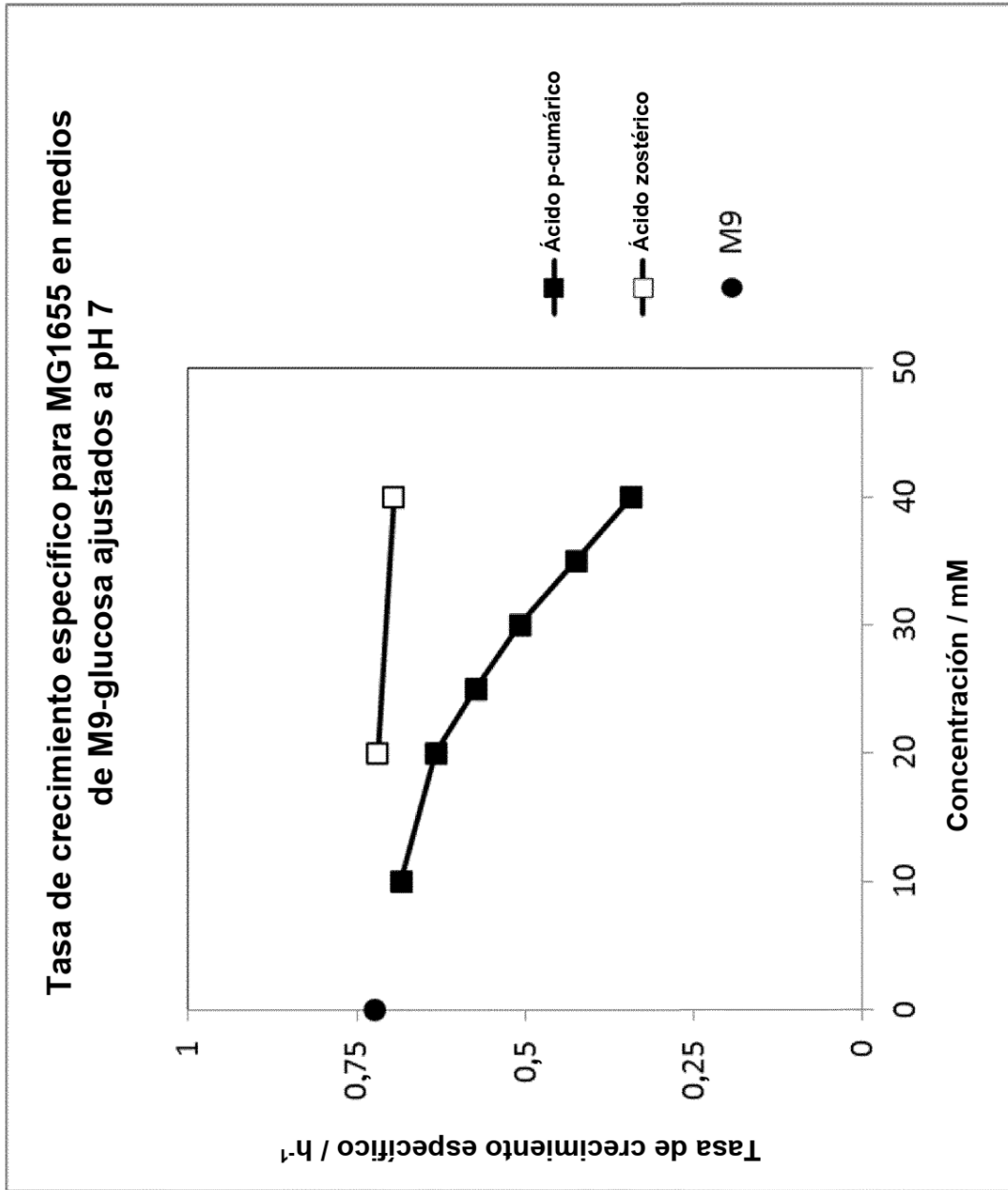


Figura 5

