



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA

⑪ Número de publicación: **2 336 858**

⑤① Int. Cl.:
C12N 15/53 (2006.01)
C12N 15/55 (2006.01)
C12N 15/60 (2006.01)
C12P 7/18 (2006.01)
C12N 9/04 (2006.01)
C12N 9/16 (2006.01)
C12N 9/88 (2006.01)

⑫

TRADUCCIÓN DE PATENTE EUROPEA

T3

⑨⑥ Número de solicitud europea: **97947395 .6**
⑨⑥ Fecha de presentación : **10.11.1997**
⑨⑦ Número de publicación de la solicitud: **0942989**
⑨⑦ Fecha de publicación de la solicitud: **22.09.1999**

⑤④ Título: **Método para la producción de 1,3-propanodiol mediante organismos recombinantes.**

③⑩ Prioridad: **13.11.1996 US 30601 P**

④⑤ Fecha de publicación de la mención BOPI:
16.04.2010

④⑤ Fecha de la publicación del folleto de la patente:
16.04.2010

⑦③ Titular/es:
E.I. DU PONT DE NEMOURS AND COMPANY
1007 Market Street
Wilmington, Delaware 19898, US
GENENCOR INTERNATIONAL Inc.

⑦② Inventor/es: **Gatenby, Anthony, Arthur;**
Haynie, Sharon, Loretta;
Nagarajan, Vasantha;
Nair, Ramesh, V.;
Nakamura, Charles, E.;
Payne, Mark, Scott;
Picataggio, Stephen, Kenneth;
Diaz-Torres, Maria;
Hsu, Amy, Kuang-Hua;
Lareau, Richard, D.;
Trimbur, Donald, E. y
Whited, Gregory, M.

⑦④ Agente: **Elzaburu Márquez, Alberto**

ES 2 336 858 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

Método para la producción de 1,3-propanodiol mediante organismos recombinantes.

5 **Campo de la invención**

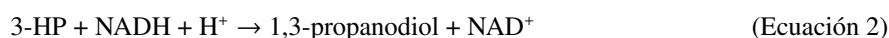
La presente invención se refiere al campo de la biología molecular y al uso de organismos recombinantes para la producción de compuestos deseados. Más específicamente, describe la expresión de genes clonados de glicerol-3-fosfato deshidrogenasa (G3PDH) y de glicerol-3-fosfatasa (G3P fosfatasa), glicerol deshidratasa (*dhaB*), y 1,3-propanodiol oxidorreductasa (*dhaT*) para la producción incrementada de 1,3-propanodiol.

Antecedentes

1,3-propanodiol es un monómero que tiene utilidad potencial en la producción de fibras de poliéster y en la preparación de poliuretanos y compuestos cíclicos.

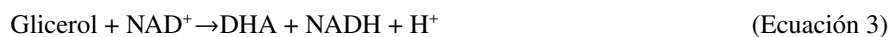
Se conocen varias rutas químicas para el 1,3-propanodiol. Por ejemplo, el óxido de etileno puede convertirse en 1,3-propanodiol sobre un catalizador en presencia de fosfina, agua, monóxido de carbono, hidrógeno y un ácido, mediante hidratación catalítica en fase de solución de acroleína seguida de reducción, o a partir de hidrocarburos tales como glicerol, hecho reaccionar en presencia de monóxido de carbono e hidrógeno sobre catalizadores con átomos del grupo VIII de la tabla periódica. Aunque es posible generar 1,3-propanodiol por estos métodos, son costosos y generan corrientes residuales que contienen contaminantes medioambientales.

Se ha conocido desde hace más de un siglo que el 1,3-propanodiol puede producirse en la fermentación del glicerol. Se han descubierto cepas bacterianas capaces de producir 1,3-propanodiol, por ejemplo, en los grupos *Citrobacter*, *Clostridium*, *Enterobacter*, *Ilyobacter*, *Klebsiella*, *Lactobacillus*, y *Pelobacter*. En cada caso estudiado, el glicerol se convierte en 1,3-propanodiol en una secuencia de reacción enzimática catalizada en dos etapas. En la primera etapa, una deshidratasa cataliza la conversión de glicerol en 3-hidroxi propionaldehído (3-HP) y agua (Ecuación 1). En la segunda etapa, 3-HP es reducido a 1,3-propanodiol por una oxidorreductasa asociada a NAD^+ (Ecuación 2).



El 1,3-propanodiol no se metaboliza posteriormente y, como resultado, se acumula a una concentración elevada en el medio. La reacción completa consume un equivalente reductor en forma de un cofactor, dinucleótido reducido de β -nicotinamida adenina (NADH), que se oxida a dinucleótido de nicotinamida adenina (NAD^+).

La producción de 1,3-propanodiol a partir de glicerol se realiza generalmente en condiciones anaerobias utilizando glicerol como única fuente de carbono y en ausencia de otros aceptores exógenos reductores equivalentes. En estas condiciones, por ejemplo, en cepas de *Citrobacter*, *Clostridium*, y *Klebsiella*, opera una ruta paralela para el glicerol que en primer lugar implica la oxidación de glicerol a dihidroxiacetona (DHA) mediante una glicerol deshidrogenasa asociada a NAD^+ (o NADP^+) (Ecuación 3). El DHA, tras la fosforilación hasta dihidroxiacetona fosfato (DHAP) mediante una DHA quinasa (Ecuación 4), está disponible para la biosíntesis y para intervenir en la generación de ATP a través, por ejemplo, de la glucólisis.



A diferencia de la ruta del 1,3-propanodiol, esta ruta puede proporcionar carbono y energía a la célula y produce NADH en vez de consumirlo.

En *Klebsiella pneumoniae* y *Citrobacter freundii*, los genes que codifican las actividades funcionalmente relacionadas de la glicerol deshidratasa (*dhaB*), 1,3-propanodiol oxidorreductasa (*dhaT*), glicerol deshidrogenasa (*dhaD*), y dihidroxiacetona quinasa (*dhaK*) están incluidos en el regulón *dha*. Los regulones *dha* de *Citrobacter* y *Klebsiella* se han expresado en *Escherichia coli* y se ha demostrado que convierten el glicerol en 1,3-propanodiol.

Se conocen procedimientos biológicos para la preparación de glicerol. La gran mayoría de productores de glicerol son levaduras, pero se sabe también que ciertas bacterias, otros hongos y algas producen glicerol. Tanto las bacterias como las levaduras producen glicerol convirtiendo la glucosa u otros carbohidratos a través de la ruta de fructosa-1,6-bisfosfato en la glucólisis o mediante la ruta de Embden Meyerhof Parnas, mientras que determinadas algas convierten el dióxido de carbono disuelto o el bicarbonato en los cloroplastos en compuestos intermedios de 3 carbonos del ciclo de Calvin. En una serie de etapas, el compuesto intermedio de 3 carbonos, ácido fosfoglicérico, se convierte en gliceraldehído 3-fosfato que puede interconvertirse fácilmente en su isómero ceto, dihidroxiacetona fosfato y por último en glicerol.

De manera específica, las bacterias *Bacillus licheniformis* y *Lactobacillus lycopersica* sintetizan glicerol, y se halla la producción de glicerol en las algas halotolerantes *Dunaliella sp.* y *Asteromonas gracilis* para la protección contra concentraciones salinas externas elevadas (Ben-Amotz *et al.*, *Experientia* 38, 49-52, (1982)). De manera similar, diversas levaduras osmotolerantes sintetizan glicerol como medida protectora. La mayoría de las cepas de *Saccharomyces* producen cierta cantidad de glicerol durante la fermentación alcohólica, y esto se puede incrementar fisiológicamente mediante la aplicación de estrés osmótico (Albertyn *et al.*, *Mol. Cell. Biol.* 14, 4135-4144, (1994)). Anteriormente en este siglo, la producción comercial de glicerol se llevaba a cabo mediante el uso de cultivos de *Saccharomyces* en los que se añadían “reactivos de dirección” tales como sulfitos o álcalis. Por medio de la formación de un complejo inactivo, los agentes de dirección bloquean o inhiben la conversión de acetaldehído a etanol; así, los equivalentes reductores (NADH) en exceso están disponibles o “se dirigen” hacia el DHAP para la reducción para producir glicerol. Este método está limitado por la inhibición parcial del crecimiento de las levaduras debido a los sulfitos. Esta limitación se puede superar parcialmente mediante el uso de álcalis que crean equivalentes de NADH en exceso mediante un mecanismo diferente. En esta práctica, los álcalis inician una desproporción de Cannizzaro para producir etanol y ácido acético a partir de dos equivalentes de acetaldehído.

El gen que codifica la glicerol-3-fosfato deshidrogenasa (DAR1, GPD1) se ha clonado y secuenciado a partir de *S. diastaticus* (Wang *et al.*, *J. Bact.* 176, 7091-7095, (1994)). El gen DAR1 se clonó en un vector transportador y se usó para transformar *E. coli*, en donde la expresión produjo la enzima activa. Wang *et al.* (anteriormente mencionado) reconoce que DAR1 está regulado por el medio osmótico celular, pero no propone cómo se podría usar el gen para incrementar la producción de 1,3-propanodiol en un organismo recombinante.

Se han aislado otras enzimas de glicerol-3-fosfato deshidrogenasa: por ejemplo, la sn-glicerol-3-fosfato deshidrogenasa se ha clonado y secuenciado a partir de *S. cerevisiae* (Larason *et al.*, *Mol. Microbiol.* 10, 1101, (1993)) y Albertyn *et al.*, (*Mol. Cell. Biol.* 14, 4135, (1994)) enseña la clonación de GPD1 que codifica una glicerol-3-fosfato deshidrogenasa de *S. cerevisiae*. Como Wang *et al.* (anteriormente mencionado), tanto Albertyn *et al.* como Larason *et al.* reconocen la osmo-sensibilidad de la regulación de este gen, pero no sugieren cómo se podría usar el gen en la producción de 1,3-propanodiol en un organismo recombinante.

Como con G3PDH, la glicerol-3-fosfatasa se ha aislado a partir de *Saccharomyces cerevisiae*, y se ha identificado que la proteína está codificada por los genes GPP1 y GPP2 (Norbeck *et al.*, *J. Biol. Chem.* 271, 13875 (1996)). Al igual que los genes que codifican G3PDH, parece que GPP2 es osmosensible.

Aunque se conocen métodos biológicos tanto de producción de glicerol como de 1,3-propanodiol, no se ha demostrado nunca que el procedimiento completo pueda realizarse mediante un solo organismo recombinante.

Ni los métodos químicos ni los biológicos descritos anteriormente para la producción de 1,3-propanodiol son muy adecuados para la producción a escala industrial, ya que los procedimientos químicos requieren mucha energía y los procedimientos biológicos requieren un material de partida costoso, glicerol. Se necesita un método que requiera poco aporte de energía y un material de partida económico. Un procedimiento más deseable incorporaría un microorganismo que tendría capacidad para convertir las fuentes básicas de carbono tales como los carbohidratos o azúcares en el producto final deseado 1,3-propanodiol.

Aunque sería deseable una conversión con un único organismo a partir de una fuente de carbono fermentable distinta de glicerol o dihidroxiacetona hasta 1,3-propanodiol, se ha documentado que existen dificultades significativas que superar en dicha empresa. Por ejemplo, Gottschalk *et al.* (patente EP 373 230) da a conocer que el crecimiento de la mayoría de cepas útiles para la producción de 1,3-propanodiol, incluyendo *Citrobacter freundii*, *Clostridium auto-butylicum*, *Clostridium butylicum* y *Klebsiella pneumoniae*, está alterado por la presencia de un donante de hidrógeno tal como fructosa o glucosa. Las cepas de *Lactobacillus brevis* y *Lactobacillus buchneri*, que producen 1,3-propanodiol en cofermentaciones de glicerol y fructosa o glucosa, no crecen cuando se proporciona glicerol como única fuente de carbono, y, aunque se ha demostrado que las células en reposo pueden metabolizar glucosa o fructosa, no producen 1,3-propanodiol. (Veiga DA Cunha *et al.*, *J. Bacteriol.* 174, 1013 (1992)). Asimismo, se ha demostrado que una cepa de *Ilyobacter polytropus*, que produce 1,3-propanodiol cuando se proporcionan glicerol y acetato, no producirá 1,3-propanodiol a partir de sustratos de carbono distintos de glicerol, incluyendo fructosa y glucosa. (Steib *et al.*, *Arch. Microbiol.* 140, 139 (1984)). Finalmente, Tong *et al.* (*Appl. Biochem. Biotech.* 34, 149 (1992)) ha dado a conocer que *Escherichia coli* recombinante transformada con el regulón *dha* que codifica la glicerol deshidratasa no produce 1,3-propanodiol a partir de glucosa o de xilosa en ausencia de glicerol exógeno.

Se han comunicado los intentos para mejorar el rendimiento de 1,3-propanodiol a partir de glicerol donde los cosustratos capaces de proporcionar equivalentes reductores, por lo general hidratos de carbono fermentables, se incluyen en el procedimiento. Se han reivindicado mejoras del rendimiento para células en reposo de *Citrobacter freundii* y *Klebsiella pneumoniae* DSM 4270 mediante cofermentación con glicerol y glucosa (Gottschalk *et al.*, anteriormente mencionado, y Tran-Dinh *et al.*, documento DE 3734 764); pero no para células en crecimiento de *Klebsiella pneumoniae* ATCC 25955 que cofermenta con glicerol y glucosa, que no producían 1,3-propanodiol (I-T. Tong, Tesis doctoral, Universidad de Wisconsin-Madison (1992)). Se han publicado aumentos de rendimientos para la cofermentación de glicerol y glucosa o fructosa por una *Escherichia coli* recombinante; sin embargo, no se produce 1,3-propanodiol en ausencia de glicerol (Tong *et al.*, anteriormente mencionado.). En estos sistemas, los organismos individuales utilizan los carbohidratos como fuente de generación de NADH a la vez que proporcionan energía y carbono para el mantenimiento o crecimiento celular. Estas descripciones sugieren que los hidratos de carbono no entran en la corriente de

carbono que produce 1,3-propanodiol. En ningún caso se produce 1,3-propanodiol en ausencia de una fuente exógena de glicerol. Por lo tanto, la mayor parte de la bibliografía indica claramente que la producción de 1,3-propanodiol a partir de una fuente de carbohidratos por un solo organismo no es posible.

- 5 El documento WO 96/35796 describe un método para la bioconversión de un sustrato de carbono en 1,3-propanodiol con un único microorganismo. El microorganismo se transforma con un fragmento de ADN que codifica *dhaB1*, *dhaB2* y *dhaB3* (que juntos codifican la glicerol deshidratasa) y *dhaT* (que codifica la 1,3-propanodiol oxidorreductasa).
- 10 El problema que debe resolver la presente invención es la producción biológica de 1,3-propanodiol mediante un solo organismo recombinante a partir de un sustrato de carbono económico tal como glucosa u otros hidratos de carbono. La producción biológica de 1,3-propanodiol requiere glicerol como sustrato para una reacción secuencial en dos etapas en la que una enzima deshidratasa (por lo general una deshidratasa dependiente de la coenzima B₁₂) convierte el glicerol en un compuesto intermedio, 3-hidroxi propionaldehído, que es reducido a continuación a 1,3-propanodiol por una oxidorreductasa dependiente de NADH (o NADPH). La complejidad de la necesidad del cofactor requiere la utilización de un catalizador de célula completa para un procedimiento industrial que utiliza esta secuencia de reacción para la producción de 1,3-propanodiol. Además, a fin de hacer el procedimiento económicamente viable, se necesita una materia prima menos costosa que el glicerol o la dihidroxiacetona. La glucosa y otros carbohidratos son sustratos adecuados, pero, como se expuso anteriormente, son conocidos por interferir con la producción de 1,3-propanodiol. Como resultado se ha demostrado que ningún organismo individual convierte la glucosa en 1,3-propanodiol.

Los solicitantes han resuelto el problema planteado, y la presente invención proporciona la bioconversión de una fuente de carbono fermentable directamente en 1,3-propanodiol utilizando un solo organismo recombinante. La glucosa se utiliza como sustrato modelo y la bioconversión es aplicable a cualquier microorganismo existente. Los microorganismos que albergan los genes que codifican la glicerol-3-fosfato deshidrogenasa (G3PDH), glicerol-3-fosfatasa (G3P fosfatasa), glicerol deshidratasa (*dhaB*), y 1,3-propanodiol oxidorreductasa (*dhaT*), son capaces de convertir la glucosa y otros hidratos de carbono por medio de la ruta de degradación del glicerol hasta 1,3-propanodiol con buenos rendimientos y selectividades. Además, la presente invención puede aplicarse en general para que incluya cualquier sustrato de carbono que se convierte fácilmente en 1) glicerol, 2) dihidroxiacetona o 3) compuestos de C₃ en el estado de oxidación de glicerol (p. ej., glicerol 3-fosfato) o 4) compuestos de C₃ en el estado de oxidación de la dihidroxiacetona (p. ej., fosfato de dihidroxiacetona o gliceraldehído 3-fosfato).

Compendio de la invención

35 La presente invención proporciona un método para la producción de 1,3-propanodiol a partir de un organismo recombinante, que comprende:

40 (i) transformar un organismo hospedador adecuado con uno o más casetes de transformación, cada uno de los cuales comprende al menos uno de (a) un gen que codifica una actividad de glicerol-3-fosfato deshidrogenasa; (b) un gen que codifica una actividad de glicerol-3 fosfatasa; (c) genes que codifican una actividad de deshidratasa; y (d) un gen que codifica una actividad de 1,3-propanodiol oxidorreductasa, en los que todos los genes (a)-(d) se introducen en el microorganismo hospedador;

45 (ii) cultivar el organismo hospedador transformado en condiciones adecuadas en presencia de al menos una fuente de carbono seleccionada del grupo que consiste en monosacáridos, oligosacáridos, polisacáridos, o un sustrato de carbono mediante el cual se produce 1,3-propanodiol; y

50 (iii) recuperar el 1,3-propanodiol.

La invención también se refiere a hospedadores transformados que comprenden casetes de expresión capaces de expresar actividades de glicerol-3-fosfato deshidrogenasa, glicerol-3-fosfatasa, glicerol deshidratasa y 1,3-propanodiol oxidorreductasa para la producción de 1,3-propanodiol, tal como se define en la reivindicación 13.

55 El organismo hospedador adecuado usado en el método se selecciona del grupo que consiste en bacterias, levaduras, y hongos filamentosos. El organismo hospedador adecuado se selecciona más en particular del grupo de géneros que consisten en *Citrobacter*, *Enterobacter*, *Clostridium*, *Klebsiella*, *Aerobacter*, *Lactobacillus*, *Aspergillus*, *Saccharomyces*, *Schizosaccharomyces*, *Zygosaccharomyces*, *Pichia*, *Kluyveromyces*, *Candida*, *Hansenula*, *Debaryomyces*, *Mucor*, *Torulopsis*, *Methylobacter*, *Escherichia*, *Salmonella*, *Bacillus*, *Streptomyces* y *Pseudomonas*. Más en particular, el organismo hospedador adecuado se selecciona del grupo que consiste en *E. coli*, *Klebsiella spp.*, y *Saccharomyces spp.* Los organismos hospedadores transformados particulares usados en el método son 1) un *Saccharomyces spp.* transformado con un casete de transformación que comprende los genes *dhaB1*, *dhaB2*, *dhaB3*, y *dhaT*, en el que los genes están integrados de manera estable en el genoma de *Saccharomyces spp.*; y 2) una *Klebsiella spp.* transformada con un casete de transformación que comprende los genes GPD1 y GPD2.

La fuente de carbono preferida de la invención es glucosa.

ES 2 336 858 T3

El método usa además el gen que codifica una enzima glicerol-3-fosfato deshidrogenasa seleccionado del grupo que consiste en los genes que corresponden a las secuencias de aminoácidos proporcionadas en SEQ ID NO: 11, en SEQ ID NO: 12, y en SEQ ID NO: 13, y las secuencias de aminoácidos abarcan las sustituciones, deleciones o adiciones de aminoácidos que no alteran la función de la enzima glicerol-3-fosfato deshidrogenasa. El método usa también el gen que codifica una enzima glicerol-3-fosfatasa seleccionado del grupo que consiste en los genes que corresponden a las secuencias de aminoácidos proporcionadas en SEQ ID NO:33 y en SEQ ID NO: 17, y las secuencias de aminoácidos abarcan las sustituciones, deleciones o adiciones de aminoácidos que no alteran la función de la enzima glicerol-3-fosfatasa. El método usa también el gen que codifica una enzima glicerol quinasa que corresponde a una secuencia de aminoácidos proporcionada en SEQ ID NO:18, la secuencia de aminoácidos que abarca las sustituciones, deleciones o adiciones de aminoácidos que no alteran la función de la enzima glicerol quinasa. El método usa también los genes que codifican una enzima deshidratasa que comprenden dhaB1, dhaB2 y dhB3, genes que corresponden respectivamente a las secuencias de aminoácidos proporcionadas en SEQ ID NO:34, SEQ ID NO:35, y SEQ ID NO:36, y las secuencias de aminoácidos abarcan las sustituciones, deleciones o adiciones de aminoácidos que no alteran la función de la enzima deshidratasa. El método usa también el gen que codifica una enzima 1,3-propanodiol oxidorreductasa que corresponde a una secuencia de aminoácidos proporcionada en SEQ ID NO:37, la secuencia de aminoácidos que abarca las sustituciones, deleciones o adiciones de aminoácidos que no alteran la función de la enzima 1,3-propanodiol oxidorreductasa.

La invención se realiza también en una célula hospedadora transformada que comprende:

(1) un gen heterólogo que codifica una enzima glicerol-3-fosfato deshidrogenasa que corresponde a la secuencia de aminoácidos proporcionada en SEQ ID NO: 11;

(2) un gen heterólogo que codifica una enzima glicerol-3-fosfatasa que corresponde a la secuencia de aminoácidos proporcionada en SEQ ID NO: 17;

(3) un gen heterólogo que codifica la subunidad α de la enzima glicerol deshidratasa que corresponde a la secuencia de aminoácidos proporcionada en SEQ ID NO:34;

(4) un gen heterólogo que codifica la subunidad β de la enzima glicerol deshidratasa que corresponde a la secuencia de aminoácidos proporcionada en SEQ ID NO:35;

(5) un gen heterólogo que codifica la subunidad γ de la enzima glicerol deshidratasa que corresponde a la secuencia de aminoácidos proporcionada en SEQ ID NO:36; y

(6) un gen heterólogo que codifica la enzima 1,3-propanodiol oxidorreductasa que corresponde a la secuencia de aminoácidos proporcionada en SEQ ID NO:37, las secuencias de aminoácidos respectivas de (1)-(6) que abarcan las sustituciones, deleciones o adiciones de aminoácidos que no alteran la función de las enzimas de los genes (1)-(6), y

por lo que la célula hospedadora transformada produce 1,3-propanodiol con al menos un sustrato seleccionado del grupo que consiste en monosacáridos, oligosacáridos, y polisacáridos o a partir de un sustrato de un carbono.

Breve descripción de depósitos

45 *Biológicos y listado de secuencias*

La cepa transformada de *E. coli* W2042 (que comprende el hospedador de *E. coli* W1485 y los plásmidos pDT20 y pAH42) que contiene los genes que codifican la glicerol-3-fosfato deshidrogenasa (G3PDH) y glicerol-3-fosfatasa (G3P fosfatasa), glicerol deshidratasa (*dhaB*), y 1,3-propanodiol oxidorreductasa (*dhaT*) se depositó el 26 de septiembre de 1996 con la ATCC de acuerdo con el Tratado de Budapest sobre el Reconocimiento Internacional del Depósito de Microorganismos a los Fines del Procedimiento en Materia de Patentes, y se denominó ATCC 98188.

S. cerevisiae YPH500 que alberga los plásmidos pMCK10, pMCK17, pMCK30 y pMCK35 que contienen los genes que codifican la glicerol-3-fosfato deshidrogenasa (G3PDH) y glicerol-3-fosfatasa (G3P fosfatasa), glicerol deshidratasa (*dhaB*), y 1,3-propanodiol oxidorreductasa (*dhaT*) se depositó el 26 de septiembre 1996 con la ATCC de acuerdo con el Tratado de Budapest sobre el Reconocimiento Internacional del Depósito de Microorganismos a los Fines del Procedimiento en Materia de Patentes, y se denominó ATCC 74392.

“ATCC” se refiere al depositario internacional American Type Culture Collection situado en 12301 Parklawn Drive, Rockville, MD 20852 EE.UU. Las denominaciones se refieren al número de registro del material depositado.

Los solicitantes han proporcionado 49 secuencias en conformidad con las Normas para la Representación Estándar de Secuencias de Nucleótidos y Aminoácidos en Solicitudes de Patente (Anexos I y II a la Decisión del Presidente de la EPO, publicados en el Suplemento n° 2 a OJ EPO, 12/1992) y con 37 C.F.R. 1.821-1.825 y Apéndices A y B (Requisitos para Descripciones de Solicitud que Contienen Secuencias de Nucleótidos y/o Aminoácidos).

Descripción detallada de la invención

La presente invención proporciona un método para una producción biológica de 1,3-propanodiol a partir de una fuente de carbono fermentable en un único organismo recombinante. El método incorpora un microorganismo que contiene genes que codifican la glicerol-3-fosfato deshidrogenasa (G3PDH), glicerol-3-fosfatasa (G3P fosfatasa), glicerol deshidratasa (*dhaB*), y 1,3-propanodiol oxidorreductasa (*dhaT*). El microorganismo recombinante se pone en contacto con un sustrato de carbono, y se aísla 1,3-propanodiol del medio de cultivo.

El presente método proporciona una fuente rápida, económica y responsable con el medio ambiente de monómero 1,3-propanodiol útil en la producción de poliésteres y otros polímeros.

Se deben usar las siguientes definiciones para interpretar las reivindicaciones y la memoria descriptiva.

Las expresiones “glicerol deshidratasa” o “enzima deshidratasa” se refieren a el/los polipéptido(s) responsable(s) de una actividad enzimática que es capaz de isomerizar o convertir una molécula de glicerol en el producto 3-hidroxi-propionaldehído. Para los fines de la presente invención, las enzimas deshidratasas incluyen una glicerol deshidratasa (GenBank U09771, U30903) y una diol deshidratasa (GenBank D45071) que tienen como sustratos preferidos glicerol y 1,2-propanodiol, respectivamente. La glicerol deshidratasa de *K. pneumoniae* ATCC 25955 está codificada por los genes *dhaB1*, *dhaB2*, y *dhaB3* identificados como SEQ ID NOS:1, 2 y 3, respectivamente. Los genes *dhaB1*, *dhaB2*, y *dhaB3* codifican las subunidades α , β , y γ de la enzima glicerol deshidratasa, respectivamente.

Las expresiones “oxidorreductasa” o “1,3-propanodiol oxidorreductasa” se refieren a el/los polipéptido(s) responsable(s) de una actividad enzimática que es capaz de catalizar la reducción de 3-hidroxi-propionaldehído a 1,3-propanodiol. La 1,3-propanodiol oxidorreductasa incluye, por ejemplo, el polipéptido codificado por el gen *dhaT* (GenBank U09771, U30903) y se identifica como SEQ ID NO:4.

Las expresiones “glicerol-3-fosfato deshidrogenasa” o “G3PDH” se refieren a el/los polipéptido(s) responsable(s) de una actividad enzimática capaz de catalizar la conversión de dihidroxiacetona fosfato (DHAP) a glicerol-3-fosfato (G3P). *In vivo* G3PDH puede ser dependiente de NADH, NADPH, o FAD. Los ejemplos de esta actividad enzimática incluyen los siguientes: Las enzimas dependientes de NADH (EC 1.1.1.8) están codificadas por varios genes que incluyen GPD1 (GenBank Z74071x2) o GPD2 (GenBank Z35169x1) o GPD3 (GenBank G984182) o DAR1 (GenBank Z74071 x2); una enzima dependiente de NADPH (EC 1.1.1.94) está codificada por *gpsA* (GenBank U32164, G466746 (cds 197911-196892), y L45246); y las enzimas dependientes de FAD (EC 1.1.99.5) están codificadas por GUT2 (GenBank Z47047x23) o glpD (GenBank G147838) o glpABC (GenBank M20938).

Las expresiones “glicerol-3-fosfatasa” o “sn-glicerol-3-fosfatasa” o “d.l-glicerol fosfatasa” o “G3P fosfatasa” se refieren a el/los polipéptido(s) responsable(s) de una actividad enzimática que es capaz de catalizar la conversión de glicerol-3-fosfato a glicerol. La G3P fosfatasa incluye, por ejemplo, los polipéptidos codificados por GPP1 (GenBank Z47047x125) o GPP2 (GenBank U18813x11).

La expresión “glicerol quinasa” se refiere a el/los polipéptido(s) responsable(s) de una actividad enzimática capaz de catalizar la conversión de glicerol a glicerol-3-fosfato o glicerol-3-fosfato a glicerol, dependiendo de las condiciones de reacción. La glicerol quinasa incluye, por ejemplo, el polipéptido codificado por GUT1 (GenBank U11583x19).

Las expresiones “GPD1”, “DAR1”, “OSG1”, “D2830”, y “YDL022W” se usarán de manera intercambiable, y se refieren a un gen que codifica una glicerol-3-fosfato deshidrogenasa citosólica, y caracterizada por la secuencia de bases proporcionada en SEQ ID NO:5.

La expresión “GPD2” se refiere a un gen que codifica una glicerol-3-fosfato deshidrogenasa citosólica, y caracterizada por la secuencia de bases proporcionada en SEQ ID NO: 6.

Las expresiones “GUT2” y “YIL 155C” se usan de manera intercambiable, y se refieren a un gen que codifica una glicerol-3-fosfato deshidrogenasa mitocondrial, y caracterizado por la secuencia de bases proporcionada en SEQ ID NO:7.

Las expresiones “GPP1”, “RHR2” y “YIL053W” se usan de manera intercambiable, y se refieren a un gen que codifica una glicerol-3-fosfatasa citosólica, y caracterizado por la secuencia de bases proporcionada en SEQ ID NO:8.

Las expresiones “GPP2”, “HOR2” y “YER062C” se usan de manera intercambiable, y se refieren a un gen que codifica una glicerol-3-fosfatasa citosólica, y caracterizado por la secuencia de bases proporcionada en SEQ ID NO:9.

La expresión “GUT1” se refiere a un gen que codifica una glicerol quinasa citosólica, y caracterizado por la secuencia de bases proporcionada en SEQ ID NO:10.

Las expresiones “función” o “función enzimática” se refieren a la actividad catalítica de una enzima al alterar la energía necesaria para llevar a cabo una reacción química específica. Se entiende que tal actividad puede ser aplicable a una reacción en equilibrio, en la que la producción de cualquier producto o sustrato se puede llevar a cabo en condiciones adecuadas.

ES 2 336 858 T3

Las expresiones “polipéptido” y “proteína” se usan de manera intercambiable.

Las expresiones “sustrato de carbono” y “fuente de carbono” se refieren a una fuente de carbono capaz de ser metabolizada por organismos hospedadores de la presente invención, y en particular fuentes de carbono seleccionadas del grupo que consiste en monosacáridos, oligosacáridos, polisacáridos, y sustratos de un carbono o mezclas de los mismos.

Las expresiones “célula hospedadora” u “organismo hospedador” se refieren a un microorganismo capaz de recibir genes exógenos o heterólogos y de expresar esos genes para producir un producto génico activo.

Las expresiones “gen exógeno”, “ADN exógeno”, “gen heterólogo” y “ADN heterólogo” se refieren a material genético nativo para un organismo que se ha colocado en un organismo hospedador por diversos medios.

Las expresiones “organismo recombinante” y “hospedador transformado” se refieren a cualquier organismo que se ha transformado con genes heterólogos o exógenos. Los organismos recombinantes de la presente invención expresan genes exógenos que codifican la glicerol-3-fosfato deshidrogenasa (G3PDH) y glicerol-3-fosfatasa (G3P fosfatasa), glicerol deshidratasa (*dhaB*), y 1,3-propanodiol oxidoreductasa (*dhaT*) para la producción de 1,3-propanodiol a partir de sustratos de carbono adecuados.

“Gen” se refiere a un fragmento de ácido nucleico que expresa una proteína específica, que incluye secuencias reguladoras anteriores (no codificantes en 5') y posteriores a la zona de codificación (no codificantes en 3'). Los términos “nativo” y “de tipo natural” se refieren a un gen tal como se halla en la naturaleza con sus propias secuencias reguladoras.

Las expresiones “que codifica” y “codificante” se refieren al proceso mediante el cual un gen, por medio de los mecanismos de transcripción y traducción, produce una secuencia de aminoácidos. Se entiende que el proceso de codificar una secuencia de aminoácidos específica incluye secuencias de ADN que pueden implicar cambios de bases que no provocan un cambio en el aminoácido codificado, o que implican cambios de bases que pueden alterar uno o más aminoácidos, pero que no afectan a las propiedades funcionales de la proteína codificada por la secuencia de ADN. Se sobrentiende por lo tanto que la invención abarca más que las secuencias ejemplares específicas. Las modificaciones en la secuencia, tales como eliminaciones, inserciones o sustituciones en la secuencia que producen cambios imperceptibles que no afectan sustancialmente a las propiedades funcionales de la molécula proteica resultante también están incluidas. Por ejemplo, se contempla la modificación en la secuencia génica que refleja la degeneración del código genético, o que da lugar a la producción de un aminoácido químicamente equivalente en un punto dado; así, un codón para el aminoácido alanina, un aminoácido hidrófobo, puede ser sustituido por un codón que codifique otro resto menos hidrófobo tal como glicina, o un resto más hidrófobo, tal como valina, leucina o isoleucina. Asimismo, es de esperar que los cambios que dan lugar a la sustitución de un resto con carga negativa por otro, tales como ácido aspártico por ácido glutámico, o un resto con carga positiva por otro, tales como lisina por arginina, produzcan un producto biológicamente equivalente. Tampoco es de esperar que los cambios de nucleótidos que causan la alteración de las porciones N-terminal y C-terminal de la molécula proteica modifiquen la actividad de la proteína. En algunos casos, puede, de hecho, ser deseable preparar mutantes de la secuencia con el fin de estudiar el efecto de la modificación sobre la actividad biológica de la proteína. Cada una de las modificaciones propuestas es correcta dentro de la pericia rutinaria en la técnica, como es la determinación de la retención de la actividad biológica en los productos codificados. Además, el técnico experto reconoce que las secuencias incluidas en esta invención se definen también por su capacidad de hibridar, bajo condiciones rigurosas (SSC 0,1X, 0,1% de SDS, 65°C), con las secuencias ejemplificadas en la presente memoria.

