

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 667 469**

51 Int. Cl.:

C12N 9/02 (2006.01)
C12N 9/88 (2006.01)
C12P 7/64 (2006.01)
C10L 1/02 (2006.01)
C12P 7/04 (2006.01)
C12P 7/24 (2006.01)
C12P 5/02 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

- 86 Fecha de presentación y número de la solicitud internacional: **18.05.2009 PCT/US2009/044403**
- 87 Fecha y número de publicación internacional: **19.11.2009 WO09140695**
- 96 Fecha de presentación y número de la solicitud europea: **18.05.2009 E 09747776 (4)**
- 97 Fecha y número de publicación de la concesión europea: **07.03.2018 EP 2288694**

54 Título: **Métodos y composiciones para producir hidrocarburos**

30 Prioridad:

16.05.2008 US 53955 P

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
11.05.2018

73 Titular/es:

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

72 Inventor/es:

**SCHIRMER, ANDREAS;
RUDE, MATHEW y
BRUBAKER, SHANE**

74 Agente/Representante:

MILTENYI, Peter

ES 2 667 469 T3

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

DESCRIPCIÓN

Métodos y composiciones para producir hidrocarburos

La presente invención se refiere a métodos de producción de un alcano o alqueno a partir de una célula huésped modificada genéticamente.

5 **Antecedentes de la invención**

El petróleo es un recurso natural limitado que se encuentra en la Tierra en formas líquida, gaseosa o sólida. El petróleo está compuesto principalmente por hidrocarburos, que están compuestos principalmente de carbono e hidrógeno. También contiene cantidades significativas de otros elementos, tales como nitrógeno, oxígeno o azufre, en diferentes formas.

10 El petróleo es un recurso valioso, aunque los productos del petróleo se desarrollan a considerables costes, tanto financieros como medioambientales. En primer lugar, deben descubrirse fuentes de petróleo. La exploración de petróleo es una empresa cara y arriesgada. El coste de exploración de pozos de aguas profundas puede exceder los 100 millones de dólares. Además, no hay ninguna garantía de que estos pozos contengan petróleo. Se estima que sólo el 40% de los pozos perforados conducen a pozos productores que generan hidrocarburos comerciales.
15 Además del coste económico, la exploración de petróleo conlleva un alto coste medioambiental. Por ejemplo, la exploración mar adentro altera los entornos marinos circundantes.

Tras descubrirse un pozo productor, el petróleo debe extraerse de la tierra con gran coste. Durante la recuperación primaria, la presión natural subterránea es suficiente para extraer aproximadamente el 20% del petróleo en el pozo. A medida que esta presión natural desciende, se emplean métodos de recuperación secundarios, si son económicos. Generalmente, la recuperación secundaria implica aumentar la presión del pozo mediante, por ejemplo, inyección de agua, inyección de gas natural o levantamiento por gas. Usando métodos de recuperación secundarios, se recupera de un 5% a un 15% adicional de petróleo. Una vez que se agotan los métodos de recuperación secundarios, pueden usarse métodos de recuperación terciarios, si son económicos. Los métodos terciarios implican reducir la viscosidad del petróleo para hacer que sea más fácil de extraer. Usando métodos de recuperación terciarios, se recupera de un 5% a un 15% adicional de petróleo. Por tanto, incluso en las mejores circunstancias, sólo puede extraerse el 50% del petróleo en un pozo. La extracción de petróleo también conlleva un coste medioambiental. Por ejemplo, la extracción de petróleo puede dar como resultado grandes fugas de petróleo que emerge a la superficie. Además, la perforación mar adentro implica dragar el lecho marino, lo que altera o destruye el entorno marino circundante.

30 Puesto que los depósitos de petróleo no se encuentran uniformemente por toda la Tierra, el petróleo debe transportarse a lo largo de grandes distancias desde regiones productoras de petróleo hasta regiones consumidoras de petróleo. Además de los costes de transporte, existe también el riesgo medioambiental de derrames de petróleo devastadores.

35 En su forma natural, el petróleo crudo extraído de la Tierra tiene pocos usos comerciales. Es una mezcla de hidrocarburos (por ejemplo, parafinas (o alcanos), olefinas (o alquenos), alquinos, naftenos (o cicloalcanos), compuestos alifáticos, compuestos aromáticos, etc.) de longitud y complejidad variables. Además, el petróleo crudo contiene otros compuestos orgánicos (por ejemplo, compuestos orgánicos que contienen nitrógeno, oxígeno, azufre, etc.) e impurezas (por ejemplo, azufre, sal, ácido, metales, etc.).

40 Por tanto, el petróleo crudo debe refinarse y purificarse antes de que pueda usarse comercialmente. Debido a su alta densidad de energía y su fácil transportabilidad, la mayoría del petróleo se refina en combustibles, tales como combustibles de transporte (por ejemplo, gasolina, diésel, combustible de aviación, etc.), gasóleo de calefacción, gas licuado de petróleo, etc.

45 El petróleo crudo es también una fuente primaria de materiales de partida para producir productos petroquímicos. Las dos clases principales de materiales de partida derivados del petróleo son olefinas de cadena corta (por ejemplo, etileno y propileno) y compuesto aromáticos (por ejemplo, isómeros de xileno y benceno). Estos materiales de partida se derivan de hidrocarburos de cadena más larga en petróleo crudo sometiendo a craqueo con un gasto considerable usando una variedad de métodos, tales como craqueo catalítico, craqueo con vapor o reformación catalítica. Estos materiales de partida se usan para producir productos petroquímicos, que no pueden refinarse directamente del petróleo crudo, tales como monómeros, disolventes, detergentes o adhesivos.

50 Un ejemplo de un material de partida derivado de petróleo crudo es etileno. Se usa etileno para producir productos petroquímicos, tales como polietileno, etanol, óxido de etileno, etilenglicol, poliéster, glicol éter, etoxilato, acetato de vinilo, 1,2-dicloroetano, tricloroetileno, tetracloroetileno, cloruro de vinilo y poli(cloruro de vinilo). Un ejemplo adicional de un material de partida es propileno, que se usa para producir alcohol isopropílico, acrilonitrilo, polipropileno, óxido de propileno, propilenglicol, glicol éteres, butileno, isobutileno, 1,3-butadieno, elastómeros sintéticos, poliolefinas,
55 alfa-olefinas, alcoholes grasos, ácido acrílico, polímeros acrílicos, cloruro de alilo, epiclohidrina y resinas epoxídicas.

- 5 Estos productos petroquímicos pueden usarse entonces para producir productos químicos especializados, tales como plásticos, resinas, fibras, elastómeros, productos farmacéuticos, lubricantes o geles. Productos químicos especializados particulares que pueden producirse a partir de materiales de partida petroquímicos son: ácidos grasos, hidrocarburos (por ejemplo, de cadena larga, de cadena ramificada, saturados, insaturados, etc.), alcoholes grasos, ésteres, aldehídos grasos, cetonas, lubricantes, etc.
- 10 Los productos químicos especializados tienen muchos usos comerciales. Los ácidos grasos se usan comercialmente como tensioactivos, por ejemplo, en detergentes y jabones. También pueden usarse como aditivos en combustibles, aceites lubricantes, pinturas, lacas, velas, aceite para ensaladas, mantequilla, cosméticos y emulsionantes. Además, los ácidos grasos se usan como activadores aceleradores en productos de caucho. Los ácidos grasos también pueden usarse como materia prima para producir ésteres metílicos, amidas, aminas, cloruros de ácido, anhídridos, dímeros de ceteno, y peroxiácidos y ésteres.
- 15 Los hidrocarburos tienen muchos usos comerciales. Por ejemplo, se usan alcanos de cadena más corta como combustibles. El metano y el etano son los principales constituyentes del gas natural. Se usan alcanos de cadena más larga (por ejemplo, de desde cinco hasta dieciséis carbonos) como combustibles de transporte (por ejemplo, gasolina, diésel o combustible de aviación). Los alcanos que tienen más de dieciséis átomos de carbono son componentes importantes de gasóleos y aceites lubricantes. Pueden usarse alcanos incluso más largos, que son sólidos a temperatura ambiente, por ejemplo, como cera de parafina. Los alcanos que contienen aproximadamente treinta y cinco carbonos se encuentran en alquitrán, que se usa para la pavimentación de carreteras. Además, pueden someterse a craqueo alcanos de cadena más larga para producir hidrocarburos de cadena más corta útiles comercialmente.
- 20 Como los alcanos de cadena corta, los alquenos de cadena corta se usan en combustibles de transporte. Se usan alquenos de cadena más larga en plásticos, lubricantes y lubricantes sintéticos. Además, se usan alquenos como materia prima para producir alcoholes, ésteres, plastificantes, tensioactivos, aminas terciarias, agentes de recuperación de petróleo potenciada, ácidos grasos, tioles, anhídridos alqueniilsuccínicos, epóxidos, alcanos clorados, alquenos clorados, ceras, aditivos para combustibles y reductores del flujo por arrastre.
- 25 Los alcoholes grasos tienen muchos usos comerciales. Los alcoholes grasos de cadena más corta se usan en las industrias alimentaria y cosmética como emulsionantes, emolientes y espesantes. Debido a su naturaleza anfífila, los alcoholes grasos se comportan como tensioactivos no iónicos, que son útiles como detergentes. Además, los alcoholes grasos se usan en ceras, gomas, resinas, lociones y bálsamos farmacéuticos, aditivos para aceites lubricantes, agentes de acabado y antiestáticos textiles, plastificantes, cosméticos, disolventes industriales y disolventes para grasas.
- 30 Los ésteres tienen muchos usos comerciales. Por ejemplo, el biodiésel, un combustible alternativo, está compuesto por ésteres (por ejemplo, éster metílico de ácidos grasos, ésteres etílicos de ácidos grasos, etc.). Algunos ésteres de bajo peso molecular son volátiles con un olor agradable que los hace útiles como fragancias o agentes saborizantes. Además, se usan ésteres como disolventes para lacas, pinturas y barnices. Además, algunas sustancias que se producen de manera natural, tales como ceras, grasas y aceites están compuestas por ésteres. También se usan ésteres como agentes de ablandamiento en resinas y plásticos, plastificantes, retardadores de la llama y aditivos en gasolina y gasoil. Además, pueden usarse ésteres en la fabricación de polímeros, películas, materiales textiles, colorantes y productos farmacéuticos.
- 35 Se usan aldehídos para producir muchos productos químicos especializados. Por ejemplo, se usan aldehídos para producir polímeros, resinas (por ejemplo, baquelita), colorantes, sabores, plastificantes, perfumes, productos farmacéuticos y otros productos químicos. Algunos se usan como disolventes, conservantes o desinfectantes. Algunos compuestos naturales y sintéticos, tales como vitaminas y hormonas, son aldehídos. Además, muchos azúcares contienen grupos aldehído.
- 40 Se usan cetonas comercialmente como disolventes. Por ejemplo, se usa frecuentemente acetona como disolvente, pero es también un material de partida para producir polímeros. También se usan cetonas en lacas, pinturas, explosivos, perfumes y procesamiento de materiales textiles. Además, se usan cetonas para producir alcoholes, alquenos, alcanos, iminas y enaminas.
- 45 Además, el petróleo crudo es una fuente de lubricantes. Los lubricantes derivados del petróleo están compuestos normalmente por olefinas, particularmente poliolefinas y alfa-olefinas. Los lubricantes o bien pueden refinarse del petróleo crudo o bien fabricarse usando materiales de partida refinados del petróleo crudo.
- 50 La obtención de estos productos químicos especializados a partir de petróleo crudo requiere una inversión financiera significativa así como un gran aporte de energía. También es un proceso ineficaz porque frecuentemente los hidrocarburos de cadena larga en petróleo crudo se someten a craqueo para producir monómeros más pequeños. Estos monómeros se usan entonces como material de partida para fabricar los productos químicos especializados más complejos.
- 55 Además de los problemas con la exploración, la extracción, el transporte y el refinado del petróleo, el petróleo es un recurso cada vez más escaso y limitado. Una estimación del consumo de petróleo mundial es de 30000 millones de

barriles al año. Según algunas estimaciones, se predice que a los niveles de producción actuales, las reservas de petróleo mundiales podrían agotarse antes del año 2050.

Finalmente, la combustión de combustibles a base de petróleo libera gases invernadero (por ejemplo, dióxido de carbono) y otras formas de contaminación del aire (por ejemplo, monóxido de carbono, dióxido de azufre, etc.). A medida que la demanda mundial de combustible aumenta, la emisión de gases invernadero y otras formas de contaminación del aire también aumenta. La acumulación de gases invernadero en la atmósfera conduce a un aumento del calentamiento global. Por tanto, además de dañar el medio ambiente localmente (por ejemplo, derrames de petróleo, dragado de entornos marinos, etc.), la combustión del petróleo también daña el medio ambiente globalmente.

Debido a los desafíos inherentes que plantea el petróleo, hay una necesidad de una fuente de petróleo renovable que no necesite explorarse, extraerse, transportarse a lo largo de grandes distancias o refinarse sustancialmente como el petróleo. Hay también una necesidad de una fuente de petróleo renovable que pueda producirse económicamente sin crear el tipo de daño medioambiental producido por la industria del petróleo y la combustión de combustibles a base de petróleo. Por motivos similares, hay también una necesidad de una fuente renovable de productos químicos que se derivan normalmente del petróleo. La entrada de la base de datos UniProt Q54764 da a conocer una secuencia polipeptídica idéntica en el 100% a la presente SEQ ID NO: 2. Fehler *et al.* (Biochemistry (1970), Vol. 9, n. ° 2, págs. 418-22) describen la biosíntesis de hidrocarburos, en particular alcanos, en *Anabaena variabilis*, y el análisis de espectrometría de masas de los mismos. El documento WO 2007/136762 se refiere a la producción de derivados de ácidos grasos, incluyendo la producción de alcanos y alquenos, en organismos modificados genéticamente, incluyendo *E. coli* modificado genéticamente. Phung *et al.* (Abstracts of the General Meeting of the American Society for Microbiology, 1 de enero de 1995, pág. 524, Abstract H-182-ISSN: 1060-2011) describen la identificación y secuenciación de cuatro polipéptidos implicados en las etapas iniciales de la biosíntesis de ácidos grasos en *Synechococcus sp.*, concretamente la proteína transportadora de biotina carboxilo (BCCP), una biotina carboxilasa, la subunidad alfa de carboxiltransferasa, y una beta-cetoaciltransferasa. Ladygina *et al.* (Process Biochemistry (2006), Vol. 41, n. ° 5, págs. 1359-5113) están revisando la biosíntesis microbiana de hidrocarburos.

Sumario de la invención

La invención se basa, al menos en parte, en la identificación de genes de cianobacterias que codifican para polipéptidos de biosíntesis de hidrocarburos. La invención se define por las reivindicaciones. En particular, la invención proporciona un método de producción de un alcano o alqueno a partir de una célula huésped modificada genéticamente, comprendiendo el método: (a) transformar una célula huésped para que se exprese introduciendo (i) un polinucleótido de biosíntesis de aldehídos que codifica para un polipéptido de biosíntesis de aldehídos, y (ii) un polinucleótido de biosíntesis de alcanos y alquenos que codifica para un polipéptido de biosíntesis de alcanos y alquenos; y (b) cultivar dicha célula huésped modificada genéticamente en un medio de cultivo que contiene una fuente de carbono en condiciones eficaces para producir un alcano o alqueno en el medio de cultivo, en el que dicho polipéptido de biosíntesis de aldehídos tiene actividad acil-ACP reductasa y procede de una cianobacteria, y en el que dicho polipéptido de biosíntesis de alcanos y alquenos tiene actividad descarboxilasa y procede de una cianobacteria. En algunas realizaciones de la divulgación, el método comprende producir en una célula huésped un polipéptido que comprende la secuencia de aminoácidos de SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34 ó 36, o una variante del mismo, y aislar el hidrocarburo de la célula huésped.

En algunas realizaciones de la divulgación, el polipéptido comprende una secuencia de aminoácidos que tiene una identidad de al menos aproximadamente el 70%, al menos aproximadamente el 80%, al menos aproximadamente el 85%, al menos aproximadamente el 90%, al menos aproximadamente el 91%, al menos aproximadamente el 92%, al menos aproximadamente el 93%, al menos aproximadamente el 94%, al menos aproximadamente el 95%, al menos aproximadamente el 96%, al menos aproximadamente el 97%, al menos aproximadamente el 98%, o al menos aproximadamente el 99% con SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34 ó 36.

En algunas realizaciones de la divulgación, el polipéptido comprende la secuencia de aminoácidos de SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34 ó 36 con una o más sustituciones, adiciones, inserciones o deleciones de aminoácido. El polipéptido tiene actividad descarboxilasa y procede de una cianobacteria. Todavía en otras realizaciones de la divulgación, el polipéptido comprende la secuencia de aminoácidos de SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34 ó 36, con una o más sustituciones de aminoácidos conservativas. Por ejemplo, el polipéptido comprende 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 lleva un grupo amida, tal como asparagina y glutamina, por otro residuo que lleva 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 de la divulgación, el polipéptido 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. El polipéptido tiene actividad descarboxilasa y procede de una cianobacteria.

En otras realizaciones de la divulgación, el polipéptido comprende la secuencia de aminoácidos de: (i) SEQ ID NO: 37 o SEQ ID NO: 38 o SEQ ID NO: 39; o (ii) SEQ ID NO: 40 y una cualquiera de (a) SEQ ID NO: 37, (b) SEQ ID NO: 38, y (c) SEQ ID NO: 39; o (iii) SEQ ID NO: 41 o SEQ ID NO: 42 o SEQ ID NO: 43 o SEQ ID NO: 44. El polipéptido tiene actividad descarboxilasa y procede de una cianobacteria.

- 5 En algunas realizaciones de la divulgación, el método comprende expresar en una célula huésped un polinucleótido que comprende una secuencia de nucleótidos que tiene una identidad de secuencia de al menos aproximadamente el 70%, al menos aproximadamente el 75%, al menos aproximadamente el 80%, al menos aproximadamente el 85%, al menos aproximadamente el 90%, al menos aproximadamente el 91%, al menos aproximadamente el 92%, al menos aproximadamente el 93%, al menos aproximadamente el 94%, al menos aproximadamente el 95%, al menos aproximadamente el 96%, al menos aproximadamente el 97%, al menos aproximadamente el 98% o al menos aproximadamente el 99% SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33 ó 35. En algunas realizaciones de la divulgación, la secuencia de nucleótidos es SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33 ó 35. En algunas realizaciones de la divulgación, el método comprende además aislar el hidrocarburo de la célula huésped.
- 10
- 15 En otras realizaciones de la divulgación, la secuencia de nucleótidos hibrida con un complemento de SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33 ó 35, o con un fragmento del mismo, por ejemplo, en condiciones de baja rigurosidad, rigurosidad media, alta rigurosidad o muy alta rigurosidad.

- 20 En otras realizaciones de la divulgación, la secuencia de nucleótidos codifica para un polipéptido que comprende: (i) la secuencia de aminoácidos de SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34 ó 36; o (ii) la secuencia de aminoácidos de SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34 ó 36 con una o más sustituciones, adiciones, inserciones o deleciones de aminoácidos. En algunas realizaciones de la divulgación, el polipéptido comprende la secuencia de aminoácidos de SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34 ó 36 con una o más sustituciones de aminoácidos conservativas. El polipéptido tiene actividad descarboxilasa y procede de una cianobacteria.

- 25 En otras realizaciones de la divulgación, la secuencia de nucleótidos codifica para un polipéptido que tiene la misma actividad biológica que un polipéptido que comprende la secuencia de aminoácidos de SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34 ó 36. En algunas realizaciones, la secuencia de nucleótidos es SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33 ó 35 o un fragmento de la misma. En otras realizaciones de la divulgación, la secuencia de nucleótidos hibrida con un complemento de SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33 ó 35 o con un fragmento del mismo, por ejemplo, en condiciones de baja rigurosidad, rigurosidad media, alta rigurosidad o muy alta rigurosidad. La actividad biológica es actividad descarboxilasa y el polipéptido codificado procede de una cianobacteria.
- 30

- 35 En algunas realizaciones de la divulgación, el método comprende transformar una célula huésped con un vector recombinante que comprende una secuencia de nucleótidos que tiene una identidad de secuencia de al menos aproximadamente el 70%, al menos aproximadamente el 75%, al menos aproximadamente el 80%, al menos aproximadamente el 85%, al menos aproximadamente el 90%, al menos aproximadamente el 91%, al menos aproximadamente el 92%, al menos aproximadamente el 93%, al menos aproximadamente el 94%, al menos aproximadamente el 95%, al menos aproximadamente el 96%, al menos aproximadamente el 97%, al menos aproximadamente el 98%, o al menos aproximadamente el 99% con SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33 ó 35. En algunas realizaciones de la divulgación, el vector recombinante comprende además un promotor operativamente unido a la secuencia de nucleótidos. En algunas realizaciones de la divulgación, 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 realizaciones particulares de la divulgación, el vector recombinante comprende al menos una secuencia seleccionada del grupo que consiste en (a) una secuencia reguladora acoplada operativamente a la secuencia de nucleótidos; (b) un marcador de selección acoplado operativamente a la secuencia de nucleótidos; (c) una secuencia de marcador acoplada operativamente a la secuencia de nucleótidos; (d) un resto de purificación acoplado operativamente a la secuencia de nucleótidos; (e) una secuencia de secreción acoplada operativamente a la secuencia de nucleótidos; y (f) una secuencia de direccionamiento acoplada operativamente a la secuencia de nucleótidos. En determinadas realizaciones de la divulgación, la secuencia de nucleótidos se incorpora de manera estable en el ADN genómico de la célula huésped, y la expresión de la secuencia de nucleótidos está bajo el control de una región promotora regulada.
- 40
- 45
- 50

En cualquiera de los aspectos descritos en el presente documento, la célula huésped puede seleccionarse del grupo que consiste en una célula de mamífero, célula vegetal, célula de insecto, célula de levadura, célula de hongo, célula de hongos filamentosos y célula bacteriana.

- 55 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 se selecciona del género *Escherichia*, *Bacillus*, *Lactobacillus*, *Rhodococcus*, *Pseudomonas*, *Aspergillus*, *Trichoderma*, *Neurospora*, *Fusarium*, *Humicola*, *Rhizomucor*, *Kluyveromyces*, *Pichia*, *Mucor*, *Myceliophthora*, *Penicillium*, *Phanerochaete*, *Pleurotus*, *Trametes*, *Chrysosporium*,

Saccharomyces, *Stenotrophomonas*, *Schizosaccharomyces*, *Yarrowia* o *Streptomyces*.

5 En realizaciones particulares, 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 pumilus*, 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 amiloliquefaciens*.

10 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 michei*.

Aún en otras realizaciones, la célula huésped es una célula de *Streptomyces lividans* o una célula de *Streptomyces murinus*. En otras realizaciones, la célula huésped es una célula de *Actinomycetes*.

15 En algunas realizaciones de la divulgación, la célula huésped es una célula CHO, una célula COS, una célula VERO, una célula BHK, una célula HeLa, una célula Cv1, una célula MDCK, una célula 293, una célula 3T3 o una célula PC12.

En realizaciones particulares, la célula huésped es una célula de *E. coli*, tal como una célula de *E. coli* de cepa B, una de cepa C, una de cepa K o una de cepa W.

20 En otras realizaciones, la célula huésped es una célula huésped cianobacteriana. En realizaciones particulares, la célula huésped cianobacteriana es una célula enumerada en la tabla 1.

En algunas realizaciones, los alcanos y alquenos se secretan de la célula huésped.

25 En determinadas realizaciones de la divulgación, la célula huésped sobreexpresa un sustrato descrito en el presente documento. En algunas realizaciones de la divulgación, el método incluye además transformar la célula huésped con un ácido nucleico que codifica para una enzima descrita en el presente documento, y la célula huésped sobreexpresa un sustrato descrito en el presente documento. En otras realizaciones de la divulgación, el método incluye además cultivar la célula huésped en presencia de al menos un sustrato descrito en el presente documento. En algunas realizaciones de la divulgación, el sustrato es un derivado de ácido graso, una acil-ACP, un ácido graso, una acil-CoA, un aldehído graso, un alcohol graso o un éster graso.

30 En algunas realizaciones de la divulgación, el sustrato de derivado de ácido graso es un sustrato de derivado de ácido graso insaturado, un sustrato de derivado de ácido graso monoinsaturado o un sustrato de derivado de ácido graso saturado. En otras realizaciones, el sustrato de derivado de ácido graso es un sustrato de derivado de ácido graso de cadena lineal, un sustrato de derivado de ácido graso de cadena ramificada o un sustrato de derivado de ácido graso que incluye un resto cíclico.

35 En determinadas realizaciones de los aspectos descritos en el presente documento, el hidrocarburo producido es un alcano. En algunas realizaciones, el alcano es un alcano C₃-C₂₅. Por ejemplo, el alcano es un alcano C₃, C₄, C₅, C₆, C₇, C₈, C₉, C₁₀, C₁₁, C₁₂, C₁₃, C₁₄, C₁₅, C₁₆, C₁₇, C₁₈, C₁₉, C₂₀, C₂₁, C₂₂, C₂₃, C₂₄ o C₂₅. En algunas realizaciones, el alcano es tridecano, metiltridecano, nonadecano, metilnonadecano, heptadecano, metilheptadecano, pentadecano o metilpentadecano.

40 En algunas realizaciones, el alcano es un alcano de cadena lineal, un alcano de cadena ramificada o un alcano cíclico.

45 En determinadas realizaciones, el método comprende además cultivar la célula huésped en presencia de un derivado de ácido graso saturado, y el hidrocarburo producido es un alcano. En determinadas realizaciones, el derivado de ácido graso saturado es un sustrato derivado de ácido graso C₆-C₂₆. Por ejemplo, el sustrato derivado de ácido graso es un sustrato derivado de ácido graso C₆, C₇, C₈, C₉, C₁₀, C₁₁, C₁₂, C₁₃, C₁₄, C₁₅, C₁₆, C₁₇, C₁₈, C₁₉, C₂₀, C₂₁, C₂₂, C₂₃, C₂₄, C₂₅ o C₂₆. En realizaciones particulares, el sustrato derivado de ácido graso es 2-metilicosanal, icosanal, octadecanal, tetradecanal, 2-metiloctadecanal, estearaldehído o palmitaldehído.

En algunas realizaciones, el método incluye además aislar el alcano de la célula huésped o del medio de cultivo. En otras realizaciones de la divulgación, el método incluye además someter el alcano a craqueo o refino.

50 En determinadas realizaciones de los aspectos descritos en el presente documento, el hidrocarburo producido es un alqueno. En algunas realizaciones de la divulgación, el alqueno es un alqueno C₃-C₂₅. Por ejemplo, el alqueno es un alqueno C₃, C₄, C₅, C₆, C₇, C₈, C₉, C₁₀, C₁₁, C₁₂, C₁₃, C₁₄, C₁₅, C₁₆, C₁₇, C₁₈, C₁₉, C₂₀, C₂₁, C₂₂, C₂₃, C₂₄ o C₂₅. En algunas realizaciones, el alqueno es pentadeceno, heptadeceno, metilpentadeceno o metilheptadeceno.

En algunas realizaciones, el alqueno es un alqueno de cadena lineal, un alqueno de cadena ramificada o un alqueno

cíclico.

5 En determinadas realizaciones, el método comprende además cultivar la célula huésped en presencia de un derivado de ácido graso insaturado, y el hidrocarburo producido es un alqueno. En determinadas realizaciones, el derivado de ácido graso insaturado es un sustrato derivado de ácido graso C₆-C₂₆. Por ejemplo, el sustrato derivado de ácido graso es un sustrato derivado de ácido graso insaturado C₆, C₇, C₈, C₉, C₁₀, C₁₁, C₁₂, C₁₃, C₁₄, C₁₅, C₁₆, C₁₇, C₁₈, C₁₉, C₂₀, C₂₁, C₂₂, C₂₃, C₂₄, C₂₅ o C₂₆. En realizaciones particulares, el sustrato derivado de ácido graso es octadecenal, hexadecenal, metilhexadecenal o metiloctadecenal.

10 En otro aspecto, en el presente documento se describe un microorganismo modificado genéticamente que comprende una secuencia de control exógena incorporada de manera estable en el ADN genómico del microorganismo. En una realización, la secuencia de control se integra en el sentido de 5' de un polinucleótido que comprende una secuencia de nucleótidos que tiene una identidad de secuencia de al menos aproximadamente el 70% con SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33 ó 35. En algunos aspectos, la secuencia de nucleótidos tiene una identidad de secuencia de al menos aproximadamente el 75%, al menos aproximadamente el 80%, al menos aproximadamente el 85%, al menos aproximadamente el 90%, al menos aproximadamente el 91%, al menos aproximadamente el 92%, al menos aproximadamente el 93%, al menos aproximadamente el 94%, al menos aproximadamente el 95%, al menos aproximadamente el 96%, al menos aproximadamente el 97%, al menos aproximadamente el 98%, o al menos aproximadamente el 99% con SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33 ó 35. En algunos aspectos, la secuencia de nucleótidos es SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33 ó 35.

20 En algunas realizaciones de la divulgación, el microorganismo expresa un nivel aumentado de un hidrocarburo en relación con una cianobacteria silvestre.

En otro aspecto, en el presente documento se describe un método de preparación de un hidrocarburo, comprendiendo el método cultivar un microorganismo modificado genéticamente descrito en el presente documento en condiciones adecuadas para la expresión génica, y aislar el hidrocarburo.

25 En algunas realizaciones de la divulgación, el polipéptido (i) tiene una identidad de al menos el 70% con la secuencia de aminoácidos de SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34 ó 36, o una variante de la misma; (ii) se codifica por una secuencia de nucleótidos que tiene una identidad de al menos el 70% con SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33 ó 35, o una variante de la misma; (iii) comprende la secuencia de aminoácidos de SEQ ID NO: 37, 38 ó 39; (iv) comprende la secuencia de aminoácidos de SEQ ID NO: 40 y una cualquiera de (a) SEQ ID NO: 37, (b) SEQ ID NO: 38 y (c) SEQ ID NO: 39; o (v) es SEQ ID NO: 41, 42, 43 ó 44. El polipéptido tiene actividad descarboxilasa y procede de una cianobacteria.

30 En algunas realizaciones de la divulgación, el polipéptido tiene una identidad de al menos aproximadamente el 80%, al menos aproximadamente el 85%, al menos aproximadamente el 90%, al menos aproximadamente el 91%, al menos aproximadamente el 92%, al menos aproximadamente el 93%, al menos aproximadamente el 94%, al menos aproximadamente el 95%, al menos aproximadamente el 96%, al menos aproximadamente el 97%, al menos aproximadamente el 98%, o al menos aproximadamente el 99% con SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34 ó 36. En algunas realizaciones de la divulgación, el polipéptido tiene la secuencia de aminoácidos de SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34 ó 36.

35 En algunas realizaciones de la divulgación, el polipéptido se codifica por una secuencia de nucleótidos que tiene una identidad de secuencia de al menos aproximadamente el 75%, al menos aproximadamente el 80%, al menos aproximadamente el 85%, al menos aproximadamente el 90%, al menos aproximadamente el 91%, al menos aproximadamente el 92%, al menos aproximadamente el 93%, al menos aproximadamente el 94%, al menos aproximadamente el 95%, al menos aproximadamente el 96%, al menos aproximadamente el 97%, al menos aproximadamente el 98%, o al menos aproximadamente el 99% con SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33 ó 35. En algunas realizaciones de la divulgación, el polipéptido se codifica por una secuencia de nucleótidos que tiene SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33 ó 35.

En algunas realizaciones, el sustrato biológico es un derivado de ácido graso, una acil-ACP, un ácido graso, una acil-CoA, un aldehído graso, un alcohol graso, o un éster graso.

50 En algunas realizaciones, el sustrato es un derivado de ácido graso saturado, y el hidrocarburo es un alcano, por ejemplo, un alcano C₃-C₂₅. Por ejemplo, el alcano es un alcano C₃, C₄, C₅, C₆, C₇, C₈, C₉, C₁₀, C₁₁, C₁₂, C₁₃, C₁₄, C₁₅, C₁₆, C₁₇, C₁₈, C₁₉, C₂₀, C₂₁, C₂₂, C₂₃, C₂₄ o C₂₅. En algunas realizaciones, el alcano es tridecano, metiltridecano, nonadecano, metilnonadecano, heptadecano, metilheptadecano, pentadecano o metilpentadecano.

En algunas realizaciones, el alcano es un alcano de cadena lineal, un alcano de cadena ramificada o un alcano cíclico.

55 En algunas realizaciones, el derivado de ácido graso saturado es 2-metilicosanal, icosanal, octadecanal, tetradecanal, 2-metiloctadecanal, estearaldehído o palmitaldehído.

En otras realizaciones de la divulgación, el sustrato biológico es un derivado de ácido graso insaturado, y el hidrocarburo es un alqueno, por ejemplo, un alqueno C₃-C₂₅. Por ejemplo, el alqueno es un alqueno C₃, C₄, C₅, C₆, C₇, C₈, C₉, C₁₀, C₁₁, C₁₂, C₁₃, C₁₄, C₁₅, C₁₆, C₁₇, C₁₈, C₁₉, C₂₀, C₂₁, C₂₂, C₂₃, C₂₄ o C₂₅. En algunas realizaciones, el alqueno es pentadeceno, heptadeceno, metilpentadeceno o metilheptadeceno.

5 En algunas realizaciones, el alqueno es un alqueno de cadena lineal, un alqueno de cadena ramificada o un alqueno cíclico.

En algunas realizaciones, el derivado de ácido graso insaturado es octadecenal, hexadecenal, metilhexadecenal o metiloctadecenal.

10 En otro aspecto, en el presente documento se describe un hidrocarburo producido mediante cualquiera de los métodos o microorganismos descritos en el presente documento. El hidrocarburo puede ser un alcano o un alqueno que tiene una $\delta^{13}\text{C}$ de aproximadamente -15,4 o mayor. Por ejemplo, el alcano o alqueno tiene una $\delta^{13}\text{C}$ de aproximadamente -15,4 a aproximadamente -10,9, por ejemplo, de aproximadamente -13,92 a aproximadamente -13,84. El alcano o alqueno puede tener una $f_M^{14}\text{C}$ de al menos aproximadamente 1,003. Por ejemplo, el alcano o alqueno tiene una $f_M^{14}\text{C}$ de al menos aproximadamente 1,01 o al menos aproximadamente 1,5. El alcano o alqueno puede tener una $f_M^{14}\text{C}$ de aproximadamente 1,111 a aproximadamente 1,124.

15 En otro aspecto, en el presente documento se describe un biocombustible que incluye un hidrocarburo producido mediante cualquiera de los métodos o microorganismos descritos en el presente documento. El hidrocarburo puede ser un alcano o alqueno que tiene una $\delta^{13}\text{C}$ de aproximadamente -15,4 o mayor. Por ejemplo, el alcano o alqueno tiene una $\delta^{13}\text{C}$ de aproximadamente -15,4 a aproximadamente -10,9, por ejemplo, de aproximadamente -13,92 a aproximadamente -13,84. El alcano o alqueno puede tener una $f_M^{14}\text{C}$ de al menos aproximadamente 1,003. Por ejemplo, el alcano o alqueno tiene una $f_M^{14}\text{C}$ de al menos aproximadamente 1,01 o al menos aproximadamente 1,5. El alcano o alqueno tiene una $f_M^{14}\text{C}$ de aproximadamente 1,111 a aproximadamente 1,124. El biocombustible es diésel, gasolina, o combustible para motores a reacción.

20 En otro aspecto, en el presente documento se describe un ácido nucleico aislado que consiste en no más de aproximadamente 500 nucleótidos de SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33 ó 35. El ácido nucleico puede consistir en no más de aproximadamente 300 nucleótidos, no más de aproximadamente 350 nucleótidos, no más de aproximadamente 400 nucleótidos, no más de aproximadamente 450 nucleótidos, no más de aproximadamente 550 nucleótidos, no más de aproximadamente 600 nucleótidos, o no más de aproximadamente 650 nucleótidos, de SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33 ó 35. El ácido nucleico codifica para un polipéptido que tiene actividad descarboxilasa y que procede de una cianobacteria.

25 En otro aspecto, en el presente documento se describe un ácido nucleico aislado que consiste en no más de aproximadamente el 99%, no más de aproximadamente el 98%, no más de aproximadamente el 97%, no más de aproximadamente el 96%, no más de aproximadamente el 95%, no más de aproximadamente el 94%, no más de aproximadamente el 93%, no más de aproximadamente el 92%, no más de aproximadamente el 91%, no más de aproximadamente el 90%, no más de aproximadamente el 85%, o no más de aproximadamente el 80% de los nucleótidos de SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33 ó 35. El ácido nucleico codifica para un polipéptido que tiene actividad descarboxilasa y que procede de una cianobacteria.

30 En otro aspecto, en el presente documento se describe un polipéptido aislado que consiste en no más de aproximadamente 200, no más de aproximadamente 175, no más de aproximadamente 150, o no más de aproximadamente 100 de los aminoácidos de SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34 ó 36. El polipéptido tiene actividad descarboxilasa y procede de una cianobacteria.

35 En otro aspecto, en el presente documento se describe un polipéptido aislado que consiste en no más de aproximadamente el 99%, no más de aproximadamente el 98%, no más de aproximadamente el 97%, no más de aproximadamente el 96%, no más de aproximadamente el 95%, no más de aproximadamente el 94%, no más de aproximadamente el 93%, no más de aproximadamente el 92%, no más de aproximadamente el 91%, no más de aproximadamente el 90%, no más de aproximadamente el 85%, o no más de aproximadamente el 80% de los aminoácidos de SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34 ó 36. El polipéptido tiene actividad descarboxilasa y procede de una cianobacteria.

Definiciones

50 A lo largo de toda la memoria descriptiva, puede hacerse referencia usando un nombre de polipéptido o nombre de gen abreviado, pero se entiende que tal nombre de polipéptido o nombre de gen abreviado representa el género de genes o polipéptidos. Tales nombres de gen incluyen todos los genes que codifican para el mismo polipéptido y polipéptidos homólogos que tienen la misma función fisiológica. Los nombres de polipéptido incluyen todos los polipéptidos que tienen la misma actividad (por ejemplo, que catalizan la misma reacción química fundamental).

55 Los números de registro a los que se hace referencia en el presente documento se derivan de la base de datos de NCBI (National Center for Biotechnology Information) mantenida por el National Institute of Health, EE.UU. A menos que se indique lo contrario, los números de registro son tal como se proporcionan en la base de datos de abril de

2009.

Los números EC se establecen por el Nomenclature Committee de la International Union of Biochemistry and Molecular Biology (NC-IUBMB) (disponible en <http://www.chem.qmul.ac.uk/iubmb/enzyme/>). Los números EC a los que se hace referencia en el presente documento se derivan de la base de datos KEGG Ligand, mantenida por la Kyoto Encyclopedia of Genes and Genomics, subvencionada en parte por la Universidad de Tokio. A menos que se indique lo contrario, los números EC son tal como se proporcionan en la base de datos en marzo de 2008.

Los artículos “un” y “una” se usan en el presente documento para referirse a uno o a más de uno (es decir, a al menos uno) del objeto gramatical del artículo. A modo de ejemplo, “un elemento” significa un elemento o más de un elemento.

El término “aproximadamente” se usa en el presente documento para significar un valor $\pm 20\%$ de un valor numérico dado. Por tanto, “aproximadamente el 60%” significa un valor de entre $60 \pm (20\% \text{ de } 60)$ (es decir, entre 48 y 70).

Tal como se usa en el presente documento, el término “aldehído” significa un hidrocarburo que tiene la fórmula RCHO caracterizado por un grupo carbonilo insaturado (C=O). En una realización preferida, el aldehído es cualquier aldehído compuesto por un ácido graso o derivado de ácido graso. En una realización, el grupo R tiene al menos aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 ó 20 carbonos de longitud.

Tal como se usa en el presente documento, un “gen de biosíntesis de aldehídos” o un “polinucleótido de biosíntesis de aldehídos” es un ácido nucleico que codifica para un polipéptido de biosíntesis de aldehídos.

Tal como se usa en el presente documento, un “polipéptido de biosíntesis de aldehídos” es un polipéptido que forma parte de la ruta de biosíntesis de un aldehído. Tales polipéptidos pueden actuar sobre un sustrato biológico para producir un aldehído. El polipéptido de biosíntesis de aldehídos tiene actividad reductasa y procede de una cianobacteria.

Tal como se usa en el presente documento, el término “alcano” significa un hidrocarburo que contiene sólo enlaces carbono-carbono sencillos.

Tal como se usa en el presente documento, un “gen de biosíntesis de alcanos” o un “polinucleótido de biosíntesis de alcanos” es un ácido nucleico que codifica para un polipéptido de biosíntesis de alcanos.

Tal como se usa en el presente documento, un “polipéptido de biosíntesis de alcanos” es un polipéptido que forma parte de la ruta de biosíntesis de un alcano. Tales polipéptidos pueden actuar sobre un sustrato biológico para producir un alcano. El polipéptido de biosíntesis de alcanos tiene actividad descarboxilasa y procede de una cianobacteria.

Tal como se usa en el presente documento, un “gen de biosíntesis de alquenos” o un “polinucleótido de biosíntesis de alquenos” es un ácido nucleico que codifica para un polipéptido de biosíntesis de alquenos.

Tal como se usa en el presente documento, un “polipéptido de biosíntesis de alquenos” es un polipéptido que forma parte de la ruta de biosíntesis de un alqueno. Tales polipéptidos pueden actuar sobre un sustrato biológico para producir un alqueno. El polipéptido de biosíntesis de alquenos tiene actividad descarboxilasa y procede de una cianobacteria.

Tal como se usa en el presente documento, el término “atenuar” significa debilitar, reducir o disminuir. Por ejemplo, un polipéptido puede atenuarse modificando el polipéptido para reducir su actividad (por ejemplo, modificando una secuencia de nucleótidos que codifica para el polipéptido).

Tal como se usa en el presente documento, el término “biodiésel” significa un biocombustible que puede ser un sustituto de diésel, que se deriva de petróleo. Puede usarse biodiésel en motores diésel de combustión interna o bien en forma pura, que se denomina biodiésel “no mezclado”, o bien como mezcla en cualquier concentración con diésel a base de petróleo. Biodiésel puede incluir ésteres o hidrocarburos, tales como aldehídos y alcanos.

Tal como se usa en el presente documento, el término “biocombustible” se refiere a cualquier combustible derivado a partir de biomasa. Los combustibles a base de petróleo pueden sustituirse por biocombustibles. Por ejemplo, los biocombustibles incluyen combustibles de transporte (por ejemplo, gasolina, diésel, combustible para motores a reacción, etc.), combustibles para calefacción y combustibles para la generación de electricidad. Los biocombustibles son una fuente de energía renovable.

Tal como se usa en el presente documento, el término “biomasa” se refiere a una fuente de carbono derivada de material biológico. La biomasa puede convertirse en un biocombustible. Una fuente de biomasa a modo de ejemplo es materia vegetal. Por ejemplo puede usarse maíz, caña de azúcar o césped como biomasa. Otro ejemplo no limitativo de biomasa es materia animal, por ejemplo estiércol de vaca. La biomasa también incluye productos de desecho procedentes de la industria, agricultura, silvicultura y desechos domésticos. Ejemplos de tales productos de desecho que pueden usarse como biomasa son desechos de fermentación, paja, madera, aguas residuales, basura

y restos de comida. La biomasa también incluye fuentes de carbono, tales como hidratos de carbono (por ejemplo, monosacáridos, disacáridos o polisacáridos).

5 Tal como se usa en el presente documento, la expresión “fuente de carbono” se refiere a un sustrato o compuesto adecuado para usarse como fuente de carbono para el crecimiento de células procariotas 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 CO₂). Estas incluyen, por ejemplo, diversos monosacáridos, tales como glucosa, fructosa, manosa y galactosa; oligosacáridos, tales como fructo-oligosacárido y galacto-oligosacárido; polisacáridos tales como xilosa y arabinosa; disacáridos, tales como sacarosa, maltosa y turanosa; material celulósico, tal como metilcelulosa y carboximetilcelulosa de sodio; ésteres de ácidos grasos saturados o insaturados, tales como succinato, lactato y acetato; alcoholes, tales como etanol, o mezclas de los mismos. La fuente de carbono también puede ser un producto de fotosíntesis, incluyendo, pero sin limitarse a, glucosa. Una fuente de carbono preferida es biomasa. Otra fuente de carbono preferida es glucosa.

15 Tal como se usa en el presente documento, un “aditivo que reduce el punto de turbidez” es un aditivo añadido a una composición para disminuir o reducir el punto de turbidez de una disolución.

20 Tal como se usa en el presente documento, la expresión “punto de turbidez de un fluido” significa la temperatura a la que los sólidos disueltos ya no son completamente solubles. Por debajo de esta temperatura, los sólidos empiezan a precipitar como una segunda fase dando al fluido una apariencia turbia. En la industria petrolera, el punto de turbidez se refiere a la temperatura por debajo de la cual un material solidificado u otro hidrocarburo pesado cristaliza para dar un aceite crudo, aceite refinado o combustible para formar una apariencia turbia. La presencia de materiales solidificados influye en el comportamiento de flujo del fluido, la tendencia del fluido a obstruir los filtros de combustible, inyectores, etc., la acumulación de materiales solidificados sobre superficies frías (por ejemplo, incrustación de una tubería o un intercambiador de calor) y las características de emulsión del fluido con agua.

25 Una secuencia de nucleótidos es “complementaria” a otra secuencia de nucleótidos si cada una de las bases de las dos secuencias se aparea (es decir, puede formar pares de bases de Watson Crick). El término “hebra complementaria” se usa en el presente documento de manera intercambiable con el término “complemento”. El complemento de una hebra de ácido nucleico puede ser el complemento de una hebra codificante o el complemento de una hebra no codificante.

30 Tal como se usa en el presente documento, el término “condiciones suficientes para permitir la expresión” significa cualquier condición que permite a una célula huésped producir un producto deseado, tal como un polipéptido, aldehído o alcano descrito en el presente documento. Las condiciones adecuadas incluyen, por ejemplo, condiciones de fermentación. Las condiciones de fermentación pueden comprender muchos parámetros, tales como intervalos de temperatura, niveles de aireación y composición de medios. Cada una de estas condiciones, individualmente y en combinación, permite que la célula huésped crezca. Los medios de cultivo a modo de ejemplo incluyen caldos o geles. Generalmente, el medio incluye una fuente de carbono, tal como glucosa, fructosa, celulosa, o similares, 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.

40 Para determinar si las condiciones son suficientes para permitir la expresión, puede cultivarse una célula huésped, por ejemplo, durante aproximadamente 4, 8, 12, 24, 36 ó 48 horas. Durante y/o tras el cultivo, pueden obtenerse muestras y analizarse para determinar si las condiciones permiten la expresión. Por ejemplo, las células huésped en la muestra o el medio en el que se hicieron crecer las células huésped pueden someterse a prueba para determinar la presencia de un producto deseado. Cuando se somete a prueba para determinar la presencia de un producto, pueden usarse ensayos, tales como, pero sin limitarse a, CCF, HPLC, CG/FID, CG/EM, CL/EM, EM.

45 Se entiende que los polipéptidos descritos en el presente documento pueden tener sustituciones de aminoácidos conservativas o no esenciales adicionales, que no tienen un efecto sustancial sobre las funciones del polipéptido. Puede determinarse si una sustitución particular se tolerará (es decir, no afectará de manera adversa a las propiedades biológicas deseadas, tales como actividad descarboxilasa) tal como se describe en Bowie *et al.* Science (1990) 247:1306 1310. Una “sustitución de aminoácido conservativa” es una en la que el residuo de aminoácido se reemplaza por un residuo de aminoácido que tiene una cadena lateral similar. En la técnica se han definido las 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 no polares (por ejemplo, alanina, valina, leucina, isoleucina, prolina, fenilalanina, metionina, triptófano), cadenas laterales ramificadas en beta (por ejemplo, treonina, valina, isoleucina) y cadenas laterales aromáticas (por ejemplo, tirosina, fenilalanina, triptófano, histidina).

55 Tal como se usa en el presente documento, “elemento de control” significa un elemento de control de la transcripción. Los elementos de control incluyen promotores y potenciadores. El término “elemento promotor”, “promotor” o “secuencia promotora” se refiere a una secuencia de ADN que funciona como interruptor que activa la

expresión de un gen. Si se activa el gen, se dice que se transcribe o que participa en la transcripción. La transcripción implica la síntesis de ARNm a partir del gen. Por tanto, un promotor sirve como un elemento regulador de la transcripción y también proporciona un sitio para el inicio de la transcripción del gen para dar ARNm. Los elementos de control interaccionan específicamente con proteínas celulares implicadas en la transcripción (Maniatis *et al*, Science 236:1237, 1987).

Tal como se usa en el presente documento, el término “éster sintasa” significa un péptido que puede producir ésteres grasos. Más específicamente, una éster sintasa es un péptido que convierte un tioéster en un éster graso. En una realización preferida, la éster sintasa convierte un tioéster (por ejemplo, acil-CoA) en un éster graso.

En una realización alternativa, una éster sintasa usa un tioéster y un alcohol como sustratos para producir un éster graso. Las éster sintasas pueden usar tioésteres de cadena corta y larga como sustratos. Además, las éster sintasas pueden usar alcoholes de cadena corta y larga como sustratos.

Los ejemplos no limitativos de éster sintasas son cera sintasas, cera-éster sintasas, acil CoA:alcohol transacilasas, aciltransferasas y acil graso-coenzima A:alcohol graso aciltransferasas. Las éster sintasas a modo de ejemplo se clasifican en el número de clasificación de enzimas EC 2.3.1.75. En la figura 40 se proporcionan números de registro de GenBank a modo de ejemplo.

Tal como se usa en el presente documento, 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 estar saturados, monoinsaturados o poliinsaturados. En una realización preferida, el ácido graso se prepara a partir de una ruta de biosíntesis de ácidos grasos.

Tal como se usa en el presente documento, el término “ruta de biosíntesis de ácidos grasos” significa una ruta de biosíntesis que produce ácidos grasos. La ruta de biosíntesis de ácidos grasos incluye enzimas de ácido graso que pueden modificarse por ingeniería genética, tal como se describe en el presente documento, para producir ácidos grasos, y en algunas realizaciones pueden expresarse con enzimas adicionales para producir ácidos grasos que tienen características de la cadena de carbono deseadas.

Tal como se usa en el presente documento, el término “derivado de ácido graso” significa productos preparados en parte a partir de la ruta de biosíntesis de ácidos grasos del microorganismo huésped de producción. “Derivado de ácido graso” también incluye productos preparados en parte a partir de acil-ACP o derivados de acil-ACP. La ruta de biosíntesis de ácidos grasos incluye enzimas ácido graso sintasa que pueden modificarse por ingeniería genética tal como se describe en el presente documento para producir derivados de ácido graso, y en algunos ejemplos pueden expresarse con enzimas adicionales para producir derivados de ácido graso que tienen características de cadena de carbono deseadas. Los derivados de ácido graso a modo de ejemplo incluyen, por ejemplo, ácidos grasos, acil-CoA, aldehído graso, alcoholes de cadena corta y larga, hidrocarburos, alcoholes grasos y ésteres (por ejemplo, ceras, ésteres de ácidos grasos o ésteres grasos).

Tal como se usa en el presente documento, el término “enzimas de derivado de ácido graso” significa todas las enzimas que pueden expresarse o sobreexpresarse en la producción de derivados de ácido graso. Estas enzimas se denominan colectivamente en el presente documento enzimas de derivado de ácido graso. Estas enzimas pueden ser parte de la ruta de biosíntesis de ácidos grasos. Los ejemplos no limitativos de enzimas de derivado de ácido graso incluyen ácido graso sintasas, tioesterasas, acil-CoA sintasas, acil-CoA reductasas, alcohol deshidrogenasas, alcohol aciltransferasas, acil-CoA reductasa que forma alcohol graso, éster sintasas, polipéptidos de biosíntesis de aldehídos y polipéptidos de biosíntesis de alcanos. Las enzimas de derivado de ácido graso convierten un sustrato en un derivado de ácido graso. En algunos ejemplos, el sustrato puede ser un derivado de ácido graso que la enzima de derivado de ácido graso convierte en un derivado de ácido graso diferente.

Tal como se usa en el presente documento, el término “péptidos que forman alcohol graso” significa un péptido que puede catalizar la conversión de acil-CoA en alcohol graso, incluyendo acil-CoA reductasa que forma alcohol graso (FAR, EC 1.1.1.*), acil-CoA reductasa (EC 1.2.1.50) o alcohol deshidrogenasa (EC 1.1.1.1). Adicionalmente, un experto habitual en la técnica apreciará que algunos péptidos que forman alcohol graso catalizarán también otras reacciones. Por ejemplo, algunos péptidos de acil-CoA reductasa aceptarán otros sustratos además de ácidos grasos. Por tanto, también se incluyen tales péptidos no específicos. Se conocen en la técnica secuencias de ácido que codifican para péptidos que forman alcohol graso, y tales péptidos están disponibles públicamente. En la figura 40 se proporcionan números de registro de GenBank a modo de ejemplo.

Tal como se usa en el presente documento, “enzima de ácido graso” significa cualquier enzima implicada en la biosíntesis de ácidos grasos. Pueden expresarse o sobreexpresarse enzimas de ácido graso en células huésped para producir ácidos grasos. Los ejemplos no limitativos de enzimas de ácido graso incluyen ácido graso sintasas y tioesterasas.

Tal como se usa en el presente documento, el término “éster graso” significa un éster. En una realización preferida, un éster graso es cualquier éster compuesto por un ácido graso, por ejemplo un éster de ácido graso. En una realización, un éster graso contiene un lado A (es decir, la cadena de carbono unida al oxígeno del carboxilato) y un

- lado B (es decir, la cadena de carbono que comprende el carboxilato original). En una realización preferida, cuando el éster graso se deriva de la ruta de biosíntesis de ácidos grasos, al lado A contribuye un alcohol, y al lado B contribuye un ácido graso. Puede usarse cualquier alcohol para formar el lado A de los ésteres grasos. Por ejemplo, el alcohol puede derivarse de la ruta de biosíntesis de ácidos grasos. Alternativamente, el alcohol puede producirse a través de rutas de biosíntesis que no son de ácidos grasos. Además, el alcohol puede proporcionarse de manera exógena. Por ejemplo, el alcohol puede suministrarse en el caldo de fermentación en casos en los que el éster graso se produce por un microorganismo. Alternativamente, puede suministrarse de manera exógena un ácido carboxílico, tal como un ácido graso o ácido acético, en casos en los que el éster graso se produce por un microorganismo que también puede producir alcohol.
- Las cadenas de carbono que comprenden el lado A o el lado B pueden ser de cualquier longitud. En una realización, el lado A del éster tiene al menos aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 10, 12, 14, 16 ó 18 carbonos de longitud. El lado B del éster tiene al menos aproximadamente 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24 ó 26 carbonos de longitud. El lado A y/o el lado B pueden ser de cadena lineal o ramificada. Las cadenas ramificadas pueden tener uno o más puntos de ramificación. Además, las cadenas ramificadas pueden incluir ramificaciones cíclicas. Además, el lado A y/o el lado B puede estar saturado o insaturado. Si está insaturado, el lado A y/o el lado B puede tener uno o más puntos de insaturación.
- En una realización, el éster graso se produce de manera biosintética. En esta realización, en primer lugar el ácido graso se "activa". Ejemplos no limitativos de ácidos grasos "activados" son acil-CoA, acil-ACP y fosfato de acilo. Acil-CoA puede ser un producto directo de la biosíntesis o degradación de ácidos grasos. Además, acil-CoA puede sintetizarse a partir de un ácido graso libre, una CoA o un trifosfato de nucleótido de adenosina (ATP). Un ejemplo de una enzima que produce acil-CoA es acil-CoA sintasa.
- Tras activarse el ácido graso, puede transferirse fácilmente a un nucleófilo receptor. Nucleófilos a modo de ejemplo son alcoholes, tioles o fosfatos.
- En una realización, el éster graso es una cera. La cera puede derivarse de un alcohol de cadena larga y un ácido graso de cadena larga. En otra realización, el éster graso puede derivarse de un acil graso-tioéster y un alcohol. En otra realización, el éster graso es un tioéster de ácido graso, por ejemplo acil graso-coenzima A (CoA). En otras realizaciones, el éster graso es un pantotenato de acilo graso, una proteína portadora de acilo (ACP) o un éster de fosfato graso. Los ésteres grasos tienen muchos usos. Por ejemplo, pueden usarse ésteres grasos como biocombustible.
- Tal como se usa en el presente documento, "fracción de carbono moderno" o " f_M " tiene el mismo significado que el definido por el National Institute of Standards and Technology (NIST), Standard Reference Materials (SRM) 4990B y 4990C, conocido como patrones de ácidos oxálicos HOxI y HOxII, respectivamente. La definición fundamental se refiere a 0,95 veces la razón de isótopos $^{14}\text{C}/^{12}\text{C}$ de HOxI (en referencia a AD 1950). Esto es aproximadamente equivalente a madera anterior a la revolución industrial corregida para la descomposición. Para la biosfera viva actual (material vegetal), f_M es de aproximadamente 1,1.
- Pueden realizarse cálculos de "homología" entre dos secuencias tal como sigue. Las secuencias se alinean con fines de comparación óptima (por ejemplo, pueden introducirse huecos en uno o ambos de una primera y una segunda secuencia de aminoácidos o ácido nucleico para el alineamiento óptimo y puede no tenerse en cuenta las secuencias no homólogas con fines de comparación). En una realización preferida, la longitud de una secuencia de referencia que se alinea con fines de comparación es de al menos aproximadamente el 30%, preferiblemente al menos aproximadamente del 40%, más preferiblemente al menos aproximadamente del 50%, incluso más preferiblemente al menos aproximadamente del 60%, e incluso más preferiblemente al menos aproximadamente del 70%, al menos aproximadamente del 80%, al menos aproximadamente del 90% o aproximadamente del 100% de la longitud de la secuencia de referencia. Entonces se comparan los residuos de aminoácido o nucleótidos en las posiciones de aminoácido o posiciones de nucleótido correspondientes. 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 (tal como se usa en el presente documento, "identidad" de aminoácido o ácido nucleico es equivalente a "homología" de aminoácido o ácido nucleico). El porcentaje de identidad entre las dos secuencias es una función del número de posiciones idénticas que comparten las secuencias, teniendo en cuenta el número de huecos y la longitud de cada hueco, que es necesario introducir para el alineamiento óptimo de las dos secuencias.
- La comparación de secuencias y la determinación del porcentaje de homología entre dos secuencias pueden lograrse usando un algoritmo matemático. En una realización preferida, el porcentaje de homología entre dos secuencias de aminoácidos se determina usando el algoritmo de Needleman y Wunsch (1970), J. Mol. Biol. 48:444-453, que se ha incorporado en el programa GAP en el paquete de software 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. Aún en otra realización preferida, el porcentaje de homología entre dos secuencias de nucleótidos se determina usando el programa GAP en el paquete de software 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 conjunto particularmente preferido de parámetros (y el que debería usarse si el técnico no está seguro de qué parámetros deben aplicarse para determinar si una

molécula está dentro de una limitación de homología de las reivindicaciones) son la matriz de puntuación Blossum 62 con una penalización de hueco de 12, una penalización de extensión de hueco de 4 y una penalización de hueco de desplazamiento de marco de 5.

5 Tal como se usa en el presente documento, una “célula huésped” es una célula usada para producir un producto descrito en el presente documento (por ejemplo, un aldehído o alcano descrito en el presente documento). Una célula huésped puede modificarse para expresar o sobreexpresar genes seleccionados o para que tenga una expresión atenuada de genes seleccionados. Los ejemplos no limitativos de células huésped incluyen células vegetales, animales, humanas, de bacterias, levaduras u hongos filamentosos.

