



19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA

11 Número de publicación: **2 345 469**

51 Int. Cl.:

**C12N 15/53** (2006.01)

**C12N 15/55** (2006.01)

**C12P 7/20** (2006.01)

**C12N 1/15** (2006.01)

**C12N 1/19** (2006.01)

**C12N 1/21** (2006.01)

**C12N 9/04** (2006.01)

**C12N 9/16** (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Número de solicitud europea: **97948194 .2**

96 Fecha de presentación : **10.11.1997**

97 Número de publicación de la solicitud: **0948628**

97 Fecha de publicación de la solicitud: **13.10.1999**

54

Título: **Método para la producción de glicerol por organismos recombinantes.**

30

Prioridad: **13.11.1996 US 30602 P**

45

Fecha de publicación de la mención BOPI:  
**23.09.2010**

45

Fecha de la publicación del folleto de la patente:  
**23.09.2010**

73

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

72

Inventor/es: **Bulhuis, Ben, A.;**  
**Gatenby, Anthony, Arthur;**  
**Haynie, Sharon, Loretta;**  
**Hsu, Amy, Kuang-Hua y**  
**Lareau, Richard, D.**

74

Agente: **Elzaburu Márquez, Alberto**

ES 2 345 469 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 glicerol por 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 para glicerol-3-fosfato deshidrogenasa (G3PDH) y glicerol-3-fosfatasa (G3P fosfatasa) para la producción reforzada de glicerol.

10 **Antecedentes**

El glicerol es un compuesto muy demandado por la industria para uso en cosméticos, jabones líquidos, alimentos, productos farmacéuticos, lubricantes, soluciones anti-congelantes, y en numerosas otras aplicaciones. Los ésteres de glicerol son importantes en la industria de las grasas y del aceite.

No todos los organismos tienen una capacidad natural de sintetizar glicerol. Sin embargo, la producción biológica de glicerol es conocida para algunas especies de bacterias, algas y levaduras. Las bacterias *Bacillus licheniformis* y *Lactobacillus lycopersica* sintetizan glicerol. La producción de glicerol se encuentra en las algas halotolerantes *Dunaliella sp.* y *Asteromonas gracilis* para la protección frente a concentraciones salinas externas elevadas (Ben-Amotz *et al.*, (1982) *Experientia* 38:49-52). 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.*, (1994) *Mol. Cell.Biol.* 14, 4135-4144). En los primeros años de este siglo, el glicerol se produjo comercialmente con cultivos de *Saccharomyces* a los cuales se añadieron 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 dihidroxiacetona fosfato (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 glicerol-3-fosfato deshidrogenasa (DAR1,GPD1) ha sido clonado y secuenciado a partir de *Saccharomyces diastaticus* (Wang *et al.*, (1994), *J. Bact.* 176:7091-7095). El gen DAR1 se clonó en un vector lanzadera 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 glicerol en un organismo recombinante.

Se han aislado otras enzimas de glicerol-3-fosfato deshidrogenasa. Por ejemplo, sn-glicerol-3-fosfato deshidrogenasa ha sido clonada y secuenciada a partir de *S. cerevisiae* (Larason *et al.*, (1993) *Mol. Microbiol.*, 10:1101, (1993)). Albertyn *et al.*, (1994) *Mol. Cell. Biol.* 14: 4135) enseñan la clonación de GPD1 que codifica una glicerol-3-fosfato deshidrogenasa de *S. cerevisiae*. Al igual que Wang *et al.*, tanto Albertyn *et al.* como Larason *et al.* reconocen la osmosensibilidad de la regulación de este gen, pero no sugieren cómo se podría usar el gen en la producción de glicerol en un organismo recombinante.

Como con G3DPH, 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.*, (1996) *J. Biol. Chem.* 271:13875). Al igual que los genes que codifican G3DPH, parece que GPP2 está osmóticamente inducido.

No existe técnica conocida que enseñe que la producción de glicerol a partir de organismos recombinantes con G3PDH/G3P fosfatasa se exprese junto o por separado. Tampoco existe técnica conocida que enseñe que la producción de glicerol a partir de cualquier organismo de tipo salvaje con estas dos actividades de enzima no requiere aplicar un cierto estrés (una sal o un osmolito) a la célula. Eustace ((1987), *Can. J. Microbiol.*, 33:112-117)) enseñan conseguir la producción de glicerol mediante técnicas de ADN recombinante. Por técnicas de cultivo selectivas, estos investigadores crearon una cepa de levadura hibridada que producía glicerol a niveles más altos que las cepas parentales; sin embargo, la actividad de G3PDH permanecía constante o era ligeramente más baja.

Nevoight *et al.* (Yeast 12:1331-1337, 1996) describe que la sobre-expresión de glicerol-3-fosfato deshidrogenasa en *Saccharomyces cerevisiae* se puede combinar con un nivel incrementado de fosfatasa para la producción de glicerol.

Un microorganismo capaz de producir glicerol en condiciones fisiológicas es industrialmente deseable, en especial cuando el propio glicerol será utilizado como un sustrato *in vivo* como parte de una vía catabólica o biosintética más compleja que podría ser perturbada por el estrés osmótico o la adición de agentes de dirección.

Por lo tanto, el problema a resolver es cómo dirigir el flujo de carbono hacia la producción de glicerol mediante la adición o el refuerzo de determinadas actividades enzimáticas, en especial G3PDH y G3P fosfatasa que catalizan, respectivamente, la conversión de dihidroxiacetona fosfato (DHAP) en glicerol-3-fosfato (G3P) y luego en glicerol.

Este procedimiento no ha sido descrito previamente para un organismo recombinante y requería el aislamiento de genes que codifiquen las dos enzimas y su expresión subsiguiente. Una dificultad sorprendente y no anticipada con la que se topó era la toxicidad de G3P fosfatasa al hospedante que requería un control cuidadoso de sus niveles de expresión para evitar la inhibición del crecimiento.

### Sumario de la invención

La presente invención proporciona un método para la producción de glicerol a partir de un organismo recombinante, que comprende:

(i) transformar una célula de *E.coli* con una casete de expresión, que comprende:

(a) un gen que codifica una enzima glicerol-3-fosfato deshidrogenasa NADH-dependiente o una enzima glicerol-3-fosfato deshidrogenasa NADH-dependiente; y

(b) un gen que codifica una enzima glicerol-3-fosfato fosfatasa;

(ii) cultivar la célula de *E.coli* de (i) en presencia de al menos una fuente de carbono seleccionada del grupo que consiste en monosacáridos, oligosacáridos, polisacáridos y sustratos de un solo carbono, con lo que se produce glicerol; y

(iii) recuperar el glicerol. La glucosa es la fuente de carbono más preferida.

La invención proporciona, además, células hospedantes de *E. coli* transformadas, que comprenden:

(a) un gen que codifica una enzima glicerol-3-fosfato deshidrogenasa NADH-dependiente o una enzima glicerol-3-fosfato deshidrogenasa NADH-dependiente; y

(b) un gen que codifica una enzima glicerol-3-fosfato fosfatasa.

### Breve descripción de depósitos biológicos y listado de secuencias

La solicitante hizo los siguientes depósitos biológicos bajo los términos del Tratado de Budapest sobre el Reconocimiento Internacional del Depósito de Microorganismos para los Fines del Proceso de Patentes:

Referencia de Identificación del Depositador	Designación Int. del Depositorio	Fecha de Depósito
<i>Escherichia coli</i> pAH21/DH5α (que contiene el gen GPP2)	ATCC 98187	26 de septiembre de 1996
<i>Escherichia coli</i> (pDAR 1 A/AA200) (que contiene el gen DAR1)	ATCC 98248	6 de noviembre de 1996

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

La solicitante ha proporcionado 23 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 la producción biológica de glicerol a partir de una fuente de carbono fermentable en un organismo recombinante. El método proporciona una fuente de glicerol rápida, económica y mediambientalmente responsable en las industrias cosmética y farmacéutica. El método utiliza un microorganismo que contiene genes homólogos o heterólogos clonados que codifican glicerol-3-fosfato deshidrogenasa (G3PDH) y glicerol-3-fosfatasa (G3P fosfatasa). El microorganismo se pone en contacto con una fuente de carbono y el glicerol se aísla a partir de los medios acondicionados. Los genes se pueden incorporar en el microorganismo hospedante, por separado o juntos, para la producción de glicerol.

Tal como se utiliza en la presente memoria las siguientes expresiones pueden utilizarse para la interpretación de las reivindicaciones y de la memoria.

## ES 2 345 469 T3

Las expresiones “glicerol-3-fosfato deshidrogenasa” y “G3PDH” se refieren a un polipéptido responsable de una actividad enzimática que cataliza la conversión de dihidroxiacetona fosfato (DHAP) en glicerol-3-fosfato (G3P). G3PDH *in vivo* puede ser NADH; NADPH; o FAD-dependiente. La enzima NADH-dependiente (EC 1.1.1.8) está codificada por varios genes que incluyen GPD1 (GenBank Z74071x2) o GPD2 (GenBank Z35169x1) o GPD3 (GenBank G984182) o DART1 (GenBank Z74071 x2). La enzima NADPH-dependiente (EC 1.1.1.94) está codificada por *gpsA* (GenBank U321643, cds 197911-196892) G466746 y L45246). La enzima FAD-dependiente (EC 1.1.99.5) está codificada por GUT2 (GenBank Z47047x23) o *glpD* (GenBank G147838) o *glpABC* (GenBank M20938).

Las expresiones “glicerol-3-fosfatasa”, “sn-glicerol-3-fosfatasa” o “d,l-glicerol fosfatasa” y “G3P fosfatasa” se refieren a un polipéptido responsable de una actividad enzimática que cataliza la conversión de glicerol-3-fosfato en glicerol. G3P fosfatasa es codificada por GPP1 (GenBank Z47047x125) o GPP2 (GenBank U18813x11).

La expresión “glicerol quinasa” se refiere a un polipéptido responsable de una actividad enzimática que cataliza la conversión de glicerol en glicerol-3-fosfato o glicerol-3-fosfato en glicerol, dependiendo de las condiciones de reacción. Glicerol-quinasa es codificada por GUT1 (GenBank U11583x19).

Los términos “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 se caracteriza por la secuencia de bases proporcionada como SEQ ID NO:1.

El término “GPD2” se refiere a un gen que codifica una glicerol-3-fosfato deshidrogenasa citosólica, y se caracteriza por la secuencia de bases proporcionada en SEQ ID NO:2.

Los términos “GUT2” y “YIL155C” se usan de manera indistinta, y se refieren a un gen que codifica una glicerol-3-fosfato deshidrogenasa mitocondrial, y se caracteriza por la secuencia de bases proporcionada en SEQ ID NO:3.

Los términos “GPP1”, “RHR2” y “YIL053W” se usan de manera indistinta, y se refieren a un gen que codifica una glicerol-3-fosfatasa citosólica, y se caracteriza por la secuencia de bases proporcionada en SEQ ID NO:4.

Los términos “GPP2”, “HOR2” y “YER062C” se usan de manera indistinta, y se refieren a un gen que codifica una glicerol-3-fosfatasa citosólica, y se caracteriza por la secuencia de bases proporcionada en SEQ ID NO:5.

La expresión “GUT1” se refiere a un gen que codifica una glicerol quinasa citosólica, y se caracteriza por la secuencia de bases proporcionada en SEQ ID NO:6.

Tal como se utiliza en esta memoria, las expresiones “función” y “función enzimática” se refieren a la actividad catalítica de una enzima para 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 tanto de producto como de sustrato se puede llevar a cabo en condiciones adecuadas.

Los términos “polipéptido” y “proteína” se usan de manera indistinta en esta memoria.

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 significan 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” y “organismo hospedador” se refieren a un microorganismo capaz de recibir genes exógenos o heterólogos y de expresar esos genes para preparar un producto génico activo.

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

Las expresiones “organismo recombinante” y “hospedador transformado” se refieren a cualquier organismo transformado con genes heterólogos o exógenos. Los organismos recombinantes de la presente invención expresan genes exógenos que codifican G3PDH y G3P-fosfatasa para la producción de glicerol 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'). Las expresiones gen “nativo” y “de tipo salvaje” se refieren al gen tal como se halla en la naturaleza con sus propias secuencias reguladoras.

Tal como se utiliza en esta memoria, 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. El proceso de codificar una secuencia de aminoácidos específica pretende incluir 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. Por lo tanto, la invención abarca más de las secuencias específicas dadas a modo de ejemplo. También se contemplan modificaciones en la secuencia, tales como deleciones, 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 contemplan alteraciones en la secuencia génica que reflejan la degeneración del código genético, o que dan lugar a la producción de un aminoácido químicamente equivalente en un lugar 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, también 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 alteración 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, SDS al 0,1%, 65°C), con las secuencias ejemplificadas en esta 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”, tal como se utilizan en esta memoria, se refieren a un elemento cromosómico extra que a menudo porta genes que no son parte del metabolismo principal de la célula, y están 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 a la célula que resulta de una transformación.