El término “expresión” se refiere a la transcripción y traducción al producto génico a partir de un gen que codifica la secuencia del producto génico.

Los términos “plásmido”, “vector”, y “casete” se refieren a un elemento cromosómico extra que a menudo porta genes que no son parte del metabolismo principal de la célula, y normalmente en forma de moléculas de ADN bicatenarias circulares. Tales elementos pueden ser secuencias que se replican de manera autónoma, secuencias de integración en el genoma, secuencias de fagos o nucleotídicas, lineales o circulares, de un ADN o ARN monocatenario o bicatenario, procedentes de cualquier fuente, en las que se han unido o recombinado varias secuencias nucleotídicas en una única construcción que es capaz de introducir un fragmento promotor y una secuencia de ADN para un producto génico seleccionado junto con una secuencia sin traducir en 3' apropiada en una célula. “Casete de transformación” se refiere a un vector específico que contiene un gen exógeno y que tiene elementos además del gen exógeno, que facilitan la transformación de una célula hospedadora particular. “Casete de expresión” se refiere a un vector específico que contiene un gen exógeno y que tiene elementos además del gen exógeno que permiten la expresión incrementada de ese gen en un hospedador exógeno.

Los términos “transformación” y “transfección” se refieren a la obtención de nuevos genes en una célula tras la incorporación de ácido nucleico. Los genes obtenidos pueden estar integrados en el ADN cromosómico o introducirse como secuencias replicantes extracromosómicas. El término “transformante” se refiere al producto de una transformación.

La expresión “modificado genéticamente” se refiere al proceso de cambio del material hereditario por transformación o mutación.

ES 2 336 858 T3

Construcción de organismos recombinantes

Los organismos recombinantes que contienen los genes necesarios que codificarán la ruta enzimática para la conversión de un sustrato de carbono en 1,3-propanodiol pueden construirse utilizando métodos bien conocidos en la técnica. En la presente invención, los genes que codifican la glicerol-3-fosfato deshidrogenasa (G3PDH), glicerol-3-fosfatasa (G3P fosfatasa), glicerol deshidratasa (*dhaB*), y 1,3-propanodiol oxidorreductasa (*dhaT*) se aislaron a partir de un hospedador nativo tal como *Klebsiella* o *Saccharomyces* y se usaron para transformar cepas hospedadoras tales como *E. coli* DH5 α , ECL707, AA200, o W1485; la cepa de *Saccharomyces cerevisiae* YPH500; o las cepas de *Klebsiella pneumoniae* ATCC 25955 o ECL 2106.

Aislamiento de Genes

Los métodos de obtención de los genes deseados procedentes de un genoma bacteriano son frecuentes y muy conocidos en la técnica de biología molecular. Por ejemplo, si se conoce la secuencia del gen, pueden crearse genotecas adecuadas por digestión con endonucleasas de restricción y pueden cribarse con sondas complementarias a la secuencia génica deseada. Una vez se aísla la secuencia, el ADN puede ampliarse utilizando métodos de ampliación dirigidos por un cebador convencional, tal como la reacción en cadena de polimerasa (PCR) (documento U.S. n° 4.683.202) para obtener cantidades de ADN adecuado para la transformación utilizando vectores apropiados.

Alternativamente, pueden crearse bancos de cósmidos en los que grandes segmentos de ADN genómico (35-45 kb) pueden rellenarse dentro de los vectores y utilizarse para transformar hospedadores apropiados. Los vectores cósmidos son únicos en poder acomodar grandes cantidades de ADN. En general, los vectores cósmidos tienen al menos una copia de la secuencia *cos* de ADN que se necesita para el encapsulado y la subsiguiente ciclación del ADN exógeno. Además de la secuencia *cos* estos vectores contendrán también un origen de replicación tal como ColE1 y marcadores con resistencia a fármacos tal como un gen resistente a la ampicilina o a neomicina. Los métodos de utilización de vectores cósmidos para la transformación de hospedadores bacterianos adecuados están bien descritos en Sambrook *et al.*, Molecular Cloning: A Laboratory Manual, Segunda Edición (1989) Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY (1989).

Por lo general, para clonar cósmidos, se aísla y se liga el ADN exógeno, utilizando las endonucleasas de restricción apropiadas, en posición adyacente a la región *cos* del vector cósmido. Los vectores cósmidos que contienen el ADN exógeno linealizado se hacen reaccionar a continuación con un vehículo de encapsulación de ADN tal como el bacteriófago λ . Durante el proceso de encapsulación las secuencias *cos* se escinden y el ADN exógeno se encapsula en la parte de cabeza de la partícula vírica bacteriana. Estas partículas se utilizan a continuación para transfectar células hospedadoras adecuadas tales como *E. coli*. Una vez inyectado en la célula, el ADN exógeno se circulariza bajo la influencia de los extremos adhesivos *cos*. De esta manera pueden introducirse y expresarse segmentos grandes de ADN exógeno en células hospedadoras recombinantes.

Aislamiento y clonación de genes que codifican la glicerol deshidratasa (*dhaB*) y 1,3-propanodiol oxidorreductasa (*dhaT*)

Los vectores cósmidos y los métodos de transformación de cósmidos se utilizaron dentro del contexto de la presente invención para clonar grandes segmentos de ADN genómico procedente de genes bacterianos conocidos por poseer genes capaces de transformar el glicerol hasta 1,3-propanodiol. De manera específica, se aisló el ADN genómico de *K. pneumoniae* ATCC 25955 mediante métodos bien conocidos en la técnica y se digirió con la enzima de restricción Sau3A para la inserción en un vector cósmido Supercos 1, y se encapsuló mediante el uso de extractos de encapsulación GigapackII. Después de la construcción del vector las células XL1-Blue MR de *E. coli* se transformaron con el ADN del cósmido. Se identificó la capacidad de los transformantes para convertir glicerol en 1,3-propanodiol cultivando las células en presencia de glicerol y analizando la formación de 1,3-propanodiol en el medio.

Dos de los transformantes positivos para 1,3-propanodiol se analizaron y los cósmidos se denominaron pKP1 y pKP2. La secuenciación del ADN reveló una gran homología respecto del gen de glicerol deshidratasa (*dhaB*) de *C. freundii*, lo que demuestra que estos transformantes contenían ADN que codificaba el gen de glicerol deshidratasa. Se analizaron otros transformantes positivos para 1,3-propanodiol y los cósmidos se denominaron pKP4 y pKP5. La secuenciación del ADN puso de manifiesto que estos cósmidos llevaban ADN que codifica un gen de diol deshidratasa.

Aunque la presente invención utiliza los genes aislados del interior de un cósmido de *Klebsiella*, las fuentes alternativas de genes de deshidratasa incluyen, pero sin limitación, *Citrobacter*, *Clostridia*, y *Salmonella*.

Genes que codifican G3PDH y G3P fosfatasa

La presente invención implica genes adecuados para la expresión de las actividades de G3PDH y G3P fosfatasa en una célula hospedadora.

ES 2 336 858 T3

Se conocen los genes que codifican G3PDH. Por ejemplo, se ha aislado GPD1 de *Saccharomyces*, y tiene la secuencia de bases proporcionada en SEQ ID NO:5, que codifica la secuencia de aminoácidos proporcionada en SEQ ID NO:11 (Wang *et al.*, anteriormente mencionado). De manera similar, también se ha aislado la actividad de G3PDH de *Saccharomyces* codificada por GPD2 que tiene la secuencia de bases proporcionada en SEQ ID NO:6, que codifica la secuencia de aminoácidos proporcionada en SEQ ID NO:12 (Eriksson *et al.*, Mol. Microbiol. 17, 95, (1995)).

Se considera que cualquier gen que codifique un polipéptido responsable de la actividad de G3PDH es adecuado para los fines de la presente invención, en el que esa actividad es capaz de catalizar la conversión de dihidroxiacetona fosfato (DHAP) a glicerol-3-fosfato (G3P). Además, se considera que cualquier gen que codifique la secuencia de aminoácidos de G3PDH, tal como se proporciona mediante cualquiera de SEQ ID NOS:11, 12, 13, 14, 15 y 16, que corresponden a los genes GPD1, GPD2, GUT2, *gpsA*, *glpD*, y la subunidad α de *glpABC*, respectivamente, será funcional en la presente invención, en la que esa secuencia de aminoácidos incluye las sustituciones, delecciones o adiciones de aminoácidos que no alteran la función de la enzima. Los expertos apreciarán que los genes que codifican G3PDH aislados de otras fuentes también son adecuados para el uso en la presente invención. Por ejemplo, los genes aislados de procariotas incluyen los registros de GenBank M34393, M20938, L06231, U12567, L45246, L45323, L45324, L45325, U32164, y U39682; los genes aislados de hongos incluyen los registros de GenBank U30625, U30876 y X56162; los genes aislados de insectos incluyen los registros de GenBank X61223 y X14179; y los genes aislados de fuentes de origen mamífero incluyen los registros de GenBank U12424, M25558 y X78593.

Se conocen los genes que codifican la G3P fosfatasa. Por ejemplo, se ha aislado GPP2 de *Saccharomyces cerevisiae*, y tiene la secuencia de bases proporcionada en SEQ ID NO:9 que codifica la secuencia de aminoácidos proporcionada en SEQ ID NO: 17 (Norbeck *et al.*, J Biol. Chem. 271, pág. 13875, 1996).

Se considera que cualquier gen que codifique una actividad de G3P fosfatasa es adecuado para los fines de la presente invención, en donde esa actividad es capaz de catalizar la conversión de glicerol-3-fosfato a glicerol. Además, se considera que cualquier gen que codifique la secuencia de aminoácidos de G3P fosfatasa proporcionada en SEQ ID NOS:33 y 17 será funcional en la presente invención, en donde esa secuencia de aminoácidos abarca las sustituciones, delecciones o adiciones de aminoácidos que no alteran la función de la enzima. El experto apreciará que los genes que codifican la G3P fosfatasa aislados de otras fuentes también son adecuados para el uso en la presente invención. Por ejemplo, la desfosforilación de glicerol-3-fosfato para proporcionar glicerol se puede llevar a cabo con una o más de las siguientes fosfatasas generales o específicas: fosfatasa alcalina (EC 3.1.3.1) [GenBank M19159, M29663, U02550 o M33965]; fosfatasa ácida (EC 3.1.3.2) [GenBank U51210, U19789, U28658 o L20566]; glicerol-3-fosfatasa (EC 3.1.3.-) [GenBank Z38060 o U18813x11]; glucosa-1-fosfatasa (EC 3.1.3.10) [GenBank M33807]; glucosa-6-fosfatasa (EC 3.1.3.9) [GenBank U00445]; fructosa-1,6-bisfosfatasa (EC 3.1.3.11) [GenBank X12545 o J03207] o fosfatidil glicerofosfato fosfatasa (EC 3.1.3.27) [GenBank M23546 y M23628].

Se conocen los genes que codifican la glicerol quinasa. Por ejemplo, se ha aislado y secuenciado GUT1, que codifica la glicerol quinasa de *Saccharomyces* (Pavlik *et al.*, Curr. Genet, 24, 21, (1993)), y la secuencia de bases se proporciona en SEQ ID NO:10, que codifica la secuencia de aminoácidos proporcionada en SEQ ID NO:18. El técnico experto apreciará que aunque la glicerol quinasa cataliza la degradación de glicerol en la naturaleza, la misma enzima será capaz de funcionar en la síntesis de glicerol para convertir glicerol-3-fosfato en glicerol en las condiciones de energía de reacción apropiadas. Existen pruebas de la producción de glicerol por medio de una glicerol quinasa. En condiciones anaerobias o de inhibición de la respiración, *Trypanosoma brucei* da lugar a glicerol en presencia de Glicerol-3-P y ADP. La reacción transcurre en el compartimento del glucosoma (D. Hamond, J. Biol. Chem. 260, 15646-15654, (1985)).

Células hospedadoras

Las células hospedadoras adecuadas para la producción recombinante de glicerol mediante la expresión de G3PDH y G3P fosfatasa pueden ser procarióticas o eucarióticas, y estarán limitadas solamente por su capacidad de expresar enzimas activas. Los hospedadores preferidos serán aquellos generalmente útiles para la producción de glicerol o 1,3-propanodiol, tales como *Citrobacter*, *Enterobacter*, *Clostridium*, *Klebsiella*, *Aerobacter*, *Lactobacillus*, *Aspergillus*, *Saccharomyces*, *Schizosaccharomyces*, *Zygosaccharomyces*, *Pichia*, *Kluyveromyces*, *Candida*, *Hansenula*, *Debaryomyces*, *Mucor*, *Torulopsis*, *Methylobacter*, *Escherichia*, *Salmonella*, *Bacillus*, *Streptomyces* y *Pseudomonas*. Los más preferidos en la presente invención son *E. coli*, las especies de *Klebsiella* y las especies de *Saccharomyces*.

La adenosil-cobalamina (coenzima B₁₂) es un cofactor esencial para la actividad de la glicerol deshidratasa. La coenzima es el producto natural no polimérico más complejo conocido, y su síntesis *in vivo* se dirige mediante el uso de los productos de alrededor de 30 genes. La síntesis de la coenzima B₁₂ se halla en los procariotas, algunos de los cuales son capaces de sintetizar el compuesto *de novo*, mientras otros pueden llevar a cabo reacciones parciales. *E. coli*, por ejemplo, no puede fabricar la estructura del anillo de corrina, pero puede catalizar la conversión de cobinamida a corrinoide y puede introducir el grupo 5'-desoxiadenosilo.

Los eucariotas no pueden sintetizar la coenzima B₁₂ *de novo*, y en lugar de ello transportan la vitamina B₁₂ desde el medio extracelular, con la conversión posterior del compuesto a su forma funcional mediante las enzimas celulares. Se han descrito tres actividades enzimáticas para esta serie de reacciones.

ES 2 336 858 T3

1) acuocobalamina reductasa (EC 1.6.99.8) reduce Co(III) a Co(II);

2) cob(II)alamina reductasa (EC 1.6.99.9) reduce Co(II) a Co(I); y

5 3) cob(I)alamina adenosiltransferasa (EC 2.5.1.17) transfiere un resto 5'-desoxiadenosina desde el ATP al corri-
noide reducido. Esta última actividad enzimática es la mejor caracterizada de las tres, y está codificada por *cobA* en
S. typhimurium, *btuR* en *E. coli* y *cobO* en *P. denitrificans*. Se han clonado y secuenciado estos tres genes de cob(I)
alamina adenosiltransferasa. Se ha detectado la actividad de cob(I)alamina adenosiltransferasa en fibroblastos huma-
nos y en mitocondrias de rata aisladas (Fenton *et al.*, Biochem. Biophys. Res. Commun. 98, 283-9, (1981)). Las dos
10 enzimas implicadas en la reducción de cobalto están escasamente caracterizadas, y las secuencias génicas no están
disponibles. Existen informes de una acuocobalamina reductasa de *Euglena gracilis* (Watanabe *et al.*, Arch. Biochem.
Biophys. 305, 421-7, (1993)), y una cob(III)alamina reductasa microsomal está presente en las fracciones de las mem-
branas internas microsomales y mitocondriales de fibroblastos de rata (Pezacka, Biochim. Biophys. Acta, 1157, 167-
77, (1993)).

15 La complementación de los medios de cultivo con vitamina B₁₂ puede satisfacer la necesidad de producir coenzima
B₁₂ para la actividad de glicerol deshidratasa en muchos microorganismos, pero en ciertos casos se puede tener que
añadir o incrementar *in vivo* actividades catalíticas adicionales. La síntesis incrementada de coenzima B₁₂ en eucariotas
puede ser especialmente deseable. Dadas las secuencias publicadas para los genes que codifican la cob(I)alamina
20 adenosiltransferasa, un experto en la técnica podría llevar a cabo la clonación y la expresión de este gen. Por ejemplo, se
considera que se podría construir una levadura, tal como *Saccharomyces*, para que contuviese los genes que codifican
cob(I)alamina adenosiltransferasa además de los genes necesarios para llevar a cabo la conversión de un sustrato de
carbono tal como glucosa a 1,3-propanodiol. La clonación y la expresión de los genes para la reducción de cobalto
requieren una aproximación diferente. Esto se podría basar en una selección en *E. coli* en función del crecimiento
25 con etanolamina como la única fuente de N₂. En presencia de coenzima B₁₂, la etanolamina amoníaco-liasa permite
el crecimiento de las células en ausencia de otras fuentes de N₂. Si las células de *E. coli* contienen un gen clonado
para cob(I)alamina adenosiltransferasa y ADN clonado aleatoriamente de otro organismo, se debería incrementar y
seleccionar el crecimiento con etanolamina en presencia de acuocobalamina si el ADN clonado aleatoriamente codifica
propiedades de reducción de cobalto para facilitar la adenosilación de acuocobalamina.

30 Además de *E. coli* y *Saccharomyces*, *Klebsiella* es un hospedador especialmente preferido. Se sabe que las cepas
de *Klebsiella pneumoniae* producen 1,3-propanodiol cuando se cultivan con glicerol como la única fuente de carbono.
Se considera que *Klebsiella* se puede alterar genéticamente para producir 1,3-propanodiol a partir de monosacáridos,
oligosacáridos, polisacáridos, o sustratos de un carbono.

35 Para modificar tales cepas, será ventajoso suministrar al hospedador de *Klebsiella* los genes que facilitan la con-
versión de dihidroxiacetona fosfato a glicerol y la conversión de glicerol a 1,3-propanodiol por separado o juntos, bajo
el control transcripcional de uno o más promotores constitutivos o inducibles. La introducción de los genes DAR1
y GPP2 que codifican la glicerol-3-fosfato deshidrogenasa y glicerol-3-fosfatasa, respectivamente, proporcionará a
40 *Klebsiella* la maquinaria genética para producir 1,3-propanodiol a partir de un sustrato de carbono apropiado.

Los genes (p. ej., de G3PDH, G3P fosfatasa, *dhaB* y/o *dhaT*) se pueden introducir en cualquier vector plasmídico
capaz de replicarse en *K. pneumoniae*, o se pueden integrar en el genoma de *K. pneumoniae*. Por ejemplo, se sabe que
K. pneumoniae ATCC 25955 y *K. pneumoniae* ECL 2106 son sensibles a tetraciclina o cloranfenicol; así, se pueden
45 usar vectores plasmídicos que sean capaces tanto de replicarse en *K. pneumoniae* como de codificar la resistencia a uno
o ambos antibióticos para introducir estos genes en *K. pneumoniae*. Los métodos de transformación de *Klebsiella* con
genes de interés son habituales y bien conocidos en la técnica, y se pueden hallar protocolos adecuados, que incluyen
vectores y técnicas de expresión apropiadas, en Sambrook, anteriormente mencionado.

50 Vectores y casetes de expresión

La presente invención se puede usar conjuntamente con una diversidad de vectores y casetes de transformación y
de expresión adecuados para la clonación, la transformación y la expresión de G3PDH y G3P fosfatasa en una célula
55 hospedadora adecuada. Los vectores adecuados serán aquellos que sean compatibles con la bacteria empleada. Los
vectores adecuados pueden proceder, por ejemplo, de una bacteria, un virus (tal como el bacteriófago T7 o un fago
derivado de M-13), un cósmido, una levadura o una planta. Los expertos en la técnica conocen los protocolos para la
obtención y el uso de tales vectores. (Sambrook *et al.*, Molecular Cloning: A Laboratory Manual - volúmenes 1,2,3
(Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, (1989)).

60 En general, el vector o el casete contiene las secuencias que dirigen la transcripción y la traducción del gen relevan-
te, un marcador seleccionable, y las secuencias que permiten la replicación autónoma o la integración cromosómica.
Los vectores adecuados comprenden una región en 5' del gen que alberga los controles de la iniciación transcripcio-
nal y una región en 3' del fragmento de ADN que controla la terminación transcripcional. Preferentemente, ambas
65 regiones de control proceden de genes homólogos de la célula hospedadora transformada, aunque se debe entender
que no es necesario que tales regiones de control procedan de los genes nativos de la especie específica elegida como
hospedador para la producción.

Las regiones o promotores de control de la iniciación que son útiles para controlar la expresión de los genes de G3PDH y G3P fosfatasa en la célula hospedadora deseada son numerosos y conocidos para los expertos en la técnica. Prácticamente cualquier promotor capaz de controlar estos genes es adecuado para la presente invención, lo que incluye, pero sin limitación, CYC1, HIS3, GAL1, GAL10, ADH1, PGK, PHO5, GAPDH, ADC1, TRP1, URA3, LEU2, ENO, TPI (útiles para la expresión en *Saccharomyces*); AOX1 (útil para la expresión en *Pichia*); y lac, trp, λP_L , λP_R , T7, tac, y trc (útiles para la expresión en *E. coli*).

Las regiones de control de la terminación pueden proceder también de diversos genes nativos para los hospedadores preferidos. Opcionalmente, puede ser innecesario un sitio de terminación, sin embargo, se prefiere su inclusión.

Para la expresión eficaz de las presentes enzimas, el ADN que codifica las enzimas se une de manera operable por medio de codones de iniciación a las regiones de control de la expresión seleccionadas, de forma que la expresión da como resultado la formación del ARN mensajero apropiado.

Transformación de hospedadores adecuados y expresión de genes para la producción de 1,3-propanodiol

Una vez que se construyen casetes adecuados, se usan para transformar las células hospedadoras apropiadas. La introducción del casete que contiene los genes que codifican la glicerol-3-fosfato deshidrogenasa (G3PDH) y glicerol-3-fosfatasa (G3P fosfatasa), glicerol deshidratasa (*dhaB*), y 1,3-propanodiol oxidoreductasa (*dhaT*), por separado o juntos en la célula hospedadora se puede llevar a cabo mediante procedimientos conocidos, tales como mediante transformación (p. ej., mediante el uso de células permeabilizadas con calcio, electroporación) o mediante transfección con el uso de un virus bacteriófago recombinante. (Sambrook *et al.*, anteriormente mencionado).

En la presente invención, se creó *E. coli* W2042 (ATCC 98188) que contenía los genes que codifican la glicerol-3-fosfato deshidrogenasa (G3PDH) y glicerol-3-fosfatasa (G3P fosfatasa), glicerol deshidratasa (*dhaB*), y 1,3-propanodiol oxidoreductasa (*dhaT*). Además, se construyó *S. cerevisiae* YPH500 (ATCC 74392) que alberga los plásmidos pMCK10, pMCK17, pMCK30 y pMCK35 que contienen los genes que codifican la glicerol-3-fosfato deshidrogenasa (G3PDH) y glicerol-3-fosfatasa (G3P fosfatasa), glicerol deshidratasa (*dhaB*), y 1,3-propanodiol oxidoreductasa (*dhaT*). Las *E. coli* y *Saccharomyces* anteriormente mencionadas representan realizaciones preferidas de la invención.

Medio y sustratos de carbono

Los medios de fermentación en la presente invención deben contener sustratos de carbono adecuados. Los sustratos adecuados pueden incluir, pero sin limitación, monosacáridos tales como glucosa y fructosa, oligosacáridos tales como lactosa o sacarosa, polisacáridos tales como almidón o celulosa, o sus mezclas, y mezclas no purificadas a partir de materias primas renovables tales como filtrado de suero de queso, licor de maíz fermentado, melazas de remolacha azucarera y malta de cebada. Además, el sustrato de carbono puede ser también sustratos de un carbono tales como dióxido de carbono o metanol, para los cuales se ha demostrado una conversión metabólica en intermedios bioquímicos claves.

Se ha informado la producción de glicerol a partir de fuentes de un carbono (p. ej., metanol, formaldehído, o formiato) en levaduras metilótrofas (Yamada *et al.*, Agric. Biol. Chem., 53(2) 541-543, (1989)) y en bacterias (Hunter *et al.*, Biochemistry, 24, 4148-4155, (1985)). Estos organismos pueden asimilar compuestos de un solo carbono, que varían en el estado de oxidación desde el metano hasta el formiato y producen glicerol. La ruta de la asimilación del carbono puede ser a través de monofosfato de ribulosa, a través de serina o a través de monofosfato de xilulosa (Gottschalk, Bacterial Metabolism, segunda edición, Springer-Verlag: Nueva York (1986)). La ruta de monofosfato de ribulosa implica la condensación de formiato con ribulosa-5-fosfato para formar un hidrato de carbono de 6 carbonos que se convierte en fructosa y opcionalmente el producto de tres carbonos gliceraldehído-3-fosfato. Asimismo, la ruta de serina asimila el compuesto de un carbono en la ruta glucolítica a través del metilentetrahidrofolato.

Además de la utilización de sustratos de uno y dos carbonos, se sabe que los organismos metilótrofos utilizan otros compuestos que contienen carbono tales como metilamina, glucosamina y una diversidad de aminoácidos para la actividad metabólica. Por ejemplo, las levaduras metilótrofas son conocidas por utilizar el carbono de la metilamina para formar trehalosa o glicerol (Bellion *et al.*, Microb. Growth C1 Compd., [Int. Symp.], 7^a (1993), 415-32. Editor(es): Murrell, J. Collin; Kelly, Don P. Publisher: Intercept, Andover, UK). De manera similar, diversas especies de *Candida* metabolizarán alanina o ácido oleico (Sulter *et al.*, Arch. Microbiol., 153(5), 485-9 (1990)). Por lo tanto, la fuente de carbono utilizada en la presente invención puede abarcar una amplia diversidad de sustratos que contienen carbono, y estará limitada solamente por los requerimientos del organismo hospedador.

Aunque se considera que todos los sustratos de carbono mencionados anteriormente y las mezclas de los mismos son adecuados en la presente invención, los sustratos de carbono preferidos son monosacáridos, oligosacáridos, polisacáridos, y sustratos de un carbono. Los más preferidos son hidratos de carbono tales como glucosa, fructosa, sacarosa y los sustratos de un único átomo de carbono, tales como metanol y dióxido de carbono. El más preferido es glucosa.

ES 2 336 858 T3

Además de una fuente de carbono apropiada, el medio de fermentación debe contener minerales, sales, cofactores, tampones y otros componentes adecuados, conocidos por los expertos en la técnica, adecuados para el crecimiento de los cultivos y la promoción de la ruta enzimática necesaria para la producción de glicerol. Se presta particular atención a las sales de Co (II) y/o a la vitamina B₁₂ o a sus precursores.

5

Condiciones de cultivo

En general, las células se cultivan a 30°C en un medio apropiado. Los medios de cultivo preferidos en la presente invención son medios habituales preparados comercialmente tales como el caldo de cultivo Luria Bertani (LB), el caldo de cultivo con dextrosa Sabouraud (SD) o el caldo de cultivo con extracto de malta de levadura (YM). También pueden utilizarse otros medios de cultivo definidos o sintéticos, y el medio apropiado para el cultivo del microorganismo específico será conocido para un experto en la técnica de microbiología o la ciencia de fermentación. También se puede incorporar el uso de agentes que se sabe que modulan la represión mediante catabolitos directamente o indirectamente, p. ej. 2':3'-monofosfato de adenosina cíclico o 2':5'-monofosfato de adenosina cíclico, en el medio de reacción. De manera similar, se pueden usar agentes que se sabe que modulan las actividades enzimáticas (p. ej., sulfitos, bisulfitos y álcalis) que conducen al incremento de la producción de glicerol junto con, o como alternativa, a las manipulaciones genéticas.

El intervalo de pH adecuado para la fermentación es de pH 5,0 a pH 9,0, en donde se prefiere como intervalo de pH 6,0 a pH 8,0 para el estado inicial.

Las reacciones pueden realizarse en condiciones aerobias o anaerobias donde se prefieren las condiciones anaerobias o microaerobias.

25

Fermentaciones por lotes y continuas

El presente procedimiento emplea un método de fermentación por lotes. Una fermentación por lotes clásica es un sistema cerrado donde la composición del medio se fija al comienzo de la fermentación y no está sujeta a alteraciones artificiales durante la fermentación. Por lo tanto, al comienzo de la fermentación el medio se inocula con el organismo u organismos deseados y se deja que se produzca la fermentación sin añadir nada al sistema. Por lo general, sin embargo, una fermentación por lotes es "por lotes" con respecto a la adición de la fuente de carbono, y a menudo se hacen intentos para controlar factores tales como el pH y la concentración de oxígeno. El metabolito y las composiciones de la biomasa del sistema por lotes cambian constantemente hasta el momento en el que se interrumpe la fermentación. En los cultivos por lotes las células pasan por una fase de latencia estática hasta una fase logarítmica de alto crecimiento, y por último hasta una fase estacionaria donde el ritmo de crecimiento disminuye o se detiene. Si no se tratan, las células en la fase estacionaria finalmente morirán. Las células en fase logarítmica generalmente son responsables de la mayor parte de la producción del producto final o intermedio.

40

Una variación del sistema por lotes estándar es el sistema de fermentación semicontinuo, que también es adecuado para la presente invención. En esta variación de un sistema por lotes típico, el sustrato se añade por incrementos a medida que progresa la fermentación. Los sistemas semicontinuos son útiles cuando la represión de catabolitos es apta para inhibir el metabolismo de las células y cuando se desea tener cantidades limitadas de sustrato en el medio. La medición de la concentración real de sustrato en sistemas semicontinuos es difícil y se estima por consiguiente sobre la base de los cambios de factores medibles tales como el pH, el oxígeno disuelto y la presión parcial de los gases residuales, tales como el CO₂. Las fermentaciones por lotes y semicontinuas son habituales y bien conocidas en la técnica, y se pueden encontrar ejemplos en Brock, anteriormente mencionado.

También se contempla que el método sea adaptable a los métodos de fermentación continua. La fermentación continua es un sistema abierto donde un medio de fermentación definido se añade continuamente a un biorreactor y una cantidad igual de medio acondicionado se elimina simultáneamente durante el tratamiento. La fermentación continua generalmente mantiene los cultivos a una alta densidad constante en la que las células están principalmente en crecimiento en fase logarítmica.

55

La fermentación continua permite la modulación de un factor o de cualquier número de factores que afectan al crecimiento celular o a la concentración del producto final. Por ejemplo, un método mantendrá un nutriente limitante tal como la fuente de carbono o la concentración de nitrógeno en una cantidad fija, y permitirá moderar todos los demás parámetros. En otros sistemas, numerosos factores que afectan al crecimiento pueden alterarse continuamente mientras que la concentración celular, medida por la turbidez media, se mantenga constante. Los sistemas continuos se esfuerzan en mantener las condiciones de crecimiento del estado estacionario y así la pérdida celular debida al medio que se está extrayendo debe equilibrarse frente a la velocidad de crecimiento celular en la fermentación. Los métodos de modulación de los nutrientes y de los factores de crecimiento para los procedimientos de fermentación continua así como las técnicas para maximizar la velocidad de formación del producto son bien conocidos en la técnica de la microbiología industrial y varios métodos son detallados por Brock, anteriormente mencionado.

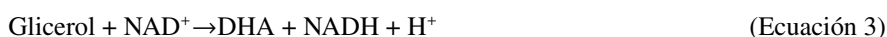
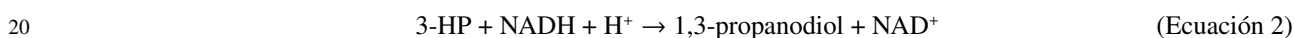
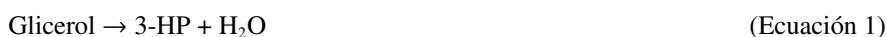
65

La presente invención se puede poner en práctica mediante el uso de los procedimientos por lotes, semicontinuos o continuos, y cualquier modo conocido de fermentación sería adecuado. Además, se contempla que las células pueden

inmovilizarse en un sustrato como catalizadores de células completas, y someterse a condiciones de fermentación para la producción de 1,3-propanodiol.

5 Alteraciones en la ruta de producción de 1,3-propanodiol

Ruta enzimática representativa. La producción de 1,3-propanodiol a partir de glucosa puede realizarse mediante la siguiente serie de etapas. Esta serie es representativa de numerosas rutas conocidas por los expertos en la técnica. La glucosa es convertida en una serie de etapas por enzimas de la ruta glucolítica en dihidroxiacetona fosfato (DHAP) y 3-fosfogliceraldehído (3-PG). El glicerol se forma a continuación por hidrólisis de DHAP a dihidroxiacetona (DHA) seguido de reducción, o reducción de DHAP a glicerol 3-fosfato (G3P) seguido de hidrólisis. La etapa de hidrólisis puede ser catalizada por cualquier número de fosfatasa celulares que se sabe que son específicas o inespecíficas con respecto a sus sustratos, o la actividad puede introducirse en el hospedador mediante recombinación. La etapa de reducción puede ser catalizada por una enzima del hospedador asociada a NAD⁺ (o NADP⁺), o la actividad puede ser introducida en el hospedador por recombinación. Es de destacar que el regulón *dha* contiene una glicerol deshidrogenasa (E.C. 1.1.1.6) que cataliza la reacción reversible de la Ecuación 3.



El glicerol se convierte en 1,3-propanodiol mediante el compuesto intermedio 3-hidroxi-propionaldehído (3-HP) como se ha descrito con detalle anteriormente. El intermedio 3-HP se produce a partir de glicerol (Ecuación 1) mediante una enzima deshidratasa que puede estar codificada por el hospedador, o puede introducirse en el hospedador mediante recombinación. Esta deshidratasa puede ser la glicerol deshidratasa (E.C. 4.2.1.30), diol deshidratasa (E.C. 4.2.1.28), o cualquier otra enzima capaz de catalizar esta transformación. El regulón *dha* codifica la glicerol deshidratasa, pero no la diol deshidratasa. El 1,3-propanodiol se produce a partir de 3-HP (Ecuación 2), mediante una enzima del hospedador asociada a NAD⁺ (o NADP⁺), o la actividad se puede introducir en el hospedador mediante recombinación. Esta reacción final en la producción de 1,3-propanodiol puede ser catalizada por la 1,3-propanodiol deshidrogenasa (E.C. 1.1.1.202) u otras alcohol deshidrogenasas.