10 Tal como se usa en el presente documento, el término “se hibrida en condiciones de baja rigurosidad, rigurosidad media, alta rigurosidad o muy alta rigurosidad” 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. En esta referencia se describen métodos acuosos y no acuosos y puede usarse cualquiera de los métodos. Condiciones de hibridación específicas a las que se hace referencia en el presente documento son tal como siguen: 1) condiciones de hibridación de baja rigurosidad en 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 las condiciones de baja rigurosidad); 2) condiciones de hibridación de rigurosidad media en 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 alta rigurosidad en 6X SSC a aproximadamente 45°C, seguido por uno o más lavados en 0,2X SSC, SDS al 0,1% a 65°C; y preferiblemente 4) condiciones de hibridación de muy alta rigurosidad son fosfato de sodio 0,5 M, SDS al 7% a 65°C, seguido por uno o más lavados a 0,2X SSC, SDS al 1% a 65°C. Las condiciones de muy alta rigurosidad (4) son las condiciones preferidas a menos que se especifique lo contrario.

25 El término “aislado” tal como se usa en el presente documento con respecto a ácidos nucleicos, tales como ADN o ARN, se refiere a moléculas separadas de otros ADN o ARN, respectivamente, que están presentes en la fuente natural del ácido nucleico. Además, un “ácido nucleico aislado” incluye fragmentos de ácido nucleico, tales como fragmentos que no se producen de manera natural. El término “aislado” también se usa en el presente documento para referirse a polipéptidos, que se aíslan de otras proteínas celulares y abarca tanto polipéptidos purificados endógenos como polipéptidos recombinantes. El término “aislado” tal como se usa en el presente documento también se refiere a un ácido nucleico o polipéptido que está sustancialmente libre de material celular, material viral o medio de cultivo cuando se produce mediante técnicas de ADN recombinante. El término “aislado” tal como se usa en el presente documento también se refiere a un ácido nucleico o polipéptido que está sustancialmente libre de precursores químicos u otros productos químicos cuando se sintetiza químicamente.

35 Tal como se usa en el presente documento, el “nivel de expresión de un gen en una célula” se refiere al nivel de ARNm, transcripto(s) incipiente(s) pre-ARNm, productos intermedios de procesamiento de transcritos, ARNm maduro(s) y productos de degradación codificados por el gen en la célula.

Tal como se usa en el presente documento, el término “microorganismo” significa especies microbianas procariontas y eucariontas de los dominios *Archaea*, *Bacteria* y *Eucarya*, incluyendo este último levaduras y hongos filamentosos, protozoos, algas o protistas superiores. El término “células microbianas”, tal como se usa en el presente documento, significa una célula de un microorganismo.

40 Tal como se usa en el presente documento, el término “ácido nucleico” se refiere a polinucleótidos, tales como ácido desoxirribonucleico (ADN) y, cuando sea apropiado, ácido ribonucleico (ARN). El término también incluye análogos de cualquiera de ARN o ADN preparados a partir de análogos de nucleótidos, y tal como puede aplicarse a la realización que está describiéndose, polinucleótidos mono (sentido o antisentido) y bicatenarios, EST, cromosomas, ADNc, ARNm y ARNr.

45 Tal como se usa en el presente documento, el término “operativamente unido” significa que una secuencia de nucleótidos seleccionada (por ejemplo, que codifica para un polipéptido descrito en el presente documento) está en proximidad con un promotor para permitir que el promotor regule la expresión de la secuencia de nucleótidos seleccionada. Además, el promotor se localiza en el sentido de 5' de la secuencia de nucleótidos seleccionada en cuanto al sentido de transcripción y traducción. Por “operativamente unido” quiere decirse que una secuencia de nucleótidos y una(s) secuencia(s) reguladora(s) están conectadas de tal manera como para permitir la expresión génica cuando las moléculas apropiadas (por ejemplo, proteínas activadoras de la transcripción) se unen a la(s) secuencia(s) reguladora(s).

El término “o” se usa en el presente documento para significar, y se usa de manera intercambiable con, el término “y/o”, a menos que el contexto indique claramente lo contrario.

55 Tal como se usa en el presente documento, “sobreexpresar” significa expresar o provocar que se exprese un ácido nucleico, polipéptido o hidrocarburo en una célula a una concentración superior a la que se expresa normalmente en una célula silvestre correspondiente. Por ejemplo, un polipéptido puede “sobreexpresarse” en una célula huésped recombinante cuando el polipéptido está presente en una concentración superior en la célula huésped recombinante

en comparación con su concentración en una célula huésped no recombinante de la misma especie.

Tal como se usa en el presente documento, "coeficiente de reparto" o "P", se define como la concentración en equilibrio de un compuesto en una fase orgánica dividida entre la concentración en equilibrio en una fase acuosa (por ejemplo, caldo de fermentación). En una realización de un sistema bifásico descrita en el presente documento, la fase orgánica se forma mediante el aldehído o alcano durante el procedimiento de producción. Sin embargo, en algunos ejemplos, puede proporcionarse una fase orgánica, tal como proporcionando una fase de octano, para facilitar la separación del producto. Cuando se describe un sistema de dos fases, las características de reparto de un compuesto pueden describirse como logP. Por ejemplo, un compuesto con un logP de 1 se repartirá 10:1 en la fase orgánica. Un compuesto con un logP de -1 se repartirá 1:10 en la fase orgánica. Eligiendo un caldo de fermentación y una fase orgánica apropiados, un aldehído o alcano con un valor de logP alto puede separarse en la fase orgánica incluso a concentraciones muy bajas en el recipiente de fermentación.

Tal como se usa en el presente documento, el término "purificar", "purificado" o "purificación" significa la retirada o el aislamiento de una molécula de su entorno mediante, por ejemplo, aislamiento o separación. Las moléculas "sustancialmente purificadas" están libres al menos aproximadamente al 60%, preferiblemente libres al menos aproximadamente al 75% y más preferiblemente libres al menos aproximadamente al 90% de otros componentes con los que están asociadas. Tal como se usa en el presente documento, estos términos también se refieren a la eliminación de contaminantes de una muestra. Por ejemplo, la eliminación de contaminantes puede dar como resultado un aumento en el porcentaje de aldehídos o alcanos en una muestra. Por ejemplo, cuando se producen aldehídos o alcanos en una célula huésped, los aldehídos o alcanos pueden purificarse mediante la eliminación de proteínas de la célula huésped. Tras la purificación, aumenta el porcentaje de aldehídos o alcanos en la muestra.

Los términos "purificar", "purificado" y "purificación" no requieren pureza absoluta. Son términos relativos. Por tanto, por ejemplo, cuando se producen aldehídos o alcanos en células huésped, un aldehído purificado o alcano purificado es uno que está sustancialmente separado de otros componentes celulares (por ejemplo, ácidos nucleicos, polipéptidos, lípidos, hidratos de carbono u otros hidrocarburos). En otro ejemplo, una preparación de aldehído purificado o alcano purificado es una en la que el aldehído o alcano está sustancialmente libre de contaminantes, tales como los que pueden estar presentes tras la fermentación. En algunas realizaciones, un aldehído o un alcano está purificado cuando al menos aproximadamente el 50% en peso de una muestra está compuesto por el aldehído o alcano. En otras realizaciones, un aldehído o un alcano está purificado cuando al menos aproximadamente el 60%, el 70%, el 80%, el 85%, el 90%, el 92%, el 95%, el 98% o el 99% o más en peso de una muestra está compuesto por el aldehído o alcano.

Tal como se usa en el presente documento, el término "polipéptido recombinante" se refiere a un polipéptido que se produce mediante técnicas de ADN recombinante, en las que de manera general se inserta ADN que codifica para la el ARN o polipéptido expresado en un vector de expresión adecuado y que a su vez se usa para transformar una célula huésped para producir el polipéptido o ARN.

Tal como se usa en el presente documento, el término "sustancialmente idéntico" (o "sustancialmente homólogo") se usa para referirse a una primera secuencia de aminoácidos o de nucleótidos que contiene un número suficiente de residuos de aminoácido o nucleótidos idénticos o equivalentes (por ejemplo, con una cadena lateral similar, por ejemplo, sustituciones de aminoácido conservadas) a una segunda secuencia de aminoácidos o de nucleótidos de manera que la primera y la segunda secuencias de aminoácidos o de nucleótidos tienen actividades similares.

Tal como se usa en el presente documento, el término "sintasa" significa una enzima que cataliza un proceso de síntesis. Tal como se usa en el presente documento, el término sintasa incluye sintasas, sintetasas y ligasas.

Tal como se usa en el presente documento, el término "transfección" significa la introducción de un ácido nucleico (por ejemplo, mediante un vector de expresión) en una célula receptora mediante transferencia de genes mediada por ácido nucleico.

Tal como se usa en el presente documento, "transformación" se refiere a un proceso en el que se cambia el genotipo de una célula como resultado de la captación celular de ácido nucleico exógeno. Esto puede dar como resultado que la célula transformada exprese una forma recombinante de un ARN o polipéptido. En el caso de expresión antisentido a partir del gen transferido, se altera la expresión de una forma que se produce de manera natural del polipéptido.

Tal como se usa en el presente documento, una "proteína de transporte" es un polipéptido que facilita el movimiento de uno o más compuestos al interior y/o al exterior de un orgánulo celular y/o una célula.

Tal como se usa en el presente documento, una "variante" de polipéptido X se refiere a un polipéptido que tiene la secuencia de aminoácidos de polipéptido X en la que uno o más residuos de aminoácido están alterados. La variante puede tener cambios conservativos o cambios no conservativos. Puede encontrarse orientación en la determinación de qué residuos de aminoácido pueden sustituirse, insertarse o delecionarse sin afectar a la actividad biológica usando programas informáticos bien conocidos en la técnica, por ejemplo, el software LASERGENE (DNASTAR).

El término "variante", cuando se usa en el contexto de una secuencia de polinucleótido, puede abarcar una secuencia de polinucleótido relacionada con la de un gen o la secuencia codificante del mismo. Esta definición también puede incluir, por ejemplo, variantes "alélicas", "de corte y empalme", "de especies" o "polimórficas". Una variante de corte y empalme puede tener identidad significativa con respecto a un polinucleótido de referencia, pero generalmente tendrá un número mayor o menor de polinucleótidos debido al corte y empalme alternativo de exones durante el procesamiento de ARNm. El polipéptido correspondiente puede presentar dominios funcionales adicionales o una ausencia de dominios. Las variantes de especies son secuencias de polinucleótido que varían de una especie a otra. Los polipéptidos resultantes tendrán generalmente identidad de aminoácido significativa unos con respecto a otros. Una variante polimórfica es una variación en la secuencia de polinucleótido de un gen particular entre individuos de una especie dada.

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 al que se ha unido. Un tipo de vector útil es un episoma (es decir, un ácido nucleico que puede realizar replicación fuera de los cromosomas). Los vectores útiles son los que pueden realizar replicación y/o expresión autónoma de ácidos nucleicos a los que están unidos. Los vectores que pueden dirigir la expresión de genes a los que están operativamente unidos se denominan en el presente documento "vectores de expresión". En general, vectores de expresión de utilidad en técnicas de ADN recombinante están a menudo en forma de "plásmidos", que se refieren generalmente a bucles de ADN bicatenario circular que, en su forma de vector, no están unidos al cromosoma. En la presente memoria descriptiva, "plásmido" y "vector" se usan de manera intercambiable, ya que el plásmido es la forma de vector más comúnmente usada. Sin embargo, también se incluyen otras formas de vectores de expresión que sirvan para funciones equivalentes y que lleguen a conocerse en la técnica posteriormente a esto.

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 esta invención. Aunque pueden usarse métodos y materiales similares o equivalentes a los descritos en el presente documento en la práctica o las pruebas de la presente invención, se describen a continuación métodos y materiales adecuados. En caso de conflicto, prevalecerá la presente memoria descriptiva, incluyendo las definiciones. Además, los materiales, métodos y ejemplos sólo son ilustrativos y no pretenden ser limitativos.

Otras características y ventajas de la invención resultarán evidentes a partir de la siguiente descripción detallada y a partir de las reivindicaciones.

30 Breve descripción de los dibujos

La figura 1A es un perfil de CG/EM de hidrocarburos producidos por células de *Prochlorococcus marinus* CCMP1986. La figura 1B es un patrón de fragmentación de masas del pico a 7,55 min de la figura 1A.

La figura 2A es un perfil de CG/EM de hidrocarburos producidos por células de *Nostoc punctiforme* PCC73102. La figura 2B es un patrón de fragmentación de masas del pico a 8,73 min de la figura 2A.

35 La figura 3A es un perfil de CG/EM de hidrocarburos producidos por células de *Gloeobaceter violaceus* ATCC29082. La figura 3B es un patrón de fragmentación de masas del pico a 8,72 min de la figura 3A.

La figura 4A es un perfil de CG/EM de hidrocarburos producidos por células de *Synechocystic sp.* PCC6803. La figura 4B es un patrón de fragmentación de masas del pico a 7,36 min de la figura 4A.

40 La figura 5A es un perfil de CG/EM de hidrocarburos producidos por células silvestres de *Synechocystis sp.* PCC6803. La figura 5B es un perfil de CG/EM de hidrocarburos producidos por células de *Synechocystis sp.* PCC6803 con una delección de los genes *slI0208* y *slI0209*.

La figura 6A es un perfil de CG/EM de hidrocarburos producidos por células silvestres de *E. coli* MG1655. La figura 6B es un perfil de CG/EM de hidrocarburos producidos por células de *E. coli* MG1655 que expresan YP_400611 de *Synechococcus elongatus* PCC7942 (Synpcc7942_1594) (SEQ ID NO: 65).

45 La figura 7 es un perfil de CG/EM de hidrocarburos producidos por células de *E. coli* que expresan *cce_1430* de *Cyanothece sp.* ATCC51142 (YP_001802846) (SEQ ID NO: 69).

La figura 8A es un perfil de CG/EM de hidrocarburos producidos por células de *E. coli* que expresan YP_400611 de *Synechococcus elongatus* PCC7942 (Synpcc7942_1594) (SEQ ID NO: 65) e YP_400610 de *Synechococcus elongatus* PCC7942 (Synpcc7942_1593) (SEQ ID NO: 1). La figura 8B representa patrones de fragmentación de masas del pico a 6,98 min de la figura 8A y de pentadecano. La figura 8C representa patrones de fragmentación de masas del pico a 8,12 min de la figura 8A y de 8-heptadeceno.

50 La figura 9 es un perfil de CG/EM de hidrocarburos producidos por células de *E. coli* MG1655 que expresan YP_400611 de *Synechococcus elongatus* PCC7942 (Synpcc7942_1594) (SEQ ID NO: 65) y Npun02004178 de *Nostoc punctiforme* PCC73102 (ZP_00108838) (SEQ ID NO: 5).

ES 2 667 469 T3

- La figura 10 es un perfil de CG/EM de hidrocarburos producidos por células de *E. coli* MG1655 que expresan YP_400611 de *Synechococcus elongatus* PCC7942 (Synpcc7942_1594) (SEQ ID NO: 65) y sll0208 de *Synechocystis* sp. PCC6803 (NP_442147) (SEQ ID NO: 3).
- 5 La figura 11 es un perfil de CG/EM de hidrocarburos producidos por células de *E. coli* MG1655 que expresan YP_400611 de *Synechococcus elongatus* PCC7942 (Synpcc7942_1594) (SEQ ID NO: 65) y alr5283 de *Nostoc* sp. PCC7210 (NP_489323) (SEQ ID NO: 7).
- La figura 12 es un perfil de CG/EM de hidrocarburos producidos por células de *E. coli* MG1655 que expresan YP_400611 de *Synechococcus elongatus* PCC7942 (Synpcc7942_1594) (SEQ ID NO: 65) y AM1_4041 de *Acaryochloris marina* MBIC11017 con codones optimizados (YP_001518340) (SEQ ID NO: 46).
- 10 La figura 13 es un perfil de CG/EM de hidrocarburos producidos por células de *E. coli* MG1655 que expresan YP_400611 de *Synechococcus elongatus* PCC7942 (Synpcc7942_1594) (SEQ ID NO: 65) y tll1313 de *Thermosynechococcus elongatus* BP-1 con codones optimizados (NP_682103) (SEQ ID NO: 47).
- La figura 14 es un perfil de CG/EM de hidrocarburos producidos por células de *E. coli* MG1655 que expresan YP_400611 de *Synechococcus elongatus* PCC7942 (Synpcc7942_1594) (SEQ ID NO: 65) y CYA_0415 de *Synechococcus* sp. JA-3-3Ab con codones optimizados (YP_473897) (SEQ ID NO: 48).
- 15 La figura 15 es un perfil de CG/EM de hidrocarburos producidos por células de *E. coli* MG1655 que expresan YP_400611 de *Synechococcus elongatus* PCC7942 (Synpcc7942_1594) (SEQ ID NO: 65) y gll3146 de *Gloeobacter violaceus* PCC7421 (NP_926092) (SEQ ID NO: 15).
- La figura 16 es un perfil de CG/EM de hidrocarburos producidos por células de *E. coli* MG1655 que expresan YP_400611 de *Synechococcus elongatus* PCC7942 (Synpcc7942_1594) (SEQ ID NO: 65) y PMT1231 de *Prochlorococcus marinus* MIT9313 con codones optimizados (NP_895059) (SEQ ID NO: 49).
- 20 La figura 17 es un perfil de CG/EM de hidrocarburos producidos por células de *E. coli* MG1655 que expresan YP_400611 de *Synechococcus elongatus* PCC7942 (Synpcc7942_1594) (SEQ ID NO: 65) y PMM0532 de *Prochlorococcus marinus* CCMP1986 (NP_892650) (SEQ ID NO: 19).
- 25 La figura 18 es un perfil de CG/EM de hidrocarburos producidos por células de *E. coli* MG1655 que expresan YP_400611 de *Synechococcus elongatus* PCC7942 (Synpcc7942_1594) (SEQ ID NO: 65) y PMN2A_1863 de *Prochlorococcus marinus* NATL2A con codones optimizados (YP_293054) (SEQ ID NO: 51).
- La figura 19 es un perfil de CG/EM de hidrocarburos producidos por células de *E. coli* MG1655 que expresan YP_400611 de *Synechococcus elongatus* PCC7942 (Synpcc7942_1594) (SEQ ID NO: 65) y RS9917_09941 de *Synechococcus* sp. RS9917 con codones optimizados (ZP_01079772) (SEQ ID NO: 52).
- 30 La figura 20 es un perfil de CG/EM de hidrocarburos producidos por células de *E. coli* MG1655 que expresan YP_400611 de *Synechococcus elongatus* PCC7942 (Synpcc7942_1594) (SEQ ID NO: 65) y RS9917_12945 de *Synechococcus* sp. RS9917 con codones optimizados (ZP_01080370) (SEQ ID NO: 53).
- La figura 21 es un perfil de CG/EM de hidrocarburos producidos por células de *E. coli* MG1655 que expresan YP_400611 de *Synechococcus elongatus* PCC7942 (Synpcc7942_1594) (SEQ ID NO: 65) y cce_0778 de *Cyanothece* sp. ATCC51142 (YP_001802195) (SEQ ID NO: 27).
- 35 La figura 22 es un perfil de CG/EM de hidrocarburos producidos por células de *E. coli* MG1655 que expresan YP_400611 de *Synechococcus elongatus* PCC7942 (Synpcc7942_1594) (SEQ ID NO: 65) y Cyan7425_0398 de *Cyanothece* sp. PCC7425 (YP_002481151) (SEQ ID NO: 29).
- 40 La figura 23 es un perfil de CG/EM de hidrocarburos producidos por células de *E. coli* MG1655 que expresan YP_400611 de *Synechococcus elongatus* PCC7942 (Synpcc7942_1594) (SEQ ID NO: 65) y Cyan7425_2986 de *Cyanothece* sp. PCC7425 (YP_002483683) (SEQ ID NO: 31).
- La figura 24A es un perfil de CG/EM de hidrocarburos producidos por células de *E. coli* MG1655 que expresan PMM0533 de *Prochlorococcus marinus* CCMP1986 (NP_892651) (SEQ ID NO: 71). La figura 24B es un perfil de CG/EM de hidrocarburos producidos por células de *E. coli* MG1655 que expresan PMM0533 de *Prochlorococcus marinus* CCMP1986 (NP_892651) (SEQ ID NO: 71) y PMM0532 de *Prochlorococcus marinus* CCMP1986 (NP_892650) (SEQ ID NO: 19).
- 45 La figura 25A es un perfil de CG/EM de hidrocarburos producidos por células de *E. coli* MG1655 $\Delta fadE lacZ::P_{trc}$ *'tesA-fadD*. La figura 25B es un perfil de CG/EM de hidrocarburos producidos por células de *E. coli* MG1655 $\Delta fadE lacZ::P_{trc}$ *'tesA-fadD* que expresan YP_400611 de *Synechococcus elongatus* PCC7942 (Synpcc7942_1594) (SEQ ID NO: 65) y AM1_4041 de *Acaryochloris marina* MBIC11017 (YP_001518340) (SEQ ID NO: 9).
- 50 La figura 26A es un perfil de CG/EM de hidrocarburos producidos por células de *E. coli* MG1655 $\Delta fadE lacZ::P_{trc}$ *'tesA-fadD* que expresan sll0209 de *Synechocystis* sp. PCC6803 (NP_442146) (SEQ ID NO: 67). La figura 26B es

- un perfil de CG/EM de hidrocarburos producidos por células de *E. coli* MG1655 $\Delta fadE$ *lacZ::P_{trc}'tesA-fadD* que expresan sll0209 de *Synechocystis* sp. PCC6803 (NP_442146) (SEQ ID NO: 67) y sll0208 de *Synechocystis* sp. PCC6803 (NP_442147) (SEQ ID NO: 3).
- 5 La figura 27A es un perfil de CG/EM de hidrocarburos producidos por células de *E. coli* MG1655 $\Delta fadD$ *lacZ::P_{trc}'tesA* que expresan MSMEG_5739 de *M. smegmatis* cepa MC2 155 (YP_889972) (SEQ ID NO: 85). La figura 27B es un perfil de CG/EM de hidrocarburos producidos por células de *E. coli* MG1655 $\Delta fadD$ *lacZ::P_{trc}'tesA* que expresan MSMEG_5739 de *M. smegmatis* cepa MC2 155 (YP_889972) (SEQ ID NO: 85) y Npun02004178 de *Nostoc punctiforme* PCC73102 (ZP_00108838) (SEQ ID NO: 5).
- 10 La figura 28 es una representación gráfica de hidrocarburos producidos por células de *E. coli* MG1655 $\Delta fadD$ *lacZ::P_{trc}'tesA* que expresan MSMEG_5739 de *M. smegmatis* cepa MC2 155 (YP_889972) (SEQ ID NO: 85) o bien solas o bien en combinación con alr5283 de *Nostoc* sp. PCC7120 (SEQ ID NO: 7), Npun02004178 de *Nostoc punctiforme* PCC73102 (SEQ ID NO: 5), PMM0532 de *P. marinus* CCMP1986 (SEQ ID NO: 19), gl13146 de *G. violaceus* PCC7421 (SEQ ID NO: 15), RS9917_09941 de *Synechococcus* sp. (SEQ ID NO: 23), RS9917_12945 de *Synechococcus* sp. (SEQ ID NO: 25) o AM1_4041 de *A. marina* MBIC11017 (SEQ ID NO: 9).
- 15 La figura 29A es una representación de la estructura tridimensional de una proteína de subunidad β de ribonucleasa reductasa de clase I, RNR β . La figura 29B es una representación de la estructura tridimensional de PMT1231 de *Prochlorococcus marinus* MIT9313 (NP_895059) (SEQ ID NO: 17). La figura 29C es una representación de la estructura tridimensional del sitio activo de PMT1231 de *Prochlorococcus marinus* MIT9313 (NP_895059) (SEQ ID NO: 17).
- 20 La figura 30A es un perfil de CG/EM de hidrocarburos producidos por células de *E. coli* MG1655 que expresan Npun02004178 de *Nostoc punctiforme* PCC73102 (ZP_00108838) (SEQ ID NO: 5). La figura 30B es un perfil de CG/EM de hidrocarburos producidos por células de *E. coli* MG1655 que expresan la variante Y123F de Npun02004178 de *Nostoc punctiforme* PCC73102 (ZP_00108838). La figura 30C es un perfil de CG/EM de hidrocarburos producidos por células de *E. coli* MG1655 que expresan la variante Y126F de Npun02004178 de *Nostoc punctiforme* PCC73102 (ZP_00108838).
- 25 La figura 31 representa perfiles de CG/EM de hidrocarburos producidos *in vitro* usando Npun02004178 de *Nostoc punctiforme* PCC73102 (ZP_00108838) (SEQ ID NO: 6) y octadecanal (A); Npun02004178 (ZP_00108838) (SEQ ID NO: 6), octadecanal, ferredoxina reductasa de espinaca y NADPH (B); octadecanal, ferredoxina de espinaca, ferredoxina reductasa de espinaca y NADPH (C); o Npun02004178 (ZP_00108838) (SEQ ID NO: 6), ferredoxina de espinaca y ferredoxina de espinaca (D).
- 30 La figura 32 representa perfiles de CG/EM de hidrocarburos producidos *in vitro* usando Npun02004178 de *Nostoc punctiforme* PCC73102 (ZP_00108838) (SEQ ID NO: 6), NADPH, octadecanal y o bien (A) ferredoxina de espinaca y ferredoxina reductasa de espinaca; (B) Npun02003626 de *N. punctiforme* PCC73102 (ZP_00109192) (SEQ ID NO: 88) y Npun02001001 de *N. punctiforme* PCC73102 (ZP_00111633) (SEQ ID NO: 90); (C) Npun02003626 (ZP_00109192) (SEQ ID NO: 88) y Npun02003530 de *N. punctiforme* PCC73102 (ZP_00109422) (SEQ ID NO: 92); o bien (D) Npun02003626 (ZP_00109192) (SEQ ID NO: 88) y Npun02003123 de *N. punctiforme* PCC73102 (ZP_00109501) (SEQ ID NO: 94).
- 35 La figura 33A es un perfil de CG/EM de hidrocarburos producidos *in vitro* usando octadecanoil-CoA, YP_400611 de *Synechococcus elongatus* PCC7942 (Synpcc7942_1594) (SEQ ID NO: 66), NADH y Mg²⁺. La figura 33B es un perfil de CG/EM de hidrocarburos producidos *in vitro* usando octadecanoil-CoA, YP_400611 de *Synechococcus elongatus* PCC7942 (Synpcc7942_1594) (SEQ ID NO: 66), NADPH y Mg²⁺. La figura 33C es un perfil de CG/EM de hidrocarburos producidos *in vitro* usando octadecanoil-CoA, YP_400611 de *Synechococcus elongatus* PCC7942 (Synpcc7942_1594) (SEQ ID NO: 66) y NADPH.
- 40 La figura 34A es un perfil de CG/EM de hidrocarburos producidos *in vitro* usando octadecanoil-CoA, NADPH marcado, YP_400611 de *Synechococcus elongatus* PCC7942 (Synpcc7942_1594) (SEQ ID NO: 66) y NADPH no marcado. La figura 34B es un perfil de CG/EM de hidrocarburos producidos *in vitro* usando octadecanoil-CoA, NADPH marcado, YP_400611 de *Synechococcus elongatus* PCC7942 (Synpcc7942_1594) (SEQ ID NO: 66) y S-(4-²H)NADPH. La figura 34C es un perfil de CG/EM de hidrocarburos producidos *in vitro* usando octadecanoil-CoA, NADPH marcado, YP_400611 de *Synechococcus elongatus* PCC7942 (Synpcc7942_1594) (SEQ ID NO: 66) y R-(4-²H)NADPH.
- 45 La figura 35 es un perfil de CG/EM de hidrocarburos en el sobrenadante libre de células producido por células de *E. coli* MG1655 $\Delta fadE$ en medios Che-9 que expresan YP_400611 de *Synechococcus elongatus* PCC7942 (Synpcc7942_1594) (SEQ ID NO: 65).
- 50 La figura 36 es un perfil de CG/EM de hidrocarburos en el sobrenadante libre de células producido por células de *E. coli* MG1655 $\Delta fadE$ en medios Che-9 que expresan YP_400611 de *Synechococcus elongatus* PCC7942 (Synpcc7942_1594) (SEQ ID NO: 65) y Npun02004178 de *Nostoc punctiforme* PCC73102 (ZP_00108838) (SEQ ID NO: 5).

La figura 37 es un perfil de CG/EM de hidrocarburos producidos por células de *E. coli* MG1655 que expresan alr5283 de *Nostoc* sp. PCC7120 (NP_489323) (SEQ ID NO: 7) y alr5284 de *Nostoc* sp. PCC7120 (NP_489324) (SEQ ID NO: 81).

5 La figura 38 es una lista de ejemplos de homólogos de YP_400610 de *Synechococcus elongatus* PCC7942 (Synpcc7942_1593) (SEQ ID NO: 1) a partir de una base de datos metagenómica.

La figura 39 es una lista de ejemplos de homólogos de YP_400611 de *Synechococcus elongatus* PCC7942 (Synpcc7942_1594) (SEQ ID NO: 65) a partir de una base de datos metagenómica.

La figura 40 es una tabla que identifica diversos genes que pueden expresarse, sobreexpresarse o atenuarse para aumentar la producción de sustratos particulares.

10 Descripción detallada

15 La invención se refiere a composiciones y métodos de producción de aldehídos, alcoholes grasos e hidrocarburos (tales como alcanos, alquenos y alquinos) a partir de sustratos, por ejemplo, un sustrato de acil-ACP, uno de ácido graso, uno de acil-CoA, uno de aldehído graso o uno de alcohol graso (por ejemplo, tal como se describe en el documento PCT/US08/058788). Tales aldehídos, alcanos y alquenos son útiles como biocombustibles (por ejemplo, sustitutos para gasolina, diésel, combustible para motores a reacción, etc.), productos químicos especializados (por ejemplo, lubricantes, aditivos para combustibles, etc.) o materia prima para conversión química adicional (por ejemplo, combustibles, polímeros, plásticos, materiales textiles, disolventes, adhesivos, etc.). La invención se basa, en parte, en la identificación de genes que están implicados en la biosíntesis de aldehídos, alcanos y alquenos.

20 Tales genes de biosíntesis de alcanos y alquenos incluyen, por ejemplo, Synpcc7942_1593 de *Synechococcus elongatus* PCC7942 (SEQ ID NO: 1), sll0208 de *Synechocystis* sp. PCC6803 (SEQ ID NO: 3), Npun02004178 de *Nostoc punctiforme* PCC 73102 (SEQ ID NO: 5), alr5283 de *Nostoc* sp. PCC 7120 (SEQ ID NO: 7), AM1_4041 de *Acaryochloris marina* MBIC11017 (SEQ ID NO: 9), tll1313 de *Thermosynechococcus elongatus* BP-1 (SEQ ID NO: 11), CYA_0415 de *Synechococcus* sp. JA-3-3A (SEQ ID NO: 13), gll3146 de *Gloeobacter violaceus* PCC 7421 (SEQ ID NO: 15), PM123 de *Prochlorococcus marinus* MIT9313 (SEQ ID NO: 17), PMM0532 de *Prochlorococcus marinus* subesp. *pastoris* cepa CCMP1986 (SEQ ID NO: 19), PMN2A_1863 de *Prochlorococcus marinus* cepa NATL2A (SEQ ID NO: 21), RS9917_09941 de *Synechococcus* sp. RS9917 (SEQ ID NO: 23), RS9917_12945 de *Synechococcus* sp. RS9917 (SEQ ID NO: 25), cce_0778 de *Cyanothece* sp. ATCC51142 (SEQ ID NO: 27), Cyan7425DRAFT_1220 de *Cyanothece* sp. PCC7245 (SEQ ID NO: 29), cce_0778 de *Cyanothece* sp. PCC7245 (SEQ ID NO: 31), YP_323043 de *Anabaena variabilis* ATCC29413 (Ava_2533) (SEQ ID NO: 33) e YP_170760 de *Synechococcus elongatus* PCC6301 (syc0050_d) (SEQ ID NO: 35). En la tabla 1 y la figura 38 se enumeran otros genes de biosíntesis de alcanos y alquenos.

35 Los genes de biosíntesis de aldehídos incluyen, por ejemplo, Synpcc7942_1594 de *Synechococcus elongatus* PCC7942 (SEQ ID NO: 65), sll0209 de *Synechocystis* sp. PCC6803 (SEQ ID NO: 67), cce_1430 de *Cyanothece* sp. ATCC51142 (SEQ ID NO: 69), PMM0533 de *Prochlorococcus marinus* subesp. *pastoris* cepa CCMP1986 (SEQ ID NO: 71), NP_96091 de *Gloeobacter violaceus* PCC7421 (gll3145) (SEQ ID NO: 73), ZP_00108837 de *Nostoc punctiforme* PCC73102 (Npun02004176) (SEQ ID NO: 75), YP_323044 de *Anabaena variabilis* ATCC29413 (Ava2534) (SEQ ID NO: 77), YP_170761 de *Synechococcus elongatus* PCC6301 (syc0051_d) (SEQ ID NO: 79) y alr5284 de *Nostoc* sp. PCC 7120 (SEQ ID NO: 81). En la tabla 1 y la figura 39 se enumeran otros genes de biosíntesis de aldehídos.

40 Usando los métodos descritos en el presente documento, pueden prepararse aldehídos, alcoholes grasos, alcanos, y alquenos usando uno o más polipéptidos o genes de biosíntesis de aldehídos, alcanos y/o alquenos descritos en el presente documento, o variantes de los mismos, utilizando células huésped o métodos libres de células.

Tabla 1: Homólogos de genes de biosíntesis de aldehídos y alcanos en genomas de cianobacterias.

Cianobacteria	N.º de registro de gen de biosíntesis de alcanos	% de ID	N.º de registro de gen de biosíntesis de aldehídos	% de ID

<i>Synechococcus elongatus</i> PCC 7942	YP_400610	100	YP_400611	100
<i>Synechococcus elongatus</i> PCC 6301	YP_170760	100	YP_170761	100
<i>Microcoleus chthonoplastes</i> PCC 7420	EDX75019	77	EDX74978	70
<i>Arthrospira maxima</i> CS-328	EDZ94963	78	EDZ94968	68
<i>Lyngbya</i> sp. PCC 8106	ZP_01619575	77	ZP_01619574	69
<i>Nodularia spumigena</i> CCY9414	ZP_01628096	77	ZP_01628095	70
<i>Trichodesmium erythraeum</i> IMS101	YP_721979	76	YP_721978	69
<i>Microcystis aeruginosa</i> NIES-843	YP_001660323	75	YP_001660322	68
<i>Microcystis aeruginosa</i> PCC 7806	CAO90780	74	CAO90781	67
<i>Nostoc</i> sp. PCC 7120	NP_489323	74	NP_489324	72
<i>Nostoc azollae</i> 0708	EEG05692	73	EEG05693	70
<i>Anabaena variabilis</i> ATCC 29413	YP_323043	74	YP_323044	73
<i>Crocospaera watsonii</i> WH 8501	ZP_00514700	74	ZP_00516920	67
<i>Synechocystis</i> sp. PCC 6803	NP_442147	72	NP_442146	68
<i>Synechococcus</i> sp. PCC 7335	EDX86803	73	EDX87870	67
<i>Cyanothece</i> sp. ATCC 51142	YP_001802195	73	YP_001802846	67
<i>Cyanothece</i> sp. CCY0110	ZP_01728578	72	ZP_01728620	68
<i>Nostoc punctiforme</i> PCC 73102	ZP_00108838	72	ZP_00108837	71
<i>Acarynchloris marina</i> MBIC11017	YP_001518340	71	YP_001518341	66
<i>Cyanothece</i> sp. PCC 7425	YP_002481151	71	YP_002481152	70
<i>Cyanolthece</i> sp. PCC 8801	ZP_02941459	70	ZP_02942716	69
<i>Thermosynechococcus elongatus</i> BP-1	NP_682103	70	NP_682102	70
<i>Synechococcus</i> sp. JA-2-3B'a(2-13)	YP_478639	68	YP_478638	63
<i>Synechococcus</i> sp. RCC307	YP_001227842	67	YP_001227841	64
<i>Synechococcus</i> sp. WH 7803	YP_001224377	68	YP_001224378	65
<i>Synechococcus</i> sp. WH 8102	NP_897829	70	NP_897828	65
<i>Synechococcus</i> sp. WH 7805	ZP_01123214	68	ZP_01123215	65
<i>Synechococcus</i> GOM 3012 tipo A marino no cultivado	ABD96376	70	ABD96375	65
<i>Synechococcus</i> sp. JA-3-3Ab	YP_473897	68	YP_473896	62
<i>Synechococcus</i> GOM 306 tipo A marino no cultivado	ABD96328	70	ABD96327	65
<i>Synechococcus</i> GOM 3M9 tipo A marino no cultivado	ABD96275	68	ABD96274	65
<i>Synechococcus</i> sp. CC9311	YP_731193	63	YP_731192	63
<i>Synechococcus</i> 5B2 tipo A marino no cultivado	ABB92250	69	ABB92249	64
<i>Synechococcus</i> sp. WH 5701	ZP_01085338	66	ZP_01085337	67
<i>Gloeobacter violaceus</i> PCC 7421	NP_926092	63	NP_926091	67
<i>Synechococcus</i> sp. RS9916	ZP_01472594	69	ZP_01472595	66
<i>Synechococcus</i> sp. RS9917	ZP_01079772	68	ZP_01079773	65
<i>Synechococcus</i> sp. CC9605	YP_381055	66	YP_381056	66
<i>Cyanobium</i> sp. PCC 7001	EDY39806	64	EDY38361	64
<i>Prochlorococcus marinus</i> cepa MIT 9303	YP_001016795	63	YP_001016797	66
<i>Prochlorococcus marinus</i> cepa MIT9313	NP_895059	63	NP_895058	65
<i>Synechococcus</i> sp. CC9902	YP_377637	66	YP_377636	65
<i>Prochlorococcus marinus</i> cepa MIT 9301	YP_001090782	62	YP_001090783	62
<i>Synechococcus</i> sp. BL107	ZP_01469468	65	ZP_01469469	65
<i>Prochlorococcus marinus</i> cepa AS9601	YP_001008981	62	YP_001008982	61
<i>Prochlorococcus marinus</i> cepa MIT9312	YP_397029	62	YP_397030	61
<i>Prochlorococcus marinus</i> subesp. <i>pastoris</i> cepa CCMP1986	NP_892650	60	NP_892651	63
<i>Prochlorococcus marinus</i> cepa MIT 9211	YP_001550420	61	YP_001550421	63
<i>Cyanothece</i> sp. PCC 7425	YP_002483683	59	-	-
<i>Prochlorococcus marinus</i> cepa NATL2A	YP_293054	59	YP_293055	62
<i>Prochlorococcus marinus</i> cepa NATL1A	YP_001014415	59	YP_001014416	62
<i>Prochlorococcus marinus</i> subesp. <i>marinus</i> cepa CCMP1375	NP_874925	59	NP_874926	64
<i>Prochlorococcus marinus</i> cepa MIT 9515_05961	YP_001010912	57	YP_001010913	63
<i>Prochlorococcus marinus</i> cepa MIT 9215_06131	YP_001483814	59	YP_001483815	62
<i>Synechococcus</i> sp. RS9917	ZP_01080370	43	-	-
<i>Synechococcus</i> GOM 5D20 tipo A marino no cultivado			ABD96480	65

Genes de biosíntesis de aldehídos, alcanos y alquenos y variantes

Los métodos y las composiciones descritos en el presente documento incluyen, por ejemplo, genes de biosíntesis de alcanos o alquenos que tienen la secuencia de nucleótidos de SEQ ID NO: 1; 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33 ó 35, así como variantes de polinucleótido de los mismos. En algunos casos, el gen de biosíntesis de alcanos o alquenos codifica para uno o más de los motivos de aminoácido descritos en el presente documento. Por ejemplo, el gen de biosíntesis de alcanos o alquenos puede codificar para un polipéptido que comprende SEQ ID NO: 37, 38, 39, 41, 42, 43 ó 44. El gen de biosíntesis de alcanos o alquenos también puede incluir un polipéptido

5

que comprende SEQ ID NO: 40 y también uno cualquiera de SEQ ID NO: 37, 38 ó 39.

Los métodos y las composiciones descritos en el presente documento también incluyen, por ejemplo, genes de biosíntesis de aldehídos que tienen la secuencia de nucleótidos de SEQ ID NO: 65, 67, 69, 71, 73, 75, 77, 79 u 81, así como variantes de polinucleótido de los mismos. En algunos casos, el gen de biosíntesis de aldehídos codifica para uno o más de los motivos de aminoácido descritos en el presente documento. Por ejemplo, el gen de biosíntesis de aldehídos puede codificar para un polipéptido que comprende SEQ ID NO: 54, 55, 56, 57; 58, 59, 60, 61, 62, 63 ó 64.

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 y técnicas de clonación convencionales. Alternativamente, tales variantes, fragmentos, análogos o derivados pueden crearse usando procedimientos de síntesis química o modificación.

En la técnica se conocen bien métodos de preparación de variantes. Estos incluyen procedimientos en los que se modifican secuencias de ácido nucleico obtenidas a partir de aislados naturales 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 de variantes que tienen una o más diferencias de nucleótido con respecto a la secuencia obtenida a partir del aislado natural. Normalmente, estas diferencias de nucleótido dan como resultado cambios de aminoácido con respecto a los polipéptidos codificados por los ácidos nucleicos a partir de aislados naturales.

Por ejemplo, pueden crearse variantes usando PCR propensa a errores (véase, por ejemplo, Leung *et al.*, Technique 1:11-15, 1989; y Caldwell *et al.*, PCR Methods Applic. 2:28-33, 1992). En PCR propensa a errores, 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 de biosíntesis de aldehídos o alcanos), con cebadores de PCR, tampón de reacción, MgCl₂, MnCl₂, Taq polimerasa y una concentración apropiada de dNTP para lograr una alta tasa de mutación puntual a lo largo de toda la longitud del producto de PCR. Por ejemplo, puede realizarse la reacción usando 20 fmoles de ácido nucleico que va a mutarse (por ejemplo, una secuencia de polinucleótido de biosíntesis de aldehídos o alcanos), 30 pmoles de cada cebador de PCR, un tampón de reacción que comprende KCl 50 mM, Tris HCl 10 mM (pH 8,3) y gelatina al 0,01%, MgCl₂ 7 mM, MnCl₂ 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 la PCR durante 30 ciclos de 94°C durante 1 min, 45°C durante 1 min y 72°C durante 1 min. Sin embargo, se apreciará que estos parámetros pueden variarse según sea apropiado. Entonces se clonan los ácidos nucleicos mutagenizados en un vector apropiado y se evalúan las actividades de los polipéptidos codificados por los ácidos nucleicos mutagenizados.

También pueden crearse variantes usando mutagénesis dirigida a oligonucleótidos para generar mutaciones específicas de sitio en cualquier ADN clonado de interés. La mutagénesis de oligonucleótidos se describe, por ejemplo, en 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 mutarse (por ejemplo, una secuencia de polinucleótido de biosíntesis de aldehídos o alcanos). Se recuperan clones que contienen el ADN mutado, y se evalúan las actividades de los polipéptidos para los que codifican.

Otro método para generar variantes es PCR de ensamblaje. La PCR de ensamblaje implica el ensamblaje de un producto de PCR a partir de una mezcla de fragmentos de ADN pequeños. Un gran número de diferentes reacciones de PCR se producen en paralelo en el mismo vial, cebando los productos de una reacción a los productos de otra reacción. La PCR de ensamblaje se describe, por ejemplo, en la patente estadounidense n.º 5.965.408.

Todavía otro método de generar variantes es mutagénesis mediante PCR sexual. En la mutagénesis mediante PCR sexual, se produce la recombinación homóloga forzada entre moléculas de ADN de secuencias de ADN diferentes, 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 la fijación del cruzamiento mediante extensión del cebador en una reacción de PCR. La mutagénesis mediante PCR sexual se describe, por ejemplo, en Stemmer, PNAS, USA 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 de biosíntesis de aldehídos o alcanos) en una de estas cepas generará finalmente mutaciones al azar dentro del ADN. Las cepas mutadoras adecuadas para su uso para la mutagénesis *in vivo* se describen, por ejemplo, en la publicación PCT n.º WO 91/16427.

También pueden generarse variantes usando mutagénesis de casete. En la mutagénesis de casete, se reemplaza una región pequeña de un molécula de ADN bicatenario por un "casete" de oligonucleótido sintético que difiere de la secuencia nativa. El oligonucleótido a menudo contiene una secuencia nativa completa y/o parcialmente al azar.

5 También puede usarse mutagénesis de conjunto recursivo para generar variantes. La mutagénesis de conjunto recursivo es un algoritmo para el diseño por ingeniería genética 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 retroalimentación para controlar las tandas sucesivas de mutagénesis de casete combinatoria. La mutagénesis de conjunto recursivo se describe, por ejemplo, en Arkin *et al.*, PNAS, USA 89:7811-7815, 1992.

10 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. La mutagénesis de conjunto exponencial se describe, por ejemplo, en Delegrave *et al.*, Biotech. Res. 11: 1548-1552, 1993. La mutagénesis dirigida al sitio y al azar se describen, por ejemplo, en Arnold, Curr. Opin. Biotech. 4:450-455, 1993.

15 En algunas realizaciones, se crean variantes usando procedimientos de intercambio en los que se fusionan entre sí porciones 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, por ejemplo, en las patentes estadounidenses n.ºs 5.965.408 y 5.939.250.

20 Las variantes de polinucleótido también incluyen análogos de ácido nucleico. Los análogos de ácido nucleico pueden modificarse en el resto básico, resto azúcar o estructura principal de fosfato para mejorar, por ejemplo, la estabilidad, hibridación o solubilidad del ácido nucleico. Las modificaciones en el resto básico incluyen desoxiuridina para desoxitimidina y 5-metil-2'-desoxicitidina o 5-bromo-2'-doxicidina para desoxicitidina. Las modificaciones del resto azúcar incluyen modificación del 2'-hidroxilo del azúcar ribosa para formar azúcares 2'-O-metilo o 2'-O-alilo. La estructura principal de fosfato de desoxirribosa puede modificarse para producir ácidos morfolino-nucleicos, en los que cada resto básico está unido a un anillo morfolino de seis miembros, o ácidos nucleicos peptídicos, en los que la estructura principal de desoxifosfato se reemplaza por una estructura principal de pseudopéptido y se mantienen las cuatro bases. (Véase, por ejemplo, Summerton *et al.*, Antisense Nucleic Acid Drug Dev. (1997) 7:187-195; e Hyrup *et al.*, Bioorgan. Med. Chem. (1996) 4:5-23). Además, la estructura principal de desoxifosfato puede remplazarse por, por ejemplo, una estructura principal de fosfortioato o fosforoditioato, una fosforoamida o una estructura principal de fosfotriéster de alquilo.

30 Los polipéptidos de biosíntesis de aldehídos y alcanos Synpcc7942_1594 (SEQ ID NO: 66) y Synpcc7942_1593 (SEQ ID NO: 2) tienen homólogos en otras cianobacterias (en la tabla 1 se representan ejemplos no limitativos). Por tanto, puede usarse cualquier secuencia de polinucleótido que codifique para un homólogo enumerado en la tabla 1, o una variante del mismo, como polinucleótido de biosíntesis de aldehídos o alcanos en los métodos descritos en el presente documento. Cada cianobacteria enumerada en la tabla 1 tiene copias de ambos genes. El nivel de identidad de secuencia de los productos génicos oscila entre el 61% y el 73% para Synpcc7942_1594 (SEQ ID NO: 66) y entre el 43% y el 78% para Synpcc7942_1593 (SEQ ID NO: 2).

40 En la figura 39 se enumeran homólogos adicionales del polipéptido de biosíntesis de aldehídos Synpcc7942_1594 (SEQ ID NO: 66), y puede usarse cualquier secuencia de polinucleótido que codifique para un homólogo enumerado en la figura 39, o una variante del mismo, como polinucleótido de biosíntesis de aldehídos en los métodos descritos en el presente documento. En la figura 38 se enumeran homólogos adicionales del polipéptido de biosíntesis de alcanos Synpcc7942_1593 (SEQ ID NO: 2), y puede usarse cualquier secuencia de polinucleótido que codifique para un homólogo enumerado en la figura 38, o una variante del mismo, como polinucleótido de biosíntesis de alcanos en los métodos descritos en el presente documento.

45 En determinados casos, un gen de biosíntesis de aldehídos, alcanos y/o alquenos tiene los codones optimizados para su 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).

Polipéptidos de biosíntesis de aldehídos, alcanos y alquenos y variantes

50 Los métodos y las composiciones descritos en el presente documento también incluyen polipéptidos de biosíntesis de alcanos o alquenos que tienen la secuencia de aminoácidos de SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34 ó 36, así como variantes de polipéptido de los mismos. En algunos casos, un polipéptido de biosíntesis de alcanos o alquenos es uno que incluye uno o más de los motivos de aminoácido descritos en el presente documento. Por ejemplo, el polipéptido de biosíntesis de alcanos o alquenos puede incluir la secuencia de aminoácidos de SEQ ID NO: 37, 38, 39, 41, 42, 43 ó 44. El polipéptido de biosíntesis de alcanos o alquenos también puede incluir la secuencia de aminoácidos de SEQ ID NO: 40 y también una cualquiera de SEQ ID NO: 37, 38 ó 39.

Los métodos y las composiciones descritos en el presente documento también incluyen polipéptidos de biosíntesis de aldehídos que tienen la secuencia de aminoácidos de SEQ ID NO: 66, 68, 70, 72, 74, 76, 78, 80 u 82, así como

variantes de polipéptido de los mismos. En algunos casos, un polipéptido de biosíntesis de aldehídos es uno que incluye uno o más de los motivos de aminoácido descritos en el presente documento. Por ejemplo, el polipéptido de biosíntesis de aldehídos puede incluir la secuencia de aminoácidos de SEQ ID NO: 54, 55, 56, 57, 58, 59, 60, 61, 62, 63 ó 64.

5 Las variantes de polipéptidos de biosíntesis de aldehídos, alcanos y alquenos pueden ser variantes en las que uno o más residuos de aminoácido se sustituyen por un residuo de aminoácido conservado o no conservado (preferiblemente un residuo de aminoácido conservado). Tal residuo de aminoácido sustituido puede ser o no uno codificado por el código genético.

10 Las sustituciones conservativas son las que sustituyen un aminoácido dado en un polipéptido por otro aminoácido de características similares. Las sustituciones conservativas típicas son los siguientes remplazos: remplazo de un aminoácido alifático, tal como alanina, valina, leucina e isoleucina, por otro aminoácido alifático; remplazo de una serina por una treonina o viceversa; remplazo de un residuo ácido, tal como ácido aspártico y ácido glutámico, por otro residuo ácido; remplazo 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 remplazo de un residuo aromático, tal como fenilalanina y tirosina, por otro residuo aromático.

15 Otras variantes de polipéptido son aquellas en las que uno o más residuos de aminoácido incluyen un grupo sustituyente. Todavía otras variantes de polipéptido son aquellas en las que en el polipéptido está asociado con otro compuesto, tal como un compuesto para aumentar la semivida del polipéptido (por ejemplo, polietilenglicol).

20 Variantes de polipéptido adicionales son aquellas en las que aminoácidos adicionales están fusionados con el polipéptido, tal como una secuencia líder, una secuencia secretora, una secuencia de proproteína o una secuencia que facilita la purificación, el enriquecimiento o la estabilización del polipéptido.

25 En algunos casos, la variante de polipéptido de biosíntesis de alcanos o alquenos conserva la misma función biológica que un polipéptido que tiene la secuencia de aminoácidos de SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34 ó 36 (por ejemplo, conserva actividad de biosíntesis de alcanos o alquenos) y tiene una secuencia de aminoácidos sustancialmente idéntica al mismo.

30 En otros casos, las variantes de polipéptido de biosíntesis de alcanos o alquenos tienen una homología de al menos aproximadamente el 50%, al menos aproximadamente el 55%, al menos aproximadamente el 60%, al menos aproximadamente el 65%, al menos aproximadamente el 70%, al menos aproximadamente el 75%, al menos aproximadamente el 80%, al menos aproximadamente el 85%, al menos aproximadamente el 90%, al menos aproximadamente el 95% o más de aproximadamente el 95% con la secuencia de aminoácidos de SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34 ó 36. En otra realización, las variantes de polipéptido incluyen un fragmento que comprende al menos aproximadamente 5, 10, 15, 20, 25, 30, 35, 40, 50, 75, 100 ó 150 aminoácidos consecutivos de las mismas.

35 En algunos casos, una variante de polipéptido de biosíntesis de aldehídos conserva la misma función biológica que un polipéptido que tiene la secuencia de aminoácidos de SEQ ID NO: 66, 68, 70, 72, 74, 76, 78, 80 u 82 (por ejemplo, conserva actividad de biosíntesis de aldehídos) y tiene una secuencia de aminoácidos sustancialmente idéntica al mismo.

40 Aún en otros casos, las variantes de polipéptido de biosíntesis de aldehídos tienen una homología de al menos aproximadamente el 50%, al menos aproximadamente el 55%, al menos aproximadamente el 60%, al menos aproximadamente el 65%, al menos aproximadamente el 70%, al menos aproximadamente el 75%, al menos aproximadamente el 80%, al menos aproximadamente el 85%, al menos aproximadamente el 90%, al menos aproximadamente el 95% o más de aproximadamente el 95% con la secuencia de aminoácidos de SEQ ID NO: 66, 68, 70, 72, 74, 76, 78, 80 u 82. En otra realización, las variantes de polipéptido incluyen un fragmento que comprende al menos aproximadamente 5, 10, 15, 20, 25, 30, 35, 40, 50, 75, 100 ó 150 aminoácidos consecutivos de las mismas.

45 Las variantes de polipéptido o fragmentos de las mismas pueden obtenerse aislando ácidos nucleicos que codifican para las mismas usando técnicas descritas en el presente documento o expresando ácidos nucleicos sintéticos que codifican para las mismas. Alternativamente, pueden obtenerse variantes de polipéptido o fragmentos de las mismas a través de procedimientos de enriquecimiento o purificación bioquímicos. La secuencia de variantes de polipéptido o fragmentos pueden determinarse mediante digestión proteolítica, electroforesis en gel y/o microsecuenciación. La secuencia de las variantes de polipéptido o fragmentos de biosíntesis de alcanos o alquenos puede compararse entonces con la secuencia de aminoácidos de SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34 ó 36 usando cualquiera de los programas descritos en el presente documento. La secuencia de las variantes de polipéptido o fragmentos de biosíntesis de aldehídos puede compararse con la secuencia de aminoácidos de SEQ ID NO: 66, 68, 70, 72, 74, 76, 78, 80 u 82 usando cualquiera de los programas descritos en el presente documento.

55 Las variantes de polipéptido y fragmentos de las mismas pueden someterse a ensayo para determinar la actividad productora de aldehídos, alcoholes grasos, alcanos y/o alquenos usando métodos de rutina. Por ejemplo, las variantes de polipéptido o fragmento pueden ponerse en contacto con un sustrato (por ejemplo, un sustrato de

derivado de ácido graso u otro sustrato descrito en el presente documento) en condiciones que permiten que la variante de polipéptido funcione. Puede medirse una disminución en el nivel del sustrato o un aumento en el nivel de un aldehído, alcano o alqueno para determinar la actividad productora de aldehídos, alcoholes grasos, alcanos o alquenos, respectivamente.

5 Anticuerpos anti-polipéptido de biosíntesis de aldehídos, alcoholes grasos, alcanos y alquenos

Los polipéptidos de biosíntesis de aldehídos, alcoholes grasos, alcanos y alquenos descritos en el presente documento también pueden usarse para producir anticuerpos dirigidos frente a polipéptidos de biosíntesis de aldehídos, alcoholes grasos, alcanos y alquenos. Pueden usarse tales anticuerpos, por ejemplo, para detectar la expresión de un polipéptido de biosíntesis de aldehídos, alcoholes grasos, alcanos o alquenos usando métodos conocidos en la técnica. El anticuerpo puede ser, por ejemplo, un anticuerpo policlinal; un anticuerpo monoclonal o fragmento de unión a antígeno del mismo; un anticuerpo modificado tal como un anticuerpo quimérico, anticuerpo reconformado, anticuerpo humanizado, o fragmento de los mismos (por ejemplo, Fab', Fab, F(ab')₂); o un anticuerpo biosintético, por ejemplo, un anticuerpo de cadena sencilla, anticuerpo de un solo dominio (DAB), Fv, Fv de cadena sencilla (scFv), o similares.

15 Se describen métodos de preparación y uso de anticuerpos policlonales y monoclonales, por ejemplo, en Harlow *et al.*, *Using Antibodies: A Laboratory Manual: Portable Protocol I*. Cold Spring Harbor Laboratory (1 de diciembre de 1998). En la técnica se conocen métodos para preparar anticuerpos modificados y fragmentos de anticuerpo (por ejemplo, anticuerpos quiméricos, anticuerpos reconformados, anticuerpos humanizados, o fragmentos de los mismos, por ejemplo, fragmentos Fab', Fab, F(ab')₂); o anticuerpos biosintéticos (por ejemplo, anticuerpos de cadena sencilla, anticuerpos de un solo dominio (DAB), Fv, Fv de cadena sencilla (scFv), y similares), y pueden encontrarse, por ejemplo, en Zola, *Monoclonal Antibodies: Preparation and Use of Monoclonal Antibodies and Engineered Antibody Derivatives*, Springer Verlag (15 de diciembre del 2000; 1ª edición).

20 Sustratos

25 Las composiciones y los métodos descritos en el presente documento pueden usarse para producir aldehídos, alcoholes grasos, alcanos y/o alquenos a partir de un sustrato apropiado. Sin querer restringirse a ninguna teoría particular, se cree que los polipéptidos de biosíntesis de alcanos o alquenos descritos en el presente documento producen alcanos o alquenos a partir de sustratos por medio de un mecanismo de descarbonilación. En algunos casos, el sustrato es un derivado de ácido graso, por ejemplo, un aldehído graso, y puede producirse un alcano que tiene patrones de ramificación y una longitud de la cadena de carbono particulares a partir de un derivado de ácido graso, por ejemplo, un aldehído graso, que tiene esas características particulares. En otros casos, el sustrato es un derivado de ácido graso insaturado, por ejemplo, un aldehído graso insaturado, y puede producirse un alqueno que tiene patrones de ramificación y una longitud de la cadena de carbono particulares a partir de un derivado de ácido graso insaturado, por ejemplo, un aldehído graso insaturado, que tiene esas características particulares.

35 Sin querer restringirse a ninguna teoría particular, se cree que los polipéptidos de biosíntesis de aldehídos descritos en el presente documento producen aldehídos a partir de sustratos por medio de un mecanismo de reducción. En determinados casos, el sustrato es una acil-ACP.

Sin querer restringirse a ninguna teoría particular, se cree que los alcoholes grasos descritos en el presente documento se producen a partir de sustratos por medio de un mecanismo de reducción. En determinados casos, el sustrato es un aldehído graso.

40 Por consiguiente, cada etapa dentro de una ruta de biosíntesis que conduce a la producción de estos sustratos puede modificarse para producir o sobreproducir el sustrato de interés. Por ejemplo, genes conocidos implicados en la ruta de biosíntesis de ácidos grasos, la ruta de aldehídos grasos y la ruta de alcoholes grasos pueden expresarse, sobreexpresarse o atenuarse en células huésped para producir un sustrato deseado (véase, por ejemplo, el documento PCT/US08/058788. En la figura 40 se proporcionan genes a modo de ejemplo.

45 Síntesis de sustratos

Ácido graso sintasa (FAS) es un grupo de polipéptidos que catalizan el inicio y la elongación de cadenas de acilo (Marrakchi *et al.*, *Biochemical Society*, 30:1050-1055, 2002). La proteína portadora de acilo (ACP) junto con las enzimas en la ruta de FAS controlan la longitud, el grado de saturación y la ramificación de los derivados de ácidos grasos producidos. La ruta de biosíntesis de ácidos grasos implica los precursores acetil-CoA y malonil-CoA. Las etapas en esta ruta se catalizan por enzimas de las familias de genes de la biosíntesis de ácidos grasos (*fab*) y acetil-CoA carboxilasa (*acc*) (véase, por ejemplo, Heath *et al.*, *Prog. Lipid Res.* 40(6):467-97 (2001)).

