

19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 613 553**

51 Int. Cl.:

**C12N 9/02** (2006.01)

**C12P 7/04** (2006.01)

**C12P 7/64** (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **02.04.2013 PCT/US2013/035040**

87 Fecha y número de publicación internacional: **10.10.2013 WO2013152052**

96 Fecha de presentación y número de la solicitud europea: **02.04.2013 E 13766728 (3)**

97 Fecha y número de publicación de la concesión europea: **16.11.2016 EP 2834350**

54 Título: **Enzimas CAR y producción mejorada de alcoholes grasos**

30 Prioridad:

**02.04.2012 US 201261619309 P**

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

**24.05.2017**

73 Titular/es:

**REG LIFE SCIENCES, LLC (100.0%)  
600 Gateway Boulevard  
South San Francisco, CA 94080, US**

72 Inventor/es:

**GREENFIELD, DEREK L.;  
CLARKE, ELIZABETH J.;  
GROBAN, ELI S.;  
ARLAGADDA, VIKRANTH;  
LEE, SUNGWON;  
LI, XUEZHI;  
HU, ZHIHAO y  
ZHU, BAOLONG**

74 Agente/Representante:

**MILTENYI, Peter**

ES 2 613 553 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

Enzimas CAR y producción mejorada de alcoholes grasos

**Campo de la divulgación**

5 La divulgación se refiere a enzimas ácido carboxílico reductasa (CAR) variantes para la producción mejorada de alcoholes grasos en células huésped recombinantes. La divulgación se refiere además a ácidos nucleicos y polipéptidos de CAR variantes así como a células huésped recombinantes y cultivos celulares. Se engloban además métodos de preparación de composiciones de alcoholes grasos.

**Antecedentes de la divulgación**

10 Los alcoholes grasos constituyen una categoría importante de productos bioquímicos industriales. Estas moléculas y sus derivados tienen numerosos usos, incluyendo como tensioactivos, lubricantes, plastificantes, disolventes, emulsionantes, emolientes, espesantes, sabores, fragancias y combustibles. En la industria, los alcoholes grasos se producen por medio de hidrogenación catalítica de ácidos grasos producidos a partir de fuentes naturales, tales como aceite de coco, aceite de palma, aceite de palmiste, sebo y manteca de cerdo, o mediante hidratación química de alfa-olefinas producidas a partir de materias primas petroquímicas. Los alcoholes grasos derivados de fuentes naturales tienen longitudes de cadena variables. La longitud de cadena de los alcoholes grasos es importante con respecto a aplicaciones particulares. En la naturaleza, los alcoholes grasos también se preparan mediante enzimas que pueden reducir moléculas de acil-ACP o acil-CoA hasta los correspondientes alcoholes primarios (véanse, por ejemplo, las publicaciones de patente estadounidense n.ºs 20100105955, 20100105963 y 20110250663).

20 El documento US 2010/0298612 describe la producción de alcoholes grasos de cadena larga mediante células huésped recombinantes que expresan enzimas ácido carboxílico reductasa heterólogas. El documento US 2010/0105963 describe métodos para producir alcoholes grasos expresando una ácido carboxílico reductasa en una célula huésped. El documento WO 20111047101 describe microorganismos recombinantes para la producción de 1,2-butanodiol y otros alcoholes, pero no usos específicos para la producción de alcoholes grasos.

25 Las tecnologías actuales para producir alcoholes grasos implican la reducción mediada por catalizador inorgánico de ácidos grasos hasta los correspondientes alcoholes primarios, que es costosa, requiere mucho tiempo y es engorrosa. Los ácidos grasos usados en este procedimiento se derivan de fuentes naturales (por ejemplo, grasas y aceites animales y vegetales, citados anteriormente). La deshidratación de alcoholes grasos para dar alfa-olefinas también puede lograrse mediante catálisis química. Sin embargo, esta técnica no es renovable y está asociada con un alto coste de funcionamiento y desechos químicos peligrosos para el medio ambiente. Por tanto, existe la necesidad de métodos mejorados para la producción de alcoholes grasos y la presente divulgación aborda esta necesidad.

**Sumario**

35 Un aspecto de la divulgación proporciona un polipéptido de ácido carboxílico reductasa (CAR) variante que comprende una secuencia de aminoácidos que tiene una identidad de secuencia de al menos el 90% con SEQ ID NO: 7, en el que el polipéptido de CAR variante se modifica por ingeniería genética para que tenga al menos una mutación en una posición de aminoácido seleccionada del grupo de S3R, D18R, D18L, D18T, D18P, E20V, E20S, E20R, S22R, S22N, S22G, L80R, R87G, R87E, V191S, F288R, F288S, F288G, Q473L, Q473W, Q473Y, Q473I, Q473H, A535S, D750A, R827C, R827A, I870L, R873S, V926A, V926E, S927K, S927G, M930K, M930R y L1128W, en la que la expresión del polipéptido de CAR variante en una célula huésped recombinante da como resultado un título superior de las composiciones de alcoholes grasos en comparación con una célula huésped recombinante que expresa el polipéptido de CAR silvestre de SEQ ID NO: 7. En un aspecto relacionado, el polipéptido de CAR es un polipéptido de CarB.

45 En realizaciones preferidas, el polipéptido de CAR incluye la mutación A535S; o las mutaciones E20R, F288G, Q473I y A535S; o las mutaciones E20R, F288G, Q473H, A535S, R827A y S927G; o las mutaciones E20R, S22R, F288G, Q473H, A535S, R827A y S927G; o las mutaciones S3R, E20R, S22R, F288G, Q473H, A535S, R873S, S927G, M930R y L1128W; o E20R, S22R, F288G, Q473H, A535S, R873S, S927G, M930R y L1128W; o las mutaciones D18R, E20R, S22R, F288G, Q473I, A535S, S927G, M930K y L1128W; o las mutaciones E20R, S22R, F288G, Q473I, A535S, R827C, V926E, S927K y M930R; o las mutaciones D18R, E20R, 288G, Q473I, A535S, R827C, V926E, M930K y L1128W; o las mutaciones E20R, S22R, F288G, Q473H, A535S, R827C, V926A, S927K y M930R; o las mutaciones E20R, S22R, F288G, Q473H, A535S y R827C; o las mutaciones E20R, S22R, F288G, Q473I, A535S, R827C y M930R; o las mutaciones E20R, S22R, F288G, Q473I, A535S, I870L, S927G y M930R; o las mutaciones E20R, S22R, F288G, Q473I, A535S, I870L y S927G; o las mutaciones D18R, E20R, S22R, F288G, Q473I, A535S, R827C, I870L, V926A y S927G; o las mutaciones E20R, S22R, F288G, Q473H, A535S, R827C, I870L y L1128W; o las mutaciones D18R, E20R, S22R, F288G, Q473H, A535S, R827C, I870L, S927G y L1128W; o las mutaciones E20R, S22R, F288G, Q473I, A535S, R827C, I870L, S927G y L1128W; o las mutaciones E20R, S22R, F288G, Q473I, A535S, R827C, I870L, S927G, M930K y L1128W; o las mutaciones E20R, S22R, F288G, Q473H, A535S, R827C, I870L, M930K; o las mutaciones E20R, S22R, F288G, Q473H, A535S, S927G, M930K y L1128W; o las mutaciones D18R, E20R, S22R,

F288G, Q473I, A535S, S927G y L1128W; o las mutaciones E20R, S22R, F288G, Q473I, A535S, R827C, I870L y S927G; o las mutaciones D18R, E20R, S22R, F288G, Q473I, A535S, R827C, I870L, S927G y L1128W; o las mutaciones D18R, E20R, S22R, F288G, Q473I, A535S, S927G, M930R y L1128W; o las mutaciones E20R, S22R, F288G, Q473H, A535S, V926E, S927G y M930R; o las mutaciones E20R, S22R, F288G, Q473H, A535S, R827C, I870L, V926A y L1128W; o combinaciones de las mismas.

Otro aspecto de la divulgación proporciona una célula huésped que incluye una secuencia de polinucleótido que codifica para un polipéptido de ácido carboxílico reductasa (CAR) variante que tiene una identidad de secuencia de al menos el 90% con SEQ ID NO: 7 y que tiene al menos una mutación en una posición de aminoácido seleccionada del grupo de S3R, D18R, D18L, D18T, D18P, E20V, E20S, E20R, S22R, S22N, S22G, L80R, R87G, R87E, V191S, F288R, F288S, F288G, Q473L, Q473W, Q473Y, Q473I, Q473H, A535S, D750A, R827C, R827A, I870L, R873S, V926A, V926E, S927K, S927G, M930K, M930R y L1128W en la que la célula huésped modificada por ingeniería genética produce una composición de alcohol graso a un rendimiento o título superior que una célula huésped que expresa un polipéptido de CAR silvestre correspondiente cuando se cultiva en un medio que contiene una fuente de carbono en condiciones eficaces para expresar el polipéptido de CAR variante, y en la que la SEQ ID NO: 7 es el polipéptido de CAR silvestre correspondiente. En una realización preferida, la célula huésped recombinante incluye además un polinucleótido que codifica para un polipéptido de tioesterasa. En otra realización preferida, la célula huésped recombinante incluye además un polinucleótido que codifica para un polipéptido de FabB y un polipéptido de FadR. En otra realización preferida, la célula huésped recombinante incluye además un polinucleótido que codifica para una aldehído graso reductasa, por ejemplo AlrA y un cultivo celular que la contiene.

En otra realización preferida, la célula huésped modificada por ingeniería genética tiene un título que es al menos 3 veces mayor que el título de una célula huésped que expresa el polipéptido de CAR silvestre correspondiente cuando se cultiva en las mismas condiciones que la célula huésped modificada por ingeniería genética. En una realización preferida, la célula huésped modificada por ingeniería genética tiene un título de desde 30 g/l hasta 250 g/l. En otra realización preferida, la célula huésped modificada por ingeniería genética tiene un título de desde 90 g/l hasta 120 g/l.

En otra realización preferida, la célula huésped modificada por ingeniería genética tiene un rendimiento que es al menos 3 veces mayor que el rendimiento de una célula huésped que expresa el polipéptido de CAR silvestre correspondiente cuando se cultiva en las mismas condiciones que la célula huésped modificada por ingeniería genética. En una realización preferida, la célula huésped modificada por ingeniería genética tiene un rendimiento de desde el 10% hasta el 40%.

La divulgación engloba además un cultivo celular que incluye la célula huésped recombinante tal como se describe en el presente documento. En una realización preferida, el cultivo celular tiene una productividad que es al menos aproximadamente 3 veces mayor que la productividad de un cultivo celular que expresa el polipéptido de CAR silvestre correspondiente. En otra realización preferida, la productividad oscila entre 0,7 mg/l/h y 3 g/l/h. En otra realización preferida, el medio de cultivo comprende una composición de alcohol graso. La composición de alcohol graso se produce de manera extracelular. La composición de alcohol graso puede incluir uno o más de un alcohol graso C6, C8, C10, C12, C13, C14, C15, C16, C17 o C18; o un alcohol graso insaturado C10:1, C12:1, C14:1, C16:1 o C18:1. En otra realización preferida, la composición de alcohol graso comprende alcoholes grasos C12 y C14. En aún otra realización preferida, la composición de alcohol graso comprende alcoholes grasos C12 y C14 a una razón de aproximadamente 3:1. En todavía otra realización preferida, la composición de alcohol graso engloba alcoholes grasos insaturados. Además, la composición de alcohol graso puede incluir un alcohol graso que tiene un doble enlace en la posición 7 en la cadena de carbono entre C7 y C8 desde el extremo reducido del alcohol graso. En otra realización preferida, la composición de alcohol graso incluye alcoholes grasos saturados. En otra realización preferida, la composición de alcohol graso incluye alcoholes grasos de cadena ramificada.

La divulgación contempla además un método de preparación de una composición de alcohol graso a un alto título, rendimiento o productividad, que incluye las etapas de modificar por ingeniería genética la célula huésped recombinante de la divulgación; cultivar la célula huésped recombinante en un medio que incluye una fuente de carbono; y opcionalmente aislar la composición de alcohol graso del medio.

### Breve descripción de los dibujos

La presente divulgación se entiende mejor cuando se lee junto con las figuras adjuntas, que sirven para ilustrar las realizaciones preferidas. Sin embargo, se entiende que la divulgación no se limita a las realizaciones específicas dadas a conocer en las figuras.

La figura 1 es una visión general esquemática de una ruta de biosíntesis a modo de ejemplo para su uso en la producción de acil-CoA como precursor para dar derivados de ácido graso en una célula huésped recombinante. El ciclo se inicia mediante la condensación de malonil-ACP y acetil-CoA.

La figura 2 es una visión general esquemática de un ciclo de biosíntesis de ácidos grasos a modo de ejemplo, en el que se produce malonil-ACP mediante la transacilación de malonil-CoA con malonil-ACP (catalizada por malonil-CoA:ACP transacilasa; fabD), luego la  $\beta$ -cetoacil-ACP sintasa III (fabH) inicia la condensación de malonil-ACP con

acetil-CoA. Los ciclos de elongación comienzan con la condensación de malonil-ACP y una acil-ACP catalizada por  $\beta$ -cetoacil-ACP sintasa I (fabB) y  $\beta$ -cetoacil-ACP sintasa II (fabF) para producir una  $\beta$ -ceto-acil-ACP, luego la  $\beta$ -ceto-acil-ACP se reduce por una  $\beta$ -cetoacil-ACP reductasa dependiente de NADPH (fabG) para producir una  $\beta$ -hidroxi-acil-ACP, que se deshidrata para dar una *trans*-2-enoil-acil-ACP por  $\beta$ -hidroxiacil-ACP deshidratasa (fabA o fabZ).  
 5 FabA también puede isomerizar *trans*-2-enoil-acil-ACP para dar *cis*-3-enoil-acil-ACP, que puede sortear a fabI y puede usarse por fabB (normalmente para una longitud de cadena alifática de hasta C16) para producir  $\beta$ -ceto-acil-ACP. La etapa final en cada ciclo está catalizada por una enoil-ACP reductasa dependiente de NADH o NADPH (fabI) que convierte *trans*-2-enoil-acil-ACP en acil-ACP. En los métodos descritos en el presente documento, se produce la terminación de la síntesis de ácidos grasos mediante la retirada por tioesterasa del grupo acilo de acil-  
 10 ACP para liberar ácidos grasos libres (FFA, *free fatty acid*). Las tioesterasas (por ejemplo, tesA) hidrolizan enlaces tioéster, que aparecen entre cadenas de acilo y ACP a través de enlaces sulfhidrilo.

La figura 3 ilustra la estructura y función del complejo enzimático de acetil-CoA carboxilasa (accABCD). La biotina carboxilasa está codificada por el gen *accC*, mientras que la proteína transportadora de biotina y carboxilo (BCCP) está codificada por el gen *accB*. Las dos subunidades implicadas en la actividad carboxiltransferasa están codificadas por los genes *accA* y *accD*. La biotina unida covalentemente de BCCP transporta el resto carboxilato. El gen *birA* (no mostrado) biotinila holo-*accB*.  
 15

La figura 4 presenta una visión general esquemática de una ruta de biosíntesis a modo de ejemplo para la producción de alcohol graso partiendo de acil-ACP, en la que la producción de aldehído graso está catalizada por la actividad enzimática de acil-ACP reductasa (AAR) o tioesterasa y ácido carboxílico reductasa (Car). El aldehído graso se convierte en alcohol graso por aldehído reductasa (también denominada alcohol deshidrogenasa). Esta ruta no incluye acil graso-CoA sintetasa (*fadD*).  
 20

La figura 5 ilustra la producción de derivados de ácido graso (especie grasa total) por la cepa de *E. coli* MG1655 con el gen de *fadE* atenuado (es decir, delecionado) en comparación con la producción de derivados de ácido graso por *E. coli* MG1655. Los datos presentados en la figura 5 muestran que la atenuación del gen de *fadE* no afectó a la producción de derivados de ácido graso.  
 25

Las figuras 6A y 6B muestran datos para la producción de "especie grasa total" a partir de exámenes en placa por duplicado cuando se transformó el plásmido pCL-WT TRC WT TesA en cada una de las cepas mostradas en las figuras se realizó una fermentación en medios FA2 con 20 horas desde la inducción hasta la cosecha a tanto 32°C (figura 6A) como 37°C (figura 6B).

Las figuras 7A y 7B proporcionan una representación esquemática del locus *iFAB138*, que incluye un diagrama del promotor *cat-loxP-T5* integrado en frente de *FAB138* (7A); y un diagrama de *iT5\_138* (7B). La secuencia del promotor *cat-loxP-T5* integrado en frente de *FAB138* con homología de 50 pares de bases mostrada en cada lado de la región promotora *cat-loxP-T5* se proporciona como SEQ ID NO: 1 y la secuencia de la región promotora *iT5\_138* con homología de 50 pares de bases en cada lado se proporciona como SEQ ID NO: 2.  
 30

La figura 8 muestra el efecto de corregir los genes de *rph* e *ilvG*. Se transformaron EG149 (*rph- ilvG-*) y V668 (EG149 *rph+ ilvG+*) con pCL-tesA (un plásmido pCL1920 que contiene  $P_{TRC}$ -tesA) obtenido a partir de D191. La figura muestra que corregir los genes de *rph* e *ilvG* en la cepa EG149 permite un nivel superior de producción de FFA que en la cepa V668 en la que los genes de *rph* e *ilvG* no se corrigieron.  
 35

La figura 9 es una representación esquemática de una inserción de casete de transposón en el gen de *yjP* de la cepa LC535 (acierto de transposón 68F11). Se muestran los promotores internos para el casete de transposón, y pueden tener efectos sobre la expresión de genes adyacentes.  
 40

La figura 10 muestra la conversión de ácidos grasos libres en alcoholes grasos mediante CarB60 en la cepa V324. Las figuras muestran que las células que expresan CarB60 a partir del cromosoma (barras oscuras) convierten una mayor fracción de ácidos grasos libres C12 y C14 en alcohol graso en comparación con CarB (barras claras).  
 45

La figura 11 muestra que las células que expresan CarB60 a partir del cromosoma convierten una mayor fracción de ácidos grasos libres C12 y C14 en alcohol graso en comparación con CarB.

La figura 12 muestra la producción de alcohol graso tras la fermentación de mutantes en bibliotecas de combinación.

La figura 13 muestra la producción de alcohol graso por variantes de *carB* en plásmido de producción (*carB1* y *CarB2*) tras la fermentación en matraz de agitación.

La figura 14 muestra la producción de alcohol graso por variantes de *carB* integradas de copia única (*icarB1*, *icarB2*, *icarB3* e *icarB4*) tras la fermentación en matraz de agitación.  
 50

La figura 15 muestra los resultados del sistema de examen de doble plásmido para variantes de CarB mejoradas tal como se valida mediante fermentación en matraz de agitación.

La figura 16 muestra variantes de CarB novedosas para la producción mejorada de alcoholes grasos en

biorreactores.

### Descripción detallada

#### Visión general

5 La presente divulgación se refiere a enzimas ácido carboxílico reductasa (CAR) variantes novedosas así como a sus secuencias de ácido nucleico y proteína. Se engloban además por la divulgación células huésped recombinantes y cultivos celulares que incluyen las enzimas CAR variantes para la producción de alcoholes grasos. La presente invención se define por las reivindicaciones.

10 Con el fin de que la producción de alcoholes grasos a partir de biomasa o azúcares fermentables sea comercialmente viable, el procedimiento debe optimizarse para lograr una conversión y recuperación eficaces del producto. La presente divulgación aborda esta necesidad proporcionando composiciones y métodos para la producción mejorada de alcoholes grasos usando enzimas variantes modificadas por ingeniería genética y células huésped recombinantes modificadas por ingeniería genética. Las células huésped sirven como biocatalizadores que dan como resultado la producción de alto título de alcoholes grasos usando procedimientos de fermentación. Como tales, se dan a conocer en el presente documento métodos para crear células huésped fotosintéticas y heterotróficas que producen alcoholes grasos y alfa-olefinas de longitudes de cadena específicas directamente de manera que no es necesaria la conversión catalítica de ácidos grasos purificados. Este nuevo método proporciona ventajas de coste y calidad del producto.

20 Más específicamente, la producción de una composición de alcohol graso deseada puede potenciarse modificando la expresión de uno o más genes implicados en una ruta de biosíntesis para la producción, degradación y/o secreción de alcohol graso. La divulgación proporciona células huésped recombinantes, que se han modificado por ingeniería genética para proporcionar una biosíntesis de alcohol graso potenciada en relación con células huésped nativas o no modificadas por ingeniería genética (por ejemplo, mejoras de cepas). La divulgación también proporciona polinucleótidos útiles en las células huésped recombinantes, métodos y composiciones de la divulgación. Sin embargo se reconocerá que no es necesaria la identidad de secuencia absoluta para tales polinucleótidos. Por ejemplo, pueden hacerse cambios en una secuencia de polinucleótido particular el polipéptido codificado puede evaluarse para determinar su actividad. Tales cambios comprenden normalmente mutaciones conservativas y mutaciones silenciosas (por ejemplo, optimización de codones). Pueden examinarse polinucleótidos modificados o mutados (es decir, mutantes) y polipéptidos variantes codificados para detectar una función deseada, tal como, una función mejorada en comparación con el polipéptido parental, incluyendo pero sin limitarse a actividad catalítica aumentada, estabilidad aumentada o inhibición disminuida (por ejemplo, inhibición por retroalimentación disminuida), usando métodos conocidos en la técnica.

35 La divulgación identifica actividades enzimáticas implicadas en diversas etapas (es decir, reacciones) de las rutas de biosíntesis de ácidos grasos descritas en el presente documento según el número de clasificación de enzimas (EC), y proporciona polipéptidos a modo de ejemplo (es decir, enzimas) clasificadas mediante tales números de EC, y polinucleótidos a modo de ejemplo que codifican para tales polipéptidos. Tales polipéptidos y polinucleótidos a modo de ejemplo, que se identifican en el presente documento mediante números de registro y/o números de identificador de secuencia (SEQ ID NO), son útiles para modificar por ingeniería genética rutas de ácidos grasos en células huésped parentales para obtener las células huésped recombinantes descritas en el presente documento. Sin embargo, debe entenderse que los polipéptidos y polinucleótidos descritos en el presente documento son a modo de ejemplo y no limitativos. Las secuencias de homólogos de polipéptidos a modo de ejemplo descritas en el presente documento están disponibles para los expertos en la técnica usando bases de datos (por ejemplo, las bases de datos Entrez proporcionadas por el Centro Nacional para la Información Biotecnológica (NCBI), las bases de datos ExPasy proporcionadas por el Instituto Suizo de Bioinformática, la base de datos BRENDA proporcionada por la Universidad Técnica de Braunschweig, y la base de datos KEGG proporcionada por el Centro de Bioinformática de la Universidad de Kyoto y la Universidad de Tokyo, todas las cuales están disponibles en Internet).

50 Una variedad de células huésped pueden modificarse para contener enzimas de biosíntesis de alcoholes grasos tales como las descritas en el presente documento, dando como resultado células huésped recombinantes adecuadas para la producción de composiciones de alcoholes grasos. Se entiende que una variedad de células pueden proporcionar fuentes de material genético, incluyendo secuencias de polinucleótido que codifican para polipéptidos adecuados para su uso en una célula huésped recombinante proporcionada en el presente documento.

#### Definiciones

55 A menos que se defina otra cosa, todos los términos técnicos y científicos usados en el presente documento tienen el mismo significado que entiende comúnmente un experto habitual en la técnica a la que pertenece la divulgación. Aunque pueden usarse otros métodos y materiales similares, o equivalentes, a los descritos en el presente documento en la práctica de la presente divulgación, los materiales y métodos preferidos se describen en el presente documento. En la descripción y reivindicación de la presente divulgación, se usará la siguiente terminología según las definiciones expuestas a continuación.

Números de registro: Los números de registro de secuencias en la totalidad de esta descripción se obtuvieron de

bases de datos proporcionadas por el NCBI (Centro Nacional para la Información Biotecnológica) mantenidas por los Institutos Nacionales de Salud, EE.UU. (que se identifican en el presente documento como “números de registro de NCBI” o alternativamente como “números de registro de GenBank”), y de las bases de datos UniProt Knowledgebase (UniProtKB) y Swiss-Prot proporcionadas por el Instituto Suizo de Bioinformática (que se identifican en el presente documento como “números de registro de UniProtKB”).

Números de clasificación de enzimas (EC): Se establecen los números de clasificación de enzimas (EC) por el Comité de Nomenclatura de la Unión Internacional de Bioquímica y Biología Molecular (IUBMB), cuya descripción está disponible en el sitio web de Nomenclatura de Enzimas del IUBMB en Internet. Los números EC clasifican las enzimas según la reacción que catalizan.

Tal como se usa en el presente documento, el término “nucleótido” se refiere a una unidad monomérica de un polinucleótido que consiste en una base heterocíclica, un azúcar y uno o más grupos fosfato. Las bases que se producen de manera natural (guanina, (G), adenina, (A), citosina, (C), timina, (T) y uracilo (U)) son normalmente derivados de purina o pirimidina, aunque debe entenderse que también están incluidos análogos de bases que se producen de manera natural y no natural. El azúcar que se produce de manera natural es la pentosa (azúcar de cinco carbonos), la desoxirribosa (que forma ADN) o la ribosa (que forma ARN), aunque debe entenderse que también están incluidos análogos de azúcares que se producen de manera natural y no natural. Los ácidos nucleicos se unen normalmente mediante enlaces fosfato para formar ácidos nucleicos o polinucleótidos, aunque se conocen en la técnica muchas otras uniones (por ejemplo, fosforotioatos, boranofosfatos, y similares).

Tal como se usa en el presente documento, el término “polinucleótido” se refiere a un polímero de ribonucleótidos (ARN) o desoxirribonucleótidos (ADN), que puede ser monocatenario o bicatenario y que puede contener nucleótidos no naturales o alterados. Los términos “polinucleótido”, “secuencia de ácido nucleico” y “secuencia de nucleótidos” se usan de manera intercambiable en el presente documento para referirse a una forma polimérica de nucleótidos de cualquier longitud, o bien ARN o bien ADN. Estos términos se refieren a la estructura primaria de la molécula, y por tanto incluyen ADN bi y monocatenario y ARN bi y monocatenario. Los términos incluyen, como equivalentes, análogos de o bien ARN o bien ADN producidos a partir de análogos de nucleótidos y polinucleótidos modificados tales como, aunque no limitados a polinucleótidos metilados y/o de extremos ocupados. El polinucleótido puede estar en cualquier forma, incluyendo pero sin limitarse a, plásmido, viral, cromosómico, EST, ADNc, ARNm y ARNr.

Tal como se usa en el presente documento, los términos “polipéptido” y “proteína” se usan de manera intercambiable para referirse a un polímero de residuos de aminoácido. El término “polipéptido recombinante” se refiere a un polipéptido que se produce mediante técnicas recombinantes, en las que generalmente se inserta ADN o ARN que codifica para la proteína expresada en un vector de expresión adecuado que se usa a su vez para transformar una célula huésped para producir el polipéptido.

Tal como se usan en el presente documento, los términos “homólogo” y “homólogo/a” se refieren a un polinucleótido o un polipéptido que comprende una secuencia que es idéntica en al menos aproximadamente el 50% a la secuencia de polinucleótido o polipéptido correspondiente. Preferiblemente, los polinucleótidos o polipéptidos homólogos tienen secuencias de polinucleótido o secuencias de aminoácidos que tienen una homología de al menos aproximadamente el 80%, el 81%, el 82%, el 83%, el 84%, el 85%, el 86%, el 87%, el 88%, el 89%, el 90%, el 91%, el 92%, el 93%, el 94%, el 95%, el 96%, el 97%, el 98% o al menos aproximadamente el 99% con la secuencia de aminoácidos o secuencia de polinucleótido correspondiente. Tal como se usa en el presente documento los términos “homología” de secuencia e “identidad” de secuencia se usan de manera intercambiable.

Un experto habitual en la técnica conoce métodos para determinar la homología entre dos o más secuencias. En resumen, pueden realizarse cálculos de “homología” entre dos secuencias de la siguiente manera. Las secuencias se alinean para fines de comparación óptima (por ejemplo, pueden introducirse huecos en una o ambas de una primera y una segunda secuencias de aminoácidos o de ácido nucleico para la alineación óptima y pueden excluirse las secuencias no homólogas para fines de comparación). En una realización preferida, la longitud de una primera secuencia que se alinea para fines de comparación es de al menos aproximadamente el 30%, preferiblemente al menos aproximadamente el 40%, más preferiblemente al menos aproximadamente el 50%, incluso más preferiblemente al menos aproximadamente el 60%, e incluso más preferiblemente al menos aproximadamente el 70%, al menos aproximadamente el 80%, al menos aproximadamente el 90% o aproximadamente el 100% de la longitud de una segunda secuencia. Se comparan entonces los residuos de aminoácido o nucleótidos en posiciones de aminoácido o posiciones de nucleótido correspondientes de las secuencias primera y segunda. Cuando una posición en la primera secuencia está ocupada por el mismo residuo de aminoácido o nucleótido que la posición correspondiente en la segunda secuencia, entonces las moléculas son idénticas en esa posición. El tanto por ciento de homología entre las dos secuencias es una función del número de posiciones idénticas compartidas por las secuencias, teniendo en cuenta el número de huecos y la longitud de cada hueco, que es necesario introducir para la alineación óptima de las dos secuencias.

La comparación de secuencias y determinación del tanto por ciento de homología entre dos secuencias puede lograrse usando un algoritmo matemático, tal como BLAST (Altschul *et al.* J. Mol. Biol. 215(3):403-410 (1990)). El tanto por ciento de homología entre dos secuencias de aminoácidos también puede determinarse usando el

algoritmo de Needleman y Wunsch que se ha incorporado en el programa GAP en el paquete de software de GCG, usando o bien una matriz Blossum 62 o bien una matriz PAM250, y un peso de hueco de 16, 14, 12, 10, 8, 6 ó 4 y un peso de longitud de 1, 2, 3, 4, 5 ó 6 (Needleman y Wunsch (1970) *J. Mol. Biol.* 48:444-453). El tanto por ciento de homología entre dos secuencias de nucleótidos también puede determinarse usando el programa GAP en el paquete de software de GCG, usando una matriz NWSgapdna.CMP y un peso de hueco de 40, 50, 60, 70 u 80 y un peso de longitud de 1, 2, 3, 4, 5 ó 6. Un experto habitual en la técnica puede realizar cálculos de homología inicial y ajustar los parámetros del algoritmo en consecuencia. Un conjunto preferido de parámetros (y el que debe usarse si un profesional no está seguro de qué parámetros deben aplicarse para determinar si una molécula está dentro de una limitación de homología de las reivindicaciones) son una matriz de puntuación Blossum 62 con una penalización por hueco de 12, penalización por extensión de hueco de 4 y penalización por hueco con desplazamiento del marco de lectura de 5. Se conocen métodos adicionales de alineación de secuencias en las técnicas biotecnológicas (véanse, por ejemplo, Rosenberg (2005) *BMC Bioinformatics* 6:278; Altschul *et al.* (2005) *FEBS J.* 272(20):5101-5109).

Tal como se usa en el presente documento, el término “se hibrida en condiciones de rigurosidad baja, rigurosidad media, rigurosidad alta o condiciones de rigurosidad muy alta” describe condiciones para la hibridación y el lavado. Puede encontrarse orientación para realizar reacciones de hibridación en *Current Protocols in Molecular Biology*, John Wiley & Sons, N.Y. (1989), 6.3.1 - 6.3.6. Se describen métodos acuosos y no acuosos en esa referencia y puede usarse cualquier método. Condiciones de hibridación específicas a las que se hace referencia en el presente documento son de la siguiente manera: (1) condiciones de hibridación de rigurosidad baja -- 6X cloruro de sodio/citrato de sodio (SSC) a aproximadamente 45°C, seguido por dos lavados en 0,2X SSC, SDS al 0,1% al menos a 50°C (la temperatura de los lavados puede aumentarse hasta 55°C para condiciones de rigurosidad baja); (2) condiciones de hibridación de rigurosidad media -- 6X SSC a aproximadamente 45°C, seguido por uno o más lavados en 0,2X SSC, SDS al 0,1% a 60°C; (3) condiciones de hibridación de rigurosidad alta -- 6X SSC a aproximadamente 45°C, seguido por uno o más lavados en 0,2X SSC, SDS al 0,1% a 65°C; y (4) condiciones de hibridación de rigurosidad muy alta -- fosfato de sodio 0,5 M, 7% SDS a 65°C, seguido por uno o más lavados a 0,2X SSC, SDS al 1% a 65°C. Las condiciones de rigurosidad muy alta (4) son las condiciones preferidas a menos que se especifique de otro modo.

Un polipéptido “endógeno” se refiere a un polipéptido codificado por el genoma de la célula parental (también denominada “célula huésped”) a partir de la cual se modifica por ingeniería genética la célula recombinante (o “derivada”).

Un polipéptido “exógeno” se refiere a un polipéptido que no está codificado por el genoma de la célula parental. Un polipéptido (es decir, mutante) es un ejemplo de un polipéptido exógeno.

El término “heterólogo” significa en general derivados de una especie diferente o derivados de un organismo diferente. Tal como se usa en el presente documento, se refiere a una secuencia de nucleótidos o una secuencia de polinucleótido que no está presente de manera natural en un organismo particular. Expresión heteróloga significa que una proteína o polipéptido se añade experimentalmente a una célula que no expresa normalmente esa proteína. Como tal, heterólogo se refiere al hecho de que una proteína transferida se derivó inicialmente a partir de un tipo celular diferente o una especie diferente del receptor. Por ejemplo, puede introducirse una secuencia de polinucleótido endógena para una célula vegetal en una célula huésped bacteriana mediante métodos recombinantes, y el polinucleótido vegetal es entonces un polinucleótido heterólogo en una célula huésped bacteriana recombinante.

Tal como se usa en el presente documento, el término “fragmento” de un polipéptido se refiere a una parte más corta de un polipéptido o una proteína de longitud completa cuyo tamaño oscila entre cuatro residuos de aminoácido y toda la secuencia de aminoácidos menos un residuo de aminoácido. En determinadas realizaciones de la divulgación, un fragmento se refiere a toda la secuencia de aminoácidos de un dominio de un polipéptido o una proteína (por ejemplo, un dominio de unión a sustrato o un dominio catalítico).

Tal como se usa en el presente documento, el término “mutagénesis” se refiere a un procedimiento mediante el cual se cambia la información genética de un organismo de manera estable. La mutagénesis de una secuencia de ácido nucleico que codifica para proteína produce una proteína mutante. Mutagénesis también se refiere a cambios en secuencias de ácido nucleico no codificantes que dan como resultado actividad de proteína modificada.

Tal como se usa en el presente documento, el término “gen” se refiere a secuencias de ácido nucleico que codifican para o bien un producto de ARN o bien un producto de proteína, así como secuencias de ácido nucleico operativamente unidas que afectan a la expresión del ARN o la proteína (por ejemplo, tales secuencias incluyen pero no se limitan a secuencias promotoras o potenciadoras) o secuencias de ácido nucleico operativamente unidas que codifican para secuencias que afectan a la expresión del ARN o la proteína (por ejemplo, tales secuencias incluyen pero no se limitan a sitios de unión al ribosoma o secuencias de control de la traducción).

Se conocen en la técnica secuencias de control de la expresión e incluyen, por ejemplo, promotores, potenciadores, señales de poliadenilación, terminadores de la transcripción, sitios internos de entrada al ribosoma (IRES), y similares, que proporcionan la expresión de la secuencia de polinucleótido en una célula huésped. Las secuencias

de control de la expresión interaccionan específicamente con proteínas celulares implicadas en la transcripción (Maniatis *et al.* (1987) *Science* 236:1237-1245 (1987)). Se describen secuencias de control de la expresión a modo de ejemplo, por ejemplo, en Goeddel, *Gene Expression Technology: Methods in Enzymology*, vol. 185, Academic Press, San Diego, Calif. (1990).

- 5 En los métodos de la divulgación, una secuencia de control de la expresión está operativamente unida a una secuencia de polinucleótido. Por “operativamente unida” quiere decirse que una secuencia de polinucleótido y una(s) secuencia(s) de control de la expresión se conectan de tal manera que se permita la expresión génica cuando se unen las moléculas apropiadas (por ejemplo, proteínas activadoras de la transcripción) a la(s) secuencia(s) de control de la expresión. Están ubicados promotores operativamente unidos en el sentido de 5' de la secuencia de polinucleótido seleccionada en cuanto a la dirección de transcripción y traducción. Pueden estar ubicados potenciadores operativamente unidos en el sentido de 5', dentro de o en el sentido de 3' del polinucleótido seleccionado.

15 Tal como se usa en el presente documento, el término “vector” se refiere a una molécula de ácido nucleico que puede transportar otro ácido nucleico, es decir, una secuencia de polinucleótido, a la que se ha unido. Un tipo de vector útil es un episoma (es decir, un ácido nucleico que puede dar replicación extracromosómica). Los vectores útiles son aquellos que pueden dar replicación y/o expresión autónoma de ácidos nucleicos a los que se unen. Los vectores que pueden dirigir la expresión de genes a los que se unen operativamente se denominan en el presente documento “vectores de expresión”. En general, los vectores de expresión de utilidad en técnicas de ADN recombinante están a menudo en forma de “plásmidos”, que se refieren en general a bucles de ADN bicatenario circular que, en su forma de vector, no se unen al cromosoma. Los términos “plásmido” y “vector” se usan de manera intercambiable en el presente documento, en la medida en que un plásmido es la forma más comúnmente usada de vector. Sin embargo, también se incluyen otras de tales formas de vectores de expresión que sirven para funciones equivalentes y que se han conocido en la técnica posteriormente al presente documento. En algunas realizaciones, vector recombinante comprende al menos una secuencia que incluye (a) una secuencia de control de la expresión acoplada operativamente a la secuencia de polinucleótido; (b) un marcador de selección acoplado operativamente a la secuencia de polinucleótido; (c) una secuencia de marcador acoplada operativamente a la secuencia de polinucleótido; (d) un resto de purificación acoplado operativamente a la secuencia de polinucleótido; (e) una secuencia de secreción acoplada operativamente a la secuencia de polinucleótido; y (f) una secuencia de selección como diana acoplada operativamente a la secuencia de polinucleótido. Los vectores de expresión descritos en el presente documento incluyen una secuencia de polinucleótido descrita en el presente documento en una forma adecuada para la expresión de la secuencia de polinucleótido en una célula huésped. Los expertos en la técnica apreciarán que el diseño del vector de expresión puede depender de factores tales como la elección de la célula huésped que va transformarse, el nivel de expresión de polipéptido deseado, etc. Los vectores de expresión descritos en el presente documento pueden introducirse en células huésped para producir polipéptidos, incluyendo polipéptidos de fusión, codificados por las secuencias de polinucleótido tal como se describe en el presente documento.

La expresión de genes que codifican para polipéptidos en procariotas, por ejemplo, *E. coli*, se lleva a cabo con la mayor frecuencia con vectores que contienen promotores constitutivos o inducibles que dirigen la expresión de polipéptidos o bien de fusión o bien de no fusión. Los vectores de fusión añaden un número de aminoácidos a un polipéptido codificado en el mismo, habitualmente al extremo amino-terminal o carboxi-terminal del polipéptido recombinante. Tales vectores de fusión sirven normalmente para uno o más de los tres propósitos siguientes: (1) aumentar expresión del polipéptido recombinante; (2) aumentar la solubilidad del polipéptido recombinante; y (3) ayudar en la purificación del polipéptido recombinante actuando como ligando en purificación por afinidad. A menudo, en los vectores de expresión de fusión, se introduce un sitio de escisión proteolítica en la unión del resto de fusión y el polipéptido recombinante. Esto permite la separación del polipéptido recombinante del resto de fusión después de la purificación del polipéptido de fusión. En determinadas realizaciones, una secuencia de polinucleótido de la divulgación está operativamente unida a un promotor derivado del bacteriófago T5. En determinadas realizaciones, la célula huésped es una célula de levadura, y el vector de expresión es un vector de expresión de levadura. Los ejemplos de vectores para la expresión en la levadura *S. cerevisiae* incluyen pYepSec1 (Baldari *et al.* (1987) *EMBO J.* 6:229-234 (1987)); pMFa (Kurjan *et al.* (1982) *Cell* 30:933-943 (1982)); pJRY88 (Schultz *et al.* (1987) *Gene* 54: 113-123 (1987)); pYES2 (Invitrogen Corp., San Diego, CA) y picZ (Invitrogen Corp., San Diego, CA). En otras realizaciones, la célula huésped es una célula de insecto, y el vector de expresión es un vector de expresión de baculovirus. Los vectores de baculovirus disponibles para la expresión de proteínas en células de insecto en cultivo (por ejemplo, células Sf9) incluyen, por ejemplo, la serie de pAc (Smith *et al.* *Mol. Cell Biol.* 3:2156-2165 (1983)) y la serie de pVL (Lucklow *et al.* (1989) *Virology* 170:31-39 (1989)). En aún otra realización, las secuencias de polinucleótido descritas en el presente documento pueden expresarse en células de mamífero usando un vector de expresión de mamífero. Se conocen bien en la técnica otros sistemas de expresión adecuados para células tanto procariotas como eucariotas; véase, por ejemplo, Sambrook *et al.*, “Molecular Cloning: A Laboratory Manual”, segunda edición, Cold Spring Harbor Laboratory, (1989).

60 Tal como se usa en el presente documento, el término “acil-CoA” se refiere a un acil-tioéster formado entre el carbono de carbonilo de la cadena de alquilo y el grupo sulfhidrilo del resto 4'-fosfopantetionilo de la coenzima A (CoA), que tiene la fórmula R-C(O)S-CoA, en la que R es cualquier grupo alquilo que tiene al menos 4 átomos de carbono.

