

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 646 815**

51 Int. Cl.:

C12N 9/10	(2006.01)
C12N 9/02	(2006.01)
C12P 7/64	(2006.01)
C07K 14/21	(2006.01)
C12N 15/53	(2006.01)
C12N 15/54	(2006.01)
C12N 1/15	(2006.01)
C12N 1/21	(2006.01)
C10L 1/02	(2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

- 86 Fecha de presentación y número de la solicitud internacional: **20.05.2008 PCT/US2008/064274**
- 87 Fecha y número de publicación internacional: **04.12.2008 WO08147781**
- 96 Fecha de presentación y número de la solicitud europea: **20.05.2008 E 08769538 (3)**
- 97 Fecha y número de publicación de la concesión europea: **02.08.2017 EP 2160460**

54 Título: **Genes que producen hidrocarburos y métodos para su uso**

30 Prioridad:

22.05.2007 US 931370 P
25.05.2007 US 931939 P
25.07.2007 US 951944 P
24.09.2007 US 974810 P

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
18.12.2017

73 Titular/es:

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

72 Inventor/es:

FRIEDMAN, LISA y
DA COSTA, BERNARDO

74 Agente/Representante:

MILTENYI, Peter

Observaciones :

Véase nota informativa (Remarks, Remarques o Bemerkungen) en el folleto original publicado por la Oficina Europea de Patentes

ES 2 646 815 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

Genes que producen hidrocarburos y métodos para su uso

Antecedentes de la invención

5 Los desarrollos en la tecnología han ido acompañados por un aumento de la dependencia de las fuentes de combustible, y tales fuentes de combustible están haciéndose cada vez más limitadas y difíciles de adquirir. Teniendo lugar el quemado de combustibles fósiles a una tasa sin precedentes, es probable que la demanda de combustible mundial pronto sea mayor que los suministros de combustible actuales.

10 La mayoría de las fuentes de combustible producidas actualmente proceden de fuentes de petróleo o del procesamiento químico de aceites vegetales. Las fuentes de petróleo se enfrentan a varios problemas: son recursos no renovables, requieren un periodo de tiempo prolongado (años) para formarse, y la formación se restringe a ubicaciones distintas. A medida que el petróleo se consume como fuente de energía, los recursos de petróleo geológicos finalmente se agotarán.

15 Como resultado, se han dirigido esfuerzos hacia la utilización de fuentes de energía renovable, tales como luz solar, agua, viento y biomasa. El uso de biomasa para producir nuevas fuentes de combustible que no se derivan de fuentes de petróleo (es decir, biocombustible) ha surgido como opción alternativa. El biocombustible (por ejemplo, biodiesel) es un combustible biodegradable compuesto por ésteres y alcanos de cadena larga. El biodiesel puede usarse en la mayor parte de los motores diesel de combustión interna o bien en forma pura, que se denomina biodiesel "puro", o bien como mezcla en cualquier concentración con diesel de petróleo normal.

20 En teoría, el biocombustible puede producirse a partir de cualquier fuente de carbono biológica. La fuente de carbono biológica más común, de lejos, son plantas fotosintéticas que capturan energía solar. Se usan muchas plantas diferentes y materiales derivados de plantas para fabricar biocombustibles. Uno de los mayores desafíos técnicos es desarrollar modos para convertir la energía de biomasa específicamente en combustibles líquidos para su transporte. Una estrategia frecuentemente usada para producir biocombustible es cultivar cultivos azucareros (por ejemplo, caña de azúcar o remolacha azucarera) o cultivos de almidón (por ejemplo, maíz) y luego usar la fermentación con levaduras para producir etanol (es decir, alcohol etílico). Otra estrategia frecuentemente usada para producir biocombustible es cultivar plantas que producen de manera natural aceites, tales como palma de aceite, soja o jatrofa. Una fuente alternativa de aceite producido de manera natural es a partir de organismos, tales como algas. Cuando se calientan estos aceites, su viscosidad se reduce, y pueden quemarse directamente en un motor diesel. Alternativamente, los aceites pueden procesarse químicamente para producir combustibles, tales como biodiesel. Los métodos actuales de producción de biodiesel implican la transesterificación de triacilglicéridos (por ejemplo, aceite vegetal o grasa animal) que conduce a una mezcla de ésteres grasos y el producto secundario no deseado glicerina. Esto da como resultado un producto que es heterogéneo y un producto de desecho, lo que conduce a ineficacias económicas.