Mutaciones y transformaciones que afectan a la canalización del carbono. Una variedad de organismos mutantes que comprende variaciones en la ruta de producción de 1,3-propanodiol será útil en la presente invención. La introducción de una mutación en una triosa fosfato isomerasa (*tpi*-) en el microorganismo es un ejemplo del uso de una mutación para mejorar el rendimiento mediante la canalización del carbono. De manera alternativa, las mutaciones que disminuyen la producción de etanol (*adh*) o de lactato (*ldh*) incrementarán la disponibilidad de NADH para la producción de 1,3-propanodiol. Las mutaciones adicionales en las etapas de las glucólisis después del gliceraldehído-3-fosfato tales como la fosfoglicerato mutasa (*pgm*) serían útiles para incrementar el flujo de carbono hacia la ruta de producción de 1,3-propanodiol. Las mutaciones que afectan al transporte de glucosa tales como PTS, que evitarían la pérdida de PEP, también resultarían útiles. Las mutaciones que bloquean las rutas alternativas para los intermedios de la producción de 1,3-propanodiol tales como la ruta catabólica de glicerol (*glp*) también serían útiles para la presente invención. La mutación puede dirigirse hacia un gen estructural para deteriorar o mejorar la actividad de una actividad enzimática, o puede dirigirse hacia un gen regulador para modular el nivel de expresión de una actividad enzimática.

Alternativamente, las transformaciones y las mutaciones pueden combinarse para controlar actividades enzimáticas específicas para la mejora de la producción de 1,3-propanodiol. Así está dentro del alcance de la presente invención anticipar modificaciones de un catalizador de célula completa que conducen a un aumento de la producción de 1,3-propanodiol.

Identificación y purificación de 1,3-propanodiol

Los métodos para la purificación de 1,3-propanodiol a partir del medio de fermentación son conocidos en la técnica. Por ejemplo, los propanodiolos se pueden obtener a partir del medio celular sometiendo la mezcla de reacción a extracción con un disolvente orgánico, destilación y cromatografía en columna (documento U.S. 5.356.812). Un disolvente orgánico particularmente adecuado para este procedimiento es el ciclohexano (documento U.S. 5.008.473).

El 1,3-propanodiol puede identificarse directamente sometiendo el medio a análisis de cromatografía líquida a alta presión (HPLC). En la presente invención se prefiere un método en el que el medio de fermentación se analiza en una columna analítica de intercambio iónico mediante el uso de una fase móvil de ácido sulfúrico 0,01 N de forma isocrática.

ES 2 336 858 T3

Identificación y purificación de G3PDH y G3P fosfatasa

Los niveles de expresión de las proteínas G3PDH y G3P fosfatasa se miden mediante ensayos enzimáticos, el ensayo de la actividad de G3PDH se basa en las propiedades espectrales del cosustrato, NADH, en la conversión de DHAP a G-3-P. NADH tiene una absorción UV/vis intrínseca, y su consumo se puede monitorizar de manera espectrofotométrica a 340 nm. La actividad de la G3P fosfatasa se puede medir mediante cualquier método de medida del fosfato inorgánico liberado en la reacción. El método de detección usado más habitualmente utiliza la determinación espectroscópica de un complejo fosfomolibdato-amonio de color azul.

10

Ejemplos

Métodos generales

15 Los procedimientos para fosforilaciones, ligaduras y transformaciones son bien conocidos en la técnica. Las técnicas adecuadas para su utilización en los ejemplos siguientes pueden encontrarse en Sambrook, J. *et al.*, Molecular Cloning: A Laboratory Manual, Segunda Edición, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY (1989).

20 Los materiales y métodos adecuados para el mantenimiento y crecimiento de cultivos bacterianos son bien conocidos en la técnica. Las técnicas adecuadas para su uso en los siguientes ejemplos se pueden encontrar en el *Manual of Methods for General Bacteriology* (Phillipp Gerhardt, R. G. E. Murray, Ralph N. Costilow, Eugene W. Nester, Willis A. Wood, Noel R. Krieg y G. Briggs Phillips, eds), American Society for Microbiology, Washington, DC. (1994)) o en Thomas D. Brock en *Biotechnology: A Textbook of Industrial Microbiology*, Segunda Edición, Sinauer Associates, Inc., Sunderland, MA (1989). Todos los reactivos y materiales utilizados para el cultivo y el mantenimiento de las células bacterianas se adquirieron de Aldrich Chemicals (Milwaukee, WI), DIFCO Laboratories (Detroit, MI), GIBCOBRL (Gaithersburg, MD), o Sigma Chemical Company (St. Louis, MO), a menos que se indique de otra manera.

25 El significado de las abreviaturas es el siguiente: “h” significa hora(s), “min” significa minuto(s), “seg” significa segundo(s), “d” significa día(s), “mL” significa mililitros, “L” significa litros.

30

Ensayos enzimáticos

35 Se determinó la actividad de la glicerol deshidratasa en extractos exentos de células mediante el uso de 1,2-propanodiol como sustrato. El ensayo, basado en la reacción de aldehídos con metilbenzo-2-tiazolona hidrazona, ha sido descrito por Forage y Foster (*Biochim. Biophys. Acta*, 569, 249 (1979)). La actividad de la 1,3-propanodiol oxidoreductasa, a veces denominada 1,3-propanodiol deshidrogenasa, se determinó en solución o en bloques de geles utilizando 1,3-propanodiol y NAD⁺ como sustratos como se ha descrito también. Johnson y Lin, *J. Bacteriol.*, 169, 2050 (1987). La actividad de la glicerol 3-fosfato deshidrogenasa (G3PDH) dependiente de NADH o NADPH se determinó de manera espectrofotométrica, siguiendo la desaparición de NADH o NADPH tal como se ha descrito. (R. M. Bell y J. E. Cronan, Jr., *J. Biol. Chem.* 250:7153-8 (1975)).

40

45 *Ensayo de glicerol-3-fosfatasa, GPP*

El ensayo de la actividad enzimática se llevó a cabo incubando el extracto con un sustrato de fosfato orgánico en un tampón bis-Tris o MES y magnesio, pH 6,5. El sustrato usado fue 1- α -glicerol fosfato; d,1- α -glicerol fosfato. Las concentraciones finales de los reactivos en el ensayo son: tampón (bis-Tris 20 mM o MES 50 mM); MgCl₂ (10 mM); y sustrato (20 mM). Si la proteína total de la muestra fue baja y no se dio una precipitación visible con un reactivo ácido, la muestra se ensayó convenientemente en la cubeta. Este método implicó incubar una muestra de enzima en una cubeta que contenía sustrato 20 mM (50 μ L, 200 mM), tampón MES 50 mM, MgCl₂ 10 mM, pH 6,5. El volumen final del ensayo de fosfatasa fue 0,5 mL. La muestra que contenía la enzima se añadió a la mezcla de reacción; el contenido de la cubeta se mezcló y después se colocó la cubeta en un baño de agua en circulación a T = 37°C durante 5 a 120 min, dependiendo de si la actividad de la fosfatasa en la muestra de enzima oscilaba de 2 a 0,02 U/mL. La reacción enzimática se paró mediante la adición del reactivo de molibdato ácido (0,4 mL). Después de añadir el reactivo de Fiske SubbaRow (0,1 mL) y agua destilada (1,5 mL), la disolución se mezcló y se dejó revelar. Después de 10 min, se leyó la absorbancia de las muestras a 660 nm mediante el uso de un espectrofotómetro Cary 219 UV/Vis. La cantidad de fosfato inorgánico liberado se comparó con una curva patrón que se preparó mediante el uso de una disolución de reserva de fosfato inorgánico (0,65 mM) y preparando 6 patrones con concentraciones finales de fosfato inorgánico que oscilaban de 0,026 a 0,130 μ mol/mL.

60

Aislamiento e identificación de 1,3-propanodiol

65

La conversión de glicerol en 1,3-propanodiol se controló mediante HPLC. Se realizaron análisis utilizando técnicas habituales y materiales disponibles para un experto en la técnica de cromatografía. Un método apropiado utilizó un sistema de HPLC Waters Maxima 820 que utiliza UV (210 nm) y detección por IR. Se inyectaron muestras sobre una

ES 2 336 858 T3

columna Shodex SH-1011 (8 mm x 300 mm, adquirida en Waters, Milford, MA) equipada con una precolumna Shodex SH-1011P (6 mm x 50 mm), temperatura controlada a 50°C, utilizando H₂SO₄ 0,01 N como fase móvil a un caudal de 0,5 ml/min. Cuando se deseaba un análisis cuantitativo, las muestras se preparaban con una cantidad conocida de ácido trimetilacético como patrón externo. Por lo general, los tiempos de retención del glicerol (detección por IR), 1,3-propanodiol (detección por IR) y ácido trimetilacético (UV y detección por IR) fueron 20,67 min., 26,08 min. y 35,03 min., respectivamente.

La producción de 1,3-propanodiol se confirmó por GC/MS. Se realizaron análisis utilizando técnicas normalizadas y materiales disponibles para un experto en materia de GC/MS. Un método apropiado utilizó un cromatógrafo de gases Hewlett Packard 5890 serie II acoplado a un detector selectivo de masas Hewlett Packard 5971 serie (EI) y una columna HP-INNOWax (30 m de longitud, 0,25 mm d.i., 0,25 micras de espesor de película). El tiempo de retención y el espectro de masas de 1,3-propanodiol generado se compararon con el del 1,3-propanodiol auténtico (m/e: 57, 58).

Un método alternativo para GC/MS implicaba la modificación de la muestra. A 1,0 ml de muestra (p. ej., sobrenadante del cultivo) se añadió 30 µl de ácido perclórico concentrado (70% v/v). Después de mezclar, se congeló y se liofilizó la muestra. Se añadió una mezcla 1:1 de bis(trimetilsilil)trifluoroacetamida:piridina (300 µl) al material liofilizado, se mezcló intensamente y se colocó a 65°C durante una h. La muestra se clarificó del material insoluble mediante centrifugación. El líquido resultante se dividió en dos fases, de las cuales la superior se utilizó para el análisis. La muestra se cromatografió en una columna DB-5 (48 m, 0,25 mm de D.I., grosor de película 0,25 µm; de J&W Scientific) y el tiempo de retención y el espectro de masas del derivado de 1,3-propanodiol obtenido de los sobrenadantes del cultivo se compararon con los obtenidos a partir de patrones auténticos. El espectro de masas de 1,3-propanodiol modificado con TMS contiene los iones característicos de 205, 177, 130 y 115 AMU.

Ejemplo 1

Clonación y transformación de células hospedadoras de E. coli con ADN de cósmido para la expresión de 1,3-propanodiol

Medios

Se utilizó medio sintético S12 en la identificación de la capacidad de los transformantes bacterianos para producir 1,3-propanodiol. El medio S12 contiene: sulfato amónico 10 mM, tampón de fosfato potásico 50 mM, pH 7,0, MgCl₂ 2 mM, CaCl₂ 0,7 mM, MnCl₂ 50 µM, FeCl₃ 1 µM, ZnCl 1 µM, CuSO₄ 1,7 µM, CoCl₂ 2,5 µM, Na₂MoO₄ 2,4 µM, e hidrocloreuro de tiamina 2 µM.

El medio A utilizado para el crecimiento y la fermentación consistía en: sulfato amónico 10 mM; tampón de MOPS/KOH 50 mM, pH 7,5; tampón de fosfato potásico 5 mM, pH 7,5; MgCl₂ 2 mM; CaCl₂ 0,7 mM; MnCl₂ 50 µM; FeCl₃ 1 µM; ZnCl 1 µM; CuSO₄ 1,72 µM; CoCl₂ 2,53 µM; Na₂MoO₄ 2,42 µM; hidrocloreuro de tiamina 2 µM; extracto de levadura al 0,01%; casaminoácidos al 0,01%; 0,8 µg/ml de vitamina B₁₂; y 50 µg/ml de amp. El medio A se complementó con glicerol al 0,2% o glicerol al 0,2% más D-glucosa al 0,2%, según fue necesario.

Células

La cepa ECL2106 de *Klebsiella pneumoniae* (Ruch *et al.*, J. Bacteriol., 124, 348 (1975)), también conocida en la bibliografía como *K. aerogenes* o *Aerobacter aerogenes*, se obtuvo de E. C. C. Lin (Harvard Medical School, Cambridge, MA) y se mantuvo como cultivo de laboratorio.

ATCC 25955 de *Klebsiella pneumoniae* se adquirió en American Type Culture Collection (Rockville, MD).

DH5α de *E. coli* se adquirió en Gibco/BRL y se transformó con el ADN de cósmido aislado de ATCC 25955 de *Klebsiella pneumoniae* que contiene un gen que codifica una enzima glicerol o diol deshidratasa. Los cósmidos que contienen la glicerol deshidratasa se identificaron como pKP1 y pKP2 y el cósmido que contiene la enzima diol deshidratasa se identificó como pKP4. Las células DH5α transformadas se identificaron como DH5α-pKP1, DH5α-pKP2 y DH5α-pKP4.

La cepa ECL707 de *E. coli* (Sprenger *et al.*, J. Gen. Microbiol., 135, 1255 (1989)) se obtuvo de E. C. C. Lin (Harvard Medical School, Cambridge, MA) y se transformó de manera similar con el ADN de cósmido de *Klebsiella pneumoniae*. Estos transformantes fueron identificados como ECL707-pKP1 y ECL707-pKP2, que contienen el gen de la glicerol deshidratasa y ECL707-pKP4 que contiene el gen de la diol deshidratasa.

La cepa AA200 de *E. coli* que contiene una mutación en el gen *tpi* (Anderson *et al.*, J. Gen. Microbiol., 62, 329 (1970)) se adquirió en el *E. coli* Genetic Stock Center, Yale University (New Haven, CT) y se transformó con ADN de cósmido de *Klebsiella* para dar los organismos recombinantes AA200-pKP1 y AA200-pKP2, que contienen el gen de la glicerol deshidratasa y AA200-pKP4, que contiene el gen de la diol deshidratasa.

DH5 α

Seis placas de transformación que contienen aproximadamente 1.000 colonias de XL1-Blue MR de *E. coli* transfectadas con ADN de *K. pneumoniae* se lavaron con 5 ml de medio LB y se centrifugaron. Se sedimentaron las bacterias y se volvieron a poner en suspensión en 5 ml de medio LB + glicerol. Se inoculó una alícuota (50 μ l) en un tubo de 15 ml que contenía medio sintético S 12 con glicerol al 0,2% + 400 ng por ml de vitamina B₁₂ + extracto de levadura al 0,001% + 50 amp. El tubo se llenó con el medio hasta arriba y se envolvió con parafilm y se incubó a 30°C. Se observó una ligera turbidez tras 48 h. Las alícuotas, analizadas para la distribución del producto como se describió anteriormente a las 78 h y 132 h, fueron positivas para 1,3-propanodiol, los últimos puntos de tiempo que contienen cantidades crecientes de 1,3-propanodiol.

Las bacterias, positivas a la prueba para la producción de 1,3-propanodiol, se diluyeron en serie y se colocaron en placas LB-50amp a fin de aislar colonias individuales. Se aislaron cuarenta y ocho colonias individuales y se comprobó de nuevo la producción de 1,3-propanodiol. El ADN de cósmido se aisló de 6 clones independientes y se transformó en la cepa de *E. coli* DH5 α . De nuevo se comprobó la producción de 1,3-propanodiol en los transformantes. Se caracterizaron adicionalmente dos transformantes y se denominaron DH5 α -pKP1 y DH5 α -pKP2.

Se secuenció un fragmento EcoRI-Sall de 12,1 kb de pKP1, se subclonó en pIBI31 (IBI Biosystem, New Haven, CT), y se denominó pHK28-26 (SEQ ID NO: 19). El secuenciado puso de manifiesto los locus de los marcos de lectura abiertos relevantes del operón *dha* que codifica la glicerol deshidratasa y los genes necesarios para la regulación. Respecto a SEQ ID NO: 19, se halla un fragmento del marco de lectura abierto de *dhaK* que codifica la dihidroxiacetona quinasa en las bases 1-399; el marco de lectura abierto *dhaD* que codifica la glicerol deshidrogenasa se encuentra en las bases 983-2107; el marco de lectura abierto *dhaR* que codifica el represor se encuentra en las bases 2209-4134; el marco de lectura abierto *dhaT* que codifica la 1,3-propanodiol oxidorreductasa se encuentra en las bases 5017-6180; el marco de lectura abierto *dhaB1* que codifica la subunidad alfa de la glicerol deshidratasa se encuentra en las bases 7044-8711; el marco de lectura abierto *dhaB2* que codifica la subunidad beta de la glicerol deshidratasa se encuentra en las bases 8724-9308; el marco de lectura abierto *dhaB3* que codifica la subunidad gamma de la glicerol deshidratasa se encuentra en las bases 9311-9736; y el marco de lectura abierto *dhaBX*, que codifica una proteína de función desconocida se encuentra en las bases 9749-11572.

Se inocularon colonias independientes de *E. coli* XL1-Blue MR transfectadas con ADN de cósmido empaquetado de *K. pneumoniae* en pocillos de microtitulación que contenían 200 μ l de medio S 15 (sulfato amónico, 10 mM; tampón de fosfato potásico pH 7,0, 1 mM; tampón de MOPS/KOH, pH 7,0, 50 mM; MgCl₂, 2 mM; CaCl₂, 0,7 mM; MnCl₂, 50 μ M; FeCl₃, 1 μ M; ZnCl₂, 1 μ M; CuSO₄, 1,72 μ M; CoCl₂, 2,53 μ M; Na₂MoO₄, 2,42 μ M; e hidrocloreuro de tiamina, 2 μ M) + glicerol al 0,2% + 400 ng/ml de vitamina B₁₂ + 0,001% de extracto de levadura + 50 μ g/ml de ampicilina. Además de los pocillos de microtitulación, se inoculó también una placa modelo que contenía amp LB-50. Tras 96 h, se extrajeron 100 μ l y se centrifugaron en un tubo de microcentrifugadora Rainin que contenía un filtro de membrana de nilón de 0,2 micras. Las bacterias quedaron retenidas y el filtrado se procesó para el análisis por HPLC. Los clones positivos que muestran la producción de 1,3-propanodiol se identificaron después de identificar aproximadamente 240 colonias. Se identificaron tres clones positivos, dos de los cuales habían crecido en amp LB-50 y uno de los cuales no. Una sola colonia, aislada a partir de uno de los dos clones positivos cultivados en amp LB-50 y en la que se verificó la producción de 1,3-propanodiol, se denominó pKP4. El ADN de cósmido se aisló a partir de cepas de *E. coli* que contenían pKP4 y se transformó la cepa DH5 α de *E. coli*. Se comprobó la producción de 1,3-propanodiol en un transformante independiente, denominado DH5 α -pKP4.

ECL707

La cepa ECL707 de *E. coli* se transformó con el ADN de cósmido de *K. pneumoniae* que correspondía a uno de pKP1, pKP2, pKP4 o el vector Supercos sólo y se denominaron ECL707-pKP1, ECL707-pKP2, ECL707-pKP4, y ECL707-sc, respectivamente. ECL707 es deficiente en *glpK*, *gld*, y *ptsD* que codifican la glicerol quinasa dependiente de ATP, la glicerol deshidrogenasa dependiente de NAD⁺, y la enzima II para la dihidroxiacetona del sistema fosfotransferasa dependiente del fosfoenolpiruvato, respectivamente.

Veinte colonias individuales de cada transformación con cósmido y cinco de la transformación con el vector Supercos sólo (control negativo), aisladas de placas LB-50 amp, se transfirieron a una placa principal LB-50 amp. En estos aislamientos se determinó asimismo la capacidad para convertir el glicerol en 1,3-propanodiol a fin de determinar si contenían actividad de deshidratasa. Los transformantes se transfirieron con un palillo estéril a placas de microtitulación que contenían 200 μ l de Medio A complementado con glicerol al 0,2% o glicerol al 0,2% más D-glucosa al 0,2%. Tras la incubación durante 48 h a 30°C, los contenidos de los pocillos de la placa de microtitulación se filtraron a través de un filtro de nailon de 0,45 micras y se sometieron a cromatografía mediante HPLC. Los resultados de estas pruebas se dan en la Tabla 1.

ES 2 336 858 T3

TABLA 1

Conversión de glicerol en 1,3-propanodiol mediante ECL707 transformadas

| | <u>Transformante</u> | <u>Glicerol*</u> | <u>Glicerol más Glucosa*</u> |
|----|----------------------|------------------|------------------------------|
| 5 | ECL707-pKP1 | 19/20 | 19/20 |
| 10 | ECL707-pKP2 | 18/20 | 20/20 |
| 15 | ECL707-pKP4 | 0/20 | 20/20 |
| | ECL707-sc | 0/5 | 0/5 |

* (Número de aislamientos positivos/número de aislamientos analizados)

AA200

La cepa AA200 de *E. coli* se transformó con el ADN de cósmido de *K. pneumoniae* que correspondía a uno de pKP1, pKP2, pKP4 y el vector Supercos sólo y se denominaron AA200-pKP1, AA200-pKP2, AA200-pKP4, y AA200-sc, respectivamente. La cepa AA200 es deficiente en triosafosfato isomerasa (*tpi*⁻).

Se aislaron veinte colonias individuales de cada transformación con cósmido y cinco de la transformación con vectores vacíos, y se determinó su capacidad para convertir el glicerol en 1,3-propanodiol como se describe para la cepa ECL707 de *E. coli*. Los resultados de estas pruebas se dan en la Tabla 2.

TABLA 2

Conversión de glicerol en 1,3-propanodiol mediante AA200 transformadas

| | <u>Transformante</u> | <u>Glicerol*</u> | <u>Glicerol más Glucosa*</u> |
|----|----------------------|------------------|------------------------------|
| 35 | AA200-pKP1 | 17/20 | 17/20 |
| 40 | AA200-pKP2 | 17/20 | 17/20 |
| 45 | AA200-pKP4 | 2/20 | 16/20 |
| | AA200-sc | 0/5 | 0/5 |

* (Número de aislamientos positivos/número de aislamientos analizados)

Ejemplo 2

Conversión de D-glucosa a 1,3-propanodiol mediante E. coli recombinante con el uso de DAR1, GPP2, dhaB, y dhaT

Construcción de plásmidos de expresión de utilidad general para el uso en la transformación de Escherichia coli

El vector de expresión pTacIQ

El vector de expresión de *E. coli*, pTacIQ, contiene el gen lacIq (Farabaugh, Nature 274, 5673 (1978)) y el promotor tac (Amann *et al.*, Gene 25, 167 (1983)) insertados en el sitio EcoRI de pBR322 (Sutcliffe *et al.*, Cold Spring Harb. Symp. Quant. Biol. 43, 77 (1979)). Un sitio de clonación múltiple y una secuencia terminadora (SEQ ID NO:20) sustituyen la secuencia de pBR322 de EcoRI a SphI.

ES 2 336 858 T3

Subclonación de los genes de la glicerol deshidratasa (dhaB1, 2, 3)

El marco de lectura abierto para el gen *dhaB3* (que incorpora un sitio EcoRI en el extremo 5' y un sitio XbaI en el extremo 3') se amplificó a partir de pHK28-26 mediante PCR con el uso de cebadores (SEQ ID NOS:21 y 22). El producto se subclonó en pLitmus29 (New England Biolab, Inc., Beverly, MA) para generar el plásmido pDHAB3 que contenía *dhaB3*.

La región que contenía la región codificante completa para los cuatro genes del operón *dhaB* de pHK28-26 se clonó en pBluescriptII KS+ (Stratagene, La Jolla, CA) mediante el uso de las enzimas de restricción KpnI y EcoRI para crear el plásmido pM7.

Se eliminó el gen *dhaBX* digiriendo el plásmido pM7, que contiene *dhaB(1, 2, 3, 4)*, con ApaI y XbaI (lo que elimina parte de *dhaB3* y la totalidad de *dhaBX*). El fragmento resultante de 5,9 kb se purificó y se ligó con el fragmento ApaI-XbaI de 325 pb del plásmido pDHAB3 (restableciendo el gen *dhaB3*) para crear el pM11, que contiene *dhaB(1,2,3)*.

El marco de lectura abierto para el gen *dhaB1* (que incorpora un sitio HindIII y un sitio de unión al ribosoma RBS consenso en el extremo 5' y un sitio XbaI en el extremo 3') se amplificó a partir de pHK28-26 mediante PCR con el uso de cebadores (SEQ ID NO:23 y SEQ ID NO:24). El producto se subclonó en pLitmus28 (New England Biolab, Inc.) para generar el plásmido pDT1 que contenía *dhaB1*.

Un fragmento NotI-XbaI procedente del pM11 que contenía parte del gen *dhaB1*, el gen *dhaB2* y el gen *dhaB3* se insertó en el pDT1 para crear el plásmido de expresión de *dhaB*, pDT2. El fragmento HindIII-XbaI que contenía los genes *dhaB(1,2,3)* del pDT2 se insertó en pTaciQ para crear el pDT3.

Subclonación del gen de la 1,3-propanodiol deshidrogenasa (dhaT)

El fragmento KpnI-SacI del pHK28-26, que contenía el gen (*dhaT*) completo de la 1,3-propanodiol deshidrogenasa, se subclonó en pBluescriptII KS+ creando el plásmido pAH1. El gen *dhaT* (que incorpora un sitio XbaI en el extremo 5' y un sitio BamHI en el extremo 3') se amplificó mediante PCR a partir de pAH1 como ADN molde mediante el uso de cebadores sintéticos (SEQ ID NO:25 con SEQ ID NO:26). El producto se subclonó en pCR-Script (Stratagene) en el sitio SrfI para generar los plásmidos pAH4 y pAH5 que contienen *dhaT*. El plásmido pAH4 contiene el gen *dhaT* en la orientación correcta para la expresión del promotor de lac en el pCR-Script y pAH5 contiene el gen *dhaT* en la orientación opuesta. El fragmento XbaI-BamHI del pAH4 que contiene el gen *dhaT* se insertó en el pTaciQ para generar el plásmido pAH8. El fragmento HindIII-BamHI del pAH8 que contiene el RBS y el gen *dhaT* se insertó en el pBluescriptII KS+ para crear el pAH11. El fragmento HindIII-SalI de pAH8 que contenía el RBS, el gen *dhaT* y el terminador se insertó en pBluescriptII SK+ para crear pAH 12.

Construcción de un casete de expresión para dhaB(1,2,3) y dhaT

Se construyó un casete de expresión para *dhaB(1, 2, 3)* y *dhaT* a partir de los subclones 1 *dhaB(1,2,3)* y *dhaT* individuales descritos anteriormente mediante el uso de métodos habituales de biología molecular. El fragmento SpeI-KpnI del pAH8 que contiene el RBS, el gen *dhaT* y el terminador se insertó en las zonas XbaI-KpnI de pDT3 para crear el pAH23. El fragmento SmaI-EcoRI entre el gen *dhaB3* y *dhaT* del pAH23 se eliminó para crear el pAH26. El fragmento SpeI-NotI que contiene una zona EcoRI del pDT2 se utilizó para reemplazar el fragmento SpeI-NotI del pAH26 para generar el pAH27.

Construcción de un casete de expresión para dhaT y dhaB(1,2,3)

Se construyó un casete de expresión para *dhaT* y *dhaB(1,2,3)* a partir de los subclones *dhaB(1, 2, 3)* y *dhaT* individuales descritos previamente mediante el uso de métodos habituales de biología molecular. Se insertó un fragmento SpeI-SacI que contenía los genes *dhaB(1,2, 3)* de pDT3 en pAH11 en los sitios SpeI-SacI para crear pAH24.

Clonación y expresión de la glicerol 3-fosfatasa para una producción de glicerol incrementada en E. coli

Se obtuvo el clon lambda 6592 del cromosoma V de *Saccharomyces cerevisiae* (Gene Bank, n° de registro U18813x11) de la ATCC. El gen de la glicerol 3-fosfato fosfatasa (GPP2) (que incorpora un sitio BamHI-RBS-XbaI en el extremo 5' y un sitio SmaI en el extremo 3') se clonó mediante PCR a partir del clon lambda como ADN objetivo mediante el uso de cebadores sintéticos (SEQ ID NO:27 con SEQ ID NO:28). El producto se subclonó en pCR-Script (Stratagene) en el sitio SrfI para generar los plásmidos pAH 15 que contenían la GPP2. El plásmido pAH 15 contiene el gen GPP2 en la orientación inactiva para la expresión a partir del promotor lac en pCR-Script SK+. El fragmento BamHI-SmaI de pAH15 que contiene el gen GPP2 se insertó en pBlueScriptII SK+ para generar el plásmido pAH19. El pAH19 contiene el gen GPP2 en la orientación correcta para la expresión a partir del promotor lac. El fragmento XbaI-PstI de pAH19 que contiene el gen GPP2 se insertó en pPHOX2 para crear el plásmido pAH21.

ES 2 336 858 T3

Plásmidos para la expresión de los genes dhaT, dhaB(1,2,3) y GPP2

Se insertó un ligador Sall-EcoRI-XbaI (SEQ ID NOS:29 y 30) en pAH5, que se digirió con las enzimas de restricción Sall-XbaI para crear pDT16. El ligador destruye el sitio XbaI. El fragmento Sall-MluI de 1 kb del pDT16 se insertó a continuación en el pAH24, sustituyendo al fragmento Sall-MluI existente para crear el pDT18.

El fragmento EcoRI-XbaI de 4,1 kb que contenía el casete de expresión para *dhaT* y *dhaB(1,2,3)* de pDT18 y el fragmento XbaI-Sall de 1,0 kb que contenía el gen GPP2 de pAH21 se insertó en el vector pMMB66EH (Füste *et al.*, *GENE*, 48, 119 (1986)) digerido con las enzimas de restricción EcoRI y Sall para crear pDT20.

Plásmidos para la sobreexpresión de DAR1 en E. coli

DAR1 se aisló mediante clonación por PCR a partir de ADN genómico de *S. cerevisiae* mediante el uso de cebadores sintéticos (SEQ ID NO:46 con SEQ ID NO:47). Los sitios de clonación mediante PCR eficaces son el sitio *NcoI* en el extremo 5' de DAR1 en donde el ATG de *NcoI* es la metionina de iniciación de DAR1. En el extremo 3' de DAR1 se introduce un sitio *BamHI* tras el terminador de la traducción. Los fragmentos de PCR se digirieron con *NcoI* + *BamHI* y se clonaron en los mismos sitios en el plásmido de expresión pTrc99A (Pharmacia, Piscataway, New Jersey) para proporcionar pDAR1A.

Para crear un sitio de unión al ribosoma más adecuado en el extremo 5' de DAR1, se insertó un ligador SpeI-RBS-*NcoI* obtenido renaturalizando cebadores sintéticos (SEQ ID NO:48 con SEQ ID NO:49) en el sitio *NcoI* de pDAR1A para crear pAH40. El plásmido pAH40 contiene el nuevo RBS y el gen DAR1 en la orientación correcta para la expresión a partir del promotor *trc* de Trc99A (Pharmacia). Se insertó el fragmento *NcoI*-*BamHI* de pDAR1A y un segundo grupo de ligadores SpeI-RBS-*NcoI* obtenidos renaturalizando cebadores sintéticos (SEQ ID NO:31 con SEQ ID NO:32) en el sitio SpeI-*BamHI* de pBluescript II-SK+ (Stratagene) para crear pAH41. La construcción pAH41 contiene un gen de resistencia a ampicilina. Se insertó el fragmento *NcoI*-*BamHI* de pDAR1A y un segundo grupo de ligadores SpeI-RBS-*NcoI* obtenidos renaturalizando cebadores sintéticos (SEQ ID NO:31 con SEQ ID NO:32) en el sitio SpeI-*BamHI* de pBC-SK+ (Stratagene) para crear pAH42. La construcción pAH42 contiene un gen de resistencia a cloranfenicol.

Construcción de un casete de expresión para DAR1 y GPP2

Se construyó un casete de expresión para DAR1 y GPP2 a partir de los subclones DAR1 y GPP2 individuales descritos anteriormente mediante el uso de métodos habituales de biología molecular. Se insertó el fragmento *BamHI*-*PstI* de pAH19 que contenía el RBS y el gen GPP2 en pAH40 para crear pAH43. Se insertó el fragmento *BamHI*-*PstI* de pAH19 que contenía el RBS y el gen DAR1 en pAH41 para crear pAH44. También se insertó el mismo fragmento *BamHI*-*PstI* de pAH19 que contenía el RBS y el gen GPP2 en pAH42 para crear pAH45.

Construcción de la cepa de E. coli

La cepa W1485 de *E. coli* es una cepa K-12 de tipo natural (ATCC 12435). Esta cepa se transformó con los plásmidos pDT20 y pAH42 y se seleccionó en placas LA (Luria Agar, Difco) complementadas con 50 µg/mL de carbencilina y 10 µg/mL de cloranfenicol.

Producción de 1,3-propanodiol a partir de glucosa

Se transfirieron W1485/pDT20/pAH42 de *E. coli* de una placa a 50 mL de un medio que contenía por litro: 22,5 g de glucosa, 6,85 g de K₂HPO₄, 6,3 g de (NH₄)₂SO₄, 0,5 g de NaHCO₃, 2,5 g de NaCl, 8 g de extracto de levadura, 8 g de triptona, 2,5 mg de vitamina B₁₂, 2,5 mL de disolución de oligoelementos de Balch modificada, 50 mg de carbencilina y 10 mg de cloranfenicol, pH final 6,8 (HCl), y después se esterilizó mediante filtración. La composición de la disolución de oligoelementos de Balch modificada puede encontrarse en *Methods for General and Molecular Bacteriology* (P. Gerhardt *et al.*, eds, pág. 158, American Society for Microbiology, Washington, DC (1994)). Después de incubar a 37°C, 300 rpm durante 6 h, se añadieron 0,5 g de glucosa e IPTG (concentración final = 0,2 mM) y la agitación se redujo a 100 rpm. Las muestras se analizaron mediante GC/MS. Después de 24 h, W1485/pDT20/pAH42 produjeron 1,1 g/L de glicerol y 195 mg/L de 1,3-propanodiol.