- 5 Tal como se usa en el presente documento, "acil-ACP" se refiere a un acil-tioéster formado entre el carbono de carbonilo de una cadena de alquilo y el grupo sulfhidrilo del resto fosfopanteteinilo de una proteína transportadora de acilo (ACP). El resto fosfopanteteinilo se une de manera postraduccional a un residuo de serina conservado en la ACP mediante la acción de la proteína transportadora de holo-acilo sintasa (ACPS), una fosfopanteteinilo transferasa. En algunas realizaciones, una acil-ACP es un producto intermedio en la síntesis de acil-ACP totalmente saturadas. En otras realizaciones, una acil-ACP es un producto intermedio en la síntesis de acil-ACP insaturadas. En algunas realizaciones, la cadena de carbonos tendrá aproximadamente 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25 ó 26 carbonos. Cada una de estas acil-ACP son sustratos para enzimas que las convierten en derivados de ácido graso.
- 10 Tal como se usa en el presente documento, el término "ácido graso o derivado del mismo" significa un "ácido graso" o un "derivado de ácido graso". El término "ácido graso" significa un ácido carboxílico que tiene la fórmula RCOOH. R representa un grupo alifático, preferiblemente un grupo alquilo. R puede comprender entre aproximadamente 4 y aproximadamente 22 átomos de carbono. Los ácidos grasos pueden ser saturados, monoinsaturados o poliinsaturados. En una realización preferida, el ácido graso se prepara a partir de una ruta de biosíntesis de ácidos grasos. El término "derivado de ácido graso" significa productos producidos en parte a partir de la ruta de biosíntesis de ácidos grasos del organismo huésped de producción. "Derivado de ácido graso" también incluye productos producidos en parte a partir de acil-ACP o derivados de acil-ACP. Los derivados de ácido graso a modo de ejemplo incluyen, por ejemplo, acil-CoA, aldehídos grasos, alcoholes de cadena corta y larga, hidrocarburos y ésteres (por ejemplo, ceras, ésteres de ácidos grasos o ésteres grasos).
- 15 Tal como se usa en el presente documento, el término "ruta de biosíntesis de ácidos grasos" significa una ruta de biosíntesis que produce derivados de ácido graso, por ejemplo alcoholes grasos. La ruta de biosíntesis de ácidos grasos incluye ácido graso sintasas que pueden modificarse por ingeniería genética para producir ácidos grasos y en algunas realizaciones pueden expresarse con enzimas adicionales para producir derivados de ácido graso, tales como alcoholes grasos que tienen características deseadas.
- 20 Tal como se usa en el presente documento, "aldehído graso" significa un aldehído que tiene la fórmula RCHO caracterizada por un grupo carbonilo (C=O). En algunas realizaciones, el aldehído graso es cualquier aldehído producido a partir de un alcohol graso. En determinadas realizaciones, el grupo R tiene al menos 5, al menos 6, al menos 7, al menos 8, al menos 9, al menos 10, al menos 11, al menos 12, al menos 13, al menos 14, al menos 15, al menos 16, al menos 17, al menos 18 o al menos 19, carbonos de longitud. Alternativamente, o además, el grupo R tiene 20 o menos, 19 o menos, 18 o menos, 17 o menos, 16 o menos, 15 o menos, 14 o menos, 13 o menos, 12 o menos, 11 o menos, 10 o menos, 9 o menos, 8 o menos, 7 o menos o 6 o menos carbonos de longitud. Por tanto, el grupo R puede tener un grupo R delimitado por dos cualesquiera de los extremos anteriores. Por ejemplo, el grupo R puede tener 6-16 carbonos de longitud, 10-14 carbonos de longitud o 12-18 carbonos de longitud. En algunas realizaciones, el aldehído graso es un aldehído graso C<sub>6</sub>, C<sub>7</sub>, C<sub>8</sub>, C<sub>9</sub>, C<sub>10</sub>, C<sub>11</sub>, C<sub>12</sub>, C<sub>13</sub>, C<sub>14</sub>, C<sub>15</sub>, C<sub>16</sub>, C<sub>17</sub>, C<sub>18</sub>, C<sub>19</sub>, C<sub>20</sub>, C<sub>21</sub>, C<sub>22</sub>, C<sub>23</sub>, C<sub>24</sub>, C<sub>25</sub>, o uno C<sub>26</sub>. En determinadas realizaciones, el aldehído graso es un aldehído graso C<sub>6</sub>, C<sub>8</sub>, C<sub>10</sub>, C<sub>12</sub>, C<sub>13</sub>, C<sub>14</sub>, C<sub>15</sub>, C<sub>16</sub>, C<sub>17</sub> o C<sub>18</sub>.
- 25 Tal como se usa en el presente documento, "alcohol graso" significa un alcohol que tiene la fórmula ROH. En algunas realizaciones, el grupo R tiene al menos 5, al menos 6, al menos 7, al menos 8, al menos 9, al menos 10, al menos 11, al menos 12, al menos 13, al menos 14, al menos 15, al menos 16, al menos 17, al menos 18 o al menos 19, carbonos de longitud. Alternativamente, o además, el grupo R tiene 20 o menos, 19 o menos, 18 o menos, 17 o menos, 16 o menos, 15 o menos, 14 o menos, 13 o menos, 12 o menos, 11 o menos, 10 o menos, 9 o menos, 8 o menos, 7 o menos o 6 o menos carbonos de longitud. Por tanto, el grupo R puede tener un grupo R delimitado por dos cualesquiera de los extremos anteriores. Por ejemplo, el grupo R puede tener 6-16 carbonos de longitud, 10-14 carbonos de longitud o 12-18 carbonos. En algunas realizaciones, el alcohol graso es un alcohol graso C<sub>6</sub>, C<sub>7</sub>, C<sub>8</sub>, C<sub>9</sub>, C<sub>10</sub>, C<sub>11</sub>, C<sub>12</sub>, C<sub>13</sub>, C<sub>14</sub>, C<sub>15</sub>, C<sub>16</sub>, C<sub>17</sub>, C<sub>18</sub>, C<sub>19</sub>, C<sub>20</sub>, C<sub>21</sub>, C<sub>22</sub>, C<sub>23</sub>, C<sub>24</sub>, C<sub>25</sub>, o uno C<sub>26</sub>. En determinadas realizaciones, el alcohol graso es un alcohol graso C<sub>6</sub>, C<sub>8</sub>, C<sub>10</sub>, C<sub>12</sub>, C<sub>13</sub>, C<sub>14</sub>, C<sub>15</sub>, C<sub>16</sub>, C<sub>17</sub> o C<sub>18</sub>.
- 30 Una "composición de alcohol graso" tal como se hace referencia en el presente documento se produce por una célula huésped recombinante y comprende normalmente una mezcla de alcoholes grasos. En algunos casos, la mezcla incluye más de un tipo de producto (por ejemplo, alcoholes grasos y ácidos grasos). En otros casos, las composiciones de derivado de ácido graso pueden comprender, por ejemplo, una mezcla de alcoholes grasos con diversas longitudes de cadena y características de saturación o ramificación. En todavía otros casos, la composición de alcohol graso comprende una mezcla de tanto más de un tipo de producto como de productos con diversas longitudes de cadena y características de saturación o ramificación.
- 35 Una célula huésped modificada por ingeniería para producir un aldehído graso convertirá normalmente parte del aldehído graso en un alcohol graso. Cuando una célula huésped que produce alcoholes grasos se modifica por ingeniería genética para expresar un polinucleótido que codifica para una éster sintasa, se producen ésteres de cera. En una realización, se producen alcoholes grasos a partir de una ruta de biosíntesis de ácidos grasos. Como ejemplo, puede convertirse acil-ACP en ácidos grasos por medio de la acción de una tioesterasa (por ejemplo, TesA de *E. coli*), que se convierten en aldehídos grasos y alcoholes grasos por medio de la acción de una ácido carboxílico reductasa (por ejemplo, CarB de *E. coli*). La conversión de aldehídos grasos en alcoholes grasos puede verse facilitada además, por ejemplo, por medio de la acción de un polipéptido de biosíntesis de alcoholes grasos.
- 40
- 45
- 50
- 55
- 60

En algunas realizaciones, un gen que codifica para un polipéptido de biosíntesis de alcoholes grasos se expresa o sobreexpresa en la célula huésped. En determinadas realizaciones, el polipéptido de biosíntesis de alcoholes grasos tiene actividad aldehído reductasa o alcohol deshidrogenasa. Los ejemplos de polipéptidos de alcohol deshidrogenasa útiles según la divulgación incluyen, pero no se limitan a AlrA de *Acinetobacter* sp. M-1 (SEQ ID NO: 3) u homólogos de AlrA, tales como AlrAadp1 (SEQ ID NO: 4) y alcohol deshidrogenasas de *E. coli* endógenas tales como YjgB, (AAC77226) (SEQ ID NO: 5), DkgA (NP\_417485), DkgB (NP\_414743), YdjL (AAC74846), YdjJ (NP\_416288), AdhP (NP\_415995), YhdH (NP\_417719), YahK (NP\_414859), YphC (AAC75598), YqhD (446856) e YbbO [AAC73595.1]. Se describen ejemplos adicionales en las publicaciones de solicitud de patente internacional n.ºs WO2007/136762, WO2008/119082 y WO2010/062480. En determinadas realizaciones, el polipéptido de biosíntesis de alcoholes grasos tiene actividad aldehído reductasa o alcohol deshidrogenasa (EC 1.1.1.1).

Tal como se usa en el presente documento, el término “alcohol deshidrogenasa” se refiere a un polipéptido que puede catalizar la conversión de un aldehído graso en un alcohol (por ejemplo, alcohol graso). Un experto habitual en la técnica apreciará que determinadas alcohol deshidrogenasas pueden catalizar otras reacciones también, y estas alcohol deshidrogenasas no específicas se abarcan también por el término “alcohol deshidrogenasa”. El grupo R de un ácido graso, aldehído graso o alcohol graso puede ser una cadena lineal o una cadena ramificada. Las cadenas ramificadas pueden tener más de un punto de ramificación y pueden incluir ramificaciones cíclicas. En algunas realizaciones, el ácido graso ramificado, aldehído graso ramificado o alcohol graso ramificado es un ácido graso ramificado, aldehído graso ramificado o alcohol graso ramificado C<sub>6</sub>, C<sub>7</sub>, C<sub>8</sub>, C<sub>9</sub>, C<sub>10</sub>, C<sub>11</sub>, C<sub>12</sub>, C<sub>13</sub>, C<sub>14</sub>, C<sub>15</sub>, C<sub>16</sub>, C<sub>17</sub>, C<sub>18</sub>, C<sub>19</sub>, C<sub>20</sub>, C<sub>21</sub>, C<sub>22</sub>, C<sub>23</sub>, C<sub>24</sub>, C<sub>25</sub>, o uno C<sub>26</sub>. En realizaciones particulares, el ácido graso ramificado, aldehído graso ramificado o alcohol graso ramificado es un ácido graso ramificado, aldehído graso ramificado o alcohol graso ramificado C<sub>6</sub>, C<sub>8</sub>, C<sub>10</sub>, C<sub>12</sub>, C<sub>13</sub>, C<sub>14</sub>, C<sub>15</sub>, C<sub>16</sub>, C<sub>17</sub> o C<sub>18</sub>. En determinadas realizaciones, el grupo hidroxilo del ácido graso ramificado, aldehído graso ramificado o alcohol graso ramificado está en la posición primaria (C<sub>1</sub>). En determinadas realizaciones, el ácido graso ramificado, aldehído graso ramificado o alcohol graso ramificado es un ácido graso iso, aldehído graso iso o alcohol graso iso, o un ácido graso anteiso, un aldehído graso anteiso o alcohol graso anteiso. En realizaciones a modo de ejemplo, el ácido graso ramificado, aldehído graso ramificado o alcohol graso ramificado se selecciona de ácido graso ramificado, aldehído graso ramificado o alcohol graso ramificado iso-C<sub>7:0</sub>, iso-C<sub>8:0</sub>, iso-C<sub>9:0</sub>, iso-C<sub>10:0</sub>, iso-C<sub>11:0</sub>, iso-C<sub>12:0</sub>, iso-C<sub>13:0</sub>, iso-C<sub>14:0</sub>, iso-C<sub>15:0</sub>, iso-C<sub>16:0</sub>, iso-C<sub>17:0</sub>, iso-C<sub>18:0</sub>, iso-C<sub>19:0</sub>, anteiso-C<sub>7:0</sub>, anteiso-C<sub>8:0</sub>, anteiso-C<sub>9:0</sub>, anteiso-C<sub>10:0</sub>, anteiso-C<sub>11:0</sub>, anteiso-C<sub>12:0</sub>, anteiso-C<sub>13:0</sub>, anteiso-C<sub>14:0</sub>, anteiso-C<sub>15:0</sub>, anteiso-C<sub>16:0</sub>, anteiso-C<sub>17:0</sub>, anteiso-C<sub>18:0</sub> y anteiso-C<sub>19:0</sub>. El grupo R de un ácido graso ramificado o no ramificado, aldehído graso ramificado o no ramificado o alcohol graso ramificado o no ramificado puede ser saturado o insaturado. Si es insaturado, el grupo R puede tener uno o más de un punto de insaturación. En algunas realizaciones, el ácido graso insaturado, aldehído graso insaturado o alcohol graso insaturado es un ácido graso monoinsaturado, aldehído graso monoinsaturado o alcohol graso monoinsaturado. En determinadas realizaciones, el ácido graso insaturado, aldehído graso insaturado o alcohol graso insaturado es un ácido graso insaturado, aldehído graso insaturado o alcohol graso insaturado C<sub>6:1</sub>, C<sub>7:1</sub>, C<sub>8:1</sub>, C<sub>9:1</sub>, C<sub>10:1</sub>, C<sub>11:1</sub>, C<sub>12:1</sub>, C<sub>13:1</sub>, C<sub>14:1</sub>, C<sub>15:1</sub>, C<sub>16:1</sub>, C<sub>17:1</sub>, C<sub>18:1</sub>, C<sub>19:1</sub>, C<sub>20:1</sub>, C<sub>21:1</sub>, C<sub>22:1</sub>, C<sub>23:1</sub>, C<sub>24:1</sub>, C<sub>25:1</sub>, o uno C<sub>26:1</sub>. En determinadas realizaciones preferidas, el ácido graso insaturado, aldehído graso insaturado o alcohol graso insaturado es C<sub>10:1</sub>, C<sub>12:1</sub>, C<sub>14:1</sub>, C<sub>16:1</sub> o C<sub>18:1</sub>. En aún otras realizaciones, el ácido graso insaturado, aldehído graso insaturado o alcohol graso insaturado es insaturado en la posición omega-7. En determinadas realizaciones, el ácido graso insaturado, aldehído graso insaturado o alcohol graso insaturado comprende un doble enlace cis.

Tal como se usa en el presente documento, una “célula huésped” recombinante o modificada por ingeniería genética es una célula huésped, por ejemplo, un microorganismo que se ha modificado de manera que produce alcoholes grasos. En algunas realizaciones, la célula huésped recombinante comprende uno o más polinucleótidos, codificando cada polinucleótido para un polipéptido que tiene actividad enzimática de biosíntesis de aldehídos grasos y/o alcoholes grasos, en la que la célula huésped recombinante produce una composición de alcohol graso cuando se cultiva en presencia de una fuente de carbono en condiciones eficaces para expresar los polinucleótidos.

Tal como se usa en el presente documento, el término “clon” se refiere normalmente a una célula o un grupo de células que descienden de y esencialmente idénticas genéticamente a un único ancestro común, por ejemplo, las bacterias de una colonia bacteriana clonada surgida a partir de una única célula bacteriana.

Tal como se usa en el presente documento, el término “cultivo” se refiere normalmente a un medio líquido que comprende células viables. En una realización, un cultivo comprende células que se reproducen en medios de cultivo predeterminados en condiciones controladas, por ejemplo, un cultivo de células huésped recombinantes hechas crecer en medios líquidos que comprenden una fuente de carbono y nitrógeno seleccionada. “Cultivar” o “cultivo” se refieren al crecimiento de una población de células de células microbianas en condiciones adecuadas en un medio líquido o sólido. En realizaciones particulares, cultivar se refiere a la bioconversión fermentativa de un sustrato en un producto final. Se conocen bien medios de cultivo y los componentes individuales de tales medios de cultivo están disponibles de fuentes comerciales, por ejemplo, con las marcas comerciales DIFCO<sup>TM</sup> y BBL<sup>TM</sup>. En un ejemplo no limitativo, el medio de nutrientes acuoso es un “medio rico” que comprende fuentes complejas de nitrógeno, sales y carbono, tal como el medio YP, que comprende 10 g/l de peptona y 10 g/l de extracto de levadura de un medio de este tipo. La célula huésped puede modificarse por ingeniería genética adicionalmente para asimilar el carbono eficazmente y usar materiales celulósicos como fuentes de carbono según los métodos descritos, por ejemplo, en las patentes estadounidenses 5.000.000; 5.028.539; 5.424.202; 5.482.846; 5.602.030 y el documento

WO2010127318. Además, la célula huésped puede modificarse por ingeniería para expresar una invertasa de modo que pueda usarse sacarosa como fuente de carbono.

Tal como se usa en el presente documento, el término “en condiciones eficaces para expresar dichas secuencias de nucleótidos heterólogas” significa cualesquiera condiciones que permiten que una célula huésped produzca un aldehído graso o alcohol graso deseado. Las condiciones adecuadas incluyen, por ejemplo, condiciones de fermentación.

Tal como se usa en el presente documento, un “nivel modificado” o “alterado de” actividad de una proteína, por ejemplo una enzima, en una célula huésped recombinante se refiere a una diferencia en una o más características en la actividad determinadas con relación a la célula huésped parental o nativa. Normalmente se determinan diferencias en actividad entre una célula huésped recombinante, que tiene actividad modificada, y la célula huésped silvestre correspondiente (por ejemplo, comparación de un cultivo de una célula huésped recombinante con relación a la célula huésped silvestre). Actividades modificadas pueden ser el resultado de, por ejemplo, cantidades modificadas de proteína expresada por una célula huésped recombinante (por ejemplo, como resultado de un aumento o una disminución del número de copias de secuencias de ADN que codifican para la proteína, un aumento o una disminución del número de transcritos de ARNm que codifican para la proteína, y/o un aumento o una disminución de las cantidades de traducción de proteína de la proteína a partir del ARNm); cambios en la estructura de la proteína (por ejemplo, cambios en la estructura primaria, tales como cambios en la secuencia codificante de la proteína que dan como resultado cambios en la especificidad de sustrato, cambios en parámetros cinéticos observados); y cambios en la estabilidad de la proteína (por ejemplo, un aumento o una disminución de la degradación de la proteína). En algunas realizaciones, el polipéptido es un mutante o una variante de cualquiera de los polipéptidos descritos en el presente documento. En determinados casos, las secuencias codificantes para los polipéptidos tal como se describe en el presente documento se someten a optimización de codones para la expresión en una célula huésped particular. Por ejemplo, para la expresión en *E. coli*, pueden optimizarse uno o más codones tal como se describe en, por ejemplo Grosjean *et al.* Gene 18:199-209 (1982).

El término “secuencias reguladoras” tal como se usa en el presente documento se refiere normalmente a una secuencia de bases en ADN, operativamente unida a secuencias de ADN que codifican para una proteína que controla en última instancia la expresión de la proteína. Los ejemplos de secuencias reguladoras incluyen, pero no se limitan a, secuencias promotoras de ARN, secuencias de unión a factor de transcripción, secuencias de terminación de la transcripción, moduladores de la transcripción (tales como elementos potenciadores), secuencias de nucleótidos que afectan a la estabilidad del ARN y secuencias reguladoras de la traducción (tales como sitios de unión al ribosoma (por ejemplo, secuencias de Shine-Dalgarno en procariotas o secuencias de Kozak en eucariotas), codones de iniciación, codones de terminación).

Tal como se usa en el presente documento, la frase “la expresión de dicha secuencia de nucleótidos se modifica con relación a la secuencia de nucleótidos silvestre”, significa un aumento o una disminución en el nivel de expresión y/o actividad de una secuencia de nucleótidos endógena o la expresión y/o actividad de una secuencia de nucleótidos que codifica para polipéptido heterólogo o no nativo. Tal como se usa en el presente documento, el término “sobrexpresar” significa expresar o hacer que se exprese un polinucleótido o polipéptido en una célula a una mayor concentración de lo que se expresa normalmente en una célula silvestre correspondiente en las mismas condiciones.

Los términos “nivel alterado de expresión” y “nivel modificado de expresión” se usan de manera intercambiable y significan que está presente un polinucleótido, polipéptido o hidrocarburo en una concentración diferente en una célula huésped modificada por ingeniería genética en comparación con su concentración en una célula silvestre correspondiente en las mismas condiciones.

Tal como se usa en el presente documento, el término “título” se refiere a la cantidad de aldehído graso o alcohol graso producido por volumen unitario de cultivo de células huésped. En cualquier aspecto de las composiciones y los métodos descritos en el presente documento, un alcohol graso se produce a un título de aproximadamente 25 mg/l, aproximadamente 50 mg/l, aproximadamente 75 mg/l, aproximadamente 100 mg/l, aproximadamente 125 mg/l, aproximadamente 150 mg/l, aproximadamente 175 mg/l, aproximadamente 200 mg/l, aproximadamente 225 mg/l, aproximadamente 250 mg/l, aproximadamente 275 mg/l, aproximadamente 300 mg/l, aproximadamente 325 mg/l, aproximadamente 350 mg/l, aproximadamente 375 mg/l, aproximadamente 400 mg/l, aproximadamente 425 mg/l, aproximadamente 450 mg/l, aproximadamente 475 mg/l, aproximadamente 500 mg/l, aproximadamente 525 mg/l, aproximadamente 550 mg/l, aproximadamente 575 mg/l, aproximadamente 600 mg/l, aproximadamente 625 mg/l, aproximadamente 650 mg/l, aproximadamente 675 mg/l, aproximadamente 700 mg/l, aproximadamente 725 mg/l, aproximadamente 750 mg/l, aproximadamente 775 mg/l, aproximadamente 800 mg/l, aproximadamente 825 mg/l, aproximadamente 850 mg/l, aproximadamente 875 mg/l, aproximadamente 900 mg/l, aproximadamente 925 mg/l, aproximadamente 950 mg/l, aproximadamente 975 mg/l, aproximadamente 1000 mg/l, aproximadamente 1050 mg/l, aproximadamente 1075 mg/l, aproximadamente 1100 mg/l, aproximadamente 1125 mg/l, aproximadamente 1150 mg/l, aproximadamente 1175 mg/l, aproximadamente 1200 mg/l, aproximadamente 1225 mg/l, aproximadamente 1250 mg/l, aproximadamente 1275 mg/l, aproximadamente 1300 mg/l, aproximadamente 1325 mg/l, aproximadamente 1350 mg/l, aproximadamente 1375 mg/l, aproximadamente 1400 mg/l, aproximadamente 1425 mg/l, aproximadamente 1450 mg/l, aproximadamente 1475 mg/l,

aproximadamente 1500 mg/l, aproximadamente 1525 mg/l, aproximadamente 1550 mg/l, aproximadamente 1575 mg/l, aproximadamente 1600 mg/l, aproximadamente 1625 mg/l, aproximadamente 1650 mg/l, aproximadamente 1675 mg/l, aproximadamente 1700 mg/l, aproximadamente 1725 mg/l, aproximadamente 1750 mg/l, aproximadamente 1775 mg/l, aproximadamente 1800 mg/l, aproximadamente 1825 mg/l, aproximadamente 1850 mg/l, aproximadamente 1875 mg/l, aproximadamente 1900 mg/l, aproximadamente 1925 mg/l, aproximadamente 1950 mg/l, aproximadamente 1975 mg/l, aproximadamente 2000 mg/l (2 g/l), 3 g/l, 5 g/l, 10 g/l, 20 g/l, 30 g/l, 40 g/l, 50 g/l, 60 g/l, 70 g/l, 80 g/l, 90 g/l, 100 g/l o un intervalo delimitado por dos cualesquiera de los valores anteriores. En otras realizaciones, se produce un aldehído graso o alcohol graso a un título de más de 100 g/l, más de 200 g/l, más de 300 g/l, o mayor, tal como 500 g/l, 700 g/l, 1000 g/l, 1200 g/l, 1500 g/l o 2000 g/l. El título preferido de aldehído graso o alcohol graso producido por una célula huésped recombinante según los métodos de la divulgación es de desde 5 g/l hasta 200 g/l, de 10 g/l a 150 g/l, de 20 g/l a 120 g/l y de 30 g/l a 100 g/l.

Tal como se usa en el presente documento, el término “rendimiento del aldehído graso o alcohol graso producido por una célula huésped” se refiere a la eficacia mediante la que una fuente de carbono de entrada se convierte en producto (es decir, alcohol graso o aldehído graso) en una célula huésped. Células huésped modificadas por ingeniería genética para producir alcoholes grasos y/o aldehídos grasos según los métodos de la divulgación tienen un rendimiento de al menos el 3%, al menos el 4%, al menos el 5%, al menos el 6%, al menos el 7%, al menos el 8%, al menos el 9%, al menos el 10%, al menos el 11%, al menos el 12%, al menos el 13%, al menos el 14%, al menos el 15%, al menos el 16%, al menos el 17%, al menos el 18%, al menos el 19%, al menos el 20%, al menos el 21%, al menos el 22%, al menos el 23%, al menos el 24%, al menos el 25%, al menos el 26%, al menos el 27%, al menos el 28%, al menos el 29%, o al menos el 30% o un intervalo delimitado por dos cualesquiera de los valores anteriores. En otras realizaciones, se produce un aldehído graso o alcohol graso con un rendimiento de más del 30%, el 40%, el 50%, el 60%, el 70%, el 80%, el 90% o más. Alternativamente, o además, el rendimiento es de aproximadamente el 30% o menos, aproximadamente el 27% o menos, aproximadamente el 25% o menos o aproximadamente el 22% o menos. Por tanto, el rendimiento puede estar delimitado por dos cualesquiera de los extremos anteriores. Por ejemplo, el rendimiento del alcohol graso o aldehído graso producido por la célula huésped recombinante según los métodos de la divulgación puede ser del 5% al 15%, del 10% al 25%, del 10% al 22%, del 15% al 27%, del 18% al 22%, del 20% al 28% o del 20% al 30%. El rendimiento preferido de alcohol graso producido por la célula huésped recombinante según los métodos de la divulgación es de desde el 10% hasta el 30%.

Tal como se usa en el presente documento, el término “productividad” se refiere a la cantidad de aldehído graso o alcohol graso producido por volumen unitario de cultivo de células huésped por tiempo unitario. En cualquier aspecto de las composiciones y los métodos descritos en el presente documento, la productividad de aldehído graso o alcohol graso producido por una célula huésped recombinante es de al menos 100 mg/l/hora, al menos 200 mg/l/hora, al menos 300 mg/l/hora, al menos 400 mg/l/hora, al menos 500 mg/l/hora, al menos 600 mg/l/hora, al menos 700 mg/l/hora, al menos 800 mg/l/hora, al menos 900 mg/l/hora, al menos 1000 mg/l/hora, al menos 1100 mg/l/hora, al menos 1200 mg/l/hora, al menos 1300 mg/l/hora, al menos 1400 mg/l/hora, al menos 1500 mg/l/hora, al menos 1600 mg/l/hora, al menos 1700 mg/l/hora, al menos 1800 mg/l/hora, al menos 1900 mg/l/hora, al menos 2000 mg/l/hora, al menos 2100 mg/l/hora, al menos 2200 mg/l/hora, al menos 2300 mg/l/hora, al menos 2400 mg/l/hora o al menos 2500 mg/l/hora. Alternativamente, o además, la productividad es de 2500 mg/l/hora o menos, 2000 mg/l/DO<sub>600</sub> o menos, 1500 mg/l/DO<sub>600</sub> o menos, 120 mg/l/hora o menos, 1000 mg/l/hora o menos, 800 mg/l/hora o menos o 600 mg/l/hora o menos. Por tanto, la productividad puede estar delimitada por dos cualesquiera de los extremos anteriores. Por ejemplo, la productividad puede ser de 3 a 30 mg/l/hora, de 6 a 20 mg/l/hora o de 15 a 30 mg/l/hora. La productividad preferida de un aldehído graso o alcohol graso producido por una célula huésped recombinante según los métodos de la divulgación se selecciona de 500 mg/l/hora a 2500 mg/l/hora, o desde 700 mg/l/hora hasta 2000 mg/l/hora.

Los términos “especie grasa total” y “producto de ácido graso total” pueden usarse de manera intercambiable en el presente documento con referencia a la cantidad total de alcoholes grasos, aldehídos grasos, ácidos grasos libres y ésteres grasos presentes en una muestra tal como se evalúa mediante CG-FID tal como se describe en la publicación de solicitud de patente internacional WO 2008/119082. Las muestras pueden contener uno, dos, tres o cuatro de estos compuestos dependiendo del contexto.

Tal como se usa en el presente documento, el término “tasa de utilización de glucosa” significa la cantidad de glucosa usada por el cultivo por tiempo unitario, notificada como gramos/litro/hora (g/l/h).

Tal como se usa en el presente documento, el término “fuente de carbono” se refiere a un sustrato o compuesto adecuado para usarse como fuente de carbono para el crecimiento de células procariontas o eucariotas simples. Las fuentes de carbono pueden estar en diversas formas, incluyendo, pero sin limitarse a polímeros, hidratos de carbono, ácidos, alcoholes, aldehídos, cetonas, aminoácidos, péptidos y gases (por ejemplo, CO y CO<sub>2</sub>). Las fuentes de carbono a modo de ejemplo incluyen, pero no se limitan a, monosacáridos, tales como glucosa, fructosa, manosa, galactosa, xilosa y arabinosa; oligosacáridos, tales como fructo-oligosacárido y galacto-oligosacárido; polisacáridos tales como almidón, celulosa, pectina y xilano; disacáridos, tales como sacarosa, maltosa, celobiosa y turanosa; material celulósico y variantes tales como hemicelulosas, metilcelulosa y carboximetilcelulosa sódica; ácidos grasos saturados o insaturados, succinato, lactato y acetato; alcoholes, tales como etanol, metanol y glicerol, o mezclas de los mismos. La fuente de carbono también puede ser un producto de fotosíntesis, tal como glucosa. En

determinadas realizaciones preferidas, la fuente de carbono es biomasa. En otras realizaciones preferidas, la fuente de carbono es glucosa. En otras realizaciones preferidas, la fuente de carbono es sacarosa.

Tal como se usa en el presente documento, el término “biomasa” se refiere a cualquier material biológico del que se derive una fuente de carbono. En algunas realizaciones, se procesa una biomasa para dar una fuente de carbono, que es adecuada para la bioconversión. En otras realizaciones, la biomasa no requiere procesamiento adicional para dar una fuente de carbono. La fuente de carbono puede convertirse en un biocombustible. Una fuente de biomasa a modo de ejemplo es vegetación o materia vegetal, tal como maíz, caña de azúcar o césped de pradera. Otra fuente de biomasa a modo de ejemplo es productos de desecho metabólicos, tales como materia animal (por ejemplo, estiércol de vaca). Las fuentes de biomasa a modo de ejemplo adicionales incluyen algas y otras plantas marinas. La biomasa también incluye productos de desecho de la industria, la agricultura, silvicultura y domésticos, incluyendo, pero sin limitarse a, residuo de fermentación, ensilado, paja, madera, aguas cloacales, basura, desechos urbanos celulósicos y restos de comidas. El término “biomasa” también puede referirse a fuentes de carbono, tales como hidratos de carbono (por ejemplo, monosacáridos, disacáridos o polisacáridos).

Tal como se usa en el presente documento, el término “aislados”, con respecto a productos (tales como ácidos grasos y derivados de los mismos) se refiere a productos que están separados de componentes celulares, medios de cultivo celular, o precursores químicos o sintéticos. Los ácidos grasos y derivados de los mismos producidos mediante los métodos descritos en el presente documento pueden ser relativamente inmiscibles en el caldo de fermentación, así como en el citoplasma. Por tanto, los ácidos grasos y derivados de los mismos pueden recogerse en una fase orgánica o bien de manera intracelular o bien de manera extracelular.

Tal como se usa en el presente documento, los términos “purificar”, “purificado” o “purificación” significan la retirada o el aislamiento de una molécula de su entorno, por ejemplo, mediante aislamiento o separación. Las moléculas “sustancialmente purificadas” están libres en al menos aproximadamente el 60% (por ejemplo, libres en al menos aproximadamente el 70%, libres en al menos aproximadamente el 75%, libres en al menos aproximadamente el 85%, libres en al menos aproximadamente el 90%, libres en al menos aproximadamente el 95%, libres en al menos aproximadamente el 97%, libres en al menos aproximadamente el 99%) de otros componentes con los que se asocian. Tal como se usa en el presente documento, estos términos también se refieren a la retirada de contaminantes de una muestra. Por ejemplo, la retirada de contaminantes puede dar como resultado un aumento en el porcentaje de un aldehído graso o un alcohol graso en una muestra. Por ejemplo, cuando se produce un aldehído graso o un alcohol graso en una célula huésped recombinante, el aldehído graso o alcohol graso puede purificarse mediante la retirada de proteínas de la célula huésped. Después de la purificación, el porcentaje de un aldehído graso o un alcohol graso en la muestra aumenta. Los términos “purificar”, “purificado” y “purificación” son términos relativos que no requieren pureza absoluta. Por tanto, por ejemplo, cuando se produce un aldehído graso o un alcohol graso en células huésped recombinantes, un aldehído graso purificado o un alcohol graso purificado es un aldehído graso o un alcohol graso que está sustancialmente separado de otros componentes celulares (por ejemplo, ácidos nucleicos, polipéptidos, lípidos, hidratos de carbono u otros hidrocarburos).

#### Mejoras de cepas

Con el fin de cumplir los objetivos muy altos para el título, el rendimiento y/o la productividad de alcoholes grasos, se hicieron varias modificaciones en las células huésped de producción. FadR es un factor regulador clave implicado en las rutas de biosíntesis de ácidos grasos y degradación de ácidos grasos (Cronan *et al.*, Mol. Microbiol., 29(4): 937-943 (1998)). La enzima ACS de *E. coli* FadD y la proteína de transporte de ácidos grasos FadL son componentes esenciales de un sistema de captación de ácidos grasos. FadL media en el transporte de ácidos grasos al interior de la célula bacteriana, y FadD media en la formación de ésteres de acil-CoA. Cuando no está disponible ninguna otra fuente de carbono, la bacteria capta ácidos grasos exógenos y los convierte en ésteres de acil-CoA, que pueden unirse al factor de transcripción FadR y desreprimir la expresión de los genes *fad* que codifican para proteínas responsables del transporte (FadL), activación (FadD) y  $\beta$ -oxidación (FadA, FadB, FadE y FadH) de ácidos grasos. Cuando están disponibles fuentes de carbono alternativas, las bacterias sintetizan ácidos grasos como acil-ACP, que se usan para la síntesis de fosfolípidos, pero no son sustratos para la  $\beta$ -oxidación. Por tanto, acil-CoA y acil-ACP son ambas fuentes independientes de ácidos grasos que pueden dar como resultado diferentes productos finales (Caviglia *et al.*, J. Biol. Chem., 279(12): 1163-1169 (2004)). La solicitud provisional estadounidense n.º 611470.989 describe métodos mejorados de producción de derivados de ácido graso en una célula huésped que se modifica por ingeniería genética para tener un nivel de expresión alterado de un polipéptido de FadR en comparación con el nivel de expresión del polipéptido de FadR en una célula huésped silvestre correspondiente.

Hay especulaciones contradictorias en la técnica en cuanto a los factores limitantes de la biosíntesis de ácidos grasos en células huésped, tales como *E. coli*. Un enfoque para aumentar el flujo a través de la biosíntesis de ácidos grasos es manipular diversas enzimas en la ruta (figuras 1 y 2). El suministro de acil-ACP a partir de acetil-CoA por medio del complejo de acetil-CoA carboxilasa (acc) (figura 3) y la ruta de biosíntesis de ácidos grasos (*fab*) puede limitar la tasa de producción de alcohol graso. En un enfoque a modo de ejemplo detallado en el ejemplo 2, el efecto de la sobreexpresión de *Corynebacterium glutamicum* accABCD ( $\pm$ birA) demostró que tales modificaciones genéticas pueden conducir a un aumento de acetil-coA y malonil-CoA en *E. coli*. Un posible motivo de una baja tasa de flujo a través de la biosíntesis de ácidos grasos es un suministro limitado de precursores, concretamente acetil-CoA y, en particular, malonil-CoA, y los principales precursores para la biosíntesis de ácidos grasos. El ejemplo 3

describe la construcción de operones de fab que codifican para enzimas en la ruta de biosíntesis para la conversión de malonil-CoA en acil-ACP y su integración en el cromosoma de una célula huésped de *E. coli*. En aún otro enfoque detallado en el ejemplo 4, se mostró que mutaciones en los genes de rph e ilvG en la célula huésped de *E. coli* daban como resultado una producción superior de ácido graso libre (FFA), que se tradujo en una producción superior de alcohol graso. En todavía otro enfoque, se realizaron mutagénesis de transposones y examen de alto rendimiento para encontrar mutaciones beneficiosas que aumentaran el título o rendimiento. El ejemplo 5 describe cómo una inserción de transposón en el gen de yjyP puede mejorar el rendimiento de alcohol graso en fermentaciones semicontinuas y en matraz de agitación.

#### Ácido carboxílico reductasa (CAR)

Se han modificado por ingeniería genética células huésped recombinantes para producir alcoholes grasos expresando una tioesterasa, que cataliza la conversión de acil-ACP en ácidos grasos libres (FFA) y una ácido carboxílico reductasa (CAR), que convierte ácidos grasos libres en aldehídos grasos. Aldehído reductasas nativas (endógenas) presentes en la célula huésped (por ejemplo, *E. coli*) pueden convertir aldehídos grasos en alcoholes grasos. Se describen tioesterasas a modo de ejemplo por ejemplo en la publicación de patente estadounidense n.º 20100154293. CarB es una ácido carboxílico reductasa a modo de ejemplo, una enzima clave en la ruta de producción de alcohol graso. El documento WO2010/062480 describe una búsqueda con BLAST usando la secuencia de aminoácidos NRRL 5646 CAR (registro de Genpept AAR91681) (SEQ ID NO: 6) como secuencia de consulta, y el uso de la misma en la identificación de aproximadamente 20 secuencias homólogas.

Los términos “ácido carboxílico reductasa”, “CAR” y “polipéptido de biosíntesis de aldehídos grasos” se usan de manera intercambiable en el presente documento. En la práctica de la divulgación, se expresa o sobreexpresa un gen que codifica para un polipéptido de ácido carboxílico reductasa en la célula huésped. En algunas realizaciones, el polipéptido de CarB tiene la secuencia de aminoácidos de SEQ ID NO: 7. En otras realizaciones, el polipéptido de CarB es una variante o un mutante de SEQ ID NO: 7. En determinadas realizaciones, el polipéptido de CarB es de una célula de mamífero, célula vegetal, célula de insecto, célula de levadura, célula fúngica, célula de hongos filamentosos, una célula bacteriana o cualquier otro organismo. En algunas realizaciones, la célula bacteriana es una micobacteria seleccionada del grupo que consiste en *Mycobacterium smegmatis*, *Mycobacterium abscessus*, *Mycobacterium avium*, *Mycobacterium bovis*, *Mycobacterium tuberculosis*, *Mycobacterium leprae*, *Mycobacterium marinum* y *Mycobacterium ulcerans*. En otras realizaciones, la célula bacteriana es de una especie de *Nocardia*, por ejemplo, *Nocardia NRRL 5646*, *Nocardia farcinica*, *Streptomyces griseus*, *Salinispora arenicola* o *Clavibacter michiganensis*. En otras realizaciones, el polipéptido de CarB es un homólogo de CarB que tiene una secuencia de aminoácidos que es al menos el 90%, el 91%, el 92%, el 93%, el 94%, el 95%, el 96%, el 97%, el 98% o el 99% idéntica a la secuencia de aminoácidos de SEQ ID NO: 7.

En otras realizaciones, el polipéptido de CarB contiene una mutación en el número de aminoácido 3, 12, 20, 28, 46, 74, 103, 191, 288, 473, 827, 926, 927, 930 ó 1128 de SEQ ID NO: 7. Se detallan mutaciones a modo de ejemplo en la tabla 10. Los fragmentos o mutantes preferidos de un polipéptido conservan parte o toda la función biológica (por ejemplo, actividad enzimática) del polipéptido silvestre correspondiente. En algunas realizaciones, el fragmento o mutante conserva al menos aproximadamente el 75%, al menos aproximadamente el 80%, al menos aproximadamente el 90%, al menos aproximadamente el 95% o al menos aproximadamente el 98% o más de la función biológica del polipéptido silvestre correspondiente. En otras realizaciones, el fragmento o mutante conserva aproximadamente el 100% de la función biológica del polipéptido silvestre correspondiente. Puede encontrarse orientación para determinar qué residuos de aminoácido pueden sustituirse, insertarse o deleccionarse sin afectar a la actividad biológica usando programas informáticos bien conocidos en la técnica, por ejemplo, el software LASERGENE™ (DNASTAR, Inc., Madison, WI).

En aún otras realizaciones, un fragmento o mutante presenta una función biológica aumentada en comparación con un polipéptido silvestre correspondiente. Por ejemplo, un fragmento o mutante puede presentar una mejora de al menos aproximadamente el 10%, al menos aproximadamente el 25%, al menos aproximadamente el 50%, al menos aproximadamente el 75% o al menos aproximadamente el 90% en la actividad enzimática en comparación con el polipéptido silvestre correspondiente. En otras realizaciones, el fragmento o mutante presenta una mejora de al menos aproximadamente el 100% (por ejemplo, al menos aproximadamente el 200%, o al menos aproximadamente el 500%) en la actividad enzimática en comparación con el polipéptido silvestre correspondiente. Se entiende que los polipéptidos descritos en el presente documento pueden tener sustituciones de aminoácidos no esenciales o conservativas adicionales, que no tienen un efecto sustancial sobre la función del polipéptido. Puede determinarse si se tolerará o no una sustitución particular (es decir, no afectará de manera adversa a la función biológica deseada, tal como unión al ADN o actividad enzimática) tal como se describe en Bowie *et al.* (Science, 247: 1306-1310 (1990)).

Como resultado de los métodos y las enzimas variantes de la presente divulgación, uno o más del título, el rendimiento y/o la productividad del ácido graso o derivado del mismo producido por la célula huésped modificada por ingeniería genética que tiene un nivel de expresión alterado de un polipéptido de CarB se aumenta en relación con el de la célula huésped silvestre correspondiente. Para permitir una conversión máxima de ácidos grasos C12 y C14 en alcoholes grasos, CarB debe expresarse a actividad suficiente. Una célula huésped recombinante mejorada tendría una enzima CAR que se expresa a partir de, por ejemplo, el cromosoma de *E. coli*. Tal como se muestra en

el ejemplo 6, las células que expresan la enzima CarB a partir del cromosoma tienen más actividad ácido carboxílico reductasa en relación con la CarB original y pueden convertir más ácidos grasos C12 y C14 en alcoholes grasos. CarB es un gen grande (3,5 kb) y aumenta el tamaño del plásmido considerablemente, lo que hace difícil usar un plásmido pCL para someter a prueba nuevos genes durante el desarrollo de cepas. Los enfoques para aumentar la actividad de CarB incluyen aumentar su solubilidad, estabilidad, expresión y/o funcionalidad. En un enfoque a modo de ejemplo, se produce una proteína de fusión que contiene 6 histidinas y un sitio de escisión de trombina en el extremo N-terminal de CarB. Esta enzima difiere de CarB en 60 nucleótidos adicionales en el extremo N-terminal, y se denomina CarB60. Cuando se expresan CarB o CarB60 a partir del cromosoma de *E. coli* bajo el control del promotor pTRC, las células que contienen CarB60 tienen una actividad ácido carboxílico reductasa celular total aumentada y convierten más ácidos grasos libres (FFA) C12 y C14 en alcoholes grasos. Un experto en la técnica apreciará que esto es un ejemplo de modificación por ingeniería genética molecular con el fin de lograr una mayor conversión de ácidos grasos libres (FFA) C12 y C14 en alcoholes grasos tal como se ilustra en el ejemplo 6 (citado anteriormente). Se engloban enfoques similares en el presente documento (véase el ejemplo 7).

Las fosfopanteteína transferasas (PPTasas) (EC 2.7.8.7) catalizan la transferencia de 4'-fosfopanteteína desde CoA hasta un sustrato. Car, CarB de *Nocardia* y varios homólogos de las mismas contienen un sitio de unión supuesto para 4'-fosfopanteteína (PPT) (He *et al.*, Appl. Environ. Microbiol., 70(3): 187 4-1881 (2004)). En algunas realizaciones de la divulgación, se expresa o sobreexpresa una PPTasa en una célula huésped modificada por ingeniería genética. En determinadas realizaciones, la PPTasa es EntD de *E. coli* MG1655 (SEQ ID NO: 8). En algunas realizaciones, se expresan o sobreexpresan una tioesterasa y una ácido carboxílico reductasa en una célula huésped modificada por ingeniería genética. En determinadas realizaciones, la tioesterasa es tesA y la ácido carboxílico reductasa es carB. En otras realizaciones, se expresan o sobreexpresan una tioesterasa, una ácido carboxílico reductasa y una alcohol deshidrogenasa en una célula huésped modificada por ingeniería genética. En determinadas realizaciones, la tioesterasa es tesA, la ácido carboxílico reductasa es carB y la alcohol deshidrogenasa es alrAadp1 (número de registro de GenPept CAG70248.1) de *Acinetobacter baylyi* ADP1 (SEQ ID NO: 4). En todavía otras realizaciones, se expresan o sobreexpresan una tioesterasa, una ácido carboxílico reductasa una PPTasa y una alcohol deshidrogenasa en la célula huésped modificada por ingeniería genética. En determinadas realizaciones, la tioesterasa es tesA, la ácido carboxílico reductasa es carB, la PPTasa es entD y la alcohol deshidrogenasa es alrAadp1. En realizaciones todavía adicionales, una célula huésped modificada que expresa una o más de una tioesterasa, una CAR, una PPTasa y una alcohol deshidrogenasa tiene también una o más mejoras de cepas. Las mejoras de cepas a modo de ejemplo incluyen, pero no se limitan a la expresión o sobreexpresión de un polipéptido de acetil-CoA carboxilasa, sobreexpresión de un polipéptido de FadR, expresión o sobreexpresión de un operón de iFAB heterólogo o la inserción de un transposón en el gen de yijP u otro gen, o enfoques similares. También se describe en el presente documento una composición de alcohol graso producida mediante cualquiera de los métodos descritos en el presente documento. Una composición de alcohol graso producida mediante cualquiera de los métodos descritos en el presente documento puede usarse directamente como materiales de partida para la producción de otros compuestos químicos (por ejemplo, polímeros, tensioactivos, plásticos, textiles, disolventes, adhesivos, etc.) o aditivos para el cuidado personal. Estos compuestos pueden usarse también como materia prima para reacciones posteriores, por ejemplo, hidrogenación, craqueo catalítico (por ejemplo, por medio de hidrogenación, pirólisis o ambas) para preparar otros productos.