35 Los aceites vegetales siguen siendo una alternativa atractiva a los combustibles fósiles porque son recursos renovables. Sin embargo, los aceites vegetales son una parte importante de la cadena alimentaria. Es improbable que puedan cultivarse suficientes cultivos como para satisfacer las necesidades de tanto alimentos como productos químicos industriales, tales como combustibles y polímeros. Además, las plantas productoras de aceites pueden estar restringidas por las condiciones medioambientales en las que florecen. Además, los aceites vegetales no son hidrocarburos (por ejemplo, alcanos o alquenos). Más bien, los aceites vegetales son principalmente triglicéridos, que contienen moléculas de oxígeno, que cuando se queman en un motor de combustión coqueficarán el motor.

40 En vista de lo anterior, sería deseable permitir la producción de combustibles a partir de fuentes alternativas con el fin de proporcionar un método mejorado de producción de biocombustibles. La invención se refiere a secuencias de ácido nucleico que codifican para polipéptidos que están implicados en la biosíntesis de hidrocarburos y productos intermedios de hidrocarburos. Además, la invención se refiere a métodos de uso de los mismos para la producción de biocombustibles. La invención descrita en el presente documento supera los problemas asociados con recursos de hidrocarburos no renovables, limitados y proporciona métodos mejorados que pueden usarse para producir biocombustibles. Estas y otras ventajas de la invención resultarán evidentes a partir de la descripción detallada proporcionada en el presente documento.

45 El documento US 2003/0233675 A1 se refiere a la expresión de proteínas microbianas en plantas para la producción de plantas con propiedades mejoradas y describe un polipéptido (SEQ ID NO: 15607) que muestra en algún grado identidad de secuencia con SEQ ID NO: 2 de la presente divulgación. Crossman *et al.* (Genome Biology, Biomed Central Ltd., Londres, 17 de abril de 2008, vol. 9, n.º 4, página R74) describen el análisis de genoma completo, comparativo y funcional de *Stenotrophomonas maltophilia*. Palleroni y Bradbury (Int. J. of Systematic Bacteriology, 31 de julio de 1993, págs. 606-609) describen *Stenotrophomonas* como un nuevo nombre de género de bacterias para *Xanthomonas maltophilia*, para restaurar de ese modo el género *Xanthomonas* a su definición anterior. Da Silva *et al.* (Nature, Nature Publishing Group, Londres, 23 de mayo de 2002, vol. 417, n.º 6887, págs. 459-463) describen una comparación de los genomas de dos patógenos de *Xanthomonas* con especificidades de huésped diferentes. Allen y Bartlett (Microbiology, 2002, vol. 148, págs. 1903-1913) describen la estructura y regulación de los genes de

ácido graso poliinsaturado omega-3 sintasa de la cepa de bacteria de aguas profundas *Photobacterium profundum* SS9. Kalscheuer *et al.* (Microbiology, 2006, vol. 152, págs. 2529-2536) se refieren a la producción de microdiesel mediante *Escherichia coli* modificada genéticamente. El número de registro de la base de datos EBI A0H249 da a conocer una secuencia de polipéptido de *Chloroflexus aggregans* DSM 9485 relacionada con SEQ ID NO: 18 de la presente divulgación. El número de registro de la base de datos EBI Q6FTD0 da a conocer una secuencia de polipéptido de *Candida glabrata* relacionada con SEQ ID NO: 64 de la presente divulgación. El número de registro de la base de datos EBI Q8PQT8 da a conocer una secuencia de polipéptido de *Xanthomonas axonopodis* relacionada con SEQ ID NO: 12 de la presente divulgación.