50 Pueden modificarse por ingeniería genética células huésped para expresar sustratos de derivados de ácidos grasos expresando o sobreexpresando de manera recombinante uno o más genes de acetil-CoA y/o malonil-CoA sintasa. Por ejemplo, para aumentar la producción de acetil-CoA, puede expresarse uno o más de los siguientes genes en una célula huésped: *pdh*, *panK*, *aceEF* (que codifica para el componente de deshidrogenasa E1p y el componente de dihidrolipoamida aciltransferasa E2p de los complejos de piruvato y 2-oxoglutarato deshidrogenasa), *fabH*, *fabB*, *fabD*, *fabG*, *acpP* y *fabF*. Números de registro de GenBank a modo de ejemplo para estos genes son: *pdh*

(BAB34380, AAC73227, AAC73226), *panK* (también conocido como *coaA*, AAC76952), *aceEF* (AAC73227, AAC73226), *fabH* (AAC74175), *fabD* (AAC74176), *fabG* (AAC74177), *acpP* (AAC74178), *fabF* (AAC74179). Adicionalmente, los niveles de expresión de *fadE*, *gpsA*, *ldhA*, *pflb*, *adhE*, *pta*, *poxB*, *ackA* y/o *ackB* pueden atenuarse o desactivarse en una célula huésped modificada por ingeniería genética mediante transformación con plásmidos condicionalmente replicativos o no replicativos que contienen mutaciones nulas o por delección de los genes correspondientes o sustituyendo secuencias de promotor o potenciador. Números de registro de GenBank a modo de ejemplo para estos genes son: *fadE* (AAC73325), *gpsA* (AAC76632), *ldhA* (AAC74462), *pflb* (AAC73989), *adhE* (AAC74323), *pta* (AAC75357), *poxB* (AAC73958), *ackA* (AAC75356) y *ackB* (BAB81430). Las células huésped resultantes tendrán niveles de producción de acetil-CoA aumentados cuando se hacen crecer en un entorno apropiado.

La sobreexpresión de malonil-CoA puede realizarse introduciendo *accABCD* (por ejemplo, número de registro AAC73296, EC 6.4.1.2) en una célula huésped. Los ácidos grasos pueden sobreexpresarse adicionalmente en células huésped introduciendo en la célula huésped una secuencia de ADN que codifica para una lipasa (por ejemplo, números de registro CAA89087, CAA98876).

Además, inhibir *PlsB* puede conducir a un aumento en los niveles de acil-ACP de cadena larga, lo que inhibirá las etapas iniciales en la ruta (por ejemplo, *accABCD*, *fabH* y *fabI*). Puede usarse la mutación D311E de *plsB* (por ejemplo, número de registro AAC77011) para aumentar la cantidad de acil-CoA disponible.

Además, puede modificarse por ingeniería genética una célula huésped para sobreexpresar un gen *sfa* (supresor de *fabA*, por ejemplo, número de registro AAN79592) para aumentar la producción de ácidos grasos monoinsaturados (Rock *et al.*, J. Bacteriology 178:5382-5387, 1996).

En algunos casos, pueden modificarse por ingeniería genética células huésped para expresar, sobreexpresar o atenuar la expresión de una tioesterasa para aumentar la producción de sustrato de ácido graso. La longitud de cadena de un sustrato de ácido graso se controla por la tioesterasa. En algunos casos, puede sobreexpresarse un gen *tes* o *fat*. En otros casos, pueden producirse ácidos grasos C₁₀ atenuando la tioesterasa C₁₈ (por ejemplo, números de registro AAC73596 y P0ADA1), que usa C_{18:1}-ACP, y expresando la tioesterasa C₁₀ (por ejemplo, número de registro Q39513), que usa C₁₀-ACP. Esto da como resultado una población de ácidos grasos relativamente homogénea que tienen una longitud de la cadena de carbono de 10. Aún en otros casos, pueden producirse ácidos grasos C₁₄ atenuando tioesterasas endógenas que producen ácidos grasos distintos de C₁₄ y expresando las tioesterasas, que usan C₁₄-ACP (por ejemplo, número de registro Q39473). En algunas situaciones, pueden producirse ácidos grasos C₁₂ expresando tioesterasas que usan C₁₂-ACP (por ejemplo, número de registro Q41635) y atenuando tioesterasas que producen ácidos grasos distintos de C₁₂. Puede verificarse la sobreproducción de acetil-CoA, malonil-CoA y ácido graso usando métodos conocidos en la técnica, por ejemplo, usando precursores radiactivos, HPLC y CG-EM posteriormente a la lisis celular. En la tabla 2 se enumeran ejemplos no limitativos de tioesterasas que pueden usarse en los métodos descritos en el presente documento.

Tabla 2: tioesterasas

Número de registro	Organismo fuente	Gen	Producto producido preferentemente
AAC73596	<i>E. coli</i>	<i>tesA</i> sin secuencia líder	C _{18:1}
AAC73555	<i>E. coli</i>	<i>tesB</i>	
Q41635, AAA34215	<i>Umbellularia californica</i>	<i>fatB</i>	C _{12:0}
Q39513; AAC49269	<i>Cuphea hookeriana</i>	<i>fatB2</i>	C _{8:0} -C _{10:0}
AAC49269; AAC72881	<i>Cuphea hookeriana</i>	<i>fatB3</i>	C _{14:0} -C _{16:0}
Q39473, AAC49151	<i>Cinnamomum camphorum</i>	<i>fatB</i>	C _{14:0}
CAA85388	<i>Arabidopsis thaliana</i>	<i>fatB</i> [M141 T]*	C _{16:1}
NP 189147; NP 193041	<i>Arabidopsis thaliana</i>	<i>fatA</i>	C _{18:1}
CAC39106	<i>Bradyrhizobium japonicum</i>	<i>fatA</i>	C _{18:1}
AAC72883	<i>Cuphea hookeriana</i>	<i>fatA</i>	C _{18:1}
AAL79361	<i>Helianthus annuus</i>	<i>fatA1</i>	

* Mayer *et al.*, BMC Plant Biology 7:1-11, 2007

Formación de aldehídos ramificados, alcoholes grasos, alcanos y alquenos

Pueden producirse aldehídos, alcoholes grasos, alcanos y alquenos que contienen puntos de ramificación usando derivados de ácidos grasos ramificados como sustratos. Por ejemplo, aunque *E. coli* produce de manera natural derivados de ácidos grasos de cadena lineal (sFA), puede modificarse por ingeniería genética *E. coli* para que produzca derivados de ácidos grasos de cadena ramificada (brFA) introduciendo y expresando o sobreexpresando genes que proporcionan precursores ramificados en la *E. coli* (por ejemplo, familias de genes *bkd*, *ilv*, *icm* y *fab*). Adicionalmente, una célula huésped puede modificarse por ingeniería genética para expresar o sobreexpresar genes que codifican para proteínas para la elongación de brFA (por ejemplo, ACP, FabF, etc.) y/o para delecionar o

atenuar genes de células huésped correspondientes que conducen normalmente a sFA.

La primera etapa en la formación de brFA es la producción de los α -cetoácidos correspondientes mediante una aminoácido de cadena ramificada aminotransferasa. Las células huésped pueden incluir de manera endógena genes que codifican para tales enzimas o tales genes pueden introducirse de manera recombinante. *E. coli*, por ejemplo, expresa de manera endógena una enzima de este tipo, IlvE (EC 2.6.1.42; registro de GenBank YP_026247). En algunas células huésped, puede no expresarse una aminoácido de cadena ramificada aminotransferasa heteróloga. Sin embargo, IlvE de *E. coli* o cualquier otra aminoácido de cadena ramificada aminotransferasa (por ejemplo, IlvE de *Lactococcus lactis* (registro de GenBank AAF34406), IlvE de *Pseudomonas putida* (registro de GenBank NP_745648) o IlvE de *Streptomyces coelicolor* (registro de GenBank NP_629657)), si no es endógena, puede introducirse y expresarse de manera recombinante.

La segunda etapa es la descarboxilación oxidativa de los α -cetoácidos para dar la correspondiente acil-CoA de cadena ramificada. Esta reacción puede catalizarse por un complejo de α -cetoácido de cadena ramificada deshidrogenasa (*bkd*; EC 1.2.4.4.) (Denoya *et al.*, J. Bacteriol. 177:3504, 1995), que consiste en las subunidades E1 α / β (descarboxilasa), E2 (dihidrolipoil transacilasa) y E3 (dihidrolipoil deshidrogenasa). Estos complejos de α -cetoácido de cadena ramificada deshidrogenasa son similares a complejos de piruvato y α -cetoglutarato deshidrogenasa. Puede usarse cualquier microorganismo que presente brFA y/o crezca sobre aminoácidos de cadena ramificada como fuente para aislar genes *bkd* para su expresión en células huésped, por ejemplo, *E. coli*. Además, *E. coli* tiene el componente E3 como parte de su complejo de piruvato deshidrogenasa (*lpd*, EC 1.8.1.4, registro de GenBank NP_414658). Por tanto, puede ser suficiente expresar sólo los genes E1 α / β y E2 *bkd*. La tabla 3 enumera ejemplos no limitativos de genes *bkd* de varios microorganismos que pueden introducirse y expresarse de manera recombinante en una célula huésped para proporcionar precursores de acil-CoA de cadena ramificada.

Tabla 3: Genes *bkd* de microorganismos seleccionados

Organismo	Gen	N.º de registro de GenBank
<i>Streptomyces coelicolor</i>	<i>bkdA1</i> (E1 α) <i>bkdB1</i> (E1 β) <i>bkdC1</i> (E2)	NP_628006 NP_628005 NP_638004
<i>Streptomyces coelicolor</i>	<i>bkdA2</i> (E1 α) <i>bkdB2</i> (E1 β) <i>bkdC2</i> (E2)	NP_733618 NP_628019 NP_628018
<i>Streptomyces avermitilis</i>	<i>bkdA</i> (E1a) <i>bkdB</i> (E1b) <i>bkdC</i> (E2)	BAC72074 BAC72075 BAC72076
<i>Streptomyces avermitilis</i>	<i>bkdF</i> (E1 α) <i>bkdG</i> (E1 β) <i>bkdH</i> (E2)	BAC72088 BAC72089 BAC72090
<i>Bacillus subtilis</i>	<i>bkdAA</i> (E1 α) <i>bkdAB</i> (E1 β) <i>bkdB</i> (E2)	NP_390288 NP_390288 NP_390288
<i>Pseudomonas putida</i>	<i>bkdA1</i> (E1 α) <i>bkdA2</i> (E1 β) <i>bkdC</i> (E2)	AAA65614 AAA65615 AAA65617

En otro ejemplo, puede prepararse isobutiril-CoA en una célula huésped, por ejemplo en *E. coli*, mediante la coexpresión de una crotonil-CoA reductasa (Ccr, EC 1.6.5.5, 1.1.1.1) e isobutiril-CoA mutasa (subunidad grande IcmA, EC 5.4.99.2; subunidad pequeña IcmB, EC 5.4.99.2) (Han y Reynolds, J. Bacteriol. 179:5157, 1997). Crotonil-CoA es un producto intermedio en la biosíntesis de ácidos grasos en *E. coli* y otros microorganismos. En la tabla 4 se enumeran ejemplos no limitativos de genes *ccr* e *icm* de microorganismos seleccionados.

Tabla 4: Genes *ccr* e *icm* de microorganismos seleccionados

Organismo	Gen	N.º de registro de GenBank
<i>Streptomyces coelicolor</i>	<i>ccr</i> <i>icmA</i> <i>icmB</i>	NP_630556 NP_629554 NP_630904
<i>Streptomyces cinnamonensis</i>	<i>ccr</i> <i>icmA</i> <i>icmB</i>	AAD53915 AAC08713 AJ246005

Además de la expresión de los genes *bkd*, el inicio de la biosíntesis de brFA usa proteína portadora de β -cetoacilo sintasa III (FabH, EC 2.3.1.41) con especificidad por acil-CoA de cadena ramificada (Li *et al.*, J. Bacteriol. 187:3795-3799, 2005). En la tabla 5 se enumeran ejemplos no limitativos de tales enzimas FabH. Pueden

expresarse en una célula huésped genes *fabH* que están implicados en la biosíntesis de ácidos grasos de cualquier microorganismo que contiene brFA. Las enzimas Bkd y FabH de células huésped que no preparan de manera natural brFA no pueden respaldar la producción de brFA. Por tanto, *bkd* y *fabH* pueden expresarse de manera recombinante. Pueden insertarse vectores que contienen los genes *bkd* y *fabH* en una célula huésped de este tipo. De manera similar, el nivel endógeno de la producción de Bkd y FabH puede no ser suficiente para producir brFA. En este caso, pueden sobreexpresarse. Adicionalmente, pueden expresarse o sobreexpresarse otros componentes de la ruta de biosíntesis de ácidos grasos, tales como proteínas portadoras de acilo (ACP) y proteína portadora de β -cetoacil-acilo sintasa II (*fabF*, EC 2.3.1.41) (en la tabla 5 se indican ejemplos no limitativos de candidatos). Además de expresar estos genes, algunos genes en la ruta de biosíntesis de ácidos grasos endógena pueden atenuarse en la célula huésped (por ejemplo, los genes de *E. coli* *fabH* (n.º de registro de GenBank NP_415609) y/o *fabF* (n.º de registro de GenBank NP_415613)).

Tabla 5: Genes *fabH*, *ACP* y *fabF* de microorganismos seleccionados con brFA

Organismo	Gen	N.º de registro de GenBank
<i>Streptomyces coelicolor</i>	<i>fabH1</i> <i>ACP</i> <i>fabF</i>	NP 626634 NP 626635 NP 626636
<i>Streptomyces avermitilis</i>	<i>fabH3</i> <i>fabC3 (ACP)</i> <i>fabF</i>	NP 823466 NP 823467 NP 823468
<i>Bacillus subtilis</i>	<i>fabH_A</i> <i>fabH_B</i> <i>ACP</i> <i>fabF</i>	NP 389015 NP 388898 NP 389474 NP 389016
<i>Stenotrophomonas maltophilia</i>	SmaIDRAFT 0818 (<i>fabH</i>) SmaIDRAFT 0821 (<i>ACP</i>) SmaIDRAFT 0822 (<i>fabF</i>)	ZP 01643059 ZP 01643063 ZP 01643064
<i>Legionella pneumophila</i>	<i>fabH</i> <i>ACP</i> <i>fabF</i>	YP 123672 YP 123675 YP 123676

Formación de aldehídos cíclicos, alcoholes grasos, alcanos y alquenos

Pueden producirse aldehídos cíclicos, alcoholes grasos, alcanos y alquenos usando derivados de ácidos grasos cíclicos como sustratos. Para producir sustratos de derivados de ácidos grasos cíclicos, pueden introducirse genes que proporcionan precursores cíclicos (por ejemplo, las familias de genes *ans*, *chc* y *plm*) en la célula huésped y expresarse para permitir el inicio de la biosíntesis de ácidos grasos a partir de precursores cíclicos. Por ejemplo, para convertir una célula huésped, tal como *E. coli*, en una que puede sintetizar derivados de ácidos grasos ω -cíclicos (cyFA), puede introducirse un gen que proporciona el precursor cíclico ciclohexilcarbonil-CoA (CHC-CoA) (Cropp *et al.*, Nature Biotech. 18:980-983, 2000) y expresarse en la célula huésped. Los ejemplos no limitativos de genes que proporcionan CHC-CoA en *E. coli* incluyen: *ansJ*, *ansK*, *ansL*, *chcA* y *ansM* de la agrupación de genes de ansatrienina de *Streptomyces collinus* (Chen *et al.*, Eur. J. Biochem. 261: 98-107, 1999) o *plmJ*, *plmK*, *plmL*, *chcA* y *plmM* de la agrupación de genes de foslactomicina B de *Streptomyces* sp. HK803 (Palaniappan *et al.*, J. Biol. Chem. 278:35552-35557, 2003) junto con el gen *chcB* (Patton *et al.*, Biochem. 39:7595-7604, 2000) de *S. collinus*, *S. avermitilis* o *S. coelicolor* (véase la tabla 6). Los genes enumerados en la tabla 5 pueden expresarse entonces para permitir el inicio y la elongación de ácidos grasos ω -cíclicos. Alternativamente, pueden aislarse los genes homólogos de microorganismos que fabrican cyFA y expresarse en una célula huésped (por ejemplo, *E. coli*).

Tabla 6: Genes para la síntesis de CHC-CoA

Organismo	Gen	N.º de registro de GenBank
<i>Streptomyces collinus</i>	<i>ansJK</i> <i>ansL</i> <i>chcA</i> <i>ansM</i> <i>chcB</i>	U72144* AF268489
<i>Streptomyces</i> sp. HK803	<i>pmlJK</i> <i>pmlL</i> <i>chcA</i> <i>pmlM</i>	AAQ84158 AAQ84159 AAQ84160 AAQ84161
<i>Streptomyces coelicolor</i>	<i>chcB/caiD</i>	NP 629292
<i>Streptomyces avermitilis</i>	<i>chcB/caiD</i>	NP 629292

*Sólo se indica *chcA* como entrada en GenBank U72144, *ansJKLM* son según Chen *et al.* (Eur. J. Biochem. 261:98-107, 1999).

Los genes enumerados en la tabla 5 (*fabH*, *ACP* y *fabF*) permiten el inicio y la elongación de derivados de ácidos grasos ω -cíclicos porque tienen una amplia especificidad de sustrato. Si la coexpresión de cualquiera de estos genes con los genes indicados en la tabla 5 no proporciona cyFA, entonces pueden aislarse homólogos de *fabH*, *ACP* y/o *fabF* de microorganismos que fabrican cyFA (por ejemplo, los indicados en la tabla 7) (por ejemplo, usando cebadores de PCR degenerados o sondas de secuencias de ADN heterólogos) y coexpresarse.

Tabla 7: Ejemplos no limitativos de microorganismos que contienen ácidos grasos ω -cíclicos

Organismo	Referencia
<i>Curtobacterium pusillum</i>	ATCC19096
<i>Alicyclobacillus acidoterrestris</i>	ATCC49025
<i>Alicyclobacillus acidocaldarius</i>	ATCC27009
<i>Alicyclobacillus cycloheptanicus</i> *	Moore, J. Org. Chem. 62: págs. 2173, 1997.

*Usa cicloheptilcarbonil-CoA y no ciclohexilcarbonil-CoA como precursor para la biosíntesis de cyFA.

Niveles de saturación de aldehídos, alcoholes grasos y alquenos

El grado de saturación en derivados de ácidos grasos puede controlarse regulando el grado de saturación de productos intermedios de derivados de ácidos grasos. Por ejemplo, las familias de genes *sfa*, *gns* y *fab* pueden expresarse o sobreexpresarse para controlar la saturación de ácidos grasos. La figura 40 enumera ejemplos no limitativos de genes en estas familias de genes que pueden usarse en los métodos y las células huésped descritos en el presente documento.

Pueden modificarse por ingeniería genética células huésped para producir ácidos grasos insaturados modificando por ingeniería genética la célula huésped para sobreexpresar *fabB* o haciendo crecer la célula huésped a bajas temperaturas (por ejemplo, inferiores a 37°C). *FabB* tiene preferencia por *cis*- δ 3-decenoil-ACP y da como resultado la producción de ácidos grasos insaturados en *E. coli*. La sobreexpresión de *fabB* da como resultado la producción de un porcentaje significativo de ácidos grasos insaturados (de Mendoza *et al.*, J. Biol. Chem. 258:2098-2101, 1983). El gen *fabB* puede insertarse y expresarse en células huésped que no tienen de manera natural el gen. Estos derivados de ácidos grasos insaturados pueden usarse entonces como productos intermedios en células huésped que se modifican por ingeniería genética para producir derivados de ácidos grasos, tales como aldehídos grasos, alcoholes grasos o alquenos.

En otros casos, puede delecionarse un represor de la biosíntesis de ácidos grasos, por ejemplo, *fabR* (registro de GenBank NP_418398), lo que también dará como resultado un aumento de la producción de ácidos grasos insaturados en *E. coli* (Zhang *et al.*, J. Biol. Chem. 277:15558, 2002). Pueden realizarse delecciones similares en otras células huésped. Puede lograrse un aumento adicional de derivados de ácidos grasos insaturados, por ejemplo, sobreexpresando *fabM* (*trans*-2, *cis*-3-decenoil-ACP isomerasa, registro de GenBank DAA05501) y expresión controlada de *fabK* (*trans*-2-enoil-ACP reductasa II, registro de GenBank NP_357969) de *Streptococcus pneumoniae* (Marrakchi *et al.*, J. Biol. Chem. 277: 44809, 2002), mientras se deleciona *fabI* de *E. coli* (*trans*-2-enoil-ACP reductasa, registro de GenBank NP_415804). En algunos ejemplos, puede atenuarse el gen *fabF* endógeno, aumentando así el porcentaje de palmitoleato (C16:1) producido.

Otros sustratos

Otros sustratos que pueden usarse para producir aldehídos, alcoholes grasos, alcanos y alquenos en los métodos descritos en el presente documento son acil-ACP, acil-CoA, un aldehído graso o un alcohol graso, que se describen en, por ejemplo, el documento PCT/US08/058788. En la figura 40 se enumeran genes a modo de ejemplo que pueden alterarse para expresar o sobreexpresar estos sustratos en células huésped. Otros genes a modo de ejemplo se describen en el documento PCT/US08/058788.

Modificación por ingeniería genética de células huésped para producir aldehídos, alcoholes grasos, alcanos y alquenos

Pueden usarse diversas células huésped para producir aldehídos, alcoholes grasos, alcanos y/o alquenos, tal como se describe en el presente documento. Una célula huésped puede ser cualquier célula procariota o eucariota. Por ejemplo, un polipéptido descrito en el presente documento puede expresarse en células bacterianas (tales como *E. coli*), células de insecto, células de levadura o de mamífero (tales como células de ovario de hámster chino (CHO), células COS, células VERO, células BHK, células HeLa, células Cv1, células MDCK, células 293, células 3T3 o células PC 12). Otras células huésped a modo de ejemplo incluyen células de los miembros del género *Escherichia*, *Bacillus*, *Lactobacillus*, *Rhodococcus*, *Pseudomonas*, *Aspergillus*, *Trichoderma*, *Neurospora*, *Fusarium*, *Humicola*, *Rhizomucor*, *Kluyveromyces*, *Pichia*, *Mucor*, *Myceliophthora*, *Penicillium*, *Phanerochaete*, *Pleurotus*, *Trametes*, *Chrysosporium*, *Saccharomyces*, *Schizosaccharomyces*, *Yarrowia* o *Streptomyces*. Aún otras células huésped a modo de ejemplo pueden ser 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*, una célula de *Bacillus amiloliquefaciens*, 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 *Rhizomucor miehei*, una célula de *Mucor michei*, una célula de *Streptomyces lividans*, una célula de *Streptomyces murinus* o una célula de *Actinomycetes*.

En la tabla 1 se enumeran otros ejemplos no limitativos de células huésped.

En una realización preferida, la célula huésped es una célula de *E. coli*. En una realización más preferida, la célula huésped es de las cepas de *E. coli* B, C, K o W. Los expertos en la técnica conocen otras células huésped adecuadas.

Pueden usarse diversos métodos bien conocidos en la técnica para modificar por ingeniería genética células huésped para producir aldehídos, alcoholes grasos, alcanos y/o alquenos. Los métodos incluyen el uso de vectores, preferiblemente vectores de expresión, que contienen un ácido nucleico que codifica para un polipéptido de biosíntesis de aldehídos, alcoholes grasos, alcanos y/o alquenos descrito en el presente documento, o una variante de polipéptido o fragmento de la misma. 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 al que se ha unido. Un tipo de vector es un "plásmido", que se refiere a un bucle de ADN bicatenario circular en el que pueden ligarse segmentos de ADN adicionales. Otro tipo de vector es un vector viral, en el que pueden ligarse segmentos de ADN adicionales en el genoma viral. Determinados vectores pueden experimentar replicación autónoma en una célula huésped en la que se introducen (por ejemplo, vectores bacterianos que tienen un origen de replicación bacteriano y vectores de mamífero episomales). Otros vectores (por ejemplo, vectores de mamífero no episomales) se integran en el genoma de una célula huésped tras la introducción en la célula huésped y de ese modo se replican junto con el genoma del huésped. Además, determinados vectores, tales como vectores de expresión, pueden dirigir la expresión de genes a los que están operativamente unidos. En general, vectores de expresión usados en técnicas de ADN recombinante están a menudo en forma de plásmidos. Sin embargo, también pueden usarse otras formas de vectores de expresión, tales como vectores virales (por ejemplo, retrovirus, adenovirus y virus adenoasociados de replicación defectuosa).

Los vectores de expresión recombinante descritos en el presente documento incluyen un ácido nucleico descrito en el presente documento en una forma adecuada para la expresión del ácido nucleico en una célula huésped. Los vectores de expresión recombinantes pueden incluir una o más secuencias de control, seleccionadas basándose en la célula huésped que va a usarse para la expresión. La secuencia de control está operativamente unida a la secuencia de ácido nucleico que va a expresarse. Tales secuencias de control se describen, por ejemplo, en Goeddel, Gene Expression Technology: Methods in Enzymology 185, Academic Press, San Diego, Calif. (1990). Las secuencias de control incluyen aquellas que dirigen la expresión constitutiva de una secuencia de nucleótidos en muchos tipos de células huésped y aquellas que dirigen la expresión de la secuencia de nucleótidos sólo en determinadas células huésped (por ejemplo, secuencias reguladoras específicas de tejido). 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 de proteína 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 los ácidos nucleicos tal como se describe en el presente documento.

Pueden diseñarse vectores de expresión recombinante para la expresión de un gen que codifica para un polipéptido de biosíntesis de aldehídos, alcoholes grasos, alcanos y/o alquenos o variante en células procariontas o eucariotas (por ejemplo, células bacterianas, tales como *E. coli*, células de insecto (por ejemplo, usando vectores de expresión de baculovirus), células de levadura o células de mamífero). Se comentan adicionalmente células huésped adecuadas en Goeddel, Gene Expression Technology: Methods in Enzymology 185, Academic Press, San Diego, Calif. (1990). Alternativamente, el vector de expresión recombinante puede transcribirse y traducirse *in vitro*, por ejemplo, usando secuencias reguladoras de promotor de T7 y polimerasa de T7.

La expresión de polipéptidos en procariontas, 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 no de fusión. Los vectores de fusión añaden varios aminoácidos a un polipéptido codificado en los mismos, habitualmente en el extremo amino-terminal del polipéptido recombinante. Tales vectores de fusión sirven normalmente para tres fines: (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 la purificación por afinidad. Con frecuencia, en 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 tras 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; Smith *et al.*, Gene (1988) 67:31-40), pMAL (New England Biolabs, Beverly, Mass.) y pRITS (Pharmacia, 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* inducibles, no de fusión, 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 génica diana a partir del vector pTrc se basa en la transcripción de ARN polimerasa del huésped a partir de un promotor de fusión de trp-lac híbrido. La expresión génica diana a partir del vector pET 11d se basa en la transcripción de un promotor de fusión de gn10-lac de T7 mediada por una ARN polimerasa viral coexpresada (gn1 de T7). Esta polimerasa viral se suministra por cepas huésped BL21(DE3) o HMS174(DE3) a partir de un profago λ residente que aloja un gen gn1 de T7 bajo el control transcripcional del promotor lacUV 5.
- Una estrategia para maximizar la expresión de polipéptido recombinante es expresar el polipéptido en una célula huésped con una capacidad alterada para escindir proteolíticamente el polipéptido recombinante (véase Gottesman, Gene Expression Technology: Methods in Enzymology 185, Academic Press, San Diego, Calif. (1990) 119-128). Otra estrategia es alterar la secuencia de ácido nucleico que va a insertarse en un vector de expresión de modo que los codones individuales para cada aminoácido son los usados preferiblemente en la célula huésped (Wada *et al.*, Nucleic Acids Res. (1992) 20:2111-2118). Tal alteración de secuencias de ácido nucleico puede llevarse a cabo mediante técnicas de síntesis de ADN convencionales.
- En otra realización, la célula huésped es una célula de levadura. En esta realización, el vector de expresión es un vector de expresión de levadura. Los ejemplos de vectores para la expresión en levadura *S. cerevisiae* incluyen pYepSec1 (Baldari *et al.*, EMBO J. (1987) 6:229-234), pMFa (Kurjan *et al.*, Cell (1982) 30:933-943), pJRY88 (Schultz *et al.*, Gene (1987) 54:113-123), pYES2 (Invitrogen Corporation, San Diego, Calif.) y picZ (Invitrogen Corp, San Diego, Calif.).
- Alternativamente, puede expresarse un polipéptido descrito en el presente documento en células de insecto usando vectores 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 pAc (Smith *et al.*, Mol. Cell Biol. (1983) 3:2156-2165) y la serie pVL (Lucklow *et al.*, Virology (1989) 170:31-39).
- En aún otra realización, los ácidos nucleicos descritos en el presente documento pueden expresarse en células de mamífero usando un vector de expresión de mamífero. Los ejemplos de vectores de expresión de mamífero incluyen pCDM8 (Seed, Nature (1987) 329:840) y pMT2PC (Kaufman *et al.*, EMBO J. (1987) 6:187-195). Cuando se usan en células de mamífero, las funciones de control del vector de expresión pueden proporcionarse por elementos reguladores virales. Por ejemplo, los promotores comúnmente usados se derivan de polioma, adenovirus 2, citomegalovirus y virus del simio 40. Otros sistemas de expresión adecuados para células tanto procariotas como eucariotas se describen en los capítulos 16 y 17 de Sambrook *et al.*, eds., Molecular Cloning: A Laboratory Manual. 2ª, ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989.
- Pueden introducirse vectores en células procariotas o eucariotas mediante técnicas de transformación o transfección convencionales. Tal como se usan en el presente documento, los términos "transformación" y "transfección" se refieren a una variedad de técnicas reconocidas en la técnica para introducir ácido nucleico foráneo (por ejemplo, ADN) en una célula huésped, incluyendo coprecipitación con fosfato de calcio o cloruro de calcio, transfección mediada por DEAE-dextrano, lipofección o electroporación. Pueden encontrarse métodos adecuados para transformar o transfectar células huésped, por ejemplo, en 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 captarán y replicarán el vector de expresión. Con el fin de identificar y seleccionar estos transformantes, puede introducirse un gen que codifica para un marcador seleccionable (por ejemplo, resistencia a antibióticos) en las células huésped junto con el gen de interés. Los marcadores seleccionables incluyen los que confieren resistencia a fármacos, tales como ampicilina, canamicina, 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. Pueden identificarse células transfectadas de manera estable con el ácido nucleico introducido mediante selección con fármaco (por ejemplo, las células que han incorporado el gen de marcador seleccionable sobrevivirán, mientras que las otras células morirán).
- Para la transfección estable de células de mamífero, se sabe que, dependiendo del vector de expresión y la técnica de transfección usados, sólo una pequeña fracción de células pueden integrar el ADN foráneo en su genoma. Con el fin de identificar y seleccionar estos integrantes, puede introducirse un gen que codifica para un marcador seleccionable (por ejemplo, resistencia a antibióticos) en las células huésped junto con el gen de interés. Los marcadores seleccionables preferidos incluyen aquellos que confieren resistencia a fármacos, tales como G418, higromicina y metotrexato. 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. Pueden identificarse células transfectadas de manera estable con el ácido nucleico introducido mediante selección con fármaco (por ejemplo, las células que han incorporado el gen de marcador seleccionable sobrevivirán, mientras que las otras células morirán).
- En determinados métodos, se coexpresan un polipéptido de biosíntesis de aldehídos y un polipéptido de biosíntesis

de alcanos o alquenos en una única célula huésped. En métodos alternativos, se coexpresan un polipéptido de biosíntesis de aldehídos y un polipéptido de alcohol deshidrogenasa en una única célula huésped.

Proteínas de transporte

5 Las proteínas de transporte pueden exportar polipéptidos e hidrocarburos (por ejemplo, aldehídos, alcanos y/o alquenos) fuera de una célula huésped. Muchas proteínas de transporte y eflujo sirven para excretar una amplia variedad de compuestos y pueden modificarse de manera natural para ser selectivas para tipos particulares de hidrocarburos.

10 Ejemplos no limitativos de proteínas de transporte adecuadas son proteínas de transporte de casete de unión a ATP (ABC), proteínas de eflujo y proteínas transportadoras de ácidos grasos (FATP). Los ejemplos no limitativos adicionales de proteínas de transporte adecuadas incluyen las proteínas de transporte de ABC de microorganismos tales como *Caenorhabditis elegans*, *Arabidopsis thaliana*, *Alkaligenes eutrophus* y *Rhodococcus erythropolis*. En la figura 40 se indican proteínas de transporte de ABC a modo de ejemplo que pueden usarse (por ejemplo, CER5, AtMRP5, AmiS2 y AtPGP1). También pueden elegirse células huésped por su capacidad endógena para secretar hidrocarburos. La eficacia de la producción de hidrocarburos y secreción en el entorno de la célula huésped (por ejemplo, medio de cultivo, caldo de fermentación) puede expresarse como una razón de producto intracelular con respecto a producto extracelular. En algunos ejemplos, la razón puede ser de aproximadamente 5:1, 4:1, 3:1, 2:1, 1:1, 1:2, 1:3, 1:4 ó 1:5.

Fermentación

20 La producción y el aislamiento de aldehídos, alcoholes grasos, alcanos y/o alquenos pueden potenciarse empleando técnicas de fermentación beneficiosas. Un método para maximizar la producción al tiempo que se reducen los costes es aumentar el porcentaje de la fuente de carbono que se convierte en productos hidrocarbonados.

25 Durante ciclos de vida celulares normales, se usa carbono en funciones celulares, tales como producción de lípidos, sacáridos, proteínas, ácidos orgánicos y ácidos nucleicos. Reducir la cantidad de carbono necesaria para actividades relacionadas con el crecimiento puede aumentar la eficacia de conversión de la fuente de carbono en producto. Esto puede lograrse, por ejemplo, haciendo crecer en primer lugar células huésped hasta una densidad deseada (por ejemplo, una densidad que se alcanza al máximo de la fase de crecimiento logarítmica). En tal punto, pueden emplearse genes de punto de control de la replicación para detener el crecimiento de células. Específicamente, pueden usarse mecanismos de percepción de quórum (revisado en Camilli *et al.*, Science 311:1113, 2006; Venturi FEMS Microbio. Rev. 30:274-291, 2006; y Reading *et al.*, FEMS Microbiol. Lett. 254:1-11, 2006) para activar genes de punto de control, tales como *p53*, *p21* u otros genes de punto de control.

30 Los genes que pueden activarse para detener la replicación y el crecimiento celular en *E. coli* incluyen genes *umuDC*. La sobreexpresión de genes *umuDC* detiene la progresión desde la fase estacionaria hasta el crecimiento exponencial (Murlí *et al.*, J. of Bact. 182: 1127, 2000). UmuC es una ADN polimerasa que puede llevar a cabo la síntesis translesión sobre lesiones no codificantes (la base mecánica de la mayoría de las mutagénesis por UV y químicas). Los productos del gen *umuDC* están implicados en el proceso de síntesis translesión y también sirven como un punto de control de daño de la secuencia de ADN. Los productos del gen *umuDC* incluyen UmuC, UmuD, umuD', UmuD₂C, UmuD₂' y UmuD₂. Simultáneamente, pueden activarse genes que producen productos, minimizando así la necesidad de usar rutas de replicación y mantenimiento mientras se prepara un aldehído, alcano y/o alqueno. También pueden modificarse por ingeniería células huésped para expresar *umuC* y *umuD* de *E. coli* en pBAD24 bajo el sistema del promotor *prpBCDE* mediante síntesis *de novo* de este gen con los genes de producción de producto final apropiados.

45 El porcentaje de carbonos introducidos convertidos en aldehídos, alcoholes grasos, alcanos y/o alquenos puede ser un impulsor de coste. Cuanto más eficaz es el proceso (es decir, mayor es el porcentaje de carbonos introducidos convertidos en aldehídos, alcoholes grasos, alcanos y/o alquenos), menos caro será el proceso. Para fuentes de carbono que contienen oxígeno (por ejemplo, fuentes a base de glucosa y otros hidratos de carbono), el oxígeno debe liberarse en forma de dióxido de carbono. Por cada 2 átomos de oxígeno liberados, también se libera un átomo de carbono conduciendo a una eficacia metabólica teórica máxima de aproximadamente el 34% (p/p) (para productos derivados de ácidos grasos). Sin embargo, esta cifra cambia para otros productos hidrocarbonados y fuentes de carbono. Las eficacias típicas en la bibliografía son aproximadamente inferiores al 5%. Las células huésped modificadas por ingeniería genética para producir aldehídos, alcanos y/o alquenos pueden tener una eficacia superior a aproximadamente el 1, el 3, el 5, el 10, el 15, el 20, el 25 y el 30%. En un ejemplo, las células huésped pueden mostrar una eficacia de aproximadamente el 10% a aproximadamente el 25%. En otros ejemplos, tales células huésped pueden mostrar una eficacia de aproximadamente el 25% a aproximadamente el 30%. En otros ejemplos, las células huésped pueden mostrar una eficacia superior al 30%.

55 La célula huésped puede modificarse por ingeniería genética adicionalmente para expresar celulosomas recombinantes, tales como los descritos en la solicitud de PCT número PCT/US2007/003736. Estos celulosomas pueden permitir que la célula huésped use material celulósico como fuente de carbono. Por ejemplo, la célula huésped puede modificarse por ingeniería genética adicionalmente para expresar invertasas (EC 3.2.1.26) de modo

que puede usarse sacarosa como fuente de carbono. De manera similar, la célula huésped puede modificarse por ingeniería genética usando las enseñanzas descritas en las patentes estadounidenses n.ºs 5.000.000; 5.028.539; 5.424.202; 5.482.846 y 5.602.030; de modo que la célula huésped puede asimilar carbono eficazmente y usar materiales celulósicos como fuentes de carbono.

5 En un ejemplo, la cámara de fermentación puede encerrar una fermentación que está experimentando una reducción continua. En este caso, puede crearse un entorno reductor estable. El equilibrio electrónico puede mantenerse mediante la liberación de dióxido de carbono (en forma gaseosa). Esfuerzos para aumentar el equilibrio de NAD/H y NADP/H también pueden facilitar la estabilización del equilibrio electrónico. También puede potenciarse la disponibilidad de NADPH intracelular modificando por ingeniería genética la célula huésped para expresar una
10 NADH:NADPH transhidrogenasa. La expresión de una o más NADH:NADPH transhidrogenasas convierte la NADH producida en la glicólisis en NADPH, lo que puede potenciar la producción de aldehídos, alcanos y/o alquenos.

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 aldehídos, alcoholes grasos, alcanos y/o alquenos deseados basándose en los genes específicos
15 codificados en los plásmidos apropiados. Por ejemplo, pueden incubarse durante la noche células de *E. coli* BL21(DE3) que albergan pBAD24 (con resistencia a ampicilina y la ruta de aldehídos, alcoholes grasos, alcanos o alquenos) así como pUMVC1 (con resistencia a kanamicina y el sistema de sobreexpresión de acetil CoA/malonil CoA) en frascos de 2 l a 37°C agitados a > 200 rpm en 500 ml de medio LB complementado con ampicilina 75 µg/ml y kanamicina 50 µg/ml hasta que los cultivos alcanzan una $DO_{600} > 0,8$. Tras lograr una $DO_{600} > 0,8$, las células
20 pueden complementarse con propionato de sodio 25 mM (pH 8,0) para activar los sistemas de genes modificados por ingeniería genética para la producción y para detener la proliferación celular activando las proteínas UmuC y UmuD. La inducción puede realizarse durante 6 h a 30°C. Tras la incubación, los medios pueden examinarse para detectar aldehídos, alcoholes grasos, alcanos y/o alquenos usando CG-EM.

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 o más; fermentarse; e inducirse para expresar aldehídos, alcoholes grasos, alcanos y/o alquenos deseados basándose en los genes específicos codificados en los plásmidos apropiados. Por ejemplo, pueden incubarse células de *E. coli* BL21(DE3) que albergan pBAD24 (con resistencia a ampicilina y la ruta de síntesis de aldehídos y/o alcanos) así como pUMVC1 (con resistencia a kanamicina y el sistema de sobreexpresión de acetil-CoA/malonil-CoA) a partir de un cultivo simiente de 500 ml para fermentaciones
25 de 10 l (5 l para fermentaciones de 100 l, etc.) en medios LB (libres de glicerol) con kanamicina 50 µg/ml y ampicilina 75 µg/ml a 37°C, y agitarse a > 200 rpm hasta que los cultivos alcanzan una $DO_{600} > 0,8$ (normalmente 16 h). Pueden complementarse de manera continua los medios para mantener propionato de sodio 25 mM (pH 8,0) para activar los sistemas de genes modificados por ingeniería genética para la producción y para detener la proliferación celular activando las proteínas umuC y umuD. Los medios pueden complementarse de manera continua con glucosa
30 para mantener una concentración de 25 g/100 ml.

Tras la primera hora de inducción, pueden retirarse alícuotas de no más del 10% del volumen celular total cada hora y permitir que se asienten sin agitación para permitir que los aldehídos, alcanos y/o alquenos suban a la superficie y experimenten una separación de fases espontánea. El componente de aldehídos, alcoholes grasos, alcanos y/o alquenos puede recogerse entonces, y devolverse la fase acuosa a la cámara de reacción. La cámara de reacción
35 puede hacerse funcionar de manera continua. Cuando la DO_{600} desciende por debajo de 0,6, las células pueden reemplazarse por un nuevo lote hecho crecer a partir de un cultivo simiente.

Producción de aldehídos, alcoholes grasos, alcanos y alquenos usando métodos libres de células

En algunos métodos descritos en el presente documento, pueden producirse un aldehído, alcoholes grasos, alcanos y/o alquenos usando un polipéptido purificado descrito en el presente documento y un sustrato descrito en el presente documento. Por ejemplo, una célula huésped puede modificarse por ingeniería genética para expresar un polipéptido de biosíntesis de aldehídos, alcoholes grasos, alcanos y/o alquenos o variante tal como se describe en el presente documento. La célula huésped puede cultivarse en condiciones adecuadas para permitir la expresión del polipéptido. Entonces pueden generarse extractos libres de células usando métodos conocidos. Por ejemplo, las células huésped pueden lisarse usando detergentes o mediante sonicación. Los polipéptidos expresados pueden
45 purificarse usando métodos conocidos. Tras obtener los extractos libres de células, pueden añadirse los sustratos descritos en el presente documento a los extractos libres de células y mantenerse en condiciones que permiten la conversión de los sustratos en aldehídos, alcoholes grasos, alcanos y/o alquenos. Los aldehídos, alcoholes grasos, alcanos y/o alquenos pueden separarse entonces y purificarse usando técnicas conocidas.

Procesamiento tras la producción

55 Los aldehídos, alcoholes grasos, alcanos y/o alquenos producidos durante la fermentación pueden separarse de los medios de fermentación. Puede usarse cualquier técnica conocida para separar aldehídos, alcoholes grasos, alcanos y/o alquenos de medios acuosos. Un procedimiento de separación a modo de ejemplo es un procedimiento de separación de dos fases (bifásico). Este procedimiento implica fermentar las células huésped modificadas por ingeniería genética en condiciones suficientes para producir aldehídos, alcoholes grasos, alcanos y/o alquenos,

permitir que los aldehídos, alcoholes grasos, alcanos y/o alquenos se recojan en una fase orgánica y separar la fase orgánica del caldo de fermentación acuoso. Este método puede ponerse en práctica en entornos de fermentación tanto continuos como discontinuos.

5 La separación bifásica usa la relativa inmiscibilidad de los aldehídos, alcoholes grasos, alcanos y/o alquenos para facilitar la separación. Inmiscible se refiere a la relativa incapacidad de un compuesto de disolverse en agua y se define por el coeficiente de reparto del compuesto. Un experto habitual en la técnica apreciará que eligiendo un caldo de fermentación y una fase orgánica, de tal manera que el aldehído, alcano y/o alqueno que está produciéndose tiene un valor de logP alto, el aldehído, alcano y/o alqueno puede separarse en la fase orgánica, incluso a concentraciones muy bajas, en el recipiente de fermentación.

10 Los aldehídos, alcoholes grasos, alcanos y/o alquenos 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 aldehídos, alcoholes grasos, alcanos y/o alquenos pueden recogerse en una fase orgánica o bien de manera intracelular o bien extracelular. La recogida de los productos en la fase orgánica puede reducir el impacto de los aldehídos, alcoholes grasos, alcanos y/o alquenos sobre la función celular y puede permitir que la célula huésped produzca más producto.

15 Los métodos descritos en el presente documento pueden dar como resultado la producción de compuestos homogéneos en los que al menos aproximadamente el 60%, el 70%, el 80%, el 90% o el 95% de los aldehídos, alcoholes grasos, alcanos y/o alquenos producidos tendrán longitudes de la cadena de carbono que varían en menos de aproximadamente 6 carbonos, menos de aproximadamente 4 carbonos o menos de aproximadamente 2 carbonos. Estos compuestos también pueden producirse con un grado de saturación relativamente uniforme. Estos compuestos pueden usarse directamente como combustibles, aditivos para combustibles, productos químicos especializados, materiales de partida para la producción de otros compuestos químicos (por ejemplo, polímeros, tensioactivos, plásticos, materiales textiles, disolventes, adhesivos, etc.), o aditivos de productos para el cuidado personal. Estos compuestos también pueden usarse como materia prima para reacciones posteriores, por ejemplo, hidrogenación, craqueo catalítico (por ejemplo, mediante hidrogenación, pirólisis o ambas), para preparar otros productos.

20 En algunas realizaciones, los aldehídos, alcoholes grasos, alcanos y/o alquenos producidos usando métodos descritos en el presente documento pueden contener entre aproximadamente el 50% y aproximadamente el 90% de carbono; o entre aproximadamente el 5% y aproximadamente el 25% de hidrógeno. En otras realizaciones, los aldehídos, alcoholes grasos, alcanos y/o alquenos producidos usando métodos descritos en el presente documento pueden contener entre aproximadamente el 65% y aproximadamente el 85% de carbono; o entre aproximadamente el 10% y aproximadamente el 15% de hidrógeno.

Composiciones de combustible y composiciones de productos químicos especializados

35 Los aldehídos, alcoholes grasos, alcanos y/o alquenos descritos en el presente documento pueden usarse como o convertirse en un combustible o como un producto químico especializado. Un experto habitual en la técnica apreciará que, dependiendo del fin previsto del combustible o producto químico especializado, pueden producirse y usarse diferentes aldehídos, alcoholes grasos, alcanos y/o alquenos. Por ejemplo, un aldehído, alcohol graso, alcano y/o alqueno ramificado puede ser deseable para combustible de automóviles que está destinado a usarse en climas fríos. Además, cuando los aldehídos, alcoholes grasos, alcanos y/o alquenos descritos en el presente documento se usan como materia prima para la producción de combustible o productos químicos especializados, un experto habitual en la técnica apreciará que las características de la materia prima de aldehído, alcohol graso, alcano y/o alqueno afectarán a las características del combustible o producto químico especializado producido. Por tanto, pueden seleccionarse las características del combustible o producto químico especializado para la producción de aldehídos, alcoholes grasos, alcanos y/o alquenos particulares para su uso como materia prima.

40 Usando los métodos descritos en el presente documento, pueden producirse biocombustibles que tienen calidades de combustible deseadas a partir de aldehídos, alcoholes grasos, alcanos y/o alquenos. Los aldehídos, alcoholes grasos, alcanos y/o alquenos producidos de manera biológica representan una nueva fuente de biocombustibles, que pueden usarse como combustible para motores a reacción, diésel o gasolina. Algunos biocombustibles preparados usando aldehídos, alcoholes grasos, alcanos y/o alquenos no se han producido a partir de fuentes renovables y son nuevas composiciones de materia. Estos nuevos combustibles o productos químicos especializados pueden distinguirse de los combustibles o productos químicos especializados derivados de carbono petroquímico basándose en la huella de carbono isotópico doble. Adicionalmente, la fuente específica de carbono de fuentes biológicas (por ejemplo, glucosa frente a glicerol) puede determinarse mediante la huella de carbono isotópico doble (véase, por ejemplo, la patente estadounidense n.º 7.169.588, en particular de la col. 4, línea 31, a la col. 6, línea 8).

55 Los aldehídos, alcoholes grasos, alcanos y/o alquenos y los biocombustibles, productos químicos especializados y mezclas asociados pueden distinguirse de sus homólogos derivados de manera petroquímica basándose en ^{14}C (f_M) y la huella de carbono isotópico doble. En algunos ejemplos, el aldehído, alcohol graso, alcano y/o alqueno en la composición de biocombustible pueden tener una fracción de carbono moderno ($f_M^{14}\text{C}$) de, por ejemplo, al menos

aproximadamente 1,003, 1,010 ó 1,5.

En algunos ejemplos, puede prepararse una composición de biocombustible que incluye aldehídos, alcoholes grasos, alcanos y/o alquenos que tienen $\delta^{13}\text{C}$ de desde aproximadamente -15,4 hasta aproximadamente -10,9, en donde los aldehídos, alcoholes grasos, alcanos y/o alquenos representan al menos aproximadamente el 85% del material de fuentes biológicas (es decir, derivados de un recurso renovable, tal como biomasa, materiales celulósicos y azúcares) en la composición.

La capacidad para distinguir estos productos derivados de manera biológica es beneficiosa en el seguimiento de estos materiales en el comercio. Por ejemplo, pueden distinguirse combustibles o productos químicos especializados que comprenden perfiles de isótopos de carbono tanto derivados de manera biológica como a base de petróleo de combustibles y productos químicos especializados compuestos sólo por materiales a base de petróleo. Por tanto, los aldehídos, alcoholes grasos, alcanos y/o alquenos descritos en el presente documento pueden seguirse en el comercio o identificarse en el comercio como biocombustible basándose en su perfil único. Además, pueden identificarse otros materiales competidores como derivados de manera biológica o derivados de una fuente petroquímica.

Se usan aditivos para combustibles para potenciar el rendimiento de un combustible o motor. Por ejemplo, pueden usarse aditivos para combustibles para alterar el punto de congelación/gelificación, el punto de turbidez, la lubricidad, la viscosidad, la estabilidad oxidativa, la calidad de ignición, el nivel de octanos y/o el punto de inflamación. En los Estados Unidos, todos los aditivos para combustibles deben registrarse en la Environmental Protection Agency. Los nombres de aditivos para combustibles y las empresas que venden los aditivos para combustibles están disponibles públicamente poniéndose en contacto con la EPA o consultando la página web de la agencia. Un experto habitual en la técnica apreciará que los biocombustibles a base de aldehídos y/o alcanos descritos en el presente documento pueden mezclarse con uno o más aditivos para combustibles para conferir una calidad deseada.

Los biocombustibles a base de aldehídos, alcoholes grasos, alcanos y/o alquenos descritos en el presente documento pueden mezclarse con otros combustibles, tales como diversos alcoholes, tales como etanol y butanol, y productos derivados de petróleo, tales como gasolina, diésel o combustible para motores a reacción.

En algunos ejemplos, la mezcla puede incluir al menos aproximadamente el 10%, el 15%, el 20%, el 30%, el 40%, el 50% o el 60% en peso del aldehído, alcoholes grasos, alcano o alqueno. En otros ejemplos, puede prepararse una composición de biocombustible que incluye al menos aproximadamente el 5%, el 10%, el 20%, el 30%, el 40%, el 50%, el 60%, el 70%, el 80%, el 85%, el 90% o el 95% de un aldehído, alcoholes grasos, alcano o alqueno que incluye una cadena de carbono que tiene 8, 10, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21 ó 22 carbonos de longitud. Tales composiciones de biocombustible pueden incluir adicionalmente al menos un aditivo seleccionado de un aditivo que reduce el punto de turbidez que puede reducir el punto de turbidez hasta menos de aproximadamente 5°C o 0°C; un tensioactivo; una microemulsión; al menos aproximadamente el 5%, el 10%, el 15%, el 20%, el 30%, el 40%, el 50%, el 60%, el 70%, el 80%, el 85%, el 90% o el 95% de combustible diésel a partir de triglicéridos; gasolina derivada de petróleo; o combustible diésel a partir de petróleo.

Ejemplos

La invención se describe adicionalmente en los siguientes ejemplos, que no limitan el alcance de la invención descrito en las reivindicaciones.

Ejemplo 1. Detección y verificación de biosíntesis de alcanos en cianobacterias seleccionadas

Se seleccionaron siete cianobacterias, cuyas secuencias genómicas completas están disponibles públicamente, para la verificación y/o la detección de biosíntesis de alcanos: *Synechococcus elongatus* PCC7942, *Synechococcus elongatus* PCC6301, *Anabaena variabilis* ATCC29413, *Synechocystis* sp. PCC6803, *Nostoc punctiforme* PCC73102, *Gloeobacter violaceus* ATCC 29082 y *Prochlorococcus marinus* CCMP 1986. Sólo se había notificado previamente que las primeras tres cepas de cianobacterias de esta lista contenían alcanos (Han *et al.*, J. Am. Chem. Soc. 91:5156-5159 (1969); Fehler *et al.*, Biochem. 9:418-422 (1970)). Se hicieron crecer las cepas de manera fotoautótrofa en frascos con agitación en 100 ml de los medios apropiados (enumerados en la tabla 8) durante 3-7 días a 30°C a una intensidad luminosa de aproximadamente 3.500 lux. Se extrajeron las células para la detección de alcanos tal como sigue: se centrifugaron células de un volumen de cultivo de 1 ml durante 1 min a 13.000 rpm, se resuspendieron los sedimentos celulares en metanol, se agitaron con vórtex durante 1 min y luego se sonicaron durante 30 min. Tras la centrifugación durante 3 min a 13.000 rpm, se transfirieron los sobrenadantes a viales nuevos y se analizaron mediante CG-EM. Se analizaron las muestras o bien en una columna capilar DP-5 de 30 m (diámetro interno de 0,25 mm) o bien en una columna capilar DP-5 de 30 m a alta temperatura (diámetro interno de 0,25 mm) usando el siguiente método.

Tras una inyección sin dividir de 1 μl (temperatura de entrada mantenida a 300°C) en la columna de CG/EM, se mantuvo el horno a 100°C durante 3 min. Se aumentó la temperatura hasta 320°C a una velocidad de 20°C/min. Se mantuvo el horno a 320°C durante 5 min adicionales. La velocidad de flujo del gas helio portador fue de 1,3 ml/min. La EM de cuadrupolo exploró desde 50 hasta 550 m/z. Se compararon los patrones de fragmentación y tiempos de

retención de los picos de producto con referencias auténticas para confirmar la identidad del pico.

De las siete cepas, seis producían principalmente heptadecano y una producía pentadecano (*P. marinus* CCMP1986); una de estas cepas producía metil-heptadecano además de heptadecano (*A. variabilis* ATCC29413) (véase la tabla 8). Por tanto, se verificó la biosíntesis de alcanos en tres cianobacterias notificadas anteriormente y se detectó la biosíntesis de alcanos en cuatro cianobacterias que no se sabía previamente que producían alcanos: *P. marinus* CCMP1986 (véase la figura 1), *N. punctiforme* PCC73102 (véase la figura 2), *G. violaceus* ATCC 29082 (véase la figura 3) y *Synechocystis sp.* PCC6803 (véase la figura 4).

La figura 1A representa el perfil de CG/EM de células de *Prochlorococcus marinus* CCMP1986 extraídas con metanol. El pico a los 7,55 min tenía el mismo tiempo de retención que el pentadecano (Sigma). En la figura 1B, se muestra el patrón de fragmentación de masas del pico de pentadecano. El pico en 212 corresponde al peso molecular del pentadecano.

La figura 2A representa el perfil de CG/EM de células de *Nostoc punctiforme* PCC73102 extraídas con metanol. El pico a los 8,73 min tiene el mismo tiempo de retención que el heptadecano (Sigma). En la figura 2B, se muestra el patrón de fragmentación de masas del pico de heptadecano. El pico en 240 corresponde al peso molecular del heptadecano.

La figura 3A representa el perfil de CG/EM de células de *Gloeobacter violaceus* ATCC29082 extraídas con metanol. El pico a los 8,72 min tiene el mismo tiempo de retención que el heptadecano (Sigma). En la figura 3B, se muestra el patrón de fragmentación de masas del pico de heptadecano. El pico en 240 corresponde al peso molecular del heptadecano.

La figura 4A representa el perfil de CG/EM de células de *Synechocystis sp.* PCC6803 extraídas con metanol. El pico a los 7,36 min tiene el mismo tiempo de retención que el heptadecano (Sigma). En la figura 4B, se muestra el patrón de fragmentación de masas del pico de heptadecano. El pico en 240 corresponde al peso molecular del heptadecano.

Tabla 8: Hidrocarburos detectados en las cianobacterias seleccionadas

Cianobacteria	n.º de la ATCC	Genoma	Medio	Alcanos	
				notificados	verificados ²
<i>Synechococcus elongatus</i> PCC7942	27144	2,7 Mb	BG-11	C17:0	C17:0 , C15:0
<i>Synechococcus elongatus</i> PCC6301	33912	2,7 Mb	BG-11	C17:0	C17:0 , C15:0
<i>Anabaena variabilis</i>	29413	6,4 Mb	BG-11	C17:0. 7- u 8-Me-C17:0	C17:0 , Me-C17:0
<i>Synechocystis sp.</i> PCC6803	27184	3,5 Mb	BG-11	-	C17:0 , C15:0
<i>Prochlorococcus marinus</i> CCMP1986 ¹	-	1,7 Mb	-	-	C15:0
<i>Nostoc punctiforme</i> PCC73102	29133	9,0 Mb	ATCC819	-	C17:0
<i>Gloeobacter violaceus</i>	29082	4,6 Mb	BG11	-	C17:0

¹ las células para la extracción fueron un obsequio de Jacob Waldbauer (MIT)

² el hidrocarburo principal está en negrita

Los análisis genómicos produjeron dos genes que estaban presentes en las cepas productoras de alcanos. En la tabla 9 se representan los homólogos de estos genes de *Synechococcus elongatus* PCC7942 y son Synpcc7942_1593 (SEQ ID NO: 1) y Synpcc7942_1594 (SEQ ID NO: 65).

Tabla 9: Genes de cianobacterias productoras de alcanos

Id del gen objeto	Etiqueta de locus	Registro de Genbank	Nombre del gen	Longitud	COG	Pfam	InterPro	Notas
637800026	Synpcc7942_1593	YP_400610	proteína hipotética	231 aa		pfam02915	IPR009078 IPR003251	rubreritina de tipo ferritina/ribonucleótido reductasa
637800027	Synpcc7942_1594	YP_400611	proteína hipotética	341 aa	COG5322	pfam00106	IPR000408 IPR016040 IPR002198	deshidrogenasa prevista deshidrogenasa de cadena corta de unión a NAD(P)

Ejemplo 2. La delección de los genes sll0208 y sll0209 en *Synechocystis sp.* PCC6803 conduce a pérdida de biosíntesis de alcanos

Se deleccionaron los genes que codifican para la supuesta descarboxilasa (sll0208; NP_442147) (SEQ ID NO: 3) y la enzima generadora de aldehídos (sll0209; NP_442146) (SEQ ID NO: 67) de *Synechocystis sp.* PCC6803 tal como sigue. Se amplificó aproximadamente 1 kb de ADN flanqueante en el sentido de 3' y de 5' usando el cebador

5 sII0208/9-KO1 (CGCGGATCCCTTGATTCTACTGCGGCGAGT) con el cebador sII0208/9-KO2 (CACGCACCTAGGTTACACTCCCATGGTATAACAGGGGCGTTGGACTCCTGTG) y el cebador sII0208/9-KO3 (GTTATACCATGGGAGTGTGAACCTAGGTGCGTGCCGACAGGATAGGGCGTGT) con el cebador sII0208/9-KO4 (CGCGGATCCAACGCATCCTCACTAGTCGGG), respectivamente. Se usaron los productos de PCR en una PCR
 10 cruzada con los cebadores sII0208/9-KO1 y sII0208/9-KO4 para amplificar el casete de delección de aproximadamente 2 kb de sII0208/sII0209, que se clonó en el sitio *Bam*HI del vector de clonación pUC19. Entonces se amplificó un casete de resistencia a kanamicina (aph, KanR) a partir del plásmido pRL27 (Larsen *et al.*, Arch. Microbiol. 178:193 (2002)) usando los cebadores Kan-aph-F (CATGCCATGGAAAGCCACGTTGTGTCTCAAATCTCTG) y Kan-aph-R (CTAGTCTAGAGCGCTGAGGTCTGCCTCGTGAA), que se cortó luego con *Nco*I y *Xba*I y se clonó en los sitios *Nco*I y *Avr*II del casete de delección de sII0208/sII0209, creando un casete de delección de sII0208/sII0209 con inserción de KanR en pUC19. El vector que contenía el casete, que no se replica en cianobacterias, se transformó en *Synechocystis* sp. PCC6803 (Zang *et al.*, 2007, J. Microbiol., vol. 45, págs. 241) y se seleccionaron transformantes (por ejemplo, integrantes cromosómicos mediante recombinación homóloga doble) en placas de agar BG-11 que contenían kanamicina 100 µg/ml en un incubador equipado con luz a 30°C. Las colonias resistentes a kanamicina volvieron a sembrarse en estrías una vez y luego se sometieron a análisis genotípico usando PCR con cebadores de diagnóstico.