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

#### *Ruta enzimática representativa*

Se contempla que el glicerol se puede producir en organismos recombinantes mediante la manipulación de la ruta biosintética del glicerol que se encuentra en la mayoría de los microorganismos. Típicamente, un sustrato de carbono, tal como glucosa, se convierte en glucosa-6-fosfato a través de hexoquinasa en presencia de ATP. La glucosa-fosfato isomerasa cataliza la conversión de glucosa-6-fosfato en fructosa-6-fosfato y luego en fructosa-1,6-difosfato a través de la acción de 6-fosfofructoquinasa. El difosfato se recoge luego en dihidroxiacetona fosfato (DHAP) a través de la aldolasa. Finalmente, G3PDH NADH-dependiente convierte DHAP en glicerol-3-fosfato que después de desfosforila para dar glicerol mediante G3P fosfatasa. (Agarwal (1990), *Adv. Biochem. Engrg.* 41:114).

#### *Rutas alternativas para la producción de glicerol*

Se ha sugerido una ruta alternativa para la producción de glicerol a partir de DHAP (Wang *et al.*, (1994) *J. Bact.* 176:7091-7095). En esta ruta propuesta, DHAP podría ser desfosforilado por una fosfatasa específica o no específica para dar dihidroxiacetona, la cual se podría luego reducir en glicerol mediante una dihidroxiacetona reductasa. La dihidroxiacetona reductasa es conocida en los procariotas y en *Schizosaccharomyces pombe*, y la clonación y expresión de actividades de este tipo, junto con una fosfatasa apropiada podría conducir a la producción de glicerol. Se ha sugerido otra ruta alternativa para la producción de glicerol a partir de DHAP (Redkar (1995), *Experimental Mycology*, 19:241, 1995). En esta ruta, DHAP se isomeriza para dar gliceraldehído-3-fosfato mediante la enzima glucolítica común triosa fosfato isomerasa. La gliceraldehído-3-fosfato se desfosforila para dar gliceraldehído, que luego se reduce mediante actividad de alcohol deshidrogenasa o glicerol deshidrogenasa NADP-dependiente. La clonación y expresión de las actividades de fosfatasa y deshidrogenasa a partir de *Aspergillus nidulans* podrían conducir a la producción de glicerol.

#### *Genes que codifican G3PDH y G3P fosfatasa*

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

## ES 2 345 469 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:1, que codifica la secuencia de aminoácidos proporcionada en SEQ ID NO:7 (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:2, que codifica la secuencia de aminoácidos proporcionada en SEQ ID NO:8 (Eriksson *et al.*, (1995) *Mol. Microbiol.*, 17:95).

Para los fines de la presente invención se contempla que sea adecuado cualquier gen que codifique un polipéptido responsable de la actividad de G3PDH, 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:7, 8, 9, 10, 11 y 12, que corresponden a los genes GPD1, GPD2, GUT2, *gpsA*, *glpD*, y la subunidad  $\alpha$  de *glpABC*, respectivamente, será funcional en la presente invención, en la que esa secuencia de aminoácidos puede abarcar las sustituciones, deleciones 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 serán 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, U32689 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 a partir de *Saccharomyces cerevisiae* y tiene la secuencia de bases dada por SEQ ID NO:5, que codifica la secuencia de aminoácidos dada en SEQ ID NO:13 (Norbeck *et al.*, (1996), *J. Biol. Chem.*, 271:13875).

Para los fines de la presente invención, cualquier gen que codifique una actividad de G3P fosfatasa es adecuado para uso en el método, en el que esa actividad es capaz de catalizar la conversión de glicerol-3-fosfato en glicerol. Además, cualquier gen que codifique la secuencia de aminoácidos de G3P fosfatasa proporcionada en SEQ ID NOS:13 y 14, correspondiente a los genes GPP2 y GPP1, respectivamente, será funcional en la presente invención, incluida cualquier secuencia de aminoácidos que abarque sustituciones, deleciones o adiciones de aminoácidos que no alteran la función de la enzima G3P fosfatasa. Los expertos apreciarán que los genes que codifican G3P fosfatasa aislados de otras fuentes también serán adecuados para el uso en la presente invención. Por ejemplo, la desfosforilación de glicerol-3-fosfato para proporcionar glicerol se puede conseguir 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.-) [GenBankZ38060 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.*, (1993), *Curr. Genet.*, 24, 21), y la secuencia de bases se proporciona en SEQ ID NO:6, que codifica la secuencia de aminoácidos proporcionada en SEQ ID NO:15. 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, convirtiendo glicerol-3-fosfato en glicerol en las condiciones de energía de reacción apropiadas. Existe evidencia 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 compartimiento del glicosoma (Hammond, (1985), *J. Biol. Chem.* 260: 15646-15654).

### *Células hospedadoras*

Células hospedadoras adecuadas para la producción recombinante de glicerol mediante la expresión de G3PDH y G3P fosfatasa son *E. coli*.

### *Vectores y casetes de expresión*

La presente invención proporciona 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 hospedadora adecuada. Vectores adecuados serán aquellos que sean compatibles con la bacteria empleada. 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. Protocolos para obtener y utilizar vectores de este tipo son conocidos en la técnica (Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual* - volúmenes 1, 2, 3 (Cold Spring Harbor Laboratory: Cold Spring Harbor, NY, 1989)).

Típicamente, el vector o la casete contiene las secuencias que dirigen la transcripción y la traducción del gen apropiado, un marcador seleccionable, y las secuencias que permiten la replicación autónoma o la integración cromosómica. Vectores adecuados comprenden una región en 5' del gen que alberga los controles de la iniciación transcripcional y una región en 3' del fragmento de ADN que controla la terminación transcripcional. Lo más preferido es cuando las dos regiones control se derivan de genes homólogos a la célula hospedadora transformada. Regiones control de este tipo no necesitan derivarse de los genes nativos para las especies específicas elegidas como un hospedador de producción.

## ES 2 345 469 T3

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 y TPI (útiles para la expresión en *Saccharomyces*); AOX1 (útil para la expresión en *Pichia*); y lac, trp,  $\lambda P_L$ ,  $\lambda P_R$ , 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 no ser necesario un sitio de terminación; sin embargo, es lo más preferido si está incluido.

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 G3PDH y G3P fosfatasa para la producción de glicerol*

Una vez que se construyen casetes adecuadas, se usan para transformar las células hospedadoras apropiadas. La introducción de la casete que contiene los genes que codifican G3PDH y/o G3P fosfatasa en la célula hospedadora se puede conseguir mediante procesos conocidos, tales como mediante transformación, p. ej., utilizando células permeabilizadas por calcio, electroporación o mediante transfección utilizando un virus de fago recombinante (Sambrook *et al.*, anteriormente mencionado).

En la presente invención se utilizaron casetes AH21 y DAR1 para transformar la DH5 $\alpha$  de *E. coli* tal como se describe completamente en los Métodos generales y los Ejemplos.

### *Medios y sustratos de carbono*

El medio de fermentación en la presente invención debe contener sustratos de carbono adecuados. Sustratos adecuados pueden incluir pero no se limitan a 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.*, (1989), *Agric. Biol. Chem.*, 53(2):541-543) y en bacterias (Hunter *et al.* (1985), *Biochemistry*, 24:4148-4155). 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, editorial Springer: 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 finalmente 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 sustratos con uno y dos carbonos los organismos metilótrofos son también conocidos por utilizar un cierto número de otros compuestos que contienen carbono tales como metilamina, glucosamina y una variedad de aminoácidos para actividad metabólica. Por ejemplo, las levaduras metilótrofas son conocidas por utilizar el carbono de la metilamina para formar trehalosa o glicerol (Bellion *et al.* (1993), *Microb. Growth CI Compd.*, [Int. Symp.], 7<sup>a</sup>, 415-32. Editor(es): Murrell, J. Collin; Kelly, Don P. Publisher: Intercept, Andover, Reino Unido). Asimismo, diversas especies de *Candida* metabolizarán la alanina o el ácido oleico (Sulter *et al.* (1990), *Arch. Microbiol.*, 153(5), 485-9). 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 la elección del organismo.

Aunque todos los sustratos de carbono mencionados anteriormente y las mezclas de los mismos son adecuados en la presente invención, sustratos de carbono preferidos son monosacáridos, oligosacáridos, polisacáridos, sustratos de un carbono o mezclas de los mismos. Los más preferidos son azúcares, tales como glucosa, fructosa, sacarosa, maltosa, lactosa y los sustratos de un solo carbono, tales como metanol y dióxido de carbono. El más preferido como sustrato de carbono es glucosa.

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.

*Condiciones de cultivo*

Típicamente, las células se cultivan a 30°C en medios apropiados. Medios de cultivo preferidos son medios comerciales comercialmente preparados tal como el caldo de cultivo Luria Bertani (LB), el caldo de cultivo con dextrosa Sabouraud (SD) o el caldo de cultivo con medio 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. El uso de agentes conocidos para modular la represión del catabolito directa o indirectamente, p. ej., 2':3'-monofosfato de adenosina cíclico, puede también incorporarse en los medios 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.

Intervalos de pH adecuados para la fermentación son entre pH 5,0 y 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.

20 *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 actividad de G3PDH se basa en las propiedades espectrales del co-sustrato, NADH, en la conversión de DHAP en 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 de fosfomolibdato-amonio de color azul.

30 *Identificación y recuperación de glicero*

El glicerol se puede identificar y cuantificar mediante cromatografía líquida de alta eficacia (HPLC) y análisis de cromatografía de gases/espectroscopía de masas (GC/MS) en los extractos exentos de células. Se prefiere un método en el que los medios de fermentación se analizan en una columna analítica de intercambio iónico mediante el uso de una fase móvil de ácido sulfúrico 0,01 N de forma isocrática.

Métodos para la recuperación de glicerol a partir de los medios de fermentación son conocidos en la técnica. Por ejemplo, el glicerol se puede obtener a partir de medios celulares, sometiendo la mezcla de reacción a la siguiente secuencia de etapas: filtración; separación de agua; extracción con disolvente orgánico; y destilación fraccional (patente de EE.UU. n° 2.986.495).

*Selección de transformantes mediante complementación*

En ausencia de una G3PDH codificada por *gpsA*-funcional, las células *E. coli* son incapaces de sintetizar G3P, un estado que conduce a un bloque en la biosíntesis de la membrana. Células con un bloque de este tipo son auxotróficas, requiriendo que glicerol o G3P estén presentes en los medios de cultivo para la síntesis de fosfolípidos de la membrana.

Un gen *gpsA* de tipo salvaje heterólogo clonado es capaz de complementar la mutación cromosomal de *gpsA* para permitir el crecimiento en medios que carecen de glicerol o G3P (Wang, *et al.* (1994), *J. Bact.* 176:7091-7095). Basado en esta estrategia de complementación, el crecimiento de células defectuosas en *gpsA* sobre glucosa sólo se produciría si poseyeran un *gpsA* codificado por plásmidos, permitiendo una selección basada en la síntesis de G3P a partir de DHAP. Células que pierden el plásmido *gpsA* recombinante durante el cultivo fracasarían en sintetizar G3P y el crecimiento de la célula sería subsiguientemente inhibido. La actividad complementaria de G3PDH puede ser expresada no sólo a partir de *gpsA*, sino también a partir de otros genes clonados que expresen la actividad de G3PDH, tal como GPD1, GPD2, GPD3, GUT2, *glpD* y *glpABC*. Estos se pueden mantener en una cepa de *E. coli* defectuosa en *gpsA* tal como BB20 (Cronan *et al.* (1974), *J. Bact.*, 118:598), paliando la necesidad de utilizar una selección de antibióticos y su coste prohibitivo en fermentaciones a gran escala.

Una estrategia relacionada se puede utilizar para la expresión y selección en mutantes osmorreguladores de *S. cerevisiae* (Larsson *et al.* (1993), *Mol. Microbiol.*, 10:1101-1111). Estos mutantes *osg 1* son incapaces de crecer a un bajo potencial de agua y muestran una capacidad disminuida de producir glicerol y una actividad reducida de G3PDH. El defecto de sensibilidad salina de *osg1* se puede complementar mediante un gen G3PDH clonado y expresado. Así, la capacidad de sintetizar glicerol se puede utilizar simultáneamente como un marcador de selección para las células productoras de glicerol deseadas.

## Ejemplos

## Métodos generales

5 Procedimientos para fosforilaciones, ligaduras y transformaciones son bien conocidos en la técnica. Técnicas adecuadas para su utilización en los ejemplos siguientes pueden encontrarse en Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual*, segunda edición, Cold Spring Harbor Laboratory Press (1989).

10 Los materiales y métodos adecuados para el mantenimiento y crecimiento de cultivos bacterianos son bien conocidos en la técnica. Técnicas adecuadas para su utilización en los ejemplos siguientes pueden encontrarse 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 *Biotechnology: A Textbook of Industrial Microbiology* (Thomas D. Brock, Segunda Edición (1989) Sinauer Associates, Inc., Sunderland, MA). Todos los reactivos y materiales utilizados para el crecimiento y mantenimiento de las células bacterianas se adquirieron en Aldrich Chemicals (Milwaukee, WI), DIFCO Laboratories (Detroit, MI), GIBCO/BRL (Gaithersburg, MD) o Sigma Chemical Company (St. Louis, MO) a menos que se especifique de otra manera.