Breve resumen de la invención

10 La presente invención se define mediante las reivindicaciones.

La invención se refiere a un ácido nucleico aislado que codifica para un polipéptido que comprende una secuencia motivo de aminoácido de OleA, OleB, OleC u OleD.

15 La invención se refiere a un ácido nucleico aislado que codifica para un polipéptido que comprende una secuencia de aminoácidos seleccionada del grupo que consiste en (a) una secuencia de aminoácidos de OleA, OleB, OleC, OleD u OleBC; (b) un homólogo de una secuencia de aminoácidos de OleA, OleB, OleC, OleD u OleBC; (c) una variante conservativa de una secuencia de aminoácidos de OleA, OleB, OleC, OleD u OleBC que comprende una o más sustituciones de aminoácidos conservados; y (d) una secuencia de aminoácidos que tiene una identidad de secuencia de al menos el 35% con una secuencia de aminoácidos de OleA, OleB, OleC, OleD u OleBC.

20 La invención se refiere a un método para producir una cetona alifática que comprende incubar un sustrato con OleA, en condiciones suficientes para producir una cetona alifática. La invención también se refiere a un método para producir un hidrocarburo que comprende incubar un sustrato con OleA, OleB, OleC, OleBC, OleD, o una combinación de los mismos en condiciones suficientes para producir un hidrocarburo.

25 La invención se refiere a un polipéptido aislado que comprende una secuencia de aminoácidos que codifica para OleA, en el que la secuencia de aminoácidos comprende una o más sustituciones, adiciones, inserciones o deleciones de aminoácidos.

La invención se refiere a un polipéptido aislado que comprende una secuencia de aminoácidos que codifica para OleC, en el que la secuencia de aminoácidos comprende una o más sustituciones, adiciones, inserciones o deleciones de aminoácidos.

30 La invención se refiere a un polipéptido aislado que comprende una secuencia de aminoácidos que codifica para OleD, en el que la secuencia de aminoácidos comprende una o más sustituciones, adiciones, inserciones o deleciones de aminoácidos.

La invención se refiere a un ácido nucleico aislado que codifica para un polipéptido que tiene la misma actividad biológica que un polipéptido que comprende la secuencia de aminoácidos que codifica para OleA, OleC u OleD.

35 La invención se refiere además a un organismo modificado genéticamente que comprende una secuencia exógena de ácido nucleico incorporada de manera estable en el genoma de un organismo en el sentido de 5' de una secuencia genómica de ácido nucleico que (a) tiene una identidad de secuencia de al menos el 35% con una secuencia de ácido nucleico que codifica para OleA, OleB, OleC u OleD y (b) codifica para un polipéptido.

40 La invención se refiere a un organismo modificado genéticamente preparado (a) proporcionando un organismo que tiene una secuencia de ácido nucleico que tiene una identidad de secuencia de al menos el 35% con OleC u OleD y (b) deleccionando o mutando la secuencia de ácido nucleico.

45 La invención también se refiere a un método para identificar una enzima útil para la producción de hidrocarburos que comprende transformar una célula que comprende polipéptidos seleccionados del grupo que consiste en (a) OleA y OleD, (b) OleA y OleC, y (c) OleC y OleD con un ácido nucleico que codifica para una enzima que se sospecha que tiene la capacidad para producir hidrocarburos; y determinar si la célula produce hidrocarburos, en el que la existencia de producción de hidrocarburos por la célula indica que el ácido nucleico codifica para un polipéptido útil para la producción de hidrocarburos.

La invención se refiere a un hidrocarburo que tiene un $\delta^{13}\text{C}$ de -28 o más. La invención también se refiere a un hidrocarburo que tiene un pMC de al menos 50.

Breve descripción de las varias vistas de los dibujos

50 La figura 1 es un cromatograma de iones totales de hidrocarburos extraídos de *E. coli* C41(DE3) ΔfadE que produce las proteínas OleA (SEQ ID NO: 2), OleC (SEQ ID NO: 6) y OleD (SEQ ID NO: 8).