20 Se cultivaron los mutantes de delección-inserción confirmados en 12 ml de medio BG11 con kanamicina 50 µg/ml durante 4 días a 30°C en un agitador-incubador equipado con luz. Entonces se centrifugó 1 ml de caldo (1 min a 13.000 g) y se extrajeron los sedimentos celulares con 0,1 ml de metanol. Tras la extracción, se centrifugaron de nuevos las muestras y se sometieron los sobrenadantes a análisis de CG-EM tal como se describe en el ejemplo 1.

25 Tal como se muestra en la figura 5, las cepas de *Synechocystis* sp. PCC6803 en las que se deleccionaron los genes sII0208 y sII0209 perdieron su capacidad para producir heptadeceno y octadecenal. Este resultado demuestra que los genes sII0208 y sII0209 en *Synechocystis* sp. PCC6803 y los genes ortólogos en otras cianobacterias (véase la tabla 1) son responsables de la biosíntesis de alcanos y aldehídos grasos en estos microorganismos.

Ejemplo 3. Producción de aldehídos grasos y alcoholes grasos en *E. coli* a través de la expresión heteróloga del orf1594 de *Synechococcus elongatus* PCC7942

30 Se amplificó el ADN genómico que codifica para el orf1594 de *Synechococcus elongatus* PCC7942 (YP_400611; supuesta enzima generadora de aldehídos) (SEQ ID NO: 65) y se clonó en los sitios *Nco*I y *Eco*RI del vector OP-80 (derivado de pCL1920) bajo el control del promotor P_{trc} . Se transformó el constructo resultante ("OP80-PCC7942_1594") en *E. coli* MG1655 y se hicieron crecer las células a 37°C en medio mínimo M9 con glucosa al 1% (p/v) como fuente de carbono y complementado con espectinomicina 100 µg/ml. Cuando el cultivo alcanzó una DO_{600} de 0,8-1,0, se indujo con IPTG 1 mM y se hicieron crecer las células durante 18-20 h adicionales a 37°C. Se extrajeron las células de 0,5 ml de cultivo con 0,5 ml de acetato de etilo. Tras la sonicación durante 60 min, se centrifugó la muestra a 15.000 rpm durante 5 min. Se analizó la capa de disolvente mediante CG-EM tal como se describe en el ejemplo 1.

40 Tal como se muestra en la figura 6, las células de *E. coli* transformadas con el vector que portaba orf1594 de *Synechococcus elongatus* PCC7942 produjeron los siguientes aldehídos grasos y alcoholes grasos: hexadecanal, octadecenal, tetradecenal, hexadecenal, hexadecanol y octadecenal. Este resultado indica que orf1594 de PCC7942 (i) genera aldehídos *in vivo* como posibles sustratos para la descarbonilación y (ii) puede reducir acil-ACP como sustratos, que son la forma más abundante de ácidos grasos activados en células de *E. coli* silvestres. Por tanto, la enzima se denominó acil-ACP reductasa. *In vivo*, los aldehídos grasos aparentemente se reducen adicionalmente para dar los correspondientes alcoholes grasos mediante una actividad aldehído reductasa endógena de *E. coli*.

45 Ejemplo 4. Producción de aldehídos grasos y alcoholes grasos en *E. coli* a través de la expresión heteróloga de cce_1430 de *Cyanothece* sp. ATCC51142

50 Se amplificó el ADN genómico que codifica para cce_1430 de *Cyanothece* sp. ATCC51142 (YP_001802846; supuesta enzima generadora de aldehídos) (SEQ ID NO: 69) y se clonó en los sitios *Nco*I y *Eco*RI del vector OP-80 (derivado de pCL1920) bajo el control del promotor P_{trc} . Se transformó el constructo resultante en *E. coli* MG1655 y se hicieron crecer las células a 37°C en medio mínimo M9 con glucosa al 1% (p/v) como fuente de carbono y complementado con espectinomicina 100 µg/ml. Se cultivaron las células y se extrajeron como en el ejemplo 3 y se analizaron mediante CG-EM tal como se describe en el ejemplo 26.

55 Tal como se muestra en la figura 7, las células de *E. coli* transformadas con el vector que portaba cce_1430 de *Cyanothece* sp. ATCC51142 produjeron los siguientes aldehídos grasos y alcoholes grasos: hexadecanal, octadecenal, tetradecenal, hexadecenal, hexadecanol y octadecenal. Este resultado indica que cce_1430 de ATCC51142 (i) genera aldehídos *in vivo* como posibles sustratos para la descarbonilación y (ii) puede reducir acil-ACP como sustratos, que son la forma más abundante de ácidos grasos activados en células de *E. coli* silvestres. Por tanto, esta enzima también es una acil-ACP reductasa.

Ejemplo 5. Producción de alcanos y alquenos en *E. coli* a través de la expresión heteróloga del orf1594 de

Synechococcus elongatus PCC7942 y el orf1593 de Synechococcus elongatus PCC7942

Se amplificó el ADN genómico que codifica para el orf1593 de *Synechococcus elongatus* PCC7942 (YP_400610; supuesta descarboxilasa) (SEQ ID NO: 1) y se clonó en los sitios *NdeI* y *XhoI* del vector OP-183 (derivado de pACYC) bajo el control del promotor P_{trc} . Se contrató el constructo resultante con OP80-PCC7942_1594 en *E. coli* MG1655 y se hicieron crecer las células a 37°C en medio mínimo M9 complementado con espectinomicina 100 µg/ml y carbenicilina 100 µg/ml. Se cultivaron las células y se extrajeron como en el ejemplo 3 y se analizaron mediante CG-EM tal como se describe en el ejemplo 1.

Tal como se muestra en la figura 8, las células de *E. coli* cotransformadas con los vectores que portaban PCC7942_1594 de *S. elongatus* y PCC7942_1593 de *S. elongatus* produjeron los mismos aldehídos grasos y alcoholes grasos que en el ejemplo 3, pero también pentadecano y heptadeceno. Este resultado indica que PCC7942_1593 en *E. coli* convierte hexadecanal y octadecenal en pentadecano y heptadeceno, respectivamente, y por tanto es una aldehído graso descarboxilasa activa.

Ejemplo 6. Producción de alcanos y alquenos en E. coli a través de la expresión heteróloga del orf1594 de Synechococcus elongatus PCC7942 y Npun02004178 de Nostoc punctiforme PCC73102

Se amplificó el ADN genómico que codifica para Npun02004178 de *Nostoc punctiforme* PCC73102 (ZP_00108838; supuesta descarboxilasa) (SEQ ID NO: 5) y se clonó en los sitios *NdeI* y *XhoI* del vector OP-183 (derivado de pACYC) bajo el control del promotor P_{trc} . Se contrató el constructo resultante con OP80-PCC7942_1594 en *E. coli* MG1655 y se hicieron crecer las células a 37°C en medio mínimo M9 complementado con espectinomicina 100 µg/ml y carbenicilina 100 µg/ml. Se cultivaron las células y se extrajeron como en el ejemplo 3 y se analizaron mediante CG-EM tal como se describe en el ejemplo 1.

Tal como se muestra en la figura 9, las células de *E. coli* cotransformadas con los vectores que portaban PCC7942_1594 de *S. elongatus* y Npun02004178 de *N. punctiforme* PCC73102 produjeron los mismos aldehídos grasos y alcoholes grasos que en el ejemplo 3, pero también tridecano, pentadeceno, pentadecano y heptadeceno. Este resultado indica que Npun02004178 en *E. coli* convierte tetradecanal, hexadecenal, hexadecanal y octadecenal en tridecano, pentadeceno, pentadecano y heptadeceno, respectivamente, y por tanto es una aldehído graso descarboxilasa activa.

Ejemplo 7. Producción de alcanos y alquenos en E. coli a través de la expresión heteróloga del orf1594 de Synechococcus elongatus PCC7942 y sll0208 de Synechocystis sp. PCC6803

Se amplificó el ADN genómico que codifica para sll0208 de *Synechocystis* sp. PCC6803 (NP_442147; supuesta descarboxilasa) (SEQ ID NO: 3) y se clonó en los sitios *NdeI* y *XhoI* del vector OP-183 (derivado de pACYC) bajo el control del promotor P_{trc} . Se contrató el constructo resultante con OP80-PCC7942_1594 en *E. coli* MG1655 y se hicieron crecer las células a 37°C en medio mínimo M9 complementado con espectinomicina 100 µg/ml y carbenicilina 100 µg/ml. Se cultivaron las células y se extrajeron como en el ejemplo 3 y se analizaron mediante CG-EM tal como se describe en el ejemplo 1.

Tal como se muestra en la figura 10, las células de *E. coli* cotransformadas con los vectores que portaban PCC7942_1594 de *S. elongatus* y sll0208 de *Synechocystis* sp. PCC6803 produjeron los mismos aldehídos grasos y alcoholes grasos que en el ejemplo 3, pero también pentadecano y heptadeceno. Este resultado indica que Npun02004178 en *E. coli* convierte hexadecanal y octadecenal en pentadecano y heptadeceno, respectivamente, y por tanto es una aldehído graso descarboxilasa activa.

Ejemplo 8. Producción de alcanos y alquenos en E. coli a través de la expresión heteróloga del orf1594 de Synechococcus elongatus PCC7942 y alr5283 de Nostoc sp. PCC7210

Se amplificó el ADN genómico que codifica para alr5283 de *Nostoc* sp. PCC7210 (NP_489323; supuesta descarboxilasa) (SEQ ID NO: 7) y se clonó en los sitios *NdeI* y *XhoI* del vector OP-183 (derivado de pACYC) bajo el control del promotor P_{trc} . Se contrató el constructo resultante con OP80-PCC7942_1594 en *E. coli* MG1655 y se hicieron crecer las células a 37°C en medio mínimo M9 complementado con espectinomicina 100 µg/ml y carbenicilina 100 µg/ml. Se cultivaron las células y se extrajeron como en el ejemplo 3 y se analizaron mediante CG-EM tal como se describe en el ejemplo 1.

Tal como se muestra en la figura 11, las células de *E. coli* cotransformadas con los vectores que portaban PCC7942_1594 de *S. elongatus* y alr5283 de *Nostoc* sp. PCC7210 produjeron los mismos aldehídos grasos y alcoholes grasos que en el ejemplo 3, pero también pentadecano y heptadeceno. Este resultado indica que alr5283 en *E. coli* convierte hexadecanal y octadecenal en pentadecano y heptadeceno, respectivamente, y por tanto es una aldehído graso descarboxilasa activa.

Ejemplo 9. Producción de alcanos y alquenos en E. coli a través de la expresión heteróloga del orf1594 de Synechococcus elongatus PCC7942 y AM1_4041 de Acaryochloris marina MBIC11017

Se realizó la optimización de codones del ADN genómico que codifica para AM1_4041 de *Acaryochloris marina*

MBIC11017 (YP_001518340; supuesta descarboxilasa) (SEQ ID NO: 9) para su expresión en *E. coli* (SEQ ID NO: 46), se sintetizó y se clonó en los sitios *NdeI* y *XhoI* del vector OP-183 (derivado de pACYC) bajo el control del promotor P_{trc} . Se contranformó el constructo resultante con OP80-PCC7942_1594 en *E. coli* MG1655 y se hicieron crecer las células a 37°C en medio mínimo M9 complementado con espectinomicina 100 µg/ml y carbenicilina 100 µg/ml. Se cultivaron las células y se extrajeron como en el ejemplo 3 y se analizaron mediante CG-EM tal como se describe en el ejemplo 26.

Tal como se muestra en la figura 12, las células de *E. coli* cotransformadas con los vectores que portaban PCR7942_1594 de *S. elongatus* y AM1_4041 de *A. marina* MBIC11017 produjeron los mismos aldehídos grasos y alcoholes grasos que en el ejemplo 3, pero también tridecano, pentadeceno, pentadecano y heptadeceno. Este resultado indica que AM1_4041 en *E. coli* convierte tetradecanal, hexadecanal, hexadecanal y octadecanal en tridecano, pentadeceno, pentadecano y heptadeceno, respectivamente, y por tanto es una aldehído graso descarboxilasa activa.

Ejemplo 10. Producción de alcanos y alquenos en *E. coli* a través de la expresión heteróloga del orf1594 de *Synechococcus elongatus* PCC7942 y tll1313 de *Thermosynechococcus elongatus* BP-1

Se realizó la optimización de codones del ADN genómico que codifica para tll1313 de *Thermosynechococcus elongatus* BP-1 (NP_682103; supuesta descarboxilasa) (SEQ ID NO: 11) para su expresión en *E. coli* (SEQ ID NO: 47), se sintetizó y se clonó en los sitios *NdeI* y *XhoI* del vector OP-183 (derivado de pACYC) bajo el control del promotor P_{trc} . Se contranformó el constructo resultante con OP80-PCC7942_1594 en *E. coli* MG1655 y se hicieron crecer las células a 37°C en medio mínimo M9 complementado con espectinomicina 100 µg/ml y carbenicilina 100 µg/ml. Se cultivaron las células y se extrajeron como en el ejemplo 3 y se analizaron mediante CG-EM tal como se describe en el ejemplo 26.

Tal como se muestra en la figura 13, las células de *E. coli* cotransformadas con los vectores que portaban PCC7942_1594 de *S. elongatus* y tll1313 de *T. elongatus* BP-1 produjeron los mismos aldehídos grasos y alcoholes grasos que en el ejemplo 3, pero también pentadecano y heptadeceno. Este resultado indica que tll1313 en *E. coli* convierte hexadecanal y octadecanal en pentadecano y heptadeceno, respectivamente, y por tanto es una aldehído graso descarboxilasa activa.

Ejemplo 11. Producción de alcanos y alquenos en *E. coli* a través de la expresión heteróloga del orf1594 de *Synechococcus elongatus* PCC7942 y CYA_0415 de *Synechococcus sp.* JA-3-3Ab

Se realizó la optimización de codones del ADN genómico que codifica para CYA_0415 de *Synechococcus sp.* JA-3-3Ab (YP_473897; supuesta descarboxilasa) (SEQ ID NO: 13) para su expresión en *E. coli* (SEQ ID NO: 48), se sintetizó y se clonó en los sitios *NdeI* y *XhoI* del vector OP-183 (derivado de pACYC) bajo el control del promotor P_{trc} . Se contranformó el constructo resultante con OP80-PCC7942_1594 en *E. coli* MG1655 y se hicieron crecer las células a 37°C en medio mínimo M9 complementado con espectinomicina 100 µg/ml y carbenicilina 100 µg/ml. Se cultivaron las células y se extrajeron como en el ejemplo 3 y se analizaron mediante CG-EM tal como se describe en el ejemplo 26.

Tal como se muestra en la figura 14, las células de *E. coli* cotransformadas con los vectores que portaban PCC7942_1594 de *S. elongatus* y CYA_0415 de *Synechococcus sp.* JA-3-3Ab produjeron los mismos aldehídos grasos y alcoholes grasos que en el ejemplo 3, pero también pentadecano y heptadeceno. Este resultado indica que Npun02004178 en *E. coli* convierte hexadecanal y octadecanal en pentadecano y heptadeceno, respectivamente, y por tanto es una aldehído graso descarboxilasa activa.

Ejemplo 12. Producción de alcanos y alquenos en *E. coli* a través de la expresión heteróloga del orf1594 de *Synechococcus elongatus* PCC7942 y gll3146 de *Gloeobacter violaceus* PCC7421

Se amplificó el ADN genómico que codifica para gll3146 de *Gloeobacter violaceus* PCC7421 (NP_926092; supuesta descarboxilasa) (SEQ ID NO: 15) y se clonó en los sitios *NdeI* y *XhoI* del vector OP-183 (derivado de pACYC) bajo el control del promotor P_{trc} . Se contranformó el constructo resultante con OP80-PCC7942_1594 en *E. coli* MG1655 y se hicieron crecer las células a 37°C en medio mínimo M9 complementado con espectinomicina 100 µg/ml y carbenicilina 100 µg/ml. Se cultivaron las células y se extrajeron como en el ejemplo 3 y se analizaron mediante CG-EM tal como se describe en el ejemplo 1.

Tal como se muestra en la figura 15, las células de *E. coli* cotransformadas con los vectores que portaban PCC7942_1594 de *S. elongatus* y gll3146 de *G. violaceus* PCC7421 produjeron los mismos aldehídos grasos y alcoholes grasos que en el ejemplo 3, pero también pentadecano y heptadeceno. Este resultado indica que gll3146 en *E. coli* convierte hexadecanal y octadecanal en pentadecano y heptadeceno, respectivamente, y por tanto es una aldehído graso descarboxilasa activa.

Ejemplo 13. Producción de alcanos y alquenos en *E. coli* a través de la expresión heteróloga del orf1594 de *Synechococcus elongatus* PCC7942 y PMT1231 de *Prochlorococcus marinus*

Se realizó la optimización de codones del ADN genómico que codifica para PMT1231 de *Prochlorococcus marinus*

- 5 MIT9313 (NP_895059; supuesta descarbonilasa) (SEQ ID NO: 17) para su expresión en *E. coli* (SEQ ID NO: 49), se sintetizó y se clonó en los sitios *Nde*I y *Xho*I del vector OP-183 (derivado de pACYC) bajo el control del promotor P_{trc}. Se contranformó el constructo resultante con OP80-PCC7942_1594 en *E. coli* MG1655 y se hicieron crecer las células a 37°C en medio mínimo M9 complementado con espectinomomicina 100 µg/ml y carbenicilina 100 µg/ml. Se cultivaron las células y se extrajeron como en el ejemplo 3 y se analizaron mediante CG-EM tal como se describe en el ejemplo 26.
- 10 Tal como se muestra en la figura 16, las células de *E. coli* cotransformadas con los vectores que portaban PCC7942_1594 de *S. elongatus* y PMT1231 de *P. marinus* MIT9313 produjeron los mismos aldehídos grasos y alcoholes grasos que en el ejemplo 3, pero también pentadecano y heptadeceno. Este resultado indica que PMT1231 en *E. coli* convierte hexadecanal y octadecenal en pentadecano y heptadeceno, respectivamente, y por tanto es una aldehído graso descarbonilasa activa.
- Ejemplo 14. Producción de alcanos y alquenos en *E. coli* a través de la expresión heteróloga del orf1594 de *Synechococcus elongatus* PCC7942 y PMM0532 de *Prochlorococcus marinus* CCMP1986
- 15 Se amplificó el ADN genómico que codifica para PMM0532 de *Prochlorococcus marinus* CCMP1986 (NP_892650; supuesta descarbonilasa) (SEQ ID NO: 19) y se clonó en los sitios *Nde*I y *Xho*I del vector OP-183 (derivado de pACYC) bajo el control del promotor P_{trc}. Se contranformó el constructo resultante con OP80-PCC7942_1594 en *E. coli* MG1655 y se hicieron crecer las células a 37°C en medio mínimo M9 complementado con espectinomomicina 100 µg/ml y carbenicilina 100 µg/ml. Se cultivaron las células y se extrajeron como en el ejemplo 3 y se analizaron mediante CG-EM tal como se describe en el ejemplo 1.
- 20 Tal como se muestra en la figura 17, las células de *E. coli* cotransformadas con los vectores que portaban PCC7942_1594 de *S. elongatus* y PMM0532 de *P. marinus* CCMP1986 produjeron los mismos aldehídos grasos y alcoholes grasos que en el ejemplo 3, pero también pentadecano y heptadeceno. Este resultado indica que PMM0532 en *E. coli* convierte hexadecanal y octadecenal en pentadecano y heptadeceno, respectivamente, y por tanto es una aldehído graso descarbonilasa activa.
- 25 Ejemplo 15. Producción de alcanos y alquenos en *E. coli* a través de la expresión heteróloga del orf1594 de *Synechococcus elongatus* PCC7942 y PMN2A_1863 de *Prochlorococcus marinus* NATL2A
- 30 Se realizó la optimización de codones del ADN genómico que codifica para PMN2A_1863 de *Prochlorococcus marinus* NATL2A (YP_293054; supuesta descarbonilasa) (SEQ ID NO: 21) para su expresión en *E. coli* (SEQ ID NO: 51), se sintetizó y se clonó en los sitios *Nde*I y *Xho*I del vector OP-183 (derivado de pACYC) bajo el control del promotor P_{trc}. Se contranformó el constructo resultante con OP80-PCC7942_1594 en *E. coli* MG1655 y se hicieron crecer las células a 37°C en medio mínimo M9 complementado con espectinomomicina 100 µg/ml y carbenicilina 100 µg/ml. Se cultivaron las células y se extrajeron como en el ejemplo 3 y se analizaron mediante CG-EM tal como se describe en el ejemplo 26.
- 35 Tal como se muestra en la figura 18, las células de *E. coli* cotransformadas con los vectores que portaban PCC7942_1594 de *S. elongatus* y PMN2A_1863 de *P. marinus* NATL2A produjeron los mismos aldehídos grasos y alcoholes grasos que en el ejemplo 3, pero también pentadecano y heptadeceno. Este resultado indica que PMN2A_1863 en *E. coli* convierte hexadecanal y octadecenal en pentadecano y heptadeceno, respectivamente, y por tanto es una aldehído graso descarbonilasa activa.
- 40 Ejemplo 16. Producción de alcanos y alquenos en *E. coli* a través de la expresión heteróloga del orf1594 de *Synechococcus elongatus* PCC7942 y RS9917_09941 de *Synechococcus* sp. RS9917
- 45 Se realizó la optimización de codones del ADN genómico que codifica para RS9917_09941 de *Synechococcus* sp. RS9917 (ZP_01079772; supuesta descarbonilasa) (SEQ ID NO: 23) para su expresión en *E. coli* (SEQ ID NO: 52), se sintetizó y se clonó en los sitios *Nde*I y *Xho*I del vector OP-183 (derivado de pACYC) bajo el control del promotor P_{trc}. Se contranformó el constructo resultante con OP80-PCC7942_1594 en *E. coli* MG1655 y se hicieron crecer las células a 37°C en medio mínimo M9 complementado con espectinomomicina 100 µg/ml y carbenicilina 100 µg/ml. Se cultivaron las células y se extrajeron como en el ejemplo 3 y se analizaron mediante CG-EM tal como se describe en el ejemplo 26.
- 50 Tal como se muestra en la figura 19, las células de *E. coli* cotransformadas con los vectores que portaban PCC7942_1594 de *S. elongatus* y RS9917_09941 de *Synechococcus* sp. RS9917 produjeron los mismos aldehídos grasos y alcoholes grasos que en el ejemplo 3, pero también pentadecano y heptadeceno. Este resultado indica que RS9917_09941 en *E. coli* convierte hexadecanal y octadecenal en pentadecano y heptadeceno, respectivamente, y por tanto es una aldehído graso descarbonilasa activa.
- Ejemplo 17. Producción de alcanos y alquenos en *E. coli* a través de la expresión heteróloga del orf1594 de *Synechococcus elongatus* PCC7942 y RS9917_12945 de *Synechococcus* sp. RS9917
- 55 Se realizó la optimización de codones del ADN genómico que codifica para RS9917_12945 de *Synechococcus* sp. RS9917 (ZP_01080370; supuesta descarbonilasa) (SEQ ID NO: 25) para su expresión en *E. coli* (SEQ ID NO: 53),

se sintetizó y se clonó en los sitios *NdeI* y *XhoI* del vector OP-183 (derivado de pACYC) bajo el control del promotor P_{trc} . Se contrató el constructo resultante con OP80-PCC7942_1594 en *E. coli* MG1655 y se hicieron crecer las células a 37°C en medio mínimo M9 complementado con espectinomicina 100 µg/ml y carbenicilina 100 µg/ml. Se cultivaron las células y se extrajeron como en el ejemplo 3 y se analizaron mediante CG-EM tal como se describe en el ejemplo 26.

Tal como se muestra en la figura 20, las células de *E. coli* cotransformadas con los vectores que portaban PCC7942_1594 de *S. elongatus* y RS9917_12945 de *Synechococcus* sp. RS9917 produjeron los mismos aldehídos grasos y alcoholes grasos que en el ejemplo 3, pero también pentadecano y heptadeceno. Este resultado indica que RS9917_12945 en *E. coli* convierte hexadecanal y octadecenal en pentadecano y heptadeceno, respectivamente, y por tanto es una aldehído graso descarboxilasa activa.

Ejemplo 18. Producción de alcanos y alquenos en *E. coli* a través de la expresión heteróloga del orf1594 de *Synechococcus elongatus* PCC7942 y cce_0778 de *Cyanothece* sp. ATCC51142

Se sintetizó el ADN genómico que codifica para cce_0778 de *Cyanothece* sp. ATCC51142 (YP_001802195; supuesta descarboxilasa) (SEQ ID NO: 27) y se clonó en los sitios *NdeI* y *XhoI* del vector OP-183 (derivado de pACYC) bajo el control del promotor P_{trc} . Se contrató el constructo resultante con OP80-PCC7942_1594 en *E. coli* MG1655 y se hicieron crecer las células a 37°C en medio mínimo M9 complementado con espectinomicina 100 µg/ml y carbenicilina 100 µg/ml. Se cultivaron las células y se extrajeron como en el ejemplo 3 y se analizaron mediante CG-EM tal como se describe en el ejemplo 26.

Tal como se muestra en la figura 21, las células de *E. coli* cotransformadas con los vectores que portaban PCC7942_1594 de *S. elongatus* y cce_0778 de *Cyanothece* sp. ATCC51142 produjeron los mismos aldehídos grasos y alcoholes grasos que en el ejemplo 3, pero también tridecano, pentadeceno, pentadecano y heptadeceno. Este resultado indica que ATCC51142 cce_0778 en *E. coli* convierte tetradecanal, hexadecenal, hexadecanal y octadecenal en tridecano, pentadeceno, pentadecano y heptadeceno, respectivamente, y por tanto es una aldehído graso descarboxilasa activa.

Ejemplo 19. Producción de alcanos y alquenos en *E. coli* a través de la expresión heteróloga del orf1594 de *Synechococcus elongatus* PCC7942 y Cyan7425_0398 de *Cyanothece* sp. PCC7425

Se sintetizó el ADN genómico que codifica para Cyan7425_0398 de *Cyanothece* sp. PCC7425 (YP_002481151; supuesta descarboxilasa) (SEQ ID NO: 29) y se clonó en los sitios *NdeI* y *XhoI* del vector OP-183 (derivado de pACYC) bajo el control del promotor P_{trc} . Se contrató el constructo resultante con OP80-PCC7942_1594 en *E. coli* MG1655 y se hicieron crecer las células a 37°C en medio mínimo M9 complementado con espectinomicina 100 µg/ml y carbenicilina 100 µg/ml. Se cultivaron las células y se extrajeron como en el ejemplo 3 y se analizaron mediante CG-EM tal como se describe en el ejemplo 26.

Tal como se muestra en la figura 22, las células de *E. coli* cotransformadas con los vectores que portaban PCC7942_1594 de *S. elongatus* y Cyan7425_0398 de *Cyanothece* sp. PCC7425 produjeron los mismos aldehídos grasos y alcoholes grasos que en el ejemplo 3, pero también tridecano, pentadeceno, pentadecano y heptadeceno. Este resultado indica que Cyan7425_0398 en *E. coli* convierte tetradecanal, hexadecenal, hexadecanal y octadecenal en tridecano, pentadeceno, pentadecano y heptadeceno, respectivamente, y por tanto es una aldehído graso descarboxilasa activa.

Ejemplo 20. Producción de alcanos y alquenos en *E. coli* a través de la expresión heteróloga del orf1594 de *Synechococcus elongatus* PCC7942 y Cyan7425_2986 de *Cyanothece* sp. PCC7425

Se sintetizó el ADN genómico que codifica para Cyan7425_2986 de *Cyanothece* sp. PCC7425 (YP_002483683; supuesta descarboxilasa) (SEQ ID NO: 31) y se clonó en los sitios *NdeI* y *XhoI* del vector OP-183 (derivado de pACYC) bajo el control del promotor P_{trc} . Se contrató el constructo resultante con OP80-PCC7942_1594 en *E. coli* MG1655 y se hicieron crecer las células a 37°C en medio mínimo M9 complementado con espectinomicina 100 µg/ml y carbenicilina 100 µg/ml. Se cultivaron las células y se extrajeron como en el ejemplo 3 y se analizaron mediante CG-EM tal como se describe en el ejemplo 26.

Tal como se muestra en la figura 23, las células de *E. coli* cotransformadas con los vectores que portaban PCC7942_1594 de *S. elongatus* y Cyan7425_2986 de *Cyanothece* sp. PCC7425 produjeron los mismos aldehídos grasos y alcoholes grasos que en el ejemplo 3, pero también tridecano, pentadeceno, pentadecano y heptadeceno. Este resultado indica que Cyan7425_2986 en *E. coli* convierte tetradecanal, hexadecenal, hexadecanal y octadecenal en tridecano, pentadeceno, pentadecano y heptadeceno, respectivamente, y por tanto es una aldehído graso descarboxilasa activa.

Ejemplo 21. Producción de alcanos y alquenos en *E. coli* a través de la expresión heteróloga de PMM0533 de *Prochlorococcus marinus* CCMP1986 y PMM0532 de *Prochlorococcus marinus* CCMP1986

Se amplificó el ADN genómico que codifica para PMM0533 de *P. marinus* CCMP1986 (NP_892651; supuesta enzima generadora de aldehídos) (SEQ ID NO: 71) y PMM0532 de *Prochlorococcus marinus* CCMP1986

(NP_892650; supuesta descarbonilasa) (SEQ ID NO: 19) y se clonó en los sitios *NcoI* y *EcoRI* del vector OP-80 y en los sitios *NdeI* y *XhoI* del vector OP-183, respectivamente. Se transformaron y se contraformaron por separado los constructos resultantes en *E. coli* MG1655 y se hicieron crecer las células a 37°C en medio mínimo M9 complementado con espectinomicina 100 µg/ml y carbenicilina 100 µg/ml. Se cultivaron las células y se extrajeron como en el ejemplo 3 y se analizaron mediante CG-EM tal como se describe en el ejemplo 26.

Tal como se muestra en la figura 24A, las células de *E. coli* transformadas sólo con el vector que portaba PMM0533 de *P. marinus* CCMP1986 no produjeron ningún aldehído graso ni alcohol graso. Sin embargo, las células de *E. coli* cotransformadas con los vectores que portaban PMM0533 y PMM0532 produjeron hexadecanol, pentadecano y heptadeceno (figura 24B). Este resultado indica que PMM0533 sólo proporciona sustratos de aldehídos grasos para la reacción de descarbonilación cuando interacciona con una descarbonilasa, tal como PMM0532.

Ejemplo 22. Producción de alcanos y alquenos en una cepa de *E. coli* productora de acil graso-CoA a través de la expresión heteróloga del orf1594 de *Synechococcus elongatus* PCC7942 y AM1_4041 de *Acaryochloris marina* MBIC11017

Se sintetizó el ADN genómico que codifica para AM1_4041 de *Acaryochloris marina* MBIC11017 (YP_001518340; supuesta aldehído graso descarbonilasa) (SEQ ID NO: 9) y se clonó en los sitios *NdeI* y *XhoI* del vector OP-183 (derivado de pACYC) bajo el control del promotor P_{trc} . Se contraformó el constructo resultante con OP80-PCC7942_1594 en *E. coli* MG1655 $\Delta fadE lacZ::P_{trc}'tesA-fadD$. Esta cepa expresa una versión citoplasmática de la tioesterasa de *E. coli*, 'TesA, y la acil-CoA sintetasa de *E. coli*, FadD, bajo el control del promotor P_{trc} , y por tanto produce acil graso-CoA. Se hicieron crecer las células a 37°C en medio mínimo M9 complementado con espectinomicina 100 µg/ml y carbenicilina 100 µg/ml. Se cultivaron las células y se extrajeron como en el ejemplo 3 y se analizaron mediante CG-EM tal como se describe en el ejemplo 1.

Tal como se muestra en la figura 25, estas células de *E. coli* cotransformadas con PCC7942_1594 de *S. elongatus* y AM1_4041 de *A. marina* MBIC11017 también produjeron alcanos y alcoholes grasos. Este resultado indica que PCC7942_1594 de *S. elongatus* puede usar acil-CoA como sustrato para producir hexadecenal, hexadecanol y octadecenal, que entonces se convierte en pentadeceno, pentadecano y heptadeceno, respectivamente, mediante AM1_4041 de *A. marina* MBIC11017.

Ejemplo 23. Producción de alcanos y alquenos en una cepa de *E. coli* productora de acil graso-CoA a través de la expresión heteróloga de sll0209 de *Synechocystis sp.* PCC6803 y sll0208 de *Synechocystis sp.* PCC6803

Se sintetizó el ADN genómico que codifica para sll0208 de *Synechocystis sp.* PCC6803 (NP_442147; supuesta aldehído graso descarbonilasa) (SEQ ID NO: 3) y se clonó en los sitios *NdeI* y *XhoI* del vector OP-183 (derivado de pACYC) bajo el control del promotor P_{trc} . Se sintetizó el ADN genómico que codifica para sll0209 de *Synechocystis sp.* PCC6803 (NP_442146; acil-ACP reductasa) (SEQ ID NO: 67) y se clonó en los sitios *NcoI* y *EcoRI* del vector OP-183 (derivado de pACYC) bajo el control del promotor P_{trc} . Se cotransformaron los constructos resultantes en *E. coli* MG1655 $\Delta fadE lacZ::P_{trc}'tesA-fadD$. Esta cepa expresa una versión citoplasmática de la tioesterasa de *E. coli*, 'TesA, y la acil-CoA sintetasa de *E. coli*, FadD, bajo el control del promotor P_{trc} , y por tanto produce acil graso-CoA. Se hicieron crecer las células a 37°C en medio mínimo M9 complementado con espectinomicina 100 µg/ml y carbenicilina 100 µg/ml. Se cultivaron las células y se extrajeron como en el ejemplo 3 y se analizaron mediante CG-EM tal como se describe en el ejemplo 26.

Tal como se muestra en la figura 26, estas células de *E. coli* transformadas con sll0209 de *Synechocystis sp.* PCC6803 no produjeron ningún aldehído graso ni alcohol graso. Sin embargo, cuando se cotransformaron con sll0208 y sll0209 de *Synechocystis sp.* PCC6803, produjeron alcanos, aldehídos grasos y alcoholes grasos. Este resultado indica que sll0209 de *Synechocystis sp.* PCC6803 puede usar acil-CoA como sustrato para producir aldehídos grasos tales como tetradecanal, hexadecanal y octadecenal, pero sólo cuando se coexpresa con una aldehído graso descarbonilasa. Los aldehídos grasos aparentemente se reducen adicionalmente para dar los correspondientes alcoholes grasos, tetradecanol, hexadecanol y octadecanol, mediante una actividad aldehído reductasa endógena de *E. coli*. En este experimento, se convirtió el octadecenal en heptadeceno mediante sll0208 de *Synechocystis sp.* PCC6803.

Ejemplo 24. Producción de alcanos y alquenos en una cepa de *E. coli* productora de aldehídos grasos a través de la expresión heteróloga de Npun02004178 de *Nostoc punctiforme* PCC73102 y varios de sus homólogos

Se amplificó el ADN genómico que codifica para Npun02004178 de *Nostoc punctiforme* PCC73102 (ZP_00108838; supuesta aldehído graso descarbonilasa) (SEQ ID NO: 5) y se clonó en los sitios *NdeI* y *XhoI* del vector OP-183 (derivado de pACYC) bajo el control del promotor P_{trc} . Se amplificó el ADN genómico que codifica para el orf MSMEG_5739 de *Mycobacterium smegmatis* cepa MC2 155 (YP_889972, supuesta ácido carboxílico reductasa) (SEQ ID NO: 85) y se clonó en los sitios *NcoI* y *EcoRI* del vector OP-180 (derivado de pCL1920) bajo el control del promotor P_{trc} . Se contraformaron los dos constructos resultantes en *E. coli* MG1655 $\Delta fadD lacZ::P_{trc}'tesA$. En esta cepa, se proporcionaron aldehídos grasos por MSMEG_5739, que reduce los ácidos grasos libres (formados mediante la acción de 'TesA) para dar aldehídos grasos. Se hicieron crecer las células a 37°C en medio mínimo M9 complementado con espectinomicina 100 µg/ml y carbenicilina 100 µg/ml. Se cultivaron las células y se extrajeron

como en el ejemplo 3 y se analizaron mediante CG-EM tal como se describe en el ejemplo 1.

Tal como se muestra en la figura 27, estas células de *E. coli* cotransformadas con los vectores que portaban Npun02004178 de *N. punctiforme* PCC73102 y MSMEG_5739 de *M. smegmatis* cepa MC2 155 produjeron tridecano, pentadeceno y pentadecano. Este resultado indica que Npun02004178 en *E. coli* convierte tetradecanal, hexadecenal y hexadecanal proporcionados por la ácido carboxílico reductasa MSMEG_5739 en tridecano, pentadeceno y pentadecano. Tal como se muestra en la figura 28, en la misma configuración experimental, las siguientes aldehído graso descarbonilasas también convirtieron aldehídos grasos proporcionados por MSMEG_5739 en los alcanos correspondientes cuando se expresaban en *E. coli* MG1655 Δ fadD lacZ::P_{trc}-tesA: alr5283 de *Nostoc* sp. PCC7210 (SEQ ID NO: 7), PMM0532 de *P. marinus* CCMP1986 (SEQ ID NO: 19), gll3146 de *G. violaceus* PCC7421 (SEQ ID NO: 15), RS9917_09941 de *Synechococcus* sp. (SEQ ID NO: 23), RS9917_12945 de *Synechococcus* sp. (SEQ ID NO: 25) y AM1_4041 de *A. marina* MBIC11017 (SEQ ID NO: 9).

Ejemplo 25: Las aldehído graso descarbonilasas de cianobacterias pertenecen a la clase de proteínas de di-hierro no hemo. La mutagénesis dirigida al sitio de histidinas conservadas para dar fenilalaninas en Npun02004178 de *Nostoc punctiforme* PCC73102 no suprime su función catalítica

Tal como se comenta en el ejemplo 13, la proteína hipotética PMT1231 de *Prochlorococcus marinus* MIT9313 (SEQ ID NO: 18) es una aldehído graso descarbonilasa activa. Basándose en la estructura tridimensional de PMT1231, que está disponible a una resolución de 1,8 Å (pdb2OC5A) (véase la figura 29B), las aldehído graso descarbonilasas de cianobacterias tienen similitud estructural con proteínas de di-hierro no hemo, en particular con las proteínas de subunidad β de ribonucleasa reductasa de clase I, RNR β (Stubbe y Riggs-Gelasco, TIBS 1998, vol. 23., págs. 438) (véase la figura 29A). La RNR β de clase Ia e Ib contiene un radical tirosilo diférrico que media en la actividad catalítica de RNR α (reducción de ribonucleótidos para dar desoxirribonucleótidos). En RNR β de *E. coli*, esta tirosina está en la posición 122 y en proximidad estrecha a una de las moléculas de hierro del sitio activo. La alineación estructural mostró que PMT1231 contenía una fenilalanina en la misma posición que la tyr122 de RNR β , lo que sugiere un mecanismo catalítico diferente para las aldehído graso descarbonilasas de cianobacterias. Sin embargo, una alineación de todas las descarbonilasas mostró que dos residuos de tirosina estaban completamente conservados en todas las secuencias, tyr135 y tyr138 con respecto a PMT1231, estando tyr135 en proximidad estrecha (5,5 Å) a una de las moléculas de hierro del sitio activo (véase la figura 29C). Para examinar si cualquiera de los dos residuos de tirosina conservados está implicado en el mecanismo catalítico de las aldehído graso descarbonilasas de cianobacterias, se reemplazaron estos residuos por fenilalanina en Npun02004178 (tyr123 y tyr126) tal como sigue.

Se clonó el ADN genómico que codifica para el ORF1594 de *S. elongatus* PCC7942 (SEQ ID NO: 65) en los sitios NcoI y EcoRI del vector OP-80 (derivado de pCL1920) bajo el control del promotor P_{trc}. También se clonó el ADN genómico que codifica para Npun02004178 de *N. punctiforme* PCC73102 (SEQ ID NO: 5) en los sitios NdeI y XhoI del vector OP-183 (derivado de pACYC177) bajo el control del promotor P_{trc}. Se usó este último constructo como molde para introducir una mutación en las posiciones 123 y 126 de la proteína descarbonilasa, cambiando las tirosinas por fenilalaninas usando los cebadores gtttgatcgagcatttaacatttacatcccgttgccgacg y gtttgatcgagcagcatataacatttcaccccgttgccgacg, respectivamente. Entonces se transformaron los constructos resultantes en *E. coli* MG1655. Se hicieron crecer las células a 37°C en medio mínimo M9 complementado con glucosa al 1% (p/v) y espectinomicina y carbenicilina 100 μ g/ml. Se cultivaron las células y se extrajeron como en el ejemplo 3.

Tal como se muestra en la figura 30, las dos variantes de proteína Npun02004178 Tyr a Phe eran activas y produjeron alcanos cuando se coexpresaron con el ORF1594 de *S. elongatus* PCC7942. Este resultado indica que en contraposición a las proteínas RNR β de clase Ia e Ib, el mecanismo catalítico de las aldehído graso descarbonilasas no implica un radical tirosilo.

Ejemplo 26: Caracterización bioquímica de Npun02004178 de *Nostoc punctiforme* PCC73102

Se clonó el ADN genómico que codifica para Npun02004178 de *N. punctiforme* PCC73102 (SEQ ID NO: 5) en los sitios NdeI y XhoI del vector pET-15b bajo el control del promotor de T7. La proteína Npun02004178 resultante contenía una cola de His N-terminal. Se transformó una cepa de *E. coli* BL21 (DE3) (Invitrogen) con el plásmido mediante técnicas de transformación química de rutina. Se llevó a cabo la expresión de proteínas inoculando en primer lugar una colonia de la cepa de *E. coli* en 5 ml de medios LB complementados con 100 mg/l de carbenicilina y se agitó durante la noche a 37°C para producir un cultivo inicial. Estos cultivos iniciales se usaron para inocular 0,5 l de medios LB complementados con 100 mg/l de carbenicilina. Se agitó el cultivo a 37°C hasta que se alcanzó un valor de DO₆₀₀ de 0,8, y entonces se añadió IPTG hasta una concentración final de 1 mM. Entonces se agitó el cultivo al 37°C durante aproximadamente 3 h adicionales. Entonces se centrifugó el cultivo a 3.700 rpm durante 20 min a 4°C. Entonces se resuspendió el sedimento en 10 ml de tampón que contenía tampón fosfato de sodio 100 mM a pH 7,2 complementado con Bacterial Protease Arrest (GBiosciences). Entonces se sonicaron las células a 12 W en hielo durante 9 s con 1,5 s de sonicación seguido por 1,5 s de descanso. Se repitió este procedimiento 5 veces con intervalos de un min entre cada ciclo de sonicación. Se centrifugó el extracto libre de células a 10.000 rpm durante 30 min a 4°C. Se añadieron 5 ml de Ni-NTA (Qiagen) al sobrenadante y se agitó suavemente la mezcla a 4°C. Se hizo pasar la suspensión sobre una columna eliminando la resina del lisado. Entonces se lavó la resina con

30 ml de tampón que contenía tampón fosfato de sodio 100 mM a pH 7,2 más imidazol 30 mM. Finalmente, se eluyó la proteína con 10 ml de tampón fosfato de sodio 100 mM a pH 7,2 más imidazol 250 mM. Se dializó la disolución de proteína con 200 volúmenes de tampón fosfato de sodio 100 mM a pH 7,2 con glicerol al 20%. Se determinó la concentración de proteína usando el ensayo de Bradford (Biorad). Se obtuvieron 5,6 mg/ml de la proteína Npun02004178.

Para sintetizar octadecanal para la reacción de la descarbonilasa, se disolvieron 500 mg de octadecanol (Sigma) en 25 ml de diclorometano. A continuación, se añadieron 200 mg de clorocromato de piridinio (TCI America) a la disolución y se agitó durante la noche. Se secó la mezcla de reacción a vacío para eliminar el diclorometano. Se resuspendieron los productos restantes en hexano y se filtraron a través de papel de filtro Whatman. Entonces se secó el filtrado a vacío y se resuspendió en 5 ml de hexano y se purificó mediante cromatografía ultrarrápida en sílice. Se cargó la mezcla en una columna alimentada por gravedad en hexano y luego se lavó con dos volúmenes de columna de hexano. Entonces se eluyó el octadecanal con una mezcla 8:1 de hexano y acetato de etilo. Se reunieron las fracciones que contenían octadecanal y se analizaron usando los métodos de CG/EM descritos a continuación. El producto final era puro en el 95% tal como se determinó mediante este método.

Para someter a prueba la proteína Npun02004178 para determinar la actividad de descarbonilación, se establecieron los siguientes ensayos enzimáticos. Se establecieron reacciones de 200 μ l en tampón fosfato de sodio 100 mM a pH 7,2 con los siguientes componentes en sus concentraciones finales respectivas: 30 μ M de proteína Npun02004178 purificada, octadecanal 200 μ M, ferredoxina de espinaca 0,11 μ g/ml (Sigma), ferredoxina reductasa de espinaca 0,05 unidades/ml (Sigma) y NADPH 1 mM (Sigma). Los controles negativos incluyeron la reacción anterior sin Npun02004178, la reacción anterior sin octadecanal y la reacción anterior sin ferredoxina de espinaca, ferredoxina reductasa y NADPH. Se incubó cada reacción a 37°C durante 2 h antes de extraerse con 100 μ l de acetato de etilo. Se analizaron las muestras mediante CG/EM usando los siguientes parámetros: tiempo de ejecución: 13,13 min; columna: HP-5-EM n.º de parte 19091S-433E (longitud de 30 metros; D.I.: 0,25 mm de calibre estrecho; película: 0,25 μ M); inyección: 1 μ l en la entrada de Agilent 6850; entrada: 300°C sin dividir; gas portador: helio; flujo: 1,3 ml/min; temperatura del horno: 75°C mantenidos 5 min, 320 a 40°C/min, 320 mantenidos 2 min; det.: Agilent 5975B VL MSD; temperatura de det.: 330°C; exploración: 50-550 M/Z. Se usó heptadecano de Sigma como referencia auténtica para determinar el patrón de fragmentación y el tiempo de retención del compuesto.

Tal como se muestra en la figura 31, se observó conversión *in vitro* de octadecanal en heptadecano en presencia de Npun02004178. La descarbonilación enzimática de octadecanal mediante Npun02004178 fue dependiente de la adición de ferredoxina reductasa de espinaca, ferredoxina y NADPH.

A continuación, se determinó si las ferredoxinas y las ferredoxina reductasas de cianobacterias pueden reemplazar a las proteínas de espinaca en el ensayo de alhído graso descarbonilasa *in vitro*. Se clonaron los cuatro genes siguientes por separado en los sitios *Nde*I y *Xho*I de pET-15b: Npun02003626 de *N. punctiforme* PCC73102 (ZP_00109192, petH de ferredoxina oxidoreductasa sin el dominio de ligador de alofococianina N-terminal) (SEQ ID NO: 87), Npun02001001 de *N. punctiforme* PCC73102 (ZP_00111633, ferredoxina 1) (SEQ ID NO: 89), Npun02003530 de *N. punctiforme* PCC73102 (ZP_00109422, ferredoxina 2) (SEQ ID NO: 91) y Npun02003123 de *N. punctiforme* PCC73102 (ZP_00109501, ferredoxina 3) (SEQ ID NO: 93). Se expresaron las cuatro proteínas y se purificaron tal como se describió anteriormente. Se obtuvieron 1 mg/ml de cada ferredoxina y 4 mg/ml de la ferredoxina oxidoreductasa. Se sometieron a prueba las tres ferredoxinas de cianobacterias con la ferredoxina oxidoreductasa de cianobacterias usando la configuración enzimática descrita anteriormente con los siguientes cambios. La concentración final de la ferredoxina reductasa era de 60 μ g/ml y las ferredoxinas estaban a 50 μ g/ml. Las reacciones enzimáticas se extrajeron con CG/EM usando los siguientes parámetros: tiempo de ejecución: 6,33 min; columna: J&W 122-5711 DB-5ht (longitud de 15 metros; D.I.: 0,25 mm de calibre estrecho, película: 0,10 μ M); inyección: 1 μ l en la entrada de Agilent 6850; entrada: 300°C sin dividir; gas portador: helio; flujo: 1,3 ml/min; temperatura del horno: 100°C mantenidos 0,5 min, 260 a 30°C/min, 260 mantenidos 0,5 min; det: Agilent 5975B VL MSD; temperatura de det.: 230°C; exploración: 50-550 M/Z.

Tal como se muestra en la figura 32, se observó la conversión *in vitro* dependiente de Npun02004178 de octadecanal en heptadecano en presencia de NADPH y la ferredoxina oxidoreductasa de cianobacterias y cualquiera de las tres ferredoxinas de cianobacterias.

Ejemplo 27. Caracterización bioquímica de orf1594 de *Synechococcus elongatus* PCC7942

Se clonó el ADN genómico que codifica para el orf1594 de *S. elongatus* PCC7492 (SEQ ID NO: 65) en los sitios *Nco*I y *Xho*I del vector pET-28b bajo el control del promotor de T7. La proteína PCC7942_orf1594 resultante contenía una cola de His C-terminal. Se transformó una cepa de *E. coli* BL21 (DE3) (Invitrogen) con el plásmido y se expresó la proteína PCC7942_orf1594 y se purificó tal como se describe en el ejemplo 22. Se almacenó la disolución de proteína en el siguiente tampón: fosfato de sodio 50 mM, pH 7,5, NaCl 100 mM, THP 1 mM, glicerol al 10%. Se determinó la concentración de proteína usando el ensayo de Bradford (Biorad). Se obtuvieron 2 mg/ml de la proteína PCC7942_orf1594.

Para someter a prueba la proteína PCC7942_orf1594 para determinar la actividad acil-ACP o acil-CoA reductasa, se establecieron los siguientes ensayos enzimáticos. Se establecieron reacciones de 100 μ l en tampón Tris-HCl 50 mM

a pH 7,5 con los siguientes componentes en sus concentraciones finales respectivas: 10 μM de proteína PCC7942_orf1594 purificada, acil-CoA o acil-ACP 0,01-1 mM, MgCl_2 2 mM, NADPH 0,2-2 mM. Se incubaron las reacciones durante 1 h a 37°C y se detuvieron añadiendo 100 μl de acetato de etilo (que contenía 1-octadeceno 5 mg/l como patrón interno). Se agitaron con vórtex las muestras durante 15 min y se centrifugaron a la velocidad máxima durante 3 min para la separación de fases. Se transfirieron 80 μl de la capa superior a viales de vidrio para CG y se analizaron mediante CG/EM tal como se describe en el ejemplo 26. Se calculó la cantidad de aldehído formada basándose en el patrón interno.

Tal como se muestra en la figura 33, PCC7942_orf1594 pudo reducir octadecanoil-CoA para dar octadecanal. La actividad reductasa requirió cationes divalentes tales como Mg^{2+} , Mn^{2+} o Fe^{2+} y NADPH como donador de electrones. El NADH no soportó la actividad reductasa. PCC7942_orf1594 también pudo reducir octadecenoil-CoA y octadecenoil-ACP para dar octadecenal. Se determinaron los valores de K_m para la reducción de octadecanoil-CoA, octadecenoil-CoA y octadecenoil-ACP en presencia de NADPH 2 mM como $45 \pm 20 \mu\text{M}$, $82 \pm 22 \mu\text{M}$ y $7,8 \pm 2 \mu\text{M}$, respectivamente. Estos resultados demuestran que PCC7942_orf1594, *in vitro*, reduce tanto acil-CoA como acil-ACP y que la enzima tiene aparentemente una afinidad superior por acil-ACP en comparación con acil-CoA. Se determinó el valor de K_m para NADPH en presencia de octadecanoil-CoA 0,5 mM para PCC7942_orf1594 como $400 \pm 80 \mu\text{M}$.

A continuación, se examinó la transferencia de hidruro estereoespecífica desde el NADPH hasta un aldehído graso catalizada por PCC7942_orf1594. Se preparó deuterio-NADPH según el siguiente protocolo. Se añadieron 5 mg de NADP^+ y 3,6 mg de D-glucosa-1-d a 2,5 ml de tampón fosfato de sodio 50 mM (pH 7,0). Se inició la producción enzimática de NADPH marcado mediante la adición de 5 unidades de glucosa deshidrogenasa de o bien *Bacillus megaterium* (USB Corporation) para la producción de R-(4- ^2H)NADPH o bien *Thermoplasma acidophilum* (Sigma) para la producción de S-(4- ^2H)NADPH. Se incubó la reacción durante 15 min a 37°C, se filtró en centrífuga usando un filtro de centrífuga Amicon Ultra de 10 kDa de punto de corte de peso molecular (Millipore), se sometió a congelación ultrarrápida en hielo seco y se almacenó a -80°C.

La reacción de ensayo *in vitro* contenía Tris-HCl 50 mM (pH 7,5), 10 μM de proteína PCC7942_orf1594 purificada, octadecanoil-CoA 1 mM, MgCl_2 2 mM y 50 μl de deuterio-NADPH (preparado tal como se describió anteriormente) en un volumen total de 100 μl . Tras una incubación de 1 h, se extrajo el producto de la reacción enzimática y se analizó tal como se describió anteriormente. El aldehído graso resultante detectado mediante CG/EM era octadecanal (véase la figura 34). Puesto que la transferencia de hidruro desde el NADPH es estereoespecífica, se sintetizaron tanto R-(4- ^2H)NADPH como S-(4- ^2H)NADPH. Se observó octadecanal con una masa unitaria de más uno usando sólo el S-(4- ^2H)NADPH. El hecho de que el aldehído graso estuviera marcado indica que el hidrógeno deuterado se ha transferido desde el NADPH marcado hasta el aldehído graso marcado. Esto demuestra que se usa NADPH en esta reacción enzimática y que la transferencia de hidruro catalizada por PCC7942_orf1594 es estereoespecífica.

Ejemplo 28. Producción intracelular y extracelular de aldehídos grasos y alcoholes grasos en *E. coli* a través de la expresión heteróloga del orf1594 de *Synechococcus elongatus* PCC7942

Se amplificó el ADN genómico que codifica para orf1594 de *Synechococcus elongatus* PCC7942 (YP_400611; acil-ACP reductasa) (SEQ ID NO: 65) y se clonó en los sitios *NcoI* y *EcoRI* del vector OP-80 (derivado de pCL1920) bajo el control del promotor P_{trc} . Se contrató el constructo resultante en *E. coli* MG1655 ΔfadE y se hicieron crecer las células a 37°C en 15 ml de medios mínimos Che-9 con glucosa al 3% (p/v) como fuente de carbono y complementados con carbenicilina y espectinomicina 100 $\mu\text{g/ml}$, respectivamente. Cuando el cultivo alcanzó una DO_{600} de 0,8-1,0, se indujo con IPTG 1 mM y se hicieron crecer las células durante 24-48 h adicionales a 37°C. El medio mínimo Che-9 se define como: Na_2HPO_4 6 g/l, KH_2PO_4 3 g/l, NaCl 0,5 g/l, NH_4Cl 2 g/l, MgSO_4 0,25 g/l x 7 H_2O , CaCl_2 11 mg/l, Fe_3Cl 27 mg/l x 6 H_2O , ZnCl 2 mg/l x 4 H_2O , Na_2MoO_4 2 mg/l x 2 H_2O , CuSO_4 1,9 mg/l x 5 H_2O , H_3BO_3 0,5 mg/l, tiamina 1 mg/l, Bis-Tris 200 mM (pH 7,25) y Triton-X100 al 0,1% (v/v). Cuando el cultivo alcanzó una DO_{600} de 1,0-1,2, se indujo con IPTG 1 mM y se permitió que las células crecieran durante 40 h adicionales a 37°C. Se extrajeron las células de 0,5 ml de cultivo con 0,5 ml de acetato de etilo para determinar la producción de hidrocarburos total tal como se describe en el ejemplo 26. Adicionalmente, se separaron las células y el sobrenadante mediante centrifugación (4.000 g a TA durante 10 min) y se extrajeron por separado.

El cultivo produjo aldehídos grasos 620 mg/l (tetradecanal, heptadecenal, heptadecanal y octadecenal) y alcoholes grasos 1670 mg/l (dodecanol, tetradecanol, tetradecanol, heptadecanol, heptadecanol y octadecanol). La figura 35 muestra el cromatograma del sobrenadante extraído. Se determinó que el 73% de los aldehídos grasos y los alcoholes grasos estaban en el sobrenadante libre de células.

Ejemplo 29. Producción intracelular y extracelular de alcanos y alquenos en *E. coli* a través de la expresión heteróloga del orf1594 de *Synechococcus elongatus* PCC7942 y Npun02004178 de *Nostoc punctiforme* PCC73102

Se amplificó el ADN genómico que codifica para orf1594 de *Synechococcus elongatus* PCC7942 (YP_400611; acil-ACP reductasa) (SEQ ID NO: 65) y se clonó en los sitios *NcoI* y *EcoRI* del vector OP-80 (derivado de pCL1920) bajo el control del promotor P_{trc} . Se amplificó el ADN genómico que codifica para Npun02004178 de *Nostoc punctiforme* PCC73102 (ZP_00108838; aldehído graso descarboxilasa) (SEQ ID NO: 5) y se clonó en los sitios *NdeI* y *XhoI* del vector OP-183 (derivado de pACYC) bajo el control del promotor P_{trc} . Se cotransformaron los constructos resultantes en *E. coli* MG1655 ΔfadE y se hicieron crecer las células a 37°C en 15 ml de medios mínimos Che9 con glucosa al

3% (p/v) como fuente de carbono y complementados con carbenicilina y espectinomicina 100 µg/ml, respectivamente. Se hicieron crecer las células, se separaron del caldo, se extrajeron y se analizaron tal como se describe en el ejemplo 28.

5 El cultivo produjo alcanos y alquenos 323 mg/l (tridecano, pentadeceno, pentadecano y heptadeceno), aldehídos grasos 367 mg/l (tetradecanal, heptadecanal, heptadecanal y octadecanal) y alcoholes grasos 819 mg/l (tetradecanol, heptadecanol, heptadecanol y octadecanol). La figura 36 muestra el cromatograma del sobrenadante extraído. Se determinó que el 86% de los alcanos, alquenos, aldehídos grasos y alcoholes grasos estaban en el sobrenadante libre de células.

10 Ejemplo 30. Producción de alcanos y alquenos en *E. coli* a través de la expresión heteróloga de alr5284 de *Nostoc* sp. PCC7210 y alr5283 de *Nostoc* sp. PCC7210

15 Se amplificó el ADN genómico que codifica para alr5284 de *Nostoc* sp. PCC7210 (NP_489324; supuesta enzima generadora de aldehídos) (SEQ ID NO: 81) y se clonó en los sitios *NcoI* y *EcoRI* del vector OP-80 (derivado de pCL1920) bajo el control del promotor P_{trc} . Se amplificó el ADN genómico que codifica para alr5283 de *Nostoc* sp. PCC7210 (NP_489323; supuesta descarboxilasa) (SEQ ID NO: 7) y se clonó en los sitios *NdeI* y *XhoI* del vector OP-183 (derivado de pACYC) bajo el control del promotor P_{trc} . Se cotransformaron los constructos resultantes en *E. coli* MG1655 y se hicieron crecer las células a 37°C en 15 ml de medios mínimos Che9 con glucosa al 3% (p/v) como fuente de carbono y complementados con carbenicilina y espectinomicina 100 µg/ml, respectivamente (tal como se describe en el ejemplo 28). Se extrajeron las células de 0,5 ml de cultivo y se analizaron tal como se describe en el ejemplo 3 y se analizaron mediante CG-EM tal como se describe en el ejemplo 26.