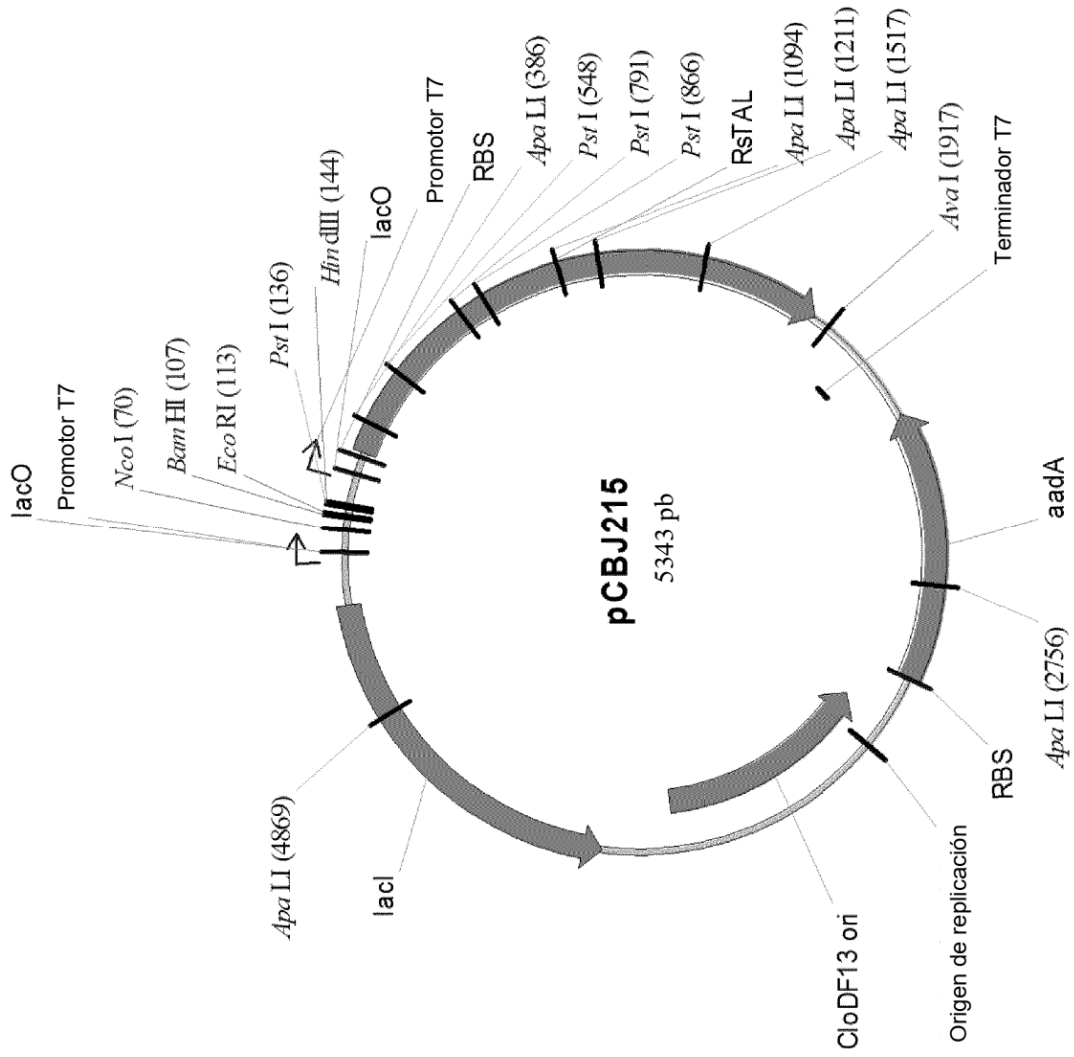


Figura 6

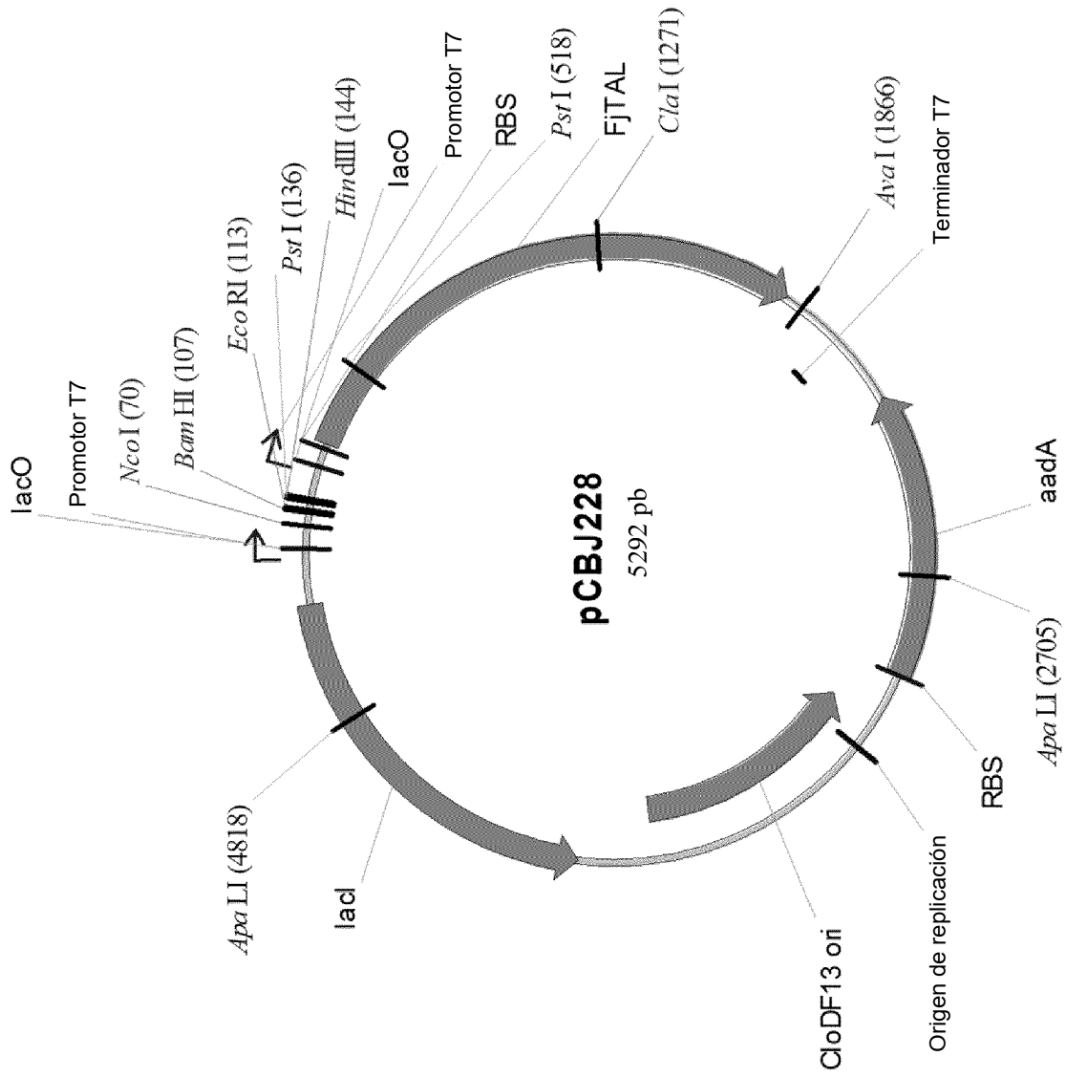