Ejemplo 3

*Clonación y expresión de dhaB y dhaT*5 *En Saccharomyces cerevisiae*

Se construyeron plásmidos de expresión que pudieran existir en forma de elementos episómicos replicantes para cada uno de los cuatro genes *dha*. Para todos los plásmidos de expresión, hubo presente un promotor de ADH1 de levadura y separado de un terminador de la transcripción de ADH1 de levadura por fragmentos de ADN que contenían
 10 sitios de reconocimiento para una o más endonucleasas de restricción. Cada plásmido de expresión contuvo también el gen de β -lactamasa para la selección en *E. coli* en medios que contenían ampicilina, un origen de replicación para el mantenimiento del plásmido en *E. coli*, y un origen de replicación de 2 micrones para el mantenimiento en *S. cerevisiae*. Los marcadores nutritivos seleccionables utilizados para levaduras y presentes en los plásmidos de expresión fueron uno de los siguientes: el gen HIS3 que codifica la imidazolglicerolfosfato deshidratasa, el gen
 15 URA3 que codifica la orotidina 5'-fosfato descarboxilasa, el gen TRP1 que codifica la N-(5'-fosforribosil)-antranilato isomerasa y el gen LEU2 que codifica la β -isopropilmalato deshidrogenasa.

Los marcos de lectura abiertos para *dhaT*, *dhaB3*, *dhaB2* y *dhaB1* se amplificaron a partir de pHK28-26 (SEQ ID NO: 19) mediante PCR con el uso de cebadores (SEQ ID NO:38 con SEQ ID NO:39, SEQ ID NO:40 con SEQ ID NO:41, SEQ ID NO:42 con SEQ ID NO:43, y SEQ ID NO:44 con SEQ ID NO:45 para *dhaT*, *dhaB3*, *dhaB2* y *dhaB1*, respectivamente) que incorporaban sitios EcoRI en los extremos 5' (Tris 10 mM pH 8,3, KCl 50 mM, MgCl₂ 1,5 mM, 0,0001% de gelatina, dATP 200 μ M, dCTP 200 μ M, dGTP 200 μ M, dTTP 200 μ M, 1 μ M de cada cebador, 1-10 ng de ADN objetivo, 25 unidades/mL de Amplitaq™ ADN polimerasa (Perkin-Elmer Cetus, Norwalk CT)). Los parámetros de la PCR fueron 1 min a 94°C, 1 min a 55°C, 1 min a 72°C, 35 ciclos. Los productos se subclonaron en el sitio EcoRI
 25 de pHL-D4 (Phillips Petroleum, Bartlesville, OK) para generar los plásmidos pMP13, pMP14, pMP20 y pMP15 que contenían *dhaT*, *dhaB3*, *dhaB2* y *dhaB1*, respectivamente.

30 *Construcción del plásmido de expresión dhaB1 pMCK10*

El plásmido de replicación pGADGH de 7,8 kb (Clontech, Palo Alto, CA) se digirió con HindIII, se desfosforiló, y se ligó al fragmento HindIII de *dhaB1* de pMP15. El plásmido resultante (pMCK10) tenía *dhaB1* orientado correctamente para la transcripción a partir del promotor ADH1 y contenía un marcador LEU2.

35 *Construcción del plásmido de expresión de dhaB2 pMCK17*

El plásmido pGADGH (Clontech, Palo Alto, CA) se digirió con HindIII y los extremos monocatenarios se convirtieron en extremos EcoRI por ligadura con los adaptadores HindIII-XmnI y EcoRI-XmnI (New England Biolabs, Beverly, MA). La selección de plásmidos con extremos EcoRI correctos se consiguió mediante ligadura a un gen de resistencia a la kanamicina en un fragmento EcoRI del plásmido pUC4K (Pharmacia Biotech, Uppsala), transformación en la cepa DH5 α de *E. coli* y selección en placas LB que contenían 25 μ g/ml de kanamicina. El plásmido resultante (pGAD/KAN2) se digirió con SnaBI y EcoRI y se aisló un fragmento de 1,8 kb con el promotor de ADH1. Se digirió el plásmido pGBT9 (Clontech, Palo Alto, CA) con SnaBI y EcoRI, y el fragmento ADH1/GAL4 de 1,5 kb se sustituyó
 45 por el fragmento del promotor de ADH1 de 1,8 kb aislado de pGAD/KAN2 mediante digestión con SnaBI y EcoRI. El vector resultante (pMCK11) es un plásmido replicante en levaduras con un promotor de ADH1 y un terminador y un marcador de TRP1. El plásmido pMCK11 se digirió con EcoRI, se desfosforiló, y se ligó al fragmento EcoRI de *dhaB2* de pMP20. El plásmido resultante (pMCK17) tenía *dhaB2* orientado correctamente para la transcripción a partir del promotor de ADH1 y contenía un marcador TRP1.

50 *Construcción del plásmido de expresión de dhaB3 pMCK30*

El plásmido pGBT9 (Clontech) se digirió con NaeI y PvuII y el gen TRP1 de 1 kb se extrajo de este vector. El gen TRP1 se sustituyó por un gen URA3 proporcionado en forma de un fragmento AatII/NaeI de 1,7 kb del plásmido pRS406 (Stratagene) para proporcionar el vector intermedio pMCK32. El promotor de ADH1 truncado presente en pMCK32 se extrajo en un fragmento SnaBI/EcoRI de 1,5 kb, y se sustituyó por un promotor de ADH1 de tamaño completo en un fragmento SnaBI/EcoRI de 1,8 kb del plásmido pGAD/KAN2 para proporcionar el vector pMCK26. El único sitio EcoRI de pMCK26 se usó para insertar un fragmento EcoRI con *dhaB3* del plásmido pMP14 para
 60 proporcionar pMCK30. El plásmido de expresión replicante pMCK30 tiene *dhaB3* orientado para la expresión a partir del promotor de ADH1, y tiene un marcador URA3.

65 *Construcción del plásmido de expresión de dhaT pMCK35*

El plásmido pGBT9 (Clontech) se digirió con NaeI y PvuII y el gen TRP1 de 1 kb se extrajo de este vector. El gen TRP1 se sustituyó por un gen HIS3 proporcionado en forma de un fragmento XmnI/NaeI del plásmido pRS403 (Stratagene) para proporcionar el vector intermedio pMCK33. El promotor de ADH1 truncado presente en pMCK33

ES 2 336 858 T3

se extrajo en un fragmento SnaBI/EcoRI de 1,5 kb, y se sustituyó por un promotor de ADH1 de tamaño completo en un fragmento SnaBI/EcoRI de 1,8 kb a partir del plásmido pGAD/KAN2 para proporcionar el vector pMCK31. El único sitio EcoRI de pMCK31 se usó para insertar un fragmento EcoRI con *dhaT* del plásmido pMP 13 para proporcionar pMCK35. El plásmido de expresión replicante pMCK35 tiene *dhaT* orientado para la expresión a partir del promotor de ADH1, y tiene un marcador HIS3.

Transformación de S. cerevisiae con plásmidos de expresión de dha

La cepa YPH500 de *S. cerevisiae* (*ura3-52 lys2-801 ade2-101 trp1 -Δ63 his3-Δ200 leu2-Δ1*) (Sikorski R. S. y Hieter P., Genetics 122, 19-27, (1989)) adquirida de Stratagene (La Jolla, CA) se transformó con 1-2 μg de ADN plasmídico mediante el uso de un equipo Frozen-EZ Yeast Transformation (nº de catálogo T2001) (Zymo Research, Orange, CA). Las colonias se cultivaron con medio mínimo complementado (SMM - 0,67% de base nitrogenada de levadura sin aminoácidos, 2% de glucosa) durante 3-4 d a 29°C con una o más de las siguientes adiciones: sulfato de adenina (20 mg/l), uracilo (20 mg/l), L-triptófano (20 mg/l), L-histidina (20 mg/l), L-leucina (30 mg/l), L-lisina (30 mg/l). Se rayaron las colonias en placas selectivas y se utilizaron para inocular el medio líquido.

Cribado de los transformantes de S. cerevisiae en función de los genes dha

Se analizó el ADN cromosómico de los transformantes URA⁺, HIS⁺, TRP⁺, LEU⁺ mediante PCR con el uso de cebadores específicos para cada gen (SEQ ID NOS:38-45). Se confirmó la presencia de los cuatro marcos de lectura abiertos.

Expresión de la actividad de dhaB y dhaT en S. cerevisiae transformadas

Se demostró la presencia de glicerol deshidratasa activa (*dhaB*) y 1,3-propanodiol oxido-reductasa (*dhaT*) utilizando análisis enzimático *in vitro*. Además, el análisis mediante transferencia de western confirmó la expresión de las proteínas de los cuatro marcos de lectura abiertos.

La cepa YPH500, transformada con el grupo de plásmidos pMCK10, pMCK17, pMCK30 y pMCK35, se cultivó en medio mínimo complementado que contenía un 0,67% de base nitrogenada de levadura sin aminoácidos, 2% de glucosa, 20 mg/L de sulfato de adenina, y 30 mg/L de L-lisina. Las células se homogeneizaron y se analizó la actividad de *dhaB* en los extractos. Se obtuvo una actividad específica de 0,12 unidades por mg de proteína para la glicerol deshidratasa, y 0,024 unidades por mg de proteína para la 1,3-propanodiol oxidoreductasa.

Ejemplo 4

Producción de 1,3-propanodiol a partir de D-glucosa

Mediante el uso de Saccharomyces cerevisiae recombinante

La cepa YPH500 de *S. cerevisiae*, que alberga los grupos de plásmidos pMCK10, pMCK17, pMCK30 y pMCK35, se cultivó en un fermentador BiostatB (B Braun Biotech, Inc.) en 1,0 L de medio mínimo que contenía inicialmente 20 g/L de glucosa, 6,7 g/L de base nitrogenada de levadura sin aminoácidos, 40 mg/L de sulfato de adenina y 60 mg/L de L-lisina HCl. Durante el transcurso del cultivo, se añadieron equivalentes adicionales de base nitrogenada de levadura, adenina y lisina. El fermentador se mantuvo a pH 5,5 con la adición de ácido fosfórico al 10% y NaOH 2 M, 30°C, y una tensión de oxígeno disuelto del 40% por medio del control de la agitación. Después de 38 h, las células (DO₆₀₀ = 5,8 UA) se recogieron mediante centrifugación y se resuspendieron en medio base (6,7 g/L de base nitrogenada de levadura sin aminoácidos, 20 mg/L de sulfato de adenina, 30 mg/L de L-lisina HCl, y tampón de fosfato potásico 50 mM, pH 7,0).

Se prepararon mezclas de reacción que contenían células (DO₆₀₀ = 20 UA) en un volumen total de 4 mL de medio base complementado con un 0,5% de glucosa, 5 μg/mL de coenzima B₁₂ y 0, 10, 20, ó 40 mM de cloroquina, en ausencia de luz y oxígeno (inyección de nitrógeno), en botellas de suero selladas de 10 mL y se incubaron a 30°C con agitación. Después de 30 h, se extrajeron alícuotas y se analizaron mediante HPLC. Los resultados se muestran en la Tabla 3.

TABLA 3

Producción de 1,3-propanodiol mediante el uso de S. cerevisiae recombinante

| Reacción | cloroquina (mM) | 1,3-propanodiol (mM) |
|----------|--------------------|-------------------------|
| 1 | 0 | 0,2 |
| 2 | 10 | 0,2 |
| 3 | 20 | 0,3 |
| 4 | 40 | 0,7 |

Ejemplo 5

Uso de un transformante doble de S. cerevisiae para la producción de 1,3-propanodiol a partir de D-glucosa en el que dhaB y dhaT están integrados en el genoma

El Ejemplo 5 demuestra la transformación de *S. cerevisiae* con *dhaB1*, *dhaB2*, *dhaB3*, y *dhaT* y la integración estable de los genes en el genoma de la levadura para la producción de 1,3-propanodiol a partir de glucosa.

Construcción de casetes de expresión

Se construyen cuatro casetes de expresión (*dhaB1*, *dhaB2*, *dhaB3*, y *dhaT*) para la expresión constitutiva a nivel elevado inducida por glucosa de estos genes en la levadura *Saccharomyces cerevisiae*. Estos casetes consisten en: (i) el promotor de la fosfoglicerato quinasa (PGK) de la cepa de *S. cerevisiae* S288C; (ii) uno de los genes *dhaB1*, *dhaB2*, *dhaB3*, o *dhaT*; y (iii) el terminador de la PGK de la cepa de *S. cerevisiae* S288C. Se usa la técnica basada en PCR de corte y empalme de genes mediante extensión solapante (Horton *et al.*, BioTechniques, 8:528-535, (1990)) para recombinar las secuencias de ADN para generar estos casetes con uniones ininterrumpidas para la expresión óptima de cada gen. Estos casetes se clonan individualmente en un vector adecuado (pLITMUS 39) con sitios de restricción que pueden someterse a clonación mediante multi-casetes en plásmidos de expresión en levaduras.

Construcción de vectores de integración de levaduras

Se construyen vectores usados para llevar a cabo la integración de los casetes de expresión en el genoma de la levadura. Estos vectores contienen los siguientes elementos: (i) una región de policlonación en la que se subclonan los casetes de expresión; (ii) un marcador único usado para seleccionar los transformantes de levadura estables; (iii) un origen de replicación y un marcador seleccionable que permite la manipulación génica en *E. coli* antes de transformar la levadura. Un vector de integración contiene el marcador auxótrofo *URA3* (YIp352b), y un segundo vector de integración contiene el marcador auxótrofo *LYS2* (pKP7).

Construcción de plásmidos de expresión de levaduras

Se subclonan los casetes de expresión para *dhaB1* y *dhaB2* en la región de policlonación de YIp352b (plásmido de expresión nº 1), y los casetes de expresión para *dhaB3* y *dhaT* se subclonan en la región de policlonación de pKP7 (plásmido de expresión nº 2).

Transformación de levaduras con los plásmidos de expresión

Se transforma *S. cerevisiae* (*ura3*, *lys2*) con el plásmido de expresión nº 1 mediante el uso del equipo Frozen-EZ Yeast Transformation (Zymo Research, Orange, CA), y los transformantes se seleccionan en placas que carecen de uracilo. Se confirma la integración de los casetes de expresión para *dhaB1* y *dhaB2* mediante análisis de PCR del ADN cromosómico. Los transformantes seleccionados se re-transforman con el plásmido de expresión nº 2 mediante el uso del equipo Frozen-EZ Yeast Transformation, y los transformantes dobles se seleccionan en placas que carecen de lisina. Se confirma la integración de los casetes de expresión para *dhaB3* y *dhaT* mediante análisis de PCR del ADN cromosómico. Se confirma la presencia de los cuatro casetes de expresión (*dhaB1*, *dhaB2*, *dhaB3*, *dhaT*) en los transformantes dobles mediante análisis de PCR del ADN cromosómico.

ES 2 336 858 T3

Producción de proteínas de las levaduras doblemente transformadas

Se confirma la producción de las proteínas codificadas por *dhaB1*, *dhaB2*, *dhaB3* y *dhaT* de las levaduras doblemente transformadas mediante análisis de transferencia de Western.

5

Actividad enzimática de las levaduras doblemente transformadas

Se confirma la glicerol deshidratasa activa y la 1,3-propanodiol deshidrogenasa activa de las levaduras doblemente transformadas mediante el ensayo enzimático descrito anteriormente en los Métodos Generales.

10

Producción de 1,3-propanodiol a partir de las levaduras doblemente transformadas

La producción de 1,3-propanodiol a partir de glucosa en las levaduras doblemente transformadas se demuestra básicamente como se describió en el Ejemplo 4.

15

Ejemplo 6

20

Construcción de plásmidos que contienen *DAR1/GPP2* o *dhaT/dhaB1-3* y transformación en la especie *Klebsiella*

K. pneumoniae (ATCC 25955), *K. pneumoniae* (ECL2106), y *K. oxytoca* (ATCC 8724) son resistentes de manera natural a la ampicilina (hasta 150 ug/mL) y kanamicina (hasta 50 ug/mL), pero sensibles a la tetraciclina (10 ug/mL) y cloranfenicol (25 ug/mL). Por lo tanto, los plásmidos replicantes que codifican la resistencia a estos dos últimos antibióticos son potencialmente útiles como vectores de clonación para estas cepas de *Klebsiella*. La *K. pneumoniae* de tipo natural (ATCC 25955), la *K. pneumoniae* desreprimida por glucosa (ECL2106), y *K. oxytoca* (ATCC 8724) se transformaron de manera eficaz con la resistencia a tetraciclina mediante electroporación con el plásmido de número de copias moderado pBR322 (New England Biolabs, Beverly, MA). Esto se llevó a cabo mediante el siguiente procedimiento: Se inocularon diez mL de un cultivo durante la noche en 1 L de LB (1% (p/v) de Bacto-triptona (Difco, Detroit, MI), 0,5% (p/v) de extracto Bacto-Yeast (Difco) y 0,5% (p/v) de NaCl (Sigma, St. Louis, MO), y el cultivo se incubó a 37°C hasta una DO₆₀₀ de 0,5-0,7. Las células se enfriaron con hielo, se recogieron mediante centrifugación a 4000 x g durante 15 min, y se resuspendieron en 1 L de glicerol al 10% helado estéril. Las células se recogieron repetidamente mediante centrifugación y se resuspendieron progresivamente en 500 mL, 20 mL y, finalmente, 2 mL de glicerol al 10% helado estéril. Para la electroporación, se mezclaron 40 uL de células con 1-2 uL de ADN en una cubeta fría de 0,2 cm y se sometieron a un pulso a 200 Ω, 2,5 kV durante 4-5 mseg mediante el uso de un Gene Pulser de BioRad (BioRad, Richmond, CA). Se añadió un μL de medio SOC (2% (p/v) de Bacto-triptona (Difco), 0,5% (p/v) de extracto Bacto-Yeast (Difco), NaCl 10 μM, MgCl₂ 10 μM, MgSO₄ 10 μM, KCl 2,5 μM y glucosa 20 μM) a las células, y, después de transferir la suspensión a un tubo de polipropileno estéril de 17 x 100 mm, el cultivo se incubó durante 1 hr a 37°C, 225 rpm. Las alícuotas se colocaron en un medio selectivo, tal como se indica. Los análisis del ADN plasmídico de los transformantes independientes resistentes a tetraciclina mostraron los patrones de digestión con endonucleasas de restricción típicos de pBR322, lo que indicó que el vector se mantenía de manera estable tras el cultivo durante la noche a 37°C en LB que contenía tetraciclina (10 ug/mL). Así, este vector, y los derivados tales como pBR329 (ATCC 37264), que codifica la resistencia a ampicilina, tetraciclina y cloranfenicol, se pueden usar para introducir los casetes de expresión de *DAR1/GPP2* y *dhaT/dhaB1-3* en *K. pneumoniae* y *K. oxytoca*.

45

Los genes *DAR1* y *GPP2* se pueden obtener mediante la amplificación mediada por PCR del genoma de *Saccharomyces cerevisiae*, basándose en su secuencia de ADN conocida. Los genes se transforman después en *K. pneumoniae* o *K. oxytoca* bajo el control de uno o más promotores que se pueden usar para dirigir su expresión en medios que contienen glucosa. Por comodidad, los genes se obtuvieron en un fragmento de ADN de 2,4 kb obtenido mediante la digestión del plásmido pAH44 con la endonucleasa de restricción *PvuII*, por lo que los genes ya estaban dispuestos en un casete de expresión bajo el control del promotor *lac* de *E. coli*. Este fragmento de ADN se ligó con pBR329 digerido con *PvuII*, lo que produjo la inactivación por inserción de su gen de resistencia a cloranfenicol. El ADN ligado se usó para transformar la cepa DH5α de *E. coli* (Gibco, Gaithersburg, MD). Los transformantes se seleccionaron por su resistencia a tetraciclina (10 ug/mL) y se cribaron en función de su sensibilidad a cloranfenicol (25 ug/mL). El análisis del ADN plasmídico de los transformantes resistentes a tetraciclina y sensibles a cloranfenicol confirmó la presencia de los plásmidos esperados, en los que el casete de expresión P_{lac}-*dar1-gpp2* se subclonó en ambas orientaciones en el sitio *PvuII* de pBR329. Estos plásmidos, denominados pJSP1A (orientación en el sentido de las agujas del reloj) y pJSP1B (orientación en sentido contrario a las agujas del reloj), se transformaron por separado mediante electroporación en *K. pneumoniae* (ATCC 25955), *K. pneumoniae* (ECL2106) y *K. oxytoca* (ATCC 8724) tal como se ha descrito. Los transformantes se seleccionaron por su resistencia a tetraciclina (10 ug/mL) y se cribaron en función de su sensibilidad a cloranfenicol (25 ug/mL). El análisis de restricción de los plásmidos aislados a partir de los transformantes independientes mostró solamente los patrones de digestión esperados, y confirmó que se mantenían de manera estable a 37°C con la selección mediante antibióticos. La expresión de los genes *DAR1* y *GPP2* se puede incrementar mediante la adición de IPTG (0,2-2,0 mM) al medio de cultivo.

65

ES 2 336 858 T3

Los cuatro genes *dhaB*(1-3) y *dhaT* de *K. pneumoniae* se pueden obtener mediante amplificación mediada por PCR a partir del genoma de *K. pneumoniae*, basándose en su secuencia ADN conocida. Estos genes se transforman después en *K. pneumoniae* bajo el control de uno o más promotores que se pueden usar para dirigir su expresión en medios que contienen glucosa. Por comodidad, los genes se obtuvieron en un fragmento de ADN de aproximadamente 4,0 kb obtenido mediante la digestión del plásmido pAH24 con las endonucleasas de restricción *KpnI/SacI*, por lo que los genes ya están dispuestos en un casete de expresión bajo el control del promotor *lac* de *E. coli*. Este fragmento de ADN se ligó con pBC-KS+ (Stratagene, LaJolla, CA) digerido de manera similar y se usó para transformar la cepa DH5 α de *E. coli*. Los transformantes se seleccionaron en función de su resistencia a cloranfenicol (25 ug/mL) y se cribaron en función del color blanco de las colonias en agar LB que contenía X-gal. El análisis de restricción del ADN plasmídico de los transformantes resistentes a cloranfenicol que mostraban el fenotipo de colonias blancas confirmó la presencia del plásmido esperado, denominado pJSP2, en el que los genes *dhaT-dhaB*(1-3) se subclonaron bajo el control del promotor *lac* de *E. coli*.

Para incrementar la conversión de glucosa a 3G, este plásmido se transformó por separado mediante electroporación en *K. pneumoniae* (ATCC 25955) (pJSP1A), *K. pneumoniae* (ECL2106) (pJSP1A) y *K. oxytoca* (ATCC 8724) (pJSP1A) que ya contenían el casete de expresión P_{lac}-*dar1-gpp2*. Los cotransformantes se seleccionaron por su resistencia tanto a tetraciclina (10 ug/mL) como a cloranfenicol (25 ug/mL). El análisis de restricción de los plásmidos aislados de los cotransformantes independientes mostró los patrones de digestión esperados tanto para pJSP1A como para pJSP2. La expresión de los genes *DAR1*, *GPP2*, *dhaB*(1-3), y *dhaT* se puede incrementar mediante la adición de IPTG (0,2-2,0 mM) al medio.

Ejemplo 7

Producción de 1,3-propanodiol a partir de glucosa mediante *K. pneumoniae*

Se cultivaron las cepas ECL 2106 y 2106-47 de *Klebsiella pneumoniae*, ambas transformadas con pJSP1A, y ATCC 25955, transformada con pJSP1A y pJSP2, en un fermentador Applikon de 5 L en diversas condiciones (véase la Tabla 4) para la producción de 1,3-propanodiol a partir de glucosa. La cepa 2104-47 es un derivado tolerante al fluoroacetato de ECL 2106 que se obtuvo a partir de una placa de selección de fluoroacetato/lactato tal como se describe en Bauer *et al.*, Appl. Environ. Microbiol. 56, 1296 (1990). En cada caso, el medio usado contuvo tampón de fosfato potásico 50-100 mM, pH 7,5, (NH₄)₂SO₄ 40 mM, 0,1% (p/v) de extracto de levadura, CoCl₂ 10 μ M, CuCl₂ 6,5 μ M, FeCl₃ 100 μ M, FeSO₄ 18 μ M, H₃BO₃ 5 μ M, MnCl₂ 50 μ M, Na₂MoO₄ 0,1 μ M, ZnCl₂ 25 μ M, MgSO₄ 0,82 mM, CaCl₂ 0,9 mM, y 10-20 g/L de glucosa. Se añadió glucosa adicional, con la glucosa residual mantenida en exceso. La temperatura se mantuvo a 37°C y el pH se mantuvo a 7,5 con KOH o NaOH 5 N. Se incluyeron antibióticos apropiados para el mantenimiento del plásmido; también se añadió IPTG (isopropil- β -D-tiogalactopiranosido) a las concentraciones indicadas. Para las fermentaciones anaeróbicas, se inyectaron 0,1 vvm de nitrógeno a través del reactor; cuando el punto de referencia de dO fue del 5%, se inyectó 1 vvm de aire a través del reactor y el medio se complementó con vitamina B12. Las concentraciones finales y los rendimientos totales (g/g) se muestran en la Tabla 4.

TABLA 4

Producción de 1,3 propanodiol a partir de glucosa mediante *K. pneumoniae*

| Organismo | dO | IPTG, mM | vitamina B12, mg/L | Título, g/L | Rendimiento, g/g |
|---------------------|----|----------|--------------------|-------------|------------------|
| 25955[pJSP1A/pJSP2] | 0 | 0,5 | 0 | 8,1 | 16% |
| 25955[pJSP1A/pJSP2] | 5% | 0,2 | 0,5 | 5,2 | 4% |
| 2106[pJSP1A] | 0 | 0 | 0 | 4,9 | 17% |
| 2106[pJSP1A] | 5% | 0 | 5 | 6,5 | 12% |
| 2106-47[pJSP1A] | 5% | 0,2 | 0,5 | 10,9 | 12% |

Ejemplo 8

Conversión de sustratos de carbono en 1,3-propanodiol mediante *K. pneumoniae* recombinante que contiene *dar1*, *gpp2*, *dhaB*, y *dT*

A. Conversión de D-fructosa a 1,3-propanodiol mediante diversas cepas recombinantes de *K. pneumoniae*

Se transfirieron colonias independientes de *K. pneumoniae* (ATCC 25955 pJSP1A), *K. pneumoniae* (ATCC 25955 pJSP1A/pJSP2), *K. pneumoniae* (ATCC 2106 pJSP1A), y *K. pneumoniae* (ATCC 2106 pJSP1A/pJSP2) de placas de agar y se subcultivaron durante la noche en distintos tubos de cultivo en caldo de cultivo Luria-Bertani (LB) que contenía el/los agente(s) antibiótico(s) apropiado(s). Un matraz de 50-mL que contiene 45 mL de un medio mínimo esterilizado mediante filtración se definió como LLMM/F, que contiene por litro: 10 g de fructosa; 1 g de extracto de levadura; 50 mmoles de fosfato potásico, pH 7,5; 40 mmoles de $(\text{NH}_4)_2\text{SO}_4$; 0,09 mmoles de cloruro cálcico; 2,38 mg de $\text{CoCl}_2 \cdot 6\text{H}_2\text{O}$; 0,88 mg de $\text{CuCl}_2 \cdot 2\text{H}_2\text{O}$; 27 mg de $\text{FeCl}_3 \cdot 6\text{H}_2\text{O}$; 5 mg de $\text{FeSO}_4 \cdot 7\text{H}_2\text{O}$; 0,31 mg de H_3BO_3 ; 10 mg de $\text{MnCl}_2 \cdot 4\text{H}_2\text{O}$; 0,023 mg de $\text{Na}_2\text{MoO}_4 \cdot 2\text{H}_2\text{O}$; 3,4 mg de ZnCl_2 ; 0,2 g de $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$. Se añadió tetraciclina a 10 $\mu\text{g}/\text{mL}$ al medio para las reacciones que usaban cualquiera de los recombinantes de un único plásmido; 10 $\mu\text{g}/\text{mL}$ de tetraciclina y 25 $\mu\text{g}/\text{mL}$ de cloranfenicol para las reacciones que usaban cualquiera de los recombinantes de plásmido doble. Se inyectó nitrógeno al medio antes de la inoculación de 2 mL del subcultivo. Se añadió IPTG (I) a una concentración final de 0,5 mM a algunos matraces. Los matraces se taparon, después se incubaron a 37°C, 100 rpm en un incubador/agitador New Brunswick Series 25. Las reacciones se llevaron a cabo durante al menos 24 horas o hasta que la mayoría del sustrato de carbono se convirtió en los productos. Las muestras se analizaron mediante HPLC. La Tabla 5 describe los rendimientos del 1,3-propanodiol producido a partir de fructosa mediante los diversos recombinantes de *Klebsiella*.

TABLA 5

Producción de 1,3-propanodiol a partir de D-fructosa mediante el uso de *Klebsiella* recombinante

| Cepa de <i>Klebsiella</i> | Medio | Conversión | [3G] (g/L) | Rendimiento de carbono (%) |
|---------------------------|------------|------------|---------------|-------------------------------|
| 2106 pBR329 | LLMM/F | 100 | 0 | 0 |
| 2106 pJSP1A | LLMM/F | 50 | 0,66 | 15,5 |
| 2106 pJSP1A | LLMM/F + I | 100 | 0,11 | 1,4 |
| 2106 pJSP1A/pJSP2 | LLMM/F | 58 | 0,26 | 5 |
| 25955 pBR329 | LLMM/F | 100 | 0 | 0 |
| 25955 pJSP1A | LLMM/F | 100 | 0,3 | 4 |
| 25955 pJSP1A | LLMM/F + I | 100 | 0,15 | 2 |
| 25955 pJSP1A/pJSP2 | LLMM/F | 100 | 0,9 | 11 |
| 25955 pJSP1A/pJSP2 | LLMM/F + I | 62 | 1,0 | 20 |

B. Conversión de diversos sustratos de carbono a 1,3-propanodiol mediante *K. pneumoniae* (ATCC 25955 pJSP1A/pJSP2)

Se transfirió una alícuota (0,1 mL) de cultivos de reserva congelados de *K. pneumoniae* (ATCC 25955 pJSP1A/pJSP2) a 50 mL de medio de cultivo para siembra en un matraz de tabiques de 250 mL. El medio de siembra contenía por litro: tampón NaK/PO₄ 0,1 molar, pH 7,0; 3 g de $(\text{NH}_4)_2\text{SO}_4$; 5 g de glucosa, 0,15 g de $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$, 10 mL de disolución de oligoelementos 100X, 25 mg de cloranfenicol, 10 mg de tetraciclina, y 1 g de extracto de levadura. La disolución de oligoelementos 100X contenía por litro: 10 g de ácido cítrico, 1,5 g de $\text{CaCl}_2 \cdot 2\text{H}_2\text{O}$, 2,8 g de $\text{FeSO}_4 \cdot 7\text{H}_2\text{O}$, 0,39 g de $\text{ZnSO}_4 \cdot 7\text{H}_2\text{O}$, 0,38 g de $\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$, 0,2 g de $\text{CoCl}_2 \cdot 6\text{H}_2\text{O}$, y 0,3 g de $\text{MnCl}_2 \cdot 4\text{H}_2\text{O}$. La disolución resultante se tituló hasta pH 7,0 con KOH o H_2SO_4 . La glucosa, los oligoelementos, los antibióticos y los extractos de levadura se esterilizaron por separado. El inóculo de siembra se cultivó durante la noche a 35°C y 250 rpm.

El diseño de la reacción fue semi-aeróbico. El sistema consistió en 130 mL de medio de reacción en matraces sellados de 125 mL que se dejaron parcialmente abiertos con una tira de papel de aluminio. El medio de reacción contenía por litro: 3 g de $(\text{NH}_4)_2\text{SO}_4$; 20 g de sustrato de carbono; tampón de NaK/PO₄ 0,15 molar, pH 7,5; 1 g de

ES 2 336 858 T3

extracto de levadura; 0,15 g de $MgSO_4 \cdot 7H_2O$; 0,5 mmoles de IPTG; 10 mL de disolución de oligoelementos 100X; 25 mg de cloranfenicol; y 10 mg de tetraciclina. La disolución resultante se tituló hasta pH 7,5 con KOH o H_2SO_4 . Las fuentes de carbono fueron: D-glucosa (Glc); D-fructosa (Frc); D-lactosa (Lac); D-Sacarosa (Suc); D-maltosa (Mal); y D-manitol (Man). Se incluyeron en el medio unas cuantas esferas de vidrio para mejorar la mezcla. Las reacciones se iniciaron mediante la adición del inóculo de siembra, de forma que la densidad óptica de la suspensión celular comenzó en 0,1 UA medida a λ_{600} nm. Los matraces se incubaron a 35°C: 250 rpm. La producción de 3G se midió mediante HPLC tras 24 hr. La Tabla 6 describe los rendimientos de 1,3-propanodiol producidos a partir de los diversos sustratos de carbono.