20 Tal como se muestra en la figura 37, las células de *E. coli* cotransformadas con los vectores que portaban alr5284 de *Nostoc* sp. PCC7210 y alr5283 de *Nostoc* sp. PCC7210 produjeron tridecano, pentadeceno, pentadecano, tetradecanol y hexadecanol. Este resultado indica que la coexpresión de alr5284 y alr5283 de *Nostoc* sp. PCC7210 es suficiente para que *E. coli* produzca alcoholes grasos, alcanos y alquenos.

25 Ha de entenderse que aunque se ha descrito la invención junto con la descripción detallada de la misma, la descripción anterior pretende ilustrar y no limitar el alcance de la invención, que está definido por el alcance de las reivindicaciones adjuntas.

Lista de secuencias

Nucleótidos de Synpcc7942_1593 de *Synechococcus elongatus* PCC7942 (YP_400610) (SEQ ID NO: 1)

```

1 atgccgcagc ttgaagccag ccttgaactg gactttcaaa gcgagtccta caaagacgct
61 tacagccgca tcaacgcgat cgtgattgaa ggcgaacaag aggcgttcga caactacaat
121 cgccttgctg agatgctgcc cgaccagcgg gatgagcttc acaagctagc caagatggaa
181 cagcgccaca tgaaaggcct tatggcctgt ggcaaaaatc tctccgtcac tcttgacatg
241 ggttttgccc agaaattttt cgagcgcttg cacgagaact tcaaagcggc ggctgcggaa
301 ggcaaggctg tcacctgcct actgattcaa tcgctaataca tcgagtgctt tgcgatcgcg
361 gcttacaaca tctacatccc agtggcggat gcttttgccc gcaaaatcac ggaggggggc
421 gtgcgcgacg aatacctgca ccgcaacttc ggtgaagagt ggctgaaggc gaattttgat
481 gcttccaaag ccgaactgga agaagccaat cgtcagaacc tgcccttggg ttgggctaatag
541 ctcaacgaag tggccgatga tgctcgcgaa ctcgggatgg agcgtgagtc gctcgtcgag
601 gactttatga ttgcctacgg tgaagctctg gaaaacatcg gcttcacaac gcgcgaaatc
661 atgcgtatgt ccgcctatgg ccttgcggcc gtttga
    
```

30 Aminoácidos de Synpcc7942_1593 de *Synechococcus elongatus* PCC7942 (YP_400610) (SEQ ID NO: 2)

```

1 mpqleaslel dfqsesykda ysrinaivie geqeafdny nrlaemlpdqr delhklakme
61 qrhmkgfmac gknlsvtpdm gfaqkfferl henfkaaaae gkvvtclliq sliiecfaia
121 ayniyipvad afarkitegv vrdeylhrnf geewlkanfd askaeleean rqnlpplvwlm
181 lnevaddare lgmereslve dfmiaygeal enigfttrei mrmsayglaa v
    
```

Nucleótidos de sll0208 de *Synechocystis* sp. PCC6803 (NP_442147) (SEQ ID NO: 3)

ES 2 667 469 T3

1 atgcccgagc ttgctgtccg caccgaatth gactattcca gcgaaattta caaagacgcc
 61 tatagccgca tcaacgccat tgtcattgaa ggccaacagg aagcctacag caactacctc
 121 cagatggcgg aactcctgcc ggaagacaaa gaagagttga cccgcttggc caaaatggaa
 181 aaccgccata aaaaaggtht ccaagcctgt ggcaacaacc tccaagtga cctgatgatg
 241 ccctatgccc aggaatttht cgcgggtctc catggcaatt tccagcacgc ttttagcgaa
 301 gggaaagtth ttacctgtht attgatccag gctthtgatta tcgaagctth tgcgatcgcc
 361 gcctataaca tatatatccc tgtggcggac gactthtgctc ggaaaatcac tgagggcgta
 421 gtcaaggacg aatacaccca cctcaactac ggggaagaat ggctaaaggc caactthgcc
 481 accgctaagg aagaactgga gcaggccaac aaagaaaacc tacccttagt gtggaaaatg
 541 ctcaaccaag tgcaggggga cgccaaggta ttgggcatgg aaaagaagc ctagtgga
 601 gattthtatga tcagctacgg cgaagcctc agtaacatcg gcttcagcac cagggaaatt
 661 atgcgtatgt cttcctacgg tttggccgga gtctag

Aminoácidos de slI0208 de *Synechocystis* sp. PCC6803 (NP_442147) (SEQ ID NO: 4)

1 mpelavrtef dysseyikda ysrinaivie geqeaysnyl qmaellpedk eeltrlakme
 61 nrhkkqfqc gnnlqvnpdm pyaqeffagl hgnfqhafse gkvvtclliq aliieafaia
 121 ayniyipvad dfarkitegv vkdeythlny geewlkanfa takeelegan kenlplvwkm
 181 lnqvqgdakv lgmekealve dfmisygeal snigfstrei mrmssyglag v

Nucleótidos de Npun02004178 de *Nostoc punctiforme* PCC73102 (ZP_00108838) (SEQ ID NO: 5)

1 atgcagcagc ttacagacca atctaaagaa ttagatttca agagcgaaac atacaagat
 61 gcttatagcc ggattaatgc gatcgtgatt gaaggggaac aagaagccca tgaaaattac
 121 atcacactag cccaactgct gccagaatct catgatgaat tgattcgcct atccaagatg
 5 181 gaaagccgcc ataagaaagg atttgaagct tgtgggcgca atttagctgt taccagat
 241 ttgcaatttg ccaagaggt tttctccggc ctacaccaa atthtcaaac agctgccgca
 301 gaagggaaag tggttacttg tctgttgatt cagtctthta ttattgaatg tthtgcatc
 361 gcagcatata acatttacat ccccgthtgc gacgatttcg cccgtaaaat tactgaagga
 421 gtagttaaag aagaatacag ccacctcaat tttggagaag tttggthtga agaacactth
 481 gcagaatcca aagctgaact tgaacttgca aatcgccaga acctaccat cgtctggaaa
 541 atgctcaacc aagtagaagg tgatgccac acaatggcaa tggaaaaaga tgctthtgta
 601 gaagacttca tgattcagta tggthgaagca ttgagtaaca ttgththttc gactcgcgat
 661 attatgcgct tctcagccta cggactcata ggtgctta

Aminoácidos de Npun02004178 de *Nostoc punctiforme* PCC73102 (ZP_00108838) (SEQ ID NO: 6)

1 mqqldtqske ldfksetykd aysrinaivi egeqeaheny itlaqllpes hdelirlskm
 61 esrhkkqfea cgrnlavtpd lqfakeffsg lhqnfqtaaa egkvvtclli qsliecfai
 121 aayniyipva ddfarkitegv vkeeyshln fgevwlkehf aeskaeela nrqnlpiwkw
 181 mlnqvvgdah tmamekdalv edfmiqygea lsnigfstred imrlsaygli ga

Nucleótidos de alr5283 de *Nostoc* sp. PCC7120 (NP_489323) (SEQ ID NO: 7)

1 atgcagcagg ttgcagccga tttagaaatt gatttcaaga gcgaaaaata taaagatgcc
 61 tatagtcgca taaatgcgat cgtgattgaa ggggaacaag aagcatacga gaattacatt
 121 caactatccc aactgctgcc agacgataaa gaagacctaa ttcgcctctc gaaaatggaa
 181 agccgtcaca aaaaaggatt tgaagcttgt ggacggaacc tacaagtatc accagatgatg
 241 gagthtgcca aagaattctt tgctggacta cacggtaact tccaaaaagc ggcggctgaa
 301 ggtaaaatcg ttacctgtct attgattcag tccctgatta ttgaatgtht tgcgatcgcc
 361 gcatacaata tctacattcc cgttgctgac gattthtgctc gtaaaatcac tgaggggtgta
 421 gtcaaagatg aatacagcca cctcaacttc ggcgaagtht ggttacagaa aaatthtgcc
 481 caatccaaag cagaattaga agaagctaat cgtcataatc tcccatagt ttggaaaatg
 541 ctcaatcaag tgcgggatga tgccgcagtc ttagctatgg aaaaagaagc ctagtcgaa
 601 gattthtatga ttcagctacgg cgaagcgtta agtaatattg gcttcacaac cagagatatt
 10 661 atgcggatgt cagcctacgg acttacagca gctta

Aminoácidos de alr5283 de *Nostoc* sp. PCC7120 (NP_489323) (SEQ ID NO: 8)

ES 2 667 469 T3

1 mqqvaadlei dfksekykda ysrinaivie geqeayenyi qlsqllpddk edlirlskme
 61 srhkkgfeac grnlqvspdm efakeffagl hgnfqkaaae gkivtclliq sliiecfaiia
 121 ayniyipvad dfarkitegv vkdeyshlnf gevwlqknfa qskaeleean rhnlpivwkm
 181 lnqvaddaav lamekealve dfmiqygeal snigfttrdi mrmsayglta a

Nucleótidos de AM1_4041 de *Acaryochloris marina* MBIC11017 (YP_001518340) (SEQ ID NO: 9)

1 atgccccaaa ctcaggctat ttcagaaatt gacttctata gtgacaccta caaagatgct
 61 tacagtcgta ttgacggcat tgtgatcgaa ggtgagcaag aagcgcgatga aaactatatt
 121 cgtcttggcg aaatgctgcc tgagcaccaa gacgacttta tccgcctgtc caagatggaa
 181 gcccgtcata agaaaggggt tgaagcctgc ggtcgcaact taaaagtaac ctgcgatcta
 241 gactttgccc ggcgtttctt ttccgactta cacaagaatt ttcaagatgc tgcagctgag
 301 gataaagtgc caacttgctt agtgattcag tccttgatca ttgagtgttt tgcgatcgca
 361 gcttacaaca tctatatccc cgtcgctgat gactttgccc gtaagattac agagtctgtg
 421 gttaaggatg agtatcaaca cctcaattat ggtgaagagt ggcttaaagc tcaactcgat
 481 gatgtgaaag cagaaatcca agaagctaatt cgcaaaaacc tccccatcgt ttggagaatg
 541 ctgaacgaag tggacaagga tgcggccggt ttaggaatgg aaaagaagc cctggttgaa
 601 gacttcatga tccagtatgg tgaagccctt agcaatattg gtttctctac aggcgaaatt
 661 atgcggatgt ctgcctatgg tcttgtggct gcgtaa

Aminoácidos de AM1_4041 de *Acaryochloris marina* MBIC11017 (YP_001518340) (SEQ ID NO: 10)

5 1 mpqtqaisei dfysdykda ysridgivie geqeahenyi rlgemlpehq ddfirlskme
 61 arhkkgfeac grnlkvtcdl dfarrffsdl hknfqdaaae dkvptclviq sliiecfaiia
 121 ayniyipvad dfarkitesv vkdeyqhlly geewlkahfd dvkaeiqean rknlpivwrm
 181 lnevdkdaav lgmekealve dfmiqygeal snigfstgei mrmsayglva a

Nucleótidos de tll1313 de *Thermosynechococcus elongatus* BP-1 (NP_682103) (SEQ ID NO: 11)

1 atgacaacgg ctaccgctac acctgttttg gactaccata gcgatcgcta caaggatgcc
 61 tacagccgca ttaacgccat tgtcattgaa ggtgaacagg aagctcacga taactatatac
 121 gatttagcca agctgctgcc acaacaccaa gaggaactca cccgccttgc caagatggaa
 181 gctcgccaca aaaagggggt tgaggcctgt ggtcgcaacc tgagcgtaac gccagatatg
 241 gaatttgcca aagccttctt tgaaaaactg cgcgctaact ttcaagggc tctggcggag
 301 ggaaaaactg cgacttgtct tctgattcaa gctttgatca tcgaatcctt tgcgatcgcg
 361 gcctacaaca tctacatccc aatggcggat cctttcgccc gtaaaattac tgagagtgtt
 421 gttaaggacg aatacagcca cctcaacttt ggcgaaatct ggctcaagga acactttgaa
 481 agcgtcaaag gagagctcga agaagccaat cgcgccaatt tacccttggc ctggaaaatg
 541 ctcaaccaag tggaaagcaga tgccaaagtg ctcggcatgg aaaagatgc ccttgtggaa
 601 gacttcatga ttcagtacag tgggtgcccta gaaaatatcg gctttaccac ccgcgaaatt
 661 atgaagatgt cagtttatgg cctcactggg gcataa

Aminoácidos de tll1313 de *Thermosynechococcus elongatus* BP-1 (NP_682103) (SEQ ID NO: 12)

10 1 mttatatpvl dyhsdrykda ysrinaivie geqeahdnyi dlakllpqhq eeltrlakme
 61 arhkkgfeac grnlsvtpdm efakaffekl ranfqrалаe gktatclliq aliiesfaia
 121 ayniyipmad pfarkitesv vkdeyshlnf geiwlkehfe svkgeleean ranlplvwkm
 181 lnqveadakv lgmekdalve dfmiqysgal enigfttrei mkmsvygltg a

Nucleótidos de CYA_0415 de *Synechococcus* sp. JA-3-3A (YP_473897) (SEQ ID NO: 13)

ES 2 667 469 T3

1 atggccccag cgaacgtcct gcccaacacc cccccgtccc ccaactgatgg gggcgggcact
 61 gccctagact acagcagccc aaggatatcg caggcctact cccgcatacaa cgggtattggt
 121 atcgaaggcg aacaagaagc ccacgacaac tacctcaagc tggccgaaat gctgccggaa
 181 gctgcagagg agctgcgcaa gctggccaag atggaattgc gccacatgaa aggcttccag
 241 gcctgcggca aaaacctgca ggtggaaccc gatgtggagt ttgcccgcgc ctttttcgcg
 301 cccttgccggg acaatttcca aagcgcgcga gcggcagggg atctgggtctc ctgttttgtc
 361 attcagtctt tgatcatcga gtgctttgcc attgccgcct acaacatcta catcccgggt
 421 gccgatgact ttgcccgcaa gatcaccgag ggggtagtta aggacgagta tctgcacctc
 481 aattttgggg agcgtggct gggcgagcac tttgccgagg ttaaagccca gatcgaagca
 541 gccaacgccc aaaatctgcc tctagttcgg cagatgctgc agcaggtaga ggcggatgtg
 601 gaagccattt acatggatcg cgaggccatt gtagaagact tcatgatcgc ctacggcgag
 661 gccctggcca gcatcggctt caacaccgcg gaggtaatgc gcctctcggc ccagggtctg
 721 cgggcccgcct ga

Aminoácidos de CYA_0415 de *Synechococcus* sp. JA-3-3A (YP_473897) (SEQ ID NO: 14)

1 mapanvlpnt ppsptdgggt aldysspryr qaysringiv iegeqeahdn ylklaemlpe
 61 aaeelrklak melrhmkqfq acgknlqvpe dvefaraffa plrdnfnqsa aagdvlvscfv
 121 iqsliiecfa iaayniyipv addfarkite gvvkdeylhl nfggerwlgeh faevkaqiea
 181 anaqnlplvr qmlqqveadv eaiymdreai vedfmiayge alasigfntr evmrlsaqgl
 241 raa

Nucleótidos de gll3146 de *Gloeobacter violaceus* PCC7421 (NP_926092) (SEQ ID NO: 15)

1 gtgaaccgaa ccgcaccgtc cagcgcgcgc cttgattacc gtcccgacac ctaccgcat
 61 gcgtactccc gcatcaatgc catcgtcctt gaaggcgagc gggaagccca cgccaactac
 121 cttaccctcg ctgagatgct gccggaccat gccgaggcgc tcaaaaaact ggccgcatg
 181 gaaaatcgcc acttcaaagg cttccagtc tgcgcccga acctcgaagt cacgccggac
 5 241 gaccggtttg caagggccta cttcgaacag ctcgacggca actttcagca ggcggcggca
 301 gaaggtgacc ttaccacctg catggtcatc caggcactga tcatcgagt cttcgcaatt
 361 gcggcctaca acgtctacat tccgggtggc gacgcgttt cccgcaaggt gaccgagggc
 421 gtcgtcaagg acgagtacac ccacctcaac tttgggcagc agtggctcaa agagcgcttc
 481 gtgaccgtgc gcgagggcat cgagcgcgcc aacgccaga atctgcccac cgtctggcgg
 541 atgctcaacg ccgtcgaagc ggacaccgaa gtgctgcaga tggataaaga agcgatcgtc
 601 gaagacttta tgatgccta cggatgaagc ttgggcgaca tcggtttttc gatgcgcgac
 661 gtgatgaaga tgtccgcccg cggccttgcc tctgcccccc gccagtga

Aminoácidos de gll3146 de *Gloeobacter violaceus* PCC7421 (NP_926092) (SEQ ID NO: 16)

1 mnrtapssaa ldysrdsdyrd aysrinaivl egereahany ltlaemlpdh aealkklaam
 61 enrhfkqfqs carnlevtpd dpfarayfeq ldgnfqgaaa egdlttcmvi galiiecfai
 121 aaynvypva dafarkvteg vvkdeythln fgqqlkerf vtvregiera naqnlpivwr
 181 mlnaveadte vlqmdkeai v edfmiaygea lgdigfsmrd vmkmsargla saprq

Nucleótidos de PM123 de *Prochlorococcus marinus* MIT9313 (NP_895059) (SEQ ID NO: 17)

ES 2 667 469 T3

1 atgtttggtc tgatcgggtca ttccaccagt tttgaagatg ccaggaagac ggcacttcag
 61 attggctatg accatctaga tggagacctg gatgtttggt gcagtgtctc tccgcagttt
 121 ttggaacaga tagaggtaga aagccttaca ggaagaaaa tagaaggcgc ttatatcgat
 181 tcatgttttg tgccagaaat gttgagtcgg ttcaaaactg cccgtcgcaa agttctcaat
 241 gccatggaaa tggcccagaa gagaggcatt cagataagtg ccctaggcgg atttacctca
 301 attatttttg agaacttcaa tcttcttaag catcaacatg tacgaaacac cacgcttgaa
 361 tgggagcgat tcacaaccgg caacactcat acggcttggg ttatttgccg tcagcttgag
 421 aacaatgccc cacttttggg cattgattta agcaaggcgc gtgtagcggg agttggtgca
 481 actggcgata taggtagtgc tgtttgccgt tggctttccg ctcgaactgg tgtcgcgtgaa
 541 cttctttttg tggctagaca gcaacaacct cttatagatc ttcaaactga attggctggc
 601 ggcagaatat tgagtcttga agaggccttg cctgaagctg atgttgtggt ttgggttggc
 661 agcatgccta gaaccctgga aatcgatatg gagagtttgc gcaagccttg cttgatgatt
 721 gatgggggct atccgaagaa cctagatgca aagtttgctg gtagtggggg gcatgttctc
 781 aaaggaggaa tcggttgagtt ctgcaatgac atcagctggg atgttggctg gattgcagag
 841 atggataagc ctgcgaggca aatgtttgcc tgttttgcag aagcaatgct gttggagttt
 901 gagaattgcc acacaaactt cagttggggg cgtaaccaga tcacccttga gaaaatggat
 961 tttatcggca tggcgtcttt gcgtcatggt ttttcaagcc tcaatctcaa tcaccaactt
 1021 caggctgctg cagcctga

Aminoácidos de PM1231 de *Prochlorococcus marinus* MIT9313 (NP_895059) (SEQ ID NO: 18)

1 mfglighsts fedarktalq igydhldgdl dvwcsappqf leqieveslt gkkiegayid
 61 scfvpemlsr fktarrkvl n amemaqkrgi qisalggfts iifenfnllk hqhvrnttle
 121 werfttgnth tawvicrql e nnapllgidl skarvavvga tgdigsavcr wlsartgvae
 181 lllvarqqqp lidlqtelag grilsleeal peadvvvwva smprtleidm eslrkpclmi
 241 dggypknlda kfagsgvhvl kggivefcnd iswdvgwia e mdkparqmfa cfaeamllef
 301 enchtnfswg rnqitlekmd figmaslrhg fsslnlhql qaaaa

5 Nucleótidos de PMM0532 de *Prochlorococcus marinus* subesp. *pastoris* cepa CCMP1986 (NP_892650) (SEQ ID NO: 19)

1 atgcaaacac tcgaatctaa taaaaaaact aatctagaaa attctattga tttaccgat
 61 tttactactg attcttacia agacgcttat agcaggataa atgcaatagt tattgaaggt
 121 gaacaagagg ctcatgataa ttacatttcc ttagcaacat taattcctaa cgaattagaa
 181 gagttaacta aattagcgaa aatggagctt aagcacaana gaggctttac tgcattgga
 241 agaaatctag gtgttcaagc tgacatgatt tttgctaaag aattcttttc caaattacat
 301 ggtaattttc aggttgcggt atctaattggc aagacaacta catgcctatt aatacaggca
 361 attttaattg aagcttttgc tatatccgcg tatcacgttt acataagagt tgctgatcct
 421 ttcgcaaaa aaattacca aggtgttgtt aaagatgaat atcttcattt aaattatgga
 481 caagaatggc taaaagaaaa tttagcgact tgtaaagatg agctaattgga agcaataag
 541 gttaaccttc cattaatcaa gaagatgta gatcaagtct cggaagatgc ttcagtacta
 601 gctatggata ggaagaatt aatggaagaa ttcattgatt cctatcagga cactctcctt
 661 gaaatagggt tagataatag agaaattgca agaattggca tggctgctat agtttaa

Aminoácidos de PMM0532 de *Prochlorococcus marinus* subesp. *pastoris* cepa CCMP1986 (NP_892650) (SEQ ID NO: 20)

10 1 mqtlesnkkt nlensidlpd ftttsykday srinaivieg eqeahdnyis latlipnele
 61 eltklakmel khkrftacg rnlgvqadmi fakeffsklh gnfqvalsng ktttclliqa
 121 ilieafaisa yhvyrvadp fakkitqgvv kdeylhlnyg qewlkenlat ckdelmeank
 181 vnlplikml dqvsedavsl amdreelme e fmiayqdtll eigldnreia rmamaaiv

Nucleótidos de PMN2A_1863 de *Prochlorococcus marinus* cepa NATL2A (YP_293054) (SEQ ID NO: 21)

ES 2 667 469 T3

```

1 atgcaagctt ttgcatccaa caatttaacc gtagaaaaag aagagctaag ttctaactct
61 cttccagatt tcacctcaga atcttacaaa gatgcttaca gcagaatcaa tgcagttgta
121 attgaagggg agcaagaagc ttatttctaatt tttcttgatc tcgctaaatt gattcctgaa
181 catgcagatg agccttgtag gctaggggag atggagaaaa agcatatgaa tgggttttgt
241 gcttgcggga gaaatcttgc tgtaaagcct gatatgcctt ttgcaaagac ctttttctca
301 aaactccata ataatttttt agaggcttct aaagtaggag atacgactac ctgtctccta
361 attcaatgca tcttgattga atcttttgca atatccgcat atcacgttta tatacgtggt
421 gctgatccat tcgccaaaag aatcacagag ggtggtgtcc aagatgaata cttgcatttg
481 aactatggtc aagaatggct taaggccaat ctagagacag ttaagaaaga tcttatgagg
541 gctaataagg aaaacttgcc tcttataaag tccatgctcg atgaagtttc aaacgacgcc
601 gaagtccttc atatggataa agaagagtta atggaggaat ttatgattgc ttatcaagat
661 tcccttcttg aaataggtct tgataataga gaaattgcaa gaatggctct tgcagcggtg
721 atataa

```

Aminoácidos de PMN2A_1863 de *Prochlorococcus marinus* cepa NATL2A (YP_293054) (SEQ ID NO: 22)

```

1 mqafasnnlt vekeelssns lpdftsesyk daysrinavv iegeqeaysn fldlaklipe
61 hadelvrlgk mekkhmgfc acgrnlavkp dmpfaktffs klhnnfleaf kvgdtttcll
121 iqciliesfa isayhvirv adpfakrite gvvqdeylhl nyggewlkan letvkkdlmr
181 ankenlplik smldevsnda evlhmdkeel meefmiayqd sllleigldnr eiarmalaav
241 i

```

Nucleótidos de RS9917_09941 de *Synechococcus* sp. RS9917 (ZP_01079772) (SEQ ID NO: 23)

```

1 atgccgaccc ttgagacgtc tgaggctgcc gttcttgaag actcgatggc ttcaggctcc
61 cggctgcctg atttcaccag cgaggcttac aaggacgcct acagccgcat caatgcgatc
121 gtgatcgagg gtgagcagga agcgcacgac aactacatcg cctcggcac gctgatcccc
181 gagcagaagg atgagctggc cegtctcgcc cgcgatggaga tgaagcacat gaagggggtc
241 acctcctgtg gccgcaatct cggcgtggag gcagacctc ctttgctaa ggaattcttc
301 gccccctgc acgggaactt ccaggcagct ctccaggagg gcaaggtggt gacctgctg
361 ttgattcagg cgctgctgat tgaagcgttc gccatttccg cctatcacat ctacatcccg
421 gtggcgggat ccttcgctcg caagatcact gaaggtgtgg tgaaggatga gtacaccac
481 ctcaattacg gccaggaatg gctgaaggcc aattttgagg ccagcaagga tgagctgatg
541 gaggccaaca aggccaatct gcctctgac cgctcgatgc tggagcaggt ggcagccgac
601 gcgcccgtgc tgcagatgga aaaggaagat ctgatcgaag atttctgat cgcttaccag
661 gaggccctct gcgagatcgg tttcagctcc cgtgacattg ctcgatggc cgccgctgcc
5 721 ctcgcggtct ga

```

Aminoácidos de RS9917_09941 de *Synechococcus* sp. RS9917 (ZP_01079772) (SEQ ID NO: 24)

```

1 mptletseva vledsmasgs rlpdftseay kdaysrinai viegeqeahd nyialgtlip
61 eqkdelarla rmemkhmkgf tscgrnlgve adlpfakeff aplhgnfqaa lqegkvvtcl
121 liqallieaf aisayhiyip vadpfarkit egvvkdeyth lnyggewlka nfeaskdelm
181 eankanlpli rsmleqvaad aavlqmeked liedfliayq ealceigfss rdiarmaaaa
241 lav

```

Nucleótidos de RS9917_12945 de *Synechococcus* sp. RS9917 (ZP_01080370) (SEQ ID NO: 25)

ES 2 667 469 T3

1 atgaccagc tgcactttgc cagtgcggcc taccgcgagg cctacagccg gatcaacggc
 61 gttgtgattg tgggcgaagg tctcgccaat cgccatttcc agatgttggc gcggcgcatt
 121 cccgctgatc gcgacgagct gcagcggctc ggacgcatgg agggagacca tgccagcgc
 181 tttgtgggct gtggtcgcaa cctcgggtgt gtggccgata tggccctggc ccggcgcctg
 241 tttcagcccc tccatgatct gttcaaacgc cacgaccacg acggcaatcg ggccgaatgc
 301 ctggtgatcc aggggttgat cgtggaatgt ttcgcccgtg cggcttaccg ccaactacctg
 361 ccggtggccg atgcctacgc ccggccgata accgcagcgg tgatgaacga tgaatcggaa
 421 cacctcgact acgctgagac ctggctgcag cgccatttcc atcaggtgaa ggcccgggtc
 481 agcgcggtgg tggtaggggc gttgccgctc accctggcga tgttgcaatc gcttgctgca
 541 gacatgcgac agatcggcat ggatccggtg gagaccctgg ccagcttcag tgaactgttt
 601 cgggaagcgt tggatcggg ggggtttgag gctgtggagg ccaggcgact gctgatgcga
 661 gcggccgccc ggatggtctg a

Aminoácidos de RS9917_12945 de *Synechococcus* sp. RS9917 (ZP_01080370) (SEQ ID NO: 26)

1 mtqldfasaa yreaysring vvivgeglan rhfqmlarri padrdelqrl grmegdhasa
 61 fvgcgrnlgv vadlplarrl fqplhldfkr hdhdgnraec lviqglivec favaaayrhyll
 121 pvadayarpi taavmdese hldyaetwlq rhfdqkav savvvealpl tlamlqslaa
 181 dmrqigmdpvetlasfself realesvgfe avearrllmr aaarmv

Nucleótidos de cce_0778 de *Cyanothece* sp. ATCC51142 (YP_001802195) (SEQ ID NO: 27)

1 atgcaagagc ttgctttacg ctcagagcct gattttaaca gcgaaaccta taaagatgct
 61 tacagtcgca tcaatgctat tgtcattgaa ggggaacaag aagcctatca aaattatctt
 121 gatatggcgc aacttctccc agaagacgag gctgagttaa ttcgtctctc caagatggaa
 181 aaccgtcaca aaaaaggctt tcaagcctgt ggcaagaatt tgaatgtgac cccagatatg
 241 gactacgctc aacaattttt tgctgaactt catggcaact tccaaaaggc aaaagccgaa
 301 ggcaaaattg tcaacttgctt attaattcaa tctttgatca tcgaagcctt tgcgatcgcc
 361 gcttataata tttatattcc tgtggcagat ccctttgctc gtaaaatcac cgaaggggta
 421 gttaaggatg aatataccca cctcaatttt ggggaagtct ggttaaaaga gcattttgaa
 481 gcctctaaag cagaattaga agacgcaaat aaagaaaatt tacccttgt ttggcaaatg
 541 ctcaaccaag ttgaaaaaga tgccgaagtg ttagggatgg agaaagaagc cttagtgga
 601 gatttcatga ttagttatgg agaagcctta agtaaatattg gtttctctac ccgtgagatc
 5 661 atgaaaatgt ctgcttacgg gctacgggct gcttaa

Aminoácidos de cce_0778 de *Cyanothece* sp. ATCC51142 (YP_001802195) (SEQ ID NO: 28)

1 mqelalrsel dfnsetykda ysrinaivie geqeayqnyl dmaqllpede aelirlskme
 61 nrhkkqfqc gknlvtpdm dyaqqffael hgnfqkakae gkivtclliq sliieafaia
 121 ayniyipvad pfarkitegv vkdeythlnf gevwlkehfe askaeledan kenlplvwqm
 181 lnqvekdaev lgmekealve dfmisygeal snigfstrei mkmsayglra a

Nucleótidos de Cyan7425_0398 de *Cyanothece* sp. PCC7245 (YP_002481151) (SEQ ID NO: 29)

1 atgcctcaag tgcagtcccc atcggctata gacttctaca gtgagaccta ccaggatgct
 61 tacagccgca ttgatgcgat cgtgatcgag ggagaacagg aagcccacga caattacctg
 121 aagctgacgg aactgctgcc ggattgtcaa gaagatctgg tccggctggc caaaatggaa
 181 gcccgtcaca aaaaagggtt tgaagcttgt ggccgcaatc tcaaggtcac acccgatatg
 241 gagtttgctc aacagttctt tgctgacctg cacaacaatt tccagaaagc tgctgcggcc
 301 aacaaaattg ccactgtct ggtgatccag gccctgatta ttgagtgtt tgccatcgcc
 361 gcttataaca tctatattcc tgtcgtgat gactttgccc gcaaaattac cgaaaacgtg
 421 gtcaaagacg aatacacca cctcaacttt ggtgaagagt ggctcaaagc taactttgat
 481 agccagcggg aagaagtgga agcggccaac cgggaaaacc tgccgatcgt ctggcggatg
 541 ctcaatcagg tagagactga tgctcacggt ttaggtatgg aaaaagaggc tttagtgga
 601 agcttcatga tccaatatgg tgaagccctg gaaaatattg gtttctctac ccgtgagatc
 10 661 atgcgcatgt ccgtttacgg cctctctgcg gcataa

Aminoácidos de Cyan7425_0398 de *Cyanothece* sp. PCC7245 (YP_002481151) (SEQ ID NO: 30)

ES 2 667 469 T3

1 mpqvqpsai dfysetyqda ysridaivie geqeahdnyl kltellpdcq edlvrlakme
 61 arhkkqfeac grnlkvtpdm efaqqffadl hnnfqkaaaa nkiatclviq aliiecfaiia
 121 ayniyipvad dfarkitenv vkdeythlnf geewlkanfd sqreeveaan renlpivwrm
 181 lnqvetaahv lgmekealve sfmiqygeal enigfstrei mrmsvyglsa a

Nucleótidos de Cyan7425_2986 de *Cyanothece* sp. PCC7245 (YP_002483683) (SEQ ID NO: 31)

1 atgtctgatt ggcaccagaa cccagccctc gactattaca gtgaaaccta ccgcaatgct
 61 taccggcggg tgaacgggat tgtgattgaa ggcgagaagc aagcctacga caactttatc
 121 cgcttagctg agctgctccc agagtatcaa gcggaattaa cccgtctggc taaaatggaa
 181 gcccgccacc agaagagctt tgttgccctg gcccaaaatc tcaaggttag cccggactta
 241 gactttgctg cacagttttt tgctgaactg catcaaattt ttgcatctgc agcaaatgctg
 301 ggccaggtgg ctacctgtct ggttgctgcaa gccctgatca ttgaatgctt tgcgatcgcc
 361 gcctacaata cctatttgcc agtagcggat gaatttgccc gtaaagtacac cgcacccggt
 421 gttcaggacg agtacagcca cctaaacttt ggtgaagtct ggctgcagaa tgcggttgag
 481 cagtgtaaag acgaaattat cacagctaac cgtcttgctc tgccgctgat ctggaaaatg
 541 ctcaaccagg tgacaggcga attgcgcatt ctgggcatgg acaagcttc tctggtagaa
 601 gacttttagca ctgctatgg agaggccctg ggccagattg gtttcaaact atctgaaatt
 661 ctctccctgt ccgttcaggg tttacaggcg gttacgcctt ag

Aminoácidos de Cyan7425_2986 de *Cyanothece* sp. PCC7245 (YP_002483683) (SEQ ID NO: 32)

1 msdcatnpal dysetyrna yrrvngivie gekqaydnfi rlaellpeyq aeltrlakme
 61 arhqksfvac gqnlkvspdl dfaaqffael hqifasaana gqvattclvvq aliiecfaiia
 121 ayntylpvad efarkvtasv vqdeyshlnf gevwlqnafe qckdeiitan rlalpliwm
 5 181 lnqvtgelri lgmdkaslve dfstrygeal gqigfklsei lsksvqglqa vtp

Nucleótidos de YP_323043 de *Anabaena variabilis* ATCC29413 (Ava_2533) (SEQ ID NO: 33)

1 atgcagcagg ttgcagccga tttagaaatc gatttcaaga gcgaaaaata taaagatgcc
 61 tatagtcgca taaatgcgat cgtgattgaa ggggaacaag aagcatatga gaattacatt
 121 caactatccc aactgctgcc agacgataaa gaagacctaa ttgcctctc gaaaatggaa
 181 agtcgccaca aaaaaggatt tgaagcttgt ggacggaacc tgcaagtatc cccagacata
 241 gagttcgcta agaattctt tgccgggcta cacggtaatt tccaaaaagc ggcagctgaa
 301 ggtaaagtgt tcaactgcct attgattcaa tccctgatta ttgaatgttt tgcgatcgcc
 361 gcatacaata tctacatccc cgtggctgac gatttcgccc gtaaaatcac tgaggggtgta
 421 gttaaagatg aatacagtca cctcaacttc ggcgaagtgt gttacagaa aaatttcgct
 481 caatcaaaag cagaactaga agaagctaac cgtcataatc ttccatagc ctggaaaatg
 541 ctcaatcaag ttgccgatga tgcggcagtc ttagctatgg aaaaagaagc cctagtgaa
 601 gattttatga ttcagtacgg cgaagcacta agtaattatg gttcacaac cagagatatt
 661 atgcggatgt cagcctacgg actcacagca gcttaa

Aminoácidos de YP_323043 de *Anabaena variabilis* ATCC29413 (Ava_2533) (SEQ ID NO: 34)

1 mqqvaadlei dfkseykda ysrinaivie geqeayenyi qlsqllpddk edlirlskme
 61 srhkkqfeac grnlqvspdi efakeffagl hgnfqkaaae gkvvtclliq sliiecfaiia
 121 ayniyipvad dfarkitegv vkdeyshlnf gevwlqknfa qskaeleean rhnlpivwkm
 181 lnqvaddaav lamekealve dfmiqygeal snigfttrdi mrmsayglta a

10 Nucleótidos de YP_170760 de *Synechococcus elongatus* PCC6301 (syc0050_d) (SEQ ID NO: 35)

1 gtgcgtaccc cctgggatcc accaaatccc acattctccc tctcatccgt gtcaggagac

```

61 cgcgactca tgcgcagct tgaagccagc cttgaactgg actttcaaag cgagtcctac
121 aaagacgctt acagecgcac caacgcgacg gtgattgaag gcgaacaaga ggcgttcgac
181 aactacaatc gccttgctga gatgctgccc gaccagcggg atgagcttca caagctagcc
241 aagatggaac agcgcacacat gaaaggcttt atggcctgtg gcaaaaatct ctccgtcact
301 cctgacatgg gttttgccc gaaatttttc gagcgccttg acgagaactt caaagcggcg
361 gctgcggaag gcaaggctgt cacctgcta ctgattcaat cgctaatcat cgagtgcctt
421 gcgatcgcgg cttacaacat ctacatccca gtggcggatg cttttgccc caaaatcacg
481 gagggggtcg tgcgcgacga atacctgcac cgcaacttcg gtgaagagtg gctgaaggcg
541 aattttgatg cttccaaagc cgaactggaa gaagccaate gtcagaacct gcccttggtt
601 tggctaatac tcaacgaagt ggccgatgat gctcgcgaac tcgggatgga gcgtgagtcg
661 ctgcgagagg actttatgat tgccatcggg gaagctctgg aaaacatcgg cttcacaacg
721 cgcgaaatca tgcgtatgtc cgcctatggc cttgcggccg tttga

```

Aminoácidos de YP_170760 de *Synechococcus elongatus* PCC6301 (syc0050_d) (SEQ ID NO: 36)

```

1 mrtpwdppnp tfslssvsdg rrlmpqleas leldfqsesy kdaysrinai viegeqeafd
61 nynrlaemlp dqrdelhkla kmeqrhmkgf macgknlsvt pdmgfaqkff erlhenfkaa
121 aaegkvvtcl liqsliecf aiaayniyip vadafarkit egvvrdeylh rnfgeewlka
181 nfdaskaele eanrqnlplv wlmlnevadd arelmeres lvedfmiayg ealenigftt
241 reimrmsayg laav

```

Motivo 1: (SEQ ID NO: 37)

5 **Y-X(2)-A-Y-X-R-X(4)-V-X(2)-G-E-X(2)-A**

Motivo 2: (SEQ ID NO: 38)

L-X(2)-M-E-X(2)-H-X(3)-F-X(2)-C

Motivo 3: (SEQ ID NO: 39)

C-X(3)-Q-X(4)-E-X-F-A-X(2)-A-Y

10 Motivo 4: (SEQ ID NO: 40)

T-X(2)-V-X(3)-E-X(2)-H-X(6)-W-L

Motivo 5: (SEQ ID NO: 41)

Y-[KRQ]-X-A-Y-[SR]-R-[IV]-[DN]-[GA]-[VI]-V-[IL]-X-G-E-X(2)-A-X(2)-[NH]-[YF]

Motivo 6: (SEQ ID NO: 42)

15 **L-[ASG]-X-M-E-X-[RDK]-H-X(2)-[ASG]-F-X-[ASG]-C-[AG]-X-N-L-X-[VI]**

Motivo 7: (SEQ ID NO: 43)

C-[LMF]-[VL]-[IV]-Q-X-[LI]-[ILM]-[IV]-E-X-F-A-[IV]-[AS]-A-Y-[NHR]-X-Y-[IL]

Motivo 8: (SEQ ID NO: 44)

D-X-[FY]-A-[RK]-X-[IV]-T-X(2)-V-[VM]-X-[DE]-E-[YS]-X-H-[LR]-[ND]-[YF]-[AG]-[EQ]-X-W-L

20 Nucleótidos de Npun02004178 de *Nostoc punctiforme* PCC73102 (ZP_00108838), sintéticos (SEQ ID NO: 45)

ES 2 667 469 T3

1 atgcagcaac tgacggatca gagcaaagaa ctggacttca aaagcgaaac ctacaaggac
 61 gcgtattctc gtatcaacgc tatcgttata gaggggtgaac aagaagcgca cgagaattac
 121 attaccctgg cgcagctgct gcctgaatcc cacgatgaac tgattcgtct gagcaaaatg
 181 gagtcgctc acaaaaaggg ttttgaggcc tgcggtcgtta acctggcggc cactccggac
 241 ctgcagttcg ctaaggagtt cttcagcggc ctgcatcaaa actttcagac ggcagcggcg
 301 gaaggtaagg ttgtcacctg cctgctgatt caaagcctga tcattgagtg tttcgtatc
 361 gcagcctata acatttacat cccgggtggc gagcattttg cacgcaagat cactgagggt
 421 gtggttaaag aagaatacag ccacctgaac ttcggtgagg tctggttgaa ggagcacttt
 481 gcggaaagca aggcggagct ggaattggca aatcgtcaaa acctgccgat cgtgtggaaa
 541 atgctgaatc aagtggaggg tgatgcacac acgatggcta tggaaaaaga cgctctggtg
 601 gaggacttca tgatccagta cggcgaggcg ctgagcaaca ttggctttag caccctgac
 661 attatgcgcc tgagcgcgta tggcctgac ggtgcgtaa

Nucleótidos de AM1_4041 de *Acaryochloris marina* MBIC11017 (YP_001518340), sintéticos (SEQ ID NO: 46)

1 atgccgcaaa cgcaagctat tagcgaaatt gatttctatt ctgacaccta taaggacgct
 61 tactctcgta tcgatggat cgtgatcgag ggtgagcaag aggcgatga gaactacatt
 121 cgtctgggtg aaatgttgcc tgagcatcaa gacgacttta tccgtttgag caagatggag
 181 gcccgtcaca agaagggtt tgaggcttgt ggtcgttaact tgaagggtgac ttgcgatctg
 241 gacttcgctc gtcgcttctt ctcggacctg cacaagaact tccaagatgc tgcggccgag
 301 gataaagttc cgacctgctt ggttattcag tccctgatca tcgaatgctt cgcgattgca
 361 gcgtataaca tttacatccc ggttgccgat gatttcgctc gtaagattac cgagagcgtc
 421 gtcaaggacg aataccagca tctgaactat ggcgaggagt ggctgaaggc ccatttcgac
 481 gacgtgaagg ccgagatcca ggaagcaaat cgcaagaatc tgccgatcgt ttggcgtatg
 541 ctgaacgagg ttgacaagga cgcagcagtg ctgggcatgg agaaggaagc gttggtttaa
 601 gacttcatga ttcaatacgg tgaggccctg tccaacattg gcttttctac cggcgagatc
 661 atgcgtatgt ctgcgtacgg tctggtggca gcctaa

Nucleótidos de fll1313 de *Thermosynechococcus elongatus* BP-1 (NP_682103), sintéticos (SEQ ID NO: 47)

1 atgaccaccg cgaccgcaac gccggtgctg gactatcaca gcgaccgcta caaggacgca
 61 tacagccgca tcaacgcgat tgtcatcgaa ggtgaacaag aggccacga caattacatt
 121 gatctggcta aactgctgcc tcaacaccaa gaagagctga cccgtctggc gaagatggag
 181 gcccgccaca agaagggtt tgaagcgtgc ggtcgcgaatc tgcctgttac cccggatatg
 241 gagttcgcga aagcgttctt tgagaagctg cgcgcgaact ttcagcgtgc cctggcggag
 301 ggtaagaccg caacctgtct gctgatccag gcggtgatca ttgaatcctt cgcaattgcc
 361 gcgtacaaca tttacatccc tatggccgat ccgtttgctc gcaagattac cgaaagcgtc
 421 gtcaaggatg aatactctca cttgaacttt ggcgaaatct ggttgaagga acatttcgag
 481 agcgtcaagg gcgagttgga ggaagctaac cgtgcgaatc tgccgctggt ttggaagatg
 541 ttgaatcagg tcgaggcaga cgaaaggtc ctgggcatgg agaaggatgc tctggtggaa
 601 gactttatga tccagtactc cgggtgcctg gagaacatcg gctttaccac ccgtgaaatc
 5 661 atgaaaatgt ctgtgtatgg cctgaccggc gcgtaa

Nucleótidos de CYA_0415 de *Synechococcus* sp. JA-3-3A (YP_473897), sintéticos (SEQ ID NO: 48)

1 atggcgctc caaacgtgct gccaaatag ccgcccagcc cgaccgatgg tgggtgtacg
 61 gccctggact acagctctcc gcgttaccgt caggcgtaca gccgatcaa tggcattggt
 121 atcgaaggcg agcaggaagc gcacgataac tacctgaagt tggcggagat gctgcctgag
 181 gctgccgagg aactgcgtaa gctggcaaa atggaattgc gtcacatgaa gggctttcag
 241 gcttgccgca agaacttgca ggtggagcct gacgtcagat ttgcccgcgc tttcttcgcg
 301 ccgctgcgca acaacttcca atccgcagca gcggccgggtg atctggtttc ctgtttcgtc
 361 atccaaagcc tgatcatcga gtgttttgcg atcgtcgcgt ataacattta catcccgtt
 421 gcagacgact tcgcccgtaa gatcacggag ggcgtgggta aggacgagta tctgcatctg
 481 aatttcggcg agcgttggtt ggggtgaacac ttcgcagagg ttaaagcaca gatcgaggca
 541 gccaatgcc agaacctgcc gctggtgcgc caaatgctgc agcaagttga ggcggacgct
 601 gaggcaatct atatggaccg tgaggcgatc gttgaggatt tcatgattgc ttatggcgaa
 661 gcgctggcaa gcattggctt caacacgcgc gaagtgatgc gtctgagcgc acagggcttg
 721 cgtgcagcat aa

ES 2 667 469 T3

Nucleótidos de PM123 de *Prochlorococcus marinus* MIT9313 (NP_895059), sintéticos (SEQ ID NO: 49)

```

1 atgccgacgt tggagatgcc ggtcgcgtgcg gtcctggaca gcacggtcggy tagctctgag
61 gcgctgccgg actttaccag cgaccgctac aaagacgctt attcgcgcat caacgcgatt
121 gtgatcgagg gtgaacaaga agcccacgac aactacatcg caattggcac cctggtgccc
181 gaccatgtgg aagaactgaa acgtctggcg aaaatggaaa tgcgtcacia gaaaggtttc
241 accgcgtgcy gtaagaactt ggggtgtgaa gccgatatgg acttcgcccg tgagttcttt
301 gccccgttgc gcgacaactt tcaaaccgcy ctgggtcaag gcaagacccc tacgtgtctg
361 ttgatccaag cgctgctgat tgaagcgttc gcgatctcgy cctaccacac ttacattccg
421 gttagcgcac cgttcgcacg taagatcact gaagggtgcy ttaaggacga atacacccat
481 ctgaactacg gtgaggcatg gctgaaggcy aatctggaga gctgccgcga ggaactgctg
541 gaagcgaacc gtgagaatct gccgctgatc cgccgcacgc tggatcaggt cgcgggcgac
601 gcggcagtc tgcagatgga taaggaagac ctgatcgaag acttcctgat tgcttaccia
661 gagagcttga ctgagatcgy ctttaacacg cgtgaaatca cccgatggc cgcagcggcy
721 ctggctcagct aa

```

Nucleótidos de PMM0532 de *Prochlorococcus marinus* subesp. *pastoris* cepa CCMP1986 (NP_892650), sintéticos (SEQ ID NO: 50)

```

1 atgcaaacc tggagagcaa caagaaaacc aacctggaaa acagcattga cctgccagat
61 ttcacgacgy acagctacia ggatgcgcat tcccgatca atgctatcgt cattgaaggt
121 gaacaggaag cccatgacia ctatatcagc ctggccacc tgatcccga tgaactggag
181 gaattgacca aactggccaa gatggagctg aacacaaac gtggctttac ggcattcggg
241 cgcaatctgg gtgttcaggc cgatatgatc tttgcgaaag agtttttctc taagctgcac
301 ggcaacttcc aagttgcgct gagcaacggt aagacgacca cctgcttgct gatccaggcc
361 atcttgattg aagccttcgc gatttcgcy taccacgtgt acattcgtgt cgcggacccg
421 tttgcgaaaa agattactca aggtgtggtg aaggatgagt acctgcacct gaactatggt
481 caggaatggt tgaaggagaa tctggcaacc tgaaggacg aactgatgga agcaaacia
541 gttaactcgc cgctgattaa gaaaatgctg gatcaggtga gcgaggatgc ctctgtgttg
601 gctatggatc gtgaggagct gatggaggag ttcattgatc cgtatcagga caccctgttg
5 661 gaaatcggtc tggacaatc tgaattgcy cgtatggcaa tggctgcgat tgtgtaa

```

Nucleótidos de PMN2A_1863 de *Prochlorococcus marinus* cepa NATL2A (YP_293054), sintéticos (SEQ ID NO: 51)

```

1 atgcaggcct tcgcaagcaa taacctgacg gtcgaaaagg aagaactgag ctccaatagc
61 ctgccggatt tcaccagcga gagctataag gatgcatact ctcgatcaa tgccgtgggt
121 atcgaaggtg aacaagaggc ttattctaac tttctggacc tggccaagct gatcccggag
181 cacgccgacg agctggtgcy cttgggtaag atggaaaaga aacacatgaa cggcttctgc
241 gcgtgtggtc gtaacttggc agttaaacca gacatgccgt tcgcgaagac gttcttttagc
301 aagctgcaca acaatttcct ggaggcgttt aagggtggcy atacgacgac ctgtttgttg
361 atccaatgca tcttgatcga gtcctttgcc atcagcgcgt accacgtgta cattcgcgtg
421 gcagatccgt ttgccaagcy tatcacgga ggtgttgttc aagacgagta cctgcatttg
481 aattacggtc aagagtggct gaaagcgaac ctggagactg tgaagaaaga cctgatgcgc
541 gcgaaciaag agaactcgc attgattaag tctatgctgy acgaagtct caacgacgct
601 gaagtgctgc acatggataa agaagagctg atggaagagt ttatgattgc atatcaggac
661 agcctgctgy aaattggcct ggacaaccgc gagatcgcac gcatggcgt ggcagcgggt
721 atttaa

```

Nucleótidos de RS9917_09941 de *Synechococcus* sp. RS9917 (ZP_01079772), sintéticos (SEQ ID NO: 52)

```

1 atgccgacc tggaaactag cgaggtggca gttctggaag actcgatggc cagcggtagc
61 cgctgccgg actttaccag cgaggcctat aaggacgcgt atagcogtat caatgcgac
121 gtgattgaag gcgagcaaga agcgcacgac aactacattg cactgggcac gctgatcca
181 gaacagaagg acgagctggc tcgctggtc cgtatggaaa tgaacacat gaagggttt
241 accagctgtg gtcgtaacct ggggtgtgaa gcggatctgc cgttcgcgaa ggagttcttc

```

301 gcaccgctgc atggtaactt tcaggcggcg ctgcaggaag gtaagggtggt gacctgtctg
 361 ctgattcagg cactgctgat tgaggcgttc gccattagcg cttatcacat ttacattccg
 421 gttgctgacc cgtttgcacg caagattacc gaagggtgtg tgaaagacga gtatacccat
 481 ctgaactacg gtcaagagtg gttgaaggcg aatttcgaag cctccaaaga cgaactgatg
 541 gaagccaaca aggcgaatct gccgctgac cgttctatgc tggaacaagt cgctgctgat
 601 gcggccgtgc tgcaaatgga gaaagaggac ctgattgaag acttctctgat cgcatatcaa
 661 gaagctctgt gtgagattgg cttctcgtcc cgtgatatcg cccgcatggc ggcagccgca
 721 ctggcggttt aa

Nucleótidos de RS9917_12945 de *Synechococcus* sp. RS9917 (ZP_01080370), sintéticos (SEQ ID NO: 53)

1 atgacccaat tggactttgc atctgcggca taccgtgagg catacagccg tatcaatggt
 61 gtcgttattg ttggcgaggg cctggcgaat cgtcacttcc aaatgctggc gcgtcgcatt
 121 ccggcagacc gtgacgaatt gcaacgtttg ggccgcatgg aggggtgacca cgcaagcgc
 181 tttggtgggt gcggtcgcaa tctgggtgtg gtcgctgatc tgccgctggc acgcccgtg
 241 ttccagccgc tgcattgatc gttcaagcgt cacgaccacg acggtaaccg tgctgaatgc
 301 ctggtgatcc agggctctgat tgttgagtgc tttgcggttg ccgctgatcg tcattacctg
 361 ccggtggcag acgctgatgc ccgtccgatc accgctgcgg ttatgaatga cgagagcgaa
 421 cacctggact acgcagaaac ctggctgcag cgccacttcc accaagttaa agcccgcgtg
 481 agcgtctgtg ttgtggaggg gctgccgctg acgctggcga tgttgcaaag cctggctgca
 541 gatatgcgcc aaatcggcat ggaccggtg gaaacgctgg cgagcttcag cgagctgttt
 601 cgtgaagcgc tggaaagcgt tggttttgaa gcggtcgaag cgcgccgttt gctgatgcgt
 661 gctgcagctc gtatggttta a

Motivo 9: (SEQ ID NO: 54)

5 **G-A-X-G-D-I-G-S-X(4)-W-X(12,13)-A-R**

Motivo 10: (SEQ ID NO: 55)

A-T-V-A-[VI]-X-G-A-T-G-D-I-G-S-A-V-[CT]-R-W-L-[DI]-X-K-X(4) -[DE]-L-[ILF]-L-[TIV]-A-R

Motivo 11: (SEQ ID NO: 56)

{K}-L-X-[FLW]-X-R-F-T-T-G-N

10 en el que {K} corresponde a cualquier aminoácido excepto Lys

Motivo 12: (SEQ ID NO: 57)

M-F-G-L-I-G-H-X(6)-A

Motivo 13: (SEQ ID NO: 58)

L-[DE]-[FLVM]-W-[CS]-X-A-P-P-[QV]-X-[LV]-[DE]-X(3)-[VI]-X-S

15 Motivo 14: (SEQ ID NO: 59)

S-X(2)-G-X(2)-I-X-G-X-Y-X(2)-S-X-F-X-PEML

Motivo 15: (SEQ ID NO: 60)

K-X-A-X-R-K-X(3)-A-M-X(3)-Q-X(5)-I-X(2)-LGGF

Motivo 16: (SEQ ID NO: 61)

20 **A-[LVM]-X(2)-[AS]-[DN]-X-[VI]-[VI]-W-V-A-S-[LMI]**

Motivo 17: (SEQ ID NO: 62)

P-X(4)-D-G-G-Y-P-K-N

Motivo 18: (SEQ ID NO: 63)

N-F-S-W-G-R-N-X-I-X(7)-I-G-X(2)S-X(2)-H-G

Motivo 19: (SEQ ID NO: 64)

FTTGNTHTA

5 Nucleótidos de Synpcc7942_1594 de *Synechococcus elongatus* PCC7942 (YP_400611) (SEQ ID NO: 65)

```

1 atgttcggtc ttatcggta tctcaccagt ttggagcagg cccgcgacgt ttctcgcagg
61 atgggctacg acgaatacgc cgatcaagga ttggagtttt ggagtagcgc tcctcctcaa
121 atcgttgatg aatcacaggt caccagtgcc acaggcaagg tgattcacgg tcgctacatc
181 gaatcgtggt tcttgccgga aatgctggcg gcgcgcccgt tcaaacacagc cacgcgcaaa
241 gttctcaatg ccatgtccca tgcccaaaaa cacggcatcg acatctcggc cttggggggc
301 tttacctcga ttattttcga gaatttcgat ttggccagtt tgcggcaagt gcgcgacact
361 accttgaggt ttgaacgggt caccaccggc aatactcaca cggcctacgt aatctgtaga
421 caggtggaag ccgctgctaa aacgctgggc atcgacatta cccaagcgac agtagcgggt
481 gtcggcgcga ctggcgatat cggtagcgcg gtctgccgct ggctcgacct caaactgggt
541 gtcggtgatt tgatcctgac ggcgcgcaat caggagcgtt tggataacct gcaggctgaa
601 ctcggccggg gcaagattct gcccttgga gccgctctgc cggagctga ctttatcgtg
661 tgggtcgcca gtatgcctca gggcgtagt atcgaccag caaccctgaa gcaaccctgc
721 gtcctaactg acgggggcta ccccaaaaac ttgggcagca aagtccaagg tgagggcatc
781 tatgtcctca atggcgggggt agttgaacat tgcttcgaca tcgactggca gatcatgtcc
841 gctgcagaga tggcgcggcc cgagcgccag atgtttgcct gctttgccga ggcgatgctc
901 ttggaatttg aaggctggca tactaacttc tcctggggcc gcaaccaa atcacgatcgag
961 aagatggaag cgatcgggtg ggcacgggt cccacccctt ggcattggca
1021 atttga
    
```

Aminoácidos de Synpcc7942_1594 de *Synechococcus elongatus* PCC7942 (YP_400611) (SEQ ID NO: 66)

```

1 mfglighlts leqardvsrr mgydeyadqg lefwssappq ivdeitvtsa tgkvihgryi
61 escflpempla arrfktatrkl vlnamshaqk hgidisalgg ftsiifenfd laslrqvrtd
121 tleferfttg nthtayvicr qveaaaktlg iditqatvav vgatgdigsa vcrwldklg
181 vgdliiltarn qerldnlqae lgrgkilple aalpeadfiv wvasmpqgvv idpatlkqpc
241 vlidggypkn lgskvqgegi yvlnggvveh cfdidwqims aaemarperq mfacfaeaml
301 lefegwhtnf swgrnqitie kmeaigeasv rhgfgplala i
    
```

Nucleótidos de sl0209 de *Synechocystis* sp. PCC6803 (NP_442146) (SEQ ID NO: 67)

```

1 ATGTTTGGTC TTATTGGTCA TCTCACGAGT TTAGAACACG CCCAAGCGGT TGCTGAAGAT
61 TTAGGCTATC CTGAGTACGC CAACCAAGGC CTGGATTTTT GGTGTTCCGC TCCTCCCCAA
121 GTGGTTGATA ATTTTCAGGT GAAAAGTGTG ACGGGGCAGG TGATTGAAGG CAAATATGTG
181 GAGTCTTGCT TTTTGCCGGA AATGTTAACC CAACGGCGGA TCAAAGCGGC CATTTCGTA
241 ATCCTCAATG CTATGGCCCT GGCCAAAAG GTGGGCTTGG ATATTACGGC CCTGGGAGGC
301 TTTTCTTCAA TCGTATTTGA AGAATTTAAC CTCAAGCAA ATAATCAAGT CCGCAATGTG
361 GAACTAGATT TTCAGCGGTT CACCACTGGT AATACCACA CCGCTTATGT GATCTGCCGT
421 CAGGTCGAGT CTGGAGCTAA ACAGTTGGGT ATTGATCTAA GTCAGGCAAC GGTAGCGGTT
481 TGTGGCGCCA CGGGAGATAT TGGTAGCGCC GTATGTCGTT GGTTAGATAG CAAACATCAA
541 GTTAAGGAAT TATTGCTAAT TGCCCGTAAC CGCCAAAGAT TGAAAATCT CCAAGAGGAA
10 601 TTGGGTCGGG GCAAATTAT GGATTTGAA ACAGCCCTGC CCCAGGCAGA TATTATIGTT
661 TGGGTGGCTA GTATGCCCAA GGGGGTAGAA ATTGCGGGGG AAATGCTGAA AAAGCCCTGT
721 TTGATTGTGG ATGGGGGCTA TCCCAAGAAT TTAGACACCA GGGTGAAAGC GGATGGGGTG
781 CATATTCTCA AGGGGGGGAT TGTAGAACAT TCCCTTGATA TTACCTGGGA AATTATGAAG
841 ATTGTGGAGA TGGATATTCC CTCCCGGCAA ATGTTGCGCT GTTTTGCGGA GGCCATTTTG
901 CTAGAGTTTG AGGGCTGGCG CACTAATTTT TCCTGGGGCC GCAACCAAAT TTCCGTTAAT
961 AAAATGGAGG CGATTGGTGA AGCTTCTGTC AAGCATGGCT TTTGCCCTTT AGTAGCTCTT
1021 TAG
    
```

Aminoácidos de sl0209 de *Synechocystis* sp. PCC6803 (NP_442146) (SEQ ID NO: 68)