Las figuras 2A-2J son una serie de diez espectros de EM para los hidrocarburos mostrados en la figura 1. La figura 2A muestra los datos de espectros de EM para trieno C27. La figura 2B muestra los datos de espectros de EM para

dieno C27. La figura 2C muestra los datos de espectros de EM para monoeno C27. La figura 2D muestra los datos de espectros de EM para dieno C28. La figura 2E muestra los datos de espectros de EM para trieno C29. La figura 2F muestra los datos de espectros de EM para dieno C29. La figura 2G muestra los datos de espectros de EM para monoeno C29. La figura 2H muestra los datos de espectros de EM para dieno C30. La figura 2I muestra los datos de espectros de EM para trieno C31. La figura 2J muestra los datos de espectros de EM para dieno C31.

La figura 3 es un cromatograma de iones totales de hidrocarburos extraídos de *S. maltophilia*.

Las figuras 4A-4E son una serie de cromatogramas de iones totales de extracciones con metanol:hexano de diversos cultivos de *Stenotrophomonas maltophilia*. La figura 4A muestra olefinas detectadas en el extracto de *S. maltophilia* ATCC 17679 silvestre. La figura 4B muestra que no se detectaron picos principales en el extracto de la cepa de *S. maltophilia* $\Delta oleA$ (delección de *oleA*). La figura 4C muestra olefinas detectadas en el extracto de *S. maltophilia* ATCC 17679. La figura 4D muestra cetonas alifáticas detectadas en los extractos de *S. maltophilia* $\Delta oleC$ (delección de *oleC*). La figura 4E muestra las cetonas alifáticas detectadas en los extractos de *S. maltophilia* $\Delta oleD$ (delección de *oleD*).

Las figuras 5A-5C son una serie de cromatogramas de iones totales de cetonas alifáticas extraídas de *E. coli* C41(DE3) que produce la proteína OleA (SEQ ID NO: 2). La figura 5A muestra cetonas alifáticas extraídas de *E. coli* C41(DE3) que produce la proteína OleA (SEQ ID NO: 2). La figura 5B muestra el extracto de la figura 5A superpuesto con 16-hentriacontano (cetona C31 saturada). La figura 5C muestra el extracto de la figura 5A superpuesto con 14-heptacosanona (cetona C27 saturada).

Las figuras 6A-6H son una serie de espectros de EM para cada cetona alifática mostrada en la figura 5A. La figura 6A muestra una cetona C27:2 (con dos enlaces insaturados). La figura 6B muestra una cetona C27:1 (con un doble enlace). La figura 6C muestra una cetona C27 (saturada). La figura 6D muestra una cetona C29:2 (con dos enlaces insaturados). La figura 6E muestra una cetona C29:1 (con un doble enlace). La figura 6F muestra una cetona C29 (saturada). La figura 6G muestra una cetona C31:2 (con dos enlaces insaturados). La figura 6H muestra una cetona C31:1 (con un doble enlace).

Las figuras 7A-7C demuestran que se detectan cetonas alifáticas en un ensayo *in vitro* que combina lisado de células de *E. coli* que expresan *oleA* (SEQ ID NO: 1) con sustratos de acil-CoA. La figura 7A es un cromatograma de iones totales de un extracto de una muestra que contiene lisado de OleA con un sustrato de miristoil-CoA. La figura 7B es un cromatograma de iones totales del extracto de control que contiene lisado de OleA sin sustrato de miristoil-CoA añadido. La figura 7C es un gráfico que compara la cantidad de cetona encontrada en cada una de las muestras. El eje y representa el área bajo la curva y el eje x representa las cetonas alifáticas indicadas. Las barras sombreadas representan la muestra con extracto celular de OleA solo y las barras negras representan extracto celular de OleA combinado con sustrato de miristoil-CoA.

Las figuras 8A y 8B representan datos de CG/EM de los extractos de un ensayo *in vitro* que contiene proteína OleA purificada, sustrato de miristoil-CoA y lisado celular de *E. coli* C41(DE3). La figura 8A es un cromatograma de iones totales que muestra un pico que eluye al tiempo de retención de 13,95 minutos en el cromatograma de iones totales, indicando la presencia de cetona alifática C27 saturada. La figura 8B representa los espectros de EM para la cetona alifática C27.