Figura 7

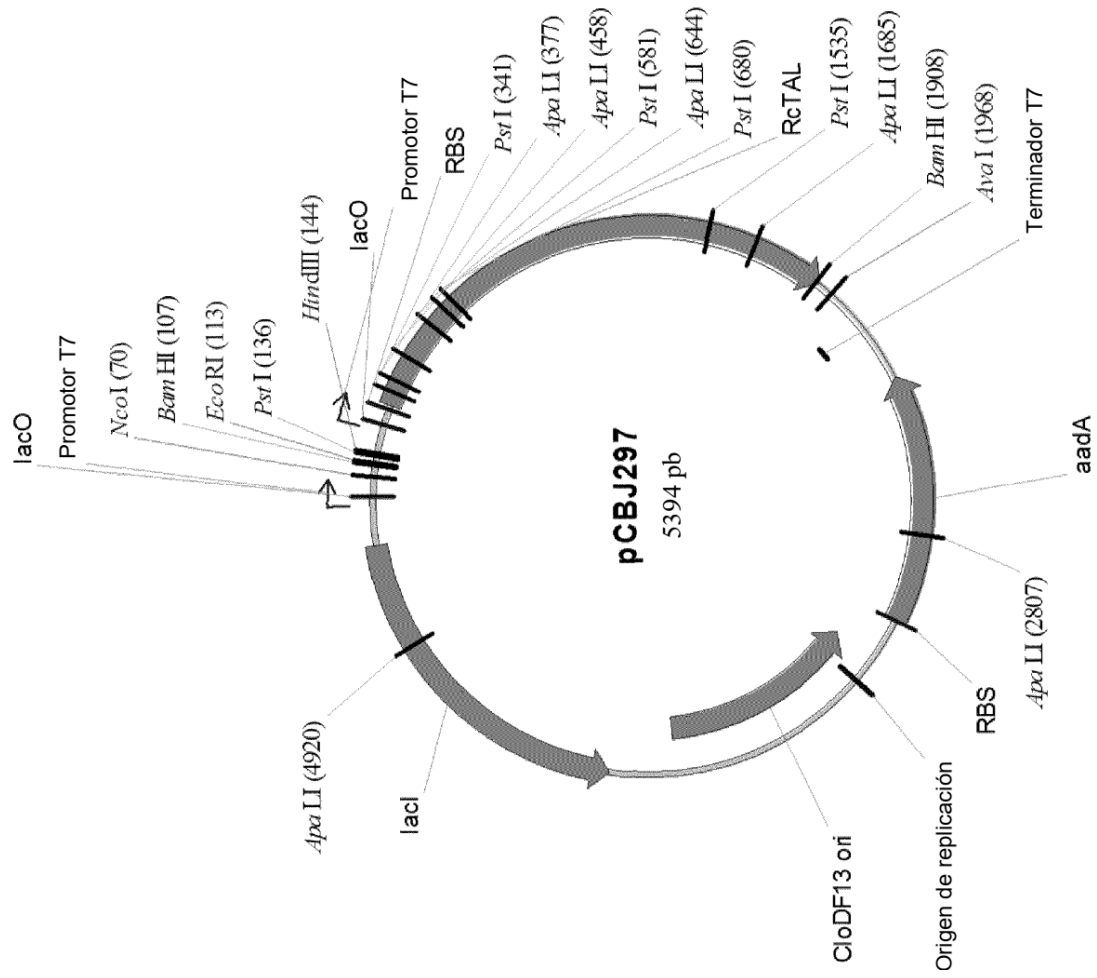