ES 2 667 469 T3

1 mfglighlts lehaqavaed lgypeyanqg ldfwcsappq vvdnfvkxsv tgqviegkyv
 61 escflpemlt qrrikaairk ilnamalagk vgllditalgg fssivfeefn lkqnnqvrnv
 121 eldfqrfttg nthtayvicr qvesgakqlg idlsqatvav cgatgdigsa vcrwldskhq
 181 vkellliarn rqrlnlqee lgrgkimdle talpqadiiv wvasmpkgve iagemlkkpc
 241 livdggypkn ldtrvkadgv hilkggiveh slditweimk ivemdipsrq mfacfaeail
 301 lefegwrtnf swgrnqisvn kmeaigeasv khgfcplval

Nucleótidos de cce_1430 de *Cyanothece* sp. ATCC51142 (YP_001802846) (SEQ ID NO: 69)

1 atgtttgggt taattgggtca tcttacaagt ttagaacacg cccactccgt tgctgatgcc
 61 tttggctatg gccatacgc cactcagga cttgatttgt ggtgttctgc tccaccccaa
 121 ttcgtcgcgc attttcatgt tactagcatc acaggacaaa ccatcgaagg aaagtatata
 181 gaatccgctt tcttaccaga aatgctgata aagcgacgga ttaaagcagc aattcgcaaa
 241 aactgcaatg cgatggcctt tgctcagaaa aataacctta acatcacagc attagggggc
 301 ttttcttcga ttatttttga agaatttaac ctcaaagaga atagacaagt tcgtaatgtc
 361 tcttttagagt ttgatcgtt caccaccgga aacaccata ctgcttatat catttgcgt
 421 caagttgaac aggcacccgc taaactaggg attgacttat cccaagcaac ggttgctatt
 481 tgcggggcaa ccggagatat tggcagtgca gtgtgtcgtt ggtagatag aaaaaccgat
 541 acccaggaac tattcttaac tgctcgaat aaagaacgat tacaacgact gcaagatgag
 601 ttgggacggg gtaaaattat gggattggag gaggctttac ccgaagcaga tattatcgtt
 661 tgggtggcga gtatgcccaa aggagtggaa attaatgccg aaactctcaa aaaaccctgt
 721 ttaattatcg atggtggta tcctaagaat ttagacacaa aaattaaaca tctgatgtc
 781 catatcctga aagggggaat ttagaacat tctctagata ttgactggaa gattatggaa
 841 actgtcaata tggatgttcc ttctcgtcaa atgtttgctt gttttgccga agccatttta
 901 ttagagtttg aacaatggca cactaatttt tcttggggac gcaatcaaat tacagtgact
 961 aaaatggaac aataggaga agcttctgtc aaacatgggt tacaaccgtt gttgagttgg
 1021 taa

Aminoácidos de cce_1430 de *Cyanothece* sp. ATCC51142 (YP_001802846) (SEQ ID NO: 70)

1 mfglighlts lehahsvada fgygpyatgg ldlwcsappq fvehfhvtsi tgqtiegkyi
 61 esaflepemli krrikaairk ilnamafaqk nlnlitalgg fssiifeefn lkenrqvrnv
 121 slefdrfttg nthtayiicr qveqasaklg idlsqatvai cgatgdigsa vcrwldrkt
 181 tqelfliarn kerlqrlqde lgrgkimgle ealpeadiiv wvasmpkgve inaetlkkpc
 241 liidggypkn ldtkikhpdv hilkggiveh sldidwkime tvnmdvpsrq mfacfaeail
 301 lefeqwhtnf swgrnqitvt kmeqigeasv khglqplls

5

Nucleótidos de PMM0533 de *Prochlorococcus marinus* subesp. *pastoris* cepa CCMP1986 (NP_892651) (SEQ ID NO: 71)

1 atgtttgggc ttataggtca ttcaactagt tttgaagatg caaaaagaaa ggcttcatta
 61 ttgggctttg atcatattgc ggatgggtgat ttagatgttt ggtgcacagc tccacctcaa
 121 ctagttgaaa atgtagaggt taaaagtgc ataggtatat caattgaagg ttcttatatt
 181 gattcatggt tcgttctga aatgctttca agatttaaaa cggcaagaag aaaagtatta
 241 aatgcaatgg aattagctca aaaaaaagggt attaataata ccgctttggg ggggttact
 301 tctatcatct ttgaaaattt taatctcctt caacataagc agattagaaa cacttacta
 361 gagtgggaaa ggtttacaac tggaataact catactcgtt ggttatttg caggcaatta
 421 gagatgaatg ctctaaaat aggtattgat cttaaagcg caacagttgc tgtagttggt
 481 gctactggag atataggcag tgctgtttgt cgatgggttaa tcaataaaac aggtattggg
 541 gaacttcttt tggtagctag gcaaaaggaa cccttggtt ctttgcaaaa ggaattagat
 601 ggtggaacta tcaaaaatct agatgaagca ttgcctgaag cagatattgt tgtatgggta
 661 gcaagtatgc caaagacaat ggaaatcgat gctaataatc ttaacaacc atgtttaatg
 721 attgatggag gttatccaaa gaatctagat gaaaaatttc aaggaaataa tatacatggt
 781 gtaaaaggag gtatagtaag attcttcaat gatataggtt ggaatatgat ggaactagct
 841 gaaatgcaaa atcccagag agaaatgttt gcatgctttg cagaagcaat gattttagaa
 901 tttgaaaaat gtcatacaaa cttagctgg ggaagaaata atatatctct cgagaaaatg
 961 gagtttattg gagctgcttc tgtaaagcat ggcttctctg caattggcct agataagcat
 1021 ccaaaagtac tagcagtttg a

ES 2 667 469 T3

Aminoácidos de PMM0533 de *Prochlorococcus marinus* subesp. *pastoris* cepa CCMP1986 (NP_892651) (SEQ ID NO: 72)

1 mfglighsts fedakrkasl lgfdhiadgd ldvwctappq lvenvevksa igisiagsyi
 61 dscfvpemls rfktarrkvl namelaqkkg initalggft siifenfnll qhkqirntsl
 121 ewerfttgnt htawvicrql emnapkigid lksatvavvg atgdigsavc rwlinktgig
 181 elllvarqke pldslqkeld ggtiknldea lpeadivvwv asmpktmeid annlkqpcml
 241 idggypknld ekfqqnnihv vkggivrffn digwnmmela emqnpqremf acfaeamile
 301 fekchtnfsw grnnislekm efigaasvkh gfsaigldkh pkvlav

Nucleótidos de NP_96091 de *Gloeobacter violaceus* PCC7421 (gll3145) (SEQ ID NO: 73)

1 atgtttggcc tgatcggaca cttgaccaat ctttcccatg cccagcgggt cgcccgcgac
 61 ctgggctacg acgagtatgc aagccacgac ctcgattct ggtgcatggc cctccccag
 121 gcggtc gatg aatcacgat caccagcgtc accggtcagg tgatccacgg tcagtacgtc
 181 gaatcgtgct ttctgccgga gatgctcgcc cagggcgcgt tcaagaccgc catgcgcaag
 241 atcctcaatg ccatggccct ggtccagaag cgcggcatcg acattacggc cctgggaggc
 301 ttctcgtcga tcattctcga gaatttcagc ctcgataaat tgctcaacgt ccgcgacatc
 361 accctcgaca tccagcgctt caccaccggc aacaccacaca cggcctacat cctttgtcag
 421 caggtcgcgc aggggtgcgg acgctacggc atcgatccgg ccaaagcgac cgtggcggta
 481 gtcggggcca ccggcgacat cggtagcgc gtctgcccgt ggtcaccga ccgcgccggc
 541 atccacgaac tcttgctggt ggcccgcgac gccgaaaggc tcgaccggct gcagcaggaa
 601 ctcggcaccg gtcggatcct gccggtcgaa gaagcacttc ccaaagccga catcgtcgtc
 661 tgggtcgcct cgatgaacca gggcatggcc atcgaccccg ccggcctgcg caccctctgc
 721 ctgctcatcg acggcggcta cccaagaac atggccggca cctgcagcg cccgggcatc
 781 catatcctcg acggcggcat ggtcgcgcac tcgctcgaca tcgactggca gatcatgctg
 841 tttctaaatg tgccaaccc cgcgccag ttcttcgcct gcttcgccga gtcgatgctg
 901 ctggaattcg aagggttca cttcaatttt tctggggcc gcaaccacat caccgctcag
 961 aagatggccc agatcggctc gctgtctaaa aaacatggct ttcgtcccct gcttgaacc
 5 1021 agtcagcgca gcggcgaact cgtacacgga taa

Aminoácidos de NP_96091 de *Gloeobacter violaceus* PCC7421 (gll3145) (SEQ ID NO: 74)

1 mfglighltn lshaqrvard lgydeyashd lefwcmappq avdeititsv tgqvihggyv
 61 escflpemla qgrfktamrk ilnamalvqk rgiditalgg fssiifenfs ldkllnvrldi
 121 tldiqrfttg nthtayilcq qveggavryg idpakatvav vgatgdigsa vcrwltdrag
 181 ihelllvard aerldrlqqe lgtgrilpve ealpkadivv wvasmnggma idpaglrtpc
 241 llidggypkn magtlqrpgi hildggmveh sldidwqims flnvpnparg ffacfaesml
 301 lefeglhfnf swgrnhitve kmaqigsllk khgfrpllep srsrgelvhg

Nucleótidos de ZP_00108837 de *Nostoc punctiforme* PCC73102 (Npun02004176) (SEQ ID NO: 75)

1 atgtttggtc taattggaca tctgactagt ttagaacacg ctcaagccgt agcccaagaa
 61 ttgggatacc cagaatatgc cgatcaaggg ctagactttt ggtgcagcgc cccgccgcaa
 121 attgctgata gtattattgt caccagtgtt actgggcaac aaattgaagg acgatatgta
 181 gaatcttgct ttttgccgga aatgctagct agtcgccgca tcaaagccgc aacacggaaa
 241 atcctcaacg ctatggccca tgcacagaag cacggcatta acatcacagc tttaggcggga

ES 2 667 469 T3

301 ttttctcga ttatTTTTga aaactttaaG ttagagcagt ttagccaagt ccgaaatatac
 361 aagctagagt ttgaacgctt caccacagga aacacgcata ctgcctacat tatttgtaag
 421 caggtggaag aagcatccaa acaactggga attaactctat caaacgcgac tgttgcggtta
 481 tgtggagcaa ctggggatat tggtagtgcc gttacacgct ggctagatgc gagaacagat
 541 gtccaagaac tctgtctaata cgcccgcgat caagaacgct tcaaagagtt gcaaggcgaa
 601 ctggggcggg ggaaaatcat gggtttgaca gaagcactac cccaagccga tgtttagatt
 661 tgggttgcta gtatgccag aggcgtggaa attgacccca ccaacttgaa acaaccctgt
 721 ttgttgattg atggtggcta tcttaaaaaac ttagcaacaa aaattcaata tcttggcgta
 781 cacgtgtaa atggtgggat tgtagagcat tccctggata ttgactggaa aattatgaaa
 841 atagtcaata tggacgtgcc agcccgctcag ttgtttgcct gttttgcgga atcaatgcta
 901 ctggaatttg agaagttata cacgaacttt tctgtggggac ggaatcagat taccgtagat
 961 aaaatggagc agattggccg ggtgtcagta aaacatggat ttagaccggt gttggtttag

Aminoácidos de ZP_00108837 de *Nostoc punctiforme* PCC73102 (Npun02004176) (SEQ ID NO: 76)

1 mfglighthts lehaqavaqe lgyypeyadqg ldfwcsappq ivdsiiivtsv tgqqiegryv
 61 escflpemla srrikaatrk ilnamahaqk hginitalgg fssiifenfk leqfsqvrni
 121 kleferfttg nthtayiick qveeaskqlg inlsnatvav cgatgdigsa vtrwldartd
 181 vqellliard qerlkelqge lgrgkimlgt ealpqadvvv wvasmprgve idpttlkqpc
 241 llidggypkn latkiqypgv hvlnnggiveh sldidwkimk ivnmdvparq lfacfaesml
 301 lefeklytnf swgrnqitvd kmeqigrvsv khgfrpllv

Nucleótidos de YP_323044 de *Anabaena variabilis* ATCC29413 (Ava_2534) (SEQ ID NO: 77)

1 atgtttggtc taattggaca tctgacaagt ttagaacacg ctcaagcggg agctcaagaa
 61 ctgggatacc cagaatacgc cgaccaaggg ctagatTTTT ggtgcagcgc tccaccgcaa
 121 atagttgacc acattaaagt tactagcatt actggtgaaa taattgaagg gaggtatgta
 181 gaatcttget ttttaccaga aatgctagcc agccgtagga ttaaagccgc aaccgcaaa
 241 gtctcaatg ctatggctca tgctcaaaaa catggcattg acatcacgcg tttgggtggt
 301 ttctcctcca ttatTTTTga aaacttcaaa ttggaacagt ttagccaagt tcgtaatgtc
 361 aactagagt ttgaacgctt cactacagge aacactcaca cagcttataat catttgctcg
 421 caggtagaac aagcatcaca acaactcggc attgaactct cccaagcaac agtagctata
 481 tgtggggcta ctggtgacat tggtagtgca gttactcget ggctggatgc caaaacagac
 541 gtaaaagaat tactgttaat cgcccgtaat caagaacgct tcaaagagtt gcaaaagcag
 601 ttgggacgcg gtaaaatcat gagcctagat gaagcattgc ctcaagctga tatttagatt
 661 tgggtagcta gtatgcctaa aggcgtggaa attaactctc aagttttgaa acaaccctgt
 721 ttattgattg atggtggtta tccgaaaaac ttgggtacaa aagttcagta tcttgggtgt
 781 tatgtactga acggaggtat cgtcgaacat tccctagata ttgactggaa aatcatgaaa
 841 atagtcaata tggatgtacc tgcacgcaa ttatttgctt gttttgcgga atctatgctc
 901 ttggaatttg agaagttgta cacgaacttt tcttgggggc gcaatcagat taccgtagac
 5 961 aaaatggagc agattggctca agcatcagtg aaacatgggt ttagaccact gctggtttag

Aminoácidos de YP_323044 de *Anabaena variabilis* ATCC29413 (Ava_2534) (SEQ ID NO: 78)

1 mfglighthts lehaqavaqe lgyypeyadqg ldfwcsappq ivdhikvtsi tgeiiegryv
 61 escflpemla srrikaatrk vlnamahaqk hgiditalgg fssiifenfk leqfsqvrnv
 121 tleferfttg nthtayiicr qveqasqqlg ielsqatvai cgatgdigsa vtrwldaktd
 181 vkellliarn qerlqelqse lgrgkimsld ealpqadivv wvasmpkgve inpqvlkqpc
 241 llidggypkn lgtkvqypgv yvlnnggiveh sldidwkimk ivnmdvparq lfacfaesml
 301 lefeklytnf swgrnqitvd kmeqigqasv khgfrpllv

Nucleótidos de YP_170761 de *Synechococcus elongatus* PCC6301 (syc0051_d) (SEQ ID NO: 79)

1 atgttcggtc ttatcggctca tctcaccagt ttggagcagg cccgcgacgt ttctcgcagg
 61 atgggctacg acgaatacgc cgatcaagga ttggagtttt ggagtagcgc tctcctcaa
 121 atcgttgatg aaatcacagt caccagtgcc acaggcaagg tgattcacgg tcgctacatac

ES 2 667 469 T3

181 gaatcgtggt tcttgccgga aatgctggcg gcgcgccgct tcaaaacagc cacgcgcaaa
 241 gttctcaatg ccatgtccca tgcccaaaaa cacggcatcg acatctcggc cttggggggc
 301 tttacctcga ttattttcga gaatttcgat ttggccagtt tgcggcaagt gcgcgacact
 361 accttgaggt ttgaacgggt caccaccggc aatactcaca cggcctacgt aatctgtaga
 421 caggtggaag ccgctgctaa aacgctgggc atcgacatta cccaagcgac agtagcgggt
 481 gtcggcgcgga ctggcgatat cggtagcgct gtctgccgct ggctcgacct caaactgggt
 541 gtcggtgatt tgatcctgac ggcgcgcaat caggagcggt tggataacct gcaggctgaa
 601 ctcgccggg gcaagattct gcccttgaa gccgctctgc cggagctga ctttatcgtg
 661 tgggtcgcca gtagcctca gggcgtagtg atcgaccag caaccctgaa gcaaccctgc
 721 gtcctaactg acgggggcta ccccaaaaac ttgggcagca aagtccaagg tgagggcac
 781 tatgtcctca atggcggggt agttgaacat tgcttcgaca tcgactggca gatcatgtcc
 841 gctgcagaga tggcgcggcc cgagcgccag atgtttgct gctttgccga ggcgatgctc
 901 ttggaatttg aaggctggca tactaacttc tcttggggcc gcaaccaa at cacgatcgag
 961 aagatggaag cgatcgggtg ggcacgggt cgccacggct tccaaccctt ggcattggca
 1021 atttga

Aminoácidos de YP_170761 de *Synechococcus elongatus* PCC6301 (syc0051_d) (SEQ ID NO: 80)

1 mfglighlts leqardvsrr mgydeyadqg lefwssappq ivdeitvtsa tgkvihgryi
 61 escflpemla arrfktatr k vlnamshaqk hgidisalgg ftsiifenfd laslrqvrtd
 121 tleferfttg nthtayvicr qveaaaktlg iditqatvav vgatgdigsa vcrwldklg
 181 vgdliiltarn qerldnlqae lgrgkilple aalpeadfiv wvasmpqgvv idpatlkqpc
 241 vlidggypkn lgskvqgegi yvlnqgvveh cfdidwqims aaemarperq mfacfaeaml
 301 lefegwhtnf swgrnqitie kmeaigeasv rhgfrplala

Nucleótidos de alr5284 de *Nostoc* sp. PCC7120 (NP_489324) (SEQ ID NO: 81)

1 atgtttggtc taattggaca tctgacaagt ttagaacacg ctcaagcggg agctcaagaa
 61 ctgggatacc cagaatacgc cgaccaaggg ctagatTTTT ggtgtagcgc tccaccgcaa
 121 atagttgacc acattaaagt tactagtatt actggtgaaa taattgaagg gaggtatgta
 181 gaatcttgct ttttaccgga gatgctagcc agtcgctcga ttaaagccgc aaccgcaaa
 241 gtctcaatg ctatggctca tgctcaaaag aatggcattg atatcacagc tttgggtggt
 301 ttctcctcca ttatttttga aaactttaaa ttggagcagt ttagccaagt tcgtaatgtg
 361 aactagagt ttgaacgctt cactacaggc aactcaca cagcatata tatttgcgg
 421 caggtagaac aagcatcaca acaactcggc attgaactct ccaagcaac agtagctata
 481 tgtggggcta ctggtgatat tggtagtgca gttactcgt ggctggatgc taaaacagac
 541 gtgaaagaat tgctgttaat cgcccgtaat caagaacgct tccaagagtt gcaagcgag
 601 ctgggacgcg gtaaaatcat gagccttgat gaagcactgc ccaagctga tatcgtagtt
 661 tgggtagcca gtagcctaa aggtgtggaa attaatectc aagtttgaa gcaaccctgt
 721 ttgctgattg atgggggtta tccgaaaaac ttgggtacaa aagttcagta tcttgggtgt
 781 tatgtactga acggcggtat cgtcgaacat tcgctggata ttgactggaa aatcatgaaa
 841 atagtcaata tggatgtacc tgcacgcaa ttatttgctt gttttgcgga atctatgctc
 901 ttggaatttg agaagttgta cacgaacttt tcttgggggc gcaatcagat taccgtagac
 5 961 aaaatggagc agattggtca agcatcagtg aaacatgggt ttagaccact gctggtttag

Aminoácidos de alr5284 de *Nostoc* sp. PCC7120 (NP_489324) (SEQ ID NO: 82)

1 mfglighlts lehaqavaqe lgypeyadqg ldfwcsappq ivdhikvtsi tgeiiegryv
 61 escflpemla srrikaatr k vlnamahaqk ngiditalgg fssiifenfk leqfsqvrnv
 121 tleferfttg nthtayiicr qveqasqqlg ielsqatvai cgatgdigsa vtrwldaktd
 181 vkellliarn qerlqelqse lgrgkimsld ealpqadivv wvasmpkgve inpqvlkqpc
 241 llidggypkn lgtkvqypgv yvlnqgiveh sldidwkimk ivnmdvparq lfacfaesml
 301 lefeklytnf swgrnqitvd kmeqigqasv khgfrpllv

Nucleótidos de Synpcc7942_1594 de *Synechococcus elongatus* PCC7942 (YP_400611), sintéticos (SEQ ID NO: 83)

ES 2 667 469 T3

1 atgtttggtc tgattgggtca cctgaccagc ttggaacaag cgcgtgacgt cagccgccgt
 61 atgggttatg atgaatacgc tgatcaaggc ctggagtttt ggagcagcgc gccaccgcag
 121 atcgtcgatg agatcacctg gacctccgca accggtaagg tcatccacgg ccgctacatt
 181 gagtcctgct tcctgcctga gatgctggca gctcgcggtt tcaaaacggc cactcgtgtaag
 241 gttctgaatg cgatgtccca tgcgcaaaag catggcattg acattagcgc cttgggcggt
 301 tttacgtcga ttatcttcga gaacttcgat ctggcctctt tgcgccaggt gcgtgacacg
 361 accttggagt ttgagcgttt taccacgggt aatacgcaca ccgcttacgt tatctgtcgc
 421 caagtgcgaag cagcagccaa aaccctgggt attgatatac cccaggccac cgtcgcctg
 481 gtgggtgcta ccggtgatata tggttccgcg gtttgccggt ggctggatct gaaactgggt
 541 gttggcgatc tgatcctgac ggcgcgtaac caggagcgtc tggacaacct gcaagccgag
 601 ttgggtcgcg gtaagatcct gccgttggag gcagcgttgc cggaggcaga cttcatcgtc
 661 tgggttgctg ctatgccgca ggggtgtgtt atcgaccgcg cgacctgaa acagccgtgc
 721 gtgctgattg atggcgggta tccgaaaaac ctgggcagca aggtccaagg cgaggggatc
 781 tatgtcctga atggcgggtg ggttgagcat tgcttcgaca ttgactggca gatcatgagc
 841 gcagcagaaa tggcgcgtcc ggagcgcaca atgtttgcct gttttgcaga agccatgctg
 901 ctggagttcg aaggctggca tacgaatttc agctggggtc gtaatcagat taccattgaa
 961 aagatggaag cgattggtga agcaagcgtg cgtcatggtt ttcagccact ggcgctggct
 1021 atttaa

Nucleótidos de PMM0533 de *Prochlorococcus marinus* subesp. *pastoris* cepa CCMP1986 (NP_892651), sintéticos (SEQ ID NO: 84)

1 atgtttggtc tgattggcca cagcacgagc tttgaggacg caaagcgtaa ggcgagcctg
 61 ctgggctttg atcatattgc tgatggcgac ctggacgtct ggtgcacggc acctccgcaa
 121 ctggttgaga atgtcgaggt gaaatcggcg attggcattt ccatcgaagg ctctacatc
 181 gacagctggt tcgtgccgga gatgttgagc cgtttcaaaa ccgcacgtcg caaagttctg
 241 aatgcaatgg agctggcaca aaagaagggc atcaacatca cggcgtcggg tggtttcacc
 301 agcattatct ttgagaactt caatctgttg cagcataaac agatccgtaa taccagcctg
 361 gagtgggaac gctttaccac gggtaacacc cacaccgctg gggatgatctg ccgccagctg
 421 gagatgaatg cgccgaaaat cggatattgac ctgaaaagcg cgacgggtggc agttgttggc
 481 gcaactggcg acattggttc ggccgtttgt cgctggctga ttaacaagac cggtatcggg
 541 gaattggttc tggctcgtcg ccagaaggag cctctggaca gcctgcaaaa agagctggac
 601 ggtggtacga tcaagaacct ggatgaagcg ctgccagaag cggacatcgt cgtctgggtc
 661 gcatctatgc cgaaaactat ggaaatcgat gccacaatc tgaacaacc gtgcctgatg
 721 atcgatggcg gctaccggaa gaacttggat gagaagtttc aaggcaataa catccacgtt
 781 gtgaaggggt gtattgtccg tttcttcaat gatctcgggt ggaacatgat ggaactggct
 841 gaaatgcaga acccgcaacg tgagatgttc gcttgttttg cggaggccat gattctggag
 901 ttcgagaaat gccataccaa tttcagctgg gctcgaaca acattagcct ggagaaaatg
 961 gagttcatcg gcgctcgcgag cgtaagcac ggcttcagcg cgattggttt ggataaacat
 1021 ccgaaggtcc tggcagttta a

5 Nucleótidos del orfMSMEG_5739 de *Mycobacterium smegmatis* cepa MC2 155 (YP_889972) (SEQ ID NO: 85)

1 atgaccagcg atgttcacga cgccacagac ggcgtcaccg aaaccgcact cgacgacgag
 61 cagtcgaccc gccgcategc cgagctgtac gccaccgatc ccgagttcgc cgccgccgca
 121 ccggtgcccg ccggtggcga cgcggcgcac aaaccggggc tgcggctggc agagatcctg
 181 cagaccctgt tcaccggcta cgggtgaccg ccggcgtgga gataccgcgc ccgtgaactg
 241 gccaccgacg agggcgggcg caccgtgacg cgtctgctgc cgcggttcga caccctcacc
 301 tacgcccagg tgtggtcgcg cgtgcaagcg gtcgccgcg cctgcgcca caacttcgcg
 361 cagccgatct accccggcga cgccgtcgcg acgatcgggt tcgcgagtc cgattacctg
 421 acgctggatc tcgtatgcgc ctacctgggc ctctgtagtg ttccgctgca gcacaacgca
 481 ccggtcagcc ggctcgcgcc gatcctggcc gaggtcgaac cgcgatcct caccgtgagc
 541 gccgaatacc tcgacctcgc agtcgaatcc gtgcgggacg tcaactcggg gtcgcagctc
 601 gtgggtgttc accatcacc cgaggtcgac gaccaccgcg acgactggc ccgcgcgctg
 661 gaacaactcg ccggcaaggg catcgcctc accaccctgg acgcgatcgc cgacgagggc
 721 gccgggctgc cggccgaacc gatctacacc gccgacctg atcagcgcct ccgcatgatc
 781 ctgtacacct cgggttcacc cggcgcaccc aagggtgcga tgtacaccga ggcgatgggtg

ES 2 667 469 T3

841 gcgcggctgt ggaccatgtc gttcatcacg ggtgacccca cgccgggtcat caacgtcaac
 901 ttcattgccgc tcaaccacct gggcggggcg atccccattt ccaccgccgt gcagaacggg
 961 ggaaccagtt acttcgtacc ggaatccgac atgtccacgc tgttcgagga tctcgcgctg
 1021 gtgcgcccga ccgaactcgg cctggttccg cgcgtcgccg acatgctcta ccagcaccac
 1081 ctcgccaccg tcgaccgcct ggtcacgcag ggcgccgacg aactgaccgc cgagaagcag
 1141 gccgggtgcc aactgcgtga gcaggtgctc ggcggacgcg tgatcaccgg attcgtcagc
 1201 accgcaccgc tggccgcgga gatgagggcg ttccctcgaca tcaccctggg cgcacacatc
 1261 gtcgacggct acgggctcac cgagaccggc gccgtgacac gcgacgggtg gatcgtgcgg
 1321 ccaccgggtga tcgactacaa gctgatcgac gttcccgaac tcggctactt cagcaccgac
 1381 aagccctacc cgcgtggcga actgctggtc aggtcgcaaa cgctgactcc cgggtactac
 1441 aagcgcgccg aggtcaccgc gaggctcttc gaccgggacg gctactacca caccggcgac
 1501 gtcattggcc agaccgcacc cgaccacctg gtgtacgtgg accgtcgcaa caacgtcctc
 1561 aaactcgcgc agggcgagtt cgtggcggtc gccaacctgg aggcgggtgt ctccggcgcg
 1621 gcgctgggtc gccagatctt cgtgtacggc aacagcgcgc gcagtttctt tctggccgtg
 1681 gtgggtcccga cgcgcgaggg gctcgcgcag tacgatccgg ccgcgctcaa ggccgcgctg
 1741 gccgactcgc tcgagcgcac cgcacgcgcac gccgaactgc aatcctacga ggtgccggcc
 1801 gatttcatcg tcgagaccga gccgttcagc gccgccaacg ggctgctgct ggggtgcgga
 1861 aaactgctgc ggccaacct caaagaccgc tacgggcagc gcctggagca gatgtacgcc
 1921 gatatcgccg ccacgcaggg caaccagtgt cgcgaactgc ggcgcgcggc cgccacacaa
 1981 ccggtgatcg acaccctcac ccaggccgct gccacgatcc tcggcaccgg gaggcgaggtg
 2041 gcatccgacg cccacttcac cgacctgggc ggggattccc tgtcggcget gacactttcg
 2101 aacctgctga gcgatttctt cggtttcgaa gttcccgtcg gcaccatcgt gaaccgggcc
 2161 accaacctcg cccaactcgc ccagcacatc gaggcgcagc gcaccgcggg tgaccgcagg
 2221 ccgagtttca ccaccgtgca cggcgcggac gccaccgaga tcggggcgag tgagctgacc
 2281 ctggacaagt tcattcgacgc cgaaacgctc cgggcccgcac cgggtctgcc caaggtcacc
 2341 accgagccac ggacgggtgt gctctcgggc gccaacggct ggctggggcc gttcctcacg
 2401 ttgcagtgcc tggaaacgct ggcacctgct ggcggcacc tcattcacgat cgtgcggggc
 2461 cgcgacgacg ccgcggcccg cgcacggetg acccaggcct acgacaccga tcccagattg
 2521 tcccgcggct tcgccgagct ggccgaccgc cacctgcggg tggtcgccgg tgacatccgc
 2581 gaccccaatc tgggcctcac acccgagatc tggcaccggc tcgcgcgccg ggtcgcacctg
 2641 gtgggtgcac cggcagcget ggtcaaccac gtgctcccct accggcagct gttcggcccc
 2701 aacgtcgtgg gcacggccga ggtgatcaag ctggccctca ccgaacggat caagcccgtc
 2761 acgtacctgt ccaccgtgct ggtggccatg gggatccccg acttcgagga ggaccggcgc
 2821 atccggaccg tgagcccggg gcgcccgctc gacggcggat acgccaacgg ctaccggcaac
 2881 agcaagtggg ccggcgaggt gctgctgcgg gaggcccacg atctgtgcgg gctgcccgtg
 2941 gcgacgttcc gctcggacat gatcctggcg catccgcgct accgcggtea ggtcaacgtg
 3001 ccagacatgt tcacgcgact cctggtgagc ctcttgatca ccggcgtcgc gccgcggctg
 3061 ttctacatcg gagacgggtga gcgcccgcgg gcgcaactacc ccggcctgac ggtcagattc
 3121 gtggccgagg cggtcacgcac gctcggcgcg cagcagcgcg agggatacgt gtectacgac
 3181 gtgatgaacc cgcacgacga cgggatctcc ctggatgtgt tcgtggactg gctgatccgg
 3241 gcggggccatc cgatcgaccg ggtcgcgcac tacgacgact ggggtgcgtcg gttcagagacc
 3301 gcggtgaccg cgcttcccga gaagcgcgcg gcacagaccg tactgcccgt gctgcacgcg
 3361 ttccgcgctc cgcaggcacc gttgcgcggc gcaccgaac ccacggaggt gttccacgcc
 3421 gcggtgcgca ccgcgaaggt gggcccggga gacatcccgc acctcgacga ggcgctgate
 3481 gacaagtaca tacgcgatct gcgtgagttc ggtctgatct ga

Aminoácidos del orfMSMEG_5739 de *Mycobacterium smegmatis* cepa MC2 155 (YP_889972) (SEQ ID NO: 86)

ES 2 667 469 T3

1 mtsdvhdad gvtetaldde qstrriaely atdpefaaaa plpavvdaah kpglrlaeil
 61 qtlftgygdr palgyrarel atdeggtrvt rllprfdtlt yaqvwsvrqa vaaalrhafa
 121 qpiypgdava tigfaspdyl tldlvcaylg lvsvplqhna pvsrlapila evepriltvs
 181 aeyldlaves vrdvnsvsqf vvfddhpevd dhrdalarar eqlagkgiav ttldaiadeg
 241 aglpaepiyt adhdqrlami lytsgstgap kgamyteamv arlwtmsfit gdptpvinvn
 301 fmplnhlggr ipistavqng gtsyfvped mstlfedlal vrptelglvp rvadmlyqhh
 361 latvdrlvtg gadeltaekq agaelreqvl ggrvitgfvs taplaaemra flditlgahi
 421 vdgygltetg avtrdgvivr ppvidyklid vpelgyfstd kpyprgellv rsqtltpggy
 481 krpevtasvf drdgyyhtgd vmaetapdhl vyvdrrnnvl klaggefvaav anleavfsga
 541 alvrqifvyg nsersfllav vvpptpealeq ydpaalkaal adslqrtard aelqsyevpa
 601 dfivetepfs aangllsgvg kllrpnlkdr ygqrleqmya diaatqanql relrraaatq
 661 pvidtltqaa atilgtgsev asdahftdlg gdsalsaltls nllsdfgfe vpvgtivnpa
 721 tnlalqlaahi eaqrtagdr psfttvhgad ateiraselt ldkfidaetl raapglpkvt
 781 teprtvllsg angwlgfllt lqwlerlapv ggtlitivrg rddaaararl tqaytdpel
 841 srrfaeladr hlrvvagdig dplngltpei whrlaaevdl vvhpaalvnh vlpqrqfpg
 901 nvvgtaevik lalterikpv tylstsvsam gipdfeedgd irtvspvrpl dggyangygn
 961 skwagevllr eahdlcglpv atfrsdmila hprryrgqvnv pdmftrllls llitgvaprs
 1021 fyigdgerpr ahypgltvdf vaeavttlga qqregyvsyd vmpnphddgis ldvfvdlir
 1081 aghpidrvdd yddwvrrfet altalpekrr aqtvplllha frapqaplrq apeptevfha
 1141 avrtakvgpg diphleali dkyirdlref gli

Nucleótidos de Npun02003626 de *Nostoc punctiforme* PCC73102 (ZP_00109192) (SEQ ID NO: 87)

1 atgactcaag cgaaagccaa aaaagaccac ggtgacgttc ctgttaacac ttaccgtccc
 61 aatgctccat ttattggcaa ggtaatatct aatgaacat tagtcaaaga aggtggtatt
 121 ggtattggtc aacaccttaa atttgaccta tctgggtggg atttgaagta tatagaaggt
 181 caaagtattg gcattattcc gccaggttta gacaagaacg gcaagcctga aaaactcaga
 241 ctatattcca tcgctcaac tcgctatggt gatgatgtag atgataagac agtatcactg
 301 tgcgtccgcc agttggagta caagcaccga gaaactggcg aaacagtcta cgggtgttgc
 361 tctacgcacc tgtgtttcct caagccaggg gaagaggtaa aaattacagg gcctgtgggt
 421 aaggaaatgt tgttaccocaa tgaccctgat gctaattgta tcatgatggc tactggaaca
 481 ggtattgccc cgatgcgggc ttaacttggg cgtcagttta aagatgcgga aagagcggct
 541 aaccagaat accaatttaa aggattctct tggcctaata ttggcgtacc tacaactcca
 601 aacctttat ataaggaaga actggaagag attcaacaaa aatctcctga gaacttccgc
 661 ctaactgctg ccatcagccg cgaacagaaa aatccccaa gcggtagaat gtatattcaa
 721 gaccgcgtag cagaacatgc tgatgaattg tggcagttga ttaaaaatga aaaaaccac
 781 acttacattt gcggtttgcg cggatgga gaaggtattg atgcagcctt aactgctgct
 841 gctgctaagg aaggcgtaac ctggagtgat taccagaagc aactcaagaa agccggtcgc
 901 tggcacgtag aaacttacta a

5 Aminoácidos de Npun02003626 de *Nostoc punctiforme* PCC73102 (ZP_00109192) (SEQ ID NO: 88)

1 mynqgavega anielgsrif vyevvlgrrg eetdqtnypi rksgsvfirv pynrmnqemr
 61 ritrlggtiv siqpitalep vngkasfgna tsvvselaks getansegng katpvnahsa
 121 eeqnkdkkgn tmtqakakkd hgdvpvntyr pnapfigkvi sneplvkegg igivqhlkfd
 181 lsggdlykie gqsigiippg ldknkpkel rlysiasthr gddvddktsv lcvrqlkykh
 241 petgetvygv csthlcflkp geevkitgpv gkemllpndp danvimmatg tgiapmrayl
 301 wrqfkdaera anpeyqfkgf swlifgvptt pnlllykeele eiqqkypenf rltaaaisreq
 361 knpqqgrmyi qdrvaehade lwqliknekt htyicglrgm eegidaalta aaakegvtws
 421 dyqkqlkkag rwhvety

Nucleótidos de Npun02001001 de *Nostoc punctiforme* PCC73102 (ZP_00111633) (SEQ ID NO: 89)

1 atgccaactt ataaagtgac actaattaac gaggctgaag ggctgaacac aacccttgat
 61 gttgaggacg atacctatat tctagacgca gctgaagaag ctggtattga cctgccctac
 121 tcttgccgcg ctggtgcttg ctctacttgt gcaggtaaac tcgtatcagg taccgtcgat
 181 caaggcgatc aatcattctt agatgacgat caaatagaag ctggatatgt actgacctgt
 241 gttgcttacc caacttctaa tgtcacgatc gaaactcaca aagaagaaga actctattaa

ES 2 667 469 T3

Aminoácidos de Npun02001001 de *Nostoc punctiforme* PCC73102 (ZP_00111633) (SEQ ID NO: 90)

1 mptykvtlin eaeglnttld veddtyilda aeeagidlpy scragacstc agklvsgtvd
61 qgdqsflddd qieagyvltc vayptsnvti ethkeeely

Nucleótidos de Npun02003530 de *Nostoc punctiforme* PCC73102 (ZP_00109422) (SEQ ID NO: 91)

1 atgtcccgta catacacaat taaagttcgc gatcgcgcca ctggcaaaac acacacccta
61 aaagtgccag aagaccgta ttcctgcac actgccgaaa aacaagggtg ggaactaccg
121 ttttcctgtc gcaacggagc ttgcaccgct tgtgctgtga ggtattgtc aggagaaatt
181 tatcaaccag aggcgatcgg attgtcacca gatttacgct agcaaggta tgccctgttg
241 tgtgtgagtt atccccgttc tgacttgaa gtagagacac aagacgaaga tgaagtctac
301 gaactccagt ttgggcgcta ttttgctaag gggaaagta aagcggggtt accgttagat
361 gaggaataa

5 Aminoácidos de Npun02003530 de *Nostoc punctiforme* PCC73102 (ZP_00109422) (SEQ ID NO: 92)

1 msrtytikvr dratgkthtl kvpedryilh taekqgvlp fscrngacta cavrvtlgei
61 yqpeaiglspl dlrqqgyall cvsyrpdle vetqdedevy elqfgryfak gkvkaglpd
121 ee

Nucleótidos de Npun02003123 de *Nostoc punctiforme* PCC73102 (ZP_00109501) (SEQ ID NO: 93)

1 atgccccaaa cttacaccgt agaaatcgat catcaaggca aaattcatac cttgcaagtt
61 cctgaaaatg aaacgatctt atcagttgcc gatgctgctg gtttggaact gccgagttct
121 tgtaatgcag gtgtttgac aacttgccc ggtcaaataa gccagggaac tgtggatcaa
181 actgatggca tgggcgtag tccagattta caaaagcaag gttacgtatt gctttgtgtt
241 gcgaaacccc tttctgattt gaaacttgaa acagaaaagg aagacatagt ttatcagtta
301 caatttgca aagacaaata a

Aminoácidos de Npun02003123 de *Nostoc punctiforme* PCC73102 (ZP_00109501) (SEQ ID NO: 94)

10 1 mpktytveid hqgkihtlqv penetilsva daaglelpss cnagvcttca gqisqgtvdq
61 tdmgvvspdl qkqgyvllcv akplsdikle tekedivyql qfgkdk

REIVINDICACIONES

1. Método de producción de un alcano o alqueno a partir de una célula huésped modificada genéticamente, comprendiendo el método:
 - (a) transformar una célula huésped para que exprese introduciendo
 - 5 (i) un polinucleótido de biosíntesis de aldehídos que codifica para un polipéptido de biosíntesis de aldehídos, y
 - (ii) un polinucleótido de biosíntesis de alcanos o alquenos que codifica para un polipéptido de biosíntesis de alcanos o alquenos; y
 - 10 (b) cultivar dicha célula huésped modificada genéticamente en un medio de cultivo que contiene una fuente de carbono en condiciones eficaces para producir un alcano o alqueno en el medio de cultivo,

en el que dicho polipéptido de biosíntesis de aldehídos tiene actividad acil-ACP reductasa y procede de una cianobacteria, y en el que dicho polipéptido de biosíntesis de alcanos o alquenos tiene actividad descarboxilasa y procede de una cianobacteria.
- 15 2. Método según la reivindicación 1, en el que dicho polipéptido de biosíntesis de alcanos o alquenos comprende la secuencia de aminoácidos de
 - (i) SEQ ID NO: 37 o SEQ ID NO: 38 o SEQ ID NO: 39, o
 - (ii) SEQ ID NO: 40 y una cualquiera de (a) SEQ ID NO: 37 o SEQ ID NO: 38 o SEQ ID NO: 39, o
 - (iii) SEQ ID NO: 41 o SEQ ID NO: 42 o SEQ ID NO: 43 o SEQ ID NO: 44.
- 20 3. Método según la reivindicación 1 ó 2, en el que la célula huésped se selecciona del grupo que consiste en una célula de levadura, célula de hongo, célula de hongos filamentosos, célula de algas y célula bacteriana.
4. Método según la reivindicación 3, en el que la célula huésped es una célula de *E. coli*, preferiblemente una célula de *E. coli* de cepa B, una de cepa C, una de cepa K o una de cepa W.
5. Método según una cualquiera de las reivindicaciones 1-4, en el que el alcano o alqueno se secreta por la célula huésped.
- 25 6. Método según una cualquiera de las reivindicaciones 1-5, en el que el alcano o alqueno es un alcano o alqueno C₁₃-C₂₁.
7. Método según una cualquiera de las reivindicaciones 1-6, en el que el alcano o alqueno se selecciona del grupo que consiste en tridecano, metiltridecano, nonadecano, metilnonadecano, heptadecano, metilheptadecano, pentadecano, metilpentadecano, pentadeceno, heptadeceno, metilpentadeceno y metilheptadeceno.
- 30 8. Método según una cualquiera de las reivindicaciones 1-7, que comprende además cultivar la célula huésped en presencia de al menos un sustrato biológico para el polipéptido o para un polipéptido codificado por la secuencia de nucleótidos.
9. Cultivo celular que comprende una célula huésped modificada genéticamente, en el que
 - 35 (a) un polinucleótido de biosíntesis de aldehídos que codifica para un polipéptido de biosíntesis de aldehídos; y
 - (b) un polinucleótido de biosíntesis de alcanos o alquenos que codifica para un polipéptido de biosíntesis de alcanos o alquenos

se incorporan y se sobreexpresan en dicha célula huésped modificada genéticamente y se produce un alcano o alqueno en el cultivo celular cuando dicha célula huésped modificada genéticamente se cultiva en un medio de cultivo que contiene una fuente de carbono en condiciones eficaces para expresar dicho polipéptido de biosíntesis de aldehídos y dicho polipéptido de biosíntesis de alcanos o alquenos,

en el que dicho polipéptido de biosíntesis de aldehídos tiene actividad acil-ACP reductasa y procede de una cianobacteria, y en el que dicho polipéptido de biosíntesis de alcanos o alquenos tiene actividad descarboxilasa y procede de una cianobacteria.
- 45 10. Cultivo celular según la reivindicación 9, en el que dicho polipéptido de biosíntesis de alcanos o alquenos comprende la secuencia de aminoácidos de

(i) SEQ ID NO: 37 o SEQ ID NO: 38 o SEQ ID NO: 39, o

(ii) SEQ ID NO: 40 y una cualquiera de (a) SEQ ID NO: 37 o SEQ ID NO: 38 o SEQ ID NO: 39, o

(iii) SEQ ID NO: 41 o SEQ ID NO: 42 o SEQ ID NO: 43 o SEQ ID NO: 44.

- 5 11. Cultivo celular según la reivindicación 9 ó 10, en el que la célula huésped modificada genéticamente se selecciona del grupo que consiste en una célula de levadura, célula de hongo, célula de hongos filamentosos, célula de algas y célula bacteriana.
12. Cultivo celular según una cualquiera de las reivindicaciones 9 a 11, en el que el alcano o alqueno es un alcano o alqueno C₁₃-C₂₁.
- 10 13. Cultivo celular según una cualquiera de las reivindicaciones 9 a 12, en el que el alcano o alqueno se selecciona del grupo que consiste en tridecano, metiltridecano, nonadecano, metilnonadecano, heptadecano, metilheptadecano, pentadecano, metilpentadecano, pentadeceno, heptadeceno, metilpentadeceno y metilheptadeceno.
14. Cultivo celular según una cualquiera de las reivindicaciones 9 a 13, que comprende un alcano o alqueno que tiene un $\delta^{13}\text{C}$ de aproximadamente -15,4 o mayor.
- 15 15. Cultivo celular según una cualquiera de las reivindicaciones 9 a 14, que comprende un alcano o alqueno que tiene una $f_M^{14}\text{C}$ de al menos aproximadamente 1,003.
16. Célula huésped modificada genéticamente, en la que
- (a) un polinucleótido de biosíntesis de aldehídos que codifica para un polipéptido de biosíntesis de aldehídos; y
- 20 (b) un polinucleótido de biosíntesis de alcanos o alquenos que codifica para un polipéptido de biosíntesis de alcanos o alquenos
- 25 se incorporan en dicha célula huésped modificada genéticamente y se produce un alcano o alqueno en la célula cuando dicha célula huésped modificada genéticamente se cultiva en un medio de cultivo que contiene una fuente de carbono en condiciones eficaces para expresar dicho polipéptido de biosíntesis de aldehídos y dicho polipéptido de biosíntesis de alcanos o alquenos,
- en la que dicho polipéptido de biosíntesis de aldehídos tiene actividad acil-ACP reductasa y procede de una cianobacteria, y en la que dicho polipéptido de biosíntesis de alcanos o alquenos tiene actividad descarboxilasa y procede de una cianobacteria.

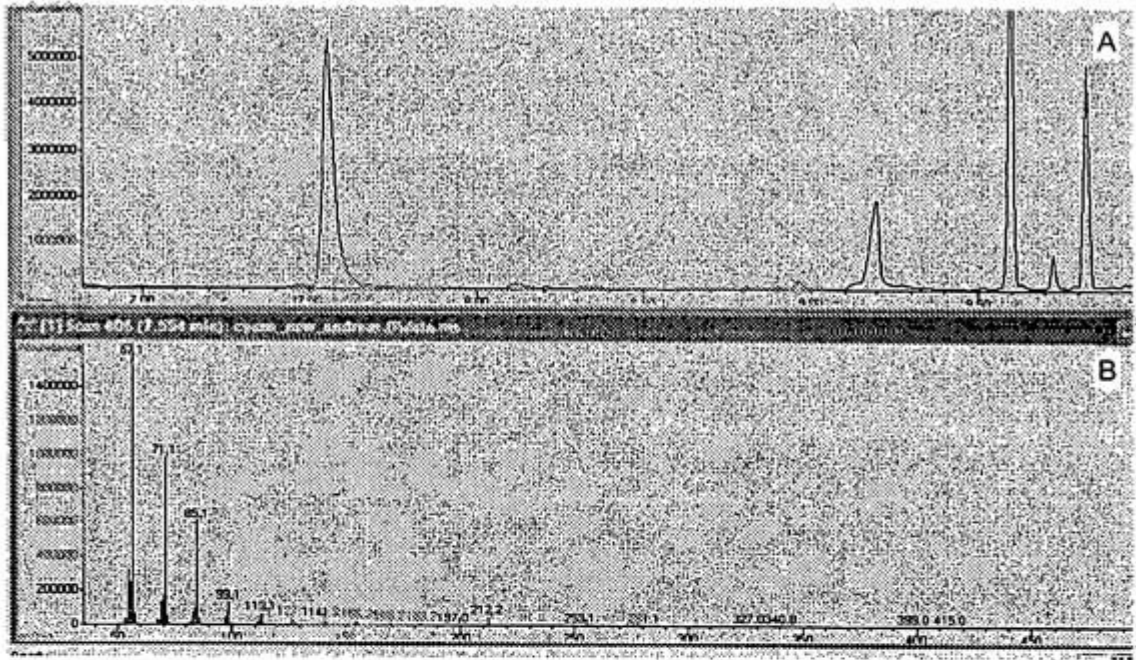


FIG. 1

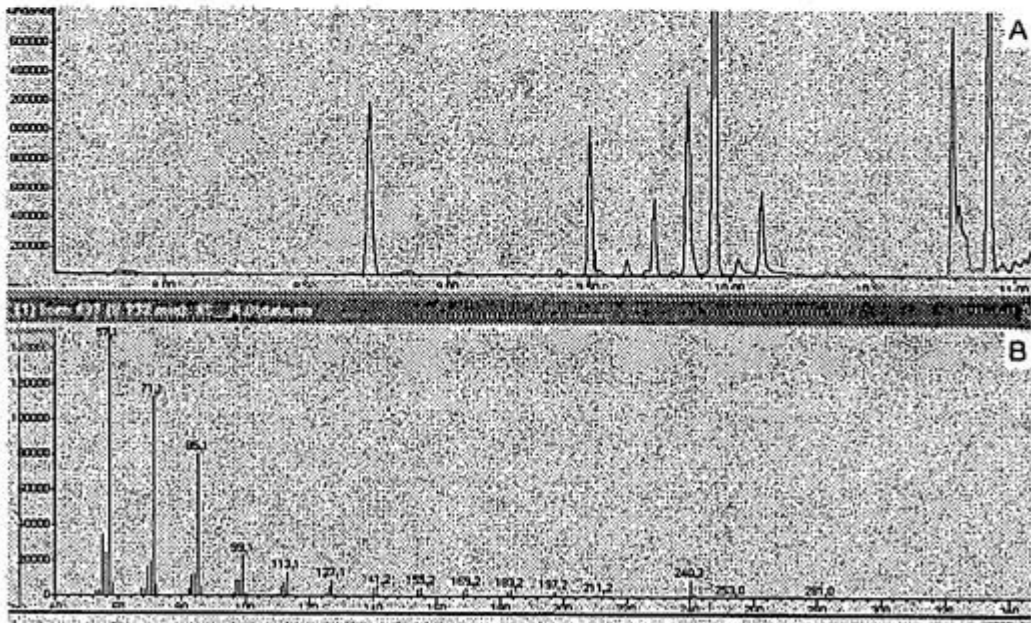
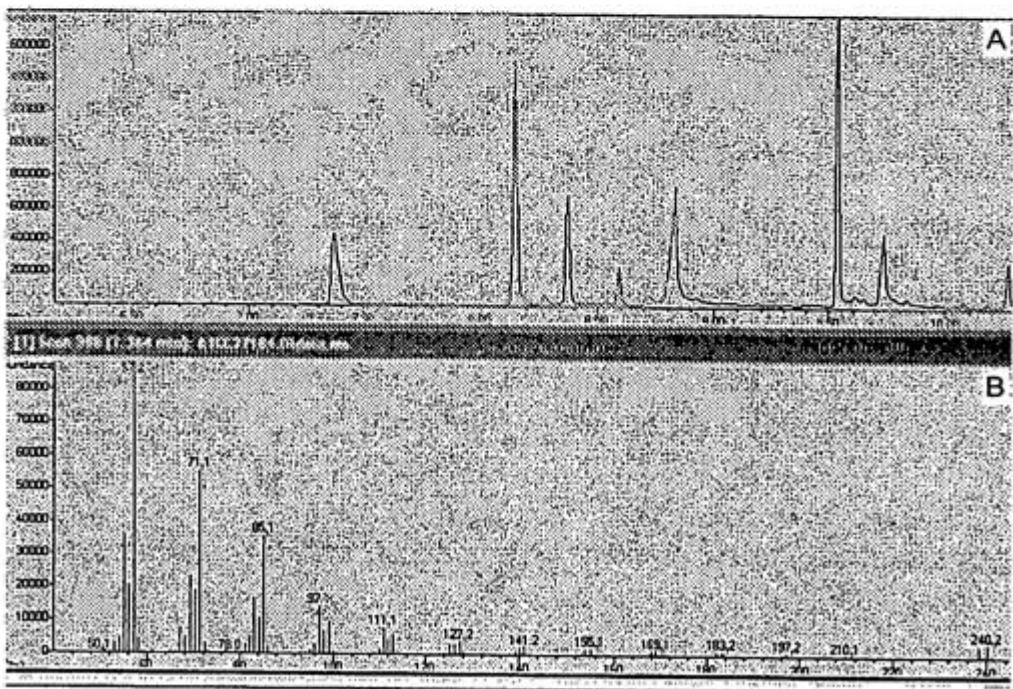
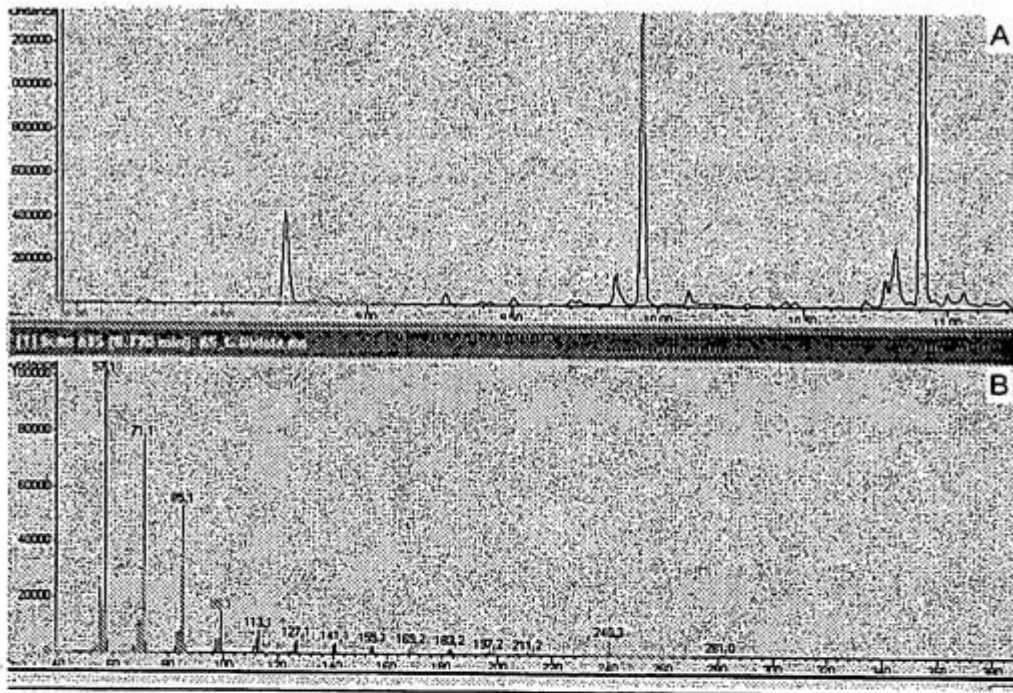