La figura 9 es un gráfico del área bajo la curva de CG para cetonas alifáticas que se detectaron en extractos de fermentaciones de *E. coli* C41(DE3) $\Delta fadE$ que contenía diversos vectores para la expresión de secuencias de proteína *oleA*. Las barras negras representan la cantidad de cetona alifática observada en extractos de una cepa que expresa un plásmido que contiene *oleA* basándose en el genoma de *Chloroflexus aggregans*, las barras con rayado diagonal representan la cantidad de cetona alifática observada en extractos de una cepa que expresa un plásmido que contiene *oleA* basándose en el genoma de *Xanthomonas axonopodis*, las barras blancas representan la cantidad de cetona alifática observada en extractos de una cepa que expresa un plásmido que contiene la secuencia de nucleótidos nativa que codifica para el polipéptido de OleA de *Stenotrophomonas maltophilia* ATCC17679, y las barras sombreadas representan la cantidad de cetona alifática observada en extractos de una cepa que expresa un plásmido que contiene una secuencia de nucleótidos sintética, con codones optimizados que codifica para el polipéptido de OleA de *Stenotrophomonas maltophilia* R551-3 (*S. maltophilia* CO). *S. maltophilia* CO se refiere a la expresión de un ADN sintético con codones optimizados que codifica para la secuencia de aminoácidos de OleA de *S. maltophilia* R551-3 (SEQ ID NO: 4).

La figura 10 es un gráfico que muestra el aumento en la producción de cetonas alifáticas que resultó debido a la expresión de *oleA* en las cepas de *E. coli* C41(DE3) con alteraciones en *fadE*, *fadD*, y/o *tesA*. Las barras blancas representan cetona C27, las barras sombreadas representan cetona C29, y las barras negras representan cetona C31. Un signo (-) indica una desactivación de la actividad de un gen y un signo (+) indica sobreexpresión del gen de interés.

La figura 11 es un gráfico que muestra el aumento en la producción de olefinas que resultó debido a la expresión de *oleA*, *oleC* y *oleD* en las cepas de *E. coli* C41(DE3) con alteraciones en *fadE*, *fadD* y/o *tesA*. Las barras blancas representan olefina C27, las barras sombreadas representan olefina C29 y las barras negras representan olefina

C31. Un signo (-) indica una desactivación de la actividad del gen y un signo (+) indica sobreexpresión del gen de interés.

5 Las figuras 12A-12E son gráficos que muestran la observación de tetradecanal, 14-heptacosanona e isómeros de heptacosano mediante cromatografía CG/EM del extracto orgánico de las reacciones *in vitro* con las enzimas OleA y OleD combinadas con miristoil coenzima A. La figura 12A es un cromatograma de iones totales de tetradecanal. La figura 12B es un cromatograma de iones totales con 2 picos que indican isómeros de heptacosano. La figura 12C es un espectro de EM del pico 1 de la figura 12B. La figura 12D es un espectro de EM del pico 2 de la figura 12B. La figura 12E es un cromatograma de iones totales de 14-heptacosanona.

10 Las figuras 13A-13C son una serie de cromatogramas de iones totales que representan la observación de tetradecanal, 14-heptacosanona e isómeros de heptacosano mediante cromatografía CG/EM del extracto orgánico de dos reacciones *in vitro* de enzimas Ole combinadas con miristoil coenzima A. Una reacción comprendía las enzimas OleA y OleD, y la otra comprendía las enzimas OleA, OleD y OleB. La figura 13A es un cromatograma de iones totales de tetradecanal. La figura 13B es un cromatograma de iones totales de 2 picos que indican isómeros de heptacosano. La figura 13C es un cromatograma de iones totales de 14-heptacosanona. Las flechas indican la reacción OleA + OleD + OleB y la reacción de OleA + OleD. La presencia de OleB potencia la formación de tetradecanal y heptacosano y disminuye los niveles de 14-heptacosanona.