TABLA 6

Producción de 1,3-propanodiol a partir de diversos sustratos de carbono mediante el uso de Klebsiella 25955 $pJSP1A/pJSP2$

| Sustrato de Carbono | 1,3-Propanodiol (g/L) | | |
|---------------------|-----------------------|---------|---------|
| | Expt. 1 | Expt. 2 | Expt. 3 |
| Glc | 0,89 | 1 | 1,6 |
| Frc | 0,19 | 0,23 | 0,24 |
| Lac | 0,15 | 0,58 | 0,56 |
| Sac | 0,88 | 0,62 | |
| Mal | 0,05 | 0,03 | 0,02 |
| Man | 0,03 | 0,05 | 0,04 |

REIVINDICACIONES

- 5 1. Un método para la producción de 1,3-propanodiol a partir de un organismo recombinante que comprende:
- (i) transformar un organismo hospedador adecuado con uno o más casetes de transformación, cada uno de los cuales comprende al menos uno de
- 10 (a) un gen que codifica una actividad de glicerol-3-fosfato deshidrogenasa;
- (b) un gen que codifica una actividad de glicerol-3-fosfatasa;
- (c) genes que codifican una actividad de deshidratasa;
- 15 (d) un gen que codifica una actividad de 1,3-propanodiol oxidorreductasa,
- en donde todos los genes (a)-(d) se introducen en el microorganismo hospedador;
- (ii) cultivar el organismo hospedador transformado en condiciones adecuadas en presencia de al menos una fuente
- 20 de carbono seleccionada del grupo que consiste en monosacáridos, oligosacáridos, polisacáridos, o un sustrato de un carbono, por lo cual se produce 1,3-propanodiol; y
- (iii) recuperar el 1,3-propanodiol.
- 25 2. El método de la Reivindicación 1, en el que el organismo hospedador adecuado se selecciona del grupo que consiste en bacterias, levaduras y hongos filamentosos.
3. El método de la Reivindicación 2, en el que el organismo hospedador adecuado se selecciona del grupo de géneros que consiste en *Citrobacter*, *Enterobacter*, *Clostridium*, *Klebsiella*, *Aerobacter*, *Lactobacillus*, *Aspergillus*, *Saccharomyces*, *Schizosaccharomyces*, *Zygosaccharomyces*, *Pichia*, *Kluyveromyces*, *Candida*, *Hansenula*, *Debaryomyces*, *Mucor*, *Torulopsis*, *Methylobacter*, *Escherichia*, *Salmonella*, *Bacillus*, *Streptomyces* y *Pseudomonas*.
- 30 4. El método de la Reivindicación 3, en el que el organismo hospedador adecuado se selecciona del grupo que consiste en *E. coli*, *Klebsiella spp.* y *Saccharomyces spp.*
- 35 5. El método de la Reivindicación 1, en el que el organismo hospedador transformado es *Saccharomyces spp.* transformado con un casete de transformación que comprende los genes *dhaB1*, *dhaB2*, *dhaB3*, y *dhaT*, en el que los genes se integran de manera estable en el genoma de *Saccharomyces spp.*
- 40 6. El método de la Reivindicación 1, en el que el organismo hospedador transformado es *Klebsiella spp.* transformado con un casete de transformación que comprende los genes glicerol-3-fosfato deshidrogenasa 1 (GPD1) y glicerol-3-fosfatasa 2 (GPP2).
- 45 7. El método de la Reivindicación 1, en el que la fuente de carbono es glucosa.
8. El método de la Reivindicación 1, en el que el gen que codifica una enzima glicerol-3-fosfato deshidrogenasa se selecciona del grupo que consiste en los genes que corresponden a las secuencias de aminoácidos proporcionadas en SEQ ID NO: 11, en SEQ ID NO: 12, y en SEQ ID NO: 13, y las secuencias de aminoácidos abarcan las sustituciones, deleciones o adiciones de aminoácidos que no alteran la función de la enzima glicerol-3-fosfato deshidrogenasa.
- 50 9. El método de la Reivindicación 1, en el que el gen que codifica una enzima glicerol-3-fosfatasa se selecciona del grupo que consiste en los genes que corresponden a las secuencias de aminoácidos proporcionadas en SEQ ID NO: 33 y en SEQ ID NO: 17, y las secuencias de aminoácidos abarcan las sustituciones, deleciones o adiciones de aminoácidos que no alteran la función de la enzima glicerol-3-fosfatasa.
- 55 10. El método de la Reivindicación 1, en el que el gen que codifica una actividad de glicerol-3-fosfatasa es una enzima glicerol quinasa que corresponde a la secuencia de aminoácidos proporcionada en SEQ ID NO: 18, y la secuencia de aminoácidos abarca las sustituciones, deleciones o adiciones de aminoácidos que no alteran la función de la enzima glicerol quinasa.
- 60 11. El método de la Reivindicación 1, en el que los genes que codifican una enzima deshidratasa comprenden *dhaB1*, *dhaB2* y *dhaB3*, y los genes corresponden respectivamente a las secuencias de aminoácidos proporcionadas en SEQ ID NO:34, SEQ ID NO:35, y SEQ ID NO:36, y las secuencias de aminoácidos abarcan las sustituciones, deleciones o adiciones de aminoácidos que no alteran la función de la enzima deshidratasa.
- 65 12. El método de la Reivindicación 1, en el que el gen que codifica una enzima 1,3-propanodiol oxidorreductasa corresponde a una secuencia de aminoácidos proporcionada en SEQ ID NO:37, y la secuencia de aminoácidos abarca

ES 2 336 858 T3

las sustituciones, deleciones o adiciones de aminoácidos que no alteran la función de la enzima 1,3-propanodiol oxidorreductasa.

13. Una célula hospedadora transformada que comprende:

(1) un gen heterólogo que codifica una enzima glicerol-3-fosfato deshidrogenasa que corresponde a la secuencia de aminoácidos proporcionada en SEQ ID NO:11;

(2) un gen heterólogo que codifica una enzima glicerol-3-fosfatasa que corresponde a la secuencia de aminoácidos proporcionada en SEQ ID NO:17;

(3) un gen heterólogo que codifica la subunidad α de la enzima glicerol deshidratasa que corresponde a la secuencia de aminoácidos proporcionada en SEQ ID NO:34;

(4) un gen heterólogo que codifica la subunidad β de la enzima glicerol deshidratasa que corresponde a la secuencia de aminoácidos proporcionada en SEQ ID NO:35;

(5) un gen heterólogo que codifica la subunidad γ de la enzima glicerol deshidratasa que corresponde a la secuencia de aminoácidos proporcionada en SEQ ID NO:36; y

(6) un gen heterólogo que codifica la enzima 1,3-propanodiol oxidorreductasa que corresponde a la secuencia de aminoácidos proporcionada en SEQ ID NO:37,

las secuencias de aminoácidos respectivas de (1)-(6) abarcan las sustituciones, deleciones o adiciones de aminoácidos que no alteran la función de las enzimas de los genes (1)-(6),

por lo que la célula hospedadora transformada produce 1,3-propanodiol con al menos un sustrato seleccionado del grupo que consiste en monosacáridos, oligosacáridos, y polisacáridos, o a partir de un sustrato de un carbono.

ES 2 336 858 T3

LISTA DE SECUENCIAS

(1) INFORMACIÓN GENERAL:

- 5 (i) SOLICITANTE:
- (A) TITULAR: E. I. DU PONT DE NEMOURS AND COMPANY
 - (B) CALLE: 1007 MARKET STREET
 - 10 (C) CIUDAD: WILMINGTON
 - (D) ESTADO: DELAWARE
 - (E) PAIS: U.S.A.
 - (F) CÓDIGO: 19898
 - 15 (G) TELEFONO: 302-892-8112
 - (H) TELEFAX: 302-773-0164
 - (I) TELEX: 6717325
- 20 (A) TITULAR: GENENCOR INTERNATIONAL, INC.
- (B) CALLE : 4 CAMBRIDGE PLACE 1870 SOUTH WINTON ROAD
 - (C) CIUDAD: ROCHESTER
 - 25 (D) ESTADO: NEW YORK
 - (E) PAIS: U.S.A.
 - (F) CÓDIGO POSTAL (CP): 14618
- 30 (ii) TÍTULO DE LA INVENCIÓN: MÉTODO PARA LA PRODUCCIÓN RECOMBINANTE DE 1,3-PRO-PANODIOL
- (iii) NÚMERO DE SECUENCIAS: 49
- 35 (iv) FORMA LEGIBLE POR ORDENADOR:
- (A) TIPO DE MEDIO: DISQUETE DE 3,5 PULGADAS
 - (B) ORDENADOR: COMPATIBLE CON IBM
 - 40 (C) SISTEMA OPERATIVO: MICROSOFT WORD PARA WINDOWS 95
 - (D) SOFTWARE: MICROSOFT WORD VERSIÓN 7.0A
- (v) DATOS DE LA SOLICITUD ACTUAL:
- 45 (A) NÚMERO DE SOLICITANTE:
 - (B) FECHA DE PRESENTACIÓN:
 - (C) CLASIFICACIÓN:
- (vi) DATOS DE LA SOLICITUD ANTERIOR:
- 50 (A) NÚMERO DE SOLICITANTE: 60/030,601
 - (B) FECHA DE PRESENTACIÓN: 13 NOVIEMBRE 1996
- (vii) INFORMACIÓN SOBRE EL AGENTE/ABOGADO:
- 55 (A) NOMBRE: FLOYD, LINDA AXAMETHY
 - (B) NÚMERO DE REGISTRO: 33,692
 - (C) NÚMERO DE REFERENCIA/EXPEDIENTE: CR-9982
- 60 (2) INFORMACIÓN PARA LA SEC. ID NO:1:
- (i) CARACTERÍSTICAS DE LA SECUENCIA:
- 65 (A) LONGITUD: 1668 pares de bases
 - (B) TIPO: ácido nucleico
 - (C) TIPO DE CADENA: simple

ES 2 336 858 T3

(D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: DNA (genómico)

(iii) HYPOTÉTICO: NO

(iv) ANTI-SENTIDO: NO

(vi) FUENTE ORIGINAL:

(A) ORGANISMO: DHAB1

(xi) DESCRIPCIÓN DE LA SECUENCIA: SEC ID NO:1:

| | | |
|----|---|------|
| 15 | ATGAAAAGAT CAAAACGATT TGCAGTACTG GCCCAGCGCC CCGTCAATCA GGACGGGCTG | 60 |
| | ATTGGCGAGT GGCCTGAAGA GGGGCTGATC GCCATGGACA GCCCCTTTGA CCCGGTCTCT | 120 |
| | TCAGTAAAG TGGACAACGG TCTGATCGTC GAAGTGGACG GCAAACGCCG GGACCAGTTT | 180 |
| 20 | GACATGATCG ACCGATTTAT CGCCGATTAC GCGATCAACG TTGAGCGCAC AGAGCAGGCA | 240 |
| | ATGCGCCTGG AGGCGGTGGA AATAGCCCGT ATGCTGGTGG ATATTCACGT CAGCCGGGAG | 300 |
| | GAGATCATTG CCATCACTAC CGCCATCACG CCGGCCAAG CGGTCGAGGT GATGGCGCAG | 360 |
| 25 | ATGAACGTGG TGGAGATGAT GATGGCGCTG CAGAAGATGC GTGCCCGCCG GACCCCTCC | 420 |
| | AACCAGTGCC ACGTCACCPA TCTCAAAGAT AATCCGGTGC AGATTGCCGC TGACGCCGCC | 480 |
| | GAGGCCGGGA TCCGCGGCTT CTCAGAACAG GAGACCACGG TCGGTATCGC GCGCTACGCG | 540 |
| 30 | CCGTTTAACG CCCTGGCGCT GTTGGTCGGT TCGCAGTGGC GCCGCCCGG CGTGTGACG | 600 |
| | CAGTGCTCGG TGGAAAGAGC CACCGAGCTG GAGCTGGGCA TGCCTGGCTT AACCAGCTAC | 660 |
| 35 | GCCGAGACGG TGTCGGTCTA CGGCACCGAA GCGGTATTTA CCGACGGCGA TGATACGCCG | 720 |
| | TGGTCAAAGG CGTTCCTCGC CTCGGCCTAC GCCTCCCGCG GGTGAAAAT GCGCTACACC | 780 |
| 40 | TCCGGCACCG GATCCGAAGC GCTGATGGGC TATTCGGAGA GCAAGTCGAT GCTCTACCTC | 840 |
| | GAATCGCGCT GCATCTTCAT TACTAAAGGC GCCGGGGTTC AGGGACTGCA AAACGGCGCG | 900 |
| | GTGAGCTGTA TCGGCATGAC CGGCGCTGTG CCGTCGGGCA TTCGGGCGGT GCTGGCGGAA | 960 |
| 45 | AACCTGATCG CCTCTATGCT CGACCTCGAA GTGGCGTCCG CCAACGACCA GACTTTCTCC | 1020 |
| | CACTCGGATA TTCGCCGCAC CGCGCGCACC CTGATGCAGA TGCTGCCGGG CACCGACTTT | 1080 |
| | ATTTTCTCCG GCTACAGCGC GGTGCCGAAC TACGACAACA TGTTCCCGG CTCGAACTTC | 1140 |
| 50 | GATGCGGAAG ATTTTGATGA TTACAACATC CTGCAGCGTG ACCTGATGGT TGACGGCGGC | 1200 |
| | CTGCGTCCGG TGACCGAGGC GGAAACCATT GCCATTCCGC AGAAAGCGGC GCGGGCGATC | 1260 |
| 55 | CAGGCGGTTT TCCGCGAGCT GGGGCTGCCG CCAATCGCCG ACGAGGAGGT GGAGGCCGCC | 1320 |
| | ACCTACGCGC ACGGCAGCAA CGAGATGCCG CCGCGTAACG TGGTGGAGGA TCTGAGTGCG | 1380 |
| | GTGGAAGAGA TGATGAAGCG CAACATCACC GGCCTCGATA TTGTCGGCGC GCTGAGCCGC | 1440 |
| 60 | AGCGGCTTTG AGGATATCGC CAGCAATATT CTCAATATGC TGCGCCAGCG GGTCACCGGC | 1500 |
| | GATTACCTGC AGACCTCGGC CATTCTCGAT CGGCAGTTCG AGGTGGTGAG TGCGGTCAAC | 1560 |
| | GACATCAATG ACTATCAGGG GCCGGGCACC GGCTATCGCA TCTCTGCCGA ACGCTGGGCG | 1620 |
| 65 | GAGATCAAAA ATATTCCGGG CGTGGTTCAG CCCGACACCA TTGAATAA | 1668 |

ES 2 336 858 T3

(2) INFORMACIÓN PARA LA SEC. ID NO:2:

(i) CARACTERÍSTICAS DE LA SECUENCIA:

- (A) LONGITUD: 585 pares de bases
- (B) TIPO: ácido nucleico
- (C) TIPO DE CADENA: simple
- (D) TIPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: DNA (genómico)

(vi) FUENTE ORIGINAL:

- (A) ORGANISMO: DHAB2

(xi) DESCRIPCIÓN DE LA SECUENCIA: SEC ID NO:2:

| | | |
|----|---|-----|
| 5 | GTGCAACAGA CAACCCAAAT TCAGCCCTCT TTTACCCTGA AAACCCGCGA GGGCGGGGTA | 60 |
| 20 | GCTTCTGCCG ATGAACGCGC CGATGAAGTG GTGATCGGCG TCGGCCCTGC CTTCGATAAA | 120 |
| 25 | CACCAGCATC AACTCTGAT CGATATGCC CATGGCGCGA TCCTCAAAGA GCTGATTGCC | 180 |
| 30 | GGGGTGAAG AAGAGGGGCT TCACGCCCGG GTGGTGC GCA TTCTGCGCAC GTCCGACGTC | 240 |
| 35 | TCCTTTATGG CCTGGGATGC GGCCAACCTG AGCGGCTCGG GGATCGGCAT CGGTATCCAG | 300 |
| 40 | TCGAAGGGGA CCACGGTCAT CCATCAGCGC GATCTGCTGC CGCTCAGCAA CCTGGAGCTG | 360 |
| 45 | TTCTCCCAGG CGCCGCTGCT GACGCTGGAG ACCTACCGGC AGATTGGCAA AAACGCTGCG | 420 |
| 50 | CGCTATGCGC GCAAAGAGTC ACCTTCGCCG GTGCCGGTGG TGAACGATCA GATGGTGGG | 480 |
| 55 | CCGAAATTTA TGGCCAAAGC CGCGCTATTT CATATCAAAG AGACCAAACA TGTGGTGCAG | 540 |
| 60 | GACGCCGAGC CCGTCACCCT GCACATCGAC TTAGTAAGGG AGTGA | 585 |

(2) INFORMACIÓN PARA LA SEC. ID NO:3:

(i) CARACTERÍSTICAS DE LA SECUENCIA:

- (A) LONGITUD: 426 pares de bases
- (B) TIPO: ácido nucleico
- (C) TIPO DE CADENA: simple
- (D) TIPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: DNA (genómico)

(vi) FUENTE ORIGINAL:

- (A) ORGANISMO: DHAB3

(xi) DESCRIPCIÓN DE LA SECUENCIA: SEC ID NO:3:

| | | |
|----|---|-----|
| 55 | ATGAGCGAGA AAACCATGCG CGTGCAGGAT TATCCGTTAG CCACCCGCTG CCCGGAGCAT | 60 |
| 60 | ATCCTGACGC CTACCGGCAA ACCATTGACC GATATTACCC TCGAGAAGGT GCTCTCTGGC | 120 |
| 65 | GAGGTGGGCC CGCAGGATGT GCGGATCTCC CGCCAGACCC TTGAGTACCA GGCAGGATT | 180 |
| 70 | GCCGAGCAGA TGCAGCGCCA TGCGGTGGCG CGCAATTTCC GCCGCGCGGC GGAGCTTATC | 240 |
| 75 | GCCATTCTTG ACGAGCGCAT TCTGGCTATC TATAACGCGC TGCGCCCGTT CCGCTCCTCG | 300 |
| 80 | CAGGCGGAGC TGCTGGCGAT CGCCGACGAG CTGGAGCACA CCTGGCATGC GACAGTGAAT | 360 |
| 85 | GCCGCCTTTG TCCGGGAGTC GGCGGAAGTG TATCAGCAGC GGCATAAGCT GCGTAAAGGA | 420 |
| 90 | AGCTAA | 426 |

ES 2 336 858 T3

(2) INFORMACIÓN PARA LA SEC. ID NO:4:

(i) CARACTERÍSTICAS DE LA SECUENCIA:

- (A) LONGITUD: 1164 pares de bases
- (B) TIPO: ácido nucleico
- (C) TIPO DE CADENA: simple
- (D) TIPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: DNA (genómico)

(vi) FUENTE ORIGINAL:

- (A) ORGANISMO: DHAT

(xi) DESCRIPCIÓN DE LA SECUENCIA: SEC ID NO:4:

| | | |
|-----|---|------|
| 5 | ATGAGCTATC GTATGTTTGA TTATCTGGTG CCAAACGTTA ACTTTTTTGG CCCCAACGCC | 60 |
| 20 | ATTTCCGTAG TCGGCGAACG CTGCCAGCTG CTGGGGGGGA AAAAAGCCCT GCTGGTCACC | 120 |
| 25 | GACAAAGGCC TGCGGGCAAT TAAAGATGGC GCGGTGGACA AAACCCTGCA TTATCTGCGG | 180 |
| 30 | GAGGCCGGGA TCGAGGTGGC GATCTTTGAC GCGGTCGAGC CGAACCCGAA AGACACCAAC | 240 |
| 35 | GTGCGCGACG GCCTCGCCGT GTTTCGCCGC GAACAGTGGC ACATCATCGT CACCGTGGGC | 300 |
| 40 | GGCGGCAGCC CGCACGATTG CGGCAAAGGC ATCGGCATCG CCGCCACCCA TGAGGGCGAT | 360 |
| 45 | CTGTACCAGT ATGCCGGAAT CGAGACCCTG ACCAACCCGC TGCCGCCTAT CGTCGCGGTC | 420 |
| 50 | AATACCACCG CCGGCACCGC CAGCGAGGTC ACCCGCCACT GCGTCCTGAC CAACACCGAA | 480 |
| 55 | ACCAAAGTGA AGTTTGTGAT CGTCAGCTGG CGCAAAGTGC CGTCGGTCTC TATCAACGAT | 540 |
| 60 | CCACTGCTGA TGATCGGTAA ACCGGCCGCC CTGACCGCGG CGACCGGGAT GGATGCCCTG | 600 |
| 65 | ACCCACGCCG TAGAGGCCTA TATCTCCAAA GACGCTAACC CCGTGACGGA CGCCGCCGCC | 660 |
| 70 | ATGCAGGCGA TCCGCCTCAT CGCCCGCAAC CTGCGCCAGG CCGTGGCCCT CGGCAGCAAT | 720 |
| 75 | CTGCAGGCGC GGGAAAACAT GGCCTATGCT TCTCTGCTGG CCGGGATGGC TTTCAATAAC | 780 |
| 80 | GCCAACCTCG GCTACGTGCA CGCCATGGCG CACCAGCTGG GCGGCCTGTA CGACATGCCG | 840 |
| 85 | CACGGCGTGG CCAACGCTGT CCTGCTGCCG CATGTGGCGC GCTACAACCT GATCGCCAAC | 900 |
| 90 | CCGGAGAAAT TCGCCGATAT CGCTGAACTG ATGGGCGAAA ATATCACCGG ACTGTCCACT | 960 |
| 95 | CTCGACGCGG CGGAAAAGC CATCGCCGCT ATCACGCGTC TGTCGATGGA TATCGGTATT | 1020 |
| 100 | CCGCAGCATC TGC GCGATCT GGGGGTAAAA GAGGCCGACT TCCCCTACAT GCGGAGATG | 1080 |
| 105 | GCTCTAAAAG ACGGCAATGC GTTCTCGAAC CCGCGTAAAG GCAACGAGCA GGAGATTGCC | 1140 |
| 110 | GCGATTTTCC GCCAGGCATT CTGA | 1164 |

(2) INFORMACIÓN PARA LA SEC. ID NO:5:

(i) CARACTERÍSTICAS DE LA SECUENCIA:

- (A) LONGITUD: 1380 pares de bases
- (B) TIPO: ácido nucleico
- (C) TIPO DE CADENA: simple
- (D) TIPOLOGÍA: lineal

ES 2 336 858 T3

(ii) TIPO DE MOLÉCULA: DNA (genómico)

(vi) FUENTE ORIGINAL:

(A) ORGANISMO: GPD1

5

(xi) DESCRIPCIÓN DE LA SECUENCIA: SEC ID NO:5:

CTTTAATTTT CTTTATCTT ACTCTCCTAC ATAAGACATC AAGAAACAAT TGATATTGT 60
10 ACACCCCCC CCTCCACAAA CACAAATATT GATAATATAA AGATGTCTGC TGCTGCTGAT 120
AGATTAAACT TAACTTCCGG CCACTTGAAT GCTGGTAGAA AGAGAAGTTC CTCTTCTGTT 180
15 TCTTTGAAGG CTGCCGAAAA GCCTTCAAG GTTACTGTGA TTGGATCTGG TAACTGGGGT 240
ACTACTATTG CCAAGGTGGT TGCCGAAAAT TGTAAGGGAT ACCCAGAAGT TTTCGCTCCA 300
ATAGTACAAA TGTGGGTGTT CGAAGAAGAG ATCAATGGTG AAAAATTGAC TGAATCATA 360
20 AATACTAGAC ATCAAAACGT GAAATACTTG CCTGGCATCA CTCTACCCGA CAATTTGGTT 420
GCTAATCCAG ACTTGATTGA TTCAGTCAAG GATGTGACACA TCATCGTTTT CAACATTCCA 480
CATCAATTTT TGCCCGTAT CTGTAGCCAA TTGAAAGGTC ATGTTGATTC ACACGTCAGA 540
25 GCTATCTCCT GTCTAAAGGG TTTTGAAGTT GGTGCTAAAG GTGTCCAATT GCTATCCTCT 600
TACATCACTG AGGAACTAGG TATTCAATGT GGTGCTCTAT CTGGTGCTAA CATTGCCACC 660
GAAGTCGCTC AAGAACACTG GTCTGAAACA ACAGTTGCTT ACCACATTCC AAAGGATTTTC 720
AGAGGCGAGG GCAAGGACGT CGACCATAAG GTTCTAAAGG CCTTGTCCA CAGACCTTAC 780
TTCCACGTTA GTGCATCGA AGATGTTGCT GGTATCTCCA TCTGTGGTGC TTTGAAGAAC 840
35 GTTGTGCTT TAGGTTGTGG TTTCGTCGAA GGTCTAGGCT GGGGTAACAA CGCTTCTGCT 900
GCCATCCAAA GAGTCGGTTT GGGTGAGATC ATCAGATTCG GTCAAATGTT TTTCCAGAA 960
TCTAGAGAAG AAACATACTA CCAAGAGTCT GCTGGTGTG CTGATTTGAT CACCACCTGC 1020
40 GCTGGTGGTA GAAACGTCAA GGTGCTAGG CTAATGGCTA CTCTGGTAA GGACGCCTGG 1080
GAATGTGAAA AGGAGTTGTT GAATGGCCAA TCCGCTCAAG GTTTAATTAC CTGCAAAGAA 1140
45 GTTCACGAAT GGTGGAAC ATGTGGCTCT GTCGAAGACT TCCCATTATT TGAAGCCGTA 1200
TACCAAATCG TTTACAACAA CTACCCAATG AAGAACCTGC CGGACATGAT TGAAGAATTA 1260
GATCTACATG AAGATTAGAT TTATTGGAGA AAGATAACAT ATCATACTTC CCCCCTTTT 1320
50 TTCGAGGCTC TTCTATATCA TATTCATAAA TTAGCATTAT GTCATTTCTC ATAACTACTT 1380

(2) INFORMACIÓN PARA LA SEC. ID NO:6:

55

(i) CARACTERÍSTICAS DE LA SECUENCIA:

(A) LONGITUD: 2946 pares de bases

(B) TIPO: ácido nucleico

60

(C) TIPO DE CADENA: simple

(D) TIPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: DNA (genómico)

65

(vi) FUENTE ORIGINAL:

(A) ORGANISMO: GPD2

ES 2 336 858 T3

(xi) DESCRIPCIÓN DE LA SECUENCIA: SEC ID NO:6:

| | | |
|----|--|------|
| | GAATTCGAGC CTGAAGTGC ^T GATTACCTTC AGGTAGACTT CATCTTGACC CATCAACCCC | 60 |
| 5 | AGCGTCAATC CTGCAAATAC ACCACCCAGC AGCACTAGGA TGATAGAGAT AATATAGTAC | 120 |
| | GTGGTAACGC TTGCCTCATC ACCTACGCTA TGGCCGGAAT CGGCAACATC CCTAGAATTG | 180 |
| | AGTACGTGTG ATCCGGATAA CAACGGCAGT GAATATATCT TCGGTATCGT AAAGATGTGA | 240 |
| 10 | TATAAGATGA TGTATACCCA ATGAGGAGCG CCTGATCGTG ACCTAGACCT TAGTGGCAAA | 300 |
| | AACGACATAT CTATTATAGT GGGGAGAGTT TCGTGCAAAT AACAGACGCA GCAGCAAGTA | 360 |
| 15 | ACTGTGACGA TATCAACTCT TTTTTTATTA TGTAATAAGC AAACAAGCAC GAATGGGGAA | 420 |
| | AGCCTATGTG CAATCACCAA GGTCGTCCCT TTTTTCCCAT TTGCTAATTT AGAATTTAAA | 480 |
| | GAAACCAAAA GAATGAAGAA AGAAAACAAA TACTAGCCCT AACCTGACT TCGTTTCTAT | 540 |
| 20 | GATAATACCC TGCTTTAATG AACGGTATGC CCTAGGGTAT ATCTCACTCT GTACGTTACA | 600 |
| | A A CTCCGGTT ATTTTATCGG AACATCCGAG CACCCGCGCC TTCCTCAACC CAGGCACCGC | 660 |
| 25 | CCAGGTAAC CGTGCGCGAT GAGCTAATCC TGAGCCATCA CCCACCCAC CCGTTGATGA | 720 |
| | CAGCAATT C G GGAGGGCGAA AATAAACTG GAGCAAGGAA TTACCATCAC CGTCACCATC | 780 |
| | ACCATCATAT CGCCTTAGCC TCTAGCCATA GCCATCATGC AAGCGTGTAT CTTCTAAGAT | 840 |
| 30 | TCAGTCATCA TCATTACCGA GTTTGTTTTT CTTACATGA TGAAGAAGGT TTGAGTATGC | 900 |
| | TCGAAACAAT AAGACGACGA TGGCTCTGCC ATTGGTTATA TTACGCTTTT GCGGCGAGGT | 960 |
| | GCCGATGGGT TGCTGAGGGG AAGAGTGTTT AGCTTACGGA CCTATTGCCA TTGTTATTCC | 1020 |
| 35 | GATTAATCTA TTGTTACGCA GCTCTTCTCT ACCCTGTCAT TCTAGTATTT TTTTTTTTTT | 1080 |
| | TTTTTGTTT TACTTTTTTT TCTTCTTGCC TTTTTTCTT GTTACTTTTT TTCTAGTTTT | 1140 |
| 40 | TTTTCTTCC ACTAAGCTTT TTCCTTGATT TATCCTGGG TTCTTCTTTC TACTCCTTTA | 1200 |
| | GATTTTTTTT TTATATATTA ATTTTAAAGT TTATGTATTT TGGTAGATTC AATTCTCTTT | 1260 |
| | CCCTTTCCTT TTCCTTCGCT CCCCTTCCTT ATCAATGCTT GCTGTCAGAA GATTAACAAG | 1320 |
| 45 | ATACACATTC CTTAAGCGAA CGCATCCGGT GTTATATACT CGTCGTGCAT AAAAAATTTT | 1380 |
| | GCCTTCAAGA TCTACTTTCC TAAGAAGATC ATTATTACAA ACACAACCTGC ACTCAAAGAT | 1440 |
| 50 | GACTGCTCAT ACTAATATCA AACAGCACAA ACACTGTCAT GAGGACCATC CTATCAGAAG | 1500 |
| | ATCGGACTCT GCCGTGTCAA TTGTACATTT GAAACGTGCG CCCTTCAAGG TTACAGTGAT | 1560 |
| | TGGTCTGGT AACTGGGGGA CCACCATCGC CAAAGTCATT GCGGAAAACA CAGAATTGCA | 1620 |
| 55 | TTCCCATATC TTCGAGCCAG AGGTGAGAA ^T GTGGGTTTTT GATGAAAAGA TCGGCGACGA | 1680 |

60

65

ES 2 336 858 T3

| | | |
|----|--|------|
| | AAATCTGACG GATATCATAA ATACAAGACA CCAGAACGTT AAATATCTAC CCAATATTGA | 1740 |
| | CCTGCCCCAT AATCTAGTGG CCGATCCTGA TCTTTTACAC TCCATCAAGG GTGCTGACAT | 1800 |
| 5 | CCTTGTTTTT AACATCCCTC ATCAATTTTT ACCAAACATA GTCAAACAAT TGCAAGGCCA | 1860 |
| | CGTGGCCCCT CATGTAAGGG CCATCTCGTG TCTAAAAGGG TTCGAGTTGG GCTCCAAGGG | 1920 |
| | TGTGCAATTG CTATCCTCCT ATGTTACTGA TGAGTTAGGA ATCCAATGTG GCGCACTATC | 1980 |
| 10 | TGGTGCAAAC TTGGCACC GG AAGTGGCCAA GGAGCATTGG TCCGAAACCA CCGTGGCTTA | 2040 |
| | CCAAC TACCA AAGGATTATC AAGGTGATGG CAAGGATGTA GATCATAAGA TTTTGAAATT | 2100 |
| 15 | GCTGTTCCAC AGACCTTACT TCCACGTCAA TGTGATCGAT GATGTTGCTG GTATATCCAT | 2160 |
| | TGCCGGTGCC TTGAAGAACG TCGTGGCACT TGCATGTGGT TTCGTAGAAG GTATGGGATG | 2220 |
| | GGGTAACAAT GCCTCCGCAG CCATTCAAAG GCTGGGTTTA GGTGAAATTA TCAAGTTCGG | 2280 |
| 20 | TAGAATGTTT TTCCCAGAAT CCAAAGTCGA GACCTACTAT CAAGAATCCG CTGGTGTTC | 2340 |
| | AGATCTGATC ACCACCTGCT CAGGCGGTAG AAACGTCAAG GTTGCCACAT ACATGGCCAA | 2400 |
| 25 | GACCGGTAAG TCAGCCTTGG AAGCAGAAAA GGAATTGCTT AACGGTCAAT CCGCCCAAGG | 2460 |
| | GATAATCACA TGCAGAGAAG TTCACGAGTG GCTACAAACA TGTGAGTTGA CCCAAGAATT | 2520 |
| | CCCAATTATT CGAGGCAGTC TACCAGATAG TCTACAACAA CGTCCGCATG GAAGACCTAC | 2580 |
| 30 | CGGAGATGAT TGAAGAGCTA GACATCGATG ACGAATAGAC ACTCTCCCC CCCCTCCCC | 2640 |
| | TCTGATCITT CCTGTTGCCT CTTTTTCCCC CAACCAATTT ATCATTATAC ACAAGTTCTA | 2700 |
| | CAACTACTAC TAGTAACATT ACTACAGTTA TTATAATTTT CTATTCTCTT TTTCTTTAAG | 2760 |
| 35 | AATCTATCAT TAACGTTAAT TTCTATATAT ACATAACTAC CATTATACAC GCTATTATCG | 2820 |
| | TTTACATATC ACATCACCGT TAATGAAAGA TACGACACCC TGTACTACTAA CACAATTAAG | 2880 |
| 40 | TAATCGCCAT AACCTTTTCT GTTATCTATA GCCCTTAAAG CTGTTTCTTC GAGCTTTTCA | 2940 |
| | CTGCAG | 2946 |

(2) INFORMACIÓN PARA LA SEC. ID NO:7:

- 45 (i) CARACTERÍSTICAS DE LA SECUENCIA:
- (A) LONGITUD: 3178 pares de bases
 - (B) TIPO: ácido nucleico
 - 50 (C) TIPO DE CADENA: simple
 - (D) TIPOLOGÍA: lineal
- (ii) TIPO DE MOLÉCULA: DNA (genómico)
- 55 (vi) FUENTE ORIGINAL:
- (A) ORGANISMO: GUT2
- (xi) DESCRIPCIÓN DE LA SECUENCIA: SEC ID NO:7:

| | | |
|----|---|-----|
| | CTGCAGAACT TCGTCTGCTC TGTGCCCATC CTCGCGGTTA GAAAGAAGCT GAATTGTTTC | 60 |
| | ATGCGCAAGG GCATCAGCGA GTGACCAATA ATCACTGCAC TAATTCCTTT TTAGCAACAC | 120 |
| 65 | ATACTTATAT ACAGCACCAG ACCTTATGTC TTTTCTCTGC TCCGATACGT TATCCCACCC | 180 |
| | AACTTTTATT TCAGTTTTGG CAGGGGAAAT TTCACAACCC CGCACGCTAA AAATCGTATT | 240 |