FIG. 2



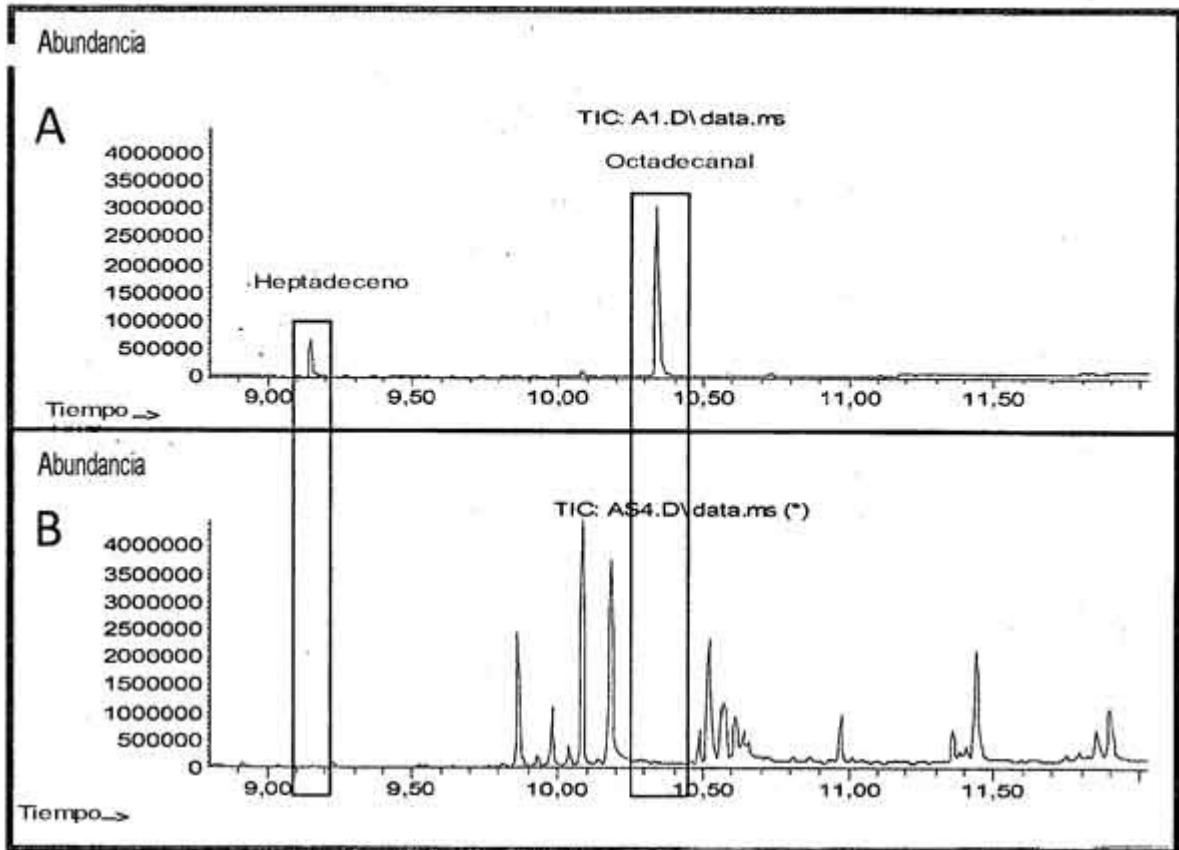


FIG. 5

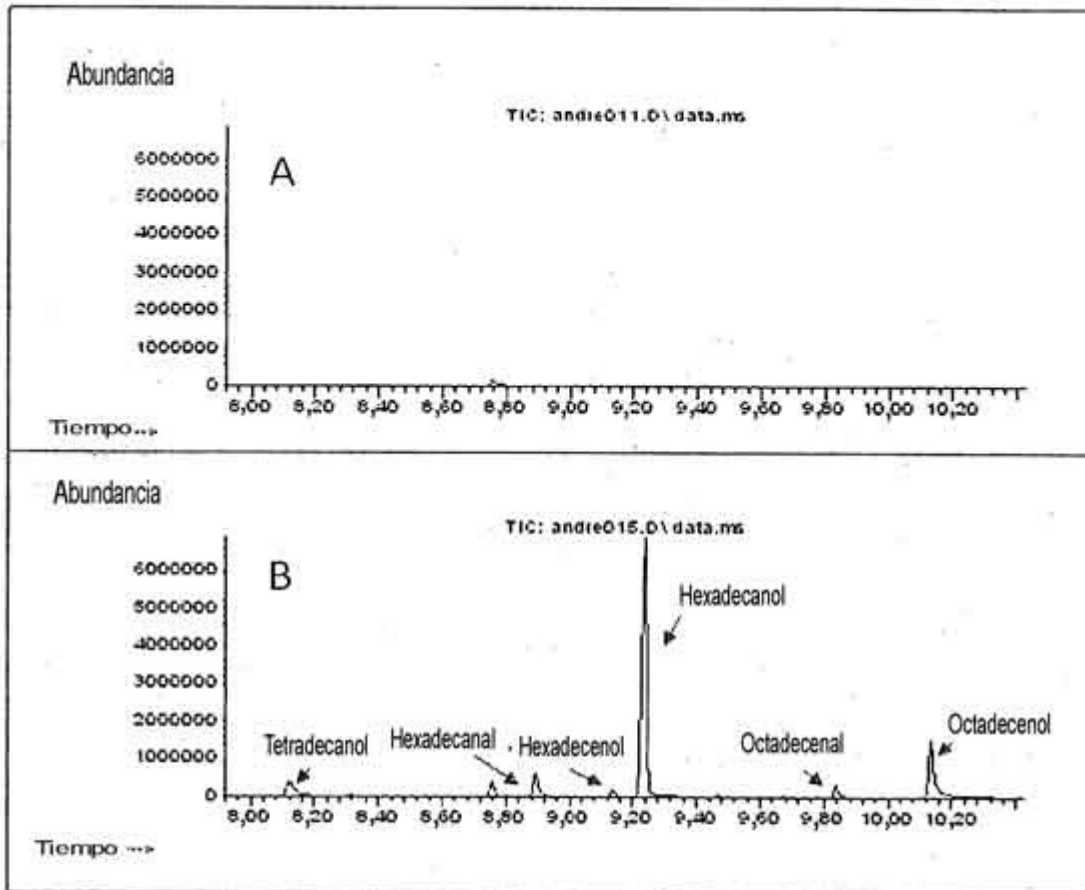


FIG. 6

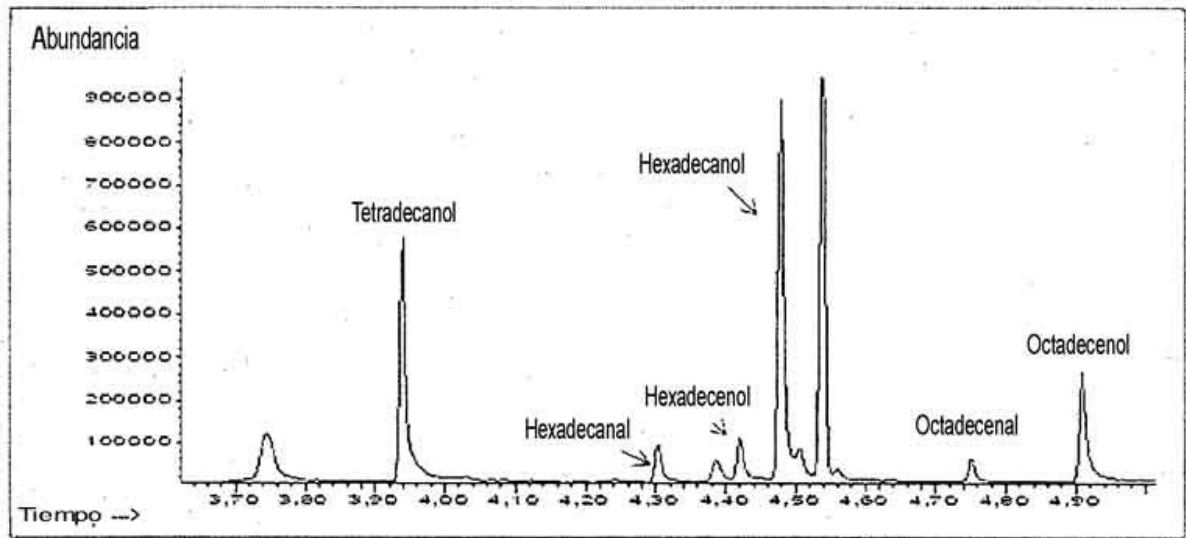


FIG. 7

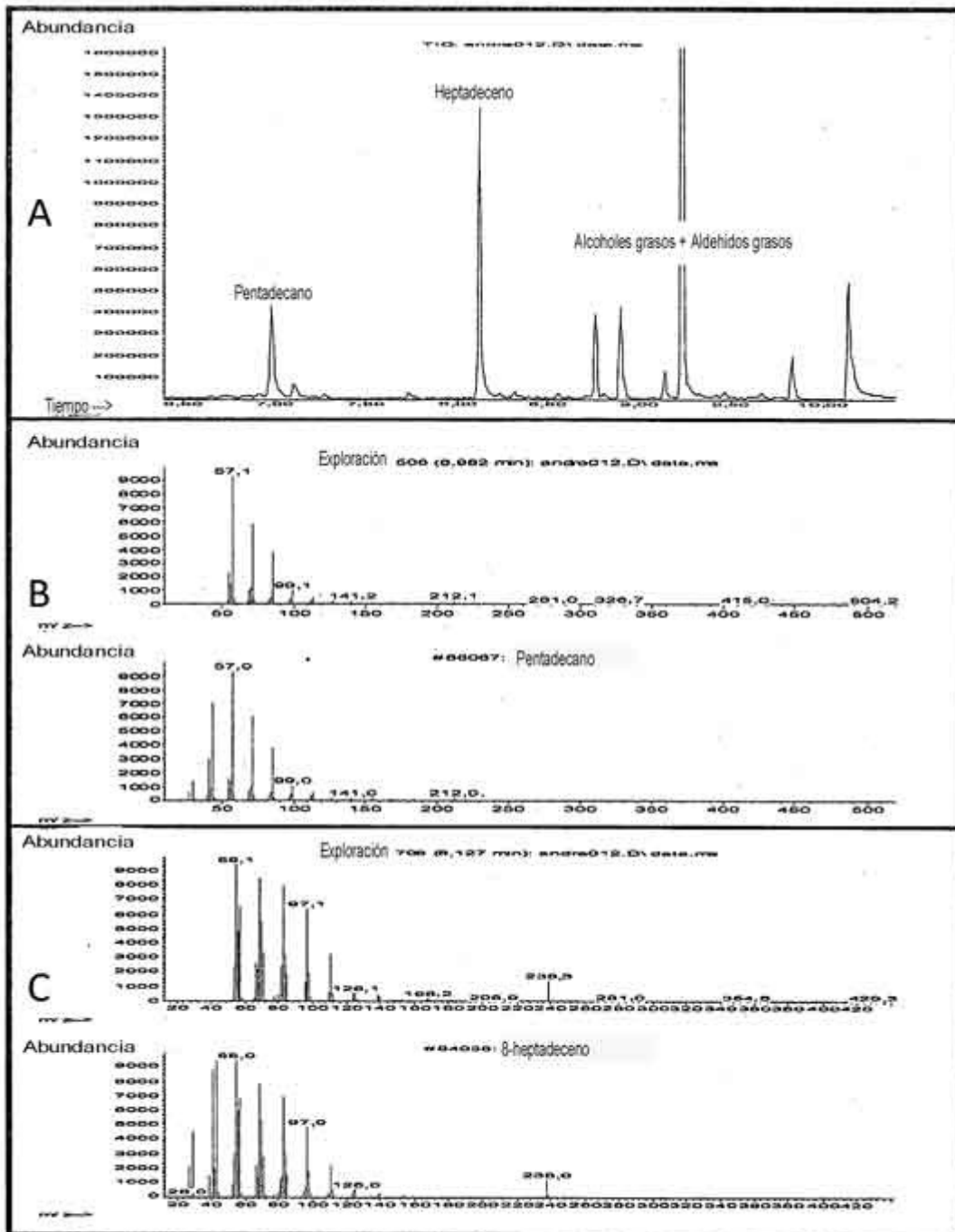


FIG. 8

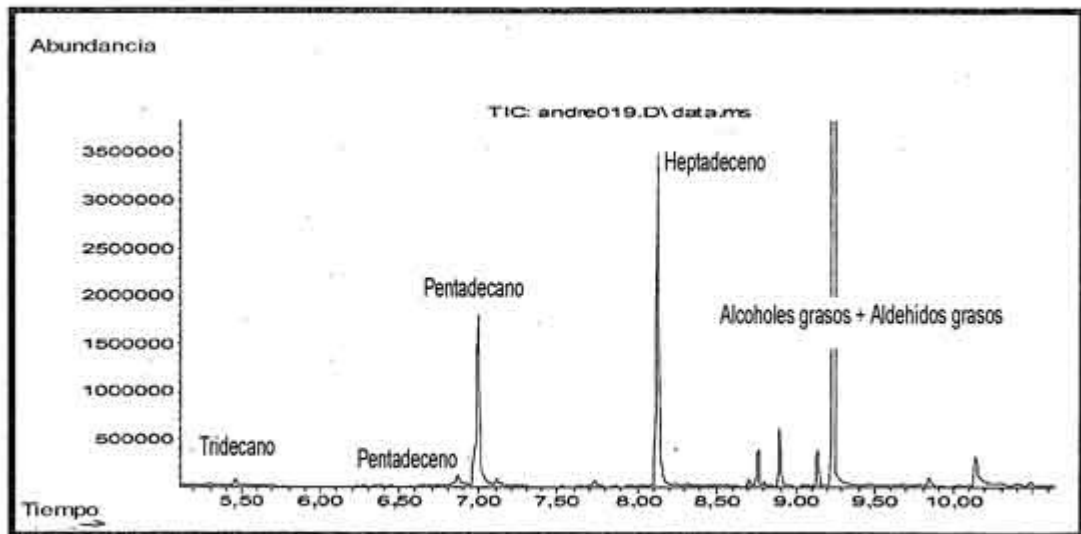


FIG. 9

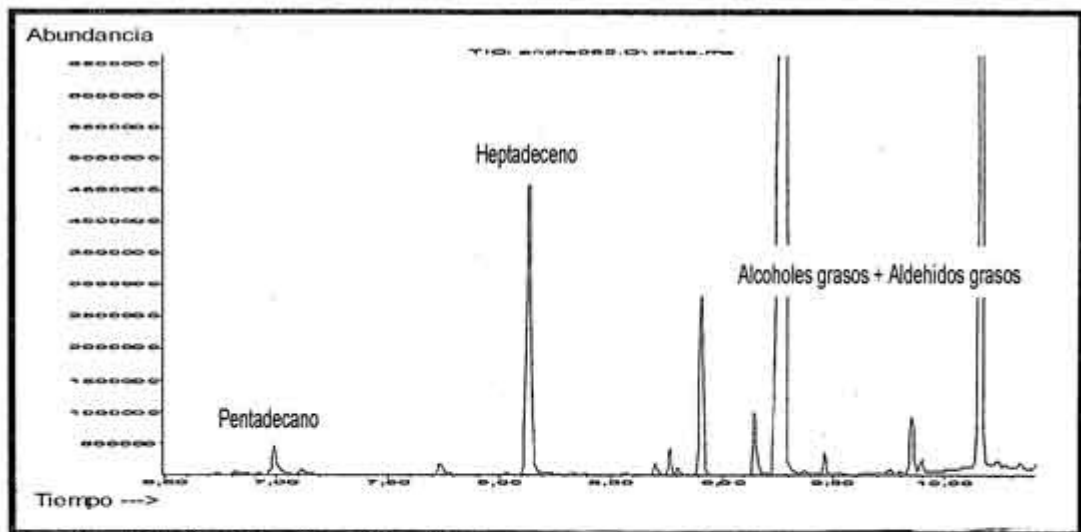


FIG. 10

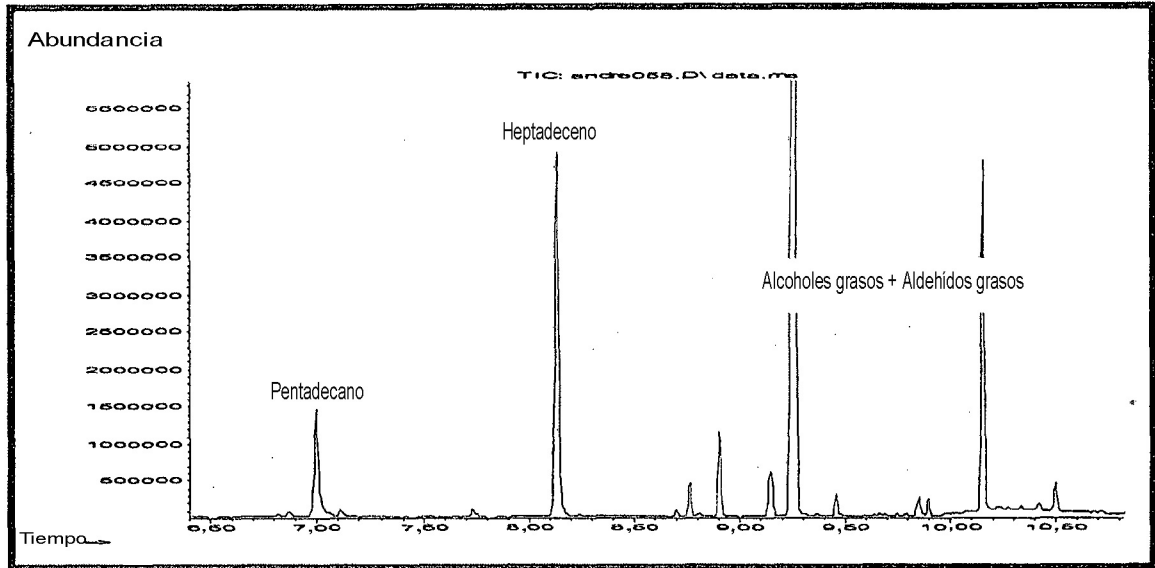


FIG. 11

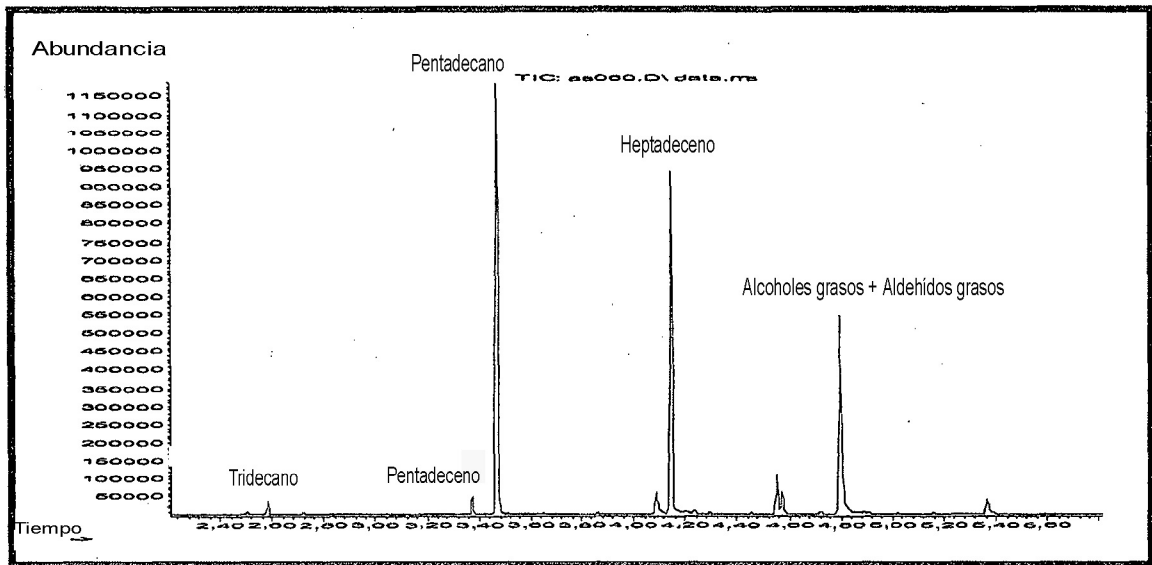


FIG. 12

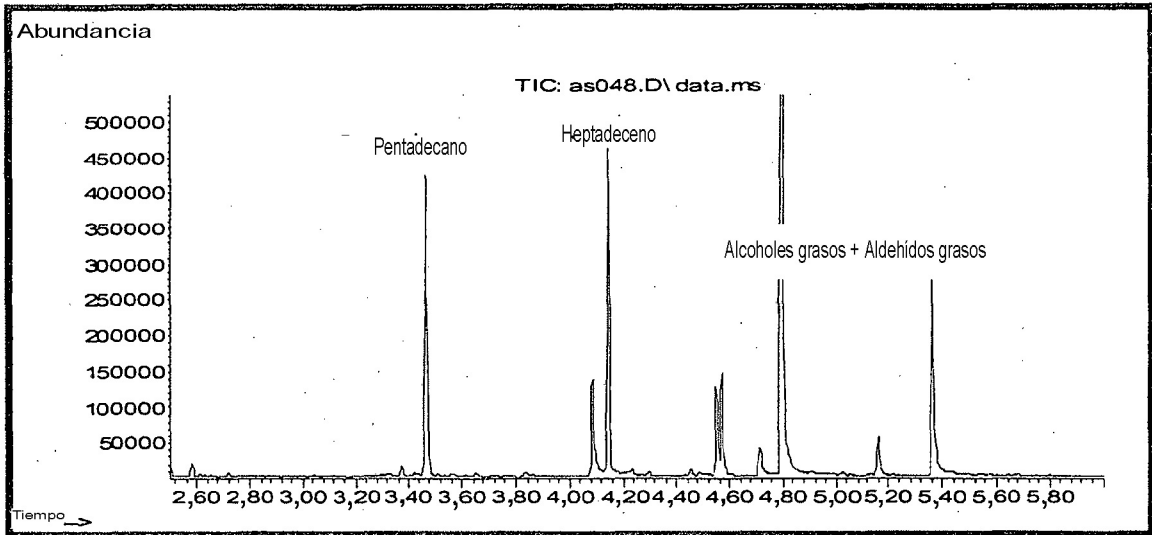


FIG. 13

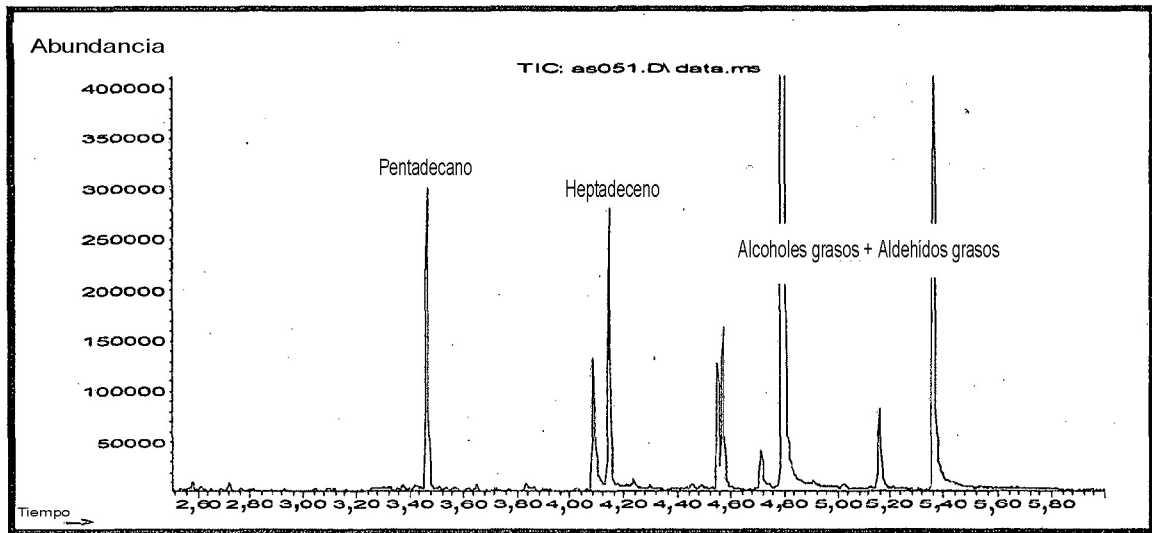


FIG. 14

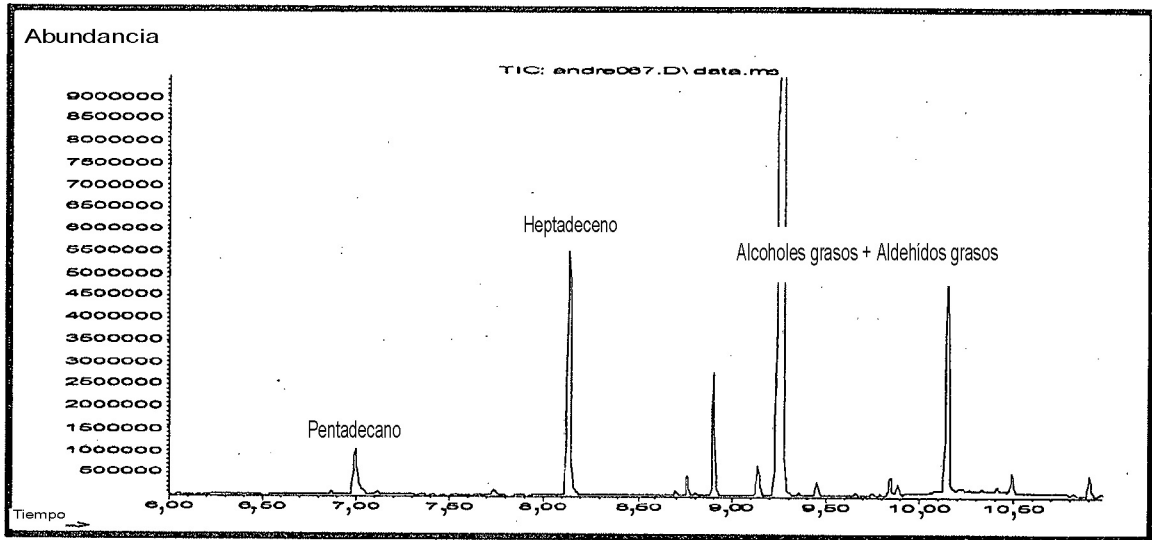


FIG. 15

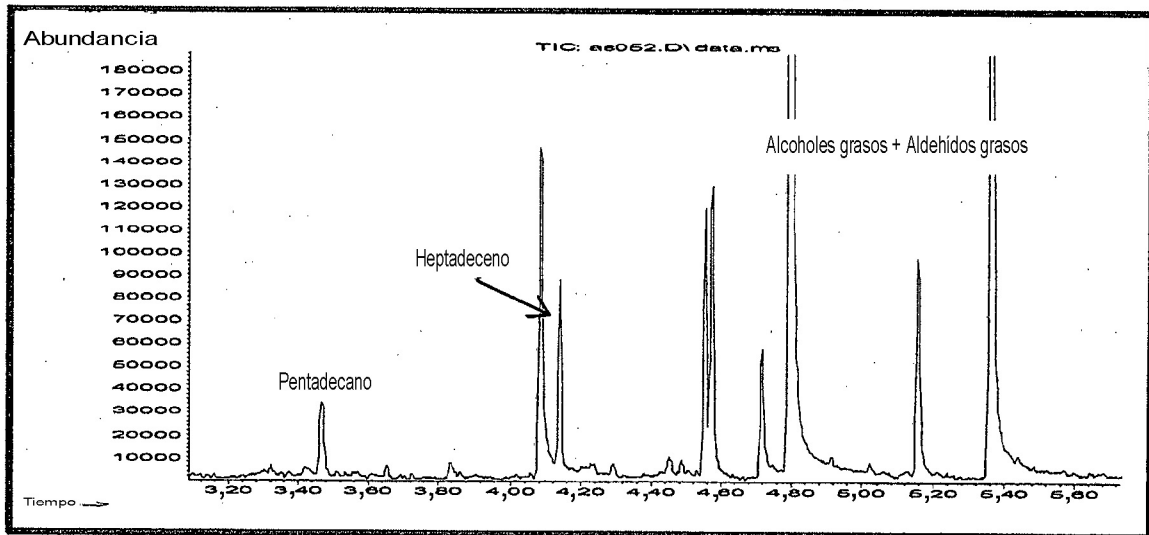


FIG. 16

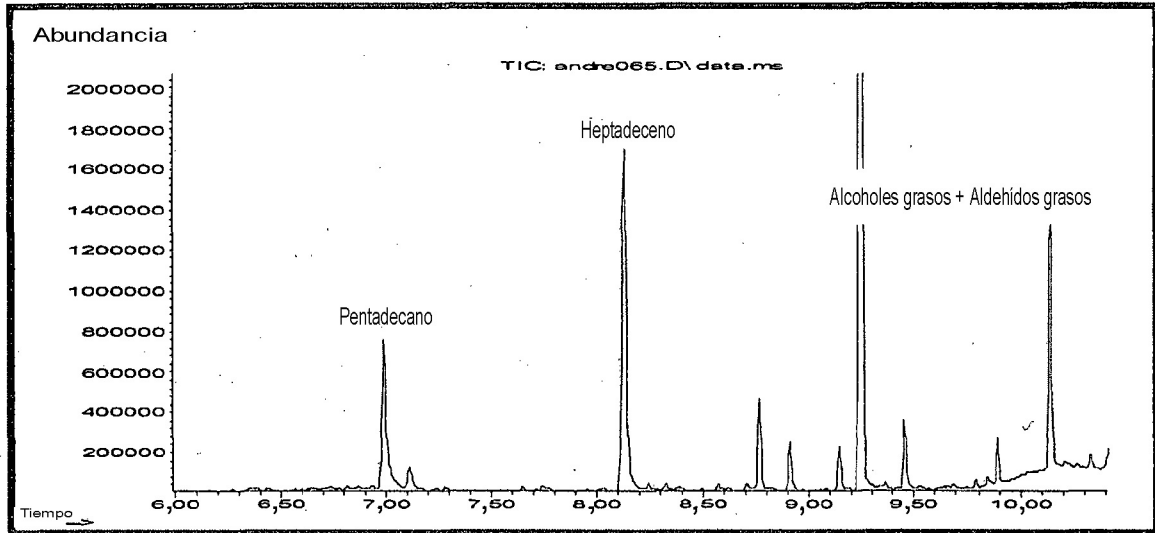


FIG. 17

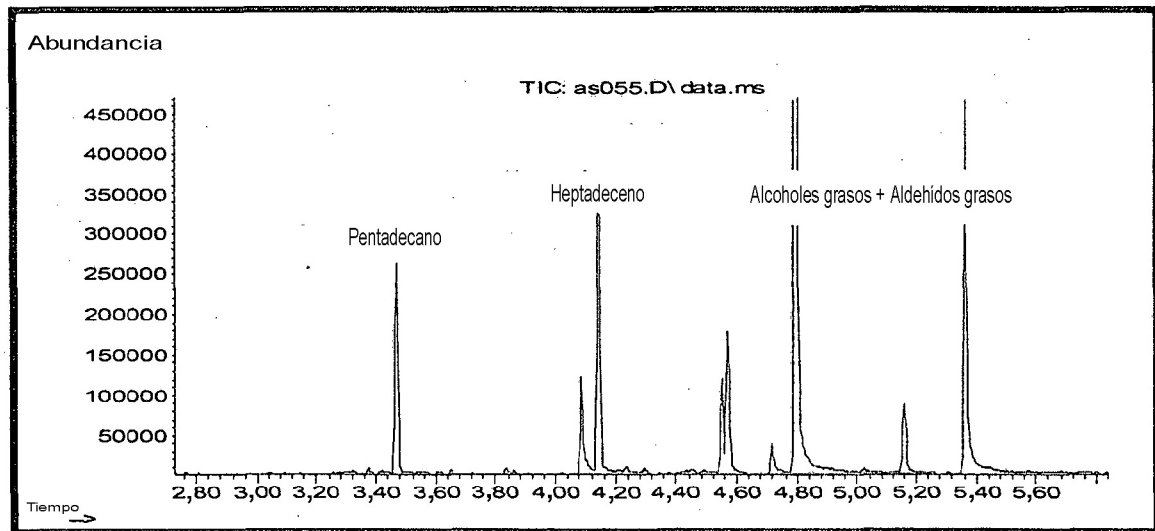


FIG. 18

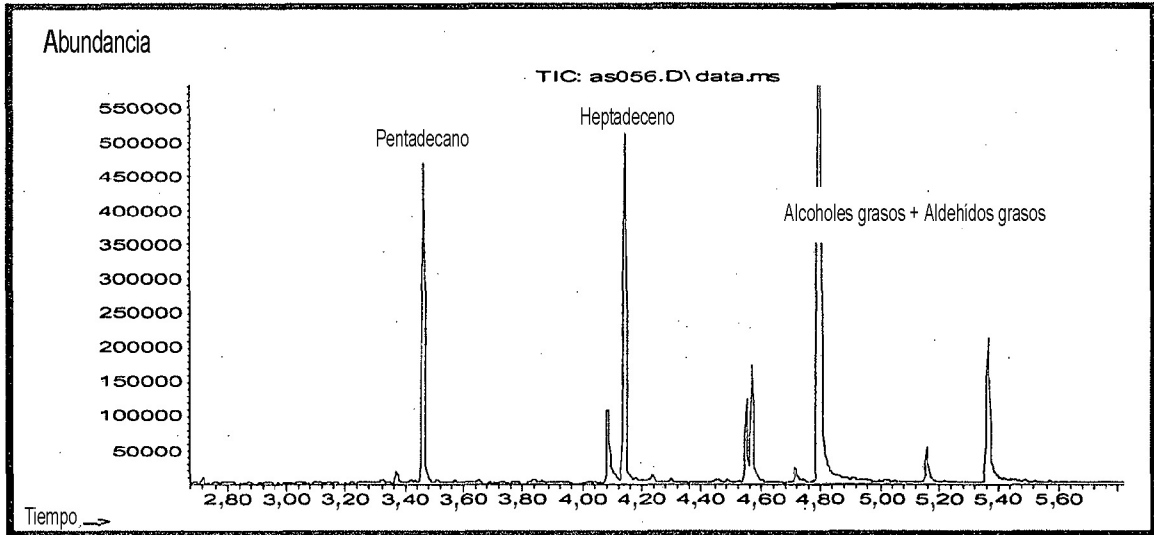


FIG. 19

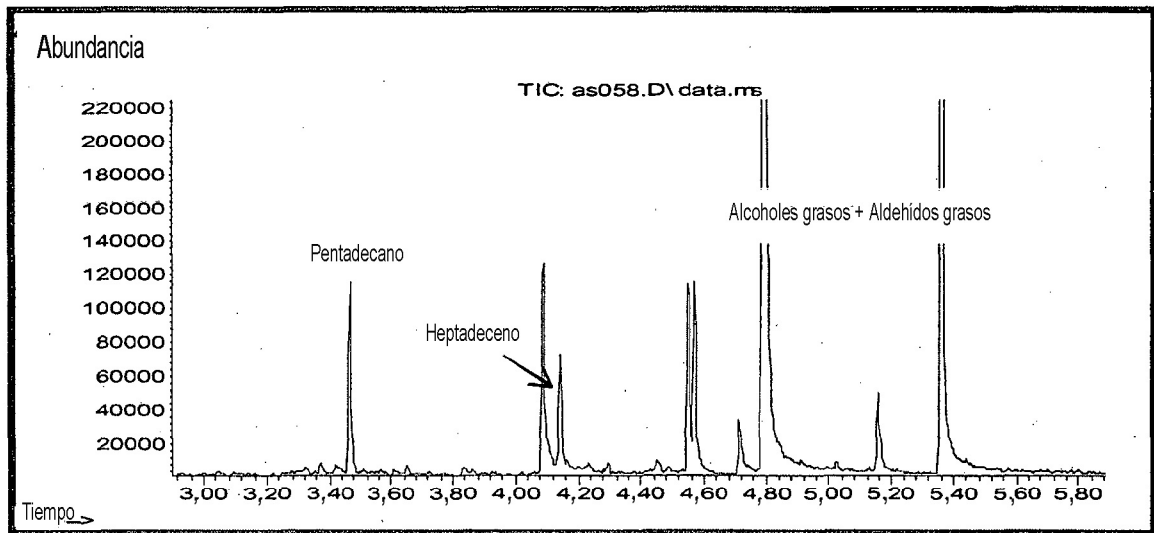


FIG. 20

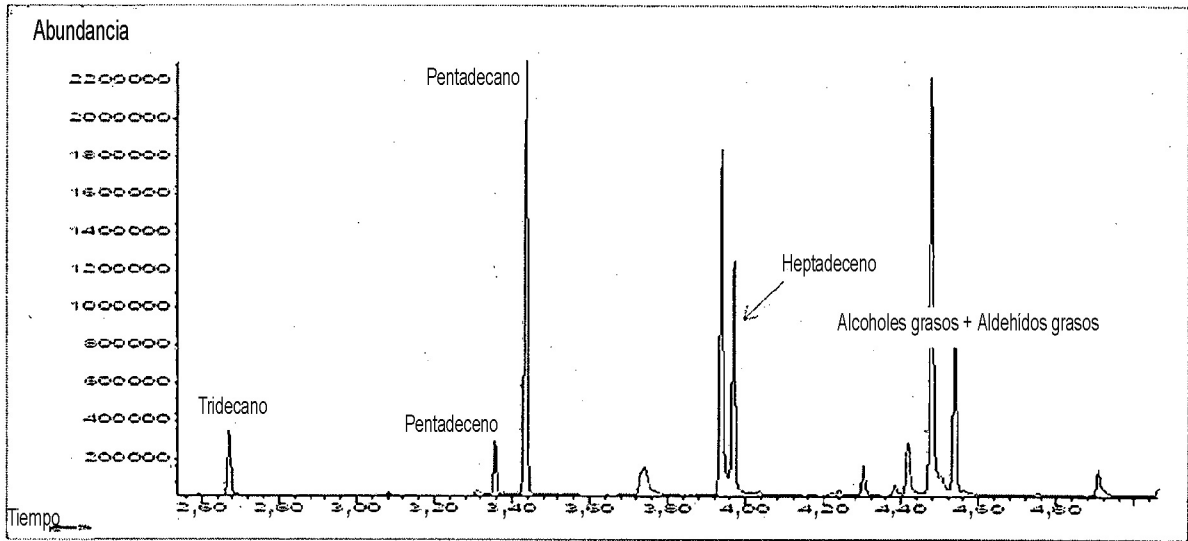


FIG. 21

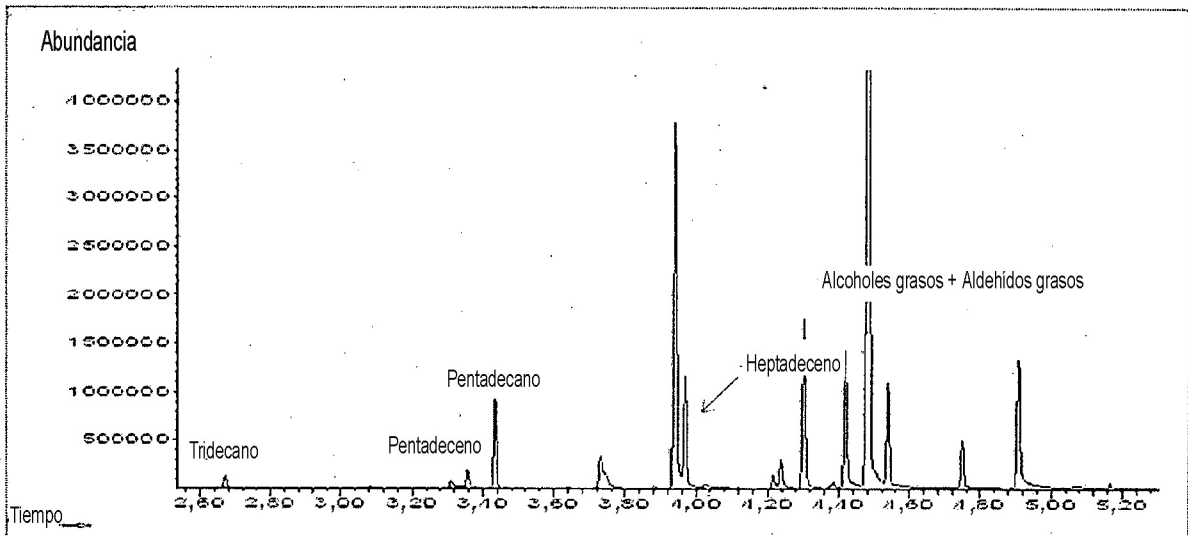


FIG. 22

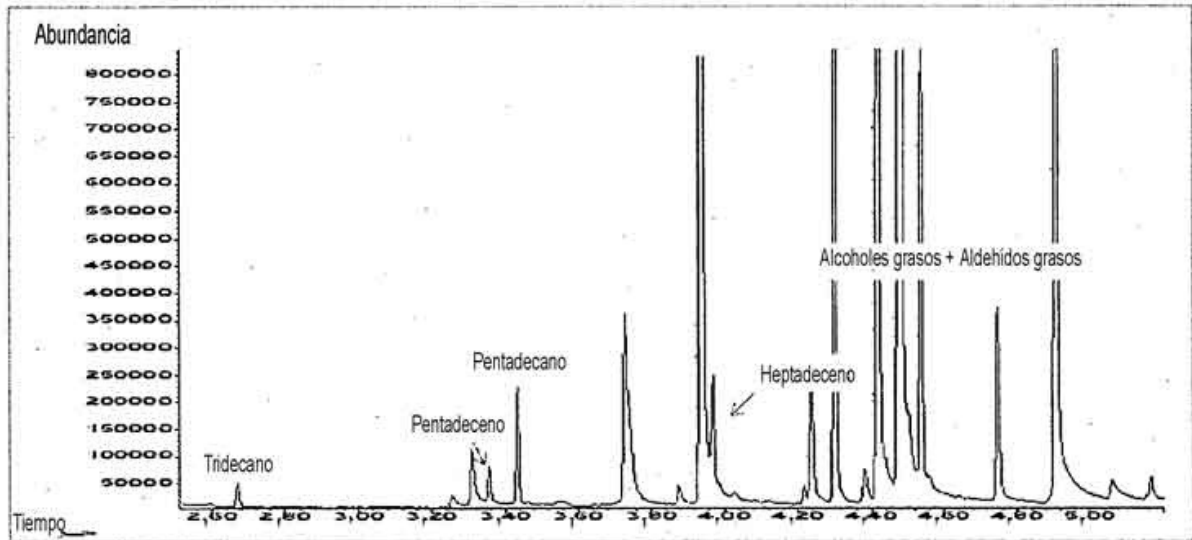


FIG. 23

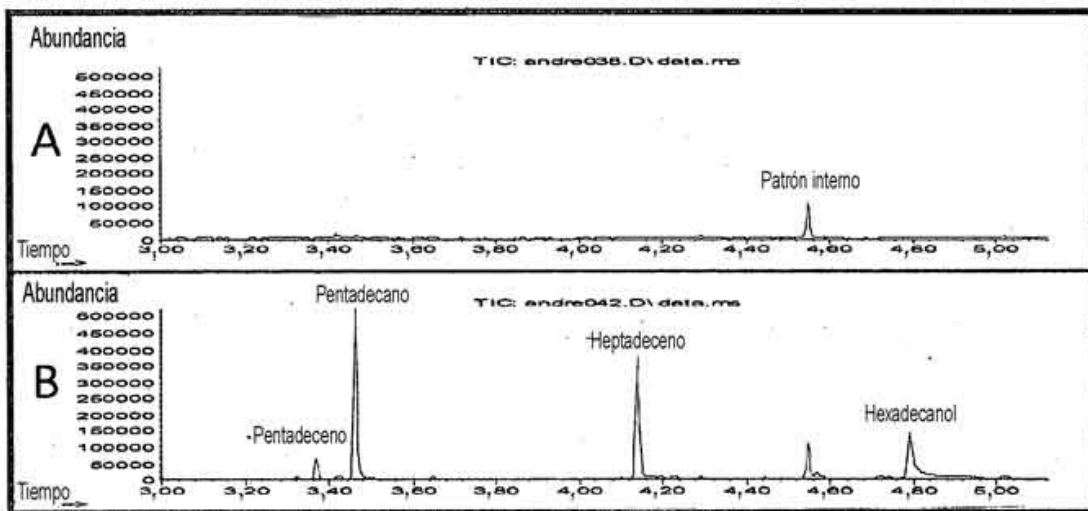


FIG. 24

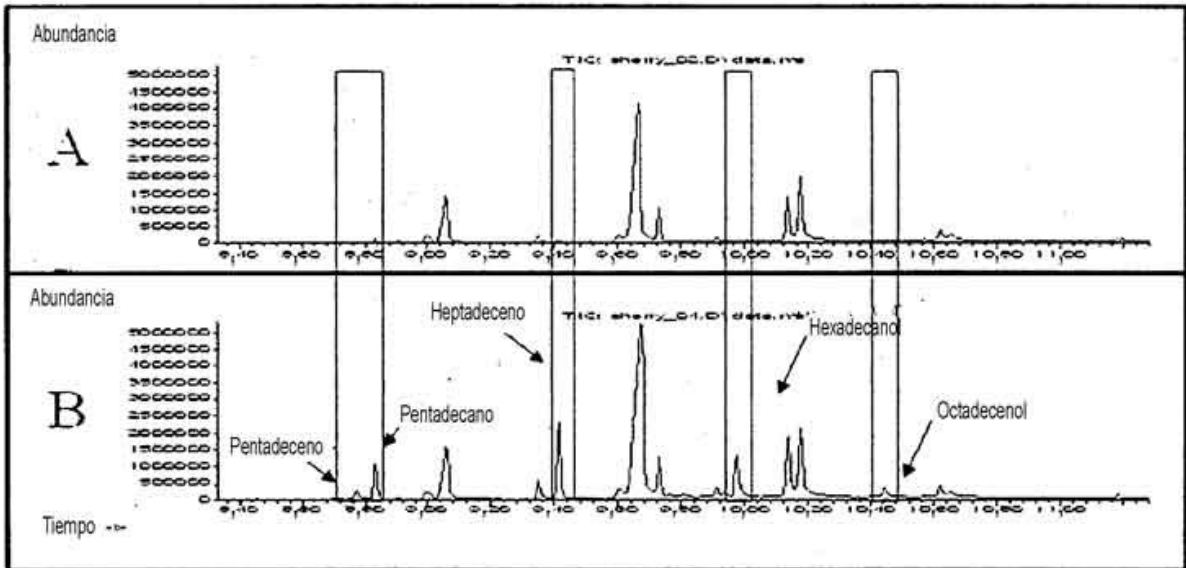


FIG. 25

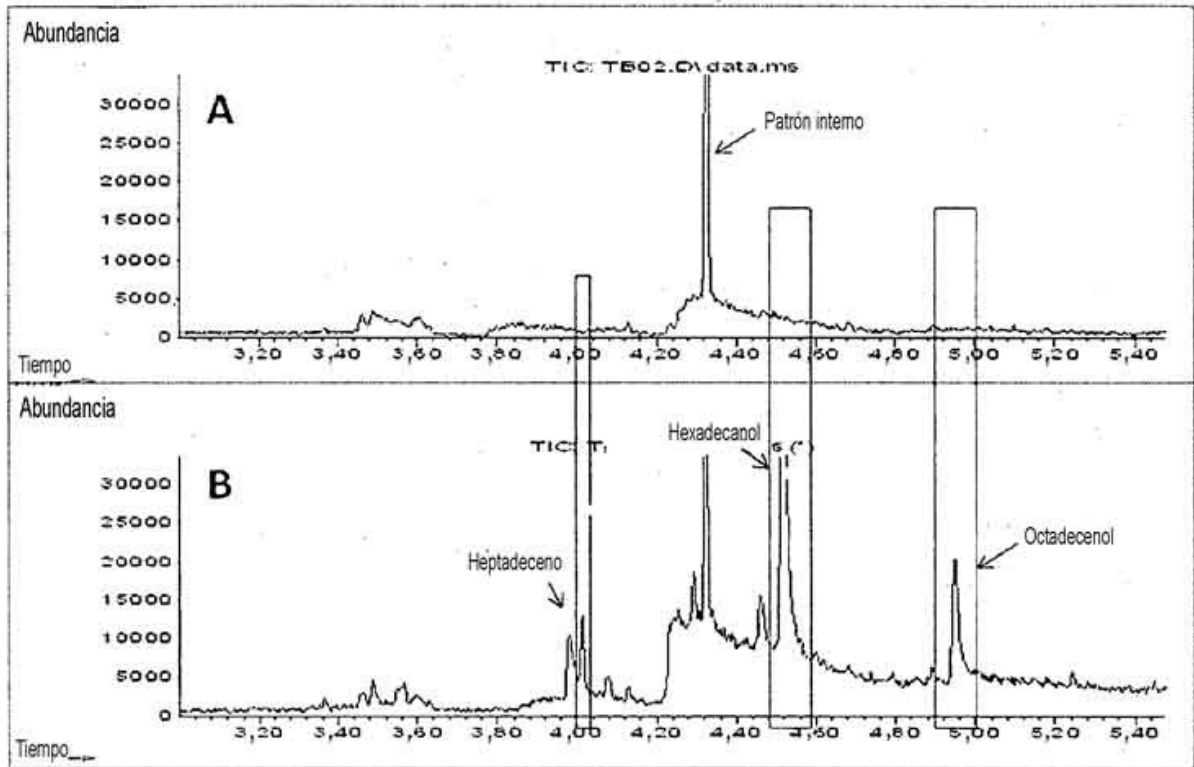


FIG. 26

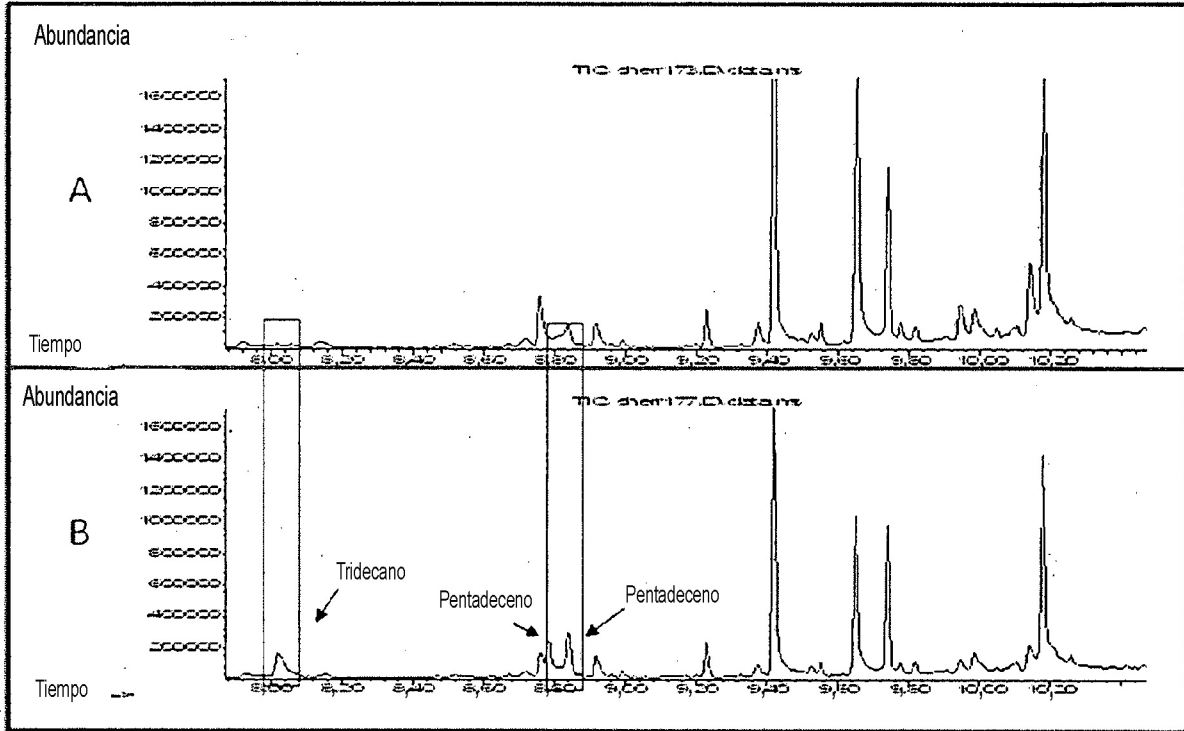


FIG. 27

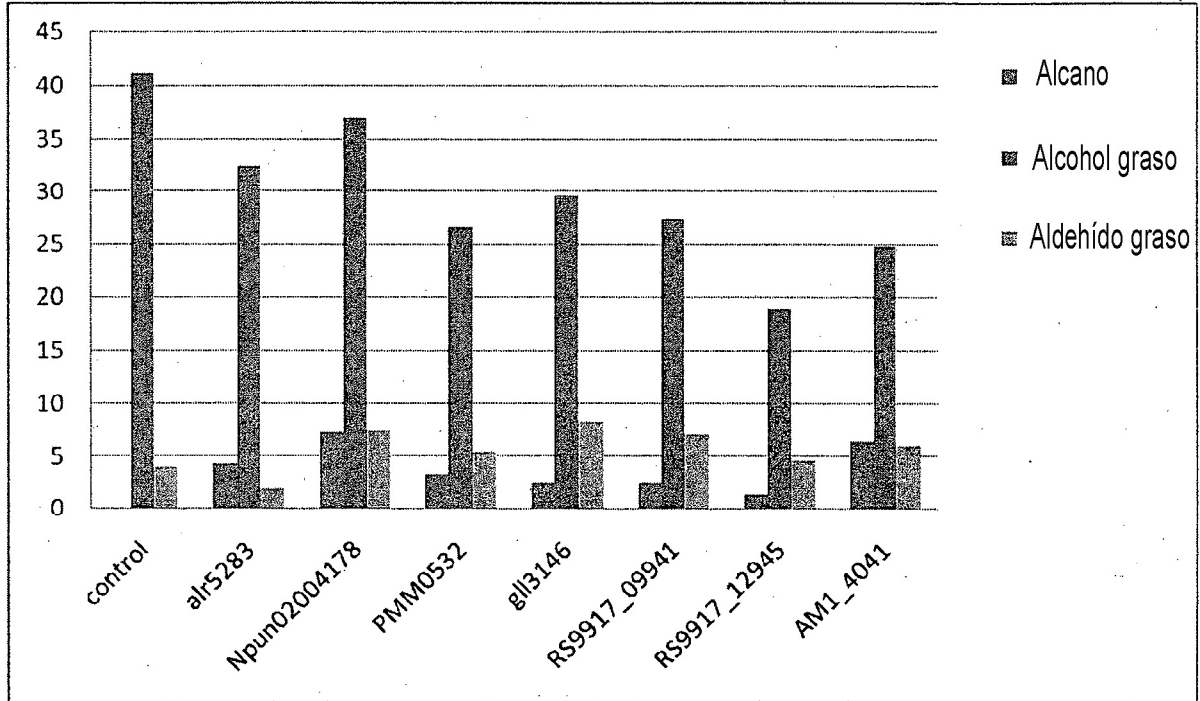


FIG. 28

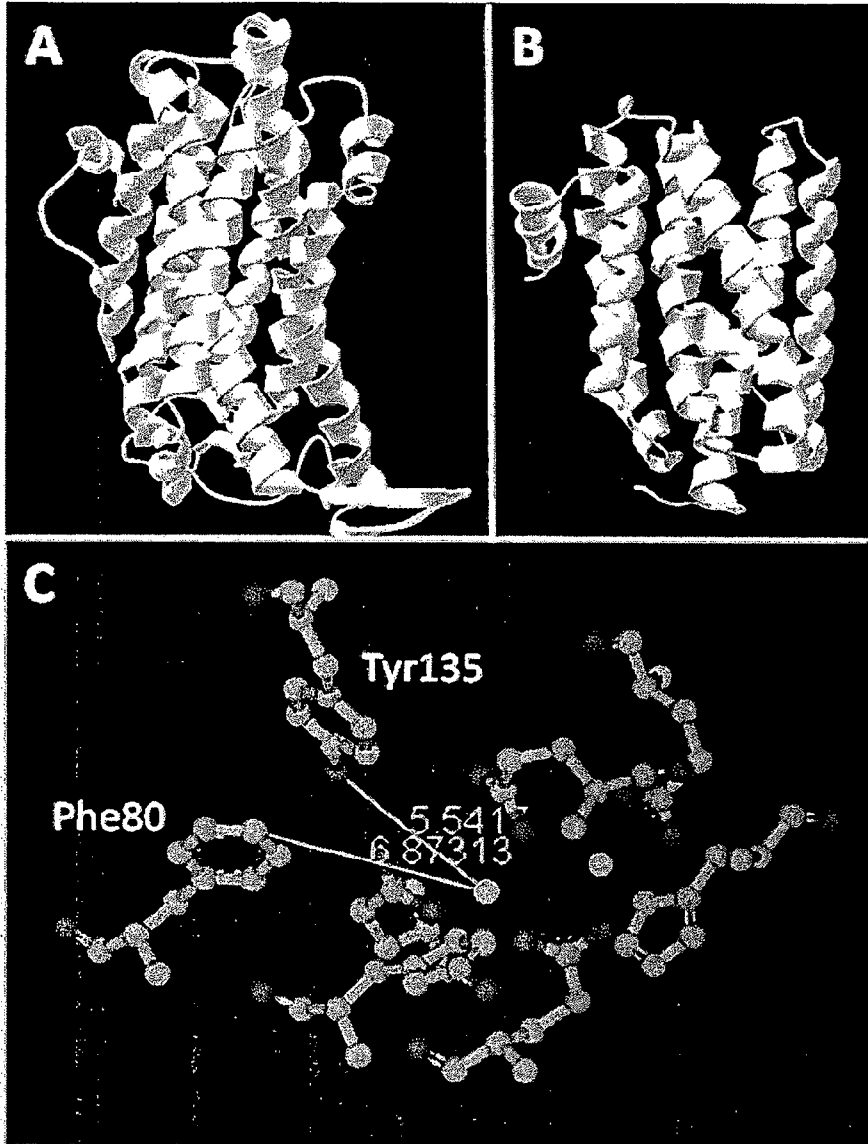


FIG. 29

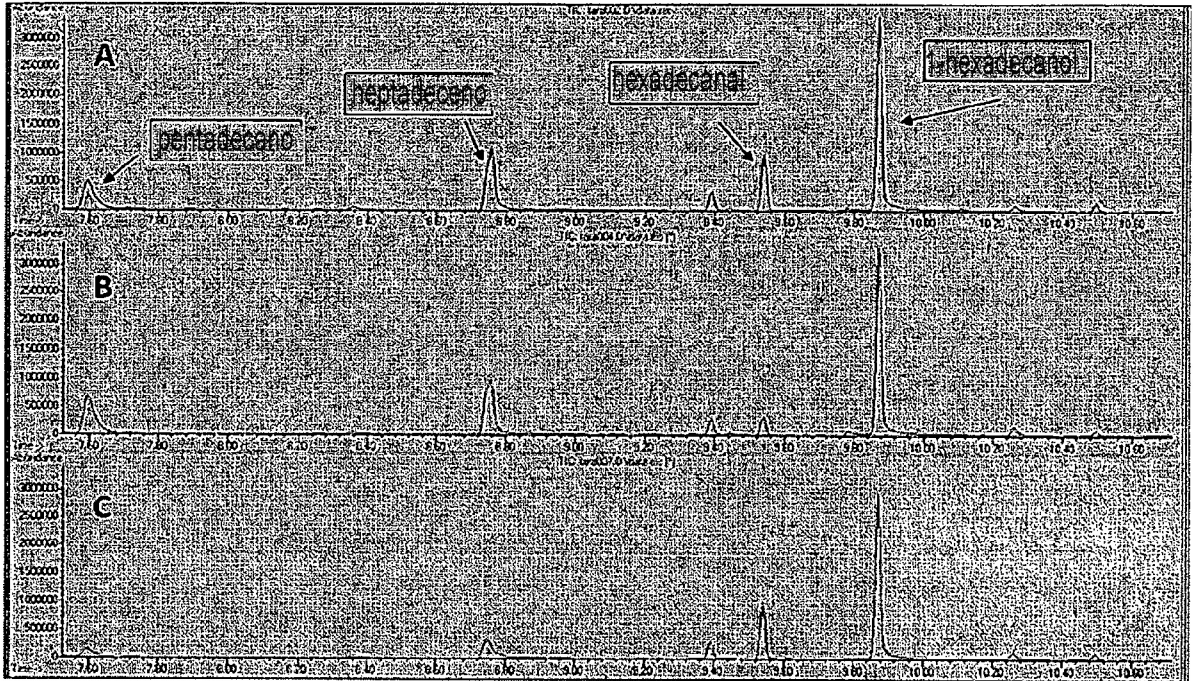


FIG. 30

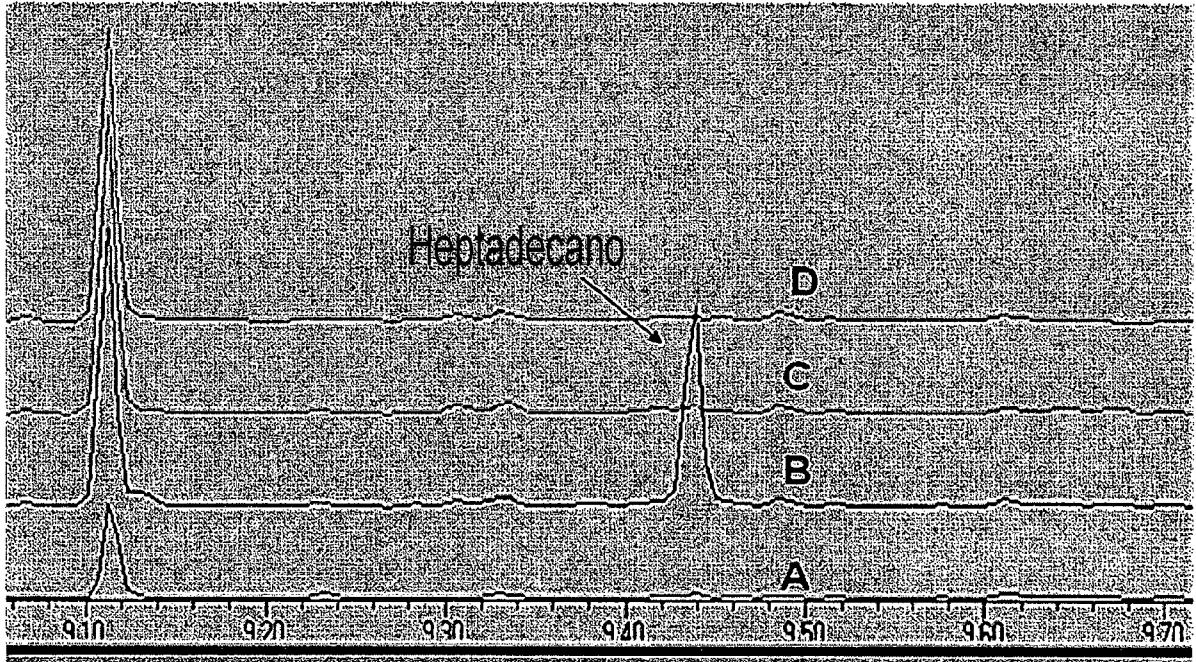


FIG. 31

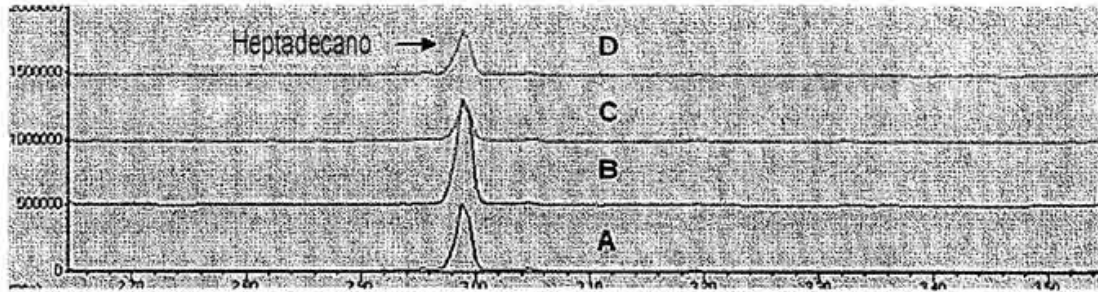


FIG. 32

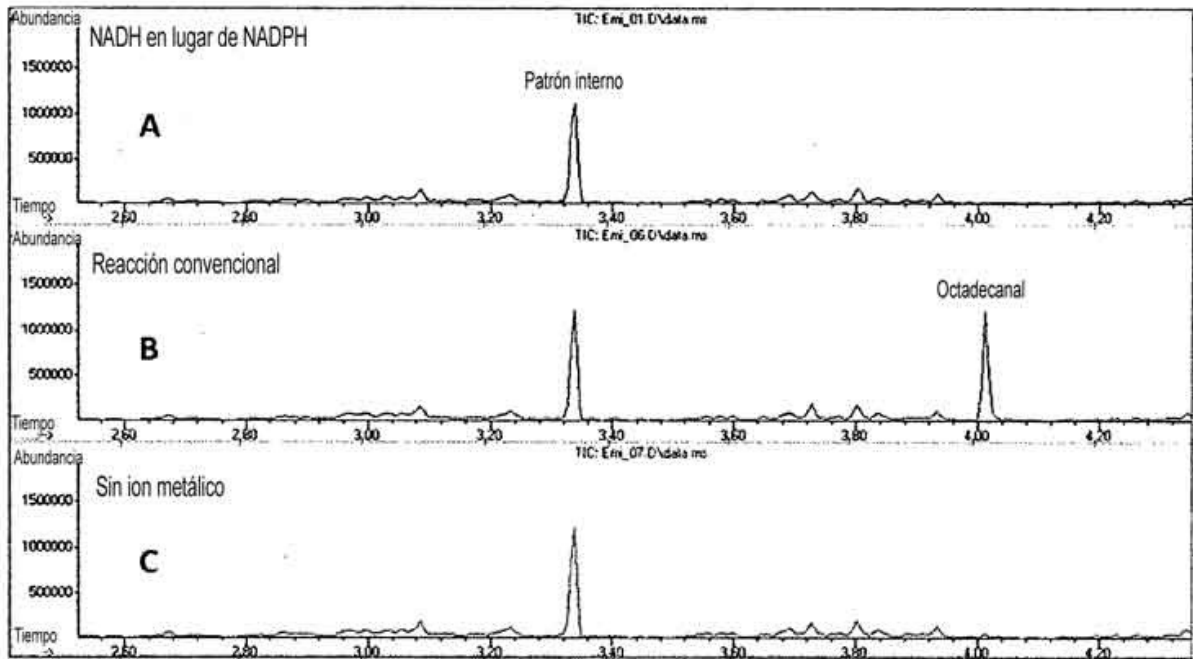


FIG. 33

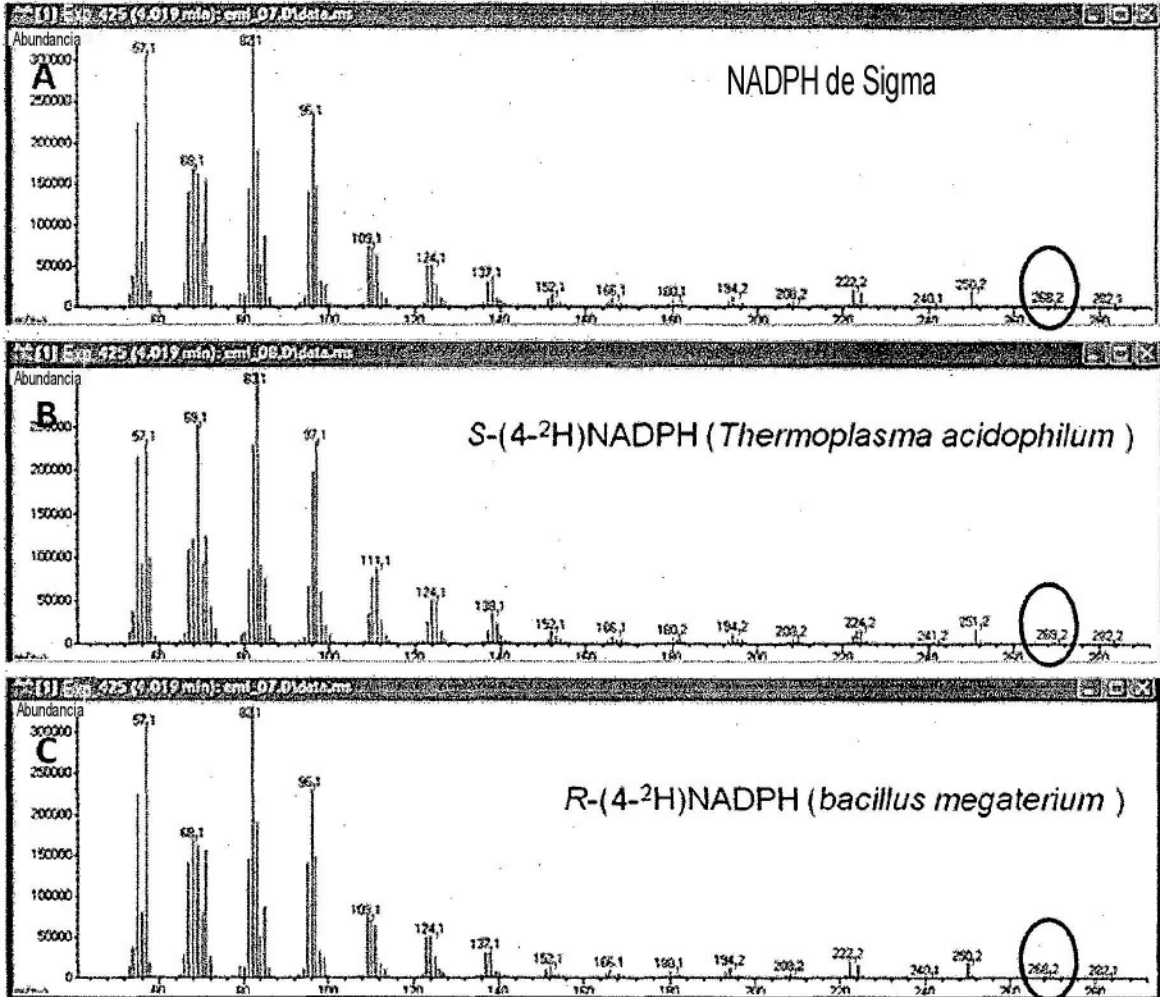


FIG. 34

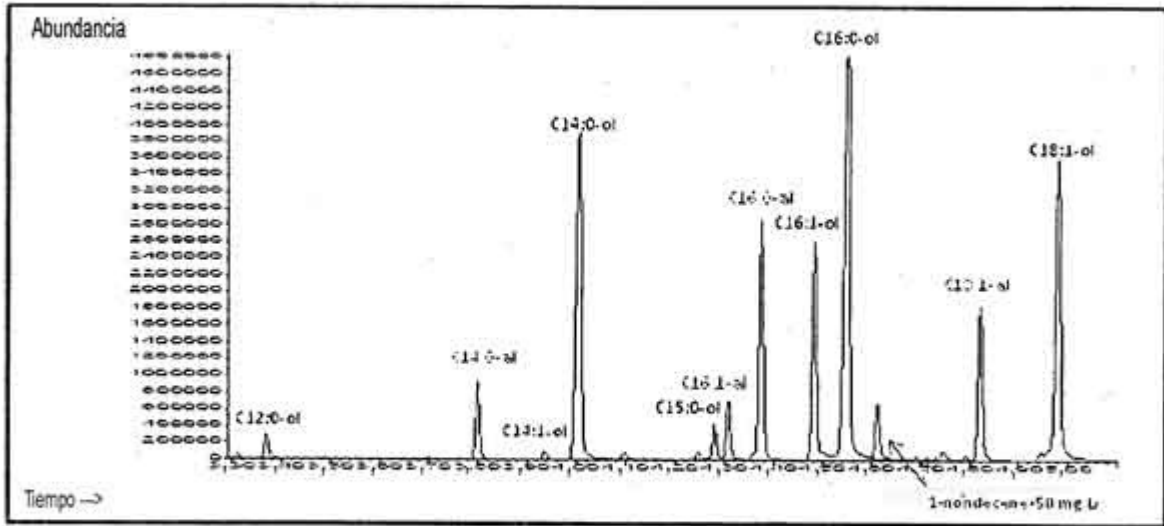


FIG. 35

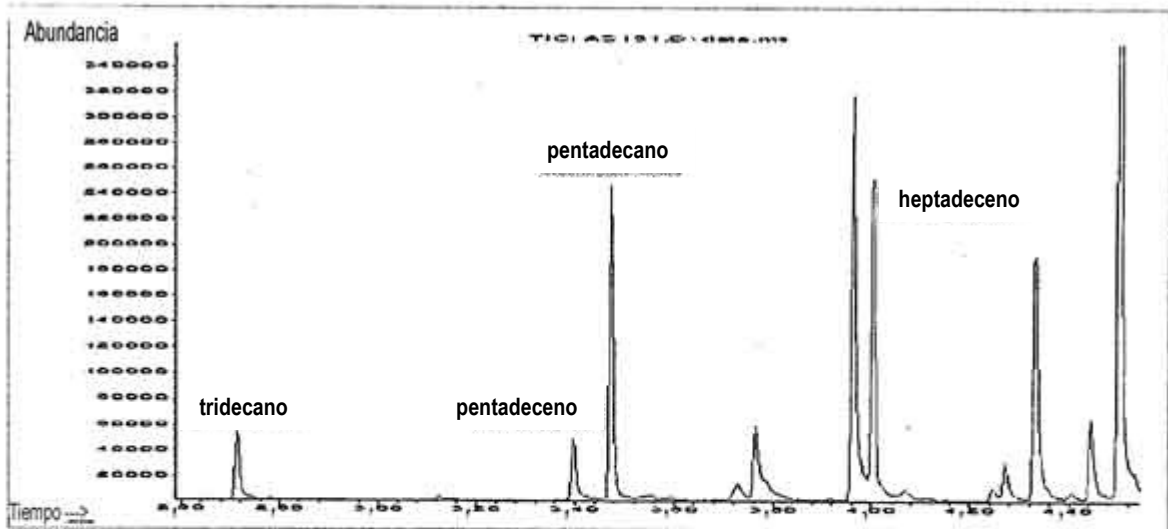


FIG. 36

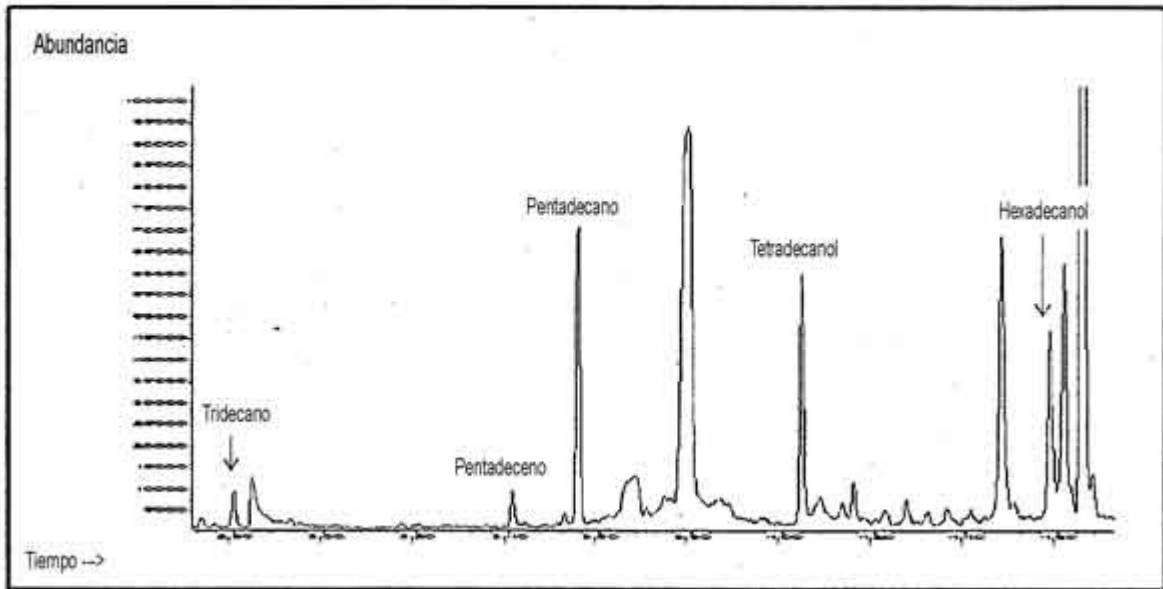


FIG. 37

FIG. 38

Números de registro el 10 de abril de 2009

Número de registro	% de identidad	% de similitud	Longitud de alineación
gi 135970898 gb EBL05614.1	59,5	74,3	237
gi 134964254 gb EBE59803.1	57,6	74,2	229
gi 142528845 gb ECY73505.1	60,4	77,9	222
gi 135713677 gb EBJ38387.1	61,1	78,7	221
gi 141225813 gb ECQ49060.1	59,7	77,8	221
gi 144115151 gb EDI97334.1	67,7	80,9	220
gi 142133005 gb ECV83152.1	67,3	80,9	220
gi 137965371 gb EBX01252.1	67,3	80,9	220
gi 134786157 gb EBD42319.1	67,3	80,9	220
gi 136216894 gb EBM66672.1	62,3	77,7	220
gi 143271262 gb EDE04654.1	63,0	78,5	219
gi 135973786 gb EBL07573.1	63,0	78,1	219
gi 140222739 gb ECK35865.1	63,0	77,6	219
gi 139710482 gb ECG93903.1	62,6	78,1	219
gi 140109767 gb ECJ60294.1	62,6	78,5	219
gi 137939755 gb EBW86789.1	62,6	78,1	219
gi 140086977 gb ECJ44922.1	62,6	78,1	219
gi 143729007 gb EDG48416.1	62,6	78,1	219
gi 143217179 gb EDD66368.1	62,6	78,1	219
gi 138728422 gb ECB60007.1	62,6	78,1	219
gi 143540790 gb EDF53461.1	62,6	78,1	219
gi 143580323 gb EDF73830.1	62,6	78,1	219
gi 137317024 gb EBT41871.1	62,6	78,1	219
gi 143567212 gb EDF67415.1	62,6	78,1	219
gi 140726723 gb ECN09681.1	62,6	78,1	219
gi 136249402 gb EBM88688.1	62,6	78,1	219
gi 141156650 gb ECQ02018.1	62,6	78,1	219
gi 143550472 gb EDF58473.1	62,6	78,1	219
gi 139581046 gb ECG04925.1	62,6	78,1	219
gi 141904835 gb ECU06854.1	62,6	78,1	219
gi 143596624 gb EDF78559.1	62,1	78,1	219
gi 142808717 gb EDA73967.1	62,1	78,1	219
gi 137639134 gb EBV19468.1	62,1	78,1	219
gi 140970945 gb ECO75236.1	62,1	77,6	219

gi 137724335 gb EBV66164.1	62,1	78,1	219
gi 143095952 gb EDC78454.1	66,1	80,3	218
gi 135919851 gb EBK71172.1	66,1	80,3	218
gi 143083445 gb EDC69308.1	66,1	80,3	218
gi 137949431 gb EBW92260.1	63,1	77,6	214
gi 136336883 gb EBN48108.1	62,2	78,0	214
gi 136008042 gb EBL28916.1	62,2	78,0	214
gi 134796061 gb EBD49256.1	60,8	76,6	214
gi 143142117 gb EDD12240.1	60,8	76,6	214
gi 141842906 gb ECT63492.1	60,3	77,6	214
gi 137436109 gb EBU09046.1	60,3	77,1	214
gi 142796298 gb EDA64685.1	71,0	83,3	210
gi 138143480 gb EBX98705.1	70,5	83,3	210
gi 136004890 gb EBL27104.1	68,6	82,4	210
gi 142206956 gb ECW39307.1	67,1	82,4	210
gi 135813588 gb EBK00762.1	62,4	78,6	210
gi 137008728 gb EBR69250.1	62,7	79,0	209
gi 141443295 gb ECR86509.1	63,0	78,9	208
gi 135999440 gb EBL23890.1	63,1	79,1	206
gi 136001501 gb EBL25083.1	62,6	77,7	206
gi 137008499 gb EBR69124.1	62,3	77,5	204
gi 135813080 gb EBK00444.1	60,8	77,5	204
gi 139947991 gb ECI56786.1	63,6	79,8	198
gi 136255251 gb EBM92608.1	61,1	77,3	198
gi 141717528 gb ECS91623.1	58,6	76,8	198
gi 141153056 gb ECP99448.1	59,3	75,3	194
gi 140654921 gb ECM59517.1	67,9	81,9	193
gi 140993407 gb ECO91072.1	63,9	79,6	191
gi 143171676 gb EDD33295.1	63,9	79,6	191
gi 143659341 gb EDG12240.1	59,5	75,8	190
gi 138539256 gb ECA29247.1	68,1	81,9	188
gi 141964470 gb ECU48335.1	64,5	80,3	183
gi 139227663 gb ECE28885.1	63,9	79,2	183
gi 135926500 gb EBK75672.1	58,2	76,4	182
gi 140708786 gb ECM97174.1	68,4	81,4	177
gi 139523141 gb ECF65392.1	62,6	79,3	174
gi 137874181 gb EBW49523.1	61,5	76,4	174
gi 143221750 gb EDD69688.1	60,9	75,9	174
gi 140086962 gb ECJ44914.1	63,4	80,2	172
gi 142781070 gb EDA53384.1	61,8	77,1	170
gi 139775004 gb ECH37282.1	61,0	76,3	169

gi 136260534 gb EBM96169.1	61,0	75,7	169
gi 137441185 gb EBU11854.1	58,1	75,5	167
gi 136330550 gb EBN43814.1	65,1	80,7	166
gi 139650149 gb ECG51660.1	67,3	80,6	165
gi 143638003 gb EDF99869.1	63,6	80,0	165
gi 137949739 gb EBW92432.1	61,7	75,9	162
gi 143382653 gb EDE68551.1	63,9	81,0	158
gi 138989189 gb ECC70595.1	63,9	80,4	158
gi 138408887 gb EBZ46853.1	64,1	78,2	156
gi 137230040 gb EBS93199.1	61,3	76,1	155
gi 141605381 gb ECS53894.1	65,6	81,8	154
gi 137858747 gb EBW40699.1	64,9	81,8	154
gi 140209383 gb ECK27191.1	66,0	79,1	153
gi 142753984 gb EDA33411.1	60,7	75,3	150
gi 137242084 gb EBS99775.1	64,9	82,4	148
gi 136229422 gb EBM75188.1	64,6	82,3	147
gi 140311369 gb ECK89744.1	73,3	85,6	146
gi 140866197 gb ECO03647.1	65,1	82,2	146
gi 139229558 gb ECE29833.1	61,4	76,6	145
gi 141659030 gb ECS68172.1	61,4	76,6	145
gi 139580852 gb ECG04786.1	65,3	81,9	144
gi 138338712 gb EBZ05758.1	65,3	81,9	144
gi 136204827 gb EBM58548.1	61,3	76,8	142
gi 139095530 gb ECD38154.1	64,0	81,3	139
gi 136351648 gb EBN58190.1	60,1	76,1	138
gi 138155154 gb EBY06350.1	67,9	80,3	137
gi 137644530 gb EBV22059.1	63,1	80,8	130
gi 143775710 gb EDG72409.1	61,5	79,2	130
gi 143500330 gb EDF32920.1	63,1	77,9	122
gi 139709584 gb ECG93249.1	71,1	85,1	121
gi 142537519 gb ECY79816.1	62,8	81,0	121
gi 137944410 gb EBW89433.1	64,2	78,3	120
gi 137387955 gb EBT81682.1	63,9	78,2	119
gi 139955976 gb ECI62054.1	71,2	84,8	118
gi 137251843 gb EBT05348.1	62,1	77,6	116
gi 138442523 gb EBZ70326.1	62,1	76,7	116
gi 141590592 gb ECS49420.1	60,7	78,6	112
gi 143187997 gb EDD45026.1	60,7	77,7	112
gi 143655969 gb EDG10472.1	61,5	78,0	109
gi 139459255 gb ECF24788.1	62,9	79,1	105
gi 141976584 gb ECU56751.1	62,9	79,1	105

ES 2 667 469 T3

gi 139233104 gb ECE31190.1	69,6	83,3	102
gi 139233107 gb ECE31193.1	70,3	82,2	101
gi 138582711 gb ECA59616.1	61,4	77,2	101
gi 138442855 gb EBZ70514.1	61,4	75,3	101
gi 137662676 gb EBV31757.1	60,4	76,2	101
gi 139846062 gb ECH87248.1	62,6	74,8	99
gi 136935327 gb EBR27657.1	60,6	74,5	94
gi 137466252 gb EBU25703.1	67,0	80,2	91
gi 137820604 gb EBW18665.1	63,7	80,2	91
gi 138539295 gb ECA29276.1	62,6	79,1	91
gi 136239262 gb EBM81844.1	64,7	81,2	85
gi 136294902 gb EBN19768.1	64,7	81,2	85
gi 137413136 gb EBT96003.1	61,2	76,5	85
gi 137641041 gb EBV20354.1	60,0	72,9	85
gi 142508710 gb ECY58869.1	64,3	82,1	84
gi 140096399 gb ECJ51008.1	62,2	81,7	82
gi 137938664 gb EBW86178.1	64,0	81,3	75
gi 137275448 gb EBT18729.1	58,1	70,3	74
gi 141955842 gb ECU42610.1	63,0	80,8	73
gi 139221707 gb ECE24659.1	66,2	83,1	71
gi 142508709 gb ECY58868.1	52,9	68,6	70
gi 137523719 gb EBU55323.1	65,2	82,6	69
gi 140781524 gb ECN46583.1	65,6	82,8	64
gi 137627577 gb EBV13553.1	60,0	73,3	60
gi 141951833 gb ECU39722.1	59,7	73,7	57
gi 137232510 gb EBS94613.1	59,7	73,7	57

Punto de corte usado: >50% de identidad con y >25% de la longitud de synpcc7942_1593

FIG. 39

Números de registro el 10 de abril de 2009

Número de registro	% de identidad	% de similitud	Longitud de alineación
gi 143288250 gb EDE13503.1	71,3	80,5	87
gi 142342310 gb ECX39602.1	71,3	80,5	87
gi 137949588 gb EBW92346.1	70,7	85,9	92
gi 139984340 gb ECI81897.1	70,6	78,8	85
gi 140249046 gb ECK54318.1	69,7	82,0	267
gi 142111437 gb ECV67406.1	69,2	81,3	182
gi 142994709 gb EDC04737.1	68,8	81,2	138
gi 143066602 gb EDC56955.1	67,7	81,5	248
gi 138840827 gb ECC11022.1	67,1	78,8	146
gi 137829071 gb EBW23606.1	66,7	77,1	96
gi 142133008 gb ECV83155.1	66,5	79,5	337
gi 143095956 gb EDC78458.1	66,4	79,1	339
gi 144115152 gb EDI97335.1	66,4	79,4	339
gi 140732156 gb ECN13587.1	66,1	79,5	254
gi 136241230 gb EBM83170.1	66,0	78,6	103
gi 140001769 gb ECI93451.1	65,9	76,9	91
gi 139305662 gb ECE48752.1	65,9	80,0	205
gi 137634503 gb EBV17219.1	65,8	80,7	114
gi 138584841 gb ECA61142.1	65,5	81,9	116

gi 135919849 gb EBK71170. 1	65,4	76,6	107
gi 141153057 gb ECP99449. 1	65,4	79,6	280
gi 141976585 gb ECU56752 .1	65,3	77,6	98
gi 142206955 gb ECW3930 6.1	65,3	78,5	340
gi 141804802 gb ECT36785. 1	65,3	76,8	95
gi 138931154 gb ECC47219. 1	65,1	76,7	86
gi 138408888 gb EBZ46854. 1	65,1	78,0	255
gi 134743188 gb EBD14908 .1	64,8	78,8	307
gi 138168794 gb EBY16028. 1	64,8	77,6	304
gi 135749749 gb EBJ60721. 1	64,8	79,1	105
gi 138338711 gb EBZ05757. 1	64,6	80,0	175
gi 142827948 gb EDA88477 .1	64,5	77,7	121
gi 138361576 gb EBZ15968. 1	64,4	78,7	267
gi 140517919 gb ECM08416 .1	64,4	77,4	115
gi 135813081 gb EBK00445. 1	64,3	79,7	143
gi 137627576 gb EBV13552. 1	64,3	80,1	171
gi 141161845 gb ECQ05757 .1	64,2	77,1	109
gi 140992134 gb ECO90156 .1	64,2	76,8	95
gi 137796334 gb EBW0459 6.1	64,2	78,9	279
gi 137619413 gb EBV08950. 1	64,1	77,6	223
gi 140517917 gb ECM08414 .1	64,1	79,6	181
gi 137232509 gb EBS94612. 1	64,1	79,0	181
gi 135811491 gb EBJ99446. 1	64,0	77,7	314

gi 141167040 gb ECQ09480 .1	64,0	78,8	250
gi 143441820 gb EDE97777. 1	64,0	78,2	261
gi 140311368 gb ECK89743. 1	64,0	79,3	261
gi 140970943 gb ECO75234 .1	64,0	76,6	111
gi 136817739 gb FBQ60666 .1	63,8	78,5	340
gi 141717529 gb ECS91624. 1	63,7	78,8	146
gi 137632337 gb EBV16047. 1	63,7	77,7	256
gi 137662677 gb EBV31758. 1	63,6	79,7	143
gi 140091056 gb ECJ47190. 1	63,6	78,2	280
gi 143217178 gb EDD66367 .1	63,5	78,8	137
gi 139984339 gb ECI81896. 1	63,5	76,3	156
gi 139382506 gb ECE73591. 1	63,4	78,9	194
gi 140096397 gb ECJ51006. 1	63,4	79,7	153
gi 134606350 gb EBC34611. 1	63,4	77,9	131
gi 140705175 gb ECM95033 .1	63,4	77,0	191
gi 139846064 gb ECH87250 .1	63,1	76,0	179
gi 137953535 gb EBW9457 2.1	63,1	78,0	241
gi 143738737 gb EDG53066 .1	63,1	76,6	111
gi 141951832 gb ECU39721 .1	63,1	76,6	111
gi 143271261 gb EDE04653. 1	63,1	76,6	111
gi 139846065 gb ECH87251 .1	63,0	78,8	146
gi 137251844 gb EBT05349. 1	63,0	78,8	146
gi 136249401 gb EBM88687 .1	63,0	78,8	146

gi 134628580 gb EBC48074. 1	63,0	77,0	100
gi 136312048 gb EBN31461 .1	62,9	76,2	143
gi 143221751 gb EDD69689 .1	62,9	78,9	194
gi 141955844 gb ECU42612 .1	62,9	75,7	140
gi 143395654 gb EDE73119. 1	62,8	77,5	218
gi 142781071 gb EDA53385 .1	62,8	79,3	164
gi 136303394 gb EBN25555 .1	62,7	77,7	319
gi 143557688 gb EDF62238. 1	62,6	77,9	131
gi 136008043 gb EBL28917. 1	62,6	77,6	294
gi 143596625 gb EDF78560. 1	62,6	77,0	318
gi 137641042 gb EBV20355. 1	62,6	79,1	163
gi 136231267 gb EBM76426 .1	62,5	76,7	339
gi 143175604 gb EDD36054 .1	62,5	76,7	339
gi 142508708 gb ECY58867. 1	62,5	76,7	339
gi 135926501 gb EBK75673. 1	62,5	76,6	320
gi 141955884 gb ECU42641 .1	62,4	77,8	189
gi 142821119 gb EDA83282 .1	62,4	78,7	202
gi 134609411 gb EBC36492. 1	62,4	75,8	194
gi 142885864 gb EDB27722 .1	62,3	75,4	207
gi 136204828 gb EBM58549 .1	62,3	77,0	318
gi 143580324 gb EDF73831. 1	62,2	76,7	339
gi 143766375 gb EDG67769 .1	62,2	77,0	339
gi 143500332 gb EDF32922. 1	62,2	76,4	339

gi 139233105 gb ECE31191. 1	62,1	76,5	132
gi 143738779 gb EDG53089 .1	62,1	76,8	314
gi 134964255 gb EBE59804. 1	62,1	77,1	153
gi 140863545 gb ECO01751 .1	62,1	77,0	269
gi 137944409 gb EBW8943 2.1	62,0	78,5	158
gi 143411619 gb EDE81261. 1	62,0	76,7	339
gi 142753988 gb EDA33415 .1	62,0	77,0	339
gi 139580853 gb ECG04787 .1	61,9	75,7	202
gi 141227933 gb ECQ50606 .1	61,9	75,1	173
gi 143659340 gb EDG12239 .1	61,8	78,3	157
gi 136935328 gb EBR27658. 1	61,8	77,7	157
gi 137275449 gb EBT18730. 1	61,8	77,5	204
gi 138585243 gb ECA61437. 1	61,8	76,5	136
gi 139195947 gb ECE06889. 1	61,7	76,2	269
gi 139424973 gb ECF02640. 1	61,7	76,3	274
gi 141380828 gb ECR42772. 1	61,7	75,8	227
gi 136351647 gb EBN58189 .1	61,7	77,0	339
gi 136304410 gb EBN26254 .1	61,6	75,7	185
gi 139948037 gb ECI56814. 1	61,6	76,8	224
gi 135970899 gb EBL05615. 1	61,6	76,8	211
gi 138627165 gb ECA90647. 1	61,5	75,5	143
gi 137395720 gb EBT86160. 1	61,5	74,8	143
gi 140086960 gb ECJ44912. 1	61,5	77,1	249

gi 141024916 gb ECP11582. 1	61,4	76,7	223
gi 139095531 gb ECD38155. 1	61,4	76,7	210
gi 141659029 gb ECS68171. 1	61,4	77,2	127
gi 139969430 gb ECI71470. 1	61,4	76,7	215
gi 136986729 gb EBR56775. 1	61,4	75,7	202
gi 143634197 gb EDF97600. 1	61,3	76,7	313
gi 135973785 gb EBL07572. 1	61,3	76,3	279
gi 143200944 gb EDD54508 .1	61,3	74,6	173
gi 137787263 gb EBV99371. 1	61,2	76,3	232
gi 139204136 gb ECE12313. 1	61,2	76,7	219
gi 136001500 gb EBL25082. 1	61,2	76,3	219
gi 141874476 gb ECT85572. 1	61,2	76,4	237
gi 137905325 gb EBW6737 5.1	61,1	74,9	175
gi 140089341 gb ECJ46519. 1	61,1	75,4	203
gi 140855194 gb ECN95754 .1	61,1	75,9	203
gi 134965622 gb EBE60718. 1	61,1	75,8	339
gi 141527125 gb ECS15588. 1	61,1	76,0	208
gi 136218988 gb EBM68086 .1	61,0	73,4	154
gi 142364499 gb ECX54765. 1	60,9	76,1	330
gi 141603393 gb ECS53340. 1	60,8	74,8	143
gi 136216893 gb EBM66671 .1	60,8	76,3	245
gi 143743653 gb EDG56305 .1	60,8	73,7	148
gi 140222741 gb ECK35867. 1	60,7	73,0	163

gi 142389887 gb ECX71634. 1	60,6	74,8	198
gi 139315697 gb ECE51398. 1	60,6	76,4	241
gi 136255250 gb EBM92607 .1	60,5	76,0	291
gi 137387954 gb EBT81681. 1	60,5	76,6	124
gi 139229559 gb ECE29834. 1	60,4	75,3	202
gi 140866196 gb ECO03646 .1	60,1	74,6	303
gi 139229561 gb ECE29836. 1	60,0	74,0	100
gi 140957440 gb ECO66006 .1	59,9	75,9	274
gi 143567213 gb EDF67416. 1	59,8	74,8	286
gi 139955973 gb ECI62051. 1	59,8	72,2	97
gi 140726724 gb ECN09682 .1	59,7	73,4	154
gi 139775003 gb ECH37281 .1	59,6	73,7	99
gi 137949740 gb EBW9243 3.1	58,7	71,7	92
gi 139650150 gb ECG51661 .1	57,7	73,2	97
gi 142528844 gb ECY73504. 1	56,7	70,0	90

Punto de corte usado: >50% de identidad con y >25% de la longitud de synpcc7942_1594

Fig. 40

Los números de registro proceden de NCBI, GenBank, versión 159.0 de 15 de abril de 2007.
 Los números EC proceden de KEGG, versión 42.0 de abril de 2007 (más actualizaciones diarias hasta marzo de 2008)

CATEGORÍA	GEN	NOMBRE	REGISTRO	NÚMERO EC	MODIFICACIÓN	USO	MICROORGANISMO
1.	Aumento de la producción de ácidos grasos / Aumento de la producción de productos.						
		aumento de acil-CoA					
		reducción del catabolismo de derivados y productos intermedios					
		reducción de la inhibición de la reutilización					
		atenuación de otras rutas que consumen ácidos grasos					
	accA	Acetil-CoA carboxilasa, subunidad A (carboxiltransferasa alfa)	AAC73296, NP_414727	6.4.1.2	Sobreespresión	aumento de la producción de Malonil-CoA	<i>Escherichia coli, Lactococci</i>
	accB	Acetil-CoA carboxilasa, subunidad B (BCCP, proteína portadora de carbono y biotina)	NP_417721	6.4.1.2	Sobreespresión	aumento de la producción de Malonil-CoA	<i>Escherichia coli, Lactococci</i>
	accC	Acetil-CoA carboxilasa, subunidad C (biotina carboxilasa)	NP_417722	6.4.1.2, 6.3.4.14	Sobreespresión	aumento de la producción de Malonil-CoA	<i>Escherichia coli, Lactococci</i>
	accD	Acetil-CoA carboxilasa, subunidad D (carboxiltransferasa beta)	NP_416819	6.4.1.2	Sobreespresión	aumento de la producción de Malonil-CoA	<i>Escherichia coli, Lactococci</i>
	accE	Privato deshidrogenasa, subunidad E1	NP_414656, AAC73226	1.2.4.1	Sobreespresión	aumento de la producción de Acetil-CoA	<i>Escherichia coli</i>

aceF	Piruvato deshidrogenasa, subunidad E2	NP 414657	2.3.1.12	Sobreexpresión	aumento de la producción de acetil-CoA	<i>Escherichia coli</i>
ackA	Acetato cinasa	AAC75356, NP 416799	2.7.2.1	Delección o reducción	aumento de la producción de acetil-CoA	<i>Escherichia coli</i>
ackB	Acetato cinasa AckB	BAB81430	2.7.2.1	Delección o reducción	aumento de la producción de acetil-CoA	<i>Escherichia coli</i>
acpP	Proteína transportadora de acilo	AAC74178	NINGUNO	Sobreexpresión	aumento de la producción de acetil-CoA	<i>Escherichia coli</i>
fadD	Acil-CoA sintasa	AP 002424	2.3.1.86, 6.2.1.3	Sobreexpresión	aumento de la producción de ácidos grasos	<i>Escherichia coli</i> W3110
adhE	Alcohol deshidrogenasa	CAA47743	1.1.1.1, 1.2.1.10	Delección o reducción	aumento de la producción de acetil-CoA	<i>Escherichia coli</i> W3111
cer1	Aldehído descarboxilasa	BAA11024	4.1.99.5	Sobreexpresión	aumento de la producción de acetil-CoA	<i>Arabidopsis thaliana</i>
fabA	Beta-hidroxicocoil tioréster deshidrasa	NP 415474	4.2.1.60	Expresión	producción de acil-CoA grasos	<i>E. coli</i> K12
fabD	S-maloniitransferasa de la proteína portadora de acilo	AAC74176	2.3.1.39	Sobreexpresión	aumento de la producción de acetil-CoA	<i>E. coli</i> K12
fabF	3-oxoacil-(proteína portadora de acilo) sintasa II	AAC74179	2.3.1.179	Delección o sobreexpresión	aumento de la producción de acetil-CoA	<i>E. coli</i> K12

	fabG	3-oxoacil-proteína portadora de acilo reductasa	AAc74177	1.1.1.100	Sobreexpresión	aumento de la producción de acetyl-CoA	<i>E. coli</i> K12
	fabH	3-oxoacil-proteína portadora de acilo sintasa III	AAc74175	2.3.1.180	Sobreexpresión	aumento de la producción de acetyl-CoA	<i>E. coli</i> K12, <i>Lactococci</i>
	fabI	enoi proteína portadora de acilo reductasa, dependiente de NADH	NP_415804	1.3.1.9	Expresión	producción de acil grasp-CoA	<i>E. coli</i> K12, <i>Lactococci</i>
	fabR	Represor de la transcripción	NP_418398	NINGUNO	Delección o reducción	modulación de la producción de ácidos grasos insaturados	<i>E. coli</i> K12
	fabZ	(3R)-hidroximiristol proteína portadora de acilo deshidratasa	NP_414722	4.2.1.-			<i>E. coli</i> K12
	fadE	acil-CoA deshidrogenasa	AAc73325	1.3.99.3, 1.3.99.-	Delección o reducción	aumento de la producción de acetyl-CoA	
	acrI	Acil graso-CoA reductasa	YP_047869, AAc45217	1.2.1.42	Sobreexpresión	para la producción de alcoholes grasos	<i>Acinetobacter sp., i.e. calcoaceticus</i>