15 La figura 14 es un cromatograma de iones totales que demuestra que la levadura *S. cerevisiae* puede producir OleA funcional. La figura 14 muestra un cromatograma de iones totales del extracto de una reacción de lisado celular de *S. cerevisiae* pESC-HIS-OleA con la adición de miristoil coenzima A. Se observan ácido mirístico, 14-heptacosanona y la adición conocida de control de hexacosano.

20 Las figuras 15A-15H son una serie de cromatogramas de iones totales y espectros de EM que demuestran que se observan olefinas y aldehídos deuterados en el extracto orgánico de un ensayo *in vitro* que contiene OleA purificada, lisado celular que contiene OleD y NADPH deuterada. La figura 15A es un cromatograma de iones totales de tetradecanal observado tras la extracción de una reacción *in vitro* que comprende OleA, lisado celular que contiene OleD, miristoil coenzima A y R-(4-²H)NADPH. La figura 15B es un espectro de EM del tetradecanal de la figura 15A. La figura 15C es un cromatograma de iones totales de tetradecanal observado tras la extracción de una reacción *in vitro* que comprende OleA, lisado celular que contiene OleD, miristoil coenzima A y S-(4-²H)NADPH. La figura 15D es un espectro de EM del tetradecanal de la figura 15C. La figura 15E es un cromatograma de iones totales de isómeros de heptacosano formados mediante la reacción de OleA y lisado celular que contiene OleD con R-(4-²H)NADPH. La figura 15F es un espectro de EM del isómero de heptacosano de la figura 15E. La figura 15G es un cromatograma de iones totales de isómeros de heptacosano observado tras la extracción de una reacción *in vitro* que comprende OleA, lisado celular que contiene OleD, miristoil coenzima A y S-(4-²H)NADPH. La figura 15H es un espectro de EM del heptacosano de la figura 15G.

25 Las figuras 16A-16C son una serie de cromatogramas de iones totales que demuestran la capacidad para sintetizar olefinas alifáticas *in vitro*. La figura 16A es un cromatograma de iones totales del extracto de una reacción realizada con un lisado celular preparado a partir de la cepa de *E. coli* C41(DE3) Δ fadE con plásmidos para la expresión de *oleA*, *oleB*, *oleC* y *oleD* con miristoil coenzima A. Las flechas indican heptacosanos C27:1, C27:2 y C27:3. La figura 16B es un cromatograma de iones totales del extracto de una reacción realizada con un lisado celular preparado a partir de la cepa de *E. coli* C41(DE3) Δ fadE con plásmidos para la expresión de *oleA*, *oleB*, *oleC* y *oleD* con la adición de miristoil coenzima A. Las flechas indican heptacosanos C27:1, C27:2 y C27:3. La figura 16C es un cromatograma de iones totales del extracto de una reacción realizada con un lisado celular preparado a partir de la cepa de *E. coli* C41(DE3) Δ fadE con plásmidos para la expresión de *oleA*, *oleB*, *oleC* y *oleD* con la adición de miristoil-ACP. Las flechas indican heptacosanos C27:1, C27:2 y C27:3.

30 Las figuras 17A-17J son una serie de cromatogramas de iones totales y espectros de EM de iones seleccionados para hidrocarburos extraídos de fermentaciones de *B. megaterium* WH320; pWH1520_OleCDAB. La figura 17A es un cromatograma de iones totales de la olefina de 25 carbonos que tiene un ion parental de 350. La figura 17B es un espectro de masas de los compuestos que eluyen a un tiempo de retención 3,092 minutos y es característico de una olefina monoinsaturada C25 ramificada. La figura 17C es un cromatograma de iones totales de la olefina de 26 carbonos que tiene un ion parental de 364. La figura 17D es el espectro de masas de los compuestos que eluyen a un tiempo de retención 3,390 minutos y es característico de una olefina monoinsaturada C26 ramificada. La figura 17E es un cromatograma de iones totales de la olefina de 27 carbonos que tiene un ion parental de 378. La figura 17F es un espectro de masas de los compuestos que eluyen a un tiempo de retención 3,567 minutos y es característico de una olefina monoinsaturada C27 ramificada. La figura 17G es un cromatograma de iones totales de la olefina de 28 carbonos que tiene un ion parental de 392. La figura 17H es un espectro de masas de compuestos que eluyen a un tiempo de retención 3,893 minutos y es característico de una olefina monoinsaturada C28 ramificada. La figura 17I es un cromatograma de iones totales de la olefina de 29 carbonos que tiene un ion parental de 406. La figura 17J es un espectro de masas de compuestos que eluyen a un tiempo de retención 4,013 minutos y es característico de una olefina monoinsaturada C29 ramificada.