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

## Cepas de células

25 Se utilizaron las siguientes cepas de *Escherichia coli* para la transformación y expresión de G3PDH y G3P fosfatasa. Las cepas se obtuvieron de *E. coli* Genetic Stock Center o de Life Technologies, Gaithersburg, MD).

AA200 (*garB10 fhuA22 ompF627 fadL701 relA1 pit-10 spoT1 tpi-1 phoM510 mcrB1*) (Anderson *et al.*, (1970), *J. Gen. Microbiol.*, 62:329).

30 BB20 (*tonA22 ΔphoA8 fadL701 relA1 glpR2 glpD3 pit-10 gpsA20 spvT1 T2R*) (Cronan *et al.*, *J. Bact.*, 118:598).

DH5α (*deoR endA1 gyrA96 hsdR17 reeA1 relA1 supE44 thi-1 Δ(lacZYA-argFV169) phi80lacZΔM15 F*) (Woodcock *et al.*, (1989), *Nucl. Acids Res.*, 17:3469).

## 35 Identificación de Glicerol

La conversión de glucosa en glicerol se vigiló mediante HPLC y/o GC. Se realizaron análisis utilizando técnicas normalizadas y materiales disponibles por un experto en materia 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 en una columna Shodex SH-1011 (8 mm x 300 mm; Waters, Milford, MA) equipada con una precolumna Shodex SH-1011P (6 mm x 50 mm), controlada en temperatura a 50°C, utilizando H<sub>2</sub>SO<sub>4</sub> 0,01 N en calidad de fase móvil a un caudal de 0,5 mL/min. Cuando se deseaba un análisis cuantitativo, las muestras se inyectaron en una columna Shodex SH-1011 (8 mm x 300 mm; Waters, Milford, MA) equipada con una precolumna Shodex SH-1011P (6 mm x 50 mm), controlada en temperatura a 50°C, utilizando H<sub>2</sub>SO<sub>4</sub> 0,01 N en calidad de fase móvil a un caudal de 0,69 mL/min. Cuando se deseaba un análisis cuantitativo, las muestras se prepararon con una cantidad conocida de ácido trimetilacético en calidad de patrón externo. Típicamente, los tiempos de retención de glicerol (detección R1) y glucosa (detección R1) eran 17,03 min y 12,66 min, respectivamente.

50 El glicerol también se analizó mediante GC/MS. La cromatografía de gases con detección por espectroscopía de masas y la cuantificación de glicerol se realizó utilizando una columna DB-WAX (30 m, 0,32 mm D.I., espesor de la película de 0,25 μm, J & W Scientific, Folsom, CA), en las condiciones siguientes: inyector: rendija, 1: 15; volumen de la muestra: 1 μL; perfil de temperaturas: temperatura inicial 150°C con una pausa de 30 s, 40°C/min a 180°C, 20°C/min a 240°C, pausa durante 2,5 min. Detección: Espectrometría de masas EI (Hewlett Packard 5971, San Fernando, CA), SIM cuantitativo utilizando iones 61 m/z y 64 m/z como iones diana para el glicerol y el glicerol-d8, y el ion 43 m/z como ion cualificador para el glicerol. Glicerol-d8 se utilizó como un patrón interno.

## Ensayo de glicerol-3-fosfatasa, GPP

60 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 utilizado era 1-α-glicerol fosfato, o d,l-α-glicerol fosfato. Las concentraciones finales de los reactivos en el ensayo son: tampón (bis-Tris 20 mM o MES 50 mM); MgCl<sub>2</sub> (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 μL, 200 mM), tampón MES 50 mM, MgCl<sub>2</sub> 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 el periodo de tiempo de si la actividad de la fosfatasa en la muestra de enzima oscilaba de 2 a 0,02 U/mL. La reacción enzimática se enfrió bruscamente mediante la adición del reactivo de molibdato ácido (0,4

## ES 2 345 469 T3

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 que se desarrollara. Al cabo de 10 min, para permitir un desarrollo completo del color, se leyó la absorbancia de las muestras a 660 nm utilizando 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 entre 0,026 y 0,130  $\mu\text{mol/mL}$ .

### *Ensayo espectrofotométrico de la actividad de glicerol 3-fosfato deshidrogenasa (G3PDH)*

Se utilizó el siguiente proceso, modificado como se muestra más abajo, a partir de un método publicado por Bell *et al.* (1975), *J. Biol. Chem.*, 250:7153-8. Este método implicaba incubar una muestra enzimática en una cubeta que contenía NADH 0,2 mM; dihidroxiacetona fosfato (DHAP) 2,0 mM y enzima en tampón Tris 0,1 M/HCl, pH 7,5 con DTT 5 mM, en un volumen total de 1,0 mL a 30°C. El espectrofotómetro se ajustó para vigilar los cambios en la absorbancia a una longitud de onda fija de 340 nm. El instrumento se ajustó a cero en una cubeta que solamente contenía tampón. Después de haber añadido la enzima a la cubeta, se hizo una lectura de la absorbancia. El primer sustrato, NADH (50  $\mu\text{L}$  NADH 4 mM; la absorbancia debería aumentar aproximadamente 1,25 UA), se añadió para determinar el ritmo de fondo. El ritmo debería seguirse durante al menos 3 min. Luego se añadió el segundo sustrato DHAP (50  $\mu\text{L}$  DHAP 40 mM), y se vigiló el cambio de la absorbancia a lo largo del tiempo durante al menos 3 min para determinar el ritmo bruto. La actividad de G3PDH se definió sustrayendo el ritmo de fondo del ritmo bruto.

### *Construcción del plásmido y construcción de la cepa*

#### *Clonación y expresión de la glicerol 3-fosfatasa para el incremento de una producción de glicerol 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) se clonó mediante PCR a partir del clon lambda como ADN objetivo mediante el uso de cebadores sintéticos (SEQ ID NO:16 con SEQ ID NO:17), incorporando un sitio BamHI-RBS-XbaI en el extremo 5' y un sitio SmaI en el extremo 3'. El producto se subclonó en pCR-Script (Stratagene, Madison, WI) en el sitio SrfI para generar los plásmidos pAH15 que contenían la GPP2. El plásmido pAH15 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. pAH21/ DH5 $\alpha$  es el plásmido de expresión.

#### *Plásmidos para la sobre-expresión 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:18 con SEQ ID NO:19). 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, NJ) 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:20 con SEQ ID NO:21) 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 pTrc99A (Pharmacia, Piscataway, NJ). Se insertó el fragmento NcoI-BamHI de pDAR1A y un segundo grupo de ligadores SpeI-RBS-NcoI obtenidos renaturalizando cebadores sintéticos (SEQ ID NO:22 con SEQ ID NO:23) en el sitio SpeI-BamHI de pBC-SK+ (Stratagene, Madison, WI) para crear el plásmido pAH42. El plásmido pAH42 contiene un gen resistente a cloranfenicol.

#### *Construcción de casetes de expresión para DAR1 y GPP2*

Se construyeron casetes 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 sitio de unión ribosomal (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 GPP2 en pAH42 para crear pAH45.

El sitio de unión al ribosoma en el extremo 5' de GPP2 se modificó como sigue. Un enlazador BamHI-RBS-SpeI, obtenido al renaturalizar cebadores sintéticos GATCCAGGAAACAGA (SEQ ID NO:24) con CTAGTCTGTTTCCTG (SEQ ID NO:25) al fragmento XbaI-PstI procedente de pAH19 que contenía el gen GPP2, se insertó en el sitio BamHI-PstI de pAH40 para crear pAH48. El plásmido pAH48 contiene el gen DAR1, el RBS modificado y el gen GPP2 en la orientación correcta para la expresión a partir del promotor trc de pTrc99A (Pharmacia, Piscataway, NJ).

## ES 2 345 469 T3

### *Transformación de E. coli*

Todos los plásmidos aquí descritos se transformaron en DH5 $\alpha$  de *E. coli* utilizando técnicas de biología molecular estándares. Los transformantes se verificaron mediante su modelo RFLP de ADN.

5

#### Ejemplo 1

##### *Producción de glicerol a partir de E. coli transformada con el gen G3PDH*

#### 10 Medios

Se utilizaron medios sintéticos para la producción anaerobia o aerobia de glicerol utilizando células de *E. coli* transformadas con pDAR1A. Los medios contenían, por litro, 6,0 g de Na<sub>2</sub>HPO<sub>4</sub>, 3,0 g de KH<sub>2</sub>PO<sub>4</sub>, 1,0 g de NH<sub>4</sub>Cl, 0,5 g de NaCl, 1 mL de MgSO<sub>4</sub> al 20%.7H<sub>2</sub>O, 8,0 g de glucosa, 40 mg de casaminoácidos, 0,5 ml de hidrocloreuro de tiamina al 1%. 100 mg de ampicilina.

15

#### *Condiciones de Crecimiento*

La cepa AA200 que alberga pDAR1A o el vector pTrec99A se hizo crecer en condiciones aerobias en 50 mL de medios, sacudiendo a 250 rpm en matraces de 250 mL a 37°C. A A<sub>600</sub> 0,2-0,3 se añadió isopropiltio- $\beta$ -D-galactósido hasta una concentración final de 1 mM y la incubación continuó durante 48 h. Para el crecimiento anaerobio, se utilizaron muestras de crecimiento de células inducidas para llenar tubos Falcon n°2054 que estaban cerrados mediante tapón y se mezclaron suavemente mediante rotación a 37°C durante 48 h. La producción de glicerol se determinó mediante análisis por HPLC de los sobrenadantes del cultivo. La cepa pDAR1A/AA200 producía 0,38 g/L de glicerol después de 48 h en condiciones anaerobias, y 0,48 g/L en condiciones aerobias.

20

#### Ejemplo 2

##### *Producción de glicerol a partir de E. coli transformada con el gen G3P fosfatasa (GPP2)*

30

#### Medios

Se utilizaron medios phoA sintéticos en matraces de sacudimiento para demostrar el incremento de glicerol mediante la expresión de GPP2 en *E. coli*. El medio phoA contenía por litro: Amisoy, 12 g; sulfato amónico, 0,62 g; MOPS, 10,5 g; citrato de Na, 1,2 g; NaOH (1 M), 10 mL; MgSO<sub>4</sub> 1 M, 12 mL; elementos traza 100X, 12 mL; glucosa al 50%, 10 mL; tiamina al 1%, 10 mL; 100 mg/mL de L-prolina, 10 mL; FeCl<sub>3</sub> 2,5 mM, 5 mL; tampón fosfatos mixtos, 2 mL (5 mL de NaH<sub>2</sub>PO<sub>4</sub> 0,2 M + 9 mL de K<sub>2</sub>HPO<sub>4</sub> 0,2 M), y pH hasta 7,0. Los elementos traza 100X para el medio phoA /L contenían: ZnSO<sub>4</sub>.7 H<sub>2</sub>O, 0,58 g; MnSO<sub>4</sub>.H<sub>2</sub>O, 0,34 g; CuSO<sub>4</sub>.5 H<sub>2</sub>O, 0,49 g; CoCl<sub>2</sub>.6 H<sub>2</sub>O, 0,47 g; H<sub>3</sub>BO<sub>3</sub>, 0,12 g, NaMoO<sub>4</sub>.2 H<sub>2</sub>O, 0,48 g.

40

#### *Experimentos en matraces de sacudimiento*

Las cepas pAH21/DH5 $\alpha$  (que contenían el gen GPP2) y pPHOX21/DH5 $\alpha$  (control) se hicieron crecer en 45 mL de medios (medios phoA, 50 ug/mL de carbenicilina y 1 ug/mL de vitamina B<sub>12</sub>) en un matraz de sacudimiento de 250 mL a 37°C. Los cultivos se hicieron crecer en condiciones aerobias (sacudimiento a 250 rpm) durante 24 h. La producción de glicerol se determinó mediante análisis por HPLC del sobrenadante del cultivo. pAH21/DH5 $\alpha$  producía 0,2 g/L de glicerol al cabo de 24 h.

45

#### Ejemplo 3

50

##### *Producción de glicerol a partir de D-glucosa utilizando E. coli recombinante que contenía tanto GPP2 como DARI*

El crecimiento para la demostración de la producción incrementada de glicerol por parte de pAH43 que contenía DH5 $\alpha$  de *E. coli* prosigue en condiciones aerobias a 37°C en cultivos en matraces de sacudimiento (matraces Erlenmeyer, volumen de líquido 1/5 del volumen total).

55

Los cultivos en medios mínimos/matraces de sacudimiento con glucosa al 1% se inician mediante inoculación a partir de cultivo durante una noche LB/glucosa al 1% con selección del antibiótico. Los medios mínimos son: medios definidos esterilizados con filtro, pH final 6,8 (HCl), contenían por litro: 12,6 g de (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 13,7 g de K<sub>2</sub>HPO<sub>4</sub>, 0,2 g de extracto de levaduras (Difco), 1 g de NaHCO<sub>3</sub>, 5 mg de vitamina B<sub>12</sub>, 5 mL de Solución de Elementos Traza Modificada de Balch (cuya composición se puede encontrar en *Methods for General and Molecular Bacteriology* (P. Gerhardt *et al.*, eds, pág. 158, American Society for Microbiology, Washington, DC (1994)). Los matraces de sacudimiento se incuban a 37°C con agitación intensa durante una noche, tras lo cual se toman muestras para análisis por GC del sobrenadante. El pAH43/DH5 $\alpha$  mostraba una producción de glicerol de 3,8 g/L al cabo de 24 h.