Figura 8

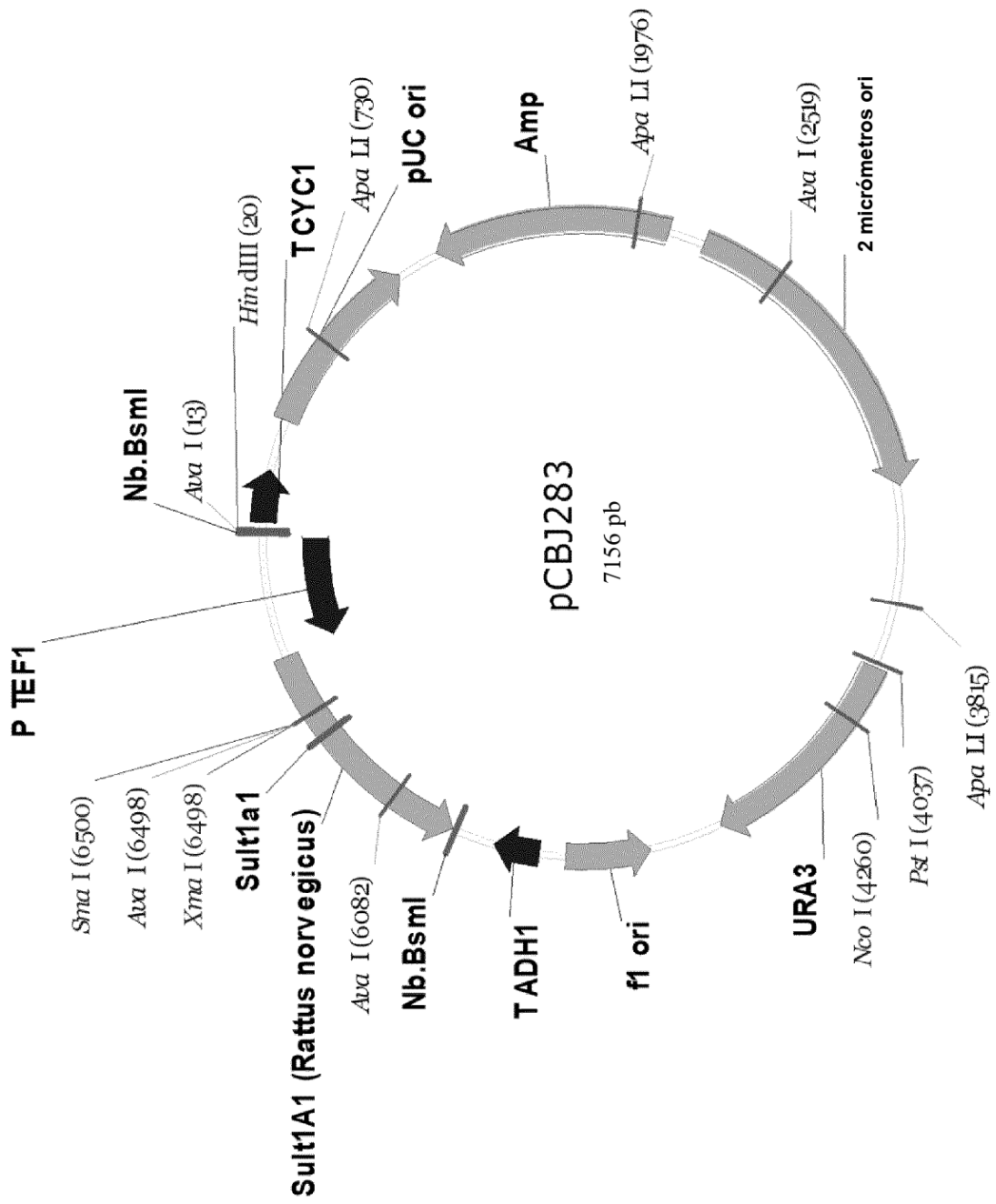