	GST, gshB	Glutación sintasa	P04425	6.3.2.3	Delección o reducción	aumento de acil-CoA	<i>E. coli</i> K12
	gpsA	sn-glicerol 3-fosfato deshidrogenasa biosintética	AAc76632, NP_418065	EC: 1.1.1.94	Delección o reducción	aumento de la producción de acetyl-CoA	<i>E. coli</i> K12
	ldhA	lactato deshidrogenasa	AAc74462, NP_415898	EC: 1.1.1.27, 1.1.1.28	Delección o reducción	aumento de la producción de Acetyl-CoA	<i>E. coli</i> K12
	Lipasa	triglicérido lipasa	CAA89087, CAA98876	3.1.1.3	Expresión	aumento de la producción de ácidos grasos	<i>Saccharomyces cerevisiae</i>
		Malonil-CoA descarboxilasa	AAA26500	4.1.1.9, 4.1.1.41	Sobreexpresión		<i>Saccharopolyspor a erythraea</i>
	panD	aspartato-1-descarboxilasa	BAB96708	4.1.1.11	Sobreexpresión	aumento de acil-CoA	<i>Escherichia coli W3110</i>

panK t.c.c. coaA	pantotenato cinasa	AAC76952	2.7.1.33	Sobreexpresión	aumento de la producción de acetil-CoA	<i>E. coli</i>
panK t.c.c. coaA, R106K	pantotenato cinasa	AAC76952	2.7.1.33	Expresión, sobreexpresión, mutación R106K	aumento de la producción de acetil-CoA	<i>E. coli</i>
pdh	Piruvato deshidrogenasa	BAB34380, AAC73226, NP 415392	1.2.4.1	Sobreexpresión	aumento de la producción de acetil-CoA	
pflB	formiato acetiltransferasa (piruvato formiato liasa)	AAC73989, P09373	EC: 2.3.1.54	Delección o reducción	aumento de la producción de acetil-CoA	
plsB	acetiltransferasa	AAC77011	2.3.1.15	Mutación D311E	reducción de los límites sobre la reserva de acil-CoA	<i>E. coli</i> K12
poxB	piruvato oxidasa	AAC73958, NP 415392	1.2.2.2	Delección o reducción	aumento de la producción de acetil-CoA	
pta	fosfoacetilasa	AAC75357, NP 416800	2.3.1.8	Delección o reducción	aumento de la producción de acetil-CoA	
udhA	nucleótido de piridina transhidrogenasa	CAA46822	1.6.1.1	Sobreexpresión	conversión de NADH en NADPH o viceversa	
fadB	3-hidroxitubitil-CoA epimerasa/ delta(3)-cis-delta(2)-trans-enoil-CoA isomerasa/ enoil-CoA hidratasa y 3-hidroxiacil-CoA deshidrogenasa fusionadas	AP_003956	4.2.1.17, 5.1.2.3, 5.3.3.8, 1.1.1.35	Delección o reducción	bloqueo de la degradación de ácidos grasos	<i>E. coli</i>

	fadJ	3-hidroxiacil-CoA deshidrogenasa; K01692 enoil-CoA hidratasa; K01782 3-hidroxiubutil-CoA epimerasa	AAC75401	1.1.1.35, 4.2.1.17, 5.1.2.3	Delección o reducción	bloqueo de la degradación de ácidos grasos	<i>E. coli</i>
	fadA	3-oxobacil-CoA tioiasa	BAE77458	2.3.1.16	Delección o reducción	bloqueo de la degradación de ácidos grasos	<i>E. coli</i>
	fadI	beta-oxobacil-CoA tioiasa	AAC75402	2.3.1.16	Delección o reducción	bloqueo de la degradación de ácidos grasos	<i>E. coli</i>
	YdiO	acil-CoA deshidrogenasa	YP 852786	1.3.99.-	Delección o reducción	bloqueo de la degradación de ácidos grasos	<i>E. coli</i>
2. Control de la estructura							
2A. Control de la longitud de cadena							
2	tesA	tioesterasa	P0ADA1	3.1.2.-, 3.1.1.5	Delección y/o expresión	Longitud de cadena C18	
	tesA sin secuencia líder	tioesterasa	AAC73596, NP 415027	3.1.2.-, 3.1.1.5	Expresión o sobreexpresión	C18:1	<i>E. coli</i>
	tesA sin secuencia líder. L109P	tioesterasa	P0ADA1	3.1.2.-, 3.1.1.5	Expresión y/o sobreexpresión de mutación L109P	Longitud de cadena <C18	<i>E. coli</i>
	fatB1 (umbellularia)	tioesterasa	Q41635	3.1.2.14	Expresión o sobreexpresión	C12:0	<i>Umbellularia californica</i>
	fatB2 (umbellularia)D ELETE	tioesterasa	AAC49269	3.1.2.14	Expresión o sobreexpresión	C8:0 - C10:0	<i>Cuphea hookeriana</i>

umbellu ria)									
fatB3	tiosterasa		AAC72881	3.1.2.14	Expresión o sobreexpresión	C14:0 - C16:0	<i>Cuphea hookeriana</i>		
fatB (cinnam onum)	tiosterasa		Q39473	3.1.2.14	Expresión o sobreexpresión	C14:0	<i>Cinnamomum camphora</i>		
fatB[M 141T]*	tiosterasa		CAA85388	3.1.2.14	Expresión o sobreexpresión	C16:1	<i>Arabidopsis thaliana</i>		
fatA1 (Heliant hus)	tiosterasa		AAL79361	3.1.2.14	Expresión o sobreexpresión	C18:1	<i>Helianthus annuus</i>		
atfata (ACIL-ACP TIOESTERA- SA DE ARABI- DOPFIS FATA)	tiosterasa		NP_189147, NP_193041	3.1.2.14	Expresión o sobreexpresión	C18:1	<i>Arabidopsis thaliana</i>		
fatA	tiosterasa		CAC39106	3.1.2.14	Expresión o sobreexpresión	C18:1	<i>Brassica juncea</i>		
fatA (cuphea)	tiosterasa		AAC72883	3.1.2.14	Expresión o sobreexpresión	C18:1	<i>Cuphea hookeriana</i>		
2B. Control de la ramificación									
atenuación de FabH									

<i>bkd</i> - E2								
subunidad dihidrolipoil transacetilasa			EC 1.2.4.4					
<i>bkdA1</i>	subunidad-a de a-cetoácido de cadena ramificada deshidrogenasa (E1a)	NP 628006	EC 1.2.4.4	Expresión o sobreexpresión	preparación de precursores de acil-CoA de cadena ramificada	<i>Streptomyces coelicolor</i>		
<i>bkdB1</i>	subunidad-a de a-cetoácido de cadena ramificada deshidrogenasa (E1b)	NP 628005	EC 1.2.4.4	Expresión o sobreexpresión	preparación de precursores de acil-CoA de cadena ramificada	<i>Streptomyces coelicolor</i>		
<i>bkdC1</i>	dihidrolipoil transacetilasa (E2)	NP 628004	EC 2.3.1.168	Expresión o sobreexpresión	preparación de precursores de acil-CoA de cadena ramificada	<i>Streptomyces coelicolor</i>		
<i>bkdA2</i>	subunidad-a de a-cetoácido de cadena ramificada deshidrogenasa (E1a)	NP 733618	EC 1.2.4.4	Expresión o sobreexpresión	preparación de precursores de acil-CoA de cadena ramificada	<i>Streptomyces coelicolor</i>		
<i>bkdB2</i>	subunidad-b de a-cetoácido de cadena ramificada deshidrogenasa (E1b)	NP 628019	EC 1.2.4.4	Expresión o sobreexpresión	preparación de precursores de acil-CoA de cadena ramificada	<i>Streptomyces coelicolor</i>		
<i>bkdC2</i>	dihidrolipoil transacetilasa (E2)	NP 628018	EC 2.3.1.168	Expresión o sobreexpresión	preparación de precursores de acil-CoA de cadena ramificada	<i>Streptomyces coelicolor</i>		

	bkdA	subunidad-a de a-cetoácido de cadena ramificada deshidrogenasa (E1a)	BAC72074		EC 1.2.4.4	Expresión o sobreexpresión	preparación de precursores de acil-CoA de cadena ramificada	<i>Streptomyces avermitilis</i>
	bkdB	subunidad-b de a-cetoácido de cadena ramificada deshidrogenasa (E1b)	BAC72075		EC 1.2.4.4	Expresión o sobreexpresión	preparación de precursores de acil-CoA de cadena ramificada	<i>Streptomyces avermitilis</i>
	bkdC	dihidrolipoil transacetilasa (E2)	BAC72076		EC 2.3.1.168	Expresión o sobreexpresión	preparación de precursores de acil-CoA de cadena ramificada	<i>Streptomyces avermitilis</i>
	bkdF	subunidad-a de a-cetoácido de cadena ramificada deshidrogenasa (E1a)	BAC72088		EC 1.2.4.4	Expresión o sobreexpresión	preparación de precursores de acil-CoA de cadena ramificada	<i>Streptomyces avermitilis</i>
	bkdG	subunidad-b de a-cetoácido de cadena ramificada deshidrogenasa (E1b)	BAC72089		EC 1.2.4.4	Expresión o sobreexpresión	preparación de precursores de acil-CoA de cadena ramificada	<i>Streptomyces avermitilis</i>
	bkdH	dihidrolipoil transacetilasa (E2)	BAC72090		EC 2.3.1.168	Expresión o sobreexpresión	preparación de precursores de acil-CoA de cadena ramificada	<i>Streptomyces avermitilis</i>
	bkdAA	subunidad-a de a-cetoácido de cadena ramificada deshidrogenasa (E1a)	NP_390285		EC 1.2.4.4	Expresión o sobreexpresión	preparación de precursores de acil-CoA de cadena ramificada	<i>Bacillus subtilis</i>
	bkdAB	subunidad-b de a-cetoácido de cadena ramificada deshidrogenasa (E1b)	NP_390284		EC 1.2.4.4	Expresión o sobreexpresión	preparación de precursores de acil-CoA de	<i>Bacillus subtilis</i>

	IlvE	aminoácido de cadena ramificada aminotransferasa	NP_629657		2.6.1.42	Expresión o sobreexpresión	preparación de a-ceto-ácido de cadena ramificada	<i>Streptomyces coelicolor</i>
	ccr	crotonil-CoA reductasa	NP_630556		1.6.5.1.1.1.1	Expresión o sobreexpresión	conversión de crotonil-CoA en butiril-CoA	<i>Streptomyces coelicolor</i>
	ccr	crotonil-CoA reductasa	AAD53915		1.6.5.1.1.1.1	Expresión o sobreexpresión	conversión de crotonil-CoA en butiril-CoA	<i>Streptomyces cinnamomensis</i>
	IcmA, isobutiril-CoA mutasa	isobutiril-CoA mutasa, subunidad A	NP_629554		5.4.99.2	Expresión o sobreexpresión	conversión de butiril-CoA en isobutiril-CoA	<i>Streptomyces coelicolor</i>
	IcmA, isobutiril-CoA mutasa	isobutiril-CoA mutasa, subunidad A	AAC08713		5.4.99.2	Expresión o sobreexpresión	conversión de butiril-CoA en isobutiril-CoA	<i>Streptomyces cinnamomensis</i>
	IcmB, isobutiril-CoA mutasa	isobutiril-CoA mutasa, subunidad B	NP_630904		5.4.99.2	Expresión o sobreexpresión	conversión de butiril-CoA en isobutiril-CoA	<i>Streptomyces coelicolor</i>
	IcmB, isobutiril-CoA mutasa	isobutiril-CoA mutasa, subunidad B	CAB59633		5.4.99.2	Expresión o sobreexpresión	conversión de butiril-CoA en isobutiril-CoA	<i>Streptomyces cinnamomensis</i>

Genes FabH, ACP y fabF con especifici- dad por Acil-CoA de cadena ramifica- da									
IlvE	Aminoácido de cadena ramificada aminotransferasa	CAC12788	EC2.6.1.42		Sobreexpresión	Aminoácido de cadena ramificada aminotransferasa	Staphylococcus carnosus		
FabH1	Beta-cetoacil-ACP sintasa III	NP_626634	2.3.1.180		Expresión o sobreexpresión	Inicio de la biosíntesis de ácidos grasos de cadena ramificada	Streptomyces coelicolor		
ACP	Proteína portadora de acilo	NP_626635	NINGUNO		Expresión o sobreexpresión	Inicio y elongación de la biosíntesis de ácidos grasos de cadena ramificada	Streptomyces coelicolor		
FabF	Beta-cetoacil-ACP sintasa II	NP_626636	2.3.1.179		Expresión o sobreexpresión	Elongación de la biosíntesis de ácidos grasos de cadena ramificada	Streptomyces coelicolor		

FabH, ACP y FabF con especificidad por acil-CoA de cadena ramificada									
IIVe	aminoalodo de cadena ramificada aminotransferasa	CAC12788	EC2.6.1.42	Sobreexpresión	aminoalodo de cadena ramificada aminotransferasa	Staphylococcus carnosus			
FabH1	beta-oxoaci-ACP sintasa III	NP_626634	2.3.1.180	Expresión o sobreexpresión	inicio de la biosíntesis de ácidos grasos de cadena ramificada	Streptomyces coelicolor			
ACP	proteína portadora de ácido	NP_626635	NINGUNO	Expresión o sobreexpresión	inicio y elongación de la biosíntesis de ácidos grasos de cadena ramificada	Streptomyces coelicolor			
FabF	beta-oxoaci-ACP sintasa II	NP_626636	2.3.1.179	Expresión o sobreexpresión	elongación de la biosíntesis de ácidos grasos de cadena ramificada	Streptomyces coelicolor			

FabH3	beta-cetoacil-ACP sintasa III	NP_823466	2.3.1.180	Expresión o sobreexpresión	inicio de la biosíntesis de ácidos grasos de cadena ramificada	<i>Streptomyces avermitilis</i>
FabC3 (ACP)	proteína portadora de acilo	NP_823467	NINGUNO	Expresión o sobreexpresión	inicio y elongación de la biosíntesis de ácidos grasos de cadena ramificada	<i>Streptomyces avermitilis</i>
FabF	beta-cetoacil-ACP sintasa II	NP_823468	2.3.1.179	Expresión o sobreexpresión	elongación de la biosíntesis de ácidos grasos de cadena ramificada	<i>Streptomyces avermitilis</i>
FabH_A	beta-cetoacil-ACP sintasa III	NP_389015	2.3.1.180	Expresión o sobreexpresión	inicio de la biosíntesis de ácidos grasos de cadena ramificada	<i>Bacillus subtilis</i>
FabH_B	beta-cetoacil-ACP sintasa III	NP_388898	2.3.1.180	Expresión o sobreexpresión	inicio de la biosíntesis de ácidos grasos de cadena ramificada	<i>Bacillus subtilis</i>

	ACP	proteína portadora de ácido	NP_389474	NINGUNO	Expresión o sobreexpresión	inicio y elongación de la biosíntesis de ácidos grasos de cadena ramificada	<i>Bacillus subtilis</i>
	FabF	beta-cetoadil-ACP sintasa II	NP_389016	2.3.1.179	Expresión o sobreexpresión	elongación de la biosíntesis de ácidos grasos de cadena ramificada	<i>Bacillus subtilis</i>
	SmaID RAFT_0818	beta-cetoadil-ACP sintasa III	ZP_01643059	2.3.1.180	Expresión o sobreexpresión	inicio de la biosíntesis de ácidos grasos de cadena ramificada	<i>Stenotrophomonas maltophilia</i>
	SmaID RAFT_0821	proteína portadora de ácido	ZP_01643063	NINGUNO	Expresión o sobreexpresión	inicio y elongación de la biosíntesis de ácidos grasos de cadena ramificada	<i>Stenotrophomonas maltophilia</i>
	SmaID RAFT_0822	beta-cetoadil-ACP sintasa II	ZP_01643064	2.3.1.179	Expresión o sobreexpresión	elongación de la biosíntesis de ácidos grasos de cadena ramificada	<i>Stenotrophomonas maltophilia</i>

FabH	beta-cetoadil-ACP sintasa III	YP_123672	2.3.1.180	Expresión o sobreexpresión	inicio de la biosíntesis de ácidos grasos de cadena ramificada	<i>Legionella pneumophila</i>
ACP	proteína portadora de ácido	YP_123675	NINGUNO	Expresión o sobreexpresión	inicio y elongación de la biosíntesis de ácidos grasos de cadena ramificada	<i>Legionella pneumophila</i>
FabF	beta-cetoadil-ACP sintasa II	YP_123676	2.3.1.179	Expresión o sobreexpresión	elongación de la biosíntesis de ácidos grasos de cadena ramificada	<i>Legionella pneumophila</i>
FabH	beta-cetoadil-ACP sintasa III	NP_415609	2.3.1.180	Delección o reducción	inicio de la biosíntesis de ácidos grasos de cadena ramificada	<i>Escherichia coli</i>
FabF	beta-cetoadil-ACP sintasa II	NP_415613	2.3.1.179	Delección o reducción	elongación de la biosíntesis de ácidos grasos de cadena ramificada	<i>Escherichia coli</i>

Para la producción de ácidos grasos cíclicos

AnsJ	deshidratasa (supuesta)	no disponible	no disponible	Expresión o sobreexpresión	biosíntesis de ciclohexilcarbonil-CoA	<i>Streptomyces collinus</i>
AnsK	CoA ligasa (supuesta)	no disponible	no disponible	Expresión o sobreexpresión	biosíntesis de ciclohexilcarbonil-CoA	<i>Streptomyces collinus</i>
AnsL	deshidrogenasa (supuesta)	no disponible	no disponible	Expresión o sobreexpresión	biosíntesis de ciclohexilcarbonil-CoA	<i>Streptomyces collinus</i>
ChcA	enol-CoA reductasa	U72144	EC 1.3.1.34	Expresión o sobreexpresión	biosíntesis de ciclohexilcarbonil-CoA	<i>Streptomyces collinus</i>
AnsM	oxidoreductasa (supuesta)	no disponible	no disponible	Expresión o sobreexpresión	biosíntesis de ciclohexilcarbonil-CoA	<i>Streptomyces collinus</i>
PImJ	deshidratasa (supuesta)	AA084158	no disponible	Expresión o sobreexpresión	biosíntesis de ciclohexilcarbonil-CoA	<i>Streptomyces sp. HK803</i>
PImK	CoA ligasa (supuesta)	AA084158	no disponible	Expresión o sobreexpresión	biosíntesis de ciclohexilcarbonil-CoA	<i>Streptomyces sp. HK803</i>
PImL	deshidrogenasa (supuesta)	AA084159	no disponible	Expresión o sobreexpresión	biosíntesis de ciclohexilcarbonil-CoA	<i>Streptomyces sp. HK803</i>

GnsA	supresores de la mutación nula secG	ABD18647.1	NINGUNO	Sobreexpresión	aumento de ésteres de ácidos grasos insaturados	<i>E. coli</i>
GnsB	supresores de la mutación nula secG	AACT4076.1	NINGUNO	Sobreexpresión	aumento de ésteres de ácidos grasos insaturados	<i>E. coli</i>
fabB	3-oxoacil-(proteína portadora de acil) sintasa I	BAA16180	EC.2.3.1.41	Sobreexpresión	modulación de la producción de ácidos grasos insaturados	<i>Escherichia coli</i>
fabK	trans-2-enoil-ACP reductasa II	AAF98273	1.3.1.9	Expresión	modulación de la producción de ácidos grasos insaturados	<i>Streptococcus pneumoniae</i>
fabL	enol-(proteína portadora de acil)-reductasa	AAU39821	1.3.1.9	Expresión	modulación de la producción de ácidos grasos insaturados	<i>Bacillus licheniformis DSM 13</i>

	fabM	trans-2, cis-3-decenol-ACP isomerasa	DAA05501	4.2.1.17	Sobreexpresión	modulación de la producción de ácidos grasos insaturados	<i>Streptococcus mutans</i>
3. Salida de productos finales							
3A. Salida de Cera							
	AT3G51970	alcohol de cadena larga C-graso aciltransferasa	NP 190765	2.3.1.26	Expresión	producción de cera	<i>Arabidopsis thaliana</i>
		tiesterasa (véase la sección de control de la longitud de cadena)			Expresión	aumento de la producción de ácidos grasos	
		acil CoA-reductasa que forma alcohol graso		1.1.1.*	Expresión	conversión de acil-CoA en alcohol graso	
	acr1	acil-CoA reductasa (ACR1)	YP 047869	1.2.1.42	Expresión	conversión de acil-CoA en alcohol graso	<i>Acinetobacter sp. ADP1</i>
	yqhD	alcohol deshidrogenasa	AP 003562	1.1.-.-	Expresión	aumento	<i>E. coli W3110</i>
	ELO1	ácido graso elongasa	BAD98251	2.3.1.-	Expresión	producción de ácidos grasos de cadena muy larga	<i>Pichia angusta</i>
	plsC	aciltransferasa	AAA16514	2.3.1.51	Expresión		<i>Saccharomyces cerevisiae</i>
	DAGA T/DGA T	diacilglicerol aciltransferasa	AAF19262	2.3.1.20	Expresión	producción de cera	<i>Arabidopsis thaliana</i>
	hws	acil-CoA cera alcohol aciltransferasa	AAx48018	2.3.1.20	Expresión	producción de cera	<i>Homo sapiens</i>

AMRP 5	asociado a resistencia a múltiples fármacos de <i>Arabidopsis thaliana</i>	NP_171908	NINGUNO	expresión	exportación de productos	<i>Arabidopsis thaliana</i>
AmiS2	transportador de ABC AmiS2	JCS491	NINGUNO	expresión	exportación de productos	<i>Rhodococcus sp.</i>
AlPGP1	GLICOPROTEÍNA 1 DE ARABIDOPSIS THALIANA P	NP_181228	NINGUNO	expresión	exportación de productos	<i>Arabidopsis thaliana</i>
AcrA	supuesta proteína acrA de transporte de eflujo de múltiples fármacos	CAF23274	NINGUNO	expresión	exportación de productos	<i>Candidatus Prochloromydia amoebophila UIWE25</i>
AcrB	probable proteína de transporte de eflujo de múltiples fármacos, acrB	CAF23275	NINGUNO	expresión	exportación de productos	<i>Candidatus Prochloromydia amoebophila UIWE25</i>
ToIC	Proteína de la membrana externa [biogénesis de la envuelta celular]	ABD59001	NINGUNO	expresión	exportación de productos	<i>Francisella tularensis subsp. novicida</i>
AcrE	proteína transmembrana que afecta a la formación del septo y a la permeabilidad de la membrana celular	YP_312213	NINGUNO	expresión	exportación de productos	<i>Shigella sonnei S5046</i>
AcrF	proteína F de resistencia a acriflavina	P24181	NINGUNO	expresión	exportación de productos	<i>Escherichia coli</i>
U11618	transportador de eflujo de múltiples fármacos	NP_682408.1	NINGUNO	expresión	exportación de productos	<i>Thermosynechococcus elongatus BP-1</i>
U11619	transportador de eflujo de múltiples fármacos	NP_682409.1	NINGUNO	expresión	exportación de productos	<i>Thermosynechococcus elongatus BP-1</i>
U10139	transportador de eflujo de múltiples fármacos	NP_680930.1	NINGUNO	expresión	exportación de productos	<i>Thermosynechococcus elongatus BP-1</i>

genes de punto de control de la replicación									
umuD	ADN polimerasa V, subunidad	YP_310132	3.4.21.-				Sobrexpresión	aumento de la eficacia de salida	<i>Shigella sonnei</i> Ss046
umuC	ADN polimerasa V, subunidad	ABC42261	2.7.7.7				Sobrexpresión	aumento de la eficacia de salida	<i>Escherichia coli</i>
NADH: NADPH transhidroge- nasa (subuni- dades alfa y beta) (pntA, pntB)									
		P07001_P0AB70	1.6.1.2				Expresión	aumento de la eficacia de salida	<i>Shigella flexneri</i>