60

65

## ES 2 345 469 T3

### Ejemplo 4

*Producción de glicerol a partir de D-glucosa utilizando E. coli recombinante que contenía tanto GPP2 como DAR1*

- 5 El Ejemplo 4 ilustra la producción de glucosa a partir de DH5 $\alpha$ /pAH48 recombinante de *E. coli*, que contenía los genes tanto GPP2 como DAR1.

La cepa DH5 $\alpha$ /pAH48 se construyó según se describe antes en los Métodos Generales.

#### 10 *Pre-Cultivo*

DH5 $\alpha$ /pAH48 se pre-cultivaron para la siembra en una operación de fermentación. Componentes y protocolos para el pre-cultivo se listan más abajo.

#### 15 *Medios de Pre-Cultivo*

	KH <sub>2</sub> PO <sub>4</sub>	30,0 g/L
	Ácido cítrico	2,0 g/L
20	MgSO <sub>4</sub> ·7H <sub>2</sub> O	2,0 g/L
	H <sub>2</sub> SO <sub>4</sub> al 98%	2,0 mL/L
	Citrato de amonio férrico	0,3 g/L
25	CaCl <sub>2</sub> ·2H <sub>2</sub> O	0,2 g/L
	Extracto de levaduras	5,0 g/L
	Metales Traza	5,0 mL/L
30	Glucosa	10,0 g/L
	Carbenicilina	100,0 mg/L

- 35 Los componentes de los medios de arriba se mezclaron juntos y el pH se ajustó a 6,8 con NH<sub>4</sub>OH. Los medios se esterilizaron luego con el filtro.

Metales trazas se utilizaron de acuerdo con la receta siguiente:

40	Ácido cítrico, monohidrato	4,0 g/L
	MgSO <sub>4</sub> ·7H <sub>2</sub> O	3,0 g/L
45	MnSO <sub>4</sub> ·H <sub>2</sub> O	0,5 g/L
	NaCl	1,0 g/L
	FeSO <sub>4</sub> ·7H <sub>2</sub> O	0,1 g/L
50	CoCl <sub>2</sub> ·6H <sub>2</sub> O	0,1 g/L
	CaCl <sub>2</sub>	0,1 g/L
	ZnSO <sub>4</sub> ·7H <sub>2</sub> O	0,1 g/L
55	CuSO <sub>4</sub> ·5 H <sub>2</sub> O	10 mg/L
	AlK(SO <sub>4</sub> ) <sub>2</sub> ·12 H <sub>2</sub> O	10 mg/L
	H <sub>3</sub> BO <sub>3</sub>	10 mg/L
60	Na <sub>2</sub> MoO <sub>4</sub> ·2H <sub>2</sub> O	10 mg/L
	NiSP4·6H <sub>2</sub> O	10 mg/L
	Na <sub>2</sub> SeO <sub>3</sub>	10 mg/L
65	Na <sub>2</sub> WO <sub>4</sub> ·2H <sub>2</sub> O	10 mg/L

## ES 2 345 469 T3

Los cultivos se iniciaron a partir de cultivo de siembra, inoculado a partir de 50  $\mu$ L de material congelado (glicerol al 15% en forma de crioprotector) a 600 mL de medio en un matraz Erlenmeyer de 2-L. Los cultivos se hicieron crecer a 30°C en un sacudidor a 250 rpm durante aproximadamente 12 h y luego se utilizaron para sembrar el fermentador.

### 5 *Crecimiento en fermentación*

#### *Recipiente*

Fermentador con tanque agitado de 15-L.

10

#### *Medio*

	$\text{KH}_2\text{PO}_4$	6,8 g/L
15	Ácido cítrico	2,0 g/L
	$\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$	2,0 g/L
20	$\text{H}_2\text{SO}_4$ al 98%	2,0 mL/L
	Citrato de amonio férrico	0,3 g/L
	$\text{CaCl}_2 \cdot 2\text{H}_2\text{O}$	0,2 g/L
25	antiespumante Mazu DF204	1,0 mL/L

Los componentes anteriores se esterilizaron juntos en el recipiente fermentador. El pH se elevó hasta 6,7 con  $\text{NH}_4\text{OH}$ . Se añadieron extracto de levaduras (5 g/L) y solución de metales trazas (5 mL/L) en condiciones asépticas a partir de soluciones patrón esterilizadas con filtro. Se añadió glucosa a partir de una alimentación del 60% para dar una concentración final de 10 g/L. Se añadió carbenicilina a 100 mg/L. El volumen tras la inoculación era de 6 L.

30

#### *Condiciones Medioambientales para la Fermentación*

La temperatura se controló a 36°C y el caudal de aire se controló a 6 litros estándar por minuto. La retropresión se controló a 0,5 bar. El agitador se ajustó a 350 rpm. Se utilizó amoniaco acuoso para el control del pH a 6,7. La velocidad de alimentación de glucosa (glucosa monohidrato al 60%) se controló para mantener la glucosa en exceso.

35

#### *Resultados*

40

Los resultados de la operación de fermentación se dan en la Tabla 1.

45

(Tabla pasa a página siguiente)

50

55

60

65

# ES 2 345 469 T3

TABLA 1

5	EFT (h)	DD550 (UA)	[Glucosa] (g/L)	[Glicerol] (g/L)	Total de glucosa alimentada (g)	Total de glicerol producido (g)
	0	0,8	9.3		25	
10	6	4,7	4,0	2,0	49	14
	8	5,4	0	3.6	71	25
15	10	6,7	0,0	4,7	116	33
	12	7.4	2,1	7,0	157	49
	14.2	10.4	0,3	10,0	230	70
20	16.2	18.1	9.7	15,5	259	106
	18.2	12.4	14.5		305	
25	20.2	11,8	17.4	17.7	353	119
	22.2	11.0	12,6		382	
	24.2	10,8	6,5	26.6	404	178
30	26,2	10,9	6.8		442	
	28.2	10.4	10,3	31.5	463	216
	30.2	10,2	13.1	30.4	493	213
35	32.2	10.1	8,1	28.2	512	196
	34.2	10,2	3,5	33.4	530	223
40	36.2	10.1	5,8		548	
	38.2	9.8	5,1	36.1	512	233

45

50

55

60

65

**REIVINDICACIONES**

1. Un método para la producción de glicerol a partir de un microorganismo recombinante, que comprende:

(i) transformar una célula de *E. coli* con:

(a) una casete de expresión que comprende:

(a-i) un gen que codifica una enzima glicerol-3-fosfato deshidrogenasa NADH-dependiente o una enzima glicerol-3-fosfato deshidrogenasa NADH-dependiente; y

(a-ii) un gen que codifica una enzima glicerol-3-fosfato fosfatasa, o

(b) una primera casete de expresión que comprende un gen que codifica una enzima glicerol-3-fosfato deshidrogenasa NADH-dependiente o una enzima glicerol-3-fosfato deshidrogenasa NADH-dependiente; y una segunda casete de expresión que comprende un gen que codifica una enzima glicerol-3-fosfato fosfatasa;

(ii) cultivar la célula de *E.coli* de (i) transformada en presencia de al menos una fuente de carbono seleccionada del grupo que consiste en monosacáridos, oligosacáridos, polisacáridos y sustratos de un solo carbono, con lo que se produce glicerol; y

(iii) recuperar el glicerol producido en (ii).

2. Un método de acuerdo con la reivindicación 1, en el que la fuente de carbono es glucosa.

3. Un método de acuerdo con la reivindicación 1, en el que el gen que codifica una enzima glicerol-3-fosfato deshidrogenasa NADH-dependiente o una enzima glicerol-3-fosfato deshidrogenasa NADPH-dependiente codifica la secuencia de aminoácidos dada en SEQ ID NO:7, SEQ ID NO:8 o SEQ ID NO: 10.

4. Un método de acuerdo con la reivindicación 1, en el que el gen que codifica una enzima de glicerol-3-fosfato fosfatasa codifica la secuencia de aminoácidos proporcionada en SEQ ID NO 13 o SEQ ID NO:14.

5. Una célula hospedadora transformada de *E.coli*, que comprende:

(a) una casete de expresión que comprende:

(a-i) un gen que codifica una enzima glicerol-3-fosfato deshidrogenasa NADH-dependiente o una enzima glicerol-3-fosfato deshidrogenasa NADH-dependiente; y

(a-ii) un gen que codifica una enzima glicerol-3-fosfato fosfatasa; o

(b) una primera casete de expresión que comprende un gen que codifica una enzima glicerol-3-fosfato deshidrogenasa NADH-dependiente o una enzima glicerol-3-fosfato deshidrogenasa NADH-dependiente; y una segunda casete de expresión que comprende un gen que codifica una enzima glicerol-3-fosfato fosfatasa.

6. Una célula hospedadora de *E. coli* transformada de acuerdo con la reivindicación 5, en la que el gen que codifica una enzima de glicerol-3-fosfato fosfatasa codifica la secuencia de aminoácidos proporcionada en SEQ ID NO 13 o SEQ ID NO:14.

# ES 2 345 469 T3

## LISTA DE SECUENCIAS

### (1) INFORMACIÓN GENERAL:

- 5 (i) SOLICITANTE:
- (A) DENOMINACIÓN: E. I. DU PONT DE NEMOURS AND COMPANY
  - (B) CALLE: 1007 MARKET STREET
  - 10 (C) CIUDAD: WILMINGTON
  - (D) ESTADO: DELAWARE
  - (E) PAÍS: EE.UU.
  - (F) CÓDIGO POSTAL (ZIP): 19898
  - 15 (G) TELÉFONO: 302-892-8112
  - (H) TELEFAX: 302-773-0164
  - (I) TELEX: 6717325
- 20 (A) DESTINATARIO: GENENCOR INTERNATIONAL, INC.
- (B) CALLE: 4 CAMBRIDGE PLACE  
1870 SOUTH WINTON ROAD
  - 25 (C) CIUDAD: ROCHESTER
  - (D) ESTADO: NUEVA YORK
  - (E) PAÍS: EE.UU.
  - (F) CÓDIGO POSTAL (ZIP): 14618
- 30 (ii) TÍTULO DE LA INVENCIÓN: MÉTODO PARA LA PRODUCCIÓN DE GLICEROL MEDIANTE ORGANISMOS RECOMBINANTES
- 35 (iii) NÚMERO DE SECUENCIAS: 25
- (iv) FORMA LEGIBLE POR ORDENADOR:
- (A) TIPO MEDIO: DISQUETE DE 3,5 PULGADAS
  - (B) ORDENADOR: IBM PC COMPATIBLE
  - 40 (C) SISTEMA OPERATIVO: MICROSOFT WORD FOR WINDOWS 95
  - (D) PROGRAMA INFORMÁTICO: MICROSOFT WORD VERSION 7.0A
- (v) DATOS DE LA SOLICITUD ACTUAL:
- 45 (A) NÚMERO DE SOLICITUD:
  - (B) FECHA DE PRESENTACIÓN:
  - (C) CLASIFICACIÓN:
- 50 (vi) DATOS DE LA SOLICITUD ANTERIOR:
- (A) NÚMERO DE SOLICITUD: 60/03602
  - (B) FECHA DE PRESENTACIÓN: 13 DE NOVIEMBRE, 1996
  - 55 (C) CLASIFICACIÓN:
- (vii) INFORMACIÓN DEL APODERADO/AGENTE:
- (A) DENOMINACIÓN: FLOYD, LINDA AXAMETHY
  - 60 (B) NÚMERO DE REGISTRO: 33.692
  - (C) NÚMERO DE REFERENCIA/ETIQUETA: CR-9981-P1

### (2) INFORMACIÓN PARA LA SEQ. ID. NO:1:

- 65 (i) CARACTERÍSTICAS DE LA SECUENCIA:
- (A) LONGITUD: 1380 pares de bases
  - (B) TIPO: ácido nucleico

## ES 2 345 469 T3

(C) TIPO DE CADENA: sencilla

(D) CONFIGURACIÓN: lineal

(ii) TIPO DE MOLÉCULA: ADN (genómico)

(xi) DESCRIPCIÓN DE LA SECUENCIA: SEQ. ID. NO:1:

5	CTTTAATTTT CTTTATCTT ACTCTCCTAC ATAAGACATC AAGAAACAAT TGTATATTGT	60
10	ACACCCCCC CCTCCACAAA CACAAATATT GATAATATAA AGATGTCTGC TGCTGCTGAT	120
	AGATTAACT TAACTTCCGG CCACTTGAAT GCTGGTAGAA AGAGAAGTTC CTCTTCTGTT	180
15	TCTTTGAAGG CTGCCGAAA GCCTTTCAAG GTTACTGTGA TTGGATCTGG TAACTGGGGT	240
	ACTACTATTG CCAAGGTGGT TGCCGAAAAT TGTAAGGGAT ACCCAGAAGT TTTGCTCCA	300
	ATAGTACAAA TGTGGGTGTT CGAAGAAGAG ATCAATGGTG AAAAATTGAC TGAAATCATA	360
20	AATACTAGAC ATCAAAACGT GAAATACTTG CCTGGCATCA CTCTACCCGA CAATTTGGTT	420
	GCTAATCCAG ACTTGATTGA TTCAGTCAAG GATGTCGACA TCATCGTTTT CAACATTCCA	480
25	CATCAATTTT TGCCCCGTAT CTGTAGCCAA TTGAAAGGTC ATGTTGATTC ACACGTCAGA	540
	GCTATCTCCT GTCTAAAGGG TTTTGAAGTT GGTGCTAAAG GTGTCCAAT TCTATCTCT	600
	TACATCACTG AGGAACTAGG TATTCAATGT GGTGCTCTAT CTGGTGCTAA CATTGCCACC	660
30	GAAGTCGCTC AAGAACACTG GTCTGAAACA ACAGTTGCTT ACCACATTCC AAAGGATTTT	720
	AGAGGCGAGG GCAAGGACGT CGACCATAAG GTTCTAAAGG CCTTGTCCA CAGACCTTAC	780
35	TCCACGTTA GTGTCATCGA AGATGTTGCT GGTATCTCCA TCTGTGGTGC TTTGAAGAAC	840
	GTTGTTGCCT TAGGTTGTGG TTTGCTCGAA GGTCTAGGCT GGGGTAACAA CGCTTCTGCT	900
	GCCATCCAAA GAGTCGGTTT GGGTGAGATC ATCAGATTCG GTCAAATGTT TTTCCAGAA	960
40	TCTAGAGAAG AAACATACTA CCAAGAGTCT GCTGGTGTG CTGATTTGAT CACCACCTGC	1020
	GCTGGTGGTA GAAACGTCAA GGTTGCTAGG CTAATGGCTA CTTCTGGTAA GGACGCCTGG	1080
45	GAATGTGAAA AGGAGTTGTT GAATGGCCAA TCCGCTCAAG GTTTAATTAC CTGCAAAGAA	1140
	GTTACGAAT GGTTGGAAC ATGTGGCTCT GTCGAAGACT TCCCATTATT TGAAGCCGTA	1200
	TACCAAATCG TTTACAACAA CTACCCAATG AAGAACCTGC CGGACATGAT TGAAGAATTA	1260
50	GATCTACATG AAGATTAGAT TTATTGGAGA AAGATAACAT ATCATACTTC CCCCACTTTT	1320
	TTGAGGCTC TTCTATATCA TATTCATAAA TTAGCATTAT GTCATTTCTC ATAACTACTT	1380

55 (2) INFORMACIÓN PARA LA SEQ. ID. NO:2:

(i) CARACTERÍSTICAS DE LA SECUENCIA:

(A) LONGITUD: 2946 pares de bases

(B) TIPO: ácido nucleico

(C) TIPO DE CADENA: sencilla

(D) CONFIGURACIÓN: lineal

(ii) TIPO DE MOLÉCULA: ADN (genómico)

# ES 2 345 469 T3

(xi) DESCRIPCIÓN DE LA SECUENCIA: SEQ. ID. NO:2:

5	GAATTCGAGC CTGAAGTGCT GATTACCTTC AGGTAGACTT CATCTTGACC CATCAACCCC	60
	AGCGTCAATC CTGCAAATAC ACCACCCAGC AGCACTAGGA TGATAGAGAT AATATAGTAC	120
	GTGGTAACGC TTGCCTCATC ACCTACGCTA TGGCCGGAAT CGGCAACATC CCTAGAATTG	180
10	AGTACGTGTG ATCCGGATAA CAACGGCAGT GAATATATCT TCGGTATCGT AAAGATGTGA	240
	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
20	GAAACCAAAA GAATGAAGAA AGAAAACAAA TACTAGCCCT AACCTGACT TCGTTTCTAT	540
	GATAATACCC TGCTTTAATG AACGGTATGC CCTAGGGTAT ATCTCACTCT GTACGTTACA	600
	AACTCCGGTT ATTTTATCGG AACATCCGAG CACCCGCGCC TTCCTCAACC CAGGCACCGC	660
25	CCCAGGTAAC CGTGCGCGAT GAGCTAATCC TGAGCCATCA CCCACCCAC CCGTTGATGA	720
	CAGCAATTGC GGAGGGCGAA AATAAACTG GAGCAAGGAA TTACCATCAC CGTCACCATC	780
30	ACCATCATAT CGCCTTAGCC TCTAGCCATA GCCATCATGC AAGCGTGTAT CTTCTAAGAT	840
	TCAGTCATCA TCATTACCGA GTTTGTTTTC CTTACATGA TGAAGAAGGT TTGAGTATGC	900
	TCGAAACAAT AAGACGACGA TGGCTCTGCC ATTGGTTATA TTACGCTTTT GCGGCGAGGT	960
35	GCCGATGGGT TGCTGAGGGG AAGAGTGFTT AGCTTACGGA CCTATTGCCA TTGTTATTCC	1020
	GATTAATCTA TTGTTGAGCA GCTCTTCTCT ACCCTGTCAT TCTAGTATTT TTTTTTTTTT	1080
40	TTTTTGTTTT TACTTTTTTT TCTTCTTGCC TTTTTTCTT GTTACTTTTT TTCTAGTTTT	1140
	TTTTCTTCC ACTAAGCTTT TTCCTTGATT TATCCTTGGG TTCTTCTTTC TACTCCTTTA	1200
	GATTTTTTTT TTATATATTA ATTTTAAAGT TTATGTATTT TGGTAGATTC AATTCTCTTT	1260
45	CCCTTTCCTT TTCCTTCGCT CCCCTTCCTT ATCAATGCTT GCTGTCAGAA GATTAACAAG	1320
	ATACACATTC CTTAAGCGAA CGCATCCGGT GTTATATACT CGTCGTGCAT ATAAAATTTT	1380

50

55

60

65

## ES 2 345 469 T3

	GCCTTCAAGA TCTACTTTCC TAAGAAGATC ATTATTACAA ACACAACCTGC ACTCAAAGAT	1440
	GACTGCTCAT ACTAATATCA AACAGCACAA ACACTGTCAT GAGGACCATC CTATCAGAAG	1500
5	ATCGGACTCT GCCGTGICAA TTGTACATTT GAAACGTGCG CCCTTCAAGG TTACAGTGAT	1560
	TGGTTCTGGT AACTGGGGGA CCACCATCGC CAAAGTCATT GCGGAAAACA CAGAATTGCA	1620
10	TTCCCATATC TTCGAGCCAG AGGTGAGAAT GTGGGTTTTT GATGAAAAGA TCGGCGACGA	1680
	AAATCTGACG GATATCATAA ATACAAGACA CCAGAACGTT AAATATCTAC CCAATATTGA	1740
	CCTGCCCAT AATCTAGTGG CCGATCCTGA TCTTTTACAC TCCATCAAGG GTGCTGACAT	1800
15	CCTTGTTTTT AACATCCCTC ATCAATTTTT ACCAAACATA GTCAAACAAT TGCAAGGCCA	1860
	CGTGGCCCT CATGTAAGGG CCATCTCGTG TCTAAAAGGG TTCGAGTTGG GCTCCAAGGG	1920
20	TGTGCAATTG CTATCCTCCT AIGTTACTGA TGAGTTAGGA ATCCAATGTG GCGCACTATC	1980
	TGGTGCAAAC TTGGCACCGG AAGTGGCCAA GGAGCATTGG TCCGAAACCA CCGTGGCTTA	2040
	CCAACTACCA AAGGATTATC AAGGTGATGG CAAGGATGTA GATCATAAGA TTTTGAAATT	2100
25	GCTGTTCCAC AGACCTTACT TCCACGTCAA TGTCATCGAT GATGTTGCTG GTATATCCAT	2160
	TGCCGGTGCC TTGAAGAACG TCGTGGCACT TGCATGTGGT TTCGTAGAAG GTATGGGATG	2220
30	GGGTAACAAT GCCTCCGCAG CCATTCAAAG GCTGGGTTTA GGTGAAATTA TCAAGTTCGG	2280
	TAGAATGTTT TTCCAGAAT CCAAAGTCGA GACCTACTAT CAAGAATCCG CTGGTGTTC	2340
	AGATCTGATC ACCACCTGCT CAGGCGGTAG AAACGTCAAG GTTGCCACAT ACATGGCCAA	2400
35	GACCGGTAAG TCAGCCTTGG AAGCAGAAAA GGAATTGCTT AACGGTCAAT CCGCCAAGG	2460
	GATAATCACA TGCAGAGAAG TTCACGAGTG GCTACAAACA TGTGAGTTGA CCCAAGAATT	2520
40	CCCAATTATT CGAGGCAGTC TACCAGATAG TCTACAACAA CGTCCGCATG GAAGACCTAC	2580
	CGGAGATGAT TGAAGAGCTA GACATCGATG ACGAATAGAC ACTCTCCCC CCCCTCCCC	2640
	TCTGATCTTT CCTGTTGCCT CTTTTTCCCC CAACCAATTT ATCATTATAC ACAAGTTCTA	2700
45	CAACTACTAC TAGTAACATT ACTACAGTTA TTATAATTTT CTATTCTCTT TTTCTTTAAG	2760
	AATCTATCAT TAACGTTAAT TTCTATATAT ACATAACTAC CATTATACAC GCTATTATCG	2820
50	TTTACATATC ACATCACCGT TAATGAAAGA TACGACACCC TGTACACTAA CACAATTAAA	2880
	TAATCGCCAT AACCTTTTCT GTTATCTATA GCCCTTAAAG CTGTTTCTTC GAGCTTTTCA	2940
55	CTGCAG	2946

(2) INFORMACIÓN PARA LA SEQ. ID. NO:3:

(i) CARACTERÍSTICAS DE LA SECUENCIA:

- (A) LONGITUD: 3178 pares de bases
- (B) TIPO: ácido nucleico
- (C) TIPO DE CADENA: sencilla
- (D) CONFIGURACIÓN: lineal

(ii) TIPO DE MOLÉCULA: ADN (genómico)

# ES 2 345 469 T3

(xi) DESCRIPCIÓN DE LA SECUENCIA: SEQ. ID. NO:3:

5	CTGCAGAACT TCGTCTGCTC TGTGCCCATC CTCGCGGTTA GAAAGAAGCT GAATTGTTTC	60
	ATGCGCAAGG GCATCAGCGA GTGACCAATA ATCACTGCAC TAATTCCTTT TTAGCAACAC	120
	ATACTTATAT ACAGCACCAG ACCTTATGTC TTTTCTCTGC TCCGATACGT TATCCCACCC	180
10	AACTTTTATT TCAGTTTTGG CAGGGGAAAT TTCACAACCC CGCACGCTAA AAATCGTATT	240
	TAAACTTAAA AGAGAACAGC CACAAATAGG GAACTTTGGT CTAAACGAAG GACTCTCCCT	300
	CCCTTATCTT GACCGTGCTA TTGCCATCAC TGCTACAAGA CTAAATACGT ACTAATATAT	360
15	GTTTTCGGTA ACGAGAAGAA GAGCTGCCGG TGCAGCTGCT GCCATGGCCA CAGCCACGGG	420
	GACGCTGTAC TGGATGACTA GCCAAGGTGA TAGGCCGTTA GTGCACAATG ACCCGAGCTA	480
20	CATGGTGCAA TCCCCACCG CCGCTCCACC GGCAGGTCTC TAGACGAGAC CTGCTGGACC	540
	GTCTGGACAA GACGCATCAA TTCGACGTGT TGATCATCGG TGGCGGGGCC ACGGGGACAG	600
	GATGTGCCCT AGATGCTGCG ACCAGGGGAC TCAATGTGGC CCTTGTGAA AAGGGGGATT	660
25	TTGCCTCGGG AACGTCGTCC AAATCTACCA AGATGATTCA CGGTGGGGTG CGGTACTTAG	720
	AGAAGGCCTT CTGGGAGTTC TCCAAGGCAC AACTGGATCT GGTTCATCGAG GCACTCAACG	780
30	AGCGTAAACA TCTTATCAAC ACTGCCCTC ACCTGTGCAC GGTGCTACCA ATTCTGATCC	840
	CCATCTACAG CACCTGGCAG GTCCCGTACA TCTATATGGG CTGTAAATTC TACGATTTCT	900
	TTGGCGGTTT CCAAACCTTG AAAAAATCAT ACCTACTGTC CAAATCCGCC ACCGTGGAGA	960
35	AGGCTCCCAT GCTTACCACA GACAATTTAA AGGCCTCGCT TGTGTACCAT GATGGGTCCT	1020
	TTAACGACTC GCGTTTGAAC GCCACTTTAG CCATCACGGG TGTGGAGAAC GGCCTACCG	1080
40	TCTTGATCTA TGTCGAGGTA CAAAAATTGA TCAAAGACCC AACTTCTGGT AAGGTTATCG	1140
	GTGCCGAGGC CCGGGACGTT GAGACTAATG AGCTTGTGAG AATCAACGCT AAATGTGTGG	1200
	TCAATGCCAC GGGCCCATAC AGTGACGCCA TTTTGAAAT GGACCGCAAC CCATCCGGTC	1260
45	TGCCGGACTC CCCGCTAAAC GACAACTCCA AGATCAAGTC GACTTTCAAT CAAATCTCCG	1320
	TCATGGACCC GAAAATGGTC ATCCCATCTA TTGGCGTTCA CATCGTATTG CCCTCTTTTT	1380
50	ACTCCCCGAA GGATATGGCT TTGTTGGACG TCAGAACCTC TGATGGCAGA GTGATGTTCT	1440
	TTTTACCTTG GCAGGGCAA GTCCTTGCCG GCACCACAGA CATCCCCTA AAGCAAGTCC	1500