Figura 9

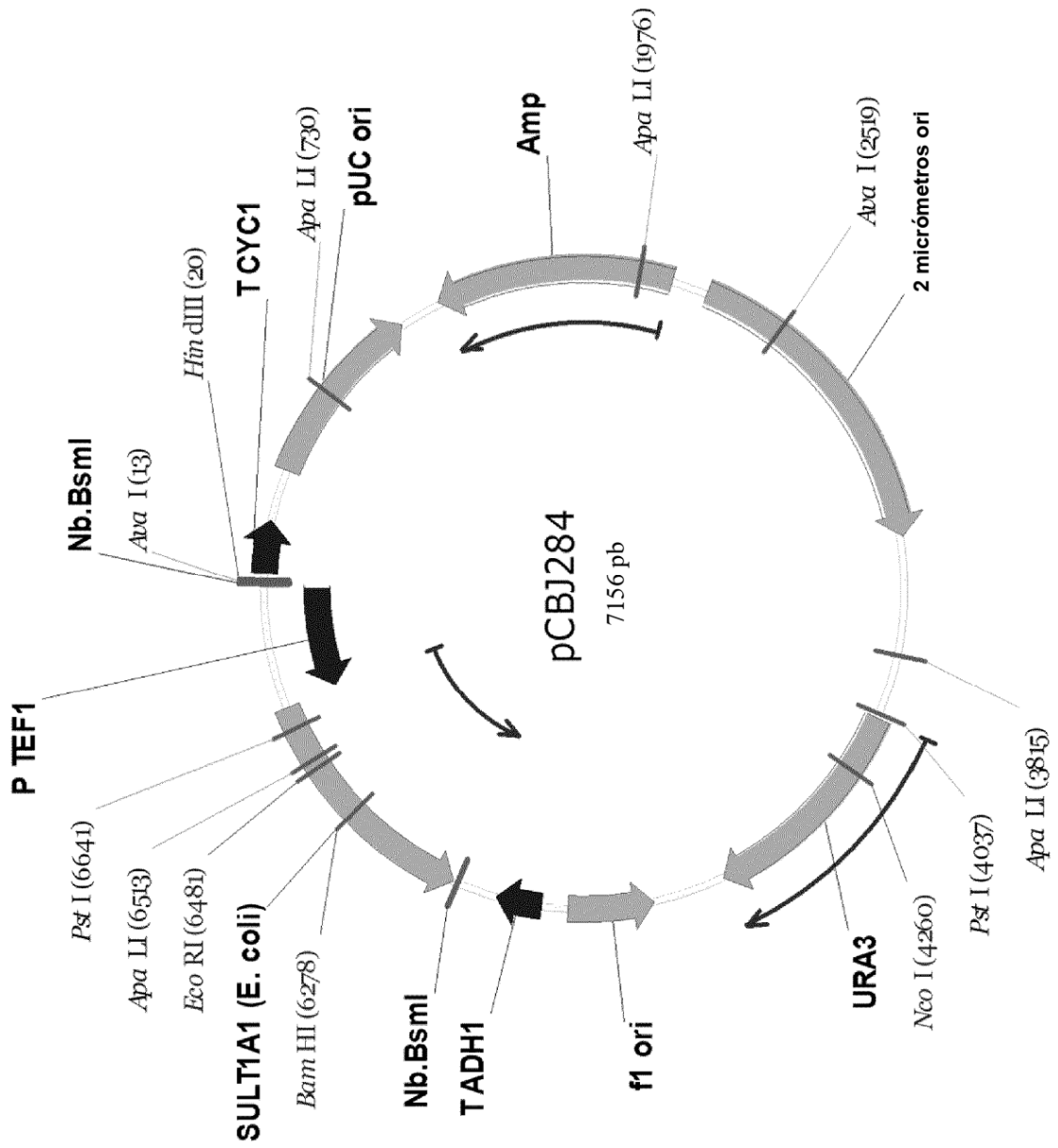


Figura 10