#### 40 Mutantes o variantes

En algunas realizaciones, el polipéptido expresado en una célula huésped recombinante es un mutante o una variante de cualquiera de los polipéptidos descritos en el presente documento. Los términos "mutante" y "variante" tal como se usan en el presente documento se refieren a un polipéptido que tiene una secuencia de aminoácidos que difiere de un polipéptido silvestre en al menos un aminoácido. Por ejemplo, el mutante puede comprender una o más de las siguientes sustituciones de aminoácidos conservativas: reemplazo de un aminoácido alifático, tal como alanina, valina, leucina e isoleucina, por otro aminoácido alifático; reemplazo de una serina por una treonina; reemplazo de una treonina por una serina; reemplazo de un residuo ácido, tal como ácido aspártico y ácido glutámico, por otro residuo ácido; reemplazo de un residuo que porta un grupo amida, tal como asparagina y glutamina, por otro residuo que porta un grupo amida; intercambio de un residuo básico, tal como lisina y arginina, por otro residuo básico; y reemplazo de un residuo aromático, tal como fenilalanina y tirosina, por otro residuo aromático. En algunas realizaciones, el polipéptido mutante tiene aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 30, 40, 50, 60, 70, 80, 90, 100 o más sustituciones, adiciones, inserciones o deleciones de aminoácidos. Algunos fragmentos o mutantes preferidos de un polipéptido conservan parte o la totalidad de la función biológica (por ejemplo, actividad enzimática) del polipéptido silvestre correspondiente. En algunas realizaciones, el fragmento conserva al menos aproximadamente el 75%, al menos aproximadamente el 80%, al menos aproximadamente el 90%, al menos aproximadamente el 95% o al menos aproximadamente el 98% o más de la función biológica del polipéptido silvestre correspondiente. En otras realizaciones, el fragmento o mutante conserva aproximadamente el 100% de la función biológica del polipéptido silvestre correspondiente. Puede encontrarse orientación sobre la determinación de qué residuos de aminoácido pueden sustituirse, insertarse o deleccionarse sin afectar a la biológica actividad usando programas informáticos bien conocidos en la técnica, por ejemplo, el software LASERGENE™ (DNASTAR, Inc., Madison, WI).

En aún otras realizaciones, un fragmento o mutante presenta un aumento de la función biológica en comparación con un polipéptido silvestre correspondiente. Por ejemplo, un fragmento o mutante puede presentar una mejora de al

menos un 10%, al menos un 25%, al menos un 50%, al menos un 75% o al menos un 90% de la actividad enzimática en comparación con el polipéptido silvestre correspondiente. En otras realizaciones, el fragmento o mutante presenta una mejora de al menos el 100%, (por ejemplo, al menos el 200%, o al menos el 500%) de la actividad enzimática en comparación con el polipéptido silvestre correspondiente. Se entiende que los polipéptidos descritos en el presente documento pueden tener sustituciones de aminoácidos no esenciales o conservativas adicionales que no tienen un efecto sustancial sobre la función del polipéptido. Puede determinarse si se tolerará o no una sustitución particular (es decir, no afectará de manera adversa a la función biológica deseada, tal como actividad ácido carboxílico reductasa) tal como se describe en Bowie *et al.* (Science, 247:1306-1310 (1990)). Una sustitución de aminoácidos conservativa es una en la que el residuo de aminoácido se reemplaza por un residuo de aminoácido que tiene una cadena lateral similar. Se han definido en la técnica familias de residuos de aminoácido que tienen cadenas laterales similares. Estas familias incluyen aminoácidos con cadenas laterales básicas (por ejemplo, lisina, arginina, histidina), cadenas laterales ácidas (por ejemplo, ácido aspártico, ácido glutámico), cadenas laterales polares no cargadas (por ejemplo, glicina, asparagina, glutamina, serina, treonina, tirosina, cisteína), cadenas laterales apolares (por ejemplo, alanina, valina, leucina, isoleucina, prolina, fenilalanina, metionina, triptófano), cadenas laterales con ramificación beta (por ejemplo, treonina, valina, isoleucina) y cadenas laterales aromáticas (por ejemplo, tirosina, fenilalanina, triptófano, histidina). Las variantes pueden producirse de manera natural o crearse *in vitro*. En particular, tales variantes pueden crearse usando técnicas de ingeniería genética, tales como mutagénesis dirigida al sitio, mutagénesis química al azar, procedimientos de delección con exonucleasa III o técnicas de clonación convencionales. Alternativamente, tales variantes, mutantes, fragmentos, análogos o derivados pueden crearse usando procedimientos de modificación o síntesis química.

Se conocen bien en la técnica métodos de producción de variantes. Estos incluyen procedimientos en los que secuencias de ácido nucleico obtenidas a partir de aislados naturales se modifican para generar ácidos nucleicos que codifican para polipéptidos que tienen características que potencian su valor en aplicaciones industriales o de laboratorio. En tales procedimientos, se generan y caracterizan un gran número de secuencias variantes que tienen una o más diferencias de nucleótidos con respecto a la secuencia obtenida a partir del aislado natural. Normalmente, estas diferencias de nucleótidos dan como resultado cambios de aminoácidos con respecto a los polipéptidos codificados por los ácidos nucleicos de los aislados naturales. Por ejemplo, pueden prepararse variantes usando mutagénesis al azar y dirigida al sitio. La mutagénesis al azar y dirigida al sitio se describen en, por ejemplo Arnold Curr. Opin. Biotech. 4:450-455 (1993). Puede lograrse la mutagénesis al azar usando PCR propensa a error (véanse, por ejemplo, Leung *et al.* Technique 1:11-15 (1989); y Caldwell *et al.* PCR Methods Applic. 2: 28-33 (1992)). En la PCR propensa a error, la se realiza PCR en condiciones en las que la fidelidad de copiado de la ADN polimerasa es baja, de manera que se obtiene una alta tasa de mutaciones puntuales a lo largo de toda la longitud del producto de PCR. En resumen, en tales procedimientos, se mezclan ácidos nucleicos que van a mutagenizarse (por ejemplo, una secuencia de polinucleótido que codifica para una enzima ácido carboxílico reductasa) con cebadores de PCR, tampón de reacción, MgCl<sub>2</sub>, MnCl<sub>2</sub>, Taq polimerasa, y una concentración apropiada de dNTP para lograr una alta tasa de mutaciones puntuales a lo largo de toda la longitud del producto de PCR. Por ejemplo, la reacción puede realizarse usando 20 fmoles de ácido nucleico que va a mutagenizarse, 30 pmoles de cada cebador de PCR, un tampón de reacción que comprende KCl 50 mM, Tris HCl 10 mM (pH 8,3), gelatina al 0,01%, MgCl<sub>2</sub> 7 mM, MnCl<sub>2</sub> 0,5 mM, 5 unidades de Taq polimerasa, dGTP 0,2 mM, dATP 0,2 mM, dCTP 1 mM y dTTP 1 mM. Puede realizarse PCR durante 30 ciclos de 94°C durante 1 min, 45°C durante 1 min y 72°C durante 1 min. Sin embargo, los expertos en la técnica apreciarán que estos parámetros pueden variarse según sea apropiado. Los ácidos nucleicos mutagenizados se clonan entonces en un vector apropiado, y se evalúan las actividades de los polipéptidos codificados por los ácidos nucleicos mutagenizados (véase el ejemplo 7). Puede lograrse la mutagénesis dirigida al sitio usando mutagénesis dirigida por oligonucleótidos para generar mutaciones específicas de sitio en cualquier ADN clonado de interés. La mutagénesis por oligonucleótidos se describe en, por ejemplo, Reidhaar-Olson *et al.* Science 241:53-57 (1988). En resumen, en tales procedimientos, se sintetizan una pluralidad de oligonucleótidos bicatenarios que portan una o más mutaciones que van a introducirse en el ADN clonado, y se insertan en el ADN clonado que va a mutagenizarse (por ejemplo, una secuencia de polinucleótido que codifica para un polipéptido de CAR). Se recuperan los clones que contienen el ADN mutagenizado, y se evalúan las actividades de los polipéptidos para los que codifican. Otro método para generar variantes es la PCR de ensamblaje. La PCR de ensamblaje implica el ensamblaje de un producto de PCR de una mezcla de fragmentos de ADN pequeños. Un gran número de diferentes reacciones PCR se producen en paralelo en el mismo vial, cebando con los productos de una reacción los productos de otra reacción. La PCR de ensamblaje se describe en, por ejemplo, la patente estadounidense 5.965.408. Todavía otro método de generación de variantes es la mutagénesis por PCR sexual. En la mutagénesis por PCR sexual, se produce recombinación homóloga forzada entre moléculas de ADN de diferentes secuencias de ADN, pero altamente relacionadas, *in vitro* como resultado de fragmentación al azar de la molécula de ADN basándose en la homología de secuencia. Esto va seguido por fijación del cruzamiento mediante extensión de cebadores en una reacción PCR. La mutagénesis por PCR sexual se describe en, por ejemplo, Stemmer Proc. Natl. Acad. Sci. U.S.A. 91:10747-10751 (1994).

También pueden crearse variantes mediante mutagénesis *in vivo*. En algunas realizaciones, se generan mutaciones al azar en una secuencia de ácido nucleico propagando la secuencia en una cepa bacteriana, tal como una cepa de *E. coli*, que porta mutaciones en una o más de las rutas de reparación de ADN. Tales cepas "mutadoras" tienen una mayor tasa de mutación al azar que la de una cepa silvestre. La propagación de una secuencia de ADN (por ejemplo, una secuencia de polinucleótido que codifica para un polipéptido de CAR) en una de estas cepas generará

eventualmente mutaciones al azar dentro del ADN. Se describen cepas mutadoras adecuadas para su uso para mutagénesis *in vivo* en, por ejemplo, la publicación de solicitud de patente internacional n.º WO1991/016427). También pueden generarse variantes usando mutagénesis de casete. En la mutagénesis de casete, una pequeña región de una molécula de ADN bicatenario se reemplaza por un "casete" de oligonucleótido sintético que difiere de la secuencia nativa. El oligonucleótido contiene a menudo una secuencia nativa completa y/o parcialmente aleatorizada. También puede usarse mutagénesis de conjunto recursivo para generar variantes. La mutagénesis de conjunto recursivo es un algoritmo para la modificación por ingeniería de proteínas (es decir, mutagénesis de proteínas) desarrollado para producir diversas poblaciones de mutantes relacionados fenotípicamente cuyos miembros difieren en la secuencia de aminoácidos. Este método usa un mecanismo de realimentación para controlar tandas sucesivas de mutagénesis de casete combinatoria. La mutagénesis de conjunto recursivo se describe en, por ejemplo, Arkin *et al.* Proc. Natl. Acad. Sci., U.S.A. 89:7811-7815 (1992). En algunas realizaciones, se crean variantes usando mutagénesis de conjunto exponencial. La mutagénesis de conjunto exponencial es un procedimiento para generar bibliotecas combinatorias con un alto porcentaje de mutantes únicos y funcionales, en los que se aleatorizan pequeños grupos de residuos en paralelo para identificar, en cada posición alterada, aminoácidos que conducen a proteínas funcionales (véase, por ejemplo, Delegrave *et al.* (1993) Biotech. Res. 11:1548-1552). En algunas realizaciones, se crean variantes usando procedimientos de intercambio en los que se fusionan entre sí partes de una pluralidad de ácidos nucleicos que codifican para distintos polipéptidos para crear secuencias de ácido nucleico quiméricas que codifican para polipéptidos quiméricos tal como se describe en, por ejemplo, las patentes estadounidenses 5.965.408 y 5.939.250.

La mutagénesis por inserción es mutagénesis de ADN mediante la inserción de una o más bases. Las mutaciones por inserción pueden producirse de manera natural, mediadas por virus o transposón, o pueden crearse artificialmente para fines de investigación en el laboratorio, por ejemplo, mediante mutagénesis de transposones. Cuando se integra ADN exógeno en el del huésped, la gravedad de cualquier mutación consiguiente depende totalmente de la ubicación dentro del genoma del huésped en el que se inserta el ADN. Por ejemplo, pueden ser evidentes efectos significativos si se inserta un transposón en el medio de un gen esencial, en una región promotora o en una región represora o potenciadora. Se realizaron mutagénesis de transposones y examen de alto rendimiento para encontrar mutaciones beneficiosas que aumentarían el título o rendimiento de alcohol graso. Se describen en el presente documento células huésped recombinantes que comprenden (a) una secuencia de polinucleótido que codifica para una ácido carboxílico reductasa que comprende una secuencia de aminoácidos que tiene una identidad de secuencia de al menos el 80%, el 81%, el 82%, el 83%, el 84%, el 85%, el 86%, el 87%, el 88%, el 89%, el 90%, el 91%, el 92%, el 93%, el 94%, el 95%, el 96%, el 97%, el 98% o el 99% con la secuencia de aminoácidos de SEQ ID NO: 7 y (b) un polinucleótido que codifica para un polipéptido que tiene actividad ácido carboxílico reductasa, en las que la célula huésped recombinante puede producir un aldehído graso o un alcohol graso.

#### Modificación por ingeniería genética de células huésped

En algunas realizaciones, se proporciona una secuencia de polinucleótido (o gen) a una célula huésped por medio de un vector recombinante, que comprende un promotor operativamente unido a la secuencia de polinucleótido. En determinadas realizaciones, el promotor es un promotor regulado por el desarrollo, uno específico de orgánulo, uno específico de tejido, uno inducible, uno constitutivo o uno específico de célula. En algunas realizaciones, el vector recombinante incluye (a) una secuencia de control de la expresión acoplada operativamente a la secuencia de polinucleótido; (b) un marcador de selección acoplado operativamente a la secuencia de polinucleótido; (c) una secuencia de marcador acoplada operativamente a la secuencia de polinucleótido; (d) un resto de purificación acoplado operativamente a la secuencia de polinucleótido; (e) una secuencia de secreción acoplada operativamente a la secuencia de polinucleótido; y (f) una secuencia de direccionamiento acoplada operativamente a la secuencia de polinucleótido. Los vectores de expresión descritos en el presente documento incluyen una secuencia de polinucleótido descrita en el presente documento en una forma adecuada para la expresión de la secuencia de polinucleótido en una célula huésped. Los expertos en la técnica apreciarán que el diseño del vector de expresión puede depender de factores tales como la elección de la célula huésped que va a transformarse, el nivel de expresión del polipéptido deseado, etc. Los vectores de expresión descritos en el presente documento pueden introducirse en las células huésped para producir polipéptidos, incluyendo polipéptidos de fusión, codificados por las secuencias de polinucleótido descritas en el presente documento. La expresión de genes que codifican para polipéptidos en procariontes, por ejemplo, *E. coli*, se lleva a cabo lo más a menudo con vectores que contienen promotores constitutivos o inducibles que dirigen la expresión de polipéptidos o bien de fusión o bien de no fusión. Los vectores de fusión añaden un número de aminoácidos a un polipéptido codificado en el mismo, habitualmente al extremo amino-terminal o carboxi-terminal del polipéptido recombinante. Tales vectores de fusión sirven normalmente para uno o más de los tres propósitos siguientes: (1) aumentar la expresión del polipéptido recombinante; (2) aumentar la solubilidad del polipéptido recombinante; y (3) ayudar en la purificación del polipéptido recombinante actuando como ligando en purificación por afinidad. A menudo, en los vectores de expresión de fusión, se introduce un sitio de escisión proteolítica en la unión del resto de fusión y el polipéptido recombinante. Esto permite la separación del polipéptido recombinante del resto de fusión después de la purificación del polipéptido de fusión. Los ejemplos de tales enzimas, y sus secuencias de reconocimiento relacionadas, incluyen factor Xa, trombina y enterocinasa. Los vectores de expresión de fusión a modo de ejemplo incluyen pGEX (Pharmacia Biotech, Inc., Piscataway, NJ; Smith *et al.*, Gene, 67: 31-40 (1988)), pMAL (New England Biolabs, Beverly, MA) y pRITS (Pharmacia Biotech, Inc., Piscataway, N.J.), que fusionan glutatión-S-transferasa (GST), proteína de unión a

maltosa E o proteína A, respectivamente, al polipéptido recombinante diana.

Los ejemplos de vectores de expresión de *E. coli* de no fusión, inducibles incluyen pTrc (Amann *et al.*, Gene (1988) 69:301-315) y pET 11d (Studier *et al.*, Gene Expression Technology: Methods in Enzymology 185, Academic Press, San Diego, Calif. (1990) 60-89). La expresión del gen diana a partir del vector pTrc se basa en la transcripción por la ARN polimerasa del huésped a partir de un promotor de fusión trp-lac híbrido. La expresión del gen diana a partir del vector pET 11d se basa en la transcripción a partir de un promotor de fusión T7 gn10-lac mediada por una ARN polimerasa viral coexpresada (T7 gn1). Esta polimerasa viral la suministran cepas huésped BL21(DE3) o HMS174(DE3) a partir de un profago  $\lambda$  residente que alberga un gen de T7 gn1 bajo el control transcripcional del promotor lacUV 5. Se conocen bien en la técnica sistemas de expresión adecuados para células tanto procariontas como eucariotas; véase, por ejemplo, Sambrook *et al.*, "Molecular Cloning: A Laboratory Manual", segunda edición, Cold Spring Harbor Laboratory, (1989). Los ejemplos de vectores de expresión de *E. coli* de no fusión, inducibles incluyen pTrc (Amann *et al.*, Gene, 69: 301-315 (1988)) y PET 11d (Studier *et al.*, Gene Expression Technology: Methods in Enzymology 185, Academic Press, San Diego, CA, págs. 60-89 (1990)). En determinadas realizaciones, una secuencia de polinucleótido de la divulgación está operativamente unida a un promotor derivado del bacteriófago T5. En una realización, la célula huésped es una célula de levadura. En esta realización, el vector de expresión es un vector de expresión de levadura. Pueden introducirse vectores en células procariontas o eucariotas por medio de técnicas reconocidas en la técnica para introducir ácido nucleico foráneo (por ejemplo, ADN) en una célula huésped. Pueden encontrarse métodos adecuados para transformar o transfectar células huésped en, por ejemplo, Sambrook *et al.* (citado anteriormente). Para la transformación estable de células bacterianas, se sabe que, dependiendo del vector de expresión y la técnica de transformación usados, sólo una pequeña fracción de células captará y replicará el vector de expresión. En algunas realizaciones, con el fin de identificar y seleccionar estos transformantes, se introduce un gen que codifica para un marcador seleccionable (por ejemplo, resistencia a un antibiótico) en las células huésped junto con el gen de interés. Los marcadores seleccionables incluyen los que confieren resistencia a fármacos tales como, pero sin limitarse a, ampicilina, kanamicina, cloranfenicol o tetraciclina. Pueden introducirse ácidos nucleicos que codifican para un marcador seleccionable en una célula huésped en el mismo vector que el que codifica para un polipéptido descrito en el presente documento o pueden introducirse en un vector separado. Las células transformadas de manera estable con el ácido nucleico introducido pueden identificarse mediante crecimiento en presencia de un fármaco de selección apropiado.

Producción de composiciones de alcoholes grasos por células huésped recombinantes

Las estrategias para aumentar la producción de alcoholes grasos por células huésped recombinantes incluyen el aumento del flujo a través de la ruta de biosíntesis de ácidos grasos mediante la sobreexpresión de genes de biosíntesis de ácidos grasos nativos y la expresión de genes de biosíntesis de ácidos grasos exógenos de diferentes organismos en un huésped de producción modificado por ingeniería genética. También pueden emplearse actividad potenciada de enzimas relevantes en la ruta de biosíntesis de alcoholes grasos, por ejemplo, CAR, así como otras estrategias para optimizar el crecimiento y la productividad de la célula huésped para maximizar la producción. En algunas realizaciones, la célula huésped recombinante comprende un polinucleótido que codifica para un polipéptido (una enzima) que tiene una actividad de biosíntesis de alcoholes grasos (es decir, un polipéptido de biosíntesis de alcoholes grasos o una enzima de biosíntesis de alcoholes grasos), y se produce un alcohol graso por la célula huésped recombinante. Puede producirse una composición que comprende alcoholes grasos (una composición de alcohol graso) cultivando la célula huésped recombinante en presencia de una fuente de carbono en condiciones eficaces para expresar una enzima de biosíntesis de alcoholes grasos. En algunas realizaciones, la composición de alcohol graso comprende alcoholes grasos, sin embargo, una composición de alcohol graso puede comprender otros derivados de ácido graso. Normalmente, la composición de alcohol graso se recupera del entorno extracelular de la célula huésped recombinante, es decir, el medio de cultivo celular. En un enfoque, se han modificado por ingeniería genética células huésped recombinantes para producir alcoholes grasos expresando una tioesterasa, que cataliza la conversión de acil-ACP en ácidos grasos libres (FFA) y una ácido carboxílico reductasa (CAR), que convierte ácidos grasos libres en aldehídos grasos. Las aldehído reductasas nativas (endógenas) presentes en la célula huésped (por ejemplo, *E. coli*) pueden convertir los aldehídos grasos en alcoholes grasos. En algunas realizaciones, el alcohol graso se produce expresando o sobreexpresando en la célula huésped recombinante un polinucleótido que codifica para un polipéptido que tiene actividad de biosíntesis de alcoholes grasos que convierte un aldehído graso en un alcohol graso. Por ejemplo, puede usarse una alcohol deshidrogenasa (también denominada en el presente documento aldehído reductasa, por ejemplo, EC 1.1.1.1), en la práctica de la divulgación. Tal como se usa en el presente documento, el término "alcohol deshidrogenasa" se refiere a un polipéptido que puede catalizar la conversión de un aldehído graso en un alcohol (por ejemplo, un alcohol graso). Un experto habitual en la técnica apreciará que determinadas alcohol deshidrogenasas también pueden catalizar otras reacciones, y estas alcohol deshidrogenasas no específicas también se engloban mediante el término "alcohol deshidrogenasa". Los ejemplos de polipéptidos de alcohol deshidrogenasa útiles según la divulgación incluyen, pero no se limitan a AlrAadp1 (SEQ ID NO: 4) u homólogos de AlrA y alcohol deshidrogenasas de *E. coli* endógenas tales como YjgB, (AAC77226) (SEQ ID NO: 5), DkgA (NP\_417485), DkgB (NP\_414743), YdjL (AAC74846), YdjJ (NP\_416288), AdhP (NP\_415995), YhdH (NP\_417719), YahK (NP\_414859), YphC (AAC75598), YqhD (446856) e YbbO [AAC73595.1]. Se describen ejemplos adicionales en las publicaciones de solicitud de patente internacional n.<sup>os</sup> WO2007/136762, WO2008/119082 y WO 2010/062480. En determinadas realizaciones el polipéptido de biosíntesis de alcoholes grasos tiene actividad aldehído reductasa o alcohol deshidrogenasa (EC 1.1.1.1). En otro enfoque, se han

modificado por ingeniería genética células huésped recombinantes para producir alcoholes grasos expresando acil-CoA reductasas o acilo graso reductasas (FAR) que forman alcoholes grasos que convierten sustratos de acil graso-tioéster (por ejemplo, acil graso-CoA o acil graso-ACP) en alcoholes grasos. En algunas realizaciones, el alcohol graso se produce expresando o sobreexpresando un polinucleótido que codifica para un polipéptido que tiene actividad acil-CoA reductasa (FAR) que forma alcoholes grasos en una célula huésped recombinante. Se describen ejemplos de polipéptidos de FAR útiles según esta realización en la publicación PCT n.º WO2010/062480.

El alcohol graso puede producirse por medio de una ruta dependiente de acil-CoA que utiliza productos intermedios de acil graso-ACP y acil graso-CoA y una ruta independiente de acil-CoA que utiliza productos intermedios de acil graso-ACP pero no un producto intermedio de acil graso-CoA. En realizaciones particulares, la enzima codificada por el gen sobreexpresado se selecciona de una ácido graso sintasa, una acil-ACP tioesterasa, una acil graso-CoA sintasa y una acetil-CoA carboxilasa. En algunas realizaciones, la proteína codificada por el gen sobreexpresado es endógena para la célula huésped. En otras realizaciones la proteína codificada por el gen sobreexpresado es heteróloga para la célula huésped. Se preparan también alcoholes grasos en la naturaleza por enzimas que pueden reducir diversas moléculas de acil-ACP o acil-CoA hasta los correspondientes alcoholes primarios. Véanse también las publicaciones de patente estadounidense n.ºs 20100105963 y 20110206630 y la patente estadounidense n.º 8097439. Tal como se usa en el presente documento, una célula huésped recombinante o una célula huésped modificada por ingeniería genética se refiere a una célula huésped cuya constitución genética se ha alterado en relación con la célula huésped silvestre correspondiente, por ejemplo, mediante la introducción deliberada de nuevos elementos genéticos y/o la modificación deliberada de elementos genéticos presentes de manera natural en la célula huésped. La descendencia de tales células huésped recombinantes también contiene estos elementos genéticos nuevos y/o modificados. En cualquiera de los aspectos de la divulgación descrita en el presente documento, la célula huésped puede seleccionarse del grupo que consiste en una célula vegetal, célula de insecto, célula fúngica (por ejemplo, un hongo filamentoso, tales como *Candida* sp., o una levadura gemante, tal como *Saccharomyces* sp.), una célula algal y una célula bacteriana. En una realización preferida, las células huésped recombinantes son células microbianas recombinantes. Los ejemplos de células huésped que son células microbianas incluyen pero no se limitan a células del género *Escherichia*, *Bacillus*, *Lactobacillus*, *Zymomonas*, *Rhodococcus*, *Pseudomonas*, *Aspergillus*, *Trichoderma*, *Neurospora*, *Fusarium*, *Humicola*, *Rhizomucor*, *Kluyveromyces*, *Pichia*, *Mucor*, *Myceliophthora*, *Penicillium*, *Phanerochaete*, *Pleurotus*, *Trametes*, *Chrysosporium*, *Saccharomyces*, *Stenotrophomonas*, *Schizosaccharomyces*, *Yarrowia* o *Streptomyces*. En algunas realizaciones, la célula huésped es una célula bacteriana Gram-positiva. En otras realizaciones, la célula huésped es una célula bacteriana Gram-negativa. En algunas realizaciones, la célula huésped es una célula de *E. coli*. En otras realizaciones, la célula huésped es una célula de *Bacillus lentus*, una célula de *Bacillus brevis*, una célula de *Bacillus stearothermophilus*, una célula de *Bacillus licheniformis*, una célula de *Bacillus alkalophilus*, una célula de *Bacillus coagulans*, una célula de *Bacillus circulans*, una célula de *Bacillus pumilis*, una célula de *Bacillus thuringiensis*, una célula de *Bacillus clausii*, una célula de *Bacillus megaterium*, una célula de *Bacillus subtilis* o una célula de *Bacillus amyloliquefaciens*. En otras realizaciones la célula huésped es una célula de *Trichoderma koningii*, una célula de *Trichoderma viride*, una célula de *Trichoderma reesei*, una célula de *Trichoderma longibrachiatum*, una célula de *Aspergillus awamori*, una célula de *Aspergillus fumigates*, una célula de *Aspergillus foetidus*, una célula de *Aspergillus nidulans*, una célula de *Aspergillus niger*, una célula de *Aspergillus oryzae*, una célula de *Humicola insolens*, una célula de *Humicola lanuginosa*, una célula de *Rhodococcus opacus*, una célula de *Rhizomucor miehei* o una célula de *Mucor miceli*.

En aún otras realizaciones, la célula huésped es una célula de *Streptomyces lividans* o una célula de *Streptomyces murinus*. En aún otras realizaciones, la célula huésped es una célula de *Actinomycetes*. En algunas realizaciones la célula huésped es una célula de *Saccharomyces cerevisiae*. En algunas realizaciones, la célula huésped es una célula de *Saccharomyces cerevisiae*. En otras realizaciones, la célula huésped es una célula de una planta eucariota, alga, cianobacteria, bacteria verde del azufre, bacteria verde no del azufre, bacteria púrpura del azufre, bacteria púrpura no del azufre, extremófilo, levadura, hongo, un organismo modificado por ingeniería genética de los mismos, o un organismo sintético. En algunas realizaciones, la célula huésped depende de la luz o fija carbono. En algunas realizaciones, la célula huésped depende de la luz o fija carbono. En algunas realizaciones, la célula huésped tiene actividad autótrofa. En algunas realizaciones, la célula huésped tiene actividad fotoautótrofa, tal como en presencia de luz. En algunas realizaciones, la célula huésped es heterótrofa o mixótrofa en ausencia de luz. En determinadas realizaciones, la célula huésped es una célula de *Arabidopsis thaliana*, *Panicum virgatum*, *Miscanthus giganteus*, *Zea mays*, *Botryococcus braunii*, *Chlamydomonas reinhardtii*, *Dunaliella salina*, *Synechococcus* Sp. PCC 7002, *Synechococcus* Sp. PCC 7942, *Synechocystis* Sp. PCC 6803, *Thermosynechococcus elongatus* BP-1, *Chlorobium tepidum*, *Chloroflexus auranticus*, *Chromatium vinosum*, *Rhodospirillum rubrum*, *Rhodobacter capsulatus*, *Rhodopseudomonas palustris*, *Clostridium ljungdahlii*, *Clostridium thermocellum*, *Penicillium chrysogenum*, *Pichia pastoris*, *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Pseudomonas fluorescens* o *Zymomonas mobilis*.

#### Cultivo de células huésped recombinantes y fermentación

Tal como se usa en el presente documento, el término fermentación se refiere ampliamente a la conversión de materiales orgánicos en sustancias diana por células huésped, por ejemplo, la conversión de una fuente de carbono por células huésped recombinantes en ácidos grasos o derivados de los mismos mediante la propagación de un cultivo de las células huésped recombinantes en un medio que comprende la fuente de carbono. Tal como se usa en

el presente documento, las condiciones permisivas para la producción significan cualesquiera condiciones que permiten que una célula huésped produzca un producto deseado, tal como un ácido graso o derivado de ácido graso. De manera similar, las condiciones en las que se expresa la secuencia de polinucleótido de un vector significan cualesquiera condiciones que permiten que una célula huésped sintetice un polipéptido. Las condiciones adecuadas incluyen, por ejemplo, condiciones de fermentación. Las condiciones de fermentación pueden comprender muchos parámetros incluyendo, pero sin limitarse a, intervalos de temperatura, niveles de aireación, tasas de alimentación y composición de medios. Cada una de estas condiciones, individualmente y en combinación, permiten que crezca la célula huésped. La fermentación puede ser aerobia, anaerobia, o variaciones de las mismas (tales como microaerobias). Los medios de cultivo a modo de ejemplo incluyen caldos o geles. Generalmente, el medio incluye una fuente de carbono que puede metabolizarse por una célula huésped directamente. Además, pueden usarse enzimas en el medio para facilitar la movilización (por ejemplo, la despolimerización de almidón o celulosa para dar azúcares fermentables) y el metabolismo posterior de la fuente de carbono. Para la producción a pequeña escala, las células huésped modificadas por ingeniería genética pueden hacerse crecer en lotes de, por ejemplo, aproximadamente 100 ml, 500 ml, 1 l, 2 l, 5 l o 10 l; fermentarse; e inducirse para expresar una secuencia de polinucleótido deseada, tal como una secuencia de polinucleótido que codifica para un polipéptido de CAR. Para la producción a gran escala, las células huésped modificadas por ingeniería genética pueden hacerse crecer en lotes de aproximadamente 10 l, 100 l, 1000 l, 10.000 l, 100.000 l y 1.000.000 l o más; fermentarse; e inducirse para expresar una secuencia de polinucleótido deseada. Alternativamente, puede llevarse a cabo fermentación semicontinua a gran escala.

## 20 Composiciones de alcoholes grasos

Las composiciones de alcoholes grasos descritas en el presente documento se encuentran en el entorno extracelular del cultivo de células huésped recombinantes y pueden aislarse fácilmente del medio de cultivo. Una composición de alcohol graso puede secretarse por la célula huésped recombinante, transportarse al entorno extracelular o transferirse de manera pasiva al entorno extracelular del cultivo de células huésped recombinantes. La composición de alcohol graso se aísla de un cultivo de células huésped recombinantes usando métodos de rutina conocidos en la técnica. Se describen en el presente documento composiciones producidas por células huésped recombinantes o modificadas por ingeniería genética (bioproductos) que incluyen uno más aldehídos grasos y/o alcoholes grasos. Aunque un componente de alcohol graso con una longitud de cadena y grado de saturación particulares puede constituir la mayoría del bioproducto producido por una célula huésped recombinante o modificada por ingeniería genética cultivada, la composición incluye normalmente una mezcla de aldehídos grasos y/o alcoholes grasos que varían con respecto a la longitud de cadena y/o grado de saturación. Tal como se usa en el presente documento, fracción de carbono moderno o  $f_M$  tiene el mismo significado definido por los materiales de referencia patrón del Instituto Nacional de Normas y Tecnología (NIST) (SRM 4990B y 4990C, conocidos como patrones de ácido oxálico HOxI y HOxII, respectivamente). La definición fundamental se refiere a 0,95 veces la razón isotópica  $^{14}\text{C}/^{12}\text{C}$  HOxI (con referencia a AD 1950). Esto es aproximadamente equivalente a la madera antes de la Revolución Industrial corregida para la desintegración. Para la biosfera viva actual (material vegetal),  $f_M$  es aproximadamente 1,1.

Los bioproductos (por ejemplo, los alcoholes y aldehídos grasos producidos según la presente divulgación) que comprenden compuestos orgánicos producidos biológicamente y, en particular, los alcoholes y aldehídos grasos producidos biológicamente usando la ruta de biosíntesis de ácidos grasos en el presente documento, no se han producido a partir de fuentes renovables y, como tal, son nuevas composiciones de materia. Estos nuevos bioproductos pueden distinguirse de compuestos orgánicos derivados de carbono petroquímico basándose en la obtención de huella isotópica de carbono dual o la datación mediante  $^{14}\text{C}$ . Adicionalmente, la fuente específica de carbono de fuente biológica (por ejemplo, glucosa frente a glicerol) puede determinarse mediante obtención de huella isotópica de carbono dual (véase, por ejemplo, la patente estadounidense n.º 7.169.588). La capacidad para distinguir bioproductos de compuestos orgánicos basados en el petróleo es beneficiosa en la trazabilidad de estos materiales en el comercio. Por ejemplo, pueden distinguirse compuestos orgánicos o químicos que comprenden perfiles isotópicos de carbono tanto de base biológica como de base en el petróleo de compuestos orgánicos y químicos que se componen sólo de materiales basados en el petróleo. Así, los bioproductos en el presente documento pueden someterse a seguimiento o trazabilidad en el comercio basándose en su perfil isotópico de carbono único. Los bioproductos pueden distinguirse de compuestos orgánicos basados en el petróleo comparando la razón isotópica de carbono estable ( $^{13}\text{C}/^{12}\text{C}$ ) en cada muestra. La razón  $^{13}\text{C}/^{12}\text{C}$  en un bioproducto dado es consecuencia de la razón  $^{13}\text{C}/^{12}\text{C}$  en el dióxido de carbono atmosférico en el momento en que se fija el dióxido de carbono. También refleja la ruta metabólica precisa. También se producen variaciones regionales. El petróleo, las plantas  $\text{C}_3$  (las de hoja caduca), las plantas  $\text{C}_4$  (las hierbas) y carbonatos marinos muestran todas diferencias significativas en  $^{13}\text{C}/^{12}\text{C}$  y los valores de  $\delta^{13}\text{C}$  correspondientes. Además, la materia lipídica de plantas  $\text{C}_3$  y  $\text{C}_4$  se analiza de manera diferente que en materiales derivados de los componentes de hidrato de carbono de las mismas plantas como consecuencia de la ruta metabólica. Dentro de la precisión de medición,  $^{13}\text{C}$  muestra grandes variaciones debido a efectos de fraccionamiento isotópico, siendo el más significativo de ellos para bioproductos el mecanismo de fotosíntesis. La principal causa de las diferencias en la razón isotópica de carbono en plantas se asocia estrechamente con diferencias en la ruta de metabolismo de carbono por fotosíntesis en las plantas, particularmente la reacción que se produce durante la carboxilación primaria (es decir, la fijación inicial de  $\text{CO}_2$  atmosférico). Dos grandes clases de vegetación son las que incorporan el ciclo de fotosíntesis  $\text{C}_3$  (o de Calvin-

Benson) y las que incorporan el ciclo de fotosíntesis  $C_4$  (o de Hatch-Slack). En plantas  $C_3$ , la fijación primaria de  $CO_2$  o reacción de carboxilación implica la enzima ribulosa-1,5-difosfato carboxilasa, y el primer producto estable es un compuesto de 3 carbonos. Las plantas  $C_3$ , tales como frondosas y coníferas, son predominantes en las zonas de climas templados. En plantas  $C_4$ , una reacción de carboxilación adicional que implica otra enzima, fosfoenol-piruvato carboxilasa, es la reacción de carboxilación primaria. El primer compuesto carbonado estable es un ácido de 4 carbonos que se descarboxila posteriormente. El  $CO_2$  así liberado vuelve a fijarse mediante el ciclo  $C_3$ . Ejemplos de plantas  $C_4$  son hierbas tropicales, maíz y caña de azúcar. Las plantas tanto  $C_4$  como  $C_3$  muestran un intervalo de razones isotópicas  $^{13}C/^{12}C$ , pero valores típicos son de aproximadamente -7 a aproximadamente -13 por cada mil para plantas  $C_4$  y de aproximadamente -19 a aproximadamente -27 por cada mil para plantas  $C_3$  (véase, por ejemplo, Stuiver *et al.* (1977) Radiocarbon 19:355 (1977)). El carbón y el petróleo se encuentran generalmente en este último intervalo. La escala de medición de  $^{13}C$  se definió originariamente por un cero establecido por la piedra caliza belemnita de Pee Dee (PDB), facilitándose los valores en partes por mil desviaciones con respecto a este material. Los valores de  $\delta^{13}C$  se expresan en partes por milés (por mil), abreviado como, ‰, y se calculan de la siguiente manera:

$$\delta^{13}C \text{ (‰)} = [(^{13}C/^{12}C)_{\text{muestra}} - (^{13}C/^{12}C)_{\text{patrón}}] / (^{13}C/^{12}C)_{\text{patrón}} \times 1000$$

Puesto que el material de referencia (RM) de PDB se ha agotado, una serie de RM alternativos se ha desarrollado en cooperación con IAEA, USGS, NIST, y otros laboratorios de isótopos internacionales seleccionados. Las notaciones para las desviaciones por mil con respecto a PDB es  $\delta^{13}C$ . Se realizan las mediciones con  $CO_2$  mediante espectrometría de masas de razón estable de alta precisión (IRMS) con iones moleculares de masas 44, 45 y 46. Las composiciones descritas en el presente documento incluyen bioproductos producidos mediante cualquiera de los métodos descritos en el presente documento, incluyendo, por ejemplo, productos de alcohol y aldehído graso. Específicamente, el bioproducto puede tener un  $\delta^{13}C$  de aproximadamente -28 o mayor, aproximadamente -27 o mayor, -20 o mayor, -18 o mayor, -15 o mayor, -13 o mayor, -10 o mayor, o -8 o mayor. Por ejemplo, el bioproducto puede tener un  $\delta^{13}C$  de aproximadamente -30 a aproximadamente -15, de aproximadamente -27 a aproximadamente -19, de aproximadamente -25 a aproximadamente -21, de aproximadamente -15 a aproximadamente -5, de aproximadamente -13 a aproximadamente -7 o de aproximadamente -13 a aproximadamente -10. En otros casos, el bioproducto puede tener un  $\delta^{13}C$  de aproximadamente -10, -11, -12 o -12,3. Los bioproductos, incluyendo los bioproductos producidos según la divulgación en el presente documento, también pueden distinguirse de compuestos orgánicos basados en el petróleo comparando la cantidad de  $^{14}C$  en cada compuesto. Dado que el  $^{14}C$  tiene una semivida nuclear de 5730 años, los combustibles basados en el petróleo que contienen carbono más "antiguo" pueden distinguirse de bioproductos que contienen carbono "más nuevo" (véanse, por ejemplo, Currie, "Source Apportionment of Atmospheric Particles", Characterization of Environmental Particles, J. Buffle y H. P. van Leeuwen, eds., 1 de vol. I de IUPAC Environmental Analytical Chemical Series (Lewis Publishers, Inc.) 3-74, (1992)).

La suposición básica en la datación por radiocarbono es que la constancia de la concentración de  $^{14}C$  en la atmósfera conduce a la constancia de  $^{14}C$  en organismos vivos. Sin embargo, debido a las pruebas nucleares atmosféricas desde 1950 y el quemado de combustibles fósiles desde 1850, el  $^{14}C$  ha adquirido una segunda característica temporal geoquímica. Su concentración en el  $CO_2$  atmosférico, y así en la biosfera viva, se duplicó aproximadamente en el apogeo de las pruebas nucleares, a mediados de los años 1960. Desde entonces se ha vuelto gradualmente a la tasa isotópica ( $^{14}C/^{12}C$ ) inicial cosmogénica (atmosférica) de estado estacionario de aproximadamente  $1,2 \times 10^{-12}$ , con una "semivida" de relajación aproximada de 7-10 años. (Esta última semivida no ha de tomarse literalmente; más bien, debe usarse la función de entrada/desintegración nuclear atmosférica detallada para rastrear la variación de  $^{14}C$  atmosférico y de la biosfera desde el comienzo de la era nuclear). Es esta última característica temporal de  $^{14}C$  de la biosfera la que mantiene la promesa de datación anual de carbono reciente de la biosfera. Puede medirse el  $^{14}C$  mediante espectrometría de masas con acelerador (AMS), facilitándose los resultados en unidades de "fracción de carbono moderno" ( $f_M$ ).  $f_M$  se define por los materiales de referencia patrón 4990B y 4990C del Instituto Nacional de Normas y Tecnología (NIST). Tal como se usa en el presente documento, la fracción de carbono moderno ( $f_M$ ) tiene el mismo significado definido por los materiales de referencia patrón (SRM) 4990B y 4990C del Instituto Nacional de Normas y Tecnología (NIST), conocidos como patrones de ácidos oxálicos HOxI y HOxII, respectivamente). La definición fundamental se refiere a 0,95 veces la razón isotópica  $^{14}C/^{12}C$  HOxI (con referencia a 1950). Esto es aproximadamente equivalente a la madera anterior a la Revolución Industrial corregida para la desintegración. Para la biosfera viva actual (material vegetal),  $f_M$  es de aproximadamente 1,1. Esto es aproximadamente equivalente a la madera anterior a la Revolución Industrial corregida para la desintegración. Para la biosfera viva actual (material vegetal),  $f_M$  es de aproximadamente 1,1.

Las composiciones descritas en el presente documento incluyen bioproductos que pueden tener un  $f_M^{14}C$  de al menos aproximadamente 1. Por ejemplo, el bioproducto de la divulgación puede tener un  $f_M^{14}C$  de al menos aproximadamente 1,01, un  $f_M^{14}C$  de aproximadamente 1 a aproximadamente 1,5, un  $f_M^{14}C$  de aproximadamente 1,04 a aproximadamente 1,18, o un  $f_M^{14}C$  de aproximadamente 1,11 a aproximadamente 1,124. Otra medición de  $^{14}C$  se conoce como el tanto por ciento de carbono moderno (pMC). Para un arqueólogo o geólogo que usa dataciones por  $^{14}C$ , 1950 es igual a una edad de "cero años". Esto también representa 100 pMC. El "carbono de bombas" en la atmósfera alcanzó casi el doble del nivel normal en 1963 en el apogeo de las armas termonucleares. Su distribución dentro de la atmósfera se ha aproximado desde su aparición, mostrando valores que son mayores de 100 pMC para

plantas y animales que viven desde 1950. Ha disminuido gradualmente a lo largo del tiempo, siendo el valor de hoy en día próximo a 107,5 pMC. Esto significa que un material de biomasa reciente, tal como maíz, proporcionaría una firma de  $^{14}\text{C}$  próxima a 107,5 pMC. Los compuestos basados en el petróleo tendrán un valor de pMC de cero. La combinación de carbono fósil con carbono de hoy en día dará como resultado una dilución del contenido de pMC de hoy en día. Suponiendo que 107,5 pMC representa el contenido de  $^{14}\text{C}$  de materiales de biomasa de hoy en día y 0 pMC representa el contenido de  $^{14}\text{C}$  de productos basados en el petróleo, el valor de pMC medido para ese material reflejará las proporciones de los dos tipos de componente. Por ejemplo, un material derivado al 100% de soja de hoy en día proporcionará una firma de radiocarbono próxima a 107,5 pMC. Si ese material se diluyera al 50% con productos basados en el petróleo, proporcionaría una firma de radiocarbono de aproximadamente 54 pMC. Se deriva un contenido de carbono de base biológica asignando el "100%" igual a 107,5 pMC y el "0%" igual a 0 pMC. Por ejemplo, una muestra que mide 99 pMC proporcionará un contenido de carbono de base biológica equivalente del 93%. Este valor se denomina el resultado de carbono de base biológica medio y supone que todos los componentes dentro del material analizado se originan o bien a partir de material biológico de hoy en día o bien a partir de material basado en el petróleo. Un bioproducto que comprende uno o más derivados de ácido graso tal como se describe en el presente documento puede tener un pMC de al menos aproximadamente 50, 60, 70, 75, 80, 85, 90, 95, 96, 97, 98, 99 ó 100. En otros casos, un bioproducto descrito en el presente documento puede tener un pMC de entre aproximadamente 50 y aproximadamente 100; entre aproximadamente 60 y aproximadamente 100; entre aproximadamente 70 y aproximadamente 100; entre aproximadamente 80 y aproximadamente 100; entre aproximadamente 85 y aproximadamente 100; entre aproximadamente 87 y aproximadamente 98; o entre aproximadamente 90 y aproximadamente 95. En aún otros casos, un bioproducto descrito en el presente documento puede tener un pMC de aproximadamente 90, 91, 92, 93, 94 ó 94,2.