55

60

65

## ES 2 345 469 T3

	CAGAAAACCC TATGCCTACA GAGGCTGATA TTCAAGATAT CTTGAAAGAA CTACAGCACT	1560
	ATATCGAATT CCCCCTGAAA AGAGAAGACG TGCTAAGTGC ATGGGCTGGT GTCAGACCTT	1620
5	TGGTCAGAGA TCCACGTACA ATCCCCGCAG ACGGGAAGAA GGGCTCTGCC ACTCAGGGCG	1680
	TGGTAAGATC CCACTTCTTG TTCACTTCGG ATAATGGCCT AATTACTATT GCAGGTGGTA	1740
10	AATGGACTAC TTACAGACAA ATGGCTGAGG AAACAGTCGA CAAAGTTGTC GAAGTTGGCG	1800
	GATTCCACAA CCTGAAACCT TGTACACAAA GAGATATTAA GCTTGCTGGT GCAGAAGAAT	1860
	GGACGCAAAA CTATGTGGCT TTATTGGCTC AAAACTACCA TTTATCATCA AAAATGTCCA	1920
15	ACTACTTGGT TCAAACTAC GGAACCCGTT CCTCTATCAT TTGCGAATTT TTCAAAGAAT	1980
	CCATGGAAAA TAAACTGCCT TTGTCTTAG CCGACAAGGA AAATAACGTA ATCTACTCTA	2040
20	GCGAGGAGAA CAACTTGGTC AATTTTGATA CTTTCAGATA TCCATTACACA ATCGGTGAGT	2100
	TAAAGTATTC CATGCAGTAC GAATATTGTA GAACTCCCTT GGACTTCCTT TTAAGAAGAA	2160
	CAAGATTCGC CTTCTGGAC GCCAAGGAAG CTTTGAATGC CGTGCATGCC ACCGTCAAAG	2220
25	TTATGGGTGA TGAGTTCAAT TGGTCGGAGA AAAAGAGGCA GTGGGAACCT GAAAAAACTG	2280
	TGAACTTCAT CCAAGGACGT TTCGGTGTCT AAATCGATCA TGATAGTTAA GGGTGACAAA	2340
30	GATAACATTC ACAAGAGTAA TAATAATGGT AATGATGATA ATAATAATAA TGATAGTAAT	2400
	AACAATAATA ATAATGGTGG TAATGGCAAT GAAATCGCTA TTATTACCTA TTTTCCTTAA	2460
	TGGAAGAGTT AAAGTAAACT AAAAAAACTA CAAAAATATA TGAAGAAAAA AAAAAAAGA	2520
35	GGTAATAGAC TCTACTACTA CAATTGATCT TCAAATTATG ACCTTCCTAG TGTTTATATT	2580
	CTATTTCCAA TACATAATAT AATCTATATA ATCATTGCTG GTAGACTTCC GTTTAATAT	2640
40	CGTTTTAATT ATCCCCTTTA TCTCTAGTCT AGTTTTATCA TAAAATATAG AAACACTAAA	2700
	TAATATTCTT CAAACGGTCC TGGTGCATAC GCAATACATA TTTATGGTGC AAAAAAAAAA	2760
	ATGGAAAATT TTGCTAGTCA TAAACCCTTT CATAAACAA TACGTAGACA TCGCTACTTG	2820
45	AAATTTTCAA GTTTTTATCA GATCCATGTT TCCTATCTGC CTTGACAACC TCATCGTCGA	2880
	AATAGTACCA TTTAGAACGC CCAATATTCA CATTGTGTTC AAGGTCITTA TTCACCAGTG	2940
50	ACGTGTAATG GCCATGATTA ATGTGCCTGT ATGGTTAACC ACTCCAAATA GCTTATATTT	3000
	CATAGTGTC TGTTTTTTCA ATATAATGTT TAGTATCAAT GGATATGTTA CGACGGTGTT	3060
	ATTTTTCTTG GTCAAATCGT AATAAAATCT CGATAAATGG ATGACTAAGA TTTTGGTAA	3120
55	AGTTACAAAA TTTATCGTTT TCACTGTTGT CAATTTTTTG TTCTTGTAA CACTCGAG	3178

(2) INFORMACIÓN PARA LA SEQ. ID. NO:4:

- 60 (i) CARACTERÍSTICAS DE LA SECUENCIA:
- (A) LONGITUD: 816 pares de bases
  - (B) TIPO: ácido nucleico
  - 65 (C) TIPO DE CADENA: sencilla
  - (D) CONFIGURACIÓN: lineal

## ES 2 345 469 T3

(ii) TIPO DE MOLÉCULA: ADN (genómico)

(xi) DESCRIPCIÓN DE LA SECUENCIA: SEQ. ID. NO:4:

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

30 (2) INFORMACIÓN PARA LA SEQ. ID. NO:5:

(i) CARACTERÍSTICAS DE LA SECUENCIA:

- 35 (A) LONGITUD: 753 pares de bases
- (B) TIPO: ácido nucleico
- (C) TIPO DE CADENA: sencilla
- (D) CONFIGURACIÓN: lineal

40 (ii) TIPO DE MOLÉCULA: ADN (genómico)

(xi) DESCRIPCIÓN DE LA SECUENCIA: SEQ. ID. NO:5:

45	ATGGGATTGA CTAATAAACC TCTATCTTTG AAAGTTAACG CCGCTTTGTT CGACGTCGAC	60
	GGTACCATTA TCATCTCTCA ACCAGCCATT GCTGCATTCT GGAGGGATTT CGGTAAGGAC	120
	AAACCTTATT TCGATGCTGA ACACGTTATC CAAGTCTCGC ATGGTTGGAG AACGTTTGAT	180
50	GCCATTGCTA AGTTCGCTCC AGACTTTGCC AATGAAGAGT ATGTTAACAA ATTAGAAGCT	240
	GAAATCCCGG TCAAGTACGG TGAAAAATCC ATTGAAGTCC CAGGTGCAGT TAAGCTGTGC	300
	AACGCTTTGA ACGCTCTACC AAAAGAGAAA TGGGCTGTGG CAACTTCCGG TACCCGTGAT	360
55	ATGGCACAAA AATGGTTCGA GCATCTGGGA ATCAGGAGAC CAAAGTACTT CATTACCGCT	420
	AATGATGTCA AACAGGGTAA GCCTCATCCA GAACCATATC TGAAGGGCAG GAATGGCTTA	480
	GGATATCCGA TCAATGAGCA AGACCCTTCC AAATCTAAGG TAGTAGTATT TGAAGACGCT	540
60	CCAGCAGGTA TTGCCGCCGG AAAAGCCGCC GGTTGTAAGA TCATTGGTAT TGCCACTACT	600
	TTCGACTTGG ACTTCCTAAA GGAAAAAGGC TGTGACATCA TTGTCAAAAA CCACGAATCC	660
65	ATCAGAGTTG GCGGCTACAA TGCCGAAACA GACGAAGTTG AATTCATTTT TGACGACTAC	720
	TTATATGCTA AGGACGATCT GTTGAAATGG TAA	753

## ES 2 345 469 T3

(2) INFORMACIÓN PARA LA SEQ. ID. NO:6:

(i) CARACTERÍSTICAS DE LA SECUENCIA:

(A) LONGITUD: 2520 pares de bases

(B) TIPO: ácido nucleico

(C) TIPO DE CADENA: sencilla

(D) CONFIGURACIÓN: lineal

(ii) TIPO DE MOLÉCULA: ADN (genómico)

(xi) DESCRIPCIÓN DE LA SECUENCIA: SEQ. ID. NO:6:

15	<b>TGTATTGGCC ACGATAACCA CCCTTTGTAT ACTGTTTTTG TTTTTCACAT GGTAATAAAC</b>	60
	<b>GACTTTTATT AAACAACGTA TGTAATAACA TAACAAGAAT CTACCCATAC AGGCCATTTC</b>	120
20	<b>GTAATTCTTC TCTTCTAATT GGAGTAAAAC CATCAATTAA AGGGTGTGGA GTAGCATAGT</b>	180
	<b>GAGGGGCTGA CTGCATTGAC AAAAAAATTG AAAAAAAAAA AGGAAAAGGA AAGGAAAAAA</b>	240
	<b>AGACAGCCAA GACTTTTAGA ACGGATAAGG TGTAATAAAA TGTGGGGGGA TGCCTGTTCT</b>	300
25	<b>CGAACCATAT AAAATATACC ATGTGGTTTG AGTTGTGGCC GGAACATAC AAATAGTTAT</b>	360
	<b>ATGTTTCCCT CTCTCTCCG ACTTGTAGTA TTCTCAAAC GTTACATATT CCGATCAAGC</b>	420
30	<b>CAGCGCCTTT ACACTAGTTT AAAACAAGAA CAGAGCCGTA TGTCCAAAAT AATGGAAGAT</b>	480
	<b>TTACGAAGTG ACTACGTCCC GCTTATCGCC AGTATTGATG TAGGAACGAC CTCATCCAGA</b>	540
	<b>TGCATTCTGT TCAACAGATG GGGCCAGGAC GTTTCAAAC ACCAAATTGA ATATTCAACT</b>	600
35	<b>TCAGCATCGA AGGGCAAGAT TGGGGTGTCT GGCCTAAGGA GACCCTCTAC AGCCCCAGCT</b>	660
	<b>CGTGAACAC CAAACGCCGG TGACATCAAA ACCAGCGGAA AGCCCATCTT TTCTGCAGAA</b>	720
40	<b>GGCTATGCCA TTCAAGAAAC CAAATTCCTA AAAATCGAGG AATTGGACTT GGACTTCCAT</b>	780

45

50

55

60

65

ES 2 345 469 T3

AACGAACCCA CGTTGAAGTT CCCCAAACCG GGTGGGTTG AGTGCCATCC GCAGAAATTA 840  
 CTGGTGAACG TCGTCCAATG CCTTGCCTCA AGTTTGCTCT CTCTGCAGAC TATCAACAGC 900  
 5 GAACGTGTAG CAAACGGTCT CCCACCTTAC AAGGTAATAT GCATGGGTAT AGCAAACATG 960  
 AGAGAAACCA CAATTCTGTG GTCCCGCCGC ACAGGAAAAC CAATTGTTAA CTACGGTATT 1020  
 10 GTTTGAACG ACACCAGAAC GATCAAAATC GTTAGAGACA AATGGCAAAA CACTAGCGTC 1080  
 GATAGGCAAC TGCAGCTTAG ACAGAAGACT GGATTGCCAT TGCTCTCCAC GTATTTCTCC 1140  
 TGTTCCAAGC TGCCTGGTT CCTCGACAAT GAGCCTCTGT GTACCAAGGC GTATGAGGAG 1200  
 15 AACGACCTGA TGTTGGGCAC TGTGGACACA TGGCTGATTT ACCAATTAAC TAAACAAAAG 1260  
 GCGTTCGTTT CTGACGTAAC CAACGCTTCC AGAACTGGAT TTATGAACCT CTCCACTTTA 1320  
 20 AAGTACGACA ACGAGTTGCT GGAATTTTGG GGTATTGACA AGAACCTGAT TCACATGCCC 1380  
 GAAATTGTGT CCTCATCTCA ATACTACGGT GACTTTGGCA TTCCTGATTG GATAATGGAA 1440  
 AAGCTACACG ATTCGCCAAA AACAGTACTG CGAGATCTAG TCAAGAGAAA CCTGCCATA 1500  
 25 CAGGGCTGTC TGGGCGACCA AAGCGCATCC ATGGTGGGGC AACTCGCTTA CAAACCCGGT 1560  
 GCTGCAAAAT GTACTTATGG TACCGGTTGC TTTTACTGT ACAATACGGG GACCAAAAAA 1620  
 30 TTGATCTCCC AACATGGCGC ACTGACGACT CTAGCATTTT GGTTCCACA TTTGCAAGAG 1680  
 TACGGTGGCC AAAAACCAGA ATTGAGCAAG CCACATTTTG CATTAGAGGG TTCCGTCGCT 1740  
 GTGGCTGGTG CTGTGGTCCA ATGGCTACGT GATAATTTAC GATTGATCGA TAAATCAGAG 1800  
 35 GATGTCGGAC CGATTGCATC TACGGTTCCT GATTCTGGTG GCGTAGTTTT CGTCCCCGCA 1860  
 TTTAGTGGCC TATTGCTCC CTATTGGGAC CCAGATGCCA GAGCCACCAT AATGGGGATG 1920  
 40 TCTCAATTCA CTA CTACTGCCTC CCACATCGCC AGAGCTGCCG TGGAAGGTGT TTGCTTTCAA 1980  
 GCCAGGGCTA TCTTGAAGGC AATGAGTTCT GACGCGTTTG GTGAAGGTTT CAAAGACAGG 2040  
 GACTTTTTAG AGGAAATTTT CGACGTCACA TATGAAAAGT CGCCCCTGTC GGTTCTGGCA 2100  
 45 GTGGATGGCG GGATGTCGAG GTCTAATGAA GTCATGCAA TTCAAGCCGA TATCCTAGGT 2160  
 CCCTGTGTCA AAGTCAGAAG GTCTCCGACA GCGGAATGTA CCGCATTGGG GGCAGCCATT 2220  
 50 GCAGCCAATA TGGCTTTCAA GGATGTGAAC GAGCGCCAT TATGGAAGGA CCTACACGAT 2280  
 GTTAAGAAAT GGGTCTTTTA CAATGGAATG GAGAAAAACG AACAAATATC ACCAGAGGCT 2340  
 CATCCAAACC TTAAGATATT CAGAAGTGAA TCCGACGATG CTGAAAGGAG AAAGCATTGG 2400  
 55 AAGTATTGGG AAGTTGCCGT GGAAAGATCC AAAGGTTGGC TGAAGGACAT AGAAGGTGAA 2460  
 CACGAACAGG TTCTAGAAAA CTCCAATAA CAACATAAAT AATTTCTATT AACAAATGTAA 2520