ES 2 336 858 T3

TAAACTTAAA AGAGAACAGC CACAAATAGG GAACTTTGGT CTAAACGAAG GACTCTCCCT 300
 CCCTTATCTT GACCGTGCTA TTGCCATCAC TGCTACAAGA CTAAATACGT ACTAATATAT 360
 5 GTTTTCGGTA ACGAGAAGAA GAGCTGCCGG TGCAGCTGCT GCCATGGCCA CAGCCACGGG 420
 GACGCTGTAC TGGATGACTA GCCAAGGTGA TAGGCCGTTA GTGCACAATG ACCCGAGCTA 480
 10 CATGGTGCAA TTCCCCACCG CCGCTCCACC GGCAGGTCTC TAGACGAGAC CTGCTGGACC 540
 GTCTGGACAA GACGCATCAA TTCGACGTGT TGATCATCGG TGGCGGGGCC ACGGGGACAG 600
 GATGTGCCCT AGATGCTGCG ACCAGGGGAC TCAATGTGGC CTTGTTGAA AAGGGGGATT 660
 15 TTGCCTCGGG AACGTCGTCC AAATCTACCA AGATGATTCA CGGTGGGGTG CGGTACTTAG 720
 AGAAGGCCTT CTGGGAGTTC TCCAAGGCAC AACTGGATCT GGTCATCGAG GCACTCAACG 780
 20 AGCGTAAACA TCTTATCAAC ACTGCCCTC ACCTGTGCAC GGTGCTACCA ATTCTGATCC 840
 CCATCTACAG CACCTGGCAG GTCCCGTACA TCTATATGGG CTGTAAATC TACGATTTCT 900
 TTGGCGGTTT CCAAACCTTG AAAAAATCAT ACCTACTGTC CAAATCCGCC ACCGTGGAGA 960
 25 AGGCTCCCAT GCTTACCACA GACAATTTAA AGGCCTCGCT TGTGTACCAT GATGGGTCCCT 1020
 TTAACGACTC GCGTTTGAAC GCCACTTTAG CCATCACGGG TGTGGAGAAC GGCCTACCG 1080
 30 TCTTGATCTA TGTCGAGGTA CAAAAATTGA TCAAAGACCC AACTTCTGGT AAGGTTATCG 1140
 GTGCCGAGGC CCGGGACGTT GAGACTAATG AGCTTGTGAG AATCAACGCT AAATGTGTGG 1200
 TCAATGCCAC GGGCCCATAC AGTGACGCCA TTTTGCAAAT GGACCGCAAC CCATCCGGTC 1260
 35 TGCCGGACTC CCCGCTAAAC GACAACTCCA AGATCAAGTC GACTTTCAAT CAAATCTCCG 1320
 TCATGGACCC GAAAATGGTC ATCCCATCTA TTGGCGTTCA CATCGTATTG CCCTCTTTTT 1380
 40 ACTCCCCGAA GGATATGGGT TTGTTGGACG TCAGAACCTC TGATGGCAGA GTGATGTTCT 1440
 TTTTACCTTG GCAGGGCAA GTCCTTGCCG GCACCACAGA CATCCCCTA AAGCAAGTCC 1500
 CAGAAAACCC TATGCCTACA GAGGCTGATA TTCAAGATAT CTTGAAAGAA CTACAGCACT 1560
 45 ATATCGAATT CCCCCTGAAA AGAGAAGACG TGCTAAGTGC ATGGGCTGGT GTCAGACCTT 1620
 TGGTCAGAGA TCCACGTACA ATCCCCGCAG ACGGGAAGAA GGGCTCTGCC ACTCAGGGCG 1680
 50 TGSTAAGATC CCACTTCTTG TTCACTTCGG ATAATGGCCT AATTACTATT GCAGGTGGTA 1740
 AATGGACTAC TTACAGACPA ATGGCTGAGG AACAGTCSA CAAAGTTGTC GAAGTTGGCG 1800
 GATTCCACAA CCTGAAACCT TGTCACACAA GAGATATTAA GCTTGCTGGT GCAGAAGAAT 1860
 55 GGACGCAAAA CTATGTGGCT TTATTGGCTC AAAACTACCA TTTATCATCA AAAATGTCCA 1920
 ACTACTTGGT TCAAACTAC GGAACCCGTT CCTCTATCAT TTGCGAATTT TTCAAAGAAT 1980
 60 CCATGGAAAA TAAACTGCCT TTGTCCTTAG CCGACAAGGA AAATAACGTA ATCTACTCTA 2040
 GCGAGGAGAA CAACTGGTC AATTTTGATA CTTTCAGATA TCCATTACA ATCGGTGAGT 2100
 TAAAGTATTC CATGCAGTAC GAATATTGTA GAACTCCCTT GGAATCCTT TTAAGAAGAA 2160
 65 CAAGATTGCG CTTCTTGGAC GCCAAGGAAG CTTTGAATGC CGTGCATGCC ACCGTCAAAG 2220

ES 2 336 858 T3

5
10
15
20
25
30

```
TTATGGGTGA TGAGTTCAAT TGGTCGGAGA AAAAGAGGCA GTGGGAACTT GAAAAAACTG 2280
TGAAC TTCAT CCAAGGACGT TTCGGTGTCT AAATCGATCA TGATAGTTAA GGGTGACAAA 2340
GATAACATTC ACAAGAGTAA TAATAATGGT AATGATGATA ATAATAATAA TGATAGTAAT 2400
AACATAATA ATAATGGTGG TAATGGCAAT GAAATCGCTA TTATTACCTA TTTTCCTTAA 2460
TGGAAGAGTT AAAGTAAACT AAAAAAATA CAAAAATATA TGAAGAAAAA AAAAAAAGA 2520
GGTAATAGAC TCTACTACTA CAATTGATCT TCAAATTATG ACCTTCCTAG TGTTTATATT 2580
CTATTTCCAA TACATAATAT AATCTATATA ATCATTGCTG GTAGACTTCC GTTTTAATAT 2640
CGTTTTAATT ATCCCCTTTA TCTCTAGTCT AGTTTTATCA TAAAATATAG AAACACTAAA 2700
TAATATTCTT CAAACGGTCC TGGTGCATAC GCAATACATA TTTATGGTGC AAAAAAAAAA 2760
ATGGAAAATT TTGCTAGTCA TAAACCCTTT CATAAAACAA TACGTAGACA TCGCTACTTG 2820
AAATTTTCAA GTTTTTATCA GATCCATGTT TCCTATCTGC CTTGACAACC TCATCGTCGA 2880
AATAGTACCA TTTAGAACGC CCAATATTCA CATTGTGTTT AAGGTCTTTA TTCACCAGTG 2940
ACGTGTAATG GCCATGATTA ATGTGCCTGT ATGGTTAACC ACTCCAAATA GCTTATATTT 3000
CATAGTGTCA TTGTTTTTCA ATATAATGTT TAGTATCAAT GGATATGTTA CGACGGTGTT 3060
ATTTTTCTTG GTCAAATCGT AATAAAATCT CGATAAATGG ATGACTAAGA TTTTGGTAA 3120
AGTTACAAA TTTATCGTTT TCACTGTTGT CAATTTTTTG TTCTTGTAAT CACTCGAG 3178
```

35 (2) INFORMACIÓN PARA LA SEC. ID NO:8:

(i) CARACTERÍSTICAS DE LA SECUENCIA:

(A) LONGITUD: 816 pares de bases

40 (B) TIPO: ácido nucleico

(C) TIPO DE CADENA: simple

(D) TIPOLOGÍA: lineal

45 (ii) TIPO DE MOLÉCULA: DNA (genómico)

(vi) FUENTE ORIGINAL:

(A) ORGANISMO: GPP1

50 (xi) DESCRIPCIÓN DE LA SECUENCIA: SEC ID NO:8:

55

60

65

ES 2 336 858 T3

| | | |
|----|--|-----|
| | ATGAAACGTT TCAATGTTTT AAAATATATC AGAACAACAA AAGCAAATAT ACAAACCATC | 60 |
| 5 | GCAATGCCTT TGACCACAAA ACCTTTATCT TTGAAAATCA ACGCCGCTCT ATTCGATGTT | 120 |
| | GACGGTACCA TCATCATCTC TCAACCAGCC ATTGCTGCTT TCTGGAGAGA TTTCGGTAAA | 180 |
| | GACAAGCCTT ACTTCGATGC CGAACACGTT ATTCACATCT CTCACGTTG GAGAACTTAC | 240 |
| 10 | GATGCCATTG CCAAGTTCGC TCCAGACTTT GCTGATGAAG AATACGTAA CAAGCTAGAA | 300 |
| | GGTGAAATCC CAGAAAAGTA CGGTGAACAC TCCATCGAAG TTCCAGGTGC TGTCAAGTTG | 360 |
| | TGTAATGCTT TGAACGCCTT GCCAAAGGAA AAATGGGCTG TCGCCACCTC TGGTACCCGT | 420 |
| 15 | GACATGGCCA AGAAATGGTT CGACATTTTG AAGATCAAGA GACCAGAATA CTTTCATCACC | 480 |
| | GCCAATGATG TCAAGCAAGG TAAGCCTCAC CCAGAACCAT ACTTAAAGGG TAGAAACGGT | 540 |
| 20 | TTGGGTTTCC CAATTAATGA ACAAGACCCA TCCAAATCTA AGGTTGTTGT CTTTGAAGAC | 600 |
| | GCACCAGCTG GTATTGCTGC TGGAAGGCT GCTGGCTGTA AAATCGTTGG TATTGCTACC | 660 |
| 25 | ACTTTCGATT TGGACTTCTT GAAGGAAAAG GGTGTGACA TCATTGTCAA GAACCACGAA | 720 |
| | TCTATCAGAG TCGGTGAATA CAACGCTGAA ACCGATGAAG TCGAATTGAT CTTTGATGAC | 780 |
| 30 | TACTTATACG CTAAGGATGA CTTGTTGAAA TGGTAA | 816 |

(2) INFORMACIÓN PARA LA SEC. ID NO:9:

- 35 (i) CARACTERÍSTICAS DE LA SECUENCIA:
- (A) LONGITUD: 753 pares de bases
 - (B) TIPO: ácido nucleico
 - 40 (C) TIPO DE CADENA: simple
 - (D) TIPOLOGÍA: lineal
- (ii) TIPO DE MOLÉCULA: DNA (genómico)
- 45 (vi) FUENTE ORIGINAL:
- (A) ORGANISMO: GPP2
- (xi) DESCRIPCIÓN DE LA SECUENCIA: SEC ID NO:9:
- 50
- 55
- 60
- 65

ES 2 336 858 T3

ATGGGATTGA C TACTAAACC TCTATCTTTG AAAGTTAACG CCGCTTTGTT CGACGTCGAC 60
GGTACCATTA TCATCTCTCA ACCAGCCATT GCTGCATTCT GGAGGGATTT CGGTAAGGAC 120
5 AAACCTTATT TCGATGCTGA ACACGTTATC CAAGTCTCGC ATGGTTGGAG AACGTTTGAT 180
GCCATTGCTA AGTTCGCTCC AGACTTTGCC AATGAAGAGT ATGTTAACAA ATTAGAAGCT 240
10 GAAATTCGGG TCAAGTACGG TGAAAAATCC ATTGAAGTCC CAGGTGCAGT TAAGCTGTGC 300
AACGCTTTGA ACGCTCTACC AAAAGAGAAA TGGGCTGTGG CAACTTCGGG TACCCGTGAT 360
ATGGCACAAA AATGGTTCGA GCATCTGGGA ATCAGGAGAC CAAAGTACTT CATTACCGCT 420
15 AATGATGTCA AACAGGGTAA GCCTCATCCA GAACCATATC TGAAGGGCAG GAATGCCTTA 480
GGATATCCGA TCAATGAGCA AGACCCTTCC AAATCTAAGG TAGTAGTATT TGAAGACGCT 540
20 CCAGCAGGTA TTGCCGCCGG AAAAGCCGCC GGTTGTAAGA TCATTGGTAT TGCCACTACT 600
TTCGACTTGG ACTTCCTAAA GGAAAAAGGC TGTGACATCA TTGTCAAAAA CCACGAATCC 660
ATCAGAGTTG GCGGCTACAA TGCCGAAACA GACGAAGTTG AATTCATTTT TGACGACTAC 720
25 TTATATGCTA AGGACGACTCT GTTCAAATGG TAA 753

(2) INFORMACIÓN PARA LA SEC. ID NO:10:

30

(i) CARACTERÍSTICAS DE LA SECUENCIA:

(A) LONGITUD: 2520 pares de bases

(B) TIPO: ácido nucleico

35

(C) TIPO DE CADENA: simple

(D) TIPOLOGÍA: lineal

40

(ii) TIPO DE MOLÉCULA: DNA (genómico)

(vi) FUENTE ORIGINAL:

(A) ORGANISMO: GUT1

45

(xi) DESCRIPCIÓN DE LA SECUENCIA: SEC ID NO:10:

50

55

60

65

ES 2 336 858 T3

| | | |
|----|--|------|
| | TGTATTGGCC ACGATAACCA CCCTTTGTAT ACTGTTTTTG TTTTTCACAT GGTAATAAAC | 60 |
| | GACTTTTATT AAACAACGTA TGTA AAAACA TAACAAGAAT CTACCCATAC AGGCCATTTT | 120 |
| 5 | GTAATTCTTC TCTTCTAATT GGAGTAAAAC CATCAATTAA AGGGTGTGGA GTAGCATAGT | 180 |
| | GAGGGGCTGA CTGCATTGAC AAAAAAATTG AAAAAAAAAA AGGAAAAGGA AAGGAAAAAA | 240 |
| 10 | AGACAGCCAA GACTTTTAGA ACGGATAAGG TGTAATAAAA TGTGGGGGGA TGCCCTGTTCT | 300 |
| | CGAACCATAT AAAATATACC ATGTGGTTTG AGTTGTGGCC GGAACTATAC AAATAGTTAT | 360 |
| | ATGTTTCCCT CTCTCTCCG ACTTGTAGTA TTCTCCAAAC GTTACATATT CCGATCAAGC | 420 |
| 15 | CAGCGCCTTT AACTAGTTT AAAACAAGAA CAGAGCCGTA TGCCAAAAT AATGGAAGAT | 480 |
| | TTACGAAGTG ACTACGTCCC GCTTATCGCC AGTATTGATG TAGGAACGAC CTCATCCAGA | 540 |
| | TGCATTCTGT TCAACAGATG GGGCCAGGAC GTTTCAAAAC ACCAAATTGA ATATTCAACT | 600 |
| 20 | TCAGCATCGA AGGGCAAGAT TGGGGTGTCT GGCCTAAGGA GACCCTCTAC AGCCCCAGCT | 660 |
| | CGTGAAACAC CAAACGCCGG TGACATCAAA ACCAGCGGAA AGCCCATCTT TTCTGCAGAA | 720 |
| 25 | GGCTATGCCA TTCAAGAAAC CAAATTCCTA AAAATCGAGG AATTGGAATT GGACTTCCAT | 780 |
| | AACGAACCCA CGTTGAAGTT CCCCAAACCG GTTTGGGTTG AGTGCCATCC GCAGAAATTA | 840 |
| | CTGGTGAACG TCGTCCAATG CCTTGCCTCA AGTTTGCTCT CTCTGCAGAC TATCAACAGC | 900 |
| 30 | GAACGTGTAG CAAACGGTCT CCCACCTTAC AAGGTAATAT GCATGGGTAT AGCAAACATG | 960 |
| | AGAGAAACCA CAATTCTGTG GTCCCGCCGC ACAGGAAAAC CAATTGTTAA CTACGGTATT | 1020 |
| | GTTTGGAAACG ACACCAGAAC GATCAAAATC GTTAGAGACA AATGGCAAAA CACTAGCGTC | 1080 |
| 35 | GATAGGCAAC TGCAGCTTAG ACAGAAGACT GGATTGCCAT TGCTCTCCAC GTATTTCTCC | 1140 |
| | TGTTCCAAGC TCGCCTGGTT CCTCGACAAT GAGCCTCTGT GTACCAAGGC GTATGAGGAG | 1200 |
| 40 | AACGACCTGA TGTTCCGGCAC TGTGGACACA TGCTGATTT ACCAATTAAC TAAACAAAAG | 1260 |
| | GCGTTCGTTT CTGACGTAAC CAACGCTTCC AGAACTGGAT TTATGAACCT CTCCACTTTA | 1320 |
| | AAGTACGACA ACGAGTTGCT GGAATTTTGG GGTATTGACA AGAACCTGAT TCACATGCCC | 1380 |
| 45 | GAAATTGTGT CCTCATCTCA ATACTACGGT GACTTTGGCA TTCCTGATTG GATAATGGAA | 1440 |
| | AAGCTACACG ATTCCGCAAA AACAGTACTG CGAGATCTAG TCAAGAGAAA CCTGCCATA | 1500 |
| | CAGGGCTGTC TGGGCGACCA AAGCGCATCC ATGGTGGGGC AACTCGCTTA CAAACCCGGT | 1560 |
| 50 | GCTGCAAAAT GTACTTATGG TACCGGTTGC TTTTACTGT ACAATACGGG GACCAAAAAA | 1620 |
| | TTGATCTCCC AACATGGCGC ACTGACGACT CTAGCATTTT GGTCCACACA TTGCAAGAG | 1680 |
| 55 | TACGGTGGCC AAAAACCAGA ATTGAGCAAG CCACATTTG CATTAGAGGG TTCCGTCGCT | 1740 |
| | GTGGCTGGTG CTGTGGTCCA ATGGCTACGT GATAATTTAC GATTGATCGA TAAATCAGAG | 1800 |
| | GATGTCCGAC CGATTGCATC TACGGTTCTT GATTCTGGTG GCGTAGTTT CGTCCCGCA | 1860 |
| 60 | TTTAGTGGCC TATTCGCTCC CTATTGGGAC CCAGATGCCA GAGCCACCAT AATGGGGATG | 1920 |

65

ES 2 336 858 T3

TCICAATTCA CTACTGCCTC CCACATCGCC AGAGCTGCCG TGGAAGGTGT TTGCTTTCAA 1980
 GCCAGGGCTA TCTTGAAGGC AATGAGTTCT GACGCGTTTG GTGAAGGTTC CAAAGACAGG 2040
 5 GACTTTTTAG AGGAAATTC CGACGTCACA TATGAAAAGT CGCCCCTGTC GGTTCTGGCA 2100
 GTGGATGGCG GGATGTCGAG GTCTAATGAA GTCATGCAAA TTCAAGCCGA TATCCTAGGT 2160
 CCCTGTGTCA AAGTCAGAAG GTCTCCGACA GCGGAATGTA CCGCATTGGG GGCAGCCATT 2220
 10 GCAGCCAATA TGGCTTTCAA GGATGTGAAC GAGCGCCCAT TATGGAAGGA CCTACACGAT 2280
 GTTAAGAAAT GGGTCTTTTA CAATGGAATG GAGAAAAACG AACAAATATC ACCAGAGGCT 2340
 15 CATCCAAACC TTAAGATATT CAGAAGTGAA TCCGACGATG CTGAAAGGAG AAAGCATTGG 2400
 AAGTATTGGG AAGTTGCCGT GGAAAGATCC AAAGGTTGGC TGAAGGACAT AGAAGGTGAA 2460
 CACGAACAGG TTCTAGAAA CTCCAATAA CAACATAAAT AATTTCTATT AACAATGTAA 2520

20 (2) INFORMACIÓN PARA LA SEC. ID NO:11:

(i) CARACTERÍSTICAS DE LA SECUENCIA:

- (A) LONGITUD: 391 amino ácidos
- (B) TIPO: amino ácido
- (C) TIPO DE CADENA: desconocida
- (D) TIPOLOGÍA: desconocida

30 (ii) TIPO DE MOLÉCULA: proteína

(vi) FUENTE ORIGINAL:

- (A) ORGANISMO: GPD1

35 (xi) DESCRIPCIÓN DE LA SECUENCIA: SEC ID NO:11:

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

ES 2 336 858 T3

Val Gln Leu Leu Ser Ser Tyr Ile Thr Glu Glu Leu Gly Ile Gln Cys
 165 170 175
 Gly Ala Leu Ser Gly Ala Asn Ile Ala Thr Glu Val Ala Gln Glu His
 180 185 190
 5 Trp Ser Glu Thr Thr Val Ala Tyr His Ile Pro Lys Asp Phe Arg Gly
 195 200 205
 10 Glu Gly Lys Asp Val Asp His Lys Val Leu Lys Ala Leu Phe His Arg
 210 215 220
 Pro Tyr Phe His Val Ser Val Ile Glu Asp Val Ala Gly Ile Ser Ile
 225 230 235 240
 15 Cys Gly Ala Leu Lys Asn Val Val Ala Leu Gly Cys Gly Phe Val Glu
 245 250 255
 Gly Leu Gly Trp Gly Asn Asn Ala Ser Ala Ala Ile Gln Arg Val Gly
 260 265 270
 20 Leu Gly Glu Ile Ile Arg Phe Gly Gln Met Phe Phe Pro Glu Ser Arg
 275 280 285
 Glu Glu Thr Tyr Tyr Gln Glu Ser Ala Gly Val Ala Asp Leu Ile Thr
 290 295 300
 25 Thr Cys Ala Gly Gly Arg Asn Val Lys Val Ala Arg Leu Met Ala Thr
 305 310 315 320
 Ser Gly Lys Asp Ala Trp Glu Cys Glu Lys Glu Leu Leu Asn Gly Gln
 325 330 335
 30 Ser Ala Gln Gly Leu Ile Thr Cys Lys Glu Val His Glu Trp Leu Glu
 340 345 350
 Thr Cys Gly Ser Val Glu Asp Phe Pro Leu Phe Glu Ala Val Tyr Gln
 355 360 365
 35 Ile Val Tyr Asn Asn Tyr Pro Met Lys Asn Leu Pro Asp Met Ile Glu
 370 375 380
 40 Glu Leu Asp Leu His Glu Asp
 385 390

(2) INFORMACIÓN PARA LA SEC. ID NO:12:

45 (i) CARACTERÍSTICAS DE LA SECUENCIA:

(A) LONGITUD: 384 amino ácidos

(B) TIPO: amino ácido

50 (C) TIPO DE CADENA: desconocida

(D) TIPOLOGÍA: desconocida

(ii) TIPO DE MOLÉCULA: proteína

55 (vi) FUENTE ORIGINAL:

(A) ORGANISMO: GPD2

(xi) DESCRIPCIÓN DE LA SECUENCIA: SEC ID NO:12:

60 Met Thr Ala His Thr Asn Ile Lys Gln His Lys His Cys His Glu Asp
 1 5 10 15
 His Pro Ile Arg Arg Ser Asp Ser Ala Val Ser Ile Val His Leu Lys
 20 25 30
 65

ES 2 336 858 T3

Arg Ala Pro Phe Lys Val Thr Val Ile Gly Ser Gly Asn Trp Gly Thr
 35 40 45
 5 Thr Ile Ala Lys Val Ile Ala Glu Asn Thr Glu Leu His Ser His Ile
 50 55 60
 Phe Glu Pro Glu Val Arg Met Trp Val Phe Asp Glu Lys Ile Gly Asp
 65 70 75 80
 10 Glu Asn Leu Thr Asp Ile Ile Asn Thr Arg His Gln Asn Val Lys Tyr
 85 90 95
 Leu Pro Asn Ile Asp Leu Pro His Asn Leu Val Ala Asp Pro Asp Leu
 100 105 110
 15 Leu His Ser Ile Lys Gly Ala Asp Ile Leu Val Phe Asn Ile Pro His
 115 120 125
 20 Gln Phe Leu Pro Asn Ile Val Lys Gln Leu Gln Gly His Val Ala Pro
 130 135 140
 His Val Arg Ala Ile Ser Cys Leu Lys Gly Phe Glu Leu Gly Ser Lys
 145 150 155 160
 25 Gly Val Gln Leu Leu Ser Ser Tyr Val Thr Asp Glu Leu Gly Ile Gln
 165 170 175
 Cys Gly Ala Leu Ser Gly Ala Asn Leu Ala Pro Glu Val Ala Lys Glu
 180 185 190
 30 His Trp Ser Glu Thr Thr Val Ala Tyr Gln Leu Pro Lys Asp Tyr Gln
 195 200 205
 35 Gly Asp Gly Lys Asp Val Asp His Lys Ile Leu Lys Leu Leu Phe His
 210 215 220
 Arg Pro Tyr Phe His Val Asn Val Ile Asp Asp Val Ala Gly Ile Ser
 225 230 235 240
 40 Ile Ala Gly Ala Leu Lys Asn Val Val Ala Leu Ala Cys Gly Phe Val
 245 250 255
 45 Glu Gly Met Gly Trp Gly Asn Asn Ala Ser Ala Ala Ile Gln Arg Leu
 260 265 270
 Gly Leu Gly Glu Ile Ile Lys Phe Gly Arg Met Phe Phe Pro Glu Ser
 275 280 285
 50 Lys Val Glu Thr Tyr Tyr Gln Glu Ser Ala Gly Val Ala Asp Leu Ile
 290 295 300
 Thr Thr Cys Ser Gly Gly Arg Asn Val Lys Val Ala Thr Tyr Met Ala
 305 310 315 320
 55 Lys Thr Gly Lys Ser Ala Leu Glu Ala Glu Lys Glu Leu Leu Asn Gly
 325 330 335
 60 Gln Ser Ala Gln Gly Ile Ile Thr Cys Arg Glu Val His Glu Trp Leu
 340 345 350
 Gln Thr Cys Glu Leu Thr Gln Glu Phe Pro Ile Ile Arg Gly Ser Leu
 355 360 365
 65 Pro Asp Ser Leu Gln Gln Arg Pro His Gly Arg Pro Thr Gly Asp Asp
 370 375 380

ES 2 336 858 T3

(2) INFORMACIÓN PARA LA SEC. ID NO:13:

(i) CARACTERÍSTICAS DE LA SECUENCIA:

- (A) LONGITUD: 614 amino ácidos
- (B) TIPO: amino ácido
- (C) TIPO DE CADENA: desconocida
- (D) TIPOLOGÍA: desconocida

(ii) TIPO DE MOLÉCULA: proteína

(vi) FUENTE ORIGINAL:

- (A) ORGANISMO: GUT2

(xi) DESCRIPCIÓN DE LA SECUENCIA: SEC ID NO:13:

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

ES 2 336 858 T3

Asn Ser Lys Ile Lys Ser Thr Phe Asn Gln Ile Ser Val Met Asp Pro
 275 280 285
 5 Lys Met Val Ile Pro Ser Ile Gly Val His Ile Val Leu Pro Ser Phe
 290 295 300
 Tyr Ser Pro Lys Asp Met Gly Leu Leu Asp Val Arg Thr Ser Asp Gly
 305 310 315 320
 10 Arg Val Met Phe Phe Leu Pro Trp Gln Gly Lys Val Leu Ala Gly Thr
 325 330 335
 Thr Asp Ile Pro Leu Lys Gln Val Pro Glu Asn Pro Met Pro Thr Glu
 340 345 350
 15 Ala Asp Ile Gln Asp Ile Leu Lys Glu Leu Gln His Tyr Ile Glu Phe
 355 360 365
 20 Pro Val Lys Arg Glu Asp Val Leu Ser Ala Trp Ala Gly Val Arg Pro
 370 375 380
 Leu Val Arg Asp Pro Arg Thr Ile Pro Ala Asp Gly Lys Lys Gly Ser
 385 390 395 400
 25 Ala Thr Gln Gly Val Val Arg Ser His Phe Leu Phe Thr Ser Asp Asn
 405 410 415
 Gly Leu Ile Thr Ile Ala Gly Gly Lys Trp Thr Thr Tyr Arg Gln Met
 420 425 430
 30 Ala Glu Glu Thr Val Asp Lys Val Val Glu Val Gly Gly Phe His Asn
 435 440 445
 35 Leu Lys Pro Cys His Thr Arg Asp Ile Lys Leu Ala Gly Ala Glu Glu
 450 455 460
 Trp Thr Gln Asn Tyr Val Ala Leu Leu Ala Gln Asn Tyr His Leu Ser
 465 470 475 480
 Ser Lys Met Ser Asn Tyr Leu Val Gln Asn Tyr Gly Thr Arg Ser Ser
 485 490 495
 45 Ile Ile Cys Glu Phe Phe Lys Glu Ser Met Glu Asn Lys Leu Pro Leu
 500 505 510
 Ser Leu Ala Asp Lys Glu Asn Asn Val Ile Tyr Ser Ser Glu Glu Asn
 515 520 525
 50 Asn Leu Val Asn Phe Asp Thr Phe Arg Tyr Pro Phe Thr Ile Gly Glu
 530 535 540
 Leu Lys Tyr Ser Met Gln Tyr Glu Tyr Cys Arg Thr Pro Leu Asp Phe
 545 550 555 560
 Leu Leu Arg Arg Thr Arg Phe Ala Phe Leu Asp Ala Lys Glu Ala Leu
 565 570 575
 60 Asn Ala Val His Ala Thr Val Lys Val Met Gly Asp Glu Phe Asn Trp
 580 585 590
 Ser Glu Lys Lys Arg Gln Trp Glu Leu Glu Lys Thr Val Asn Phe Ile
 595 600 605
 65 Gln Gly Arg Phe Gly Val
 610

ES 2 336 858 T3

(2) INFORMACIÓN PARA LA SEC. ID NO:14:

(i) CARACTERÍSTICAS DE LA SECUENCIA:

- (A) LONGITUD: 339 amino ácidos
- (B) TIPO: amino ácido
- (C) TIPO DE CADENA: desconocida
- (D) TIPOLOGÍA: desconocida

(ii) TIPO DE MOLÉCULA: proteína

(vi) FUENTE ORIGINAL:

- (A) ORGANISMO: GPSA

(xi) DESCRIPCIÓN DE LA SECUENCIA: SEC ID NO:14:

```

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

```

ES 2 336 858 T3

| | | | | | | | | | | | | | | | | |
|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| | Val | Gln | Ser | Ala | Gln | Glu | Lys | Ile | Gly | Gln | Val | Val | Glu | Gly | Tyr | Arg |
| | | | 275 | | | | | 280 | | | | | 285 | | | |
| 5 | Asn | Thr | Lys | Glu | Val | Arg | Glu | Leu | Ala | His | Arg | Phe | Gly | Val | Glu | Met |
| | | | | | | | 295 | | | | | 300 | | | | |
| 10 | Pro | Ile | Thr | Glu | Glu | Ile | Tyr | Gln | Val | Leu | Tyr | Cys | Gly | Lys | Asn | Ala |
| | 305 | | | | | 310 | | | | | 315 | | | | 320 | |
| 15 | Arg | Glu | Ala | Ala | Leu | Thr | Leu | Leu | Gly | Arg | Ala | Arg | Lys | Asp | Glu | Arg |
| | | | | 325 | | | | | | 330 | | | | 335 | | |
| | Ser | Ser | His | | | | | | | | | | | | | |

(2) INFORMACIÓN PARA LA SEC. ID NO:15:

20 (i) CARACTERÍSTICAS DE LA SECUENCIA:

(A) LONGITUD: 501 amino ácidos

(B) TIPO: amino ácido

25 (C) TIPO DE CADENA: desconocida

(D) TIPOLOGÍA: desconocida

(ii) TIPO DE MOLÉCULA: proteína

30 (vi) FUENTE ORIGINAL:

(A) ORGANISMO: GLPD

(xi) DESCRIPCIÓN DE LA SECUENCIA: SEC ID NO:15:

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

ES 2 336 858 T3

Lys Lys Tyr Ser Trp Gln Ala Arg Gly Leu Val Asn Ala Thr Gly Pro
 195 200 205
 Trp Val Lys Gln Phe Phe Asp Asp Gly Met His Leu Pro Ser Pro Tyr
 210 215 220
 Gly Ile Arg Leu Ile Lys Gly Ser His Ile Val Val Pro Arg Val His
 225 230 235 240
 Thr Gln Lys Gln Ala Tyr Ile Leu Gln Asn Glu Asp Lys Arg Ile Val
 245 250 255
 Phe Val Ile Pro Trp Met Asp Glu Phe Ser Ile Ile Gly Thr Thr Asp
 260 265 270
 Val Glu Tyr Lys Gly Asp Pro Lys Ala Val Lys Ile Glu Glu Ser Glu
 275 280 285
 Ile Asn Tyr Leu Leu Asn Val Tyr Asn Thr His Phe Lys Lys Gln Leu
 290 295 300
 Ser Arg Asp Asp Ile Val Trp Thr Tyr Ser Gly Val Arg Pro Leu Cys
 305 310 315 320
 Asp Asp Glu Ser Asp Ser Pro Gln Ala Ile Thr Arg Asp Tyr Thr Leu
 325 330 335
 Asp Ile His Asp Glu Asn Gly Lys Ala Pro Leu Leu Ser Val Phe Gly
 340 345 350
 Gly Lys Leu Thr Thr Tyr Arg Lys Leu Ala Glu His Ala Leu Glu Lys
 355 360 365
 Leu Thr Pro Tyr Tyr Gln Gly Ile Gly Pro Ala Trp Thr Lys Glu Ser
 370 375 380
 Val Leu Pro Gly Gly Ala Ile Glu Gly Asp Arg Asp Asp Tyr Ala Ala
 385 390 395 400
 Arg Leu Arg Arg Arg Tyr Pro Phe Leu Thr Glu Ser Leu Ala Arg His
 405 410 415
 Tyr Ala Arg Thr Tyr Gly Ser Asn Ser Glu Leu Leu Leu Gly Asn Ala
 420 425 430
 Gly Thr Val Ser Asp Leu Gly Glu Asp Phe Gly His Glu Phe Tyr Glu
 435 440 445
 Ala Glu Leu Lys Tyr Leu Val Asp His Glu Trp Val Arg Arg Ala Asp
 450 455 460
 Asp Ala Leu Trp Arg Arg Thr Lys Gln Gly Met Trp Leu Asn Ala Asp
 465 470 475 480
 Gln Gln Ser Arg Val Ser Gln Trp Leu Val Glu Tyr Thr Gln Gln Arg
 485 490 495
 Leu Ser Leu Ala Ser
 500