#### Examen de composiciones de alcoholes grasos producidas por la célula huésped recombinante

Para determinar si las condiciones son suficientes para permitir la expresión, se cultiva una célula huésped recombinante que comprende un gen heterólogo o un gen nativo modificado, por ejemplo, durante aproximadamente 4, 8, 12, 24, 36 ó 48 horas. Durante y/o después del cultivo, pueden obtenerse muestras y analizarse para determinar si el nivel de producción de alcohol graso (título, rendimiento o productividad) es diferente del de la célula parental silvestre correspondiente que no se ha modificado. Por ejemplo, el medio en el que las células huésped se hicieron crecer puede someterse a prueba para detectar la presencia de un producto deseado. Cuando se somete a prueba para detectar la presencia de un producto, pueden usarse ensayos, tales como, pero sin limitarse a, CCF, HPLC, CG/FID, CG/EM, CL/EM, EM. Pueden cultivarse cepas de células huésped recombinantes en volúmenes pequeños (de 0,001 l a 1 l) de medio en placas o matraces de agitación con el fin de examinar para detectar un nivel alterado de producción de especies grasas o alcohol graso. Una vez identificadas las cepas candidatas o "aciertos" a pequeña escala, estas cepas se cultivan en volúmenes más grandes (de 1 l a 1000 l) de medio en biorreactores, tanques y plantas piloto para determinar el nivel preciso de producción de especies grasas o alcohol graso. Estas condiciones de cultivo de volumen grande las usan los expertos en la técnica para optimizar las condiciones de cultivo para obtener la producción deseada de especies grasas o alcohol graso.

#### Utilidad de composiciones de alcoholes grasos y aldehídos grasos

Se usan aldehídos para producir muchos productos químicos especializados. Por ejemplo, se usan aldehídos para producir polímeros, resinas (por ejemplo, Baquelita), colorantes, saborizantes, plastificantes, perfumes, productos farmacéuticos, y otros productos químicos, algunos de los cuales pueden usarse como disolventes, conservantes o desinfectantes. Además, determinados compuestos naturales y sintéticos, tales como vitaminas y hormonas, son aldehídos, y muchos azúcares contienen grupos aldehído. Los aldehídos grasos pueden convertirse en alcoholes grasos mediante reducción química o enzimática. Los alcoholes grasos también tienen muchos usos comerciales. Las ventas anuales a nivel mundial de alcoholes grasos y sus derivados son superiores a mil millones de USD. Los alcoholes grasos de cadena más corta se usan en las industrias cosmética y alimentaria como emulsionantes, emolientes y espesantes. Debido a su naturaleza anfífila, los alcoholes grasos se comportan como tensioactivos no iónicos, que son útiles en productos para el cuidado personal y de uso doméstico, tales como, por ejemplo, detergentes. Además, se usan alcoholes grasos en ceras, gomas, resinas, pomadas y lociones farmacéuticas, aditivos de aceites lubricantes, agentes textiles antiestáticos y de acabado, plastificantes, cosméticos, disolventes industriales y disolventes para grasas. También se describe en el presente documento una composición de tensioactivo o una composición de detergente que comprende un alcohol graso producido mediante cualquiera de los métodos descritos en el presente documento. Un experto habitual en la técnica apreciará que, dependiendo del fin previsto de la composición de tensioactivo o detergente, pueden producirse y usarse diferentes alcoholes grasos. Por ejemplo, cuando los alcoholes grasos descritos en el presente documento se usan como materia prima para la producción de tensioactivos o detergentes, un experto habitual en la técnica apreciará que las características de la materia prima de alcohol graso afectarán a las características de la composición de tensioactivo o detergente producida. Así, las características de la composición de tensioactivo o detergente pueden seleccionarse para producir alcoholes grasos particulares para su uso como materia prima. Una composición de detergente y/o tensioactivo basado en alcohol graso descrita en el presente documento puede mezclarse con otros tensioactivos y/o detergentes bien conocidos en la técnica. En algunas realizaciones, la mezcla puede incluir al menos aproximadamente el 10%, al menos aproximadamente el 15%, al menos aproximadamente el 20%, al menos aproximadamente el 30%, al menos aproximadamente el 40%, al menos aproximadamente el 50%, al menos aproximadamente el 60%, o un intervalo delimitado por dos cualesquiera de los valores anteriores, en peso del

alcohol graso. En otros ejemplos, puede producirse una composición de tensioactivo o detergente que incluye al menos aproximadamente el 5%, al menos aproximadamente el 10%, al menos aproximadamente el 20%, al menos aproximadamente el 30%, al menos aproximadamente el 40%, al menos aproximadamente el 50%, al menos aproximadamente el 60%, al menos aproximadamente el 70%, al menos aproximadamente el 80%, al menos aproximadamente el 85%, al menos aproximadamente el 90%, al menos aproximadamente el 95%, o un intervalo delimitado por dos cualesquiera de los valores anteriores, en peso de un alcohol graso que incluye una cadena de carbonos que tiene 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21 ó 22 carbonos de longitud. Tales composiciones de tensioactivo o detergente también pueden incluir al menos un aditivo, tal como una microemulsión o un tensioactivo o un detergente de fuentes no microbianas tales como aceites vegetales o petróleo, que pueden estar presentes en la cantidad de al menos aproximadamente el 5%, al menos aproximadamente el 10%, al menos aproximadamente el 15%, al menos aproximadamente el 20%, al menos aproximadamente el 30%, al menos aproximadamente el 40%, al menos aproximadamente el 50%, al menos aproximadamente el 60%, al menos aproximadamente el 70%, al menos aproximadamente el 80%, al menos aproximadamente el 85%, al menos aproximadamente el 90%, al menos aproximadamente el 95%, o un intervalo delimitado por dos cualesquiera de los valores anteriores, en peso del alcohol graso. La divulgación se ilustra adicionalmente mediante los siguientes ejemplos. Los ejemplos se proporcionan para fines ilustrativos sólo. No deben interpretarse como limitativos del alcance o contenido de la divulgación de ningún modo.

## Ejemplos

### EJEMPLO 1

#### Modificaciones del huésped de producción - Atenuación de Acil-CoA deshidrogenasa

Este ejemplo describe la construcción de una célula huésped modificada por ingeniería genética en la que la expresión de una enzima de degradación de ácidos grasos está atenuada. Se deletionó el gen de *fadE* de *Escherichia coli* MG1655 (una cepa de *E. coli* K) usando el sistema Lambda Red (también conocido como Red-Driven Integration) descrito por Datsenko *et al.*, Proc. Natl. Acad. Sci. USA 97: 6640- 6645 (2000), con las siguientes modificaciones:

Se usaron los siguientes cebadores para crear la delección de *fadE*:

Del-*fadE*-

F5'-AAAAACAGCAACAATGTGAGCTTTGTTGTAATTATATTGTAACATATTGATTCCGGGGATCCGTCGACC (SEQ ID NO: 9); y

30 Del-*fadE*-

R5'-AAACGGAGCCTTTTCGGCTCCGTTATTCATTTACGCGGCTTCAACTTTCCTGTA GGCTGGAGCTGCTTC (SEQ ID NO: 10)

Se usaron los cebadores Del-*fadE*-F y Del-*fadE*-R para amplificar el casete de resistencia a kanamicina (KmR) del plásmido pKD13 (descrito por Datsenko *et al.*, citado anteriormente) mediante PCR. Entonces se usó el producto de PCR para transformar células de *E. coli* MG1655 electrocompetentes que contenían pKD46 (descrito en Datsenko *et al.*, citado anteriormente) que se habían inducido previamente con arabinosa durante 3-4 horas. Tras un crecimiento de 3 horas en un caldo súper óptimo con medio de represión de catabolitos (SOC) a 37°C, se sembraron en placa las células en placas de agar Luria que contenían 50 µg/ml de kanamicina. Se identificaron colonias resistentes y se aislaron tras una incubación durante la noche a 37°C. Se confirmó la alteración del gen de *fadE* mediante amplificación por PCR usando los cebadores *fadE*-L2 y *fadE*-R1, que se diseñaron para flanquear el gen de *fadE* de *E. coli*.

Los cebadores de confirmación de la delección de *fadE* fueron:

*fadE*-L2 5'-CGGGCAGGTGCTATGACCAGGAC (SEQ ID NO: 11); y

*fadE*-R1 5'-CGCGGCGTTGACCGGCAGCCTGG (SEQ ID NO: 12)

45 Tras confirmarse la delección de *fadE*, se usó una única colonia para eliminar el marcador KmR usando el plásmido pCP20 tal como se describe por Datsenko *et al.*, citado anteriormente. La cepa de *E. coli* MG1655 con el gen de *fadE* deletionado y el marcador KmR eliminado resultante se denominó *E. coli* MG1655 Δ*fadE*, o *E. coli* MG1655 D1. Se comparó la producción de derivados de ácido graso ("especie grasa total") por la cepa de *E. coli* MG1655 con el gen *fadE* deletionado con la producción de derivados de ácido graso por *E. coli* MG1655. Se transformaron las células con el plásmido de producción pDG109 (pCL1920\_P<sub>TRC</sub>\_carBopt\_12H08\_alrAadp1\_fabB[A329G]\_fadR) y se fermentaron en medio mínimo de glucosa. Los datos presentados en la figura 5 muestran que la delección del gen *fadE* no afectó a la producción de derivados de ácido graso.

### EJEMPLO 2

Aumento del flujo a través de la ruta de síntesis de ácidos grasos - mediada por acetil CoA carboxilasa

Los principales precursores para la biosíntesis de ácidos grasos son malonil-CoA y acetil-CoA (figura 1). Se ha sugerido que estos precursores limitan la velocidad de biosíntesis de ácidos grasos (figura 2) en *E. coli*. En este ejemplo, se sobreexpresaron operones de acc sintéticos [accABCD de *Corynebacterium glutamicum* ( $\pm$ birA)] y las modificaciones genéticas condujeron a un aumento de la producción de acetil-coA y malonil-CoA en *E. coli*. En este estudio, se evaluaron dieciséis cepas que contienen iFAB 130 - 145. Se presenta la estructura detallada de iFAB 130 - 145 en la tabla 1 de iFAB, a continuación.

EJEMPLO 3

Aumento del flujo a través de la ruta de biosíntesis de ácidos grasos - iFAB

Producción de derivados de ácido graso:

Las estrategias para aumentar el flujo a través de la ruta de síntesis de ácidos grasos en células huésped recombinantes incluyen tanto la sobreexpresión de genes de biosíntesis de ácidos grasos de *E. coli* nativos como la expresión de genes de biosíntesis de ácidos grasos endógenos de diferentes organismos en *E. coli*. En este estudio, se combinaron genes de biosíntesis de ácidos grasos de diferentes organismos en el genoma de *E. coli* DV2. Se evaluaron dieciséis cepas que contienen iFAB 130 - 145. Se presenta la estructura detallada de iFAB 130 - 145 en la tabla 1 de iFAB, a continuación.

Tabla 1. Componentes encontrados en iFAB 130-145.

Abreviatura	Descripción completa
St fabD	Gen de fabD de <i>Salmonella typhimurium</i>
nSt_fabH	Gen de fabH de <i>Salmonella typhimurium</i> con el RBS nativo
sSt-fabH	Gen de fabH de <i>Salmonella typhimurium</i> con un RBS sintético
Cac fabF	Gen de fabF de <i>Clostridium acetobutylicum</i> (ATCC824)
St fabG	Gen de fabG de <i>Salmonella typhimurium</i>
St fabA	Gen de fabA de <i>Salmonella typhimurium</i>
St fabZ	Gen de fabZ de <i>Salmonella typhimurium</i>
BS_fabI	Gen de fabI de <i>Bacillus subtilis</i>
BS_FabL	Gen de fabL de <i>Bacillus subtilis</i>
Vc_FabV	Gen de fabV de <i>Vibrio cholerae</i>
Ec_FabI	Gen de fabI de <i>Escherichia Coli</i>

Cada "iFAB" incluía diversos genes de fab en el siguiente orden: 1) una enoil-ACP reductasa (BS\_fabI, BS\_FabL, Vc\_FabV o Ec\_FabI); 2) una b-cetoacil-ACP sintetasa III (St\_fabH); 3) una malonil-CoA-ACP transacilasa (St\_fabD); 4) una b-cetoacil-ACP reductasa (St\_fabG); 5) una 3-hidroxi-acil-ACP deshidratasa (St\_fabA o St\_fabZ); 6) una b-cetoacil-ACP sintetasa II (Cac\_fabF). Obsérvese que St\_fabA tiene también actividad trans-2, cis-3-decenoil-ACP isomerasa (ref) y que Cac\_fabF tiene actividades b-cetoacil-ACP sintetasa II y b-cetoacil-ACP sintetasa I (Zhu *et al.*, BMC Microbiology 9:119 (2009)). Véase la tabla 2, a continuación para la composición específica de iFAB 130 - 145. Véanse las figuras 7A y B que proporcionan la representación esquemática del locus iFAB138, incluyendo un diagrama del promotor cat-loxP-T5 integrado en frente de FAB138 (7A); y un diagrama de iT5\_138 (7B).

Tabla 2: Composición de iFAB 130 - 145.

iFab	BS_fabI	BS_fabL	Vc_fabV	Ec_fabI	nSt_fabH	sSt_fabH	St_fabD	St_fabG	St_fabA	St_fabZ	Cac_fabF
ifab130	1	0	0	0	1	0	1	1	1	0	1
ifab131	1	0	0	0	1	0	1	1	0	1	1
ifab132	1	0	0	0	0	1	1	1	1	0	1
ifab133	1	0	0	0	0	1	1	1	0	1	1
ifab134	0	1	0	0	1	0	1	1	1	0	1
ifab135	0	1	0	0	1	0	1	1	0	1	1
ifab136	0	1	0	0	0	1	1	1	1	0	1
ifab137	0	1	0	0	0	1	1	1	0	1	1
ifab138	0	0	1	0	1	0	1	1	1	0	1
ifab139	0	0	1	0	1	0	1	1	0	1	1
ifab140	0	0	1	0	0	1	1	1	1	0	1
ifab141	0	0	1	0	0	1	1	1	0	1	1
ifab142	0	0	0	1	1	0	1	1	1	0	1
ifab143	0	0	0	1	1	0	1	1	0	1	1
ifab144	0	0	0	1	0	1	1	1	1	0	1

ifab145	0	0	0	1	0	1	1	1	0	1	1
---------	---	---	---	---	---	---	---	---	---	---	---

Se transformó el plásmido pCL\_P<sub>trc</sub>\_tesA en cada una de las cepas y se realizó una fermentación en medio FA2 con 20 horas desde la inducción hasta la cosecha en caldo tanto a 32°C como a 37°C. En las figuras 6A y 6B se muestran los datos para la producción de especie grasa total a partir de exámenes en placa por duplicado. A partir de este examen de la biblioteca se determinó que el mejor constructo era DV2 con iFAB138. Se transfirió el constructo iFAB138 en la cepa D178 para preparar la cepa EG149. Se usó esta cepa para la modificación por ingeniería genética adicional. La secuencia de iFAB138 en el genoma de EG149 se presenta como SEQ ID NO: 13. La tabla 3 presenta la caracterización genética de varias cepas de *E. coli* en las que se introdujeron plásmidos que contenían los constructos de expresión descritos en el presente documento tal como se describe a continuación. Estas cepas y plásmidos se usaron para mostrar las células huésped recombinantes, los cultivos y los métodos de determinadas realizaciones de la presente divulgación. Las designaciones genéticas en la tabla 3 son designaciones convencionales conocidas por los expertos habituales en la técnica.

Tabla 3: Caracterización genética de cepas de *E. coli*

Cepa	Caracterización genética
DV2	MG1655 F-, λ-, ilvG-, rfb-50, rph-1, ΔfhuA::FRT, ΔfadE::FRT
DV2.1	DV2 fabB::fabB[A329V]
D178	DV2.1 entD::FRT_P <sub>T5</sub> _entD
EG149	D178 ΔinsH-11::P <sub>LACUV5</sub> -iFAB138
V642	EG149 rph+
SL313	V642 lacI <sub>Z</sub> ::P <sub>A1</sub> 'tesA/pDG109
V668	V642 ilvG <sup>+</sup>
LC397	V668 lacI <sub>Z</sub> ::P <sub>TRC</sub> 'tesA(var)_kan
SL571	V668 lacI <sub>Z</sub> ::P <sub>TRC</sub> 'tesA(var)_FRT
LC942	SL571 attTn7::P <sub>TRC</sub> 'tesA(var)
DG16	LC942/pLC56
V940	LC397/pV171.1
D851	SL571 yijP::Tn5-cat/pV171.1

Plásmidos: pDG109, pLC56 y pV171.1 son el operón pCL\_P<sub>trc</sub>\_carB\_tesA\_alrA\_fabB\_fadR con expresión variable de carB y tesA. iFAB138 es SEQ ID NO: 13.

15 **EJEMPLO 4**

**Aumento de la cantidad de producto de ácido graso libre (FFA) reparando las mutaciones de rph e ilvG**

Se corrigieron las mutaciones de ilvG y rph en esta cepa dando como resultado una producción superior de FFA. Se transformaron las cepas D178, EG149 y V668 (tabla 3) con pCL\_P<sub>trc</sub>\_tesA. Se realizó una fermentación a 32°C en medio FA2 durante 40 horas para comparar la producción de FFA de las cepas D178, EG149 y V668 con pCL\_P<sub>trc</sub>\_tesA. La corrección de las mutaciones de rph e ilvG dio como resultado un aumento del 116% en la producción de FFA de la cepa base con pCL\_P<sub>trc</sub>\_tesA. Tal como se observa en la figura 8, V668/pCL\_P<sub>trc</sub>\_tesA produce más FFA que la D178/pCL\_P<sub>trc</sub>\_tesA, o el control EG149/pCL\_P<sub>trc</sub>\_tesA. Puesto que FFA es un precursor para los productos de LS9, una producción de FFA superior es un buen indicador de que la nueva cepa puede producir niveles superiores de productos de LS9. Se ejecutó la fermentación y extracción según el protocolo de fermentación de FALC convencional ejemplificado por lo siguiente.

Se usó un vial de banco de células congelado de la cepa de *E. coli* seleccionada para inocular 20 ml de caldo LB en un matraz de agitación con deflectores de 125 ml que contenía antibiótico espectinomicina a una concentración de 115 μg/ml. Se incubó este matraz de agitación en un agitador orbital a 32°C durante aproximadamente seis horas, luego se transfirieron 1,25 ml del caldo a 125 ml de medio de siembra FA2 con bajo contenido en P (NH<sub>4</sub>Cl 2 g/l, NaCl 0,5 g/l, KH<sub>2</sub>PO<sub>4</sub> 3 g/l, MgSO<sub>4</sub>·7H<sub>2</sub>O 0,25 g/l, CaCl<sub>2</sub>·2H<sub>2</sub>O 0,015 g/l mM, glucosa 30 g/l, 1 ml/l de disolución de oligoelementos (2 g/l de ZnCl<sub>2</sub>·4H<sub>2</sub>O, 2 g/l de CaCl<sub>2</sub>·6H<sub>2</sub>O, 2 g/l de Na<sub>2</sub>MoO<sub>4</sub>·2H<sub>2</sub>O, 1,9 g/l de CuSO<sub>4</sub>·5H<sub>2</sub>O, 0,5 g/l de H<sub>3</sub>BO<sub>3</sub> y 10 ml/l de HCl concentrado), 10 mg/l de citrato férrico, 100 mM de tampón Bis-Tris (pH 7,0) y 115 μ/ml de espectinomicina), en un matraz de agitación Erlenmeyer con deflectores de 500 ml, y se incubó en un agitador durante la noche a 32°C. Se usaron 100 ml de este cultivo de siembra de FA2 con bajo contenido en P para inocular un biorreactor Biostat Aplus de 5 l (Sartorius BBI), que contenía inicialmente 1,9 l de medio de fermentación de biorreactor F1 esterilizado. Este medio se compone inicialmente de 3,5 g/l de KH<sub>2</sub>PO<sub>4</sub>, 0,5 g/l de (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 0,5 g/l de MgSO<sub>4</sub> heptahidratado, 10 g/l de glucosa filtrada estéril, citrato férrico 80 mg/l, casaminoácidos 5 g/l, 10 ml/l de la disolución de oligoelementos filtrada estéril, 1,25 ml/l de una disolución de vitaminas filtrada estéril (0,42 g/l de riboflavina, 5,4 g/l de ácido pantoténico, 6 g/l de niacina, 1,4 g/l de piridoxina, 0,06 g/l de biotina y 0,04 g/l de ácido fólico), y la espectinomicina a la concentración que se utilizó en el medio de siembra. Se mantuvo el pH del cultivo a 6,9 usando agua amoniacal al 28% p/v, la temperatura a 33°C, la velocidad de aireación 1 lpm (0,5 v/v/m), y la tensión de oxígeno disuelto al 30% de saturación, utilizando el bucle de agitación en cascada con el controlador

de DO y complementación de oxígeno. Se controló la formación de espuma mediante la adición automatizada de un antiespumante basado en emulsión de silicona (Dow Corning 1410).

Se inició una alimentación de nutrientes compuesta por  $\text{MgSO}_4$  heptahidratado 3,9 g/l y glucosa 600 g/l cuando la glucosa en el medio inicial casi se había agotado (aproximadamente 4-6 horas tras la inoculación) bajo una velocidad de alimentación exponencial de  $0,3 \text{ h}^{-1}$  hasta una velocidad de alimentación de glucosa máxima constante de 10-12 g/l/h, basándose en el volumen de fermentación nominal de 2 l. Se indujo la producción de alcohol graso en el biorreactor cuando el cultivo alcanzó una DO de 5 AU (aproximadamente 3-4 horas tras la inoculación) mediante la adición de una disolución madre de IPTG 1 M hasta una concentración final de 1 mM. Se tomaron muestras del biorreactor dos veces al día después de eso, y se cosechó aproximadamente 72 horas tras la inoculación. Se transfirió una muestra de 0,5 ml del caldo de fermentación bien mezclado a un tubo cónico de 15 ml (VWR), y se mezcló concienzudamente con 5 ml de acetato de butilo. Se invirtió el tubo varias veces para mezclar, entonces se agitó con vórtex vigorosamente durante aproximadamente dos minutos. Entonces se centrifugó el tubo durante cinco minutos para separar las fases acuosa y orgánica, y se transfirió una parte de la fase orgánica a un vial de vidrio para el análisis por cromatografía de gases.

## 15 EJEMPLO 5

### Aumento de la producción de Alcohol graso mediante mutagénesis de transposón - yijP

Para mejorar el título, el rendimiento, la productividad de la producción de alcohol graso por *E. coli*, se llevaron a cabo mutagénesis de transposón y examen de alto rendimiento y se secuenciaron mutaciones beneficiosas. Se mostró que una inserción de transposón en la cepa yijP mejoraba el rendimiento de alcohol graso de la cepa en fermentaciones tanto en matraz de agitación como semicontinuas. La cepa SL313 produce alcoholes grasos. El genotipo de esta cepa se proporciona en la tabla 3. Se sometieron los clones de transposón a examen de alto rendimiento para medir la producción de alcoholes grasos. En resumen, se recogieron colonias en placas de pocillos profundos que contenían LB, se hicieron crecer durante la noche, se inocularon en LB nuevo y se hicieron crecer durante 3 horas, se inocularon en medio FA2.1 nuevo, se hicieron crecer durante 16 horas, luego se extrajeron usando acetato de butilo. Se derivatizó el extracto en bruto con BSTFA (N,O-bis[trimetilsilil]trifluoroacetamida) y se analizó usando CG/FID. Se incluyó espectinomicina (100 mg/l) en todos los medios para mantener la selección del plásmido pDG109. Se seleccionaron aciertos eligiendo clones que producían una especie grasa total similar que la cepa control SL313, pero que tenían un mayor porcentaje de especies de alcohol graso y un menor porcentaje de ácidos grasos libres que el control. Se identificó la cepa 68F11 como un acierto y se validó en una fermentación en matraz de agitación usando el medio FA2.1. Una comparación del acierto de transposón 68F11 con la cepa control SL313 indica que 68F11 produce un mayor porcentaje de especies de alcohol graso que el control, mientras que ambas cepas producen títulos similares de especie grasa total. Se secuenció una única colonia del acierto 68F11, denominada LC535, para identificar la ubicación de la inserción del transposón. En resumen, se purificó el ADN genómico a partir de un cultivo en LB durante la noche de 10 ml usando el kit ZR Fungal/Bacterial DNA MiniPrep™ (Zymo Research Corporation, Irvine, CA) según las instrucciones del fabricante. Se secuenció el ADN genómico purificado fuera del transposón usando cebadores internos para el transposón:

DG150 5'-GCAGTTATTGGTGCCCTTAAACGCCTGGTTGCTACGCCTG-3' (SEQ ID NO: 14)

DG131 5'-GAGCCAATATGCGAGAACACCCGAGAA-3' (SEQ ID NO: 15)

Se determinó que la cepa LC535 tiene una inserción de transposón en el gen de yijP (figura 18). yijP codifica para una proteína de la membrana interna conservada cuya función no está clara. El gen de yijP está en un operón y se transcribe conjuntamente con el gen de ppc, que codifica para fosfoenolpiruvato carboxilasa, y el gen de yijO, que codifica para un regulador de la transcripción de unión a ADN predicho de función desconocida. Los promotores internos para el transposón probablemente tienen efectos sobre el nivel y el momento de la transcripción de yijP, ppc e yijO, y también pueden tener efectos sobre genes adyacentes, frwD, pflC, pflD y argE. En la figura 18 se muestran los promotores internos para el casete de transposón, y pueden tener efectos sobre la expresión génica adyacente. Se evaluó la cepa LC535 en una fermentación semicontinua en dos fechas diferentes. Ambas fermentaciones demostraron que LC535 producía alcoholes grasos con un rendimiento mayor que SL313 de control, y la mejora era del 1,3-1,9% del rendimiento absoluto basándose en la entrada de carbono. Se evaluó adicionalmente el casete de transposón de yijP en una cepa diferente, V940, que produce alcohol graso a un rendimiento mayor que la cepa SL313. Se amplificó el casete yijP::Tn5-cat de la cepa LC535 usando los cebadores:

LC277 5'-CGCTGAACGTATTGCAGGCCGAGTTGCTGCACCGCTCCCGCCAGGCAG-3' (SEQ ID NO: 16)

LC278 5'-GGAATTGCCACGGTGCGGCAGGCTCCATACGCGAGGCCAGGTTATCCAACG-3' (SEQ ID NO: 17)

Se introdujo por electroporación este ADN lineal en la cepa SL571 y se integró en el cromosoma usando el sistema de recombinación Lambda Red. Se examinaron colonias usando los cebadores fuera de la región del transposón:

55 DG407 5'-AATCACCAGCACTAAAGTGCGCGGTTGTTACCCG-3' (SEQ ID NO: 18)

DG408 5'-ATCTGCCGTGATTGCAGAGTCTATTCAGCTACG-3' (SEQ ID NO: 19)

Se transformó una colonia con el casete de transposón de yijP correcto (figura 9) con el plásmido de producción pV171.1 para producir la cepa D851. Se sometió a prueba D851 (V940 yijP::Tn5-cat) en una fermentación en matraz de agitación frente a la cepa isogénica V940 que no contiene el casete de transposón de yijP. El resultado de esta fermentación mostró que el casete de transposón de yijP confiere la producción de un mayor porcentaje de alcohol graso por la cepa D851 en relación con la cepa V940 y produce títulos similares de especie grasa total que la cepa de control V940. Se evaluó la cepa D851 en una fermentación semicontinua en dos fechas diferentes. En la tabla 4 se muestran los datos de estas fermentaciones, que ilustra que en fermentaciones semicontinuas de 5 l, las cepas con la inserción del transposón yijP::Tn5-cat tenían un rendimiento de especie grasa total ("FAS") aumentado y un aumento en el porcentaje de alcohol graso ("FALC"). Las "especies grasas" incluyen FALC y FFA.

10 Tabla 4: Efecto de la inserción del transposón de yijp sobre el título y rendimiento de FAS y FALC

Cepa	Título de FAS	Rendimiento de FAS	Porcentaje de FALC	Rendimiento de FALC
V940	68 g/l	18,70%	95,00%	17,80%
D851	70 g/l	19,40%	96,10%	18,60%
V940	64 g/l	18,40%	91,90%	16,90%
D851	67 g/l	19,00%	94,00%	17,80%

Método de fermentación en tanque:

Para evaluar la producción de ésteres de ácidos grasos en tanque, se usó un vial en glicerol de la cepa deseada para inocular 20 ml de LB + espectinomicina en matraz de agitación y se incubó a 32°C durante aproximadamente seis horas. Se usaron 4 ml de cultivo en LB para inocular 125 ml de medio de siembra con bajo contenido en PFA (a continuación), que entonces se incubó a 32°C en un agitador durante la noche. Se usaron 50 ml del cultivo durante la noche para inocular 1 l de medio de tanque. Se hicieron funcionar los tanques a pH 7,2 y 30,5°C en condiciones de pH estacionario con una velocidad de alimentación máxima de 16 g/l/h (glucosa o metanol).

15

Tabla 5: Medio de siembra con bajo contenido en PFA

Componente	Concentración
NH <sub>4</sub> Cl	2 g/l
NaCl	0,5 g/l
KH <sub>2</sub> PO <sub>4</sub>	1 g/l
MgSO <sub>4</sub> -7H <sub>2</sub> O	0,25 g/l
CaCl <sub>2</sub> -2H <sub>2</sub> O	0,015 g/l
Glucosa	20 g/l
Disolución de oligoelementos TM2	1 ml/l
Citrato férrico	10 mg/l
Tampón Bis Tris (pH 7,0)	100 mM
Espectinomicina	115 mg/l

Tabla 6: Medio de tanque

Componente	Concentración
(NH <sub>4</sub> ) <sub>2</sub> SO <sub>4</sub>	0,5 g/l
KH <sub>2</sub> PO <sub>4</sub>	3,0 g/l
Citrato férrico	0,034 g/l
Disolución de oligoelementos TM2	10 ml/l
Casaminoácidos	5 g/l
Adiciones tras la esterilidad	
MgSO <sub>4</sub> -7H <sub>2</sub> O	2,2 g/l
Disolución de vitaminas traza	1,25 ml/l
Glucosa	5 g/l
Inóculo	50 ml/l

20 **EJEMPLO 6**

**Adición de una etiqueta de fusión de 60 pb N-terminal a CarB (CarB60)**

Hay muchos modos de aumentar la solubilidad, estabilidad, expresión o funcionalidad de una proteína. En un enfoque para aumentar la solubilidad de CarB, pudo clonarse una etiqueta de fusión antes del gen. En otro enfoque para aumentar la expresión de CarB, pudo alterarse el promotor o sitio de unión al ribosoma (RBS) del gen. En este estudio, se modificó carB (SEQ ID NO: 7) mediante la adición de una etiqueta de fusión de 60 pb N-terminal. Para generar la proteína modificada (denominada en el presente documento "CarB60"), se clonó el primer lugar *carB* en el vector pET15b usando los cebadores:

25

5'-GCAATTCATATGACGAGCGATGTTACGA-3' (SEQ ID NO: 20); y

5'-CCGCTCGAGTAAATCAGACCGAACTCGCG (SEQ ID NO: 21).

El constructo de pET15b-carB contenía 60 nucleótidos directamente en el sentido de 5' del gen carB:

5'-ATGGGCAGCAGCCATCATCATCATCACAGCAGCGGCTGGTGCCGCGCGGCAGCCAT (SEQ ID NO: 22)

5 La versión con etiqueta de fusión de carB se renombró carB60. Entonces se digirió el pET15b\_carB60 usando las enzimas de restricción NcoI y HindIII y se subclonó en el vector derivado de pCL1920 OP80 que se cortó con las mismas enzimas. Se transformó este plásmido en la cepa V324 (MG1655  $\Delta$ fadE::FRT  $\Delta$ fhuA::FRT fabB::A329V entD::T5-entD lacZ::P<sub>TRC</sub>-TesA) para evaluar la producción de FALC. Se fermentaron las cepas según un procedimiento convencional (resumido más adelante) y se cuantificaron el título de especie grasa total y el título de alcohol graso total. La figura 10 muestra que CarB60 aumenta los títulos de alcohol graso y por tanto la enzima CarB60 tiene una actividad celular total mayor que CarB cuando se expresa a partir de un plásmido de múltiples copias.

15 Para evaluar la producción de alcoholes grasos en cepas de producción, se hicieron crecer transformantes en 2 ml de caldo LB complementado con antibióticos (100 mg/l) a 37°C. Tras el crecimiento durante la noche, se transfirieron 40  $\mu$ l de cultivo a 2 ml de LB nuevo complementado con antibióticos. Tras 3 horas de crecimiento, se transfirieron 2 ml de cultivo a un matraz de 125 ml que contenía 20 ml de medio M9 con glucosa al 3% complementado con 20  $\mu$ l de disolución de oligoelementos, citrato de hierro 10  $\mu$ g/l, tiamina 1  $\mu$ g/l, y antibióticos (medio FA2). Cuando la DO<sub>600</sub> del cultivo alcanzó 1,0, se añadió 1 mM de IPTG a cada matraz. Tras 20 horas de crecimiento a 37°C, se retiraron muestras de 400  $\mu$ l de cada matraz y se extrajeron los alcoholes grasos con 400  $\mu$ l de acetato de butilo. Para entender adicionalmente el mecanismo de la actividad de CarB mejorada, se purificó CarB60 a partir de la cepa D178 que no contiene 'TesA (MG1655  $\Delta$ fadE::FRT  $\Delta$ fhuA::FRT fabB::A329V entD::P<sub>T5</sub>-entD). En resumen, se transformó pCL1920\_carB60 en la cepa D178, que se había modificado por ingeniería genética para la producción de alcohol graso, y se llevó a cabo la fermentación a 37°C en medio FA-2 complementado con espectinomicina (100  $\mu$ g/ml). Cuando el cultivo alcanzó DO<sub>600</sub> de 1,6, se indujeron las células con isopropil- $\beta$ -D-tiogalactopiranosido (IPTG) 1 mM y se incubaron durante 23 h adicionales a 37°C. Para la purificación de CarB60, se cosecharon las células mediante centrifugación durante 20 min a 4°C a 4.500 rpm. Se suspendió la pasta celular (10 g) en 12 ml de mezcla maestra BugBuster (Novagen) y disolución de cóctel inhibidor de proteasas. Se rompieron las células mediante prensa francesa y se centrifugó el homogenado resultante a 10.000 rpm para eliminar los residuos celulares. Se añadió Ni-NTA a la mezcla resultante, y se agitó la suspensión a 4°C a 100 rpm durante 1 hora en un agitador rotatorio. Se vertió la suspensión en una columna, y se recogió la fracción no retenida. Se lavó la resina de Ni-NTA con imidazol 10 mM en tampón fosfato de sodio 50 mM pH 8,0 que contenía NaCl 300 mM, y se lavó adicionalmente con imidazol 20 mM en tampón fosfato de sodio 50 mM pH 8,0 que contenía NaCl 300 mM. Se eluyó la proteína CarB60 con imidazol 250 mM en tampón fosfato de sodio 50 mM pH 8,0 que contenía NaCl 300 mM, y se analizó mediante SDS-PAGE. Se dializó la proteína frente a glicerol al 20% (v/v) en tampón fosfato de sodio 50 mM pH 7,5 produciendo aproximadamente 10 mg de CarB60 por litro de cultivo. Se congeló inmediatamente la proteína y se almacenó a -80°C hasta que se necesitó.

40 La proteína CarB60 se expresaba abundantemente a partir de un plásmido de múltiples copias. El análisis de SDS-PAGE adicional mostró que la expresión de CarB60 era superior a la de CarB. El nivel de expresión superior de CarB60 sugirió que el gen de carB60 integrado en el cromosoma de *E. coli* produciría más proteína que el gen de carB en la misma ubicación. Para someter a prueba esta hipótesis, se integró el gen de carB60 en el cromosoma de *E. coli*. En resumen, se amplificó en primer lugar el gen de carB60 a partir de pCL\_carB60 usando el cebador directo:

5'-ACGGATCCCCGGAATGCGCAACGCAATTAATGTaAGTTAGCGC-3' (SEQ ID NO: 23); y el cebador inverso:

5'-TGCATCATCGCCATTGAATTCCTAAATCAGACCGAACTCGCGCAGG-3' (SEQ ID NO: 24).

Se amplificó un segundo producto de PCR a partir del vector pAH56 usando el cebador directo:

45 5'-ATTCCGGGGATCCGTCGACC-3' (SEQ ID NO: 25); y el cebador inverso:

5'-AATGGCGATGACGCATCCTCACG-3' (SEQ ID NO: 26)

50 Este fragmento contiene un casete de resistencia a kanamicina, el sitio  $\lambda$ attP, y un origen de replicación  $\lambda$ R6k. Se unieron los dos productos de PCR usando el kit InFusion (Clontech) para crear el plásmido pSL116-126. Se transformó una cepa de producción de alcohol graso que contenía una forma integrada de 'TesA12H08 y un plásmido auxiliar pINT con o bien pSL116-126 que contenía el gen de carB60 o bien el plásmido F27 que contenía el gen de carB. Se fermentaron estas cepas en medio FA2 según procedimientos convencionales para fermentaciones en matraz de agitación, tal como se describió anteriormente. Para caracterizar y cuantificar los alcoholes grasos y ésteres de ácidos grasos, se usó la detección por cromatografía de gases ("CG") acoplada con ionización de llama ("FID"). Se derivatizó el extracto en bruto con BSTFA (N,O-bis[trimetilsilil]trifluoroacetamida) y se analizó usando una CG/FID. Se llevó a cabo la cuantificación inyectando diversas concentraciones de las referencias auténticas apropiadas usando el método de CG descrito anteriormente así como ensayos que incluyen, pero no se limitan a,

cromatografía de gases (CG), espectrometría de masas (EM), cromatografía en capa fina (CCF), cromatografía de líquidos de alta resolución (HPLC), cromatografía de líquidos (CL), CG acoplada con un detector de ionización de llama (CG-FID), CG-EM y CL-EM. Cuando se somete a prueba para detectar la expresión de un polipéptido, pueden usarse técnicas tales como inmunotransferencia de tipo Western y transferencia puntual.

- 5 En la figura 11 se muestran los resultados de la fermentación tras 20 horas. Los títulos de productos grasos totales de las dos cepas son similares (especie grasa total 2,4 g/l), sin embargo CarB60 integrada convierte una mayor fracción de ácidos grasos libres de longitud de cadena C12 y C14 en alcoholes grasos, en comparación con CarB sin la etiqueta N-terminal. Estos datos sugieren que células que expresan CarB60 tienen una actividad ácido carboxílico reductasa celular total superior, y pueden convertir más FFA en alcoholes grasos. Por tanto, carB60, cuando se integra en el cromosoma, es un molde de carB mejorado que proporciona la actividad deseada para hacer evolucionar el gen de carB para identificar variantes de carB mejoradas.

### EJEMPLO 7

#### Generación de mutantes de CarB

- 15 La enzima CarB es una etapa limitante de la velocidad en la producción de alcoholes grasos en determinadas condiciones del procedimiento. Para producir alcoholes grasos de manera económica, se hicieron esfuerzos para aumentar la actividad de la enzima CarB.

Examen de la biblioteca de PCR propensa a error:

- 20 Se realizó mutagénesis al azar usando PCR propensa a error en condiciones en las que la fidelidad de copiado de la ADN polimerasa es baja. Se clonaron los ácidos nucleicos mutagenizados en un vector, y se realizó PCR propensa a error seguido por examen de alto rendimiento para encontrar mutaciones beneficiosas que aumentarían la conversión de ácidos grasos libres en alcoholes grasos (tal como se detalla a continuación). Se mutaron adicionalmente residuos importantes para dar otros aminoácidos. Varias mutaciones de aminoácidos individuales y combinaciones de mutaciones aumentaron la fracción de especies grasas que se convertían en alcoholes grasos. En resumen, se generaron mutaciones al azar en el gen de carB60opt mediante PCR propensa a error usando el kit Genemorph II (Stratagene). Se generaron mutaciones en sólo uno de dos dominios de carB60opt por separado, para facilitar la clonación. La biblioteca 1 contenía los primeros 759 residuos de carB60opt y se generó mediante PCR propensa a error usando los cebadores:

HZ117 5'-ACGGAAGGAGCTAGCACATGGGCAGCAGCCATCATCAT-3' (SEQ ID NO: 27); y

DG264 5'-GTAAAGGATGGACGGCGGTCACCCGCC-3' (SEQ ID NO: 28).

- 30 El vector para la biblioteca 1 era el plásmido pDG115 digerido con las enzimas NheI y PshAI. La biblioteca 2 contenía los últimos 435 residuos de carB60opt y se generó mediante PCR propensa a error usando los cebadores:

DG263 5'-CACGGCGGGTGACCGCCGTCCATCC-3' (SEQ ID NO: 29); y

HZ118 5'-TTAATTCCGGGGATCCCTAAATCAGACCGAACTCGCGCAGGTC-3' (SEQ ID NO: 30).

- 35 El vector para la biblioteca 2 era el plásmido pDG115 digerido con las enzimas PshAI y BamHI. Se clonaron los insertos propensos a error en los vectores usando InFusion Advantage (Clontech) y se hicieron pasar a través de la cepa de clonación NEB Turbo (New England Biolabs). Entonces se transformaron las bibliotecas en la cepa EG442 (EG149 Tn7::P<sub>TRC</sub>-ABR lacZ::P<sub>T50</sub>-ABR). Entonces se sometieron los clones de carB60opt propensos a error a examen de alto rendimiento para medir la producción de alcoholes grasos. En resumen, se recogieron colonias en placas de pocillos profundos que contenían LB, se hicieron crecer durante la noche, se inocularon en LB nuevo y se hicieron crecer durante 3 horas, se inocularon en medio FA-2.1 nuevo, se hicieron crecer durante 16 horas, luego se extrajeron usando acetato de butilo. Se derivatizó el extracto en bruto con BSTFA (N,O-bis[trimetilsilil]trifluoroacetamida) y se analizó usando un método de CG/FID convencional. Se incluyó espectinomicina (100 mg/l) en todos los medios para mantener la selección del plásmido pDG115. Se seleccionaron aciertos eligiendo clones que producían un título de ácidos grasos libres más pequeño y un título de ácidos grasos totales más grande en comparación con la cepa de control. Para comparar aciertos de diferentes exámenes de fermentación, se normalizó la conversión de ácidos grasos libres en alcoholes grasos calculando un porcentaje de ácidos grasos libres normalizado NORM FFA = porcentaje de FFA mutante / porcentaje de FFA de control en donde "porcentaje de FFA" es el título de la especie de ácidos grasos totales dividido entre el título de la especie grasa total. Se sometieron los aciertos a verificación adicional usando fermentaciones en matraz de agitación, tal como se describe a continuación.

Se secuenciaron los aciertos para identificar mutaciones beneficiosas. Se realizó la secuenciación mediante PCR de colonias de todo el gen de carB60opt usando los cebadores

SL59 5'-CAGCCGTTTATTGCCGACTGGATG-3' (SEQ ID NO: 31); y

EG479 5'-CTGTTTTATCAGACCGCTTCTGCGTTC-3' (SEQ ID NO: 32), y se secuenciaron usando cebadores

internos para la enzima carB60opt.

En la tabla 7 se muestran las mutaciones beneficiosas que mejoran la enzima CarB60opt. La columna de ácidos grasos libres normalizados (NORM FFA) indica la mejora en la enzima, indicando los valores inferiores la mejor mejora. "Pocillo n.º" indica el pocillo de examen primario en el que se encontró esta mutación. Todos los números de residuos se refieren a la secuencia de proteína CarB, que no incluye la etiqueta de 60 pb. Las mutaciones indicadas con el prefijo "Tag:" indican mutaciones en la etiqueta N-terminal de 60 pb/20 residuos.