60 (2) INFORMACIÓN PARA LA SEQ. ID. NO:7:

(i) CARACTERÍSTICAS DE LA SECUENCIA:

- 65 (A) LONGITUD: 391 aminoácidos  
 (B) TIPO: aminoácido  
 (C) TIPO DE CADENA: desconocido



ES 2 345 469 T3

Cys Gly Ala Leu Lys Asn Val Val Ala Leu Gly Cys Gly Phe Val Glu  
 245 250 255  
 5 Gly Leu Gly Trp Gly Asn Asn Ala Ser Ala Ala Ile Gln Arg Val Gly  
 260 265 270  
 10 Leu Gly Glu Ile Ile Arg Phe Gly Gln Met Phe Phe Pro Glu Ser Arg  
 275 280 285  
 15 Glu Glu Thr Tyr Tyr Gln Glu Ser Ala Gly Val Ala Asp Leu Ile Thr  
 290 295 300  
 20 Thr Cys Ala Gly Gly Arg Asn Val Lys Val Ala Arg Leu Met Ala Thr  
 305 310 315  
 25 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  
 35 Thr Cys Gly Ser Val Glu Asp Phe Pro Leu Phe Glu Ala Val Tyr Gln  
 355 360 365  
 40 Ile Val Tyr Asn Asn Tyr Pro Met Lys Asn Leu Pro Asp Met Ile Glu  
 370 375 380  
 45 Glu Leu Asp Leu His Glu Asp  
 385 390

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

(i) CARACTERÍSTICAS DE LA SECUENCIA:

- 40 (A) LONGITUD: 384 aminoácidos  
 (B) TIPO: aminoácido  
 (C) TIPO DE CADENA: desconocido  
 (D) CONFIGURACIÓN: desconocida

45 (ii) TIPO DE MOLÉCULA: proteína

(xi) DESCRIPCIÓN DE LA SECUENCIA: SEQ. ID. NO:8:

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

ES 2 345 469 T3

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

65

## ES 2 345 469 T3

(2) INFORMACIÓN PARA LA SEQ. ID. NO:9:

(i) CARACTERÍSTICAS DE LA SECUENCIA:

- (A) LONGITUD: 614 aminoácidos
- (B) TIPO: aminoácido
- (C) TIPO DE CADENA: desconocido
- (D) CONFIGURACIÓN: desconocida

(ii) TIPO DE MOLÉCULA: proteína

(xi) DESCRIPCIÓN DE LA SECUENCIA: SEQ. ID. NO:9:

```

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          20          25          30

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
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
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
115         120         125

Thr Trp Gln Val Pro Tyr Ile Tyr Met Gly Cys Lys Phe Tyr Asp Phe
130         135         140

Phe Gly Gly Ser Gln Asn Leu Lys Lys Ser Tyr Leu Leu Ser Lys Ser
145         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
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
210         215         220

Gly Ala Glu Ala Arg Asp Val Glu Thr Asn Glu Leu Val Arg Ile Asn
225         230         235         240

```

ES 2 345 469 T3

Ala Lys Cys Val Val Asn Ala Thr Gly Pro Tyr Ser Asp Ala Ile Leu  
 245 250 255

5 Gln Met Asp Arg Asn Pro Ser Gly Leu Pro Asp Ser Pro Leu Asn Asp  
 260 265 270

10 Asn Ser Lys Ile Lys Ser Thr Phe Asn Gln Ile Ser Val Met Asp Pro  
 275 280 285

Lys Met Val Ile Pro Ser Ile Gly Val His Ile Val Leu Pro Ser Phe  
 290 295 300

15 Tyr Ser Pro Lys Asp Met Gly Leu Leu Asp Val Arg Thr Ser Asp Gly  
 305 310 315 320

20 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

25 Ala Asp Ile Gln Asp Ile Leu Lys Glu Leu Gln His Tyr Ile Glu Phe  
 355 360 365

30 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

35 Ala Thr Gln Gly Val Val Arg Ser His Phe Leu Phe Thr Ser Asp Asn  
 405 410 415

40 Gly Leu Ile Thr Ile Ala Gly Gly Lys Trp Thr Thr Tyr Arg Gln Met  
 420 425 430

Ala Glu Glu Thr Val Asp Lys Val Val Glu Val Gly Gly Phe His Asn  
 435 440 445

45 Leu Lys Pro Cys His Thr Arg Asp Ile Lys Leu Ala Gly Ala Glu Glu  
 450 455 460

50 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

55 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

60 Asn Leu Val Asn Phe Asp Thr Phe Arg Tyr Pro Phe Thr Ile Gly Glu  
 530 535 540

65

ES 2 345 469 T3

Leu Lys Tyr Ser Met Gln Tyr Glu Tyr Cys Arg Thr Pro Leu Asp Phe  
545 550 555 560

5 Leu Leu Arg Arg Thr Arg Phe Ala Phe Leu Asp Ala Lys Glu Ala Leu  
565 570 575

10 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

15 Gln Gly Arg Phe Gly Val  
610

20 (2) INFORMACIÓN PARA LA SEQ. ID. NO:10:

(i) CARACTERÍSTICAS DE LA SECUENCIA:

(A) LONGITUD: 339 aminoácidos

25 (B) TIPO: aminoácido

(C) TIPO DE CADENA: desconocido

(D) CONFIGURACIÓN: desconocida

30 (ii) TIPO DE MOLÉCULA: proteína

(xi) DESCRIPCIÓN DE LA SECUENCIA: SEQ. ID. NO:10:

35 Met Asn Gln Arg Asn Ala Ser Met Thr Val Ile Gly Ala Gly Ser Tyr  
1 5 10 15

Gly Thr Ala Leu Ala Ile Thr Leu Ala Arg Asn Gly His Glu Val Val  
20 25 30

40 Leu Trp Gly His Asp Pro Glu His Ile Ala Thr Leu Glu Arg Asp Arg  
35 40 45

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

50 Val Val Val Pro Ser His Val Phe Gly Glu Val Leu Arg Gln Ile Lys  
85 90 95

55 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

60 Gly Asp Gln Ile Pro Leu Ala Val Ile Ser Gly Pro Thr Phe Ala Lys  
130 135 140

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

ES 2 345 469 T3

Gln Thr Phe Ala Asp Asp Leu Gln Gln Leu Leu His Cys Gly Lys Ser  
 165 170 175  
 5 Phe Arg Val Tyr Ser Asn Pro Asp Phe Ile Gly Val Gln Leu Gly Gly  
 180 185 190  
 10 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  
 15 Glu Met Ser Arg Leu Gly Ala Ala Leu Gly Ala Asp Pro Ala Thr Phe  
 225 230 235 240  
 20 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  
 25 Val Gln Ser Ala Gln Glu Lys Ile Gly Gln Val Val Glu Gly Tyr Arg  
 275 280 285  
 30 Asn Thr Lys Glu Val Arg Glu Leu Ala His Arg Phe Gly Val Glu Met  
 290 295 300  
 Pro Ile Thr Glu Glu Ile Tyr Gln Val Leu Tyr Cys Gly Lys Asn Ala  
 305 310 315 320  
 35 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 SEQ. ID. NO:11:

(i) CARACTERÍSTICAS DE LA SECUENCIA:

- (A) LONGITUD: 501 aminoácidos
- (B) TIPO: aminoácido
- (C) TIPO DE CADENA: desconocido
- (D) CONFIGURACIÓN: desconocida

(ii) TIPO DE MOLÉCULA: proteína

(xi) DESCRIPCIÓN DE LA SECUENCIA: SEQ. ID. NO:11:

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

ES 2 345 469 T3

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

ES 2 345 469 T3

Leu Thr Pro Tyr Tyr Gln Gly Ile Gly Pro Ala Trp Thr Lys Glu Ser  
 370 375 380  
 5 Val Leu Pro Gly Gly Ala Ile Glu Gly Asp Arg Asp Asp Tyr Ala Ala  
 385 390 395 400  
 10 Arg Leu Arg Arg Arg Tyr Pro Phe Leu Thr Glu Ser Leu Ala Arg His  
 405 410 415  
 15 Tyr Ala Arg Thr Tyr Gly Ser Asn Ser Glu Leu Leu Leu Gly Asn Ala  
 420 425 430  
 20 Gly Thr Val Ser Asp Leu Gly Glu Asp Phe Gly His Glu Phe Tyr Glu  
 435 440 445  
 25 Ala Glu Leu Lys Tyr Leu Val Asp His Glu Trp Val Arg Arg Ala Asp  
 450 455 460  
 30 Asp Ala Leu Trp Arg Arg Thr Lys Gln Gly Met Trp Leu Asn Ala Asp  
 465 470 475 480  
 35 Gln Gln Ser Arg Val Ser Gln Trp Leu Val Glu Tyr Thr Gln Gln Arg  
 485 490 495  
 40 Leu Ser Leu Ala Ser  
 500

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

35 (i) CARACTERÍSTICAS DE LA SECUENCIA:

(A) LONGITUD: 542 aminoácidos

(B) TIPO: aminoácido

40 (C) TIPO DE CADENA: desconocido

(D) CONFIGURACIÓN: desconocida

(ii) TIPO DE MOLÉCULA: proteína

45 (xi) DESCRIPCIÓN DE LA SECUENCIA: SEQ. ID. NO:12:

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

ES 2 345 469 T3

Pro Glu Asp Asp Leu Ser Phe Gln Ala Thr Phe Ile Arg Ala Cys Glu  
 100 105 110  
 5 Glu Ala Gly Ile Ser Ala Glu Ala Ile Asp Pro Gln Gln Ala Arg Ile  
 115 120 125  
 10 Ile Glu Pro Ala Val Asn Pro Ala Leu Ile Gly Ala Val Lys Val Pro  
 130 135 140  
 15 Asp Gly Thr Val Asp Pro Phe Arg Leu Thr Ala Ala Asn Met Leu Asp  
 145 150 155 160  
 20 Ala Lys Glu His Gly Ala Val Ile Leu Thr Ala His Glu Val Thr Gly  
 165 170 175  
 25 Leu Ile Arg Glu Gly Ala Thr Val Cys Gly Val Arg Val Arg Asn His  
 180 185 190  
 30 Leu Thr Gly Glu Thr Gln Ala Leu His Ala Pro Val Val Val Asn Ala  
 195 200 205  
 35 Ala Gly Ile Trp Gly Gln His Ile Ala Glu Tyr Ala Asp Leu Arg Ile  
 210 215 220  
 40 Arg Met Phe Pro Ala Lys Gly Ser Leu Leu Ile Met Asp His Arg Ile  
 225 230 235 240  
 45 Asn Gln His Val Ile Asn Arg Cys Arg Lys Pro Ser Asp Ala Asp Ile  
 245 250 255  
 50 Leu Val Pro Gly Asp Thr Ile Ser Leu Ile Gly Thr Thr Ser Leu Arg  
 260 265 270  
 55 Ile Asp Tyr Asn Glu Ile Asp Asp Asn Arg Val Thr Ala Glu Glu Val  
 275 280 285  
 60 Asp Ile Leu Leu Arg Glu Gly Glu Lys Leu Ala Pro Val Met Ala Lys  
 290 295 300  
 65 Thr Arg Ile Leu Arg Ala Tyr Ser Gly Val Arg Pro Leu Val Ala Ser  
 305 310 315 320  
 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  
 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  
 385 390 395 400