(2) INFORMACIÓN PARA LA SEC. ID NO:16:

(i) CARACTERÍSTICAS DE LA SECUENCIA:

- (A) LONGITUD: 542 amino ácidos
- (B) TIPO: amino ácido
- (C) TIPO DE CADENA: desconocida
- (D) TIPOLOGÍA: desconocida

ES 2 336 858 T3

(ii) TIPO DE MOLÉCULA: proteína

(vi) FUENTE ORIGINAL:

(A) ORGANISMO: GLPABC

5

(xi) DESCRIPCIÓN DE LA SECUENCIA: SEC ID NO:16:

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

65

ES 2 336 858 T3

Thr Arg Ile Leu Arg Ala Tyr Ser Gly Val Arg Pro Leu Val Ala Ser
 305 310 315 320
 5 Asp Asp Asp Pro Ser Gly Arg Asn Leu Ser Arg Gly Ile Val Leu Leu
 325 330 335
 Asp His Ala Glu Arg Asp Gly Leu Asp Gly Phe Ile Thr Ile Thr Gly
 340 345 350
 10 Gly Lys Leu Met Thr Tyr Arg Leu Met Ala Glu Trp Ala Thr Asp Ala
 355 360 365
 Val Cys Arg Lys Leu Gly Asn Thr Arg Pro Cys Thr Thr Ala Asp Leu
 370 375 380
 Ala Leu Pro Gly Ser Gln Glu Pro Ala Glu Val Thr Leu Arg Lys Val
 385 390 395
 20 Ile Ser Leu Pro Ala Pro Leu Arg Gly Ser Ala Val Tyr Arg His Gly
 405 410 415
 Asp Arg Thr Pro Ala Trp Leu Ser Glu Gly Arg Leu His Arg Ser Leu
 420 425 430
 Val Cys Glu Cys Glu Ala Val Thr Ala Gly Glu Val Gln Tyr Ala Val
 435 440 445
 30 Glu Asn Leu Asn Val Asn Ser Leu Leu Asp Leu Arg Arg Arg Thr Arg
 450 455 460
 Val Gly Met Gly Thr Cys Gln Gly Glu Leu Cys Ala Cys Arg Ala Ala
 465 470 475 480
 35 Gly Leu Leu Gln Arg Phe Asn Val Thr Thr Ser Ala Gln Ser Ile Glu
 485 490 495
 Gln Leu Ser Thr Phe Leu Asn Glu Arg Trp Lys Gly Val Gln Pro Ile
 500 505 510
 Ala Trp Gly Asp Ala Leu Arg Glu Ser Glu Phe Thr Arg Trp Val Tyr
 515 520 525
 45 Gln Gly Leu Cys Gly Leu Glu Lys Glu Gln Lys Asp Ala Leu
 530 535 540

50 (2) INFORMACIÓN PARA LA SEC. ID NO:17:

(i) CARACTERÍSTICAS DE LA SECUENCIA:

- (A) LONGITUD: 250 amino ácidos
- (B) TIPO: amino ácido
- (C) TIPO DE CADENA: desconocida
- (D) TIPOLOGÍA: desconocida

60 (ii) TIPO DE MOLÉCULA: proteína

(vi) FUENTE ORIGINAL:

- (A) ORGANISMO: GPP2

65 (xi) DESCRIPCIÓN DE LA SECUENCIA: SEC ID NO:17:

ES 2 336 858 T3

1 Met .Gly Leu Thr Thr Lys Pro Leu Ser Leu Lys Val Asn Ala Ala Leu
 5 Phe Asp Val Asp Gly Thr Ile Ile Ile Ser Gln Pro Ala Ile Ala Ala
 10 Phe Trp Arg Asp Phe Gly Lys Asp Lys Pro Tyr Phe Asp Ala Glu His
 15 Val Ile Gln Val Ser His Gly Trp Arg Thr Phe Asp Ala Ile Ala Lys
 20 Phe Ala Pro Asp Phe Ala Asn Glu Glu Tyr Val Asn Lys Leu Glu Ala
 25 Glu Ile Pro Val Lys Tyr Gly Glu Lys Ser Ile Glu Val Pro Gly Ala
 30 Val Lys Leu Cys Asn Ala Leu Asn Ala Leu Pro Lys Glu Lys Trp Ala
 35 Val Ala Thr Ser Gly Thr Arg Asp Met Ala Gln Lys Trp Phe Glu His
 40 Leu Gly Ile Arg Arg Pro Lys Tyr Phe Ile Thr Ala Asn Asp Val Lys
 45 Gln Gly Lys Pro His Pro Glu Pro Tyr Leu Lys Gly Arg Asn Gly Leu
 50 Gly Tyr Pro Ile Asn Glu Gln Asp Pro Ser Lys Ser Lys Val Val Val
 55 Phe Glu Asp Ala Pro Ala Gly Ile Ala Ala Gly Lys Ala Ala Gly Cys
 60 Lys Ile Ile Gly Ile Ala Thr Thr Phe Asp Leu Asp Phe Leu Lys Glu
 65 Lys Gly Cys Asp Ile Ile Val Lys Asn His Glu Ser Ile Arg Val Gly
 70 Gly Tyr Asn Ala Glu Thr Asp Glu Val Glu Phe Ile Phe Asp Asp Tyr
 75 Leu Tyr Ala Lys Asp Asp Leu Leu Lys Trp
 80 245 250

(2) INFORMACIÓN PARA LA SEC. ID NO:18:

- 55 (i) CARACTERÍSTICAS DE LA SECUENCIA:
- (A) LONGITUD: 709 amino ácidos
 - (B) TIPO: amino ácido
 - 60 (C) TIPO DE CADENA: desconocida
 - (D) TIPOLOGÍA: desconocida
- (ii) TIPO DE MOLÉCULA: proteína
- 65 (vi) FUENTE ORIGINAL:
- (A) ORGANISMO: GUT1

ES 2 336 858 T3

(xi) DESCRIPCIÓN DE LA SECUENCIA: SEC ID NO:18:

| | | | | | | | | | | | | | | | | |
|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 5 | Met | Phe | Pro | Ser | Leu | Phe | Arg | Leu | Val | Val | Phe | Ser | Lys | Arg | Tyr | Ile |
| | 1 | | | | 5 | | | | | 10 | | | | | 15 | |
| | Phe | Arg | Ser | Ser | Gln | Arg | Leu | Tyr | Thr | Ser | Leu | Lys | Gln | Glu | Gln | Ser |
| | | | | 20 | | | | | 25 | | | | | 30 | | |
| | Arg | Met | Ser | Lys | Ile | Met | Glu | Asp | Leu | Arg | Ser | Asp | Tyr | Val | Pro | Leu |
| 10 | | | 35 | | | | | 40 | | | | | 45 | | | |
| | Ile | Ala | Ser | Ile | Asp | Val | Gly | Thr | Thr | Ser | Ser | Arg | Cys | Ile | Leu | Phe |
| | | 50 | | | | | 55 | | | | | 60 | | | | |
| | Asn | Arg | Trp | Gly | Gln | Asp | Val | Ser | Lys | His | Gln | Ile | Glu | Tyr | Ser | Thr |
| 15 | 65 | | | | | 70 | | | | | 75 | | | | | 80 |
| | Ser | Ala | Ser | Lys | Gly | Lys | Ile | Gly | Val | Ser | Gly | Leu | Arg | Arg | Pro | Ser |
| | | | | | 85 | | | | | 90 | | | | | 95 | |
| | Thr | Ala | Pro | Ala | Arg | Glu | Thr | Pro | Asn | Ala | Gly | Asp | Ile | Lys | Thr | Ser |
| 20 | | | | 100 | | | | | 105 | | | | | 110 | | |
| | Gly | Lys | Pro | Ile | Phe | Ser | Ala | Glu | Gly | Tyr | Ala | Ile | Gln | Glu | Thr | Lys |
| | | | 115 | | | | | 120 | | | | | 125 | | | |
| | Phe | Leu | Lys | Ile | Glu | Glu | Leu | Asp | Leu | Asp | Phe | His | Asn | Glu | Pro | Thr |
| 25 | | 130 | | | | | | 135 | | | | 140 | | | | |
| | Leu | Lys | Phe | Pro | Lys | Pro | Gly | Trp | Val | Glu | Cys | His | Pro | Gln | Lys | Leu |
| | | 145 | | | | 150 | | | | | 155 | | | | | 160 |
| | Leu | Val | Asn | Val | Val | Gln | Cys | Leu | Ala | Ser | Ser | Leu | Leu | Ser | Leu | Gln |
| 30 | | | | | 165 | | | | | 170 | | | | | 175 | |
| | Thr | Ile | Asn | Ser | Glu | Arg | Val | Ala | Asn | Gly | Leu | Pro | Pro | Tyr | Lys | Val |
| | | | | 180 | | | | | 185 | | | | | 190 | | |
| | Ile | Cys | Met | Gly | Ile | Ala | Asn | Met | Arg | Glu | Thr | Thr | Ile | Leu | Trp | Ser |
| 35 | | | 195 | | | | | 200 | | | | | 205 | | | |
| | Arg | Arg | Thr | Gly | Lys | Pro | Ile | Val | Asn | Tyr | Gly | Ile | Val | Trp | Asn | Asp |
| | | 210 | | | | | 215 | | | | | 220 | | | | |
| | Thr | Arg | Thr | Ile | Lys | Ile | Val | Arg | Asp | Lys | Trp | Gln | Asn | Thr | Ser | Val |
| 40 | | | | | | 230 | | | | | 235 | | | | | 240 |
| | Asp | Arg | Gln | Leu | Gln | Leu | Arg | Gln | Lys | Thr | Gly | Leu | Pro | Leu | Leu | Ser |
| | | | | | 245 | | | | | 250 | | | | | 255 | |
| | Thr | Tyr | Phe | Ser | Cys | Ser | Lys | Leu | Arg | Trp | Phe | Leu | Asp | Asn | Glu | Pro |
| | | | | 260 | | | | | 265 | | | | | 270 | | |
| | Leu | Cys | Thr | Lys | Ala | Tyr | Glu | Glu | Asn | Asp | Leu | Met | Phe | Gly | Thr | Val |
| 50 | | | 275 | | | | | 280 | | | | | 285 | | | |
| | Asp | Thr | Trp | Leu | Ile | Tyr | Gln | Leu | Thr | Lys | Gln | Lys | Ala | Phe | Val | Ser |
| | | 290 | | | | | 295 | | | | | 300 | | | | |
| | Asp | Val | Thr | Asn | Ala | Ser | Arg | Thr | Gly | Phe | Met | Asn | Leu | Ser | Thr | Leu |
| 55 | | | | | 310 | | | | | | 315 | | | | | 320 |
| | Lys | Tyr | Asp | Asn | Glu | Leu | Leu | Glu | Phe | Trp | Gly | Ile | Asp | Lys | Asn | Leu |
| | | | | | 325 | | | | | 330 | | | | | 335 | |
| | Ile | His | Met | Pro | Glu | Ile | Val | Ser | Ser | Ser | Gln | Tyr | Tyr | Gly | Asp | Phe |
| 60 | | | | 340 | | | | | | 345 | | | | 350 | | |
| | Gly | Ile | Pro | Asp | Trp | Ile | Met | Glu | Lys | Leu | His | Asp | Ser | Pro | Lys | Thr |
| | | | 355 | | | | | 360 | | | | | 365 | | | |
| | Val | Leu | Arg | Asp | Leu | Val | Lys | Arg | Asn | Leu | Pro | Ile | Gln | Gly | Cys | Leu |
| 65 | | | | | | | 375 | | | | | | 380 | | | |

ES 2 336 858 T3

Ala Ala Lys Cys Thr Tyr Gly Thr Gly Cys Phe Leu Leu Tyr Asn Thr
 405 410 415

5 Gly Thr Lys Lys Leu Ile Ser Gln His Gly Ala Leu Thr Thr Leu Ala
 420 425 430

Phe Trp Phe Pro His Leu Gln Glu Tyr Gly Gly Gln Lys Pro Glu Leu
 435 440 445

10 Ser Lys Pro His Phe Ala Leu Glu Gly Ser Val Ala Val Ala Gly Ala
 450 455 460

15 Val Val Gln Trp Leu Arg Asp Asn Leu Arg Leu Ile Asp Lys Ser Glu
 465 470 475 480

Asp Val Gly Pro Ile Ala Ser Thr Val Pro Asp Ser Gly Gly Val Val
 485 490 495

20 Phe Val Pro Ala Phe Ser Gly Leu Phe Ala Pro Tyr Trp Asp Pro Asp
 500 505 510

25 Ala Arg Ala Thr Ile Met Gly Met Ser Gln Phe Thr Thr Ala Ser His
 515 520 525

Ile Ala Arg Ala Ala Val Glu Gly Val Cys Phe Gln Ala Arg Ala Ile
 530 535 540

30 Leu Lys Ala Met Ser Ser Asp Ala Phe Gly Glu Gly Ser Lys Asp Arg
 545 550 555 560

Asp Phe Leu Glu Glu Ile Ser Asp Val Thr Tyr Glu Lys Ser Pro Leu
 565 570 575

35 Ser Val Leu Ala Val Asp Gly Gly Met Ser Arg Ser Asn Glu Val Met
 580 585 590

40 Gln Ile Gln Ala Asp Ile Leu Gly Pro Cys Val Lys Val Arg Arg Ser
 595 600 605

Pro Thr Ala Glu Cys Thr Ala Leu Gly Ala Ala Ile Ala Ala Asn Met
 610 615 620

45 Ala Phe Lys Asp Val Asn Glu Arg Pro Leu Trp Lys Asp Leu His Asp
 625 630 635 640

Val Lys Lys Trp Val Phe Tyr Asn Gly Met Glu Lys Asn Glu Gln Ile
 645 650 655

50 Ser Pro Glu Ala His Pro Asn Leu Lys Ile Phe Arg Ser Glu Ser Asp
 660 665 670

55 Asp Ala Glu Arg Arg Lys His Trp Lys Tyr Trp Glu Val Ala Val Glu
 675 680 685

Arg Ser Lys Gly Trp Leu Lys Asp Ile Glu Gly Glu His Glu Gln Val
 690 695 700

60 Leu Glu Asn Phe Gln
 705

(2) INFORMACIÓN PARA LA SEC. ID NO:19:

65 (i) CARACTERÍSTICAS DE LA SECUENCIA:

(A) LONGITUD: 12145 pares de bases

ES 2 336 858 T3

- (B) TIPO: ácido nucleico
- (C) TIPO DE CADENA: simple
- (D) TIPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: DNA (genómico)

(vi) FUENTE ORIGINAL:

(A) ORGANISMO: PHK28-26

(xi) DESCRIPCIÓN DE LA SECUENCIA: SEC ID NO:19:

| | | |
|----|---|------|
| 5 | GTCGACCACC ACGGTGGTGA CTTAATGCC GCTCTCATGC AGCAGCTCGG TGGCGGTCTC | 60 |
| 15 | AAAATTCAGG ATGTCGCCGG TATAGTTTTT GATAATCAGC AAGACGCCTT CGCCGCCGTC | 120 |
| | AATTTGCATC GCGCATTCAA ACATTTTGTG CGGCGTGGC GAGGTGAATA TTTCCCCGG | 180 |
| 20 | ACAGGCGCCG GAGAGCATGC CCTGGCCGAT ATAGCCGAG TGCATCGGTT CATGTCCGCT | 240 |
| | GCCGCCGCCG GAGAGCAGGG CCACCTTGCC AGCCACCGGC GCGTCCGTGC GGGTCACATA | 300 |
| | CAGCGGGTCC TGATGCAGGG TCAGCTGCGG ATGGGCTTTA GCCAGCCCCT GTAATTGTTC | 360 |
| 25 | ATTCAGTACA TCTTCAACAC GGTAATCAG CTTTTTCATT ATTCAGTGCT CCGTTGGAGA | 420 |
| | AGTTTCGATG CCGCCTCTCT GCTGGCGGAG GCGGTCATCG CGTAGGGGTA TCGTCTGACG | 480 |
| | GTGGAGCGTG CCTGGCGATA TGATGATTCT GGCTGAGCGG ACGAAAAAAA GAATGCCCCG | 540 |
| 30 | ACGATCGGGT TTCATTACGA AACATTGCTT CCTGATTTTG TTTCTTTATG GAACGTTTTT | 600 |
| | GCTGAGGATA TGGTGAAAAT GCGAGCTGGC GCGCTTTTTT TCTTCTGCCA TAAGCGGCGG | 660 |
| 35 | TCAGGATAGC CGGCGAAGCG GGTGGGAAAA AATTTTTTGC TGATTTTCTG CCGACTGCGG | 720 |
| | GAGAAAAGGC GGTCAAACAC GGAGGATTGT AAGGGCATTG TGCGGCAAAG GAGCGGATCG | 780 |
| | GGATCGCAAT CCTGACAGAG ACTAGGGTTT TTTGTTCCAA TATGGAACGT AAAAAATTAA | 840 |
| 40 | CCTGTGTTTC ATATCAGAAC AAAAAGGCGA AAGATTTTTT TGTTCCCTGC CGGCCCTACA | 900 |
| | GTGATCGCAC TGCTCCGGTA CGCTCCGTTT AGGCCGCGCT TCACTGGCCG GCGCGGATAA | 960 |
| 45 | CGCCAGGGCT CATCATGTCT ACATGCGCAC TTATTTGAGG GTGAAAGGAA TGCTAAAAGT | 1020 |
| | TATTCAATCT CCAGCCAAAT ATCTTCAGGG TCCTGATGCT GCTGTTCTGT TCGGTCAATA | 1080 |
| | TGCCAAAAAC CTGGCGGAGA GCTTCTTCGT CATCGCTGAC GATTCGTAA TGAAGCTGGC | 1140 |
| 50 | GGGAGAGAAA GTGGTGAATG GCCTGCAGAG CCACGATATT CGCTGCCATG CGGAACGGTT | 1200 |
| | TAACGGCGAA TGCAGCCATG CGGAAATCAA CCGTCTGATG GCGATTTTGC AAAAACAGGG | 1260 |
| | CTGCCGCGGC GTGGTCGGGA TCGGCGGTGG TAAAACCCTC GATACCGCGA AGGCGATCGG | 1320 |
| 55 | TTACTACCAG AAGTGCCGG TGGTGGTGAT CCCGACCATC GCCTCGACCG ATGCGCCAAC | 1380 |
| | CAGCGCGCTG TCGGTGATCT ACACCGAAGC GGGCGAGTTT GAAGAGTATC TGATCTATCC | 1440 |
| 60 | GAAAAACCCG GATATGGTGG TGATGGACAC GGCATTATC GCCAAAGCGC CGGTACGCCT | 1500 |
| | GCTGGTCTCC GGCATGGGCG ATGCGCTCTC CACCTGGTTC GAGGCCAAAG CTTGCTACGA | 1560 |
| | TGCGCGCGCC ACCAGCATGG CCGGAGGACA GTCCACCGAG GCGGCGCTGA GCCTCGCCCG | 1620 |
| 65 | CCTGTGCTAT GATACGCTGC TGGCGGAGGG CGAAAAGGCC CGTCTGGCGG CGCAGGCCGG | 1680 |

ES 2 336 858 T3

5 GGTAGTGACC GAAGCGCTGG AGCGCATCAT CGAGGGGAAC ACTTACCTCA GCGGCATTGG 1740
 CTTTCAAAGC AGTGGCCTGG CCGCTGCCCA TGCAATCCAC AACGGTTTCA CCATTCTTGA 1800
 AGAGTGCCAT CACCTGTATC ACGGTGAGAA AGTGGCCTTC GGTACCCTGG CGCAGCTGGT 1860
 GCTGCAGAAC AGCCCGATGG ACGAGATTGA AACGGTGCAG GGCTTCTGCC AGCGCGTCGG 1920
 10 CCTGCCGGTG ACGCTCGCGG AGATGGGCGT CAAAGAGGGG ATCGACGAGA AAATCGCCGC 1980
 GGTGGCGAAA GCTACCTGGG CGGAAGGGGA AACCATCCAT AATATGCCGT TTGCGGTGAC 2040
 CCGGAGAGC GTCCATGCCG CTATCCTCAC CGCCGATCTG TTAGGCCAGC AGTGGCTGGC 2100
 15 GCGTTAATTC GCGGTGGCTA AACCCTGGC CCAGGTCAGC GGTTTTTCTT TCTCCCTCC 2160
 GGCAGTCGCT GCCGGAGGGG TTCTCTATGG TACAACGCGG AAAAGGATAT GACTGTTCAG 2220
 ACTCAGGATA CCGGAAGGC GGTCTCTTCC GTCATTGCCC AGTCATGGCA CCGCTGCAGC 2280
 20 AAGTTTATGC AGCGGAAAC CTGGCAAACG CCGCACCAGG CCCAGGGCCT GACCTTCGAC 2340
 TCCATCTGTC GCGGTAAAAC CGCGCTGCTC ACCATCGGCC AGGCGGCGCT GGAAGACGCC 2400
 TGGGAGTTTA TGGACGGCGG CCCCTGCGCG CTGTTTATTC TTGATGAGTC CGCCTGCATC 2460
 25 CTGAGCCGTT GCGGCGAGCC GCAAACCCTG GCCCAGCTGG CTGCCCTGGG ATTTGCGGAC 2520
 GGCAGCTATT GTGCGGAGAG CATTATCGGC ACCTGCGCGC TGTCGCTGGC CGCATGCAG 2580
 30 GSCCAGCCGA TCAACACCGC CGGCGATCGG CATTTAAGC AGGCGCTACA GCCATGGAGT 2640
 TTTTGCTCGA CGCCGGTGTG TGATAACCAC GGGCGGCTGT TCGGCTCTAT CTCGCTTTCG 2700
 TGTCGTGTCG AGCACCAGTC CAGCGCCGAC CTCTCCCTGA CGCTGGCCAT CGCCCGGAG 2760
 35 GTGGGTAACT CCCTGCTTAC CGACAGCCTG CTGGCGGAAT CCAACCGTCA CCTCAATCAG 2820
 ATGTACGGCC TGCTGGAGAG CATGGACGAT GGGGTGATGG CGTGGAAACGA ACAGGGCGTG 2880
 40 CTGCAGTTTC TCAATGTTCA GCGGCGGAGA CTGCTGCATC TTGATGCTCA GGCCAGCCAG 2940
 GGGAAAATA TCGCCGATCT GGTGACCCTC CCGGCGCTGC TCGCCCGCGC CATCAAACAC 3000
 GCCCGCGGCC TGAATCACGT CGAAGTCACC TTTGAAAGTC AGCATCAGTT TGTCGATGCG 3060
 45 GTGATCACCT TAAAACCGAT TGTCGAGGCG CAAGGCAACA GTTTTATTCT GCTGCTGCAT 3120
 CCGGTGGAGC AGATGCGGCA GCTGATGACC AGCCAGCTCG GTAAAGTCAG CCACACCTTT 3180
 GAGCAGATGT CTGCCGACGA TCCGAAACC CGACGCCTGA TCCACTTTGG CCGCCAGGCG 3240
 50 GCGCGCGGCG GCTTCCCGGT GCTACTGTGC GCGAAGAGG GGGTCGGGAA AGAGCTGCTG 3300
 AGCCAGGCTA TTCACAATGA AAGCGAACGG GCGGGCGGCC CCTACATCTC CGTCAACTGC 3360
 CAGCTATATG CCGACAGCGT GCTGGGCCAG GACTTTATGG GCAGCGCCCC TACCGACGAT 3420
 55 GAAAATGGTC GCCTGAGCCG CCTTGAGCTG GCCAACGGCG GCACCCTGTT TCTGGAAAAG 3480
 ATCGAGTATC TGGCGCCGGA GCTGCAGTCG GCTCTGCTGC AGGTGATTAA GCAGGGCGTG 3540
 60 CTCACCGGCC TCGACGCCCG GCGCCTGATC CCGGTGGATG TGAAGGTGAT TGCCACCACC 3600
 ACCGTGATC TGGCCAATCT GGTGGAACAG AACCGCTTTA GCCGCCAGCT GTACTATGCG 3660

65

ES 2 336 858 T3

| | | |
|----|---|------|
| | CTGCACTCCT TTGAGATCGT CATCCCGCCG CTGCGCGCCC GACGCAACAG TATTCCGTCG | 3720 |
| | CTGGTGCATA ACCGGTTGAA GAGCCTGGAG AAGCGTTTCT CTTGCGGACT GAAAGTGGAC | 3780 |
| 5 | GATGACGCGC TGGCACAGCT GGTGGCCTAC TCGTGGCCGG GGAATGATTT TGAGCTCAAC | 3840 |
| | AGCGTCATTG AGAATATCGC CATCAGCAGC GACAACGGCC ACATTCGCCT GAGTAATCTG | 3900 |
| 10 | CCGGAATATC TCTTTTCCGA GCGGCCGGGC GGGGATAGCG CGTCATCGCT GCTGCCGGCC | 3960 |
| | AGCCTGACTT TTAGCGCCAT CGAAAAGGAA GCTATTATTC ACGCCGCCCG GGTGACCAGC | 4020 |
| | GGGCGGGTGC AGGAGATGTC GCAGCTGCTC AATATCGGCC GCACCACCCT GTGGCGCAA | 4080 |
| 15 | ATGAAGCAGT ACGATATTGA CGCCAGCCAG TTCAAGCGCA AGCATCAGGC CTAGTCTCTT | 4140 |
| | CGATTCGCGC CATGGAGAAC AGGGCATCCG ACAGGCGATT GCTGTAGCGT TTGAGCGCGT | 4200 |
| | CGCGCAGCGG ATGCGCGCGG TCCATGGCCG TCAGCAGGCG TTCGAGCCGA CGGGACTGGG | 4260 |
| 20 | TGCGCGCCAC GTGCAGCTGG GCAGAGGCGA GATTCCCTCC CGGGATCACG AACTGTTTTA | 4320 |
| | ACGGGCCGCT CTCGGCCATA TTGCGGTGCA TAAGCCGCTC CAGGGCGGTG ATCTCCTCTT | 4380 |
| | CGCCGATCST CTGGCTCAGG CGGGTGAGGC CCCGCGCATC GCTGGCCAGT TCAGCCCCCA | 4440 |
| 25 | GCACGAACAG CGTCTGCTGA ATATGGTGCA GGCTTTCCCG CAGCCCGGCG TCGCGGGTCG | 4500 |
| | TGGCGTAGCA GACGCCAGC TGGGATATCA GTTCATCGAC GGTGCCGTAG GCCTCGACGC | 4560 |
| 30 | GAATATGSTC TTTCTCGATG CGGCTGCCGC CGTACAGGGC GGTGGTGCCT TTATCCCCGG | 4620 |
| | TGCGGGTATA GATACGATAC ATTCAGTTTC TCTCACTAA CGGCAGGACT TTAACCAGCT | 4680 |
| | GCCCCGCGTT GCGCCGAGC GTACGCAGTT GATCGTCGCT ATCGGTGACG TGTCGGGTAG | 4740 |
| 35 | CCAGCGGCGC GTCGCGCCGC AGCTGGGCAT GAGTGAGGGC TATCTCGCCG GACGCGCTGA | 4800 |
| | GCCCGATACC CACCCGAGG GCGGAGCTTC TGCCCGCCAG GCGCCCGAGC GCAGCGGCGT | 4860 |
| | CACCGCCTCC GTCATAGGTT ATGGTCTGGC AGGGGACCCC CTGCTCCTCC AGCCCCCAGC | 4920 |
| 40 | ACAGCTCATT GATGGCGCCG GCATGGTGCC CGCGCGGATC GTAAAACAGG CGTACGCCTG | 4980 |
| | GCGGTGAAAG CGACATGACG GTCCCCTCGT TAACACTCAG AATGCCTGGC GGAAAATCGC | 5040 |
| 45 | GGCAATCTCC TGCTCGTTGC CTTTACGCGG GTTCGAGAAC GCATTGCCGT CTTTLAGAGC | 5100 |
| | CATCTCCGCC ATGTAGGGGA AGTCGGCCTC TTTTACCCCC AGATCGCGCA GATGCTGCGG | 5160 |
| | AATACCGATA TCCATCGACA GACCGGTGAT AGCGGCGATG GCTTTTCCG CCGCGTCGAG | 5220 |
| 50 | AGTGGACAGT CCGGTGATAT TTTCGCCAT CAGTTCAGCG ATATCGGCGA ATTTCTCCGG | 5280 |
| | GTTGGCGATC AGGTGTAGC GCGCCACATG CGGCAGCAGG ACAGCGTTGG CCACGCCGTG | 5340 |
| | CGGCATGTCT TACAGGCCGC CCAGCTGGTG CGCCATGGCG TGCACGTAGC CGAGGTTGGC | 5400 |
| 55 | GTTATTGAAA GCCATCCCGG CCAGCAGAGA AGCATAGGCC ATGTTTTCCG GCGCCTGCAG | 5460 |
| | ATTGCTGCCG AGGGCCACGG CCTGGCGCAG GTTGCGGGCG ATGAGGCGGA TCGCTGCAT | 5520 |
| | GGCGCGGCGG TCCGTCACCG GGTAGCGTC TTTGGAGATA TAGGCCTCTA CGGCGTGGGT | 5580 |
| 60 | CAGGGCATCC ATCCCGTCCG CCGCGGTCAG GCGGCGCGGT TTACCGATCA TCAGCAGTGG | 5640 |

65

ES 2 336 858 T3

5
 10
 15
 20
 25
 30
 35
 40
 45
 50
 55
 60
 65

ATCGTTGATA GAGACCGACG GCAGTTTGGC CCAGCTGACG ATCA[^]CAA[^]ACT TCACTTTGGT 5700
 TTCGGTGTG GTCAGGACGC AGTGGCGGGT GACCTCGCTG GCGGTGCCGG CGGTGGTATT 5760
 GACCGCGACG ATAGGCGGCA GCGGGTTGGT CAGGGTCTCG ATTCCGGCAT ACTGGTACAG 5820
 ATCGCCCTCA TGGGTGGCGG CGATGCCGAT GCCTTTGCCG CAATCGTGCG GGCTGCCGCC 5880
 GCCCACGGTG ACGATGATGT CGCACTGTTC GCGGCGAAAC ACGGCGAGGC CGTCGCGCAC 5940
 GTTGGTGTCT TTCGGGTTCC GCTCGACGCC GTCAAAGATC GCCACCTCGA TCCCGGCCCTC 6000
 CCGCAGATAA TGCAGGGTTT TGCCACCGC GCCATCTTTA ATTGCCCGCA GGCTTTGTC 6060
 GGTGACCAGC AGGGCTTTTT TCCCCCCAG CAGCTGGCAG CGTTCGCCGA CTACGGAAAT 6120
 GCGGTGGGG CCAAAAAAGT TAACGTTGG CACCAGATAA TCAAACATAC GATAGCTCAT 6180
 AATATACCTT CTCGCTTCAG GTTATAATGC GGAAAAACAA TCCAGGGCGC ACTGGGCTAA 6240
 TAATTGATCC TGCTCGACCG TACCGCCGCT AACGCCGACG GCGCCAATTA CCTGCTCATT 6300
 AAAAATAACT GGCAGGCCGC CGCCAAAAAT AATAATTCGC TGTTGGTTGG TTAGCTGCAG 6360
 ACCGTACAGA GATTGTCCTG GCTGGACCGC TGACGTAATT TCATGGGTAC CTTGCTTCAG 6420
 GCTGCAGGCG CTCCAGGCTT TATTCAGGGA AATATCGCAG CTGGAGACGA AGGCCTCGTC 6480
 CATCCGCTGG ATAAGCAGCG TGTTGCCCTC GCGGTCAACT ACGGAAAACA CCACCGCCAC 6540
 GTTGATCTCA GTGGCTTTTT TTTCCACCGC CGCCGCCATT TGCTGGGCGG CGGCCAGGGT 6600
 GATTGTCTGA ACTTGTGGC TCTGTTCAT CATTCTCTCC CGCACCAGGA TAACGCTGGC 6660
 GCGAATAGTC AGTAGGGGGC GATAGTAAAA AACTATTACC ATTCGGTTGG CTTGCTTTAT 6720
 TTTTGTGAGC GTTATTTTET CGCCCGCCAT GATTTAGTCA ATAGGGTTAA AATAGCGTCG 6780
 GAAAAACGTA ATTAAGGGCG TTTTTATTA ATTGATTTAT ATCATTGCGG GCGATCACAT 6840
 TTTTTATTTT TGCCGCCGGA GTAAAGTTTC ATAGTGAAAC TGTCGGTAGA TTTCGTGTGC 6900
 CAAATGAAA CGAAATTA[^]A TTTATTTTTT TCACCACTGG CTCATTTAAA GTTCCGCTAT 6960
 TGCCGTAAT GGCCGGGCGG CAACGACGCT GGCCGGCGT ATTCGCTACC GTCTGCGGAT 7020
 TTCACCTTTT GAGCCGATCA ACAATGAAAA GATCAAACG ATTTGCAGTA CTGGCCAGC 7080
 GCCCGTCAA TCAGGACGGG CTGATTGGCG AGTGGCCTGA AGAGGGGCTG ATCGCCATGG 7140
 ACAGCCCTT TGACCCGGTC TCTTCAGTAA AAGTGACAA CGGTCTGATC GTCGAACTGG 7200
 ACGGCAAACG CCGGGACCAG TTTGACATGA TCGACCGATT TATCGCCGAT TACGCGATCA 7260
 ACGTTGAGCG CACAGAGCAG GCAATGCGCC TGGAGCGGT GGAAATAGCC CGTATGCTGG 7320
 TGGATATTCA CGTCAGCCGG GAGGAGATCA TTGCCATCAC TACCGCCATC ACGCCGGCCA 7380
 AAGCGTCTGA GGTGATGGCG CAGATGAACG TGGTGGAGAT GATGATGGCG CTGCAGAAGA 7440
 TGCGTGCCCG CCGGACCC[^]C TCCAACCAGT GCCACGTCAC CAATCTCAA[^]A GATAATCCGG 7500
 TGCAGATTGC CGCTGACGCC GCCGAGGCCG GGATCCGCGG CTTCTCAGAA CAGGAGACCA 7560
 CGGTCCGTAT CGCGCGCTAC GCGCCGTTA ACGCCCTGGC GCTGTTGGTC GGTTCGCAGT 7620