5

Tabla 7: Mutaciones beneficiosas en la enzima CarB identificada durante el examen propenso a error (mutaciones de TAG eliminadas)

Pocillo n.º	NORM FFA	Mutaciones de cambio de sentido	Mutaciones silenciosas
131B08	70.50%	L799M V810F S927R M1062L A1158V F1170I	CCG1116CCT
20C07	71.80%	A535S	
65B02	74.70%	M530R	ACC867ACA
54B10	75.30%	L80Q T231M F283L A418T V530M A541V G677D F712A	
67E1	78.20%	D750G R827C D985G G1026D P1149S	GCA1C31GCT GTC1073GTT
65C03	78.90%	V926A	ATT941ATA
12C10	80.30%	V46I	
65E08	80.10%	V926A	
70002	80.90%	D750G R827C D985G G1026D P1149S	GCA1C31CCT GTC1073GTT
07D01	82.40%	E20K V191A	
65G09	82.40%	R827C L1128S	ACG78CACA CTG923TTG
25402	83.50%	F288S	
05C01	85.10%	V46I	06C01
05D02	85.20%	T396S	CCG477CCT
124E03	85.00%	R827C L1128S	ACG78CACA CTG923TTG
17A04	85.20%	A574T	GCA237GCT ACC676ACT GCC529GCT
132C08	87.00%	M1062T R1080H	TTG830TTA TAC834TAT
72C09	87.30%	P809L M1062V	
10002	87.70%	E636K	
71403	88.10%	R827C L1128S	ACG78CACA CTG923TTG
38G04	88.90%	D143E A612T	GCA181GCG
92008	90.20%	I90M	CTG185CTT
65C01	90.30%	L1128S	
18C03	90.40%	Q473L	
12E02	90.60%	D19N S22N R87F L416S	CCG167CCA
28809	91.10%	E28K H212N Q473L	CCG122CCA ACG178ACA CTG283TTG CTG54CCTA ACC401ACT GCA681GCG
103E09	92.20%	E936K P1134R	CGT829CGG CTG1007CTA
03F09	93.20%	M259I	
74G11	93.80%	I870V S927I S985I I1164F	GTG1000GTC
46C01	95.60%	D18V D292N	

10 Mutagénesis de saturación (biblioteca conjunta 1 y 2 generada):

Se sometieron posiciones de aminoácidos que se consideraban beneficiosas para la producción de alcohol graso tras la PCR propensa a error a mutagénesis adicional. Se usaron cebadores que contenían los nucleótidos degenerados NNK o NNS para mutar estas posiciones para dar otros aminoácidos. Se examinaron las "bibliotecas de mutagénesis de saturación" resultantes tal como se describió anteriormente para las bibliotecas propensas a error, y se identificaron aciertos que mejoraban adicionalmente la conversión en alcohol graso (un título de ácidos grasos libres total más pequeño y un título de alcoholes grasos total más grande en comparación con la cepa "control" parental). En la tabla 8 se muestran cambios de codones/aminoácidos individuales en nueve posiciones diferentes que mejoran la producción de alcoholes grasos. Se sometieron los aciertos a verificación adicional usando fermentaciones en matraz de agitación, tal como se describe en el presente documento.

15

20 Tabla 8: Mutaciones beneficiosas en la enzima CarB identificadas durante mutagénesis de saturación de aminoácidos

Aminoácido silvestre	Codón silvestre	Aminoácido mutante	Codón mutante	NORM FFA
E20	GAG	F	TTC	92,20%
		L	CTG	94,50%
		L	TTG	96,20%
		R	CGC	86,50%
		S	TCG	87,40%
		V	GTG	86,00%

ES 2 613 553 T3

		V	GTC	85,30%
		Y	TAC	88,80%
V191	GTC	A	GCC	88,70%
		S	AGT	98,00%
F288	TTT	G	GGG	70,30%
		R	AGG	77,20%
		S	TCT	85,60%
		S	AGC	79,60%
Q473	CAA	A	GCG	89,50%
		F	TTC	89,10%
		H	CAC	84,10%
		I	ATC	77,20%
		K	AAG	90,30%
		L	CTA	90,10%
		M	ATG	89,00%
		R	AGG	88,00%
		V	GTG	89,20%
		W	TGG	84,50%
		Y	TAC	86,00%
A535	GCC	A	TCC	71,80%
R827	CGC	A	GCC	93,20%
		C	TGT	87,90%
		C	TGC	83,20%
V926	GTT	A	GCT	78,10%
		A	GCG	66,30%
		A	GCC	69,50%
		E	GAG	65,80%
		G	GGC	78,60%
S927	AGC	G	GGG	77,60%
		G	GGT	79,30%
		I	ATC	90,80%
		K	AAG	70,70%
		V	GTG	87,90%
M930	ATG	K	AAG	82,30%
		R	CGG	73,80%
		R	AGG	69,80%
L1128	TTG	A	GCG	92,70%
		G	GGG	89,70%
		K	AAG	94,80%
		M	ATG	95,80%
		P	CCG	98,40%
		R	AGG	90,90%
		R	CGG	88,50%
		S	TCG	88,90%
		T	ACG	96,30%
		V	GTG	93,90%
		W	TGG	78,80%
		Y	TAC	87,90%

A continuación se combinaron sustituciones de aminoácidos que se consideraban beneficiosas para la producción de alcohol graso. Se usó PCR para amplificar partes del gen de carBopt que contenía diversas mutaciones deseadas, y se unieron las partes entre sí usando un método basado en PCR (Horton, R.M., Hunt, H.D., Ho, S.N., Pullen, J.K. y Pease, L.R. 1989). Se examinó el gen de carBopt sin la etiqueta N-terminal de 60 pb. Las mutaciones combinadas en esta biblioteca de combinación se muestran en la tabla 9.

5

Tabla 9: Mutaciones de CarB de la primera biblioteca de combinación

Mutación	Codón
E20V	GTG
E20S	TCG
E20R	CGC
V191S	AGT
F288R	AGG
F288S	AGC

## ES 2 613 553 T3

F288G	GGG
Q473L	CTG
Q473W	TGG
Q473Y	TAC
Q473I	ATC
Q473H	CAC
A535S	TCC

5 Para facilitar el examen, se integró entonces la biblioteca de combinación de CarB resultante en el cromosoma de la cepa V668 en el locus lacZ. La secuencia del gen de carBopt en este locus se presenta como SEQ ID NO: 7. El genotipo de la cepa V668 es MG1655 ( $\Delta$ fadE::FRT  $\Delta$ fhuA::FRT  $\Delta$ fabB::A329V  $\Delta$ entD::T5-entD  $\Delta$ insH-11:: P<sub>lacUV5</sub> fab138 rph+ ilvG+) (tal como se muestra en la tabla 3 y la figura 16). Entonces se transformaron las cepas con el plásmido pVA3, que contiene TesA, una enzima CarB catalíticamente inactiva CarB[S693A] que destruye el sitio de unión de fosfopanteteína, y otros genes que aumentan la producción de ácidos grasos libres. Se examinó la biblioteca de combinación tal como se describió anteriormente para la biblioteca propensa a error. Se usó como control V668 con carB opt (A535S) integrado en la región lacZ y que contenía pVA3. Se seleccionaron aciertos que aumentaban la producción de alcoholes grasos y se sometieron a verificación adicional usando fermentaciones en matraz de agitación, tal como se describe en el ejemplo 5. En la figura 12 se muestra el porcentaje mejorado de producción de alcohol graso tras la fermentación en matraz de agitación de células huésped recombinantes que expresan mutantes de combinación de CarB.

Se amplificaron los mutantes de combinación de CarB integrados a partir de los aciertos de carB integrados mediante PCR usando los cebadores:

15 EG58 5'-GCACTCGACCGGAATTATCG (SEQ ID NO: 33); y  
EG626 5'-GCACTACGCGTACTGTGAGCCAGAG (SEQ ID NO: 34).

Se reamplificaron estos insertos usando los cebadores:

DG243 5'-GAGGAATAAACCATGACGAGCGATGTTACGACGCGACCGACGGC (SEQ ID NO: 35); y  
DG210 5'-CTAATCAGACCGAACTCGCGCAGG (SEQ ID NO: 36).

20 Usando clonación InFusion, se clonaron los mutantes de carB agrupados en un plásmido de producción, pV869, que se amplificó por PCR usando los cebadores:

DG228 5'-CATGGTTTATTCCTCCTTATTTAATCGATAC (SEQ ID NO: 37); y  
DG318 5'-TGACCTGCGCGAGTTCGGTCTGATTTAG (SEQ ID NO: 38).

25 El mutante de carB que tenía el mejor rendimiento en el examen de plásmido de fermentación en matraz de agitación (carB2; tabla 11) se designó VA101 y la cepa de control que portaba carBopt [A535S] se designó VA82. Véase la figura 13.

30 Se combinaron sustituciones de aminoácidos en el dominio de reducción de carB consideradas beneficiosas para la producción de alcohol graso con uno de los mejores aciertos de la biblioteca de combinación de carB-L, "carB3" (tabla 11). Se usó PCR para amplificar partes del gen de carBopt que contenían diversas mutaciones deseadas en el dominio de reducción, y se unieron las partes entre sí usando SOE PCR. En la tabla 10 se muestran las mutaciones combinadas en esta biblioteca de combinación.

Tabla 10: Mutaciones de CarB a partir de la segunda biblioteca de combinación

Mutación	Codón
R827C	TGC
R827A	GCA
V926A	GCG
V926E	GAG
S927K	AAG
S927G	GGG
M930K	AAG
M930R	AGG
L1128W	TGG

35 Se examinó la biblioteca de combinación tal como se describió anteriormente para la biblioteca propensa a error. Se usó como control V668 con carB3 integrado en la región de lacZ y que contenía pVA3. Se seleccionaron aciertos que presentaban producción aumentada de alcoholes grasos y se sometieron a verificación adicional usando fermentaciones en matraz de agitación, tal como se describió anteriormente. En la tabla 11 se muestran los

resultados de una fermentación en matraz de agitación que muestra un porcentaje mejorado de producción de alcohol graso usando una mutación de combinación de CarB adicional (carB4). En la figura 4 se presenta una representación gráfica de la eficacia de conversión relativa de variantes de CarB de bajo número de copias. Los resultados notificados en la tabla 11 son de ejecuciones en biorreactor llevadas a cabo en condiciones idénticas.

5 Tabla 11: Variantes de CAR

Nombre	Mutación/mutaciones	Cepa	Datos de tanque	Notas
carB	Ninguna = silvestre (E20 V191 F288 Q473)			la proteína es SEQ ID NO: 7
carB60	Ninguna + etiqueta	V324		
carB1	A535S	V940	83% de FALC; C12/C14=3,4	tiene una copia de TE cromosómico 12H08
carB2	E20R, F288G, Q473I, A535S	LH375	97% de FALC; C12/C14=3,6	tiene dos copias de TE cromosómico 12H08
carB2	E20R, F288G, Q473I, A535S	LH346	96% de FALC; C12/C14=3,7	tiene una copia de TE cromosómico 12H08
carB3	E20R, F288G, Q473H, A535S	Biblioteca combinada L	Sin ejemplos ejecutados en biorreactores hasta la fecha	
carB4	E20R, F288G, Q473H, A535S, R827A, S927G	Biblioteca combinada R (VA-219)	97% de FALC; C12/C14=3,9	tiene dos copias de TE cromosómico 12H08
carA	Ninguna	Véase la publicación de patente estadounidense n.º 20100105963		la proteína es SEQ ID NO: 39
FadD9	Ninguna	Véase la publicación de patente estadounidense n.º 20100105963		la proteína es SEQ ID NO: 40

Las secuencias de ADN de CarA, FadD9, CarB y CarB60 se presentan en el presente documento como SEQ ID NO: 41, 42, 43 y 44, respectivamente.

Identificación de mutaciones beneficiosas adicionales en la enzima CarB mediante mutagénesis de saturación:

10 Se desarrolló posteriormente un sistema de examen de plásmido doble y se validó para identificar variantes de CarB mejoradas con respecto a CarB4 para la producción de FALC. Este sistema de plásmido doble cumplía los siguientes criterios: 1) Los clones mutantes producen un alto título de FA para proporcionar flujo de ácidos grasos en exceso de la actividad de CarB. Esto se logra transformando una cepa base (V668 con dos copias de TE cromosómico) con un plásmido (pLYC4, pCL1920\_P<sub>TRC</sub>\_carDead\_tesA\_alrAadp1\_fabB[A329G]\_fadR) que porta el operón de FALC con una enzima CarB catalíticamente inactiva CarB[S693A] para potenciar la producción de ácidos grasos libres; 2) el plásmido de examen con molde de mutante de carB, preferiblemente menor de 9 kb, es propenso a procedimientos de mutagénesis de saturación y es compatible para la expresión con pLYC4; 3) el intervalo dinámico de actividad de CarB puede ajustarse. Esto se logra combinando un promotor más débil (P<sub>TRC1</sub>) y codones de iniciación alternativos (GTG o TTG) para ajustar los niveles de expresión de CarB4. 3) Se introdujo una buena estabilidad del plásmido, un módulo de toxina/antitoxina (operón de ccdBA) para mantener la estabilidad del plásmido.

En resumen, se construyó el plásmido de examen pBZ1 (pACYCDuet-1\_P<sub>TRC1</sub>-carB4GTG\_rrnBter\_ccdAB) a partir de cuatro partes usando el método de clonación In-Fusion HD (Clontech) mezclando razones molares iguales de cuatro partes (P<sub>TRC1</sub>, carB4 con codones de iniciación ATG/TTG/GTG, terminadores rrnB T1T2 con ccdAB y vector pACYCDuet-1). Se amplificaron por PCR las partes (1 a 4) mediante los siguientes pares de cebadores: (1) P<sub>TRC1</sub> - cebador directo 5'-CGGTTCTGGCAAATATTCTGAAATGAGCTGTTGACAATTAATCAAATCCGGCTCGTATAATGTGTG-3' (SEQ ID NO: 45) y cebador inverso 5'-GGTTTATTCCTCCTTATTAATCGATACAT-3' (SEQ ID NO: 46) usando el plásmido pVA232 (pCL1920\_P<sub>TRC</sub>\_carB4\_tesA\_alrAadp1\_fabB[A329G]\_fadR) como molde. (1) carB4 con codones de iniciación ATG/TTG/GTG - cebador directo carB4 ATG 5'-ATGTATCGATTAATAAGGAGGAATAAACCATGGGCACGAGCGATGTTTCACGACGCGAC-3' (SEQ ID NO: 47); carB4 GTG 5'-ATGTATCGATTAATAAGGAGGAATAAACCGTGGGCACGAGCGATGTTTCACGACGCGAC-3' (SEQ ID NO: 48); y carB4 TTG 5'-ATGTATCGATTAATAAGGAGGAATAAACCTTGGGCACGAGCGATGTTTCACGACGCGAC-3' (SEQ ID NO: 49); y cebador inverso carB4 rev 5'-TTCTAAATCAGACCGAACTCGCGCAG-3' (SEQ ID NO: 50), usando el plásmido pVA232 como molde. (3) Los terminadores rrnB T1T2 con ccdAB - cebador directo rrnB T1T2 term 5'-CTGCGCGAGTTCGGTCTGATTTAGAATTCCTCGAGGATGGTAGTGTGG-3' (SEQ ID NO: 51) y cebador inverso

ccdAB rev 5'-CAGTCGACATACGAAACGGGAATGCGG-3' (SEQ ID NO: 52), usando el plásmido pAH008 (operón de pV171\_ccdBA). (4) La estructura principal de vector pACYCDuet-1 - cebador directo vector pACYC para 5'-CCGCATTCCCGTTTCGTATGTCGACTGAAACCTCAGGCATTGAGAAGCACACGGTC-3' (SEQ ID NO: 53) y cebador inverso vector pACYC rev 5'-CTCATTTTCAGAATATTTGCCAGAACCCTTAATTTCTAATGCAGGAGTCGCATAAG-3' (SEQ ID NO: 54).

Se coexpresó el plásmido pBZ1 con pLYC4 en la cepa descrita anteriormente y se validó mediante fermentación en matraz de agitación y placa de pocillos profundos. Se optimizaron las condiciones de fermentación de manera que el molde de CarB4\_GTG tenía de manera reproducible ~65% de conversión de FALC en ambas plataformas de fermentación tal como se describe en el ejemplo 5. En la figura 15 se muestran los resultados para la fermentación en matraz de agitación.

Se sometieron sitios adicionales (18, 19, 22, 28, 80, 87, 90, 143, 212, 231, 259, 292, 396, 416, 418, 530, 541, 574, 612, 636, 677, 712, 750, 799, 809, 810, 870, 936, 985, 986, 1026, 1062, 1080, 1134, 1149, 1158, 1161, 1170) que contenían mutaciones en las variantes de CarB mejoradas (tabla 7) a mutagénesis de saturación completa. Se usaron cebadores que contenían los nucleótidos degenerados NNK o NNS para mutar estas posiciones para dar otros aminoácidos mediante un método basado en PCR (Sawano y Miyawaki 2000, Nucl. Acids Res. 28: e78). Se construyó la biblioteca de saturación usando el molde de plásmido pBZ1 (pACYCDuet-1\_P<sub>TRC1</sub>-carB4GTG\_rrnBter\_ccdAB). Se transformaron clones mutantes en cepas de clonación NEB Turbo (New England Biolab) y se aislaron y se agruparon plásmidos. Entonces se transformaron los plásmidos agrupados en una cepa basada en V668 que portaba el plásmido pLYC4 y se seleccionaron los transformantes en placas de agar LB complementado con antibióticos (espectinomicina 100 mg/l y cloranfenicol 34 mg/l).

Entonces se examinaron las variantes de CarB de la biblioteca de saturación para determinar la producción de alcoholes grasos. Se recogieron colonias individuales directamente en placas de 96 pocillos según un protocolo de fermentación en placa de pocillos profundos modificado tal como se describe en el ejemplo 5. Se seleccionaron aciertos eligiendo clones que producían un título de ácidos grasos libres total más pequeño y un título de alcoholes grasos total más grande en comparación con la cepa de control. Para comparar los aciertos de diferentes lotes de fermentación, se normalizó la conversión de ácidos grasos libres en alcoholes grasos calculando un porcentaje de ácidos grasos libres normalizado. También se usó NORM FFA (%) en la validación de aciertos tal como se describe en el ejemplo 5. NORM FFA (%) = porcentaje de FFA mutante / porcentaje de FFA de control; en donde "porcentaje de FFA" es el título de la especie de ácidos grasos totales dividido entre el título de la especie grasa total. Se sometieron los aciertos a verificación adicional usando fermentaciones en matraz de agitación tal como se describe en el ejemplo 5. La columna de ácido graso libre normalizado (NORM FFA) indica la mejora en la enzima, indicando los valores inferiores la mejor mejora. "ID de acierto" indica la posición de pocillo de placa de selección primaria en donde se encontró el fenotipo de NORM FFA inferior. Se identificaron mutaciones de aciertos secuenciando productos de PCR amplificados a partir de "aciertos" que contenían plásmidos pBZ1 usando cebadores específicos de gen de carB mutante (BZ1 para 5'-GGATCTCGACGCTCTCCCTT-3' (SEQ ID NO: 55) y cebador único BZ12\_ccdAB 5'-TCAAAAACGCCATTAACCTGATGTTCTG-3' (SEQ ID NO: 56). En la tabla 12 se resumen los valores de NORM FFA y las mutaciones identificadas en aciertos validados.

Tabla 12: Mutaciones beneficiosas en la enzima CarB4 identificadas durante la mutagénesis de saturación de aminoácidos

Aminoácido silvestre	Codón silvestre	ID de acierto (aminoácido)	Codón mutante	NORM FFA (%)
D18	GAT	P10H5(R)	AGG	75,5
		P6B4(L)	CTG	83,6
		P4H11(T)	ACG	80,8
		P8D11(P)	CCG	81,8
S22	AGC	P1F3(R)	AGG	57,7
		P2G9(R)	AGG	55,7
		P2A7(N)	AAC	90
L80	CTG	P8D7(G)	GGG	82,1
		P8H11(R)	AGG	87,4
R87	CGT	P7D7(G)	GGG	85,2
		P5D12(E)	GAG	89,4
D750	GAT	P8F11(A)	GCG	87,6
I870	ATT	P3A12(L)	CTG	76,6

Identificación de variantes novedosas de la enzima CarB mediante mutagénesis combinatoria completa:

Se construyó una biblioteca combinatoria completa para incluir los siguientes residuos de aminoácido: 18D, 18R, 22S, 22R, 473H, 473I, 827R, 827C, 870I, 870L, 926V, 926A, 926E, 927S, 927K, 927O, 930M, 930K, 930R, 1128L y 1128W. Se diseñaron cebadores que contenían codones nativos y mutantes en todas las posiciones para la construcción de la biblioteca mediante un método basado en PCR (Horton, R.M., Hunt, H.D., Ho, S.N., Pullen, J.K. y Pease, L.R. 1989). Las mutaciones beneficiosas conservadas en CarB2, CarB3 y CarB4 (20R, 288G y 535S) no se

5 cambiaron, por tanto, se usó carB2GTG clonado en pBZ1 (pBZ1\_P<sub>TRC1</sub>\_carB2GTG\_ccdAB modificado) como molde de PCR. Se completó la construcción de la biblioteca ensamblando fragmentos de PCR en los ORF de CarB que contenían las mutaciones combinatorias anteriores. Entonces se clonaron los ORF de CarB mutantes en la estructura principal de pBZ1 mediante el método In-Fusion (Clontech). Se precipitó el producto de In-Fusion y se introdujo por electroporación directamente en la cepa de examen que portaba el plásmido pLYC4. Se llevaron a cabo el examen de la biblioteca, la fermentación en matraz de agitación y placa de pocillos profundos tal como se describe en el ejemplo 5. En la tabla 13 se resumen las actividades (NORM FFA normalizado por CarB2, 100%) de mutantes de CarB con mutaciones combinatorias específicas. Se incluyen CarB2, CarB4 y CarB5 (CarB4-S22R) como controles. La columna de NORM FFA indica la mejora en la enzima CarB, indicando los valores inferiores la mejor mejora. También se muestra la mejora en veces (X-FIOC) del control (CarB2). Todas las mutaciones enumeradas son en relación con la secuencia de polipéptido de CarB silvestre (SEQ ID NO: 7). Por ejemplo, CarB1 tiene la mutación A535S, y la CarBDead (una enzima CarB catalíticamente inactiva) porta la mutación S693A que destruye el sitio de unión de fosfopanteteína.

Variantes de CarB novedosas para la producción mejorada de alcohol graso en biorreactores:

15 El fin de identificar variantes de CarB novedosas enumeradas en la tabla 13 es usarlas para la producción mejorada de alcohol graso. La variante de CarB superior (P06B6 - S3R, E20R, S22R, F288G, Q473H, A535S, R873S, S927G, M930R, L1128W) de la tabla 13 porta una mutación espontánea (AGC silvestre a AGA) en la posición 3. Ambas variantes de CarB P06B6, concretamente CarB7 (aminoácido R por AGA en la posición 3 - S3R, E20R, S22R, F288G, Q473H, A535S, R873S, S927G, M930R, L1128W), y CarB8 (aminoácido silvestre S por AGC en la posición 3 - E20R, S22R, F288G, Q473H, A535S, R873S, S927G, M930R, L1128W) se prepararon y clonaron en la estructura principal del plásmido de producción de alcohol graso de bajo número de copias pCL1920 para generar los siguientes operones de alcohol graso que difieren sólo en CarB. Se revirtió el codón de iniciación de la traducción (GTG) para todas las variantes de CarB (CarB2, CarB7 y carB8) a ATG para maximizar la expresión.

pCL1920\_P<sub>TRC</sub>\_carB2\_tesA\_alrAadp1\_fabB[A329G]\_fadR

25 pCL1920\_P<sub>TRC</sub>\_carB7\_tesA\_alrAadp1\_fabB[A329G]\_fadR

pCL1920\_P<sub>TRC</sub>\_carB8\_tesA\_alrAadp1\_fabB[A329G]\_fadR

30 Se transformaron los plásmidos descritos anteriormente en una cepa basada en V668 con una copia de TE cromosómico, y se examinaron las cepas resultantes en biorreactores tal como se describe en el ejemplo 4. En la figura 16 se muestra la mejora (medida en % de alcoholes grasos en el producto de fermentación del biorreactor) de CarB7 y CarB8 con respecto a CarB2. El orden de actividad es CarB7 > CarB8 > CarB2. La mutación de la posición 3 de CarB7 (AGC a un codón poco común AGA R) confería actividad mayor que CarB8, además, el análisis de SDS-PAGE de proteínas solubles totales reveló expresión mayor de CarB7 que CarB8 y CarB2. Los niveles de expresión de CarB2 y CarB8 eran similares. Esto concuerda con los datos de CarB60 descritos en el ejemplo 6, tanto la mutación de la posición 3 de codón poco común AGA R como la etiqueta de CarB60 en su extremo N-terminal pueden mejorar la expresión de CarB. Se entiende que CarB7 y CarB8 tendrán un mejor rendimiento que CarB2 en cepas con flujo de ácidos grasos libres aumentado o bien modificando por ingeniería genética las cepas huésped y/o bien modificando por ingeniería genética los otros componentes del operón de producción de alcohol graso.

Tabla 13: Sumario de variantes de CarB identificadas a partir de la biblioteca combinatoria en sistema de plásmido doble.

Mutantes	NORM FFA (%)	X-FIOC	Mutaciones
P06B6	16,5	6,06	S3R, E20R, S22R, F288G, Q473H, A535S, R873S, S927G, M930R, L1128W
P13A3	23,9	4,18	D18R, E20R, S22R, F288G, Q473I, A535S, S927G, M930K, L1128W
P02A2	26,5	3,77	E20R, S22R, F288G, Q473I, A535S, R827C, V926E, S927K, M930R
P05H3	26,7	3,75	D18R, E20R, 288G, Q473I, A535S, R827C, V926E, M930K, L1128W
P10F10	31,9	3,13	E20R, S22R, F288G, Q473H, A535S, R827C, V926A, S927K, M930R
P01C12	34,2	2,92	E20R, S22R, F288G, Q473H, A535S, R827C
P03B1	36,9	2,71	E20R, S22R, F288G, Q473I, A535S, R827C, M930R
P06E4	36,9	2,71	E20R, S22R, F288G, Q473I, A535S, I870L, S927G, M930R
P14C6	37,4	2,67	E20R, S22R, F288G, Q473I, A535S, I870L, S927G
P05F10	40,4	2,48	D18R, E20R, S22R, F288G, Q473I, A535S, R827C, I870L, V926A, S927G
P06C8	40,8	2,45	E20R, S22R, F288G, Q473H, A535S, R827C, I870L, L1128W
P15E4	40,8	2,45	D18R, E20R, S22R, F288G, Q473H, A535S, R827C, I870L,

ES 2 613 553 T3

			S927G, L1128W
P05H7	40,9	2,44	E20R, S22R, F288G, Q473I, A535S, R827C, I870L, S927G, L1128W
P15A6	41	2,44	E20R, S22R, F288G, Q473I, A535S, R827C, I870L, S927G, M930K, L1128W
P08F5	41,2	2,43	E20R, S22R, F288G, Q473H, A535S, I870L, S927G, M930K
P14C7	41,3	2,42	E20R, F288G, Q473I, A535S, I870L, M930K
P16H10	42,1	2,38	E20R, S22R, F288G, Q473H, A535S, S927G, M930K, L1128W
P16A1	44,1	2,27	D18R, E20R, S22R, F288G, Q473I, A535S, S927G, L1128W
P14H4	44,2	2,26	E20R, S22R, F288G, Q473I, A535S, R827C, I870L, S927G
P15C1	46,5	2,15	D18R, E20R, S22R, F288G, Q473I, A535S, R827C, I870L, S927G, L1128W
P16E5	47,2	2,12	D18R, E20R, S22R, F288G, Q473I, A535S, S927G, M930R, L1128W
P15A3	47,2	2,12	E20R, S22R, F288G, Q473H, A535S, V926E, S927G, M930R
P05A2	52,4	1,91	E20R, S22R, F288G, Q473H, A535S, R827C, I870L, V926A, L1128W
CarB2	100	1	E20R, F288G, Q473I, A535S
CarB4	77,8	1,29	E20R, F288G, Q473H, A535S, R827A, S927G
CarB5	48,9	2,04	E20R, S22R, F288G, Q473H, A535S, R827A, S927G
CarB1	ND		A535S
CarB silvestre	ND		SEQ ID NO: 7
CarBDead	ND		S693A

Todos los métodos descritos en el presente documento pueden realizarse en cualquier orden adecuado a menos que se indique lo contrario en el presente documento o se contradiga claramente por el contexto de otra forma. El uso de todos y cada uno de los ejemplos, o una expresión a modo de ejemplo (por ejemplo, "tal como") proporcionados en el presente documento, pretende meramente iluminar mejor la divulgación y no supone una limitación en el alcance de la divulgación a menos que se reivindique lo contrario. Ninguna expresión en la memoria descriptiva debe interpretarse que indica que cualquier elemento no reivindicado sea esencial para la práctica de la divulgación. Debe entenderse que la terminología usada en el presente documento es para el fin de describir realizaciones particulares sólo, y no pretende ser limitativa. Se describen en el presente documento realizaciones preferidas de esta divulgación. Variaciones de esas realizaciones preferidas pueden resultar evidentes para los expertos habituales en la técnica tras la lectura de la descripción anterior.

5

10

Lista de secuencias

<110> LS9, INC.  
 <120> PRODUCCIÓN MEJORADA DE ALCOHOLES GRASOS  
 <130> LS00039PCT  
 <140>  
 <141>  
 <150> 61/619,309  
 <151> 02-04-2012  
 <160> 56  
 <170> PatentIn versión 3.5  
 <210> 1  
 <211> 1232  
 <212> ADN  
 <213> Secuencia artificial  
 <220>  
 <221> fuente  
 <223> /nota="Descripción de secuencia artificial: Polinucleótido sintético"  
 <400> 1  
 ttgtccatct ttatataatt tgggggtagg gtggttcttta tgtaaaaaaa acgtttttagg 60  
 atgcatatgg cggccgcata acttcgtata gcatacatta tacgaagtta tctagagttg 120  
 catgcctgca ggtccgctta ttatcactta ttcaggcgta gcaaccaggc gtttaagggc 180  
 accaataact gccttaaaaa aattacgccc cgccctgccca ctcatcgag tactgttgta 240  
 attcattaag cattctgccg acatggaagc catcaciaaac ggcatgatga acctgaatcg 300  
 ccagcggcat cagcaccttg tcgccttgcg tataatattt gcccatgggtg aaaacggggg 360  
 cgaagaagtt gtccatattg gccacgttta aatcaaaaact ggtgaaactc acccagggat 420  
 tggctgagac gaaaaacata ttctcaataa acccttttagg gaaataggcc aggttttcac 480  
 cgtaacacgc cacatcttgc gaatatatgt gtagaaactg ccggaaatcg tcgtggtatt 540  
 cactccagag cgatgaaaac gtttcagttt gctcatggaa aacgggtgtaa caaggggtgaa 600  
 cactatccca taccaccagc tcaccgtctt tcattgccat acggaattcc ggatgagcat 660  
 tcatcaggcg ggcaagaatg tgaataaagg ccggataaaa cttgtgctta tttttcttta 720  
 cggctcttaa aaaggccgta atatccagct gaacggctctg gttataggta cattgagcaa 780  
 ctgactgaaa tgcctcaaaa tggtctttac gatgccattg ggatatatca acgggtggtat 840  
 atccagtgat ttttttctcc attttagctt ccttagctcc tgaaaatctc gataactcaa 900  
 aaaatacgcc cggtagtgat cttatttcat tatggtgaaa gttggaacct cttacgtgcc 960  
 gatcaacgct tcattttcgc caaaagttgg cccagggctt cccggtatca acagggacac 1020

ES 2 613 553 T3

caggatttat ttattctgcg aagtgatctt ccgtcacagg tatttattcg actctagata 1080  
 acttcgtata gcatacatta tacgaagtta tggatccagc ttatcgatac cgtcaaacia 1140  
 atcataaaaa atttatttgc tttcaggaaa atttttctgt ataatagatt caattgcat 1200  
 gacgacgaac acgcacctgc aggaggagac ca 1232

<210> 2  
 <211> 232  
 <212> ADN  
 <213> Secuencia artificial

<220>  
 <221> fuente  
 <223> /nota="Descripción de secuencia artificial: Polinucleótido sintético"

<400> 2  
 ttgtccatct ttatataatt tgggggtagg gtgttcttta tgtaaaaaaa acgttttagg 60  
 atgcatatgg cggccgcata acttcgtata gcatacatta tacgaagtta tggatccagc 120  
 ttatcgatac cgtcaaacia atcataaaaa atttatttgc tttcaggaaa atttttctgt 180  
 ataatagatt caattgcat gacgacgaac acgcacctgc aggaggagac ca 232

<210> 3  
 <211> 340  
 <212> PRT  
 <213> Acinetobacter sp.

<400> 3  
 Met Ser Asn His Gln Ile Arg Ala Tyr Ala Ala Met Gln Ala Gly Glu  
 1 5 10 15  
 Gln Val Val Pro Tyr Gln Phe Asp Ala Gly Glu Leu Lys Ala His Gln  
 20 25 30  
 Val Glu Val Lys Val Glu Tyr Cys Gly Leu Cys His Ser Asp Leu Ser  
 35 40 45  
 Val Ile Asn Asn Glu Trp Gln Ser Ser Val Tyr Pro Ala Val Ala Gly  
 50 55 60  
 His Glu Ile Ile Gly Thr Ile Ile Ala Leu Gly Ser Glu Ala Lys Gly  
 65 70 75 80  
 Leu Lys Leu Gly Gln Arg Val Gly Ile Gly Trp Thr Ala Glu Thr Cys  
 85 90 95  
 Gln Ala Cys Asp Pro Cys Ile Gly Gly Asn Gln Val Leu Cys Thr Gly  
 100 105 110

ES 2 613 553 T3

Glu Lys Lys Ala Thr Ile Ile Gly His Ala Gly Gly Phe Ala Asp Lys  
 115 120 125  
 Val Arg Ala Gly Trp Gln Trp Val Ile Pro Leu Pro Asp Asp Leu Asp  
 130 135 140  
 Pro Glu Ser Ala Gly Pro Leu Leu Cys Gly Gly Ile Thr Val Leu Asp  
 145 150 155 160  
 Pro Leu Leu Lys His Lys Ile Gln Ala Thr His His Val Gly Val Ile  
 165 170 175  
 Gly Ile Gly Gly Leu Gly His Ile Ala Ile Lys Leu Leu Lys Ala Trp  
 180 185 190  
 Gly Cys Glu Ile Thr Ala Phe Ser Ser Asn Pro Asp Lys Thr Glu Glu  
 195 200 205  
 Leu Lys Ala Asn Gly Ala Asp Gln Val Val Asn Ser Arg Asp Ala Gln  
 210 215 220  
 Ala Ile Lys Gly Thr Arg Trp Lys Leu Ile Ile Leu Ser Thr Ala Asn  
 225 230 235 240  
 Gly Thr Leu Asn Val Lys Ala Tyr Leu Asn Thr Leu Ala Pro Lys Gly  
 245 250 255  
 Ser Leu His Phe Leu Gly Val Thr Leu Glu Pro Ile Pro Val Ser Val  
 260 265 270  
 Gly Ala Ile Met Gly Gly Ala Lys Ser Val Thr Ser Ser Pro Thr Gly  
 275 280 285  
 Ser Pro Leu Ala Leu Arg Gln Leu Leu Gln Phe Ala Ala Arg Lys Asn  
 290 295 300  
 Ile Ala Pro Gln Val Glu Leu Phe Pro Met Ser Gln Leu Asn Glu Ala  
 305 310 315 320  
 Ile Glu Arg Leu His Ser Gly Gln Ala Arg Tyr Arg Ile Val Leu Lys  
 325 330 335  
 Ala Asp Phe Asp  
 340

<210> 4  
 <211> 314  
 <212> PRT

ES 2 613 553 T3

<213> Acinetobacter sp.

<400> 4

Met Ala Thr Thr Asn Val Ile His Ala Tyr Ala Ala Met Gln Ala Gly  
 1 5 10 15

Glu Ala Leu Val Pro Tyr Ser Phe Asp Ala Gly Glu Leu Gln Pro His  
 20 25 30

Gln Val Glu Val Lys Val Glu Tyr Cys Gly Leu Cys His Ser Asp Val  
 35 40 45

Ser Val Leu Asn Asn Glu Trp His Ser Ser Val Tyr Pro Val Val Ala  
 50 55 60

Gly His Glu Val Ile Gly Thr Ile Thr Gln Leu Gly Ser Glu Ala Lys  
 65 70 75 80

Gly Leu Lys Ile Gly Gln Arg Val Gly Ile Gly Trp Thr Ala Glu Ser  
 85 90 95

Cys Gln Ala Cys Asp Gln Cys Ile Ser Gly Gln Gln Val Leu Cys Thr  
 100 105 110

Gly Glu Asn Thr Ala Thr Ile Ile Gly His Ala Gly Gly Phe Ala Asp  
 115 120 125

Lys Val Arg Ala Gly Trp Gln Trp Val Ile Pro Leu Pro Asp Glu Leu  
 130 135 140

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

Asp Pro Ile Leu Lys His Gln Ile Gln Ala Ile His His Val Ala Val  
 165 170 175

Ile Gly Ile Gly Gly Leu Gly His Met Ala Ile Lys Leu Leu Lys Ala  
 180 185 190

Trp Gly Cys Glu Ile Thr Ala Phe Ser Ser Asn Pro Asn Lys Thr Asp  
 195 200 205

Glu Leu Lys Ala Met Gly Ala Asp His Val Val Asn Ser Arg Asp Asp  
 210 215 220

Ala Glu Ile Lys Ser Gln Gln Gly Lys Phe Asp Leu Leu Leu Ser Thr  
 225 230 235 240

ES 2 613 553 T3

Val Asn Val Pro Leu Asn Trp Asn Ala Tyr Leu Asn Thr Leu Ala Pro  
 245 250 255

Asn Gly Thr Phe His Phe Leu Gly Val Val Met Glu Pro Ile Pro Val  
 260 265 270

Pro Val Gly Ala Leu Leu Gly Gly Ala Lys Ser Leu Thr Ala Ser Pro  
 275 280 285

Thr Gly Ser Pro Ala Ala Leu Arg Lys Leu Leu Glu Phe Ala Ala Arg  
 290 295 300

Lys Asn Ile Ala Pro Gln Ile Glu Met Tyr  
 305 310

<210> 5  
 <211> 1020  
 <212> ADN  
 <213> Escherichia coli

<400> 5  
 atgtcgatga taaaaagcta tgccgcaaaa gaagcgggcg gcgaactgga agtttatgag 60  
 tacgatcccc gtgagctgag gccacaagat gttgaagtgc aggtggatta ctgcgggatc 120  
 tgccattccg atctgtcgat gatcgataac gaatggggat tttcacaata tccgctggtt 180  
 gccgggcatg aggtgattgg gcgcgtggtg gcaactcggga gcgccgcgca ggataaagg 240  
 ttgcaggctg gtcagcgtgt cgggattggc tggacggcgc gtagctgtgg tcaactgcgac 300  
 gcctgtatta gcggtaatca gatcaactgc gagcaagggt cggtgccgac gattatgaat 360  
 cgcggtggct ttgccgagaa gttgcgtgcg gactggcaat gggtgattcc actgccagaa 420  
 aatattgata tcgagtcgcg cgggccgctg ttgtgcggcg gtatcacggt ctttaaacca 480  
 ctgttgatgc accatatcac tgctaccagc gcgcttgggg taattggtat tggcgggctg 540  
 gggcatatcg ctataaaact tctgcacgca atgggatgcg aggtgacagc ctttagttct 600  
 aatccggcga aagagcagga agtgctggcg atgggtgccg ataaagtggg gaatagccgc 660  
 gatccgcagg cactgaaagc actggcgggg cagtttgatc tcattatcaa caccgtcaac 720  
 gtcagcctcg actggcagcc ctattttgag gcgctgacct atggcggtaa tttccatacg 780  
 gtcggtgctg ttctcacgcc gctgtctggt cggccttta cgtaattgc gggcgatcgc 840  
 agcgtctctg gttctgctac cggcaocgct tatgagctgc gtaagctgat gcgttttgcc 900  
 gcccgcagca aggttgccgc gaccaccgaa ctgttcccga tgcgaaaat taacgacgcc 960  
 atccagcatg tgccgcgacgg taaggcgcgt taccgcgtgg tgttgaaagc cgattttga 1020

<210> 6  
 <211> 1174

ES 2 613 553 T3

```

<212> PRT
<213> Secuencia artificial

<220>
<221> fuente
<223> /nota="Descripción de secuencia artificial: Polinucleótido
        sintético"

<400> 6
Met Ala Val Asp Ser Pro Asp Glu Arg Leu Gln Arg Arg Ile Ala Gln
1
                    5
                                10
                                        15

Leu Phe Ala Glu Asp Glu Gln Val Lys Ala Ala Arg Pro Leu Glu Ala
20
                    25
                                30

Val Ser Ala Ala Val Ser Ala Pro Gly Met Arg Leu Ala Gln Ile Ala
35
                    40
                                45

Ala Thr Val Met Ala Gly Tyr Ala Asp Arg Pro Ala Ala Gly Gln Arg
50
                    55
                                60

Ala Phe Glu Leu Asn Thr Asp Asp Ala Thr Gly Arg Thr Ser Leu Arg
65
                    70
                                75
                                        80

Leu Leu Pro Arg Phe Glu Thr Ile Thr Tyr Arg Glu Leu Trp Gln Arg
85
                    90
                                95

Val Gly Glu Val Ala Ala Ala Trp His His Asp Pro Glu Asn Pro Leu
100
                    105
                                110

Arg Ala Gly Asp Phe Val Ala Leu Leu Gly Phe Thr Ser Ile Asp Tyr
115
                    120
                                125

Ala Thr Leu Asp Leu Ala Asp Ile His Leu Gly Ala Val Thr Val Pro
130
                    135
                                140

Leu Gln Ala Ser Ala Ala Val Ser Gln Leu Ile Ala Ile Leu Thr Glu
145
                    150
                                155
                                        160

Thr Ser Pro Arg Leu Leu Ala Ser Thr Pro Glu His Leu Asp Ala Ala
165
                    170
                                175

Val Glu Cys Leu Leu Ala Gly Thr Thr Pro Glu Arg Leu Val Val Phe
180
                    185
                                190

Asp Tyr His Pro Glu Asp Asp Asp Gln Arg Ala Ala Phe Glu Ser Ala
195
                    200
                                205

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

```

ES 2 613 553 T3

Asp Ala Val Arg Ala Arg Gly Arg Asp Leu Pro Ala Ala Pro Leu Phe  
 225 230 235 240

Val Pro Asp Thr Asp Asp Asp Pro Leu Ala Leu Leu Ile Tyr Thr Ser  
 245 250 255

Gly Ser Thr Gly Thr Pro Lys Gly Ala Met Tyr Thr Asn Arg Leu Ala  
 260 265 270

Ala Thr Met Trp Gln Gly Asn Ser Met Leu Gln Gly Asn Ser Gln Arg  
 275 280 285

Val Gly Ile Asn Leu Asn Tyr Met Pro Met Ser His Ile Ala Gly Arg  
 290 295 300

Ile Ser Leu Phe Gly Val Leu Ala Arg Gly Gly Thr Ala Tyr Phe Ala  
 305 310 315 320

Ala Lys Ser Asp Met Ser Thr Leu Phe Glu Asp Ile Gly Leu Val Arg  
 325 330 335

Pro Thr Glu Ile Phe Phe Val Pro Arg Val Cys Asp Met Val Phe Gln  
 340 345 350

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

Asp Thr Leu Asp Arg Glu Val Lys Ala Asp Leu Arg Gln Asn Tyr Leu  
 370 375 380

Gly Gly Arg Phe Leu Val Ala Val Val Gly Ser Ala Pro Leu Ala Ala  
 385 390 395 400