ES 2 345 469 T3

Ile Ser Leu Pro Ala Pro Leu Arg Gly Ser Ala Val Tyr Arg His Gly  
 405 410 415  
 5 Asp Arg Thr Pro Ala Trp Leu Ser Glu Gly Arg Leu His Arg Ser Leu  
 420 425 430  
 10 Val Cys Glu Cys Glu Ala Val Thr Ala Gly Glu Val Gln Tyr Ala Val  
 435 440 445  
 Glu Asn Leu Asn Val Asn Ser Leu Leu Asp Leu Arg Arg Arg Thr Arg  
 450 455 460  
 15 Val Gly Met Gly Thr Cys Gln Gly Glu Leu Cys Ala Cys Arg Ala Ala  
 465 470 475 480  
 20 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  
 25 Ala Trp Gly Asp Ala Leu Arg Glu Ser Glu Phe Thr Arg Trp Val Tyr  
 515 520 525  
 30 Gln Gly Leu Cys Gly Leu Glu Lys Glu Gln Lys Asp Ala Leu  
 530 535 540

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

- 35 (i) CARACTERÍSTICAS DE LA SECUENCIA:
- (A) LONGITUD: 250 aminoácidos
  - (B) TIPO: aminoácido
  - 40 (C) TIPO DE CADENA: desconocido
  - (D) CONFIGURACIÓN: desconocida
- (ii) TIPO DE MOLÉCULA: proteína
- 45 (xi) DESCRIPCIÓN DE LA SECUENCIA: SEQ. ID. N°:13:

50

55

60

65

ES 2 345 469 T3

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

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

(i) CARACTERÍSTICAS DE LA SECUENCIA:

- (A) LONGITUD: 271 aminoácidos
- (B) TIPO: aminoácido
- (C) TIPO DE CADENA: desconocido
- (D) CONFIGURACIÓN: desconocida

(ii) TIPO DE MOLÉCULA: proteína

ES 2 345 469 T3

(xi) DESCRIPCIÓN DE LA SECUENCIA: SEQ. ID. NO:14:

5 Met Lys Arg Phe Asn Val Leu Lys Tyr Ile Arg Thr Thr Lys Ala Asn  
1 5 10 15

Ile Gln Thr Ile Ala Met Pro Leu Thr Thr Lys Pro Leu Ser Leu Lys  
20 25 30

10 Ile Asn Ala Ala Leu Phe Asp Val Asp Gly Thr Ile Ile Ser Gln  
35 40 45

Pro Ala Ile Ala Ala Phe Trp Arg Asp Phe Gly Lys Asp Lys Pro Tyr  
50 55 60

15 Phe Asp Ala Glu His Val Ile His Ile Ser His Gly Trp Arg Thr Tyr  
65 70 75 80

20 Asp Ala Ile Ala Lys Phe Ala Pro Asp Phe Ala Asp Glu Glu Tyr Val  
85 90 95

Asn Lys Leu Glu Gly Glu Ile Pro Glu Lys Tyr Gly Glu His Ser Ile  
100 105 110

25 Glu Val Pro Gly Ala Val Lys Leu Cys Asn Ala Leu Asn Ala Leu Pro  
115 120 125

Lys Glu Lys Trp Ala Val Ala Thr Ser Gly Thr Arg Asp Met Ala Lys  
130 135 140

30 Lys Trp Phe Asp Ile Leu Lys Ile Lys Arg Pro Glu Tyr Phe Ile Thr  
145 150 155 160

35 Ala Asn Asp Val Lys Gln Gly Lys Pro His Pro Glu Pro Tyr Leu Lys  
165 170 175

Gly Arg Asn Gly Leu Gly Phe Pro Ile Asn Glu Gln Asp Pro Ser Lys  
180 185 190

40 Ser Lys Val Val Val Phe Glu Asp Ala Pro Ala Gly Ile Ala Ala Gly  
195 200 205

45 Lys Ala Ala Gly Cys Lys Ile Val Gly Ile Ala Thr Thr Phe Asp Leu  
210 215 220

Asp Phe Leu Lys Glu Lys Gly Cys Asp Ile Ile Val Lys Asn His Glu  
225 230 235 240

50 Ser Ile Arg Val Gly Glu Tyr Asn Ala Glu Thr Asp Glu Val Glu Leu  
245 250 255

Ile Phe Asp Asp Tyr Leu Tyr Ala Lys Asp Asp Leu Leu Lys Trp  
55 260 265 270

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

60 (i) CARACTERÍSTICAS DE LA SECUENCIA:

- (A) LONGITUD: 709 aminoácidos
- (B) TIPO: aminoácido
- (C) TIPO DE CADENA: desconocido
- (D) CONFIGURACIÓN: desconocida

65

(ii) TIPO DE MOLÉCULA: proteína

ES 2 345 469 T3

(xi) DESCRIPCIÓN DE LA SECUENCIA: SEQ. ID. NO:15:

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

ES 2 345 469 T3

Gly Ile Pro Asp Trp Ile Met Glu Lys Leu His Asp Ser Pro Lys Thr  
 355 360 365  
 5 Val Leu Arg Asp Leu Val Lys Arg Asn Leu Pro Ile Gln Gly Cys Leu  
 370 375 380  
 10 Gly Asp Gln Ser Ala Ser Met Val Gly Gln Leu Ala Tyr Lys Pro Gly  
 385 390 395 400  
 Ala Ala Lys Cys Thr Tyr Gly Thr Gly Cys Phe Leu Leu Tyr Asn Thr  
 405 410 415  
 15 Gly Thr Lys Lys Leu Ile Ser Gln His Gly Ala Leu Thr Thr Leu Ala  
 420 425 430  
 20 Phe Trp Phe Pro His Leu Gln Glu Tyr Gly Gly Gln Lys Pro Glu Leu  
 435 440 445  
 Ser Lys Pro His Phe Ala Leu Glu Gly Ser Val Ala Val Ala Gly Ala  
 450 455 460  
 25 Val Val Gln Trp Leu Arg Asp Asn Leu Arg Leu Ile Asp Lys Ser Glu  
 465 470 475 480  
 30 Asp Val Gly Pro Ile Ala Ser Thr Val Pro Asp Ser Gly Gly Val Val  
 485 490 495  
 Phe Val Pro Ala Phe Ser Gly Leu Phe Ala Pro Tyr Trp Asp Pro Asp  
 500 505 510  
 35 Ala Arg Ala Thr Ile Met Gly Met Ser Gln Phe Thr Thr Ala Ser His  
 515 520 525  
 40 Ile Ala Arg Ala Ala Val Glu Gly Val Cys Phe Gln Ala Arg Ala Ile  
 530 535 540  
 Leu Lys Ala Met Ser Ser Asp Ala Phe Gly Glu Gly Ser Lys Asp Arg  
 545 550 555 560  
 45 Asp Phe Leu Glu Glu Ile Ser Asp Val Thr Tyr Glu Lys Ser Pro Leu  
 565 570 575  
 Ser Val Leu Ala Val Asp Gly Gly Met Ser Arg Ser Asn Glu Val Met  
 580 585 590  
 50 Gln Ile Gln Ala Asp Ile Leu Gly Pro Cys Val Lys Val Arg Arg Ser  
 595 600 605  
 55 Pro Thr Ala Glu Cys Thr Ala Leu Gly Ala Ala Ile Ala Ala Asn Met  
 610 615 620  
 60 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

65

ES 2 345 469 T3

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

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

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

Leu Glu Asn Phe Gln  
705

15

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

(i) CARACTERÍSTICAS DE LA SECUENCIA:

20

(A) LONGITUD: 51 pares de bases

(B) TIPO: ácido nucleico

(C) TIPO DE CADENA: sencilla

25

(D) CONFIGURACIÓN: lineal

(ii) TIPO DE MOLÉCULA: ADN (genómico)

(xi) DESCRIPCIÓN DE LA SECUENCIA: SEQ. ID. NO:16:

30

GCGCGGATCC AGGAGTCTAG AATTATGGGA TTGACTACTA AACCTCTATC T

51

(2) INFORMACIÓN PARA LA SEQ. ID. NO:17:

(i) CARACTERÍSTICAS DE LA SECUENCIA:

35

(A) LONGITUD: 36 pares de bases

(B) TIPO: ácido nucleico

40

(C) TIPO DE CADENA: sencilla

(D) CONFIGURACIÓN: lineal

(ii) TIPO DE MOLÉCULA: ADN (genómico)

(xi) DESCRIPCIÓN DE LA SECUENCIA: SEQ. ID. NO: 17:

45

GATACGCCCG GGTACCATT TCAACAGATC GTCCTT

36

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

50

(i) CARACTERÍSTICAS DE LA SECUENCIA:

(A) LONGITUD: 34 pares de bases

(B) TIPO: ácido nucleico

55

(C) TIPO DE CADENA: sencilla

(D) CONFIGURACIÓN: lineal

(ii) TIPO DE MOLÉCULA: ADN (genómico)

(xi) DESCRIPCIÓN DE LA SECUENCIA: SEQ. ID. NO:18:

60

TTGATAATAT AACCATGGCT GCTGCTGCTG ATAG

34

65

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

(i) CARACTERÍSTICAS DE LA SECUENCIA:

## ES 2 345 469 T3

- (A) LONGITUD: 39 pares de bases  
(B) TIPO: ácido nucleico  
(C) TIPO DE CADENA: sencilla  
5 (D) CONFIGURACIÓN: lineal
- (ii) TIPO DE MOLÉCULA: ADN (genómico)
- (xi) DESCRIPCIÓN DE LA SECUENCIA: SEQ. ID. NO:19:  
10 GTATGATATG TTATCTTGGA TCCAATAAAT CTAATCTTC 39
- (2) INFORMACIÓN PARA LA SEQ. ID. NO:20:  
15 (i) CARACTERÍSTICAS DE LA SECUENCIA:  
(A) LONGITUD: 24 pares de bases  
(B) TIPO: ácido nucleico  
20 (C) TIPO DE CADENA: sencilla  
(D) CONFIGURACIÓN: lineal
- (ii) TIPO DE MOLÉCULA: ADN (genómico)
- (xi) DESCRIPCIÓN DE LA SECUENCIA: SEQ. ID. NO:20:  
25 CATGACTAGT AAGGAGGACA ATTC 24
- (2) INFORMACIÓN PARA LA SEQ. ID. NO:21:  
30 (i) CARACTERÍSTICAS DE LA SECUENCIA:  
(A) LONGITUD: 24 pares de bases  
35 (B) TIPO: ácido nucleico  
(C) TIPO DE CADENA: sencilla  
(D) CONFIGURACIÓN: lineal
- (ii) TIPO DE MOLÉCULA: ADN (genómico)
- (xi) DESCRIPCIÓN DE LA SECUENCIA: SEQ. ID. NO:21:  
40 CATGGAATTG TCCTCCTTAC TAGT 24
- (2) INFORMACIÓN PARA LA SEQ. ID. NO:22:  
50 (i) CARACTERÍSTICAS DE LA SECUENCIA:  
(A) LONGITUD: 19 pares de bases  
(B) TIPO: ácido nucleico  
(C) TIPO DE CADENA: sencilla  
55 (D) CONFIGURACIÓN: lineal
- (ii) TIPO DE MOLÉCULA: ADN (genómico)
- (xi) DESCRIPCIÓN DE LA SECUENCIA: SEQ. ID. NO:22:  
60 CTAGTAAGGA GGACAATTC 19
- (2) INFORMACIÓN PARA LA SEQ. ID. NO:23:  
65 (i) CARACTERÍSTICAS DE LA SECUENCIA:  
(A) LONGITUD: 19 pares de bases

## ES 2 345 469 T3

- (B) TIPO: ácido nucleico
- (C) TIPO DE CADENA: sencilla
- (D) CONFIGURACIÓN: lineal

5

(ii) TIPO DE MOLÉCULA: ADN (genómico)

(xi) DESCRIPCIÓN DE LA SECUENCIA: SEQ. ID. NO:23:

10

CATGGAATTG TCCTCCTTA

19

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

15

(i) CARACTERÍSTICAS DE LA SECUENCIA:

(A) LONGITUD: 15 pares de bases

(B) TIPO: ácido nucleico

(C) TIPO DE CADENA: sencilla

20

(D) CONFIGURACIÓN: lineal

(ii) TIPO DE MOLÉCULA: otro ácido nucleico

(A) DESCRIPCIÓN: /desc = "CEBADOR"

25

(iii) HIPOTÉTICO: No

(iv) ANTI-SENTIDO: No

(xi) DESCRIPCIÓN DE LA SECUENCIA: SEQ. ID. NO:24:

30

GATCCAGGAA ACAGA

15

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

35

(i) CARACTERÍSTICAS DE LA SECUENCIA:

(A) LONGITUD: 15 pares de bases

(B) TIPO: ácido nucleico

40

(C) TIPO DE CADENA: sencilla

(D) CONFIGURACIÓN: lineal

(ii) TIPO DE MOLÉCULA: otro ácido nucleico

45

(A) DESCRIPCIÓN: /desc = "CEBADOR"

(iii) HIPOTÉTICO: No

(iv) ANTI-SENTIDO: No

50

(xi) DESCRIPCIÓN DE LA SECUENCIA: SEQ. ID. NO:25:

55

CTAGTCTGTT TCCTG

15

60

65