ES 2 336 858 T3

GCGGCCGCC CGGCGTGTG ACGCAGTGCT CGGTGGAAGA GGCCACCGAG CTGGAGCTGG 7680
 GCATGCGTGG CTTAACCAGC TACGCCGAGA CGGTGTCGGT CTACGGCACC GAAGCGGTAT 7740
 5 TTACCGACGG CGATGATACG CCGTGGTCAA AGGCGTTCCT CGCCTCGGCC TACGCCTCCC 7800
 GCGGGTTGAA AATGCGCTAC ACCTCCGGCA CCGGATCCGA AGCGCTGATG GGCTATTCCG 7860
 10 AGAGCAAGTC GATGCTCTAC CTCGAATCGC GCTGCATCTT CATTACTAAA GGCGCCGGGG 7920
 TTCAGGGACT GCAAAACGGC GCGGTGAGCT GTATCGGCAT GACCGGCGCT GTGCCGTCCG 7980
 GCATTCCGGC GGTGCTGGCG GAAAACCTGA TCGCCTCTAT GCTCGACCTC GAAGTGGCGT 8040
 15 CCGCCAACGA CCAGACTTTC TCCCCTCGG ATATTCGCCG CACCGCGCGC ACCCTGATGC 8100
 AGATGCTGCC GGGCACCGAC TTTATTTTCT CCGGCTACAG CGCGGTGCCG AACTACGACA 8160
 ACATGTTCCG CGGCTCGAAC TTCGATGCGG AAGATTTTGA TGATTACAAC ATCCTGCAGC 8220
 20 GTGACCTGAT GGTTGACGGC GGCCTGCGTC CGGTGACCGA GGCGGAAACC ATTGCCATTC 8280
 GCCAGAAAGC GGC CGGGCGG ATCCAGGCGG TTTTCCGCGA GCTGGGGCTG CCGCCAATCG 8340
 25 CCGACGAGGA GGTGGAGGCC GCCACCTACG CGCACGGCAG CAACGAGATG CCGCCGCGTA 8400
 ACGTGGTGGG GGATCTGAGT GCGGTGGAAG AGATGATGAA GCGCAACATC ACCGGCCTCG 8460
 ATATTGTCGG CGCGCTGAGC CGCAGCGGCT TTGAGGATAT CGCCAGCAAT ATTCTCAATA 8520
 30 TGCTGCGCCA GCGGGTCACC GCGGATTACC TGCAGACCTC GGCCATTCTC GATCGGCAGT 8580
 TCGAGGTGGT GAGTGCGGTC AACGACATCA ATGACTATCA GGGGCCGGGC ACCGGCTATC 8640
 GCATCTCTGC CGAACGCTGG GCGGAGATCA AAAATATTCC GGGCGTGGTT CAGCCCGACA 8700
 35 CCATTGAATA AGGCGGTATT CCTGTGCAAC AGACAACCCA AATTCAGCCC TCTTTTACCC 8760
 TGAAAACCCG CGAGGGCGGG GTAGCTTCTG CCGATGAACG CGCCGATGAA GTGGTGATCG 8820
 40 GCGTCGGCCC TGCCTTCGAT AAACACCAGC ATCACACTCT GATCGATATG CCCCATGGCG 8880
 CGATCCTCAA AGAGCTGATT GCCGGGGTGG AAGAAGAGGG GCTTCACGCC CGGGTGGTGC 8940
 GCATTCTGCG CACGTCCGAC GTCTCCTTTA TGGCCTGGGA TGCGGCCAAC CTGAGCGGCT 9000
 45 CGGGGATCCG CATCGGTATC CAGTCGAAGG GGACCACGGT CATCCATCAG CGCGATCTGC 9060
 TGCCGCTCAG CAACCTGGAG CTGTTCTCCC AGGCGCCGCT GCTGACGCTG GAGACCTACC 9120
 GGCAGATTGG CAAAAACGCT GCGCGCTATG CGCGCAAAGA GTCACCTTCG CCGGTGCCGG 9180
 50 TGGTGAACGA TCAGATGGTG CGGCCGAAAT TTATGGCCAA AGCCGCGCTA TTTCATATCA 9240
 AAGAGACCAA ACATGTGGTG CAGGACGCCG AGCCCGTCAC CCTGCACATC GACTTAGTAA 9300
 55 GGGAGTGACC ATGAGCGAGA AAACCATGCG CGTGCAGGAT TATCCGTTAG CCACCCGCTG 9360
 CCCGGAGCAT ATCCTGACGC CTACCGGCAA ACCATTGACC GATATTACCC TCGAGAAGGT 9420
 GCTCTCTGGC GAGGTGGGCC CGCAGGATGT GCGGATCTCC CGCCAGACCC TTGAGTACCA 9480
 60 GGCGCAGATT GCCGAGCAGA TGCAGGCCA TGCGGTGGCG CGCAATTTCC GCCGCGCGGC 9540
 GGAGCTTATC GCCATTCTG ACGAGCGCAT TCTGGCTATC TATAACGCGC TCGCCCGTT 9600
 65

ES 2 336 858 T3

CCGCTCCTCG CAGGCGGAGC TGCTGGCGAT CGCCGACGAG CTGGAGCACA CCTGGCATGC 9660
 5 GACAGTGAAT GCCGCCTTG TCCGGGAGTC GCGGGAAGTG TATCAGCAGC GGCATAAGCT 9720
 GCGTAAAGGA AGCTAAGCGG AGGTCAGCAT GCCGTTAATA GCCGGGATTG ATATCGGCAA 9780
 CGCCACCACC GAGGTGGCGC TGGCGTCCGA CTACCCGAG GCGAGGGCGT TTGTTGCCAG 9840
 10 CGGGATCGTC GCGACGACGG GCATGAAAGG GACGCGGGAC AATATCGCCG GGACCCTCGC 9900
 CGCGCTGGAG CAGGCCCTGG CGAAAACACC GTGGTCGATG AGCGATGTCT CTCGCATCTA 9960
 TTTTAACGAA GCCGCGCCGG TGATTGGCGA TGTGGCGATG GAGACCATCA CCGAGACCAT 10020
 15 TATCACCGAA TCGACCATGA TCGGTCATAA CCCGCAGACG CCGGGCGGGG TGGGCGTTGG 10080
 CGTGGGGACG ACTATCGCCC TCGGGCGGCT GCGGACGCTG CCGGCGGCGC AGTATGCCGA 10140
 GGGGTGGATC GTECTGATTG ACGACGCCGT CGATTTCCCT GACGCCGTGT GGTGGCTCAA 10200
 20 TGAGGCGCTC GACCGGGGGA TCAACGTGGT GCGGCGGATC CTCAAAAGG ACGACGGCGT 10260
 GCTGGTGAAC AACCGCCTGC GTAAAACCTT GCCGGTGGTG GATGAAGTGA CGCTGCTGGA 10320
 25 GCAGGTCCCC GAGGGGGTAA TGGCGGCGGT GGAAGTGCC CCGCCGGGCC AGGTGGTGCG 10380
 GATCCTGTGC AATCCCTACG GGATCGCCAC CTTCTTCGGG CTAAGCCCGG AAGAGACCCA 10440
 GGCCATCGTC CCCATCGCCC GCGCCCTGAT TGGCAACCGT TCCGCGGTGG TGCTCAAGAC 10500
 30 CCCGCAGGGG GATGTGCAGT CGCGGGTGAT CCCGGCGGGC AACCTCTACA TTAGCGGCGA 10560
 AAAGCGCCGC GGAGAGGCCG ATGTCGCCGA GGGCGCGGAA GCCATCATGC AGGCGATGAG 10620
 CGCCTGCGCT CCGGTACGCG ACATCCGCGG CGAACCGGGC ACCCACGCCG GCGGCATGCT 10680
 35 TGAGCGGGTG CGCAAGGTAA TGGCGTCCCT GACCGGCCAT GAGATGAGCG CGATATACAT 10740
 CCAGGATCTG CTGGCGGTGG ATACGTTTAT TCCGCGCAAG GTGCAGGGCG GGATGGCCGG 10800
 CGAGTGCGCC ATGGAGAAAG CCGTCGGGAT GCGGCGGATG GTGAAAGCGG ATCGTCTGCA 10860
 40 AATGCAGETT ATCGCCCGCG AACTGAGCGC CCGACTGCAG ACCGAGGTGG TGGTGGGCGG 10920
 CGTGGAGGCC AACATGGCCA TCGCCGGGGC GTTAACCACT CCGGCTGTG CGGCGCCGCT 10980
 45 GCGCATCCTC GACCTCGGCG CCGGCTCGAC GGATGCGGCG ATCGTCAACG CCGAGGGGCA 11040
 GATAACGGCG GTCCATCTCG CCGGGGCGGG GAATATGGTC AGCCTGTTGA TTAAAACCGA 11100
 CCTGGGCCCTC GAGGATCTTT CGCTGGCGGA AGCGATAAAA AAATACCCGC TGGCCAAAGT 11160
 50 GGAAAGCCTG TTCAGTATTC GTCACGAGAA TGGCGCGGTG GAGTTCCTTC GGAAGCCCT 11220
 CAGCCCGGCG GTGTTGCCA AAGTGGTGTA CATCAAGGAG GCGGAAGTGG TGCCGATCGA 11280
 55 TAACGCCAGC CCGCTGGAAA AAATTCGTCT CGTGCGCCGG CAGGCGAAG AGAAAGTGTT 11340
 TGTCACCAAC TGCCTGCGCG CGCTGCGCCA GGTCTCAGCC GCGGTTCCA TTCGCGATAT 11400
 CGCCTTTGTG GTGCTGGTGG GCGGCTCATC GCTGGACTTT GAGATCCCGC AGCTTATCAC 11460
 60 GGAAGCCTTG TCGCACTATG GCGTGGTCGC CGGGCAGGGC AATATTCGGG GAACAGAAGG 11520
 GCCGCGCAAT GCGGTCGCCA CCGGGCTGCT ACTGGCCGGT CAGGCGAATT AAACGGGCGC 11580

65

ES 2 336 858 T3

TCGCGCCAGC CTCTCTCTTT AACGTGCTAT TTCAGGATGC CGATAATGAA CCAGACTTCT 11640
 ACCTTAACCG GGCAGTGCGT GGCCGAGTTT CTTGGCACCG GATTGCTCAT TTTCTTCGGC 11700
 5 GCGGGCTGCG TCGCTGCGCT GCGGGTCGCC GGGGCCAGCT TTGGTCAGTG GGAGATCAGT 11760
 ATTATCTGGG GCCTTGGCGT CGCCATGGCC ATCTACCTGA CGGCCGGTGT CTCCGGCGCG 11820
 CACCTAAATC CGGCGGTGAC CATTGCCCTG TGGCTGTTTG CCTGTTTGA ACGCCGCAAG 11880
 10 GTGCTGCCGT TTATTGTTGC CCAGACGGCC GGGGCCCTTCT GCGCCGCCGC GCTGGTGTAT 11940
 GGGCTCTATC GCCAGCTGTT TCTCGATCTT GAACAGAGTC AGCATATCGT GCGCGGCACT 12000
 15 GCCGCCAGTC TTAACCTGGC CGGGGTCTTT TCCACGTACC CGCATCCACA TATCACTTTT 12060
 ATACAAGCGT TTGCCGTGGA GACCACCATC ACGGCAATCC TGATGGCGAT GATCATGGCC 12120
 CTGACCGACG ACGGCAACGG AATTC 12145
 20

(2) INFORMACIÓN PARA LA SEC. ID NO:20:

- (i) CARACTERÍSTICAS DE LA SECUENCIA:
- (A) LONGITUD: 94 pares de bases
 - (B) TIPO: ácido nucleico
 - (C) TIPO DE CADENA: simple
 - (D) TIPOLOGÍA: lineal
- (ii) TIPO DE MOLÉCULA: DNA (genómico)
- (xi) DESCRIPCIÓN DE LA SECUENCIA: SEC ID NO:20:

35 AGCTTAGGAG TCTAGAATAT TGAGCTCGAA TTCCCGGGCA TGCGGTACCG GATCCAGAAA 60
 AAAGCCCACA CCTGACAGTG CGGGCTTTTT TTTT 94
 40

(2) INFORMACIÓN PARA LA SEC. ID NO:21:

- (i) CARACTERÍSTICAS DE LA SECUENCIA:
- (A) LONGITUD: 37 pares de bases
 - (B) TIPO: ácido nucleico
 - (C) TIPO DE CADENA: simple
 - (D) TIPOLOGÍA: lineal
- (ii) TIPO DE MOLÉCULA: DNA (genómico)
- (xi) DESCRIPCIÓN DE LA SECUENCIA: SEC ID NO:21:

55 GGAATTCAGA TCTCAGCAAT GAGCGAGAAA ACCATGC 37

(2) INFORMACIÓN PARA LA SEC. ID NO:22:

- (i) CARACTERÍSTICAS DE LA SECUENCIA:
- (A) LONGITUD: 27 pares de bases
 - (B) TIPO: ácido nucleico
 - (C) TIPO DE CADENA: simple
 - (D) TIPOLOGÍA: lineal
- (ii) TIPO DE MOLÉCULA: DNA (genómico)

ES 2 336 858 T3

(xi) DESCRIPCIÓN DE LA SECUENCIA: SEC ID NO:22:

GCTCTAGATT AGCTTCCTTT ACGCAGC

27

5

(2) INFORMACIÓN PARA LA SEC. ID NO:23:

(i) CARACTERÍSTICAS DE LA SECUENCIA:

10

(A) LONGITUD: 33 pares de bases

(B) TIPO: ácido nucleico

(C) TIPO DE CADENA: simple

(D) TIPOLOGÍA: lineal

15

(ii) TIPO DE MOLÉCULA: DNA (genómico)

(xi) DESCRIPCIÓN DE LA SECUENCIA: SEC ID NO:23:

20

GGCCAAGCTT AAGGAGGTTA ATTAAATGAA AAG

33

(2) INFORMACIÓN PARA LA SEC. ID NO:24:

(i) CARACTERÍSTICAS DE LA SECUENCIA:

25

(A) LONGITUD: 26 pares de bases

(B) TIPO: ácido nucleico

(C) TIPO DE CADENA: simple

30

(D) TIPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: DNA (genómico)

(xi) DESCRIPCIÓN DE LA SECUENCIA: SEC ID NO:24:

35

GCTCTAGATT ATTCAATGGT GTCGGG

26

(2) INFORMACIÓN PARA LA SEC. ID NO:25:

40

(i) CARACTERÍSTICAS DE LA SECUENCIA:

(A) LONGITUD: 42 pares de bases

(B) TIPO: ácido nucleico

45

(C) TIPO DE CADENA: simple

(D) TIPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: DNA (genómico)

50

(xi) DESCRIPCIÓN DE LA SECUENCIA: SEC ID NO:25:

GCGCCGTCTA GAATTATGAG CTATCGTATG TTTGATTATC TG

42

55

(2) INFORMACIÓN PARA LA SEC. ID NO:26:

(i) CARACTERÍSTICAS DE LA SECUENCIA:

(A) LONGITUD: 36 pares de bases

60

(B) TIPO: ácido nucleico

(C) TIPO DE CADENA: simple

(D) TIPOLOGÍA: lineal

65

(ii) TIPO DE MOLÉCULA: DNA (genómico)

ES 2 336 858 T3

(xi) DESCRIPCIÓN DE LA SECUENCIA: SEC ID NO:26:

TCTGATACGG GATCCTCAGA ATGCCTGGCG GAT

36

5

(2) INFORMACIÓN PARA LA SEC. ID NO:27:

(i) CARACTERÍSTICAS DE LA SECUENCIA:

(A) LONGITUD: 51 pares de bases

(B) TIPO: ácido nucleico

(C) TIPO DE CADENA: simple

(D) TIPOLOGÍA: lineal

10

15

(ii) TIPO DE MOLÉCULA: DNA (genómico)

(xi) DESCRIPCIÓN DE LA SECUENCIA: SEC ID NO:27:

GCGCGGATCC AGGAGTCTAG AATTATGGGA TTGACTACTA AACCTCTATC T

20

51

(2) INFORMACIÓN PARA LA SEC. ID NO:28:

(i) CARACTERÍSTICAS DE LA SECUENCIA:

(A) LONGITUD: 36 pares de bases

(B) TIPO: ácido nucleico

(C) TIPO DE CADENA: simple

(D) TIPOLOGÍA: lineal

25

30

(ii) TIPO DE MOLÉCULA: DNA (genómico)

(xi) DESCRIPCIÓN DE LA SECUENCIA: SEC ID NO:28:

GATACGCCCCG GGTTACCATT TCAACAGATC GTCCTT

35

36

(2) INFORMACIÓN PARA LA SEC. ID NO:29:

(i) CARACTERÍSTICAS DE LA SECUENCIA:

(A) LONGITUD: 18 pares de bases

(B) TIPO: ácido nucleico

(C) TIPO DE CADENA: simple

(D) TIPOLOGÍA: lineal

40

45

(ii) TIPO DE MOLÉCULA: DNA (genómico)

(xi) DESCRIPCIÓN DE LA SECUENCIA: SEC ID NO:29:

TCGACGAATT CAGGAGGA

50

18

(2) INFORMACIÓN PARA LA SEC. ID NO:30:

(i) CARACTERÍSTICAS DE LA SECUENCIA:

(A) LONGITUD: 18 pares de bases

(B) TIPO: ácido nucleico

(C) TIPO DE CADENA: simple

(D) TIPOLOGÍA: lineal

55

60

(ii) TIPO DE MOLÉCULA: DNA (genómico)

65

ES 2 336 858 T3

(xi) DESCRIPCIÓN DE LA SECUENCIA: SEC ID NO:30:

CTAGTCCTCC TGAATTCG

18

5

(2) INFORMACIÓN PARA LA SEC. ID NO:31:

(i) CARACTERÍSTICAS DE LA SECUENCIA:

10

(A) LONGITUD: 19 pares de bases

(B) TIPO: ácido nucleico

(C) TIPO DE CADENA: simple

(D) TIPOLOGÍA: lineal

15

(ii) TIPO DE MOLÉCULA: DNA (genómico)

(xi) DESCRIPCIÓN DE LA SECUENCIA: SEC ID NO:31:

20

CTAGTAAGGA GGACAATTC

19

(2) INFORMACIÓN PARA LA SEC. ID NO:32:

(i) CARACTERÍSTICAS DE LA SECUENCIA:

25

(A) LONGITUD: 19 pares de bases

(B) TIPO: ácido nucleico

(C) TIPO DE CADENA: simple

30

(D) TIPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: DNA (genómico)

(xi) DESCRIPCIÓN DE LA SECUENCIA: SEC ID NO:32:

35

CATGGAATTG TCCTCCTTA

19

(2) INFORMACIÓN PARA LA SEC. ID NO:33:

40

(i) CARACTERÍSTICAS DE LA SECUENCIA:

(A) LONGITUD: 271 amino ácidos

(B) TIPO: amino ácido

45

(C) TIPO DE CADENA: desconocida

(D) TIPOLOGÍA: desconocida

(ii) TIPO DE MOLÉCULA: proteína

50

(vi) FUENTE ORIGINAL:

(A) ORGANISMO: GPP1

(xi) DESCRIPCIÓN DE LA SECUENCIA: SEC ID NO:33:

55

60

65

ES 2 336 858 T3

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

(2) INFORMACIÓN PARA LA SEC. ID NO:34:

- 60 (i) CARACTERÍSTICAS DE LA SECUENCIA:
- (A) LONGITUD: 555 amino ácidos
 - (B) TIPO: amino ácido
 - 65 (C) TIPO DE CADENA: desconocida
 - (D) TIPOLOGÍA: desconocida

ES 2 336 858 T3

(ii) TIPO DE MOLÉCULA: proteína

(vi) FUENTE ORIGINAL:

(A) ORGANISMO: DHAB1

5

(xi) DESCRIPCIÓN DE LA SECUENCIA: SEC ID NO:34:

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

55

60

65

ES 2 336 858 T3

Ser Val Tyr Gly Thr Glu Ala Val Phe Thr Asp Gly Asp Asp Thr Pro
 225 230 235 240
 5 Trp Ser Lys Ala Phe Leu Ala Ser Ala Tyr Ala Ser Arg Gly Leu Lys
 245 250 255
 10 Met Arg Tyr Thr Ser Gly Thr Gly Ser Glu Ala Leu Met Gly Tyr Ser
 260 265 270
 15 Glu Ser Lys Ser Met Leu Tyr Leu Glu Ser Arg Cys Ile Phe Ile Thr
 275 280 285
 Lys Gly Ala Gly Val Gln Gly Leu Gln Asn Gly Ala Val Ser Cys Ile
 290 295 300
 20 Gly Met Thr Gly Ala Val Pro Ser Gly Ile Arg Ala Val Leu Ala Glu
 305 310 315 320
 Asn Leu Ile Ala Ser Met Leu Asp Leu Glu Val Ala Ser Ala Asn Asp
 325 330 335
 25 Gln Thr Phe Ser His Ser Asp Ile Arg Arg Thr Ala Arg Thr Leu Met
 340 345 350
 Gln Met Leu Pro Gly Thr Asp Phe Ile Phe Ser Gly Tyr Ser Ala Val
 355 360 365
 30 Pro Asn Tyr Asp Asn Met Phe Ala Gly Ser Asn Phe Asp Ala Glu Asp
 370 375 380
 35 Phe Asp Asp Tyr Asn Ile Leu Gln Arg Asp Leu Met Val Asp Gly Gly
 385 390 395 400
 Leu Arg Pro Val Thr Glu Ala Glu Thr Ile Ala Ile Arg Gln Lys Ala
 405 410 415
 40 Ala Arg Ala Ile Gln Ala Val Phe Arg Glu Leu Gly Leu Pro Pro Ile
 420 425 430
 45 Ala Asp Glu Glu Val Glu Ala Ala Thr Tyr Ala His Gly Ser Asn Glu
 435 440 445
 Met Pro Pro Arg Asn Val Val Glu Asp Leu Ser Ala Val Glu Glu Met
 450 455 460
 50 Met Lys Arg Asn Ile Thr Gly Leu Asp Ile Val Gly Ala Leu Ser Arg
 465 470 475 480
 Ser Gly Phe Glu Asp Ile Ala Ser Asn Ile Leu Asn Met Leu Arg Gln
 485 490 495
 55 Arg Val Thr Gly Asp Tyr Leu Gln Thr Ser Ala Ile Leu Asp Arg Gln
 500 505 510
 60 Phe Glu Val Val Ser Ala Val Asn Asp Ile Asn Asp Tyr Gln Gly Pro
 515 520 525
 Gly Thr Gly Tyr Arg Ile Ser Ala Glu Arg Trp Ala Glu Ile Lys Asn
 530 535 540
 65 Ile Pro Gly Val Val Gln Pro Asp Thr Ile Glu

ES 2 336 858 T3

(2) INFORMACIÓN PARA LA SEC. ID NO:35:

(i) CARACTERÍSTICAS DE LA SECUENCIA:

- (A) LONGITUD: 194 amino ácidos
- (B) TIPO: amino ácido
- (C) TIPO DE CADENA: desconocida
- (D) TIPOLOGÍA: desconocida

(ii) TIPO DE MOLÉCULA: proteína

(vi) FUENTE ORIGINAL:

- (A) ORGANISMO: DHAB2

(xi) DESCRIPCIÓN DE LA SECUENCIA: SEC ID NO:35:

```

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

```

(2) INFORMACIÓN PARA LA SEC. ID NO:36:

(i) CARACTERÍSTICAS DE LA SECUENCIA:

- (A) LONGITUD: 140 amino ácidos
- (B) TIPO: amino ácido

ES 2 336 858 T3

(C) TIPO DE CADENA: desconocida

(D) TIPOLOGÍA: desconocida

(ii) TIPO DE MOLÉCULA: proteína

(vi) FUENTE ORIGINAL:

(A) ORGANISMO: DHAB3

(xi) DESCRIPCIÓN DE LA SECUENCIA: SEC ID NO:36:

```

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

```

(2) INFORMACIÓN PARA LA SEC. ID NO:37:

(i) CARACTERÍSTICAS DE LA SECUENCIA:

(A) LONGITUD: 387 amino ácidos

(B) TIPO: amino ácido

(C) TIPO DE CADENA: desconocida

(D) TIPOLOGÍA: desconocida

(ii) TIPO DE MOLÉCULA: proteína

(vi) FUENTE ORIGINAL:

(A) ORGANISMO: DHAT

(xi) DESCRIPCIÓN DE LA SECUENCIA: SEC ID NO:37:

ES 2 336 858 T3

Met Ser Tyr Arg Met Phe Asp Tyr Leu Val Pro Asn Val Asn Phe Phe
 1 5 10 15
 Gly Pro Asn Ala Ile Ser Val Val Gly Glu Arg Cys Gln Leu Leu Gly
 20 25 30
 Gly Lys Lys Ala Leu Leu Val Thr Asp Lys Gly Leu Arg Ala Ile Lys
 35 40 45
 Asp Gly Ala Val Asp Lys Thr Leu His Tyr Leu Arg Glu Ala Gly Ile
 50 55 60
 Glu Val Ala Ile Phe Asp Gly Val Glu Pro Asn Pro Lys Asp Thr Asn
 65 70 75 80
 Val Arg Asp Gly Leu Ala Val Phe Arg Arg Glu Gln Cys Asp Ile Ile
 85 90 95
 Val Thr Val Gly Gly Gly Ser Pro His Asp Cys Gly Lys Gly Ile Gly
 100 105 110
 Ile Ala Ala Thr His Glu Gly Asp Leu Tyr Gln Tyr Ala Gly Ile Glu
 115 120 125
 Thr Leu Thr Asn Pro Leu Pro Pro Ile Val Ala Val Asn Thr Thr Ala
 130 135 140
 Gly Thr Ala Ser Glu Val Thr Arg His Cys Val Leu Thr Asn Thr Glu
 145 150 155 160
 Thr Lys Val Lys Phe Val Ile Val Ser Trp Arg Lys Leu Pro Ser Val
 165 170 175
 Ser Ile Asn Asp Pro Leu Leu Met Ile Gly Lys Pro Ala Ala Leu Thr
 180 185 190
 Ala Ala Thr Gly Met Asp Ala Leu Thr His Ala Val Glu Ala Tyr Ile
 195 200 205
 Ser Lys Asp Ala Asn Pro Val Thr Asp Ala Ala Ala Met Gln Ala Ile
 210 215 220
 Arg Leu Ile Ala Arg Asn Leu Arg Gln Ala Val Ala Leu Gly Ser Asn
 225 230 235 240
 Leu Gln Ala Arg Glu Asn Met Ala Tyr Ala Ser Leu Leu Ala Gly Met
 245 250 255
 Ala Phe Asn Asn Ala Asn Leu Gly Tyr Val His Ala Met Ala His Gln
 260 265 270
 Leu Gly Gly Leu Tyr Asp Met Pro His Gly Val Ala Asn Ala Val Leu
 275 280 285
 Leu Pro His Val Ala Arg Tyr Asn Leu Ile Ala Asn Pro Glu Lys Phe
 290 295 300
 Ala Asp Ile Ala Glu Leu Met Gly Glu Asn Ile Thr Gly Leu Ser Thr
 305 310 315 320
 Leu Asp Ala Ala Glu Lys Ala Ile Ala Ala Ile Thr Arg Leu Ser Met
 325 330 335
 Asp Ile Gly Ile Pro Gln His Leu Arg Asp Leu Gly Val Lys Glu Ala
 340 345 350
 Asp Phe Pro Tyr Met Ala Glu Met Ala Leu Lys Asp Gly Asn Ala Phe
 355 360 365
 Ser Asn Pro Arg Lys Gly Asn Glu Gln Glu Ile Ala Ala Ile Phe Arg
 370 375 380
 Gln Ala Phe
 385

65

ES 2 336 858 T3

(2) INFORMACIÓN PARA LA SEC. ID NO:38:

(i) CARACTERÍSTICAS DE LA SECUENCIA:

(A) LONGITUD: 27 pares de bases

(B) TIPO: ácido nucleico

(C) TIPO DE CADENA: simple

(D) TIPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: DNA (genómico)

(xi) DESCRIPCIÓN DE LA SECUENCIA: SEC ID NO:38:

GCGAATTCAT GAGCTATCGT ATGTTTG

27

(2) INFORMACIÓN PARA LA SEC. ID NO:39:

(i) CARACTERÍSTICAS DE LA SECUENCIA:

(A) LONGITUD: 28 pares de bases

(B) TIPO: ácido nucleico

(C) TIPO DE CADENA: simple

(D) TIPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: DNA (genómico)

(xi) DESCRIPCIÓN DE LA SECUENCIA: SEC ID NO:39:

GCGAATTCAG AATGCCTGGC GGAAAATC

28

(2) INFORMACIÓN PARA LA SEC. ID NO:40:

(i) CARACTERÍSTICAS DE LA SECUENCIA:

(A) LONGITUD: 28 pares de bases

(B) TIPO: ácido nucleico

(C) TIPO DE CADENA: simple

(D) TIPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: DNA (genómico)

(xi) DESCRIPCIÓN DE LA SECUENCIA: SEC ID NO:40:

GGGAATTCAT GAGCGAGAAA ACCATGCG

28

(2) INFORMACIÓN PARA LA SEC. ID NO:41:

(i) CARACTERÍSTICAS DE LA SECUENCIA:

(A) LONGITUD: 27 pares de bases

(B) TIPO: ácido nucleico

(C) TIPO DE CADENA: simple

(D) TIPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: DNA (genómico)

(xi) DESCRIPCIÓN DE LA SECUENCIA: SEC ID NO:41:

GCGAATTCTT AGCTTCCTTT ACGCAGC

27

ES 2 336 858 T3

(2) INFORMACIÓN PARA LA SEC. ID NO:42:

(i) CARACTERÍSTICAS DE LA SECUENCIA:

(A) LONGITUD: 30 pares de bases

(B) TIPO: ácido nucleico

(C) TIPO DE CADENA: simple

(D) TIPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: DNA (genómico)

(xi) DESCRIPCIÓN DE LA SECUENCIA: SEC ID NO:42:

GCGAATTCAT GCAACAGACA ACCCAAATTC

30

(2) INFORMACIÓN PARA LA SEC. ID NO:43:

(i) CARACTERÍSTICAS DE LA SECUENCIA:

(A) LONGITUD: 25 pares de bases

(B) TIPO: ácido nucleico

(C) TIPO DE CADENA: simple

(D) TIPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: DNA (genómico)

(xi) DESCRIPCIÓN DE LA SECUENCIA: SEC ID NO:43:

GCGAATTCAC TCCCTTACTA AGTCG

25

(2) INFORMACIÓN PARA LA SEC. ID NO:44:

(i) CARACTERÍSTICAS DE LA SECUENCIA:

(A) LONGITUD: 30 pares de bases

(B) TIPO: ácido nucleico

(C) TIPO DE CADENA: simple

(D) TIPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: DNA (genómico)

(xi) DESCRIPCIÓN DE LA SECUENCIA: SEC ID NO:44:

GGGAATTCAT GAAAAGATCA AAACGATTTG

30

(2) INFORMACIÓN PARA LA SEC. ID NO:45:

(i) CARACTERÍSTICAS DE LA SECUENCIA:

(A) LONGITUD: 29 pares de bases

(B) TIPO: ácido nucleico

(C) TIPO DE CADENA: simple

(D) TIPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: DNA (genómico)

(xi) DESCRIPCIÓN DE LA SECUENCIA: SEC ID NO:45:

GCGAATTCTT ATTCAATGGT GTCGGGCTG

29

ES 2 336 858 T3

- (2) INFORMACIÓN PARA LA SEC. ID NO:46
- (i) CARACTERÍSTICAS DE LA SECUENCIA:
- (A) LONGITUD: 34 pares de bases
 - (B) TIPO: ácido nucleico
 - (C) TIPO DE CADENA: simple
 - (D) TIPOLOGÍA: lineal
- (ii) TIPO DE MOLÉCULA: DNA (genómico)
- (xi) DESCRIPCIÓN DE LA SECUENCIA: SEC ID NO:46:
- TTGATAATAT AACCATGGCT GCTGCTGCTG ATAG 34
- (2) INFORMACIÓN PARA LA SEC. ID NO:47
- (i) CARACTERÍSTICAS DE LA SECUENCIA:
- (A) LONGITUD: 39 pares de bases
 - (B) TIPO: ácido nucleico
 - (C) TIPO DE CADENA: simple
 - (D) TIPOLOGÍA: lineal
- (ii) TIPO DE MOLÉCULA: DNA (genómico)
- (xi) DESCRIPCIÓN DE LA SECUENCIA: SEC ID NO:47:
- GTATGATATG TTATCTTGGA TCCAATAAAT CTAATCTTC 39
- (2) INFORMACIÓN PARA LA SEC. ID NO:48:
- (i) CARACTERÍSTICAS DE LA SECUENCIA:
- (A) LONGITUD: 24 pares de bases
 - (B) TIPO: ácido nucleico
 - (C) TIPO DE CADENA: simple
 - (D) TIPOLOGÍA: lineal
- (ii) TIPO DE MOLÉCULA: DNA (genómico)
- (xi) DESCRIPCIÓN DE LA SECUENCIA: SEC ID NO:48:
- CATGACTAGT AAGGAGGACA ATTC 24
- (2) INFORMACIÓN PARA LA SEC. ID NO:49:
- (i) CARACTERÍSTICAS DE LA SECUENCIA:
- (A) LONGITUD: 24 pares de bases
 - (B) TIPO: ácido nucleico
 - (C) TIPO DE CADENA: simple
 - (D) TIPOLOGÍA: lineal
- (ii) TIPO DE MOLÉCULA: DNA (genómico)
- (xi) DESCRIPCIÓN DE LA SECUENCIA: SEC ID NO:49:
- CATGGAATTG TCCTCCTTAC TAGT 24