Glu Met Lys Thr Phe Met Glu Ser Val Leu Asp Leu Pro Leu His Asp  
 405 410 415

Gly Tyr Gly Ser Thr Glu Ala Gly Ala Ser Val Leu Leu Asp Asn Gln  
 420 425 430

Ile Gln Arg Pro Pro Val Leu Asp Tyr Lys Leu Val Asp Val Pro Glu  
 435 440 445

Leu Gly Tyr Phe Arg Thr Asp Arg Pro His Pro Arg Gly Glu Leu Leu  
 450 455 460

Leu Lys Ala Glu Thr Thr Ile Pro Gly Tyr Tyr Lys Arg Pro Glu Val



ES 2 613 553 T3

Asp Leu Ala Asn Tyr Ile Glu Ala Glu Arg Asn Ser Gly Ala Lys Arg  
 725 730 735  
 Pro Thr Phe Thr Ser Val His Gly Gly Gly Ser Glu Ile Arg Ala Ala  
 740 745 750  
 Asp Leu Thr Leu Asp Lys Phe Ile Asp Ala Arg Thr Leu Ala Ala Ala  
 755 760 765  
 Asp Ser Ile Pro His Ala Pro Val Pro Ala Gln Thr Val Leu Leu Thr  
 770 775 780  
 Gly Ala Asn Gly Tyr Leu Gly Arg Phe Leu Cys Leu Glu Trp Leu Glu  
 785 790 795 800  
 Arg Leu Asp Lys Thr Gly Gly Thr Leu Ile Cys Val Val Arg Gly Ser  
 805 810 815  
 Asp Ala Ala Ala Ala Arg Lys Arg Leu Asp Ser Ala Phe Asp Ser Gly  
 820 825 830  
 Asp Pro Gly Leu Leu Glu His Tyr Gln Gln Leu Ala Ala Arg Thr Leu  
 835 840 845  
 Glu Val Leu Ala Gly Asp Ile Gly Asp Pro Asn Leu Gly Leu Asp Asp  
 850 855 860  
 Ala Thr Trp Gln Arg Leu Ala Glu Thr Val Asp Leu Ile Val His Pro  
 865 870 875 880  
 Ala Ala Leu Val Asn His Val Leu Pro Tyr Thr Gln Leu Phe Gly Pro  
 885 890 895  
 Asn Val Val Gly Thr Ala Glu Ile Val Arg Leu Ala Ile Thr Ala Arg  
 900 905 910  
 Arg Lys Pro Val Thr Tyr Leu Ser Thr Val Gly Val Ala Asp Gln Val  
 915 920 925  
 Asp Pro Ala Glu Tyr Gln Glu Asp Ser Asp Val Arg Glu Met Ser Ala  
 930 935 940  
 Val Arg Val Val Arg Glu Ser Tyr Ala Asn Gly Tyr Gly Asn Ser Lys  
 945 950 955 960  
 Trp Ala Gly Glu Val Leu Leu Arg Glu Ala His Asp Leu Cys Gly Leu  
 965 970 975

ES 2 613 553 T3

Pro Val Ala Val Phe Arg Ser Asp Met Ile Leu Ala His Ser Arg Tyr  
 980 985 990

Ala Gly Gln Leu Asn Val Gln Asp Val Phe Thr Arg Leu Ile Leu Ser  
 995 1000 1005

Leu Val Ala Thr Gly Ile Ala Pro Tyr Ser Phe Tyr Arg Thr Asp  
 1010 1015 1020

Ala Asp Gly Asn Arg Gln Arg Ala His Tyr Asp Gly Leu Pro Ala  
 1025 1030 1035

Asp Phe Thr Ala Ala Ala Ile Thr Ala Leu Gly Ile Gln Ala Thr  
 1040 1045 1050

Glu Gly Phe Arg Thr Tyr Asp Val Leu Asn Pro Tyr Asp Asp Gly  
 1055 1060 1065

Ile Ser Leu Asp Glu Phe Val Asp Trp Leu Val Glu Ser Gly His  
 1070 1075 1080

Pro Ile Gln Arg Ile Thr Asp Tyr Ser Asp Trp Phe His Arg Phe  
 1085 1090 1095

Glu Thr Ala Ile Arg Ala Leu Pro Glu Lys Gln Arg Gln Ala Ser  
 1100 1105 1110

Val Leu Pro Leu Leu Asp Ala Tyr Arg Asn Pro Cys Pro Ala Val  
 1115 1120 1125

Arg Gly Ala Ile Leu Pro Ala Lys Glu Phe Gln Ala Ala Val Gln  
 1130 1135 1140

Thr Ala Lys Ile Gly Pro Glu Gln Asp Ile Pro His Leu Ser Ala  
 1145 1150 1155

Pro Leu Ile Asp Lys Tyr Val Ser Asp Leu Glu Leu Leu Gln Leu  
 1160 1165 1170

Leu

<210> 7  
 <211> 1173  
 <212> PRT  
 <213> Mycobacterium smegmatis

<400> 7

ES 2 613 553 T3

Met Thr Ser Asp Val His Asp Ala Thr Asp Gly Val Thr Glu Thr Ala  
1 5 10 15

Leu Asp Asp Glu Gln Ser Thr Arg Arg Ile Ala Glu Leu Tyr Ala Thr  
20 25 30

Asp Pro Glu Phe Ala Ala Ala Ala Pro Leu Pro Ala Val Val Asp Ala  
35 40 45

Ala His Lys Pro Gly Leu Arg Leu Ala Glu Ile Leu Gln Thr Leu Phe  
50 55 60

Thr Gly Tyr Gly Asp Arg Pro Ala Leu Gly Tyr Arg Ala Arg Glu Leu  
65 70 75 80

Ala Thr Asp Glu Gly Gly Arg Thr Val Thr Arg Leu Leu Pro Arg Phe  
85 90 95

Asp Thr Leu Thr Tyr Ala Gln Val Trp Ser Arg Val Gln Ala Val Ala  
100 105 110

Ala Ala Leu Arg His Asn Phe Ala Gln Pro Ile Tyr Pro Gly Asp Ala  
115 120 125

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

Val Cys Ala Tyr Leu Gly Leu Val Ser Val Pro Leu Gln His Asn Ala  
145 150 155 160

Pro Val Ser Arg Leu Ala Pro Ile Leu Ala Glu Val Glu Pro Arg Ile  
165 170 175

Leu Thr Val Ser Ala Glu Tyr Leu Asp Leu Ala Val Glu Ser Val Arg  
180 185 190

Asp Val Asn Ser Val Ser Gln Leu Val Val Phe Asp His His Pro Glu  
195 200 205

Val Asp Asp His Arg Asp Ala Leu Ala Arg Ala Arg Glu Gln Leu Ala  
210 215 220

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

Ala Gly Leu Pro Ala Glu Pro Ile Tyr Thr Ala Asp His Asp Gln Arg  
245 250 255

ES 2 613 553 T3

Leu Ala Met Ile Leu Tyr Thr Ser Gly Ser Thr Gly Ala Pro Lys Gly  
 260 265 270

Ala Met Tyr Thr Glu Ala Met Val Ala Arg Leu Trp Thr Met Ser Phe  
 275 280 285

Ile Thr Gly Asp Pro Thr Pro Val Ile Asn Val Asn Phe Met Pro Leu  
 290 295 300

Asn His Leu Gly Gly Arg Ile Pro Ile Ser Thr Ala Val Gln Asn Gly  
 305 310 315 320

Gly Thr Ser Tyr Phe Val Pro Glu Ser Asp Met Ser Thr Leu Phe Glu  
 325 330 335

Asp Leu Ala Leu Val Arg Pro Thr Glu Leu Gly Leu Val Pro Arg Val  
 340 345 350

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

Thr Gln Gly Ala Asp Glu Leu Thr Ala Glu Lys Gln Ala Gly Ala Glu  
 370 375 380

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

Thr Ala Pro Leu Ala Ala Glu Met Arg Ala Phe Leu Asp Ile Thr Leu  
 405 410 415

Gly Ala His Ile Val Asp Gly Tyr Gly Leu Thr Glu Thr Gly Ala Val  
 420 425 430

Thr Arg Asp Gly Val Ile Val Arg Pro Pro Val Ile Asp Tyr Lys Leu  
 435 440 445

Ile Asp Val Pro Glu Leu Gly Tyr Phe Ser Thr Asp Lys Pro Tyr Pro  
 450 455 460

Arg Gly Glu Leu Leu Val Arg Ser Gln Thr Leu Thr Pro Gly Tyr Tyr  
 465 470 475 480

Lys Arg Pro Glu Val Thr Ala Ser Val Phe Asp Arg Asp Gly Tyr Tyr  
 485 490 495

His Thr Gly Asp Val Met Ala Glu Thr Ala Pro Asp His Leu Val Tyr  
 500 505 510

ES 2 613 553 T3

Val Asp Arg Arg Asn Asn Val Leu Lys Leu Ala Gln Gly Glu Phe Val  
515 520 525

Ala Val Ala Asn Leu Glu Ala Val Phe Ser Gly Ala Ala Leu Val Arg  
530 535 540

Gln Ile Phe Val Tyr Gly Asn Ser Glu Arg Ser Phe Leu Leu Ala Val  
545 550 555 560

Val Val Pro Thr Pro Glu Ala Leu Glu Gln Tyr Asp Pro Ala Ala Leu  
565 570 575

Lys Ala Ala Leu Ala Asp Ser Leu Gln Arg Thr Ala Arg Asp Ala Glu  
580 585 590

Leu Gln Ser Tyr Glu Val Pro Ala Asp Phe Ile Val Glu Thr Glu Pro  
595 600 605

Phe Ser Ala Ala Asn Gly Leu Leu Ser Gly Val Gly Lys Leu Leu Arg  
610 615 620

Pro Asn Leu Lys Asp Arg Tyr Gly Gln Arg Leu Glu Gln Met Tyr Ala  
625 630 635 640

Asp Ile Ala Ala Thr Gln Ala Asn Gln Leu Arg Glu Leu Arg Arg Ala  
645 650 655

Ala Ala Thr Gln Pro Val Ile Asp Thr Leu Thr Gln Ala Ala Ala Thr  
660 665 670

Ile Leu Gly Thr Gly Ser Glu Val Ala Ser Asp Ala His Phe Thr Asp  
675 680 685

Leu Gly Gly Asp Ser Leu Ser Ala Leu Thr Leu Ser Asn Leu Leu Ser  
690 695 700

Asp Phe Phe Gly Phe Glu Val Pro Val Gly Thr Ile Val Asn Pro Ala  
705 710 715 720

Thr Asn Leu Ala Gln Leu Ala Gln His Ile Glu Ala Gln Arg Thr Ala  
725 730 735

Gly Asp Arg Arg Pro Ser Phe Thr Thr Val His Gly Ala Asp Ala Thr  
740 745 750

Glu Ile Arg Ala Ser Glu Leu Thr Leu Asp Lys Phe Ile Asp Ala Glu



ES 2 613 553 T3

Leu Ser Leu Leu Ile Thr Gly Val Ala Pro Arg Ser Phe Tyr Ile  
 1010 1015 1020  
  
 Gly Asp Gly Glu Arg Pro Arg Ala His Tyr Pro Gly Leu Thr Val  
 1025 1030 1035  
  
 Asp Phe Val Ala Glu Ala Val Thr Thr Leu Gly Ala Gln Gln Arg  
 1040 1045 1050  
  
 Glu Gly Tyr Val Ser Tyr Asp Val Met Asn Pro His Asp Asp Gly  
 1055 1060 1065  
  
 Ile Ser Leu Asp Val Phe Val Asp Trp Leu Ile Arg Ala Gly His  
 1070 1075 1080  
  
 Pro Ile Asp Arg Val Asp Asp Tyr Asp Asp Trp Val Arg Arg Phe  
 1085 1090 1095  
  
 Glu Thr Ala Leu Thr Ala Leu Pro Glu Lys Arg Arg Ala Gln Thr  
 1100 1105 1110  
  
 Val Leu Pro Leu Leu His Ala Phe Arg Ala Pro Gln Ala Pro Leu  
 1115 1120 1125  
  
 Arg Gly Ala Pro Glu Pro Thr Glu Val Phe His Ala Ala Val Arg  
 1130 1135 1140  
  
 Thr Ala Lys Val Gly Pro Gly Asp Ile Pro His Leu Asp Glu Ala  
 1145 1150 1155  
  
 Leu Ile Asp Lys Tyr Ile Arg Asp Leu Arg Glu Phe Gly Leu Ile  
 1160 1165 1170  
  
  
 <210> 8  
 <211> 209  
 <212> PRT  
 <213> Escherichia coli  
  
 <400> 8  
 Met Val Asp Met Lys Thr Thr His Thr Ser Leu Pro Phe Ala Gly His  
 1 5 10 15  
  
 Thr Leu His Phe Val Glu Phe Asp Pro Ala Asn Phe Cys Glu Gln Asp  
 20 25 30  
  
 Leu Leu Trp Leu Pro His Tyr Ala Gln Leu Gln His Ala Gly Arg Lys  
 35 40 45  
  
 Arg Lys Thr Glu His Leu Ala Gly Arg Ile Ala Ala Val Tyr Ala Leu



```

<220>
<221> fuente
<223> /nota="Descripción de secuencia artificial: Oligonucleótido
        sintético"

<400> 10
aaacggagcc ttcggctcc gttattcatt tacgcggtt caactttcct gtaggctgga      60
gctgcttc                                                                    68

<210> 11
<211> 23
<212> ADN
<213> Secuencia artificial

<220>
<221> fuente
<223> /nota="Descripción de secuencia artificial: Oligonucleótido
        sintético"

<400> 11
cgggcaggtg ctatgaccag gac                                                                    23

<210> 12
<211> 23
<212> ADN
<213> Secuencia artificial

<220>
<221> fuente
<223> /nota="Descripción de secuencia artificial: Oligonucleótido
        sintético"

<400> 12
cgcgcggttg accggcagcc tgg                                                                    23

<210> 13
<211> 5659
<212> ADN
<213> Secuencia artificial

<220>
<221> fuente
<223> /nota="Descripción de secuencia artificial: Polinucleótido
        sintético"

<400> 13
atgatcatca aacctaaaat tcgtggattt atctgtacaa caacgcaccc agtggggtgt      60
gaagcgaacg taaaagaaca aattgcctac acaaaagcac aagggtccgat caaaaacgca      120
cctaagcgcg tgttggttgt cggatcgtct agcggctatg gtctgtcatc acgcatcgct      180
gcggcgtttg gcggtggtgc ggcgacgatc ggcgatattt tcgaaaagcc gggcactgac      240
aaaaaaccag gtactgcggg tttctacaat gcagcagcgt ttgacaagct agcgcagtaa      300
gcgggcttgt acgcaaaaag cctgaacggc gatgcgttct cgaacgaagc gaagcaaaaa      360

```

ES 2 613 553 T3

gcgattgagc tgattaagca agacctcggc cagattgatt tggtaggtta ctcgttggct 420  
 tctccagtgc gtaagatgcc agacacgggt gagctagtgc gctctgcaact aaaaccgatc 480  
 ggcgaaacgt acacctctac cgcggtagat accaataaag atgtgatcat tgaagccagt 540  
 gttgaacctg cgaccgagca agaaatcgct gacactgtca ccgtgatggg cggccaagat 600  
 tgggaactgt ggatccaagc actggaagag gcgggtgttc ttgctgaagg ttgcaaaacc 660  
 gtggcgtaga gctacatcgg tactgaattg acttggccaa tttactggga tggcgcttta 720  
 ggccgtgcca agatggacct agatcgcgca gcgacagcgc tgaacgaaaa gctggcagcg 780  
 aaaggtggta ccgcgaacgt tgcagttttg aaatcagtgg tgactcaagc aagctctgcg 840  
 attcctgtga tgccgctcta catcgcaatg gtgttcaaga agatgctga acagggcgtg 900  
 catgaaggct gtatggagca gatctaccgc atgttcagtc aacgtctgta caaagaagat 960  
 ggttcagcgc cggaagtgga tgatcacaat cgtctgctgt ttgatgactg ggaactgctg 1020  
 gatgacattc agcagcactg ccgtgatctg tggccacaaa tcaccacaga gaactgctg 1080  
 gagctgaccg attacgacat gtacaaagaa gagttcatca agctgtttgg ctttggcatt 1140  
 gaaggcattg attacgatgc tgacgtcaat ccagaagtcg aatttgatgt gattgatatc 1200  
 gagtaattta gtgactgagc gtacatgtat acgaagatta ttggtactgg cagctatctg 1260  
 cccgaacaag tgcggactaa cgcgatctg gaaaaatgg ttgagacctc tgacgagtgg 1320  
 attgtcactc gtacaggtat tcgtaaacgc catatcgccg cgcgcaatga aactgtcgcg 1380  
 acgatgggct ttaccgctgc gaatcgcgcg attgagatgg cggggatcga taaagaccaa 1440  
 attggcttga ttgtggtggc taccacatca gcaacgcagc catttccaag cgcggcatgt 1500  
 cagattcaaa gtatgctcgg tattaaaggt tgcccggcgt ttgatgtcgc ggcagcgtgc 1560  
 gcaggtttca cctacgcggt aagcatcgcc gaccagtac ttaaatccgg cgcggttaa 1620  
 cacgcgctgg tggtaggttc cgatgtatta gcccgcaact gcgatcctgg cgatcgcggt 1680  
 acgatcatta ttttcggcga tggcgcaggc gcggccgtac tgagcgttc tgaagaaccg 1740  
 ggtattatct ccaactatct tcatgccgat ggccgttacg gtgaattact gaccctgccg 1800  
 aatgccgatc gcgtaaatcc ggataacccc atttacctga caatggcggg caatgaagtc 1860  
 ttaaaagtgg cggtcactga actggcgcagc attgtcgatg agacgctggc ggctaataac 1920  
 ctggatcgct cagaactcga ttggctggtg ccgcatcagg ctaacctgcg tatcattagc 1980  
 gcgacagcga aaaaactcgg catgtcgatg gacaatgtcg tcgtcacgct ggacaggcac 2040  
 ggcaataact ccgcggcttc tgtgccgtgc gcgctggatg aagccgtgcg tgacgggcca 2100  
 attaaagccg gtcagctggt attgcttgaa gccttcgggg gtggattcac ctggggctcc 2160  
 gcgctgattc gtttctagta taaggattta aacatgacgc aatttgatt tgtgttcccc 2220  
 ggtcaggggt ctacagcgtg tgggatgttg gccgagatgg cggcaatta ccctatcgta 2280

ES 2 613 553 T3

gaagaaacgt ttgctgaagc ttctgcggt ctgggatatg atctgtgggc gctcaccag 2340  
 caaggtccag cggaagaact gaataaaacc tggcagacgc agccggcggtt attaaccgct 2400  
 tccgtcgcgc tttggcgcgt ttggcagcag cagggcggtta aatgcctgc gttaatggca 2460  
 ggtcacagcc tggggaata ttccgcgctg gtttgcgctg gcgtcatcaa ctttgctgat 2520  
 gccgttcgtc tgggtgaaat gcgcggtaaa ttcatgcagg aagcggttcc ggaaggcact 2580  
 ggcggcatgt ctgcgatcat cgggctggat gatgcctcta ttgctaaagc ctgtgaagaa 2640  
 tctgccgaag ggcaggttgt ttcgccggtt aactttaact cgccgggaca ggtggttatc 2700  
 gccgggcata aagaggcggg agaacgtgcg ggcgcagcct gtaaagccgc tggcgcgaaa 2760  
 cgcgcgctgc cgctgccggt gagcgtaccg tcgcaactgc cgctgatgaa accagcggca 2820  
 gataagctgg cggttgaatt agccaaaatt acctttagcg cgccaacggg gccggtagtg 2880  
 aacaacggtg acgtgaaatg tgaaacgat gccgccgcta tccgcgatgc gctggttcgc 2940  
 cagttgtaca atccgtaca gtggacgaag agcgtggaat ttatcgcggc gcagggcggt 3000  
 gaacatcttt atgaagtggg tccaggtaaa gtccctactg gtctgacgaa acgtattgtc 3060  
 gacaccctga cagcgtcggc gctgaacgag ccggcggcgc tgtctgcggc acttacgcaa 3120  
 taaaagagga aaacatgag ctttgaagga aagattgcbc tggtgactgg tgcaagccgt 3180  
 ggcataggcc gcgcaattgc agagactctc gttgcccgcg gcgcgaaagt tatcgggact 3240  
 gcgaccagtg aaaatggtgc gaagaacatt agcgaactatt taggtgctaa cgggaaaggt 3300  
 ttgatgttga atgtgaccga tccctgcactc attgaatctg ttctggaaaa tattcgcgca 3360  
 gaatttggtg aagtggatat cctggttaat aatgccggta tcaactcgtga taatctgttg 3420  
 atgcgaatga aagatgatga gtggaacgat attatcgaaa ccaacttatc atccgttttc 3480  
 cgcctgtcaa aagcggtaat gcgcgctatg atgaaaaagc gttgtggtcg cattatcact 3540  
 attggttctg tggttgttac catgggaaat gcaggtcagg caaactacgc tcgggcgaaa 3600  
 gcgggcctga tcggtttcag taaatcactg gcgcgtgaag ttgcgtcccg tgggtattact 3660  
 gtcaatggtg tggctccggg ttttattgaa acggacatga cgcgtgcgct gtctgacgat 3720  
 cagcgtcggg gtatcctggc gcaggtgcct gcgggtcgcc tcggcggcgc tcaggaaatc 3780  
 gccagtgcgg ttgcattttt agcctctgac gaagcgagtt acatcactgg tgagactctg 3840  
 cacgtcaacg gcggaatgta catggtttaa ttttaagggt tacataaaac atggtagata 3900  
 aacgcgaatc ctatacaaaa gaagacctc ttgcctctgg tcgtggtgaa ctggttggcg 3960  
 ctaaagggcc gcaactccct gcaccgaaca tgctgatgat ggaccgcgtc gttaagatga 4020  
 ccgaaacggg cggcaatttc gacaaaggct atgtogaagc cgagctggat atcaatccgg 4080  
 atctatggtt cttcggatgc cactttatcg gcgatccggt gatgcccggg tgtctgggtc 4140

ES 2 613 553 T3

tggatgctat gtggcaattg gtgggattct acctgggctg gttgggcggc gaaggcaaag 4200  
gccgcgctct gggcgtgggc gaagtgaaat ttaccggcca ggttctgccg acagccagga 4260  
aagtcaccta tcgtattcat ttcaaacgta tcgtaaaccg tcgcctgatc atgggcctgg 4320  
cggacggtga ggttctggtg gatggtcgcc tgatctatac cgcacacgat ttgaaagtgcg 4380  
gtttgttcca ggatacttcc gcgttctaaa aggaggcaac aaaatgaatc gccgcgttgt 4440  
cattacgggt attggtgcag tgacgccggt gggtaacaac gctgatagct tctggtgcag 4500  
catcaaagag ggtaaattg gcaattgaaa gatcaaagcg tttgacgcaa ccgatttcaa 4560  
agttaagctg gctgccgaag tgaaggactt caccocggag gactttatcg acaagcgtga 4620  
ggcgaaccgt atggaccgtt ttagccagtt tgcgatcggt gcggcggatg aggcaatcaa 4680  
ggacagcaaa ctggacctg agtcogattga taagaatcgt ttcggcgtca ttggtggtag 4740  
cggcattggc ggcacgcca ccattgagaa gcaggatgaa aagctgatta ccaaaggctc 4800  
gggtcgtgtg agccctatga ctattccgat gatcattgcg aatatggcaa gcggtaatct 4860  
ggcgattcgt tatggcgcta aaggtatttg cacgaccatt gtcaccgcat gtgcgagcgc 4920  
gaacaacagc attggtgagt ccttccgtaa cattaagttt ggttatagcg acgttatgat 4980  
ctctggtggt agcgaagcag gtatcaccct gttgagcctg gcgggttttg cctcgatgaa 5040  
ggccgtgacc aaatctgagg acccgaagcg cgccagcacc ccgttcgata aggatcgcag 5100  
cggttttgtg atgggcgagg gcagcgggat cgttatcttg gaagagttgg agcacgcgct 5160  
gaagcgtggt gccaaaatct atgccgagat cgttggtat ggtgcgacct gcgacgcata 5220  
tcatatcaag agcccagcgc cgaatggtga agtggtgca cgtgcaatga aactggcaat 5280  
ggaagaagat aatgtccgcc cagaggacat ttcctatata aacgcgcacg gtacgagcac 5340  
ggcgtacaat gacagcttcg aaaccaagc gatcaagacg gtcctgggtg aatacgccta 5400  
caaagtgccg gtgtctagca ccaagagcat gaccggccac ctgctgggcg ctggcgggtgc 5460  
agtcgaagcg attatctgtg ccaaagctat tgaagagggg ttcatcctgc cgaccatcgg 5520  
ctacaaagag gcggatccgg aatgcgacct ggattacggt cctaacgagg gccgtaatgc 5580  
agaagtcaac tacgttctgt ccaacagcct gggcttcggt ggccataatg cgactctgct 5640  
gttcaaaaag tacaatga 5659

<210> 14  
<211> 40  
<212> ADN  
<213> Secuencia artificial  
<220>  
<221> fuente  
<223> /nota="Descripción de secuencia artificial: Oligonucleótido sintético"

<400> 14  
gcagttattg gtgcccttaa acgcctggtt gctacgcctg 40

<210> 15  
<211> 27  
<212> ADN  
<213> Secuencia artificial

<220>  
<221> fuente  
<223> /nota="Descripción de secuencia artificial: Oligonucleótido sintético"

<400> 15  
gagccaatat gcgagaacac ccgagaa 27

<210> 16  
<211> 48  
<212> ADN  
<213> Secuencia artificial

<220>  
<221> fuente  
<223> /nota="Descripción de secuencia artificial: Oligonucleótido sintético"

<400> 16  
cgctgaacgt attgcaggcc gagttgctgc accgctcccg ccaggcag 48

<210> 17  
<211> 51  
<212> ADN  
<213> Secuencia artificial

<220>  
<221> fuente  
<223> /nota="Descripción de secuencia artificial: Oligonucleótido sintético"

<400> 17  
ggaattgcca cggtgcgga ggctccatac gcgaggccag gttatccaac g 51

<210> 18  
<211> 35  
<212> ADN  
<213> Secuencia artificial

<220>  
<221> fuente  
<223> /nota="Descripción de secuencia artificial: Oligonucleótido sintético"

<400> 18  
aatcaccagc actaaagtgc gcggttcggt acccg 35

<210> 19  
<211> 34  
<212> ADN

<213> Secuencia artificial

<220>  
 <221> fuente  
 <223> /nota="Descripción de secuencia artificial:Oligonucleótido sintético"

<400> 19  
 atctgccgtg gattgcagag tctattcagc tacg 34

<210> 20  
 <211> 30  
 <212> ADN  
 <213> Secuencia artificial

<220>  
 <221> fuente  
 <223> /nota="Descripción de secuencia artificial:Cebador sintético"

<400> 20  
 gcaattccat atgacgagcg atgttcacga 30

<210> 21  
 <211> 29  
 <212> ADN  
 <213> Secuencia artificial

<220>  
 <221> fuente  
 <223> /nota="Descripción de secuencia artificial:Cebador sintético"

<400> 21  
 ccgctcgagt aatcagacc gaactcgcg 29

<210> 22  
 <211> 60  
 <212> ADN  
 <213> Secuencia artificial

<220>  
 <221> fuente  
 <223> /nota="Descripción de secuencia artificial:Oligonucleótido sintético"

<400> 22  
 atgggcagca gccatcatca tcatcatcac agcagcggcc tggcgccgcg cggcagccat 60

<210> 23  
 <211> 43  
 <212> ADN  
 <213> Secuencia artificial

<220>  
 <221> fuente  
 <223> /nota="Descripción de secuencia artificial: Oligonucleótido sintético"

## ES 2 613 553 T3

<400> 23  
 acggatcccc ggaatgcgca acgcaattaa tgtaagttag cgc 43

<210> 24  
 <211> 46  
 <212> ADN  
 <213> Secuencia artificial

<220>  
 <221> fuente  
 <223> /nota="Descripción de secuencia artificial: Oligonucleótido sintético"

<400> 24  
 tgcgtcatcg ccattgaatt octaaatcag accgaactcg cgcagc 46

<210> 25  
 <211> 20  
 <212> ADN  
 <213> Secuencia artificial

<220>  
 <221> fuente  
 <223> /nota="Descripción de secuencia artificial: Oligonucleótido sintético"

<400> 25  
 attccgggga tccgtcgacc 20

<210> 26  
 <211> 23  
 <212> ADN  
 <213> Secuencia artificial

<220>  
 <221> fuente  
 <223> /nota="Descripción de secuencia artificial: Oligonucleótido sintético"

<400> 26  
 aatggcgatg acgcatcctc acg 23

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

<220>  
 <221> fuente  
 <223> /nota="Descripción de secuencia artificial: Cebador sintético"

<400> 27  
 acggaaagga gctagcacat gggcagcagc catcatcat 39

<210> 28  
 <211> 27  
 <212> ADN

<213> Secuencia artificial

<220>  
 <221> fuente  
 <223> /nota="Descripción de secuencia artificial:Cebador sintético"

<400> 28  
 gtaaaggatg gacggcggtc acccgcc 27

<210> 29  
 <211> 25  
 <212> ADN  
 <213> Secuencia artificial

<220>  
 <221> fuente  
 <223> /nota="Descripción de secuencia artificial:Cebador sintético"

<400> 29  
 cacggcgggt gaccgccgtc catcc 25

<210> 30  
 <211> 43  
 <212> ADN  
 <213> Secuencia artificial

<220>  
 <221> fuente  
 <223> /nota="Descripción de secuencia artificial:Cebador sintético"

<400> 30  
 ttaattccgg g gatccctaa atcagaccga actcgcgcag gtc 43

<210> 31  
 <211> 24  
 <212> ADN  
 <213> Secuencia artificial

<220>  
 <221> fuente  
 <223> /nota="Descripción de secuencia artificial:Cebador sintético"

<400> 31  
 cagccgttta ttgccgactg gatg 24

<210> 32  
 <211> 27  
 <212> ADN  
 <213> Secuencia artificial

<220>  
 <221> fuente  
 <223> /nota="Descripción de secuencia artificial:Cebador sintético"

<400> 32  
ctgttttatc agaccgcttc tgcgttc 27

<210> 33  
<211> 20  
<212> ADN  
<213> Secuencia artificial

<220>  
<221> fuente  
<223> /nota="Descripción de secuencia artificial:Cebador sintético"

<400> 33  
gcactcgacc ggaattatcg 20

<210> 34  
<211> 25  
<212> ADN  
<213> Secuencia artificial

<220>  
<221> fuente  
<223> /nota="Descripción de secuencia artificial:Cebador sintético"

<400> 34  
gcactacgcg tactgtgagc cagag 25

<210> 35  
<211> 45  
<212> ADN  
<213> Secuencia artificial

<220>  
<221> fuente  
<223> /nota="Descripción de secuencia artificial:Cebador sintético"

<400> 35  
gaggaataaa ccatgacgag cgatgttcac gacgcgaccg acggc 45

<210> 36  
<211> 25  
<212> ADN  
<213> Secuencia artificial

<220>  
<221> fuente  
<223> /nota="Descripción de secuencia artificial:Cebador sintético"

<400> 36  
ctaaatcaga ccgaactcgc gcagg 25

<210> 37  
<211> 31  
<212> ADN

<213> Secuencia artificial

<220>

<221> fuente

<223> /nota="Descripción de secuencia artificial:Cebador sintético"

<400> 37

catggtttat tcctccttat ttaatcgata c

31

<210> 38

<211> 28

<212> ADN

<213> Secuencia artificial

<220>

<221> fuente

<223> /nota="Descripción de secuencia artificial:Cebador sintético"

<400> 38

tgacctgctc gagttcggtc tgatttag

28

<210> 39

<211> 1168

<212> PRT

<213> Mycobacterium smegmatis

<400> 39

Met Thr Ile Glu Thr Arg Glu Asp Arg Phe Asn Arg Arg Ile Asp His  
1 5 10 15

Leu Phe Glu Thr Asp Pro Gln Phe Ala Ala Ala Arg Pro Asp Glu Ala  
20 25 30

Ile Ser Ala Ala Ala Ala Asp Pro Glu Leu Arg Leu Pro Ala Ala Val  
35 40 45

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

Ala Val Glu Phe Val Thr Asp Glu Glu Gly Arg Thr Thr Ala Lys Leu  
65 70 75 80

Leu Pro Arg Phe Asp Thr Ile Thr Tyr Arg Gln Leu Ala Gly Arg Ile  
85 90 95

Gln Ala Val Thr Asn Ala Trp His Asn His Pro Val Asn Ala Gly Asp  
100 105 110

Arg Val Ala Ile Leu Gly Phe Thr Ser Val Asp Tyr Thr Thr Ile Asp  
115 120 125

ES 2 613 553 T3

Ile Ala Leu Leu Glu Leu Gly Ala Val Ser Val Pro Leu Gln Thr Ser  
130 135 140

Ala Pro Val Ala Gln Leu Gln Pro Ile Val Ala Glu Thr Glu Pro Lys  
145 150 155 160

Val Ile Ala Ser Ser Val Asp Phe Leu Ala Asp Ala Val Ala Leu Val  
165 170 175

Glu Ser Gly Pro Ala Pro Ser Arg Leu Val Val Phe Asp Tyr Ser His  
180 185 190

Glu Val Asp Asp Gln Arg Glu Ala Phe Glu Ala Ala Lys Gly Lys Leu  
195 200 205

Ala Gly Thr Gly Val Val Val Glu Thr Ile Thr Asp Ala Leu Asp Arg  
210 215 220

Gly Arg Ser Leu Ala Asp Ala Pro Leu Tyr Val Pro Asp Glu Ala Asp  
225 230 235 240

Pro Leu Thr Leu Leu Ile Tyr Thr Ser Gly Ser Thr Gly Thr Pro Lys  
245 250 255

Gly Ala Met Tyr Pro Glu Ser Lys Thr Ala Thr Met Trp Gln Ala Gly  
260 265 270

Ser Lys Ala Arg Trp Asp Glu Thr Leu Gly Val Met Pro Ser Ile Thr  
275 280 285

Leu Asn Phe Met Pro Met Ser His Val Met Gly Arg Gly Ile Leu Cys  
290 295 300

Ser Thr Leu Ala Ser Gly Gly Thr Ala Tyr Phe Ala Ala Arg Ser Asp  
305 310 315 320

Leu Ser Thr Phe Leu Glu Asp Leu Ala Leu Val Arg Pro Thr Gln Leu  
325 330 335

Asn Phe Val Pro Arg Ile Trp Asp Met Leu Phe Gln Glu Tyr Gln Ser  
340 345 350

Arg Leu Asp Asn Arg Arg Ala Glu Gly Ser Glu Asp Arg Ala Glu Ala  
355 360 365

Ala Val Leu Glu Glu Val Arg Thr Gln Leu Leu Gly Gly Arg Phe Val  
370 375 380

ES 2 613 553 T3

Ser Ala Leu Thr Gly Ser Ala Pro Ile Ser Ala Glu Met Lys Ser Trp  
385 390 395 400

Val Glu Asp Leu Leu Asp Met His Leu Leu Glu Gly Tyr Gly Ser Thr  
405 410 415

Glu Ala Gly Ala Val Phe Ile Asp Gly Gln Ile Gln Arg Pro Pro Val  
420 425 430

Ile Asp Tyr Lys Leu Val Asp Val Pro Asp Leu Gly Tyr Phe Ala Thr  
435 440 445

Asp Arg Pro Tyr Pro Arg Gly Glu Leu Leu Val Lys Ser Glu Gln Met  
450 455 460

Phe Pro Gly Tyr Tyr Lys Arg Pro Glu Ile Thr Ala Glu Met Phe Asp  
465 470 475 480

Glu Asp Gly Tyr Tyr Arg Thr Gly Asp Ile Val Ala Glu Leu Gly Pro  
485 490 495

Asp His Leu Glu Tyr Leu Asp Arg Arg Asn Asn Val Leu Lys Leu Ser  
500 505 510

Gln Gly Glu Phe Val Thr Val Ser Lys Leu Glu Ala Val Phe Gly Asp  
515 520 525

Ser Pro Leu Val Arg Gln Ile Tyr Val Tyr Gly Asn Ser Ala Arg Ser  
530 535 540

Tyr Leu Leu Ala Val Val Val Pro Thr Glu Glu Ala Leu Ser Arg Trp  
545 550 555 560

Asp Gly Asp Glu Leu Lys Ser Arg Ile Ser Asp Ser Leu Gln Asp Ala  
565 570 575

Ala Arg Ala Ala Gly Leu Gln Ser Tyr Glu Ile Pro Arg Asp Phe Leu  
580 585 590

Val Glu Thr Thr Pro Phe Thr Leu Glu Asn Gly Leu Leu Thr Gly Ile  
595 600 605

Arg Lys Leu Ala Arg Pro Lys Leu Lys Ala His Tyr Gly Glu Arg Leu  
610 615 620

Glu Gln Leu Tyr Thr Asp Leu Ala Glu Gly Gln Ala Asn Glu Leu Arg  
625 630 635 640

ES 2 613 553 T3

Glu Leu Arg Arg Asn Gly Ala Asp Arg Pro Val Val Glu Thr Val Ser  
 645 650 655  
 Arg Ala Ala Val Ala Leu Leu Gly Ala Ser Val Thr Asp Leu Arg Ser  
 660 665 670  
 Asp Ala His Phe Thr Asp Leu Gly Gly Asp Ser Leu Ser Ala Leu Ser  
 675 680 685  
 Phe Ser Asn Leu Leu His Glu Ile Phe Asp Val Asp Val Pro Val Gly  
 690 695 700  
 Val Ile Val Ser Pro Ala Thr Asp Leu Ala Gly Val Ala Ala Tyr Ile  
 705 710 715 720  
 Glu Gly Glu Leu Arg Gly Ser Lys Arg Pro Thr Tyr Ala Ser Val His  
 725 730 735  
 Gly Arg Asp Ala Thr Glu Val Arg Ala Arg Asp Leu Ala Leu Gly Lys  
 740 745 750  
 Phe Ile Asp Ala Lys Thr Leu Ser Ala Ala Pro Gly Leu Pro Arg Ser  
 755 760 765  
 Gly Thr Glu Ile Arg Thr Val Leu Leu Thr Gly Ala Thr Gly Phe Leu  
 770 775 780  
 Gly Arg Tyr Leu Ala Leu Glu Trp Leu Glu Arg Met Asp Leu Val Asp  
 785 790 795 800  
 Gly Lys Val Ile Cys Leu Val Arg Ala Arg Ser Asp Asp Glu Ala Arg  
 805 810 815  
 Ala Arg Leu Asp Ala Thr Phe Asp Thr Gly Asp Ala Thr Leu Leu Glu  
 820 825 830  
 His Tyr Arg Ala Leu Ala Ala Asp His Leu Glu Val Ile Ala Gly Asp  
 835 840 845  
 Lys Gly Glu Ala Asp Leu Gly Leu Asp His Asp Thr Trp Gln Arg Leu  
 850 855 860  
 Ala Asp Thr Val Asp Leu Ile Val Asp Pro Ala Ala Leu Val Asn His  
 865 870 875 880  
 Val Leu Pro Tyr Ser Gln Met Phe Gly Pro Asn Ala Leu Gly Thr Ala

ES 2 613 553 T3

			885						890						895
Glu	Leu	Ile	Arg	Ile	Ala	Leu	Thr	Thr	Thr	Ile	Lys	Pro	Tyr	Val	Tyr
			900						905						910
Val	Ser	Thr	Ile	Gly	Val	Gly	Gln	Gly	Ile	Ser	Pro	Glu	Ala	Phe	Val
		915					920					925			
Glu	Asp	Ala	Asp	Ile	Arg	Glu	Ile	Ser	Ala	Thr	Arg	Arg	Val	Asp	Asp
	930					935					940				
Ser	Tyr	Ala	Asn	Gly	Tyr	Gly	Asn	Ser	Lys	Trp	Ala	Gly	Glu	Val	Leu
945					950					955					960
Leu	Arg	Glu	Ala	His	Asp	Trp	Cys	Gly	Leu	Pro	Val	Ser	Val	Phe	Arg
				965					970					975	
Cys	Asp	Met	Ile	Leu	Ala	Asp	Thr	Thr	Tyr	Ser	Gly	Gln	Leu	Asn	Leu
			980						985					990	
Pro	Asp	Met	Phe	Thr	Arg	Leu	Met	Leu	Ser	Leu	Val	Ala	Thr	Gly	Ile
		995					1000							1005	
Ala	Pro	Gly	Ser	Phe	Tyr	Glu	Leu	Asp	Ala	Asp	Gly	Asn	Arg	Gln	
	1010					1015					1020				
Arg	Ala	His	Tyr	Asp	Gly	Leu	Pro	Val	Glu	Phe	Ile	Ala	Glu	Ala	
	1025					1030					1035				
Ile	Ser	Thr	Ile	Gly	Ser	Gln	Val	Thr	Asp	Gly	Phe	Glu	Thr	Phe	
	1040					1045					1050				
His	Val	Met	Asn	Pro	Tyr	Asp	Asp	Gly	Ile	Gly	Leu	Asp	Glu	Tyr	
	1055					1060					1065				
Val	Asp	Trp	Leu	Ile	Glu	Ala	Gly	Tyr	Pro	Val	His	Arg	Val	Asp	
	1070					1075					1080				
Asp	Tyr	Ala	Thr	Trp	Leu	Ser	Arg	Phe	Glu	Thr	Ala	Leu	Arg	Ala	
	1085					1090					1095				
Leu	Pro	Glu	Arg	Gln	Arg	Gln	Ala	Ser	Leu	Leu	Pro	Leu	Leu	His	
	1100					1105					1110				
Asn	Tyr	Gln	Gln	Pro	Ser	Pro	Pro	Val	Cys	Gly	Ala	Met	Ala	Pro	
	1115					1120					1125				

ES 2 613 553 T3

Thr Asp Arg Phe Arg Ala Ala Val Gln Asp Ala Lys Ile Gly Pro  
 1130 1135 1140

Asp Lys Asp Ile Pro His Val Thr Ala Asp Val Ile Val Lys Tyr  
 1145 1150 1155

Ile Ser Asn Leu Gln Met Leu Gly Leu Leu  
 1160 1165

<210> 40  
 <211> 1168  
 <212> PRT  
 <213> Mycobacterium tuberculosis

<400> 40  
 Met Ser Ile Asn Asp Gln Arg Leu Thr Arg Arg Val Glu Asp Leu Tyr  
 1 5 10 15

Ala Ser Asp Ala Gln Phe Ala Ala Ala Ser Pro Asn Glu Ala Ile Thr  
 20 25 30

Gln Ala Ile Asp Gln Pro Gly Val Ala Leu Pro Gln Leu Ile Arg Met  
 35 40 45

Val Met Glu Gly Tyr Ala Asp Arg Pro Ala Leu Gly Gln Arg Ala Leu  
 50 55 60

Arg Phe Val Thr Asp Pro Asp Ser Gly Arg Thr Met Val Glu Leu Leu  
 65 70 75 80

Pro Arg Phe Glu Thr Ile Thr Tyr Arg Glu Leu Trp Ala Arg Ala Gly  
 85 90 95

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

Arg Val Cys Val Leu Gly Phe Asn Ser Val Asp Tyr Thr Thr Ile Asp  
 115 120 125

Ile Ala Leu Ile Arg Leu Gly Ala Val Ser Val Pro Leu Gln Thr Ser  
 130 135 140

Ala Pro Val Thr Gly Leu Arg Pro Ile Val Thr Glu Thr Glu Pro Thr  
 145 150 155 160

Met Ile Ala Thr Ser Ile Asp Asn Leu Gly Asp Ala Val Glu Val Leu  
 165 170 175

Ala Gly His Ala Pro Ala Arg Leu Val Val Phe Asp Tyr His Gly Lys



ES 2 613 553 T3

Leu Val Asp Val Pro Glu Leu Gly Tyr Phe Gly Thr Asp Gln Pro Tyr  
 435 440 445

Pro Arg Gly Glu Leu Leu Val Lys Thr Gln Thr Met Phe Pro Gly Tyr  
 450 455 460

Tyr Gln Arg Pro Asp Val Thr Ala Glu Val Phe Asp Pro Asp Gly Phe  
 465 470 475 480

Tyr Arg Thr Gly Asp Ile Met Ala Lys Val Gly Pro Asp Gln Phe Val  
 485 490 495

Tyr Leu Asp Arg Arg Asn Asn Val Leu Lys Leu Ser Gln Gly Glu Phe  
 500 505 510

Ile Ala Val Ser Lys Leu Glu Ala Val Phe Gly Asp Ser Pro Leu Val  
 515 520 525

Arg Gln Ile Phe Ile Tyr Gly Asn Ser Ala Arg Ala Tyr Pro Leu Ala  
 530 535 540

Val Val Val Pro Ser Gly Asp Ala Leu Ser Arg His Gly Ile Glu Asn  
 545 550 555 560

Leu Lys Pro Val Ile Ser Glu Ser Leu Gln Glu Val Ala Arg Ala Ala  
 565 570 575

Gly Leu Gln Ser Tyr Glu Ile Pro Arg Asp Phe Ile Ile Glu Thr Thr  
 580 585 590

Pro Phe Thr Leu Glu Asn Gly Leu Leu Thr Gly Ile Arg Lys Leu Ala  
 595 600 605

Arg Pro Gln Leu Lys Lys Phe Tyr Gly Glu Arg Leu Glu Arg Leu Tyr  
 610 615 620

Thr Glu Leu Ala Asp Ser Gln Ser Asn Glu Leu Arg Glu Leu Arg Gln  
 625 630 635 640

Ser Gly Pro Asp Ala Pro Val Leu Pro Thr Leu Cys Arg Ala Ala Ala  
 645 650 655

Ala Leu Leu Gly Ser Thr Ala Ala Asp Val Arg Pro Asp Ala His Phe  
 660 665 670

Ala Asp Leu Gly Gly Asp Ser Leu Ser Ala Leu Ser Leu Ala Asn Leu  
 675 680 685

ES 2 613 553 T3

Leu His Glu Ile Phe Gly Val Asp Val Pro Val Gly Val Ile Val Ser  
 690 695 700

Pro Ala Ser Asp Leu Arg Ala Leu Ala Asp His Ile Glu Ala Ala Arg  
 705 710 715 720

Thr Gly Val Arg Arg Pro Ser Phe Ala Ser Ile His Gly Arg Ser Ala  
 725 730 735

Thr Glu Val His Ala Ser Asp Leu Thr Leu Asp Lys Phe Ile Asp Ala  
 740 745 750

Ala Thr Leu Ala Ala Ala Pro Asn Leu Pro Ala Pro Ser Ala Gln Val  
 755 760 765

Arg Thr Val Leu Leu Thr Gly Ala Thr Gly Phe Leu Gly Arg Tyr Leu  
 770 775 780

Ala Leu Glu Trp Leu Asp Arg Met Asp Leu Val Asn Gly Lys Leu Ile  
 785 790 795 800

Cys Leu Val Arg Ala Arg Ser Asp Glu Glu Ala Gln Ala Arg Leu Asp  
 805 810 815

Ala Thr Phe Asp Ser Gly Asp Pro Tyr Leu Val Arg His Tyr Arg Glu  
 820 825 830

Leu Gly Ala Gly Arg Leu Glu Val Leu Ala Gly Asp Lys Gly Glu Ala  
 835 840 845

Asp Leu Gly Leu Asp Arg Val Thr Trp Gln Arg Leu Ala Asp Thr Val  
 850 855 860

Asp Leu Ile Val Asp Pro Ala Ala Leu Val Asn His Val Leu Pro Tyr  
 865 870 875 880

Ser Gln Leu Phe Gly Pro Asn Ala Ala Gly Thr Ala Glu Leu Leu Arg  
 885 890 895

Leu Ala Leu Thr Gly Lys Arg Lys Pro Tyr Ile Tyr Thr Ser Thr Ile  
 900 905 910

Ala Val Gly Glu Gln Ile Pro Pro Glu Ala Phe Thr Glu Asp Ala Asp  
 915 920 925

Ile Arg Ala Ile Ser Pro Thr Arg Arg Ile Asp Asp Ser Tyr Ala Asn  
 930 935 940

ES 2 613 553 T3

Gly Tyr Ala Asn Ser Lys Trp Ala Gly Glu Val Leu Leu Arg Glu Ala  
 945 950 955 960

His Glu Gln Cys Gly Leu Pro Val Thr Val Phe Arg Cys Asp Met Ile  
 965 970 975

Leu Ala Asp Thr Ser Tyr Thr Gly Gln Leu Asn Leu Pro Asp Met Phe  
 980 985 990

Thr Arg Leu Met Leu Ser Leu Ala Ala Thr Gly Ile Ala Pro Gly Ser  
 995 1000 1005

Phe Tyr Glu Leu Asp Ala His Gly Asn Arg Gln Arg Ala His Tyr  
 1010 1015 1020

Asp Gly Leu Pro Val Glu Phe Val Ala Glu Ala Ile Cys Thr Leu  
 1025 1030 1035

Gly Thr His Ser Pro Asp Arg Phe Val Thr Tyr His Val Met Asn  
 1040 1045 1050

Pro Tyr Asp Asp Gly Ile Gly Leu Asp Glu Phe Val Asp Trp Leu  
 1055 1060 1065

Asn Ser Pro Thr Ser Gly Ser Gly Cys Thr Ile Gln Arg Ile Ala  
 1070 1075 1080

Asp Tyr Gly Glu Trp Leu Gln Arg Phe Glu Thr Ser Leu Arg Ala  
 1085 1090 1095

Leu Pro Asp Arg Gln Arg His Ala Ser Leu Leu Pro Leu Leu His  
 1100 1105 1110

Asn Tyr Arg Glu Pro Ala Lys Pro Ile Cys Gly Ser Ile Ala Pro  
 1115 1120 1125

Thr Asp Gln Phe Arg Ala Ala Val Gln Glu Ala Lys Ile Gly Pro  
 1130 1135 1140

Asp Lys Asp Ile Pro His Leu Thr Ala Ala Ile Ile Ala Lys Tyr  
 1145 1150 1155

Ile Ser Asn Leu Arg Leu Leu Gly Leu Leu  
 1160 1165

<210> 41

ES 2 613 553 T3

<211> 3507

<212> ADN

<213> *Mycobacterium smegmatis*

<400> 41

```

atgacgatcg aaacgcgcga agaccgcttc aaccggcgca ttgaccactt gttcgaaaacc      60
gacccgcagt tcgccgccgc ccgtcccgcg gaggggatca gcgcggctgc cgccgatccg      120
gagttgcgcc ttctctccgc ggtcaaacag attctggccg gctatgcgga ccgccctgcg      180
ctgggcaagc gcgccgtcga gttcgtcacc gacgaagaag gccgcaccac cgcaagctc      240
ctgccccgct tcgacaccat cacctaccgt cagctcgcag gccggatcca ggccgtgacc      300
aatgcctggc acaaccatcc ggtgaatgcc ggtgaccgcg tggccatcct gggtttcacc      360
agtgtcgact acacgacgat cgacatcgcc ctgctcgaac tcggcgccgt gtccgtaccg      420
ctgcagacca gtgcgccggt ggcccaactg cagccgatcg tcgccgagac cgagcccaag      480
gtgatcgcgt cgagcgtcga cttcctcgcc gacgcagtcg ctctcgtcga gtccgggccc      540
gcgccgtcgc gactggtggt gttcgactac agccacgagg tcgacgatca gcgtgaggcg      600
ttcagaggcg ccaagggcaa gctcgcaggc accggcgctc tcgtcgagac gatcaccgac      660
gactggacc gcgggcggtc actcgccgac gcaccgctct acgtgccoga cgaggccgac      720
ccgtgacccc ttctcatcta cacctccggc agcaccggca ctcccaaggg cgcgatgtac      780
cccgagtcca agaccgccac gatgtggcag gccgggtcca aggcccggtg ggacgagacc      840
ctcggcgtga tgccgtcgat caccctgaac ttcatgccca tgagtcacgt catggggcgc      900
ggcatcctgt gcagcacact cgccagcggc ggaaccgcgt acttcgccgc acgcagcgac      960
ctgtccacct tcctggagga cctcgccctc gtgcggccca cgcagctcaa cttcgttcct     1020
cgcatctggg acatgctggt ccaggagtac cagagccgcc tcgacaaccg ccgcgccgag     1080
ggatccgagg accgagccga agccgcagtc ctcgaagagg tccgcaccca actgctcggc     1140
gggcgattog tttcggccct gaccggatcg gctcccatct cggcggagat gaagagctgg     1200
gtcgaggacc tgctcgacat gcatctgctg gagggctacg gctccaccga ggccggcgcg     1260
gtgttcatcg acgggcagat ccagcgcgcc ccggtcatcg actacaagct ggtcgacgtg     1320
cccgatctcg gctacttcgc cacggaccgg ccctaccgcg gcggcgaact tctggtcaag     1380
tccgagcaga tgttccccgg ctactacaag cgtccggaga tcaccgccga gatgttcgac     1440
gaggacgggt actaccgcac cggcgacatc gtcgccgagc tcgggcccga ccatctcgaa     1500
tacctcgacc gccgcaaaa cgtgctgaaa ctgtcgcagg gcgaattcgt cacggtctcc     1560
aagctggagg cgggtgttcgg cgacagcccc ctggtacgcc agatctacgt ctacggcaac     1620
agcgcgcggt cctatctgct ggcggtcgtg gtcccgaccg aagaggcact gtcacgttgg     1680
gacggtgacg aactcaagtc gcgcatcagc gactcactgc aggacgcggc acgagccgcc     1740

```

ES 2 613 553 T3

ggattgcagt cgtatgagat cccgcgtgac ttctctcgctc agacaacacc tttcacgctg 1800  
 gagaacggcc tgctgaccgg tatccgcaag ctggcccggc cgaaactgaa ggcgactac 1860  
 ggcaacgcc tcgaacagct ctacaccgac ctggccgagg ggcaggccaa cgagttgcgc 1920  
 gagttgcgcc gcaacggagc cgaccggccc gtggctgaga ccgtcagccg cgcccgggtc 1980  
 gcaactgctc gtgcctccgt cacggatctg cggctccgatg cgcacttcac cgatctgggt 2040  
 ggagattcgt tgtcggcctt gagcttctcg aacctgttg acgagatctt cgatgtcgac 2100  
 gtgccggctg gcgtcatcgt cagcccggcc accgacctgg caggcgtcgc ggcctacatc 2160  
 gagggcgaac tgcgcggctc caagcgcctc acatacgcgt cggtgacagg gcgcgacgcc 2220  
 accgaggtgc gcgcgctga tctcgccctg ggcaagttca tcgacgcaa gacctgtcc 2280  
 gccgcgccgg gtctgcccg ttcgggcacc gagatccgca ccgctgctgct gaccggcgcc 2340  
 accgggttcc tggcccgcta tctggcgctg gaatggctgg agcgcagga cctggtggac 2400  
 ggcaaggtga tctgcctggt gcgcgcccgc agcgcgacg aggcccgggc gcgtctggac 2460  
 gccacgttct acaccgggga cgcgacactg ctcgagcact accgcgcgct ggcagccgat 2520  
 cacctcgagg tgatcgccgg tgacaaggc gaggccgatc tgggtctcga ccacgacacg 2580  
 tggcagcgac tggccgacac cgtcgatctg atcgtcgatc cggccgacct ggtcaatcac 2640  
 gtctgcccgt acagccagat gttcggacct aatgcgctcg gcaccgcca actcatccgg 2700  
 atcgcgctga ccaccacgat caagccgtac gtgtacgtct cgacgatcgg tgtgggacag 2760  
 ggcattctcc ccgaggcgtt cgtcgaggac gccgacatcc gcgagatcag cgcgacgcgc 2820  
 cgggtcgacg actcgtacgc caacggctac ggcaacagca agtgggcccg cgaggtcctg 2880  
 ctgcgggagg cgcacgactg gtgtggctcg ccggtctcgg tgttccgctg cgacatgatc 2940  
 ctggccgaca cgacctactc gggtcagctg aacctgccg acatgttcac ccgcctgatg 3000  
 ctgagcctcg tggcgaccgg catcgcgccc ggttcgttct acgaactcga tgcggacggc 3060  
 aaccggcagc gcgcccacta cgacgggctg ccggtggagt tcatcgcca ggcgatctcc 3120  
 accatcggtc cgcaggtcac cgacggattc gagacgttcc acgtgatgaa cccgtacgac 3180  
 gacggcatcg gcctcgacga gtacgtggac tggctgatcg aggcgggcta ccccgtcac 3240  
 cgcgtcgacg actacgccac ctggctgagc cggttcgaaa ccgcaactgc ggcctgccg 3300  
 gaacggcaac gtcaggcctc gctgctgccg ctgctgcaca actatcagca gccctcaccg 3360  
 cccgtgtgcg gtgccatggc acccaccgac cggttccgtg ccgcgggtgca ggacgcgaag 3420  
 atcgccccg acaaggacat tccgcacgtc acggccgacg tgatcgtcaa gtacatcagc 3480  
 aacctgcaga tgctcggatt gctgtaa 3507

<210> 42  
 <211> 3507

ES 2 613 553 T3

<212> ADN

<213> *Mycobacterium tuberculosis*

<400> 42

```

atgtcgatca acgatcagcg actgacacgc cgcgtcagagg acctatacgc cagcgacgcc      60
cagttcgccg ccgccagtcc caacgaggcg atcaccagcg cgatcgacca gcccggggtc      120
gcgcttccac agctcatccg tatggtcagtg gagggctacg ccgatcggcc ggcaactcggc      180
cagcgtgcmc tccgcttcgt caccgacccc gacagcggcc gcaccatggt cgagctactg      240
ccgcggttcg agaccatcac ctaccgcgaa ctgtgggccc gcgccggcac attggccacc      300
gcgttgagcg ctgagcccgc gatccggccg ggcgaccggg tttgcgtgct gggcttcaac      360
agcgtcgact acacaaccat cgacatcgcg ctgatccggt tgggcgccgt gtcggttcca      420
ctgcagacca gtgcgccggt caccgggttg cccccgatcg tcaccgagac cgagccgacg      480
atgatcgcca ccagcatcga caatcttggc gacgccgtcg aagtgctggc cggtcacgcc      540
ccggcccggc tggctgtatt cgattaccac ggcaagggtg acaccaccg cgaggccgctc      600
gaagccgccc gagctcgggt ggccggctcg gtgaccatcg acacacttgc cgaactgatc      660
gaacgcggca gggcgctgcc ggccacacc attgccgaca gcgccgacga cgcgctggcg      720
ctgctgattt acacctcggg tagtaccggc gcaaccaaag gcgccatgta tcgcgagagc      780
caggtgatga gcttctggcg caagtcgagt ggctggttcg agccgagcgg ttaccctcgc      840
atcacgctga acttcatgcc gatgagccac gtcggggggc gtcaggtgct ctacgggacg      900
ctttccaacg gcggtaccgc ctacttcgct gccaaagagc acctgtcgac gctgttcgag      960
gacctcgccc tggtgccggc cacagaattg tgcttcgtgc cgcgcactctg ggacatgggtg     1020
ttcgcagagt tccacagcga ggtcgaccgc cgcttgggtg acggcgccga tcgagcggcg     1080
ctggaagcgc aggtgaaggc cgagctcgcg gagaacgtgc tcggcggacg gtttgcctatg     1140
gcgctgaccg gttccgcgcc gatctccgct gagatgacgg cgtgggtcga gtcctctgctg     1200
gccgacgtgc atttggtgga gggttacggc tccaccgagg ccgggatggt cctgaacgac     1260
ggcatggtgc ggcgccccgc ggtgatcgcg tacaagctgg tcgacgtgcc cgagctgggc     1320
tacttcggca ccgatcagcc ctaccccccg ggcgagctgc tggtaagac gcaaaccatg     1380
ttccccggct actaccagcg cccggatgct accgccgagg tgttcgacct cgacggcttc     1440
taccggaccg gggacatcat ggccaaagta ggccccgacc agttcgtcta cctcgaccgc     1500
cgcaacaacg tgctaaagct ctcccagggc gagttcatcg ccgtgtcgaa gctcgaggcg     1560
gtgttcggcg acagcccgct ggtccgacag atcttcatct acggcaacag tgccccggcc     1620
taccgctggc cggtggttgt cccgtccggg gacgcgcttt ctccgatgg catcgagaat     1680
ctcaagcccg tgatcagcga gtccctgcag gaggtagcga gggcgccgg cctgcaatcc     1740
tacgagattc cacgcgactt catcatcgaa accacgccgt tcaccctgga gaacggcctg     1800

```

ES 2 613 553 T3

ctcacggca tccgcaagct ggcacgccc cagttgaaga agttctatgg cgaacgtctc 1860  
gagcggctct ataccgagct ggcgatagc caatccaacg agctgcgcga gctgcggcaa 1920  
agcgggtccg atgcgcccgt gcttccgacg ctgtgcccgt ccgcggctgc gttgctgggc 1980  
tctaccgctg cggatgtgcg gccggacgcg cacttogccg acctgggtgg tgactcgctc 2040  
tcggcgctgt cgttgccaa cctgctgcac gagatcttcg gcgtcgacgt gccgggtgggt 2100  
gtcattgtca gcccggaag cgacctgcgg gccctggccg accacatcga agcagcgcgc 2160  
accggcgtca ggcgaccag cttcgcctcg atacacggtc gctccgcgac ggaagtgcac 2220  
gccagcgacc tcacgtgga caagttcatc gacgctgcca ccctggccgc agccccgaac 2280  
ctgccggcac cgagcggcca agtgcgcacc gtactgctga ccggcgccac cggctttttg 2340  
ggtcgctacc tggcgctgga atggctcgac cgcattggacc tggccaacgg caagctgatc 2400  
tgctgggtcc gcgccagatc cgacgaggaa gcacaagccc ggctggacgc gacgttcgat 2460  
agcggcgacc cgtatttggt gcggcactac cgcgaattgg gcgccggccg cctcgaggtg 2520  
ctcgccggcg acaaggcga ggcgacctg ggcctggacc gggtcacctg gcagcggcta 2580  
gccgacacgg tggacctgat cgtggacccc gcggccctgg tcaaccacgt gctgccgat 2640  
agccagctgt tcggccaaa cggggcgggc accgcccaggt tgcttcggct ggcgctgacc 2700  
ggcaagcgca agccatacat ctacacctcg acgatcgccg tggcgagca gatcccgcg 2760  
gaggcgttca ccgaggacgc cgacatccgg gccatcagcc cgacccgcag gatcgacgac 2820  
agctacgcca acggctacgc gaacagcaag tgggcccggcg aggtgctgct gcgcgaagct 2880  
cacgagcagt gcggcctgcc ggtgacggtc ttccgctgcg acatgatcct ggccgacacc 2940  
agctataccg gtcagctcaa cctgccggac atgttcaccc ggctgatgct gagcctggcc 3000  
gctaccggca tcgcacccgg ttcgttctat gagctggatg cgcacggcaa tcggcaacgc 3060  
gccactatg acggcttgcc ggtcgaattc gtcgcagaag ccatttgac ccttgggaca 3120  
catagcccgg accgtttgt cacctaccac gtgatgaacc cctacgacga cggcatcggg 3180  
ctggacgagt tcgtcgactg gctcaactcc ccaactagcg ggtccggtg cacgatccag 3240  
cggatcgccg actacggcga gtggctgcag cggttcgaga cttcgtgcg tgccttgccg 3300  
gatcgccagc gccacgcctc gctgctgcc ttgctgcaca actaccgaga gcctgcaaag 3360  
ccgatatgcg ggtcaatcgc gccaccgac cagttccgcg ctgccgtcca agaagcga 3420  
atcgggtccg acaagacat tccgcacctc acggcggcga tcatcgcaa gtacatcagc 3480  
aacctgcgac tgctcgggt gctgtga 3507

<210> 43  
<211> 3522  
<212> ADN

ES 2 613 553 T3

<213> *Mycobacterium smegmatis*

<400> 43

atgaccagcg atgttcacga cgccacagac ggcgtcaccg aaaccgcaact cgacgacgag	60
cagtcgaccc gccgcatcgc cgagctgtac gccaccgatc ccgagttcgc cgccgccgca	120
ccgttgcccc ccgtggtcga cgcggcgcac aaaccggggc tgcggctggc agagatcctg	180
cagaccctgt tcaccggcta cggtgaccgc ccggcgctgg gataccgcbc ccgtgaactg	240
gccaccgacg agggcgggcg caccgtgacg cgtctgctgc cgcggttcga caccctcacc	300
tacgcccagg tgtggtcgcg cgtgcaagcg gtgcgcgcbg ccctgcgcca caacttcgcb	360
cagccgatct accccggcga cgccgtcgcg acgatcgggt tcgcgagtc ccgattacctg	420
acgctggatc tcgtatgcgc ctacctgggc ctcgtgagtg ttccgctgca gcacaacgca	480
ccggtcagcc ggctcgcccc gatcctggcc gaggtcgaac cgcggatcct caccgtgagc	540
gccgaatacc tcgacctcgc agtcgaatcc gtgcgggacg tcaactcggg gtcgcagctc	600
gtggtgttcg accatcacc ccgaggtcgc gaccaccgcb acgcaactgg ccgcbgcbgt	660
gaacaactcg ccggcaaggg catcgcctgc accaccctgg acgcbatcgc cgacgagggc	720
gccgggctgc cggccgaacc gatctacacc gccgaccatg atcagcbct cgcbatgatc	780
ctgtacacct cgggttcacc cggcgcaccc aagggtgcga tgtacaccga ggcgatggtg	840
gcgcbgctgt ggaccatgcb gttcatcacg ggtgacccca cgcggtcbat caactcabc	900
ttcatgcbgc tcaaccacct gggcgggcbgc atccccattt ccaccgcbgt gcagaacggt	960
ggaaccagtt acttcgtacc ggaatcbgac atgtccacgc tgttcgagga tctcgcgctg	1020
gtgcgcbcbga ccgaactcgg cctggttcgg cgcgtcgcgg acatgctcta ccagaccac	1080
ctcgcacccg tcgaccgcbt ggtcacgcbg ggcgcbgacg aactgaccgc cbgagaagcbg	1140
gccggtgcbg aactgcbtga gcaggtgctc ggcggacgcb tgatcaccgg attcgtcagc	1200
accgcbaccgc tggcbgcbga gatgagggcb ttctcgcaca tcaccctggg cgcacacatc	1260
gtcgcagcbgt acgggctcac cbgagaccgcb gccgtgacac gcgacggtgt gatcgtcgbg	1320
ccaccggtga tcgactaca gctgatcgc gttcccbgac tcggtactt cbgaccgcb	1380
aagcbctacc cgcgtgcbga actgctggtc aggtcgbcaa cgtgactcb cgggtactac	1440
aagcbcbcbg aggtcaccgc gagcbtctc gaccgggacg gctactacca caccgcbgac	1500
gtcbatggcbg agaccgcbcc cbgaccacctg gtgtacgtgg accgtcgbaa caacgtcbtc	1560
aaactcgcgc agggcbgagtt cbtggcbggtc gccaacctgg aggcggtgtt ctcbgcbgcbg	1620
gcgctggtgc gccagatcbt cbtgtacgcb aacagcbgac gcagtttcbt tctggcbgtg	1680
gtggtcbcbga cbcbggaggb gctcgbgcbg tacgatcbgg cbgcbgtcaa ggcgcbgtg	1740
gccgactcgc tgcagcbgcb cbgacgcbgac gccgaactgc aatcbtcbga ggtgcbcbgcb	1800

ES 2 613 553 T3

gatttcatcg tcgagaccga gccgttcagc gccgccaacg ggctgctgtc ggggtgtcggg 1860  
aaactgctgc ggccaacact caaagaccgc tacgggcagc gcctggagca gatgtacgcc 1920  
gatatacggg ccacgcaggc caaccagttg cgcgaactgc ggcgcgcggc cgccacacaa 1980  
ccggtgatcg acaccctcac ccaggccgct gccacgatcc tcggcaccgg gagcgaggtg 2040  
gcatccgacg cccacttcac cgacctgggc ggggattccc tgtcggcgct gacactttcg 2100  
aacctgctga gcgatttctt cggtttcgaa gttcccgtcg gcaccatcgt gaaccgggcc 2160  
accaacctcg cccaactcgc ccagcacatc gaggcgcagc gcaccgcggg tgaccgcagg 2220  
ccgagtttca ccaccgtgca cggcgcggac gccaccgaga tccgggcgag tgagctgacc 2280  
ctggacaagt tcatcgacgc cgaaacgctc cgggccgcac cgggtctgcc caaggtcacc 2340  
accgagccac ggacggtgtt gctctcgggc gccaacggct ggctgggccg gttcctcacg 2400  
ttgcagtggc tggaacgcct ggcacctgtc ggcggcacc ccatcacgat cgtgcggggc 2460  
cgcgacgacg ccgcggcccg cgcacggctg acccaggcct acgacaccga tcccagttg 2520  
tcccgcgct tcgccgagct ggccgaccgc cacctcgggg tggtcgccgg tgacatcggc 2580  
gaccgaatc tgggcctcac acccgagatc tggcaccggc tcgccccga ggtcgacctg 2640  
gtggtgcatc cggcagcgct ggtcaaccac gtgctcccct accggcagct gttcggcccc 2700  
aacgtcgtgg gcacggccga ggtgatcaag ctggccctca ccgaacggat caagcccgtc 2760  
acgtacctgt ccaccgtgtc ggtggccatg gggatccccg acttcgagga ggacggcgac 2820  
atccggaccg tgagcccggg gcgcccgtc gacggcggat acgccaacgg ctacggcaac 2880  
agcaagtggg ccggcgaggt gctgctgcgg gaggcccacg atctgtgcgg gctgcccgtg 2940  
gcgacgttcc gctcggacat gatcctggcg catccgcgct accgcggtca ggtcaacgtg 3000  
ccagacatgt tcacgcgact cctggtgagc ctcttgatca ccggcgtcgc gccgcggtcg 3060  
ttctacatcg gagacggtga gcgcccgcgg gcgcactacc ccggcctgac ggtcgatttc 3120  
gtggccgagg cggtcacgac gctcggcgcg cagcagcgcg agggatacgt gtcctacgac 3180  
gtgatgaacc cgcacgacga cgggatctcc ctggatgtgt tcgtggactg gctgatccgg 3240  
gcgggccatc cgatcgaccg ggtcgacgac tacgacgact gggcgcgtcg gttcgagacc 3300  
gcgttgaccg cgcttcccga gaagcgccgc gcacagaccg tactgccgct gctgcacgcg 3360  
ttccgcgctc cgcaggcacc gttgcgcggc gcacccgaac ccacggaggt gttccacgcc 3420  
gcggtgcgca ccgcgaaggt gggcccggga gacatcccgc acctcgacga ggcgctgatc 3480  
gacaagtaca tacgcgatct gcgtgagttc ggtctgatct ga 3522

<210> 44  
<211> 3582  
<212> ADN  
<213> Secuencia artificial

ES 2 613 553 T3

<220> fuente  
 <221> /nota="Descripción de secuencia artificial: Polinucleótido  
 <223> sintético"

<400> 44  
 atgggcagca gccatcatca tcatcatcac agcagcggcc tggtgccgcg cggcagccat 60  
 atgacgagcg atgttcacga cgcgaccgac ggcgttaccg agactgcaact ggatgatgag 120  
 cagagcaactc gtcgtattgc agaactgtac gcaacggacc cagagttcgc agcagcagct 180  
 cctctgccgg ccggttgcga tgcggcgcac aaaccgggcc tgcgtctggc ggaaatcctg 240  
 cagaccctgt tcaccggcta cggcgatcgt ccggcgctgg gctatcgtgc acgtgagctg 300  
 gcgacggagc aaggcggtcg tacggtcacg cgtctgctgc cgcgcttcga taccctgacc 360  
 tatgcacagg tgtggagccg tgttcaagca gtggctgcag cgttgctca caatttcgca 420  
 caaccgattt acccgggca cgcggtcgcg actatcggct ttgcgagccc ggactatttg 480  
 acgtggatc tgggtgctgc gtatctgggc ctggctcagc ttcctttgca gcataacgct 540  
 ccggtgtctc gcctggcccc gattctggcc gaggtggaac cgcgtattct gacgggtgagc 600  
 gcagaatacc tggacctggc ggttgaatcc gtccgtgatg tgaactccgt cagccagctg 660  
 gttgttttgc accatcatcc ggaagtggac gatcaccgtg acgcaactggc tcgcgcacgc 720  
 gagcagctgg ccggcaaagg tatcgcagtt acgaccctgg atgcgatcgc agacgaaggc 780  
 gcaggtttgc cggctgagcc gatttacacg gcggatcacg atcagcgtct ggccatgatt 840  
 ctgtatacca gcggctctac gggtgctccg aaaggcgcga tgtacaccga agcgtggtg 900  
 gctcgcctgt ggactatgag ctttatcacg ggcgacccga ccccggttat caacgtgaac 960  
 ttcattgccgc tgaaccatct gggcggctgt atcccgatta gcaccgccgt gcagaatggc 1020  
 ggtaccagct acttcgttcc ggaaagcgac atgagcaacg tgtttgagga tctggccctg 1080  
 gtccgcccta ccgaactggg tctggtgccg cgtggtgagg acatgctgta ccagcatcat 1140  
 ctggcgaccg tggatcgctt ggtgaccag ggcgaggacg aactgactgc ggaaaagcag 1200  
 gccggtgagg aactgcgtga acaggtcttg ggcggtcgtg ttatcaccgg ttttgtttcc 1260  
 accgcgccgt tggcggcaga gatgcgtgct tttctggata tcacctggg tgacacatc 1320  
 gttgacggtt acggtctgac cgaaaccggt gcggtcacc gtgatggtgt gattgttctg 1380  
 cctccggtca ttgattaaa gctgatcgt gtgccggagc tgggttactt ctccaccgac 1440  
 aaaccgtacc cgcgtggcga gctgctggtt cgtagccaaa cgttgactcc gggttactac 1500  
 aagcggccag aagtcaccgc gtccgttttc gatcgcgacg gctattacca caccggcgac 1560  
 gtgatggcag aaaccgcgcc agaccacctg gtgtatgtgg accgcgcaa caatgttctg 1620  
 aagctggcgc aaggtgaatt tgtcggcgtg gctaacctgg aggccgtttt cagcggcgtc 1680

ES 2 613 553 T3

gctctggtcc gccagattht cgtgtatggt aacagcgcgc gcagctttct gttggctggt 1740  
gttgtcccta ccccggaggc gctggagcaa tacgaccctg ccgcattgaa agcagccctg 1800  
gcggattcgc tgcagcgtac ggcgcgtgat gccgagctgc agagctatga agtgccggcg 1860  
gacttcattg ttgagactga gccttttagc gctgcgaacg gtctgctgag cgggtgtggc 1920  
aagttgctgc gtccgaatth gaaggatcgc tacggtcagc gtttgagca gatgtacgcg 1980  
gacatcgcgg ctacgcaggc gaaccaatth cgtgaactgc gccgtgctgc ggctactcaa 2040  
ccggtgatcg acacgctgac gcaagctgcg gcgaccatcc tgggtaccgg cagcgaggtt 2100  
gcaagcgaag cacactthac tgatttgggc ggtgattctc tgagcgcgct gacgttgagc 2160  
aacttgctgt ctgacttctt tggctttgaa gtcccggttg gcacgattgt taaccagcg 2220  
actaatctgg cacagctggc gcaacatath gaggcgcagc gcacggcggg tgaccgccgt 2280  
ccatccttha cgacggthca cggtgccgat gctacggaaa tccgtgcaag cgaactgact 2340  
ctggacaaat tcatcgacgc tgagactctg cgcgcagcac ctggtttgcc gaaggttacg 2400  
actgagccgc gtacggthct gttgagcggg gccaatggtt ggttgggccc cttcctgacc 2460  
ctgcagtggc tggaacgtht ggcaccggth ggcggtacc thgatccat tgtgcgcggt 2520  
cgtgacgatg cagcggcacg tgcacgtht actcaggctt acgatacggc cccagagctg 2580  
tcccgcgct tcgctgagth ggcggatcgc cacttgctg tggtggcagg tgatathggc 2640  
gatccgaath tgggcctgac cccggagath tggcaccgth tggcagcaga ggtcgathctg 2700  
gtcgtthath cagcggccct ggtcaaccac gthctgctg accgccagct gthtggthccg 2760  
aatgtgtgtg gcaccgccga agthathcaag ttggtctgca ccgagcgcath caagcctgth 2820  
acctacctgt ccacggttag cgtcgcgatg ggtathctg atthtgaggc ggacggtgac 2880  
atthctaccg thagcccggth thgtccgctg gatggtggct atgcaaatgg ctathggcaac 2940  
agcaagtggg ctggcgaggt gctgctgcgc gaggcacath acctgtgtgg cctgccggtt 3000  
gcgagctthc gtagcgacath gattctggcc caccgcgct accgtggcca agtgaathgtg 3060  
ccggacathg thaccgthct gctgctgthc thgtgatca cgggtgtggc accgcgthcc 3120  
thctacathg gtgathggca gcgtccgctg gcacathacc cgggcctgac cgtcgaththt 3180  
gthtgcggaag cggthactac cctgggtgct cagcaacgth agggthathg thctgathgac 3240  
gthathgaath cgcacgathg cggthathg thtgathgth thgtggactg gctgathctg 3300  
gcgggccacc caathgaccg tgttgacgac thathgact ggtgctgctg ththgaaacc 3360  
gcgtgaccg cthtgccgga gaaacgthc gcgcagaccg thctgccgct gctgathgthc 3420  
thctgcgcgc cacaggcgc gthtgcgthg gccctgaa cgaccgaagth gththathgca 3480  
gcggtgcgth ccgthaaagth cggthccggth gatathctgc acctgathg agcctgathc 3540  
gacaagthaca thctgathct gcgcgagthc gthctgathth ag 3582

<210> 45  
 <211> 66  
 <212> ADN  
 <213> Secuencia artificial  
  
 <220>  
 <221> fuente  
 <223> /nota="Descripción de secuencia artificial:Cebador  
 sintético"  
  
 <400> 45  
 cggttctggc aaatattctg aatgagctg ttgacaatta atcaaatccg gctcgtataa 60  
  
 tgtgtg 66  
  
 <210> 46  
 <211> 30  
 <212> ADN  
 <213> Secuencia artificial  
  
 <220>  
 <221> fuente  
 <223> /nota="Descripción de secuencia artificial:Cebador  
 sintético"  
  
 <400> 46  
 ggtttattcc tccttattta atcgatacat 30  
  
 <210> 47  
 <211> 59  
 <212> ADN  
 <213> Secuencia artificial  
  
 <220>  
 <221> fuente  
 <223> /nota="Descripción de secuencia artificial:Cebador  
 sintético"  
  
 <400> 47  
 atgtatcgat taaataagga ggaataaacc atggggcacga gcgatgttca cgacgcgac 59  
  
 <210> 48  
 <211> 59  
 <212> ADN  
 <213> Secuencia artificial  
  
 <220>  
 <221> fuente  
 <223> /nota="Descripción de secuencia artificial:Cebador  
 sintético"  
  
 <400> 48  
 atgtatcgat taaataagga ggaataaacc gtggggcacga gcgatgttca cgacgcgac 59  
  
 <210> 49  
 <211> 59  
 <212> ADN

<213> Secuencia artificial

<220>  
 <221> fuente  
 <223> /nota="Descripción de secuencia artificial:Cebador sintético"

<400> 49  
 atgtatcgat taaataagga ggaataaacc ttgggcacga gcgatgttca cgacgcgac 59

<210> 50  
 <211> 26  
 <212> ADN  
 <213> Secuencia artificial

<220>  
 <221> fuente  
 <223> /nota="Descripción de secuencia artificial:Cebador sintético"

<400> 50  
 ttctaaatca gaccgaactc gcgcag 26

<210> 51  
 <211> 48  
 <212> ADN  
 <213> Secuencia artificial

<220>  
 <221> fuente  
 <223> /nota="Descripción de secuencia artificial:Cebador sintético"

<400> 51  
 ctgcgcgagt tcggtctgat ttagaattcc tcgaggatgg tagtgtgg 48

<210> 52  
 <211> 27  
 <212> ADN  
 <213> Secuencia artificial

<220>  
 <221> fuente  
 <223> /nota="Descripción de secuencia artificial:Cebador sintético"

<400> 52  
 cagtcgacat acgaaacggg aatgcgg 27

<210> 53  
 <211> 56  
 <212> ADN  
 <213> Secuencia artificial

<220>  
 <221> fuente  
 <223> /nota="Descripción de secuencia artificial:Cebador sintético"

ES 2 613 553 T3

<400> 53  
ccgcattccc gtttcgtatg tcgactgaaa cctcaggcat tgagaagcac acggtc 56

<210> 54  
<211> 56  
<212> ADN  
<213> Secuencia artificial

<220>  
<221> fuente  
<223> /nota="Descripción de secuencia artificial:Cebador  
sintético"

<400> 54  
ctcatttcag aatatttgcc agaaccgta atttcctaat gcaggagtcg cataag 56

<210> 55  
<211> 20  
<212> ADN  
<213> Secuencia artificial

<220>  
<221> fuente  
<223> /nota="Descripción de secuencia artificial:Cebador  
sintético"

<400> 55  
ggatctcgac gctctccctt 20

<210> 56  
<211> 28  
<212> ADN  
<213> Secuencia artificial

<220>  
<221> fuente  
<223> /nota="Descripción de secuencia artificial:Cebador  
sintético"

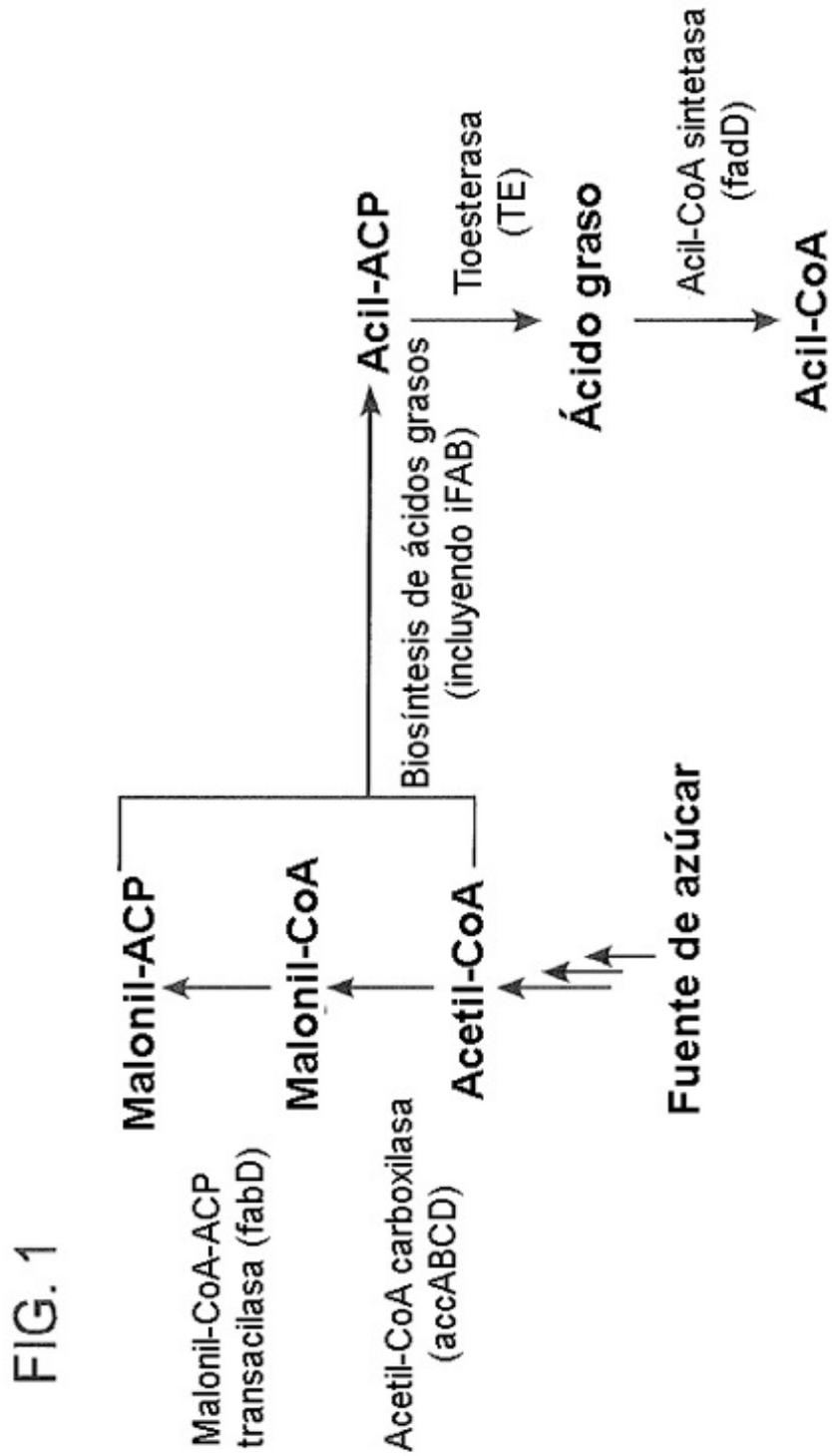
<400> 56  
tcaaaaacgc cattaacctg atgttctg 28

**REIVINDICACIONES**

1. Variante de polipéptido de ácido carboxílico reductasa (CAR) que comprende una secuencia de aminoácidos que tiene una identidad de secuencia de al menos el 90% con el polipéptido de CAR silvestre de SEQ ID NO: 7, en la que dicha variante de polipéptido de CAR se modifica por ingeniería genética para que tenga al menos una mutación en una posición de aminoácido seleccionada del grupo que consiste en S3R, D18R, D18L, D18T, D18P, E20V, E20S, E20R, S22R, S22N, S22G, L80R, R87G, R87E, V191S, F288R, F288S, F288G, Q473L, Q473W, Q473Y, Q473I, Q473H, A535S, D750A, R827C, R827A, I870L, R873S, V926A, V926E, S927K, S927G, M930K, M930R y L1128W, en la que la expresión de dicho polipéptido de CAR variante en una célula huésped recombinante da como resultado un título superior de composiciones de alcoholes grasos en comparación con una célula huésped recombinante que expresa el polipéptido de CAR silvestre de SEQ ID NO: 7.
2. Polipéptido de CAR variante según la reivindicación 1, en el que dicho polipéptido variante comprende:
  - (a) la mutación A535S;
  - (b) las mutaciones E20R, F288G, Q473I y A535S;
  - (c) las mutaciones E20R, F288G, Q473H, A535S, R827A y S927G;
  - (d) las mutaciones E20R, S22R, F288G, Q473H, A535S, R827A y S927G;
  - (e) las mutaciones S3R, E20R, S22R, F288G, Q473H, A535S, R873S, S927G, M930R y L1128W;
  - (f) las mutaciones E20R, S22R, F288G, Q473H, A535S, R873S, S927G, M930R y L1128W;
  - (g) las mutaciones D18R, E20R, S22R, F288G, Q473I, A535S, S927G, M930K y L1128W;
  - (h) las mutaciones E20R, S22R, F288G, Q473I, A535S, R827C, V926E, S927K y M930R;
  - (i) las mutaciones D18R, E20R, F288G, Q473I, A535S, R827C, V926E, M930K y L1128W;
  - (j) las mutaciones E20R, S22R, F288G, Q473H, A535S, R827C, V926A, S927K y M930R;
  - (k) las mutaciones E20R, S22R, F288G, Q473H, A535S y R827C;
  - (l) las mutaciones E20R, S22R, F288G, Q473I, A535S, R827C y M930R;
  - (m) las mutaciones E20R, S22R, F288G, Q473I, A535S, I870L, S927G y M930R;
  - (n) las mutaciones E20R, S22R, F288G, Q473I, A535S, I870L y S927G;
  - (o) las mutaciones D18R, E20R, S22R, F288G, Q473I, A535S, R827C, I870L, V926A y S927G;
  - (p) las mutaciones E20R, S22R, F288G, Q473H, A535S, R827C, I870L y L1128W;
  - (q) las mutaciones D18R, E20R, S22R, F288G, Q473H, A535S, R827C, I870L, S927G y L1128W;
  - (r) las mutaciones E20R, S22R, F288G, Q473I, A535S, R827C, I870L, S927G y L1128W;
  - (s) las mutaciones E20R, S22R, F288G, Q473I, A535S, R827C, I870L, S927G, M930K y L1128W;
  - (t) las mutaciones E20R, S22R, F288G, Q473H, A535S, I870L, S927G y M930K;
  - (u) las mutaciones E20R, F288G, Q473I, A535S, I870L, M930K;
  - (v) las mutaciones E20R, S22R, F288G, Q473H, A535S, S927G, M930K y L1128W;
  - (w) las mutaciones D18R, E20R, S22R, F288G, Q473I, A535S, S927G y L1128W;
  - (x) las mutaciones E20R, S22R, F288G, Q473I, A535S, R827C, I870L y S927G;
  - (y) las mutaciones D18R, E20R, S22R, F288G, Q473I, A535S, R827C, I870L, S927G y L1128W;
  - (z) las mutaciones D18R, E20R, S22R, F288G, Q473I, A535S, S927G, M930R y L1128W;
  - (xx) las mutaciones E20R, S22R, F288G, Q473H, A535S, V926E, S927G y M930R; o
  - (zz) las mutaciones E20R, S22R, F288G, Q473H, A535S, R827C, I870L, V926A y L1128W.
3. Célula huésped recombinante que comprende una secuencia de polinucleótido que codifica para un

- 5 polipéptido de ácido carboxílico reductasa (CAR) variante que tiene una identidad de secuencia de al menos el 90% con el polipéptido de CAR silvestre de SEQ ID NO: 7 y que tiene al menos una mutación en una posición de aminoácido seleccionada del grupo que consiste en S3R, D18R, D18L, D18T, D18P, E20V, E20S, E20R, S22R, S22N, S22G, L80R, R87G, R87E, V191S, F288R, F288S, F288G, Q473L, Q473W, Q473Y, Q473I, Q473H, A535S, D750A, R827C, R827A, I870L, R873S, V926A, V926E, S927K, S927G, M930K, M930R y L1128W, en la que la célula huésped recombinante produce una composición de alcohol graso a un título o rendimiento mayor que una célula huésped que expresa el polipéptido de CAR silvestre de SEQ ID NO: 7 cuando se cultiva en un medio que contiene una fuente de carbono en condiciones eficaces para expresar dicho polipéptido de CAR variante.
- 10 4. Célula huésped recombinante según la reivindicación 3, que comprende además un polinucleótido que codifica para un polipéptido de tioesterasa.
5. Célula huésped recombinante según la reivindicación 4, que comprende además un polinucleótido que codifica para
- 15 (i) un polipéptido de FabB y un polipéptido de FadR, o
- (ii) una aldehído graso reductasa.
6. Célula huésped recombinante según una cualquiera de las reivindicaciones 3-5, en la que dicha célula huésped recombinante tiene un título o un rendimiento que es al menos 3 veces mayor que el título o el rendimiento de una célula huésped que expresa el polipéptido de CAR silvestre correspondiente cuando se cultiva en las mismas condiciones que la célula huésped recombinante.
- 20 7. Célula huésped recombinante según la reivindicación 6, en la que dicha célula huésped recombinante tiene un título de desde 30 g/l hasta 250 g/l, o un título de desde 90 g/l hasta 120 g/l.
8. Célula huésped recombinante según la reivindicación 6, en la que dicha célula huésped recombinante tiene un rendimiento de desde el 10% hasta el 40%.
- 25 9. Cultivo celular que comprende la célula huésped recombinante según una cualquiera de las reivindicaciones 3-5.
10. Cultivo celular según la reivindicación 9, en el que dicho cultivo celular tiene una productividad que es al menos 3 veces mayor que la productividad de un cultivo celular que expresa el polipéptido de CAR silvestre correspondiente.
11. Cultivo celular según la reivindicación 10, en el que dicha productividad oscila entre 0,7 mg/l/h y 3 g/l/h.
- 30 12. Cultivo celular según la reivindicación 11, en el que el medio de cultivo comprende una composición de alcohol graso.
13. Célula huésped recombinante según una cualquiera de las reivindicaciones 3-12, en la que la composición de alcohol graso se recoge en una fase orgánica extracelularmente.
- 35 14. Cultivo celular según la reivindicación 13, en el que la composición de alcohol graso comprende uno o más de un alcohol graso C6, C8, C10, C12, C13, C14, C15, C16, C17 o C18.
15. Cultivo celular según la reivindicación 13, en el que la composición de alcohol graso comprende un alcohol graso insaturado C10:1, C12:1, C14:1, C16:1 o C18:1.
16. Cultivo celular según la reivindicación 13, en el que la composición de alcohol graso comprende alcoholes grasos C12 y C14.
- 40 17. Cultivo celular según la reivindicación 16, en el que la composición de alcohol graso comprende alcoholes grasos C12 y C14 a una razón de aproximadamente 3:1.
18. Cultivo celular según la reivindicación 13, en el que la composición de alcohol graso comprende alcoholes grasos insaturados o alcoholes grasos saturados.
- 45 19. Cultivo celular según la reivindicación 18, en el que la composición de alcohol graso comprende un alcohol graso insaturado que tiene un doble enlace en la posición 7 en la cadena de carbono entre C7 y C8 del extremo reducido del alcohol graso.
20. Cultivo celular según la reivindicación 13, en el que la composición de alcohol graso comprende alcoholes grasos de cadena ramificada.
- 50 21. Método de preparación de una composición de alcohol graso a un título, rendimiento o productividad superior, que comprende las etapas de:

- (a) modificar por ingeniería genética una célula huésped recombinante según una cualquiera de las reivindicaciones 3-8;
- (b) cultivar dicha célula huésped recombinante en un medio que comprende una fuente de carbono; y
- (c) opcionalmente aislar dicha composición de alcohol graso de dicho medio.



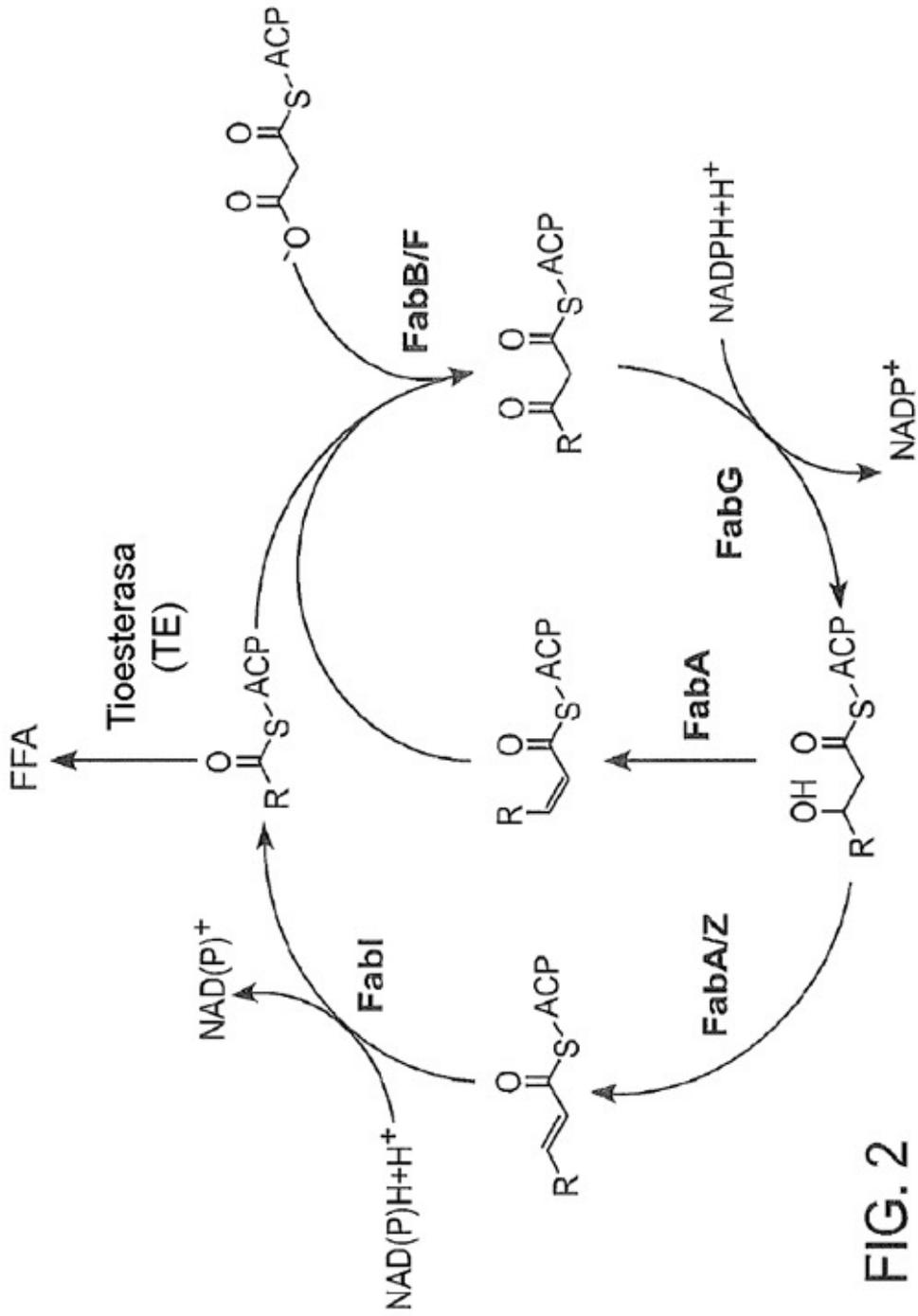


FIG. 2

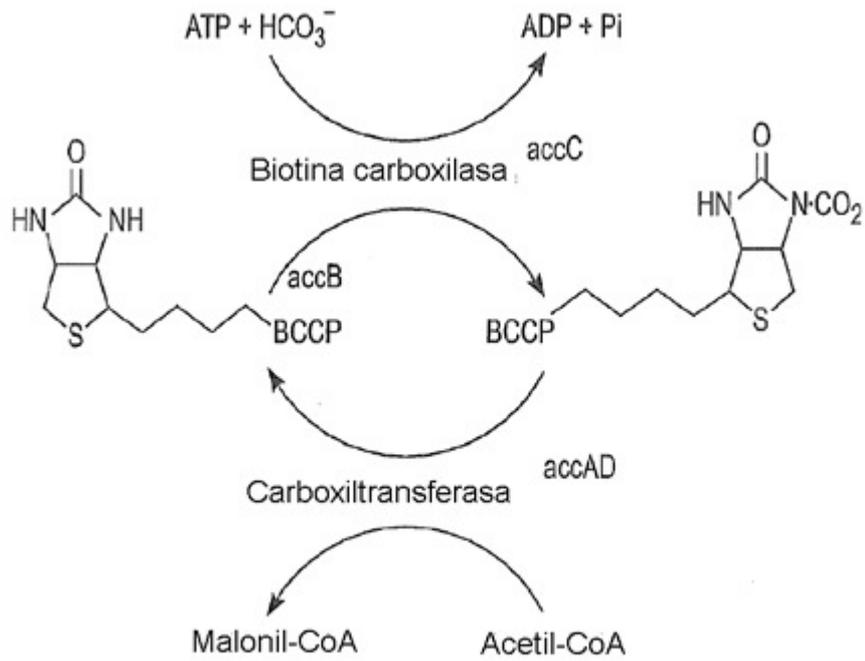


FIG. 3

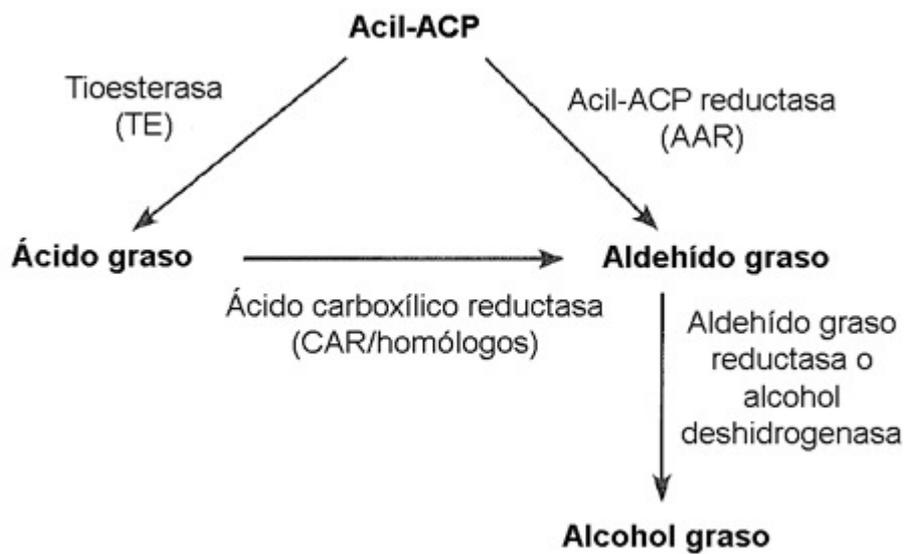


FIG. 4

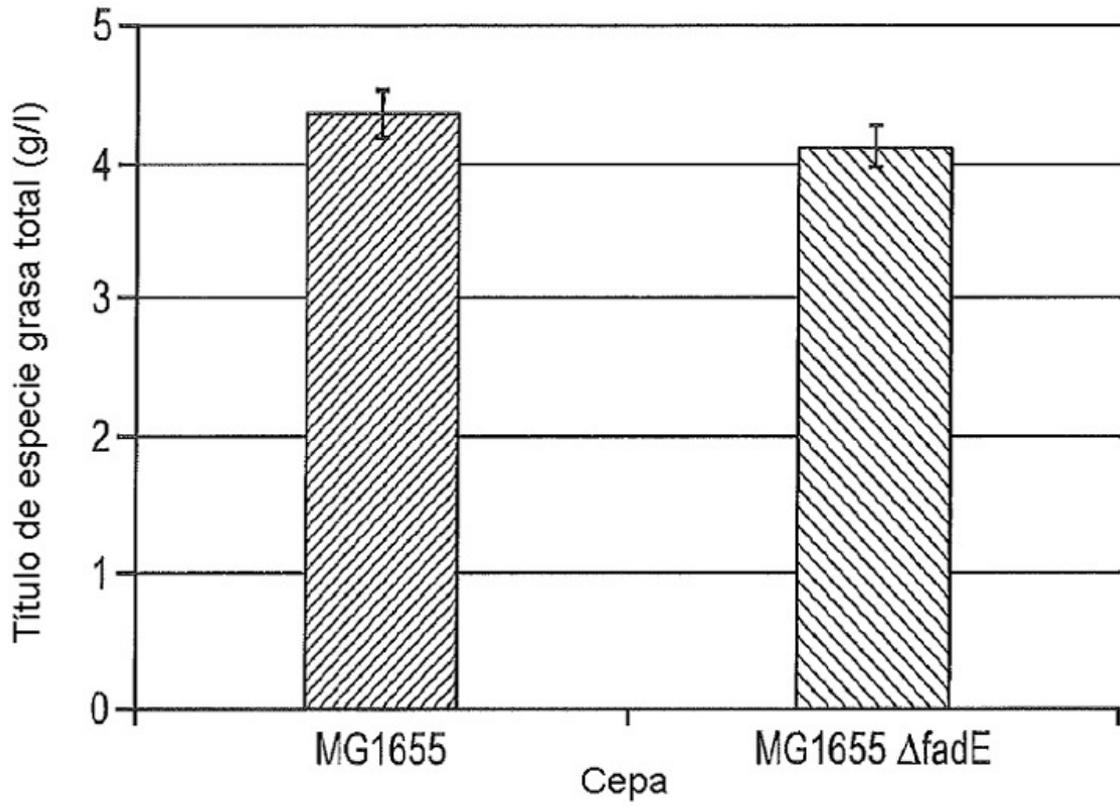
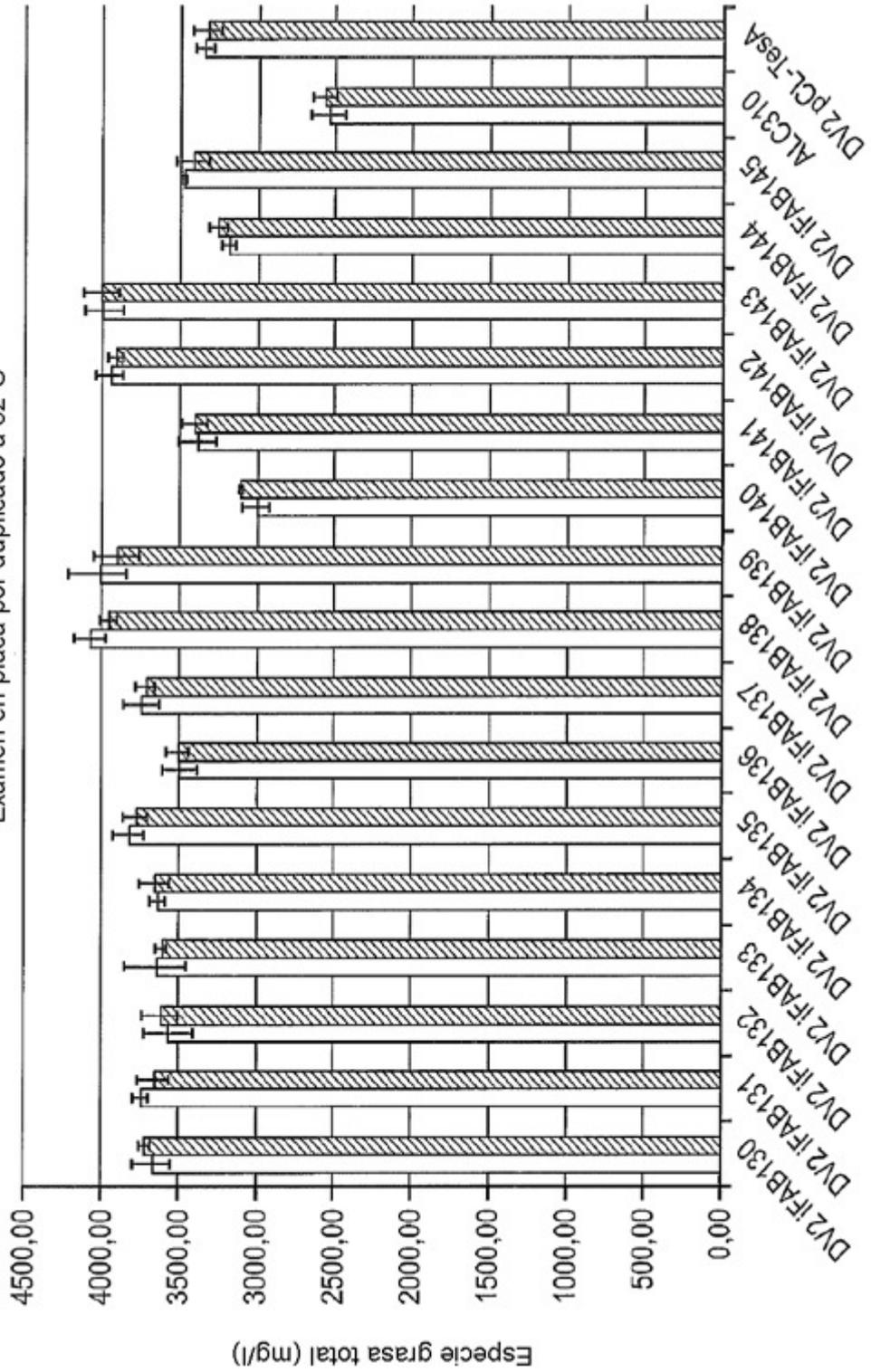
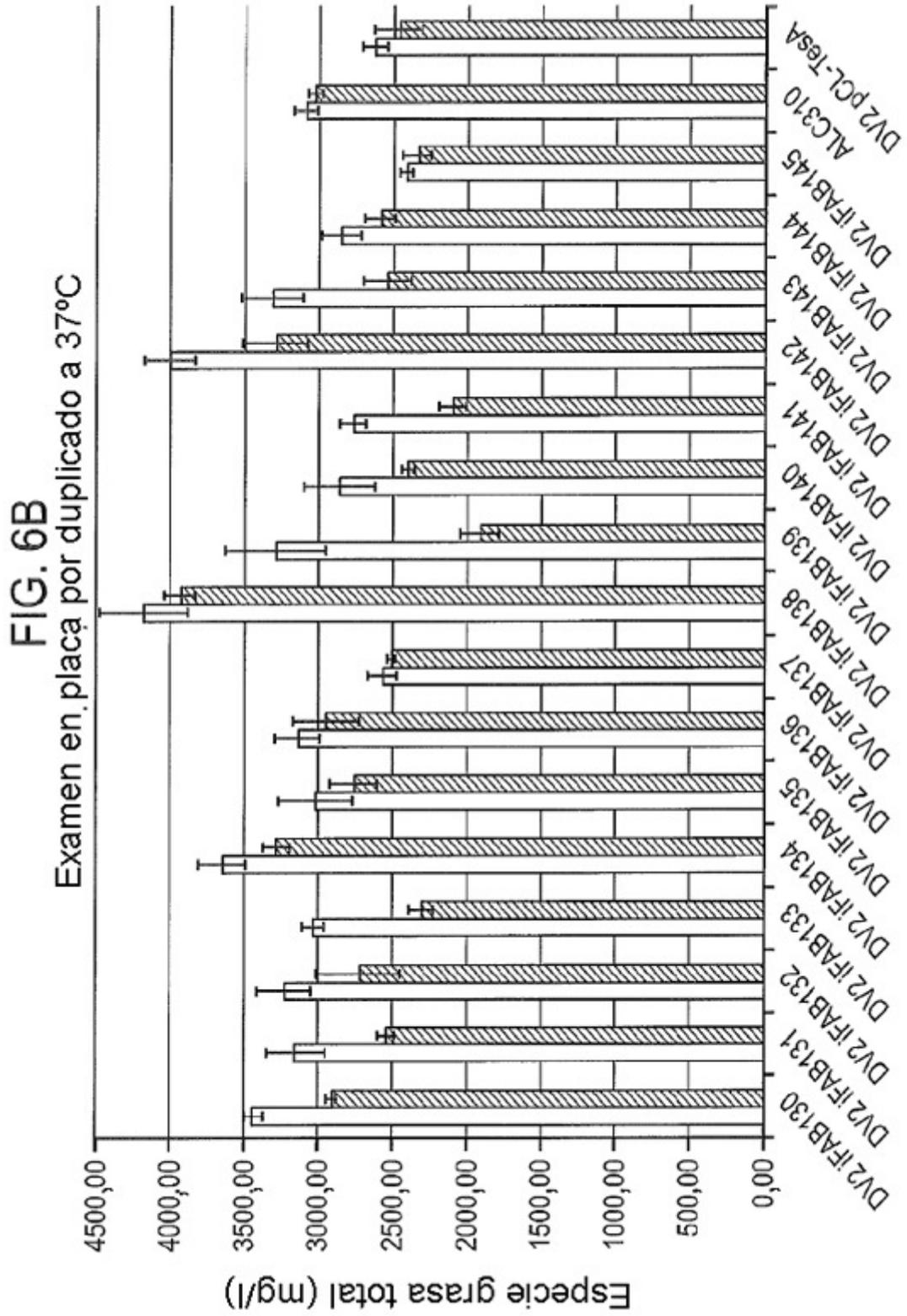


FIG. 5

FIG. 6A

Examen en placa por duplicado a 32°C





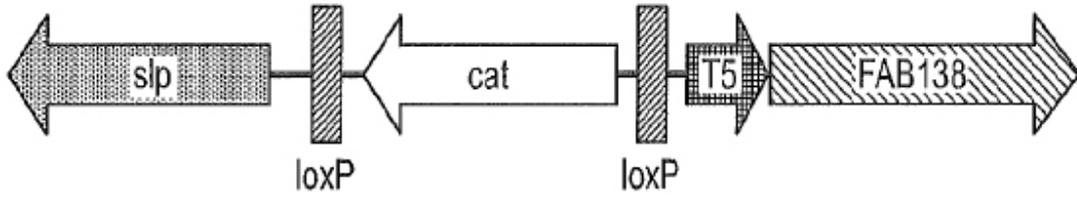


FIG. 7A

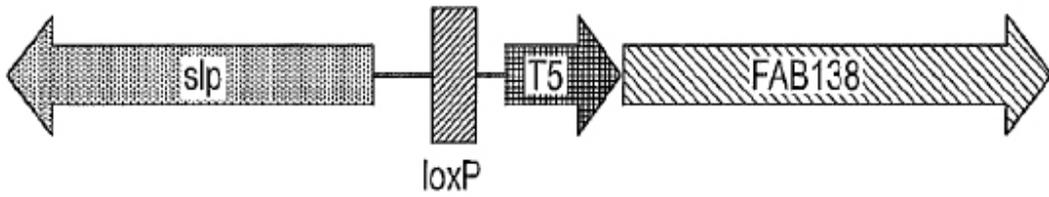


FIG. 7B

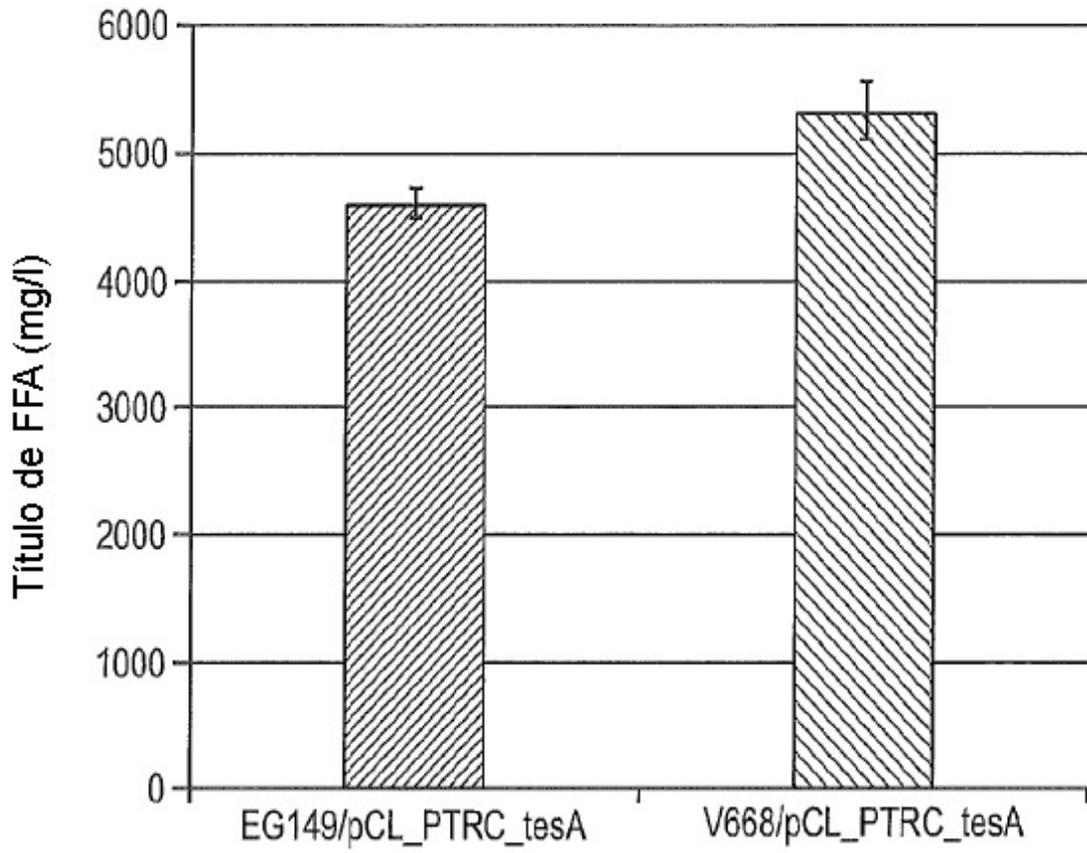
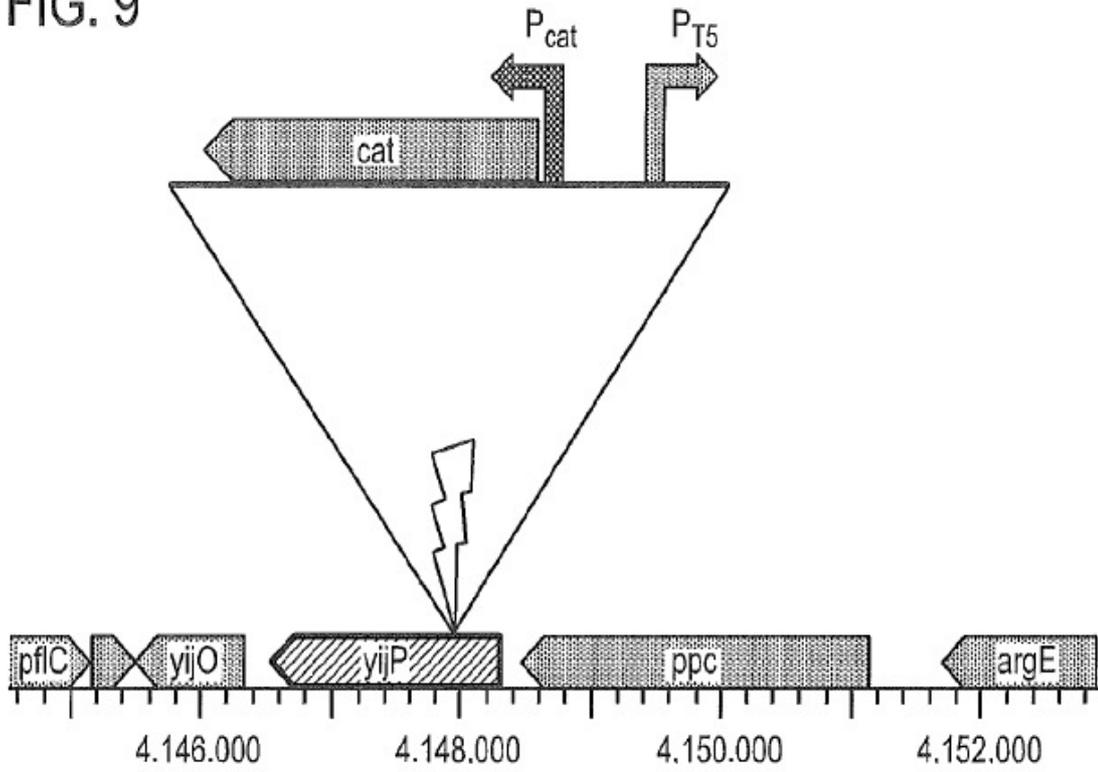


FIG. 8

FIG. 9



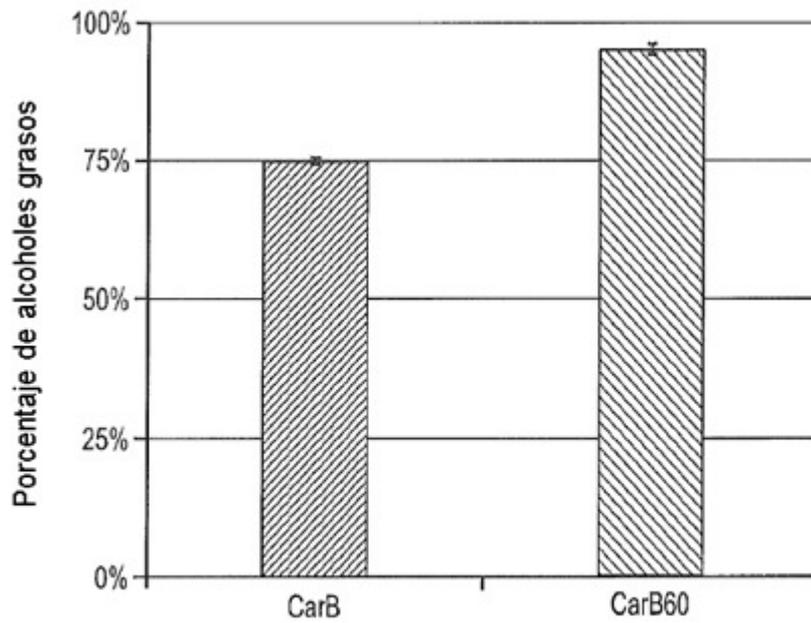


FIG. 10

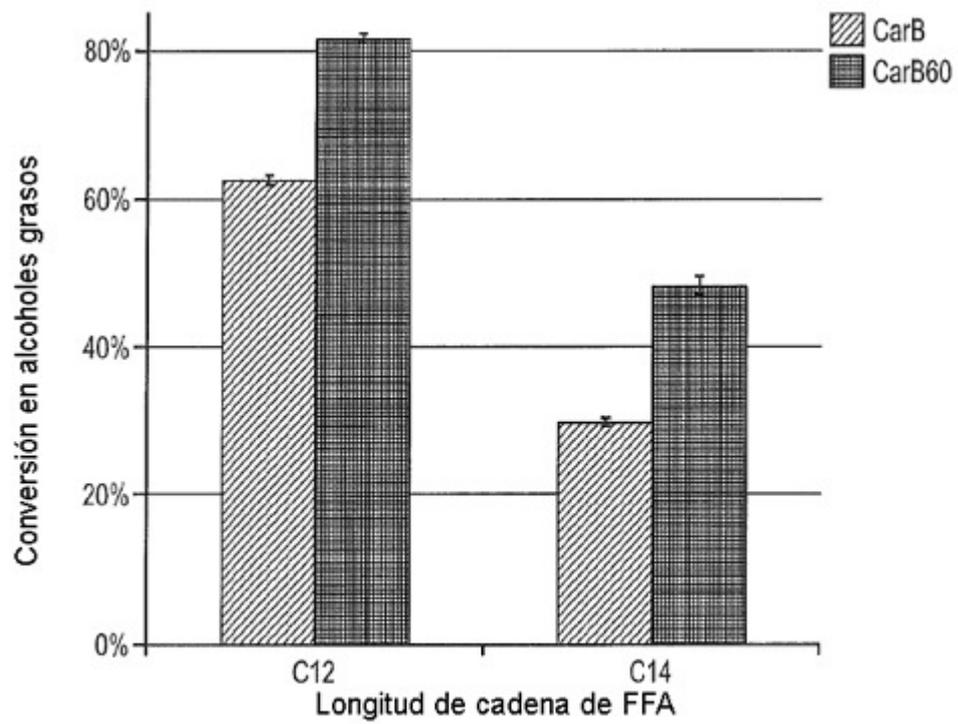
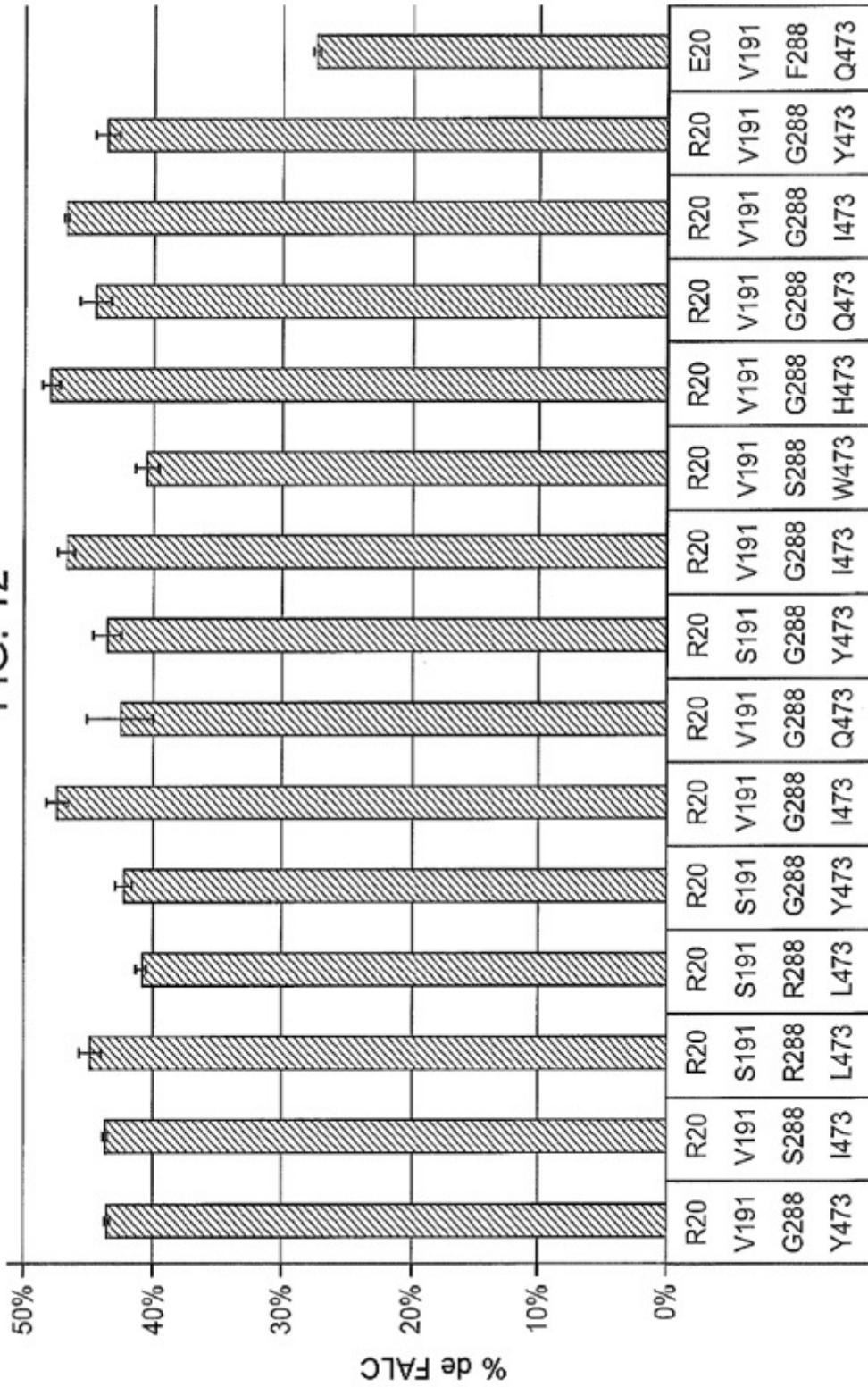


FIG. 11

FIG. 12



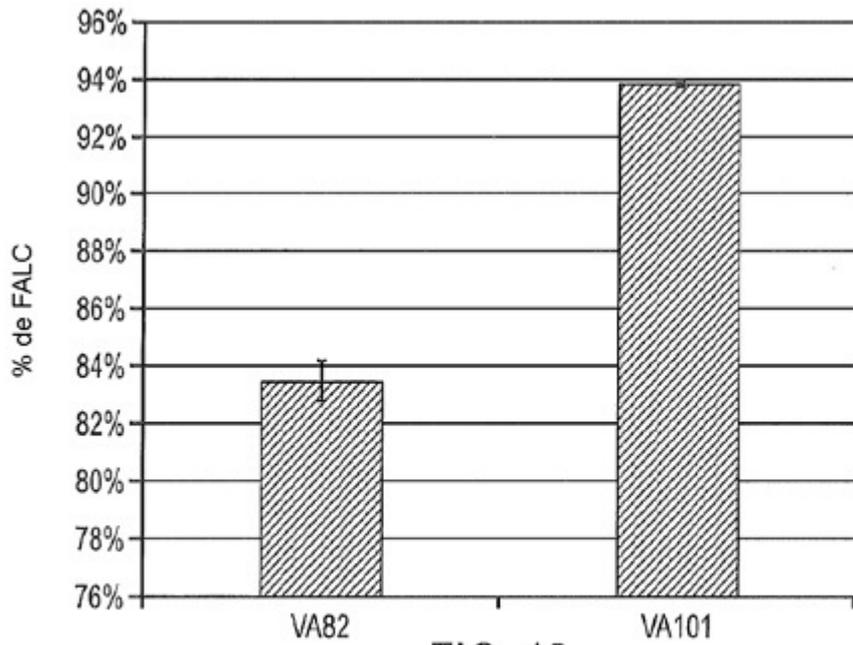


FIG. 13

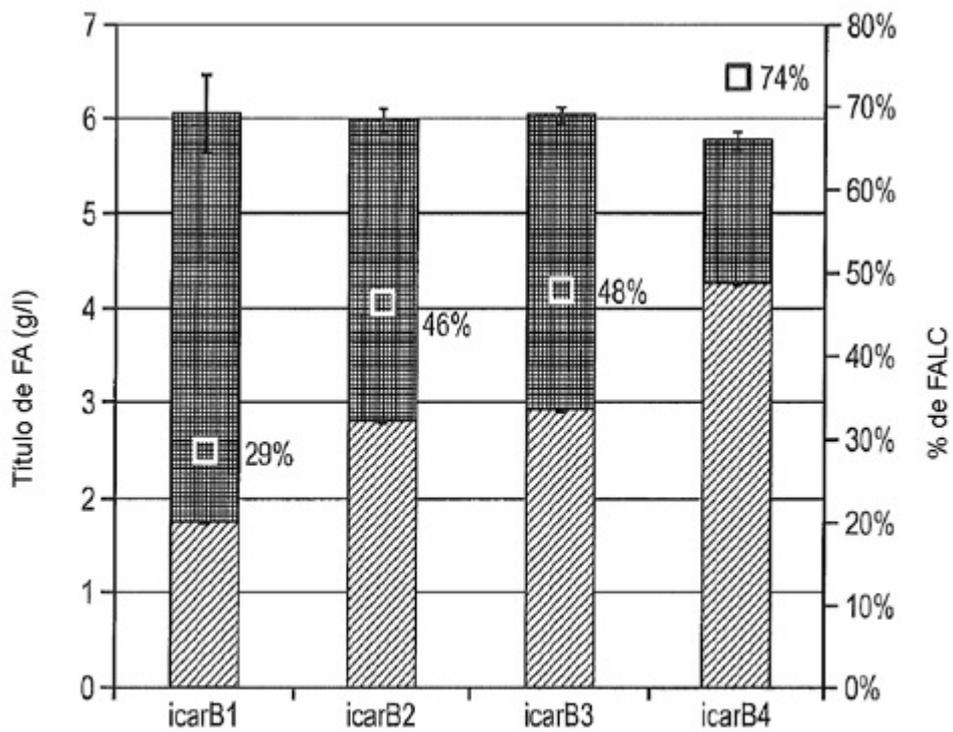


FIG. 14

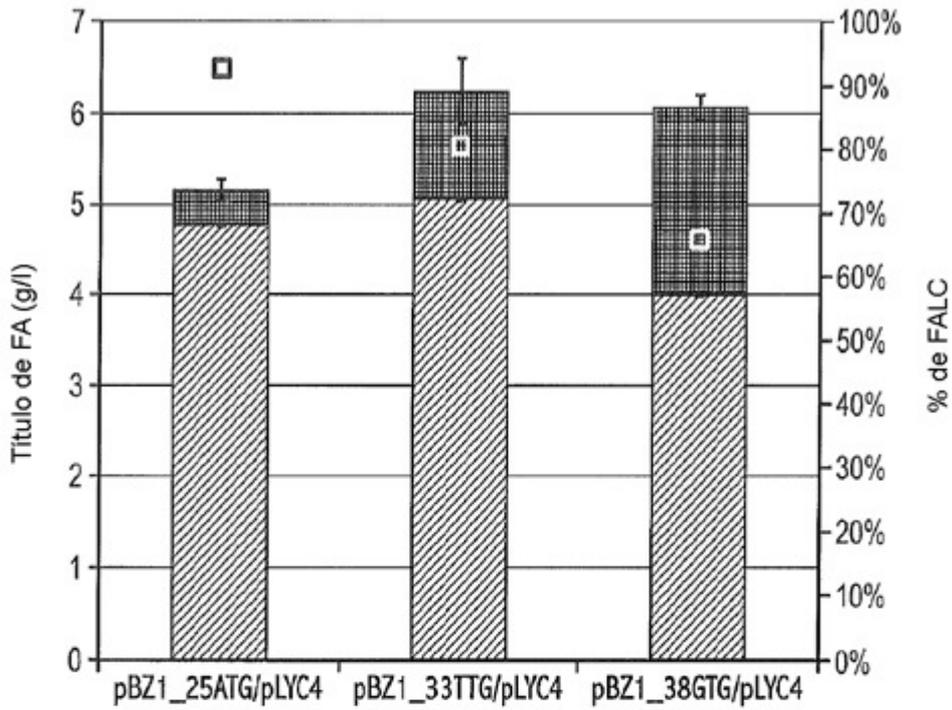


FIG. 15

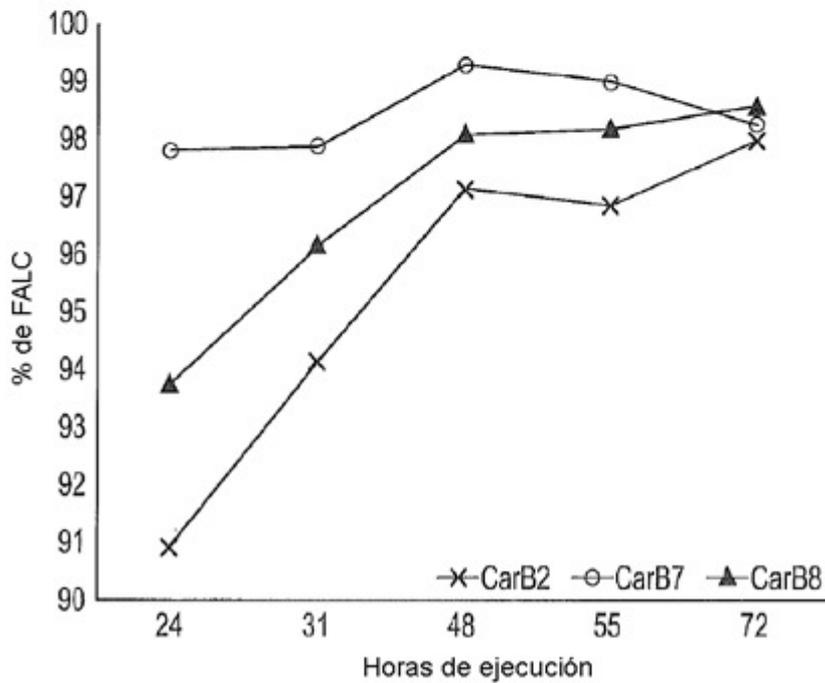


FIG. 16