35 Figuras 18A-18E son una serie de cromatogramas de iones totales y espectros de EM de extractos de *B. subtilis*. La figura 18A es un cromatograma de iones totales de cetonas alifáticas extraídas de *B. subtilis* IHA01; pHT01_OleA.

La figura 18B es un cromatograma de iones totales del extracto de *B. subtilis* IHA01; pHT01 que no produce ningún hidrocarburo alifático. La figura 18C es un espectro de masas de compuestos que eluyen a un tiempo de retención de 3,584 y es característico de cetonas C25 alifáticas ramificadas. La figura 18D es un espectro de masas de compuestos que eluyen a un tiempo de retención de 4,036 y es característico de cetonas C27 alifáticas ramificadas. La figura 18E es un espectro de masas de compuestos que eluyen a un tiempo de retención de 4,485 y es característico de cetonas C29 alifáticas ramificadas.

Las figuras 19A-19B son un cromatograma de iones totales y espectros de masas de extractos de *Arthrobacter aureescens* TC1. La figura 19A es un cromatograma de iones totales de hidrocarburos extraídos de *Arthrobacter aureescens* TC1. Los tres picos indican diversas versiones ramificadas de olefina C29 monoinsaturada. La figura 19B es un espectro de masas de compuestos que eluyen a un tiempo de retención 13,6 minutos y es característico de olefina C29 monoinsaturada.

La figura 20 es un diagrama de las rutas propuestas para la biosíntesis de ácido micólico y la síntesis de olefinas usando OleA, OleB, OleC y OleD. X puede ser o bien CoA o bien ACP.

La figura 21 es un diagrama de una ruta propuesta para la producción de olefinas de cadena larga en organismos productores de ácido micólico que expresan *oleC* y *oleD*.

La figura 22 es una ruta propuesta para la producción de olefinas de cadena larga en organismos productores de ácido micólico que expresan *oleC*.

Descripción detallada de la invención

La invención se basa, al menos en parte, en el descubrimiento de varios genes que codifican para proteínas que están implicadas en la biosíntesis de hidrocarburos. Pueden usarse organismos transformados con una o más secuencias de ácido nucleico que codifican para una o más de las proteínas identificadas en el presente documento para generar hidrocarburos, tales como olefinas, y productos intermedios de hidrocarburos, tales como cetonas alifáticas. Los términos "olefina" y "alqueno" se usan de manera intercambiable en el presente documento. Tal como se describe en el presente documento, estas secuencias de ácido nucleico se han identificado como genes que están implicados en la biosíntesis de hidrocarburos. Por tanto, pueden usarse células que se han transformado con uno o más de estos genes como fuente para la producción de hidrocarburos, incluyendo olefinas, y sus precursores (por ejemplo, cetonas alifáticas). Este descubrimiento proporciona una fuente para hidrocarburos que pueden usarse como combustible en lugar de recursos de hidrocarburos no renovables, limitados (por ejemplo, combustibles a base de petróleo). Además, permite la producción de una amplia gama de productos de olefinas y cetonas alifáticas específicos diseñados para aplicaciones particulares. Controlando el organismo huésped y/o los sustratos de reacción (por ejemplo, controlando la longitud de cadena, ramificación, saturación y/o la ubicación de dobles enlaces), pueden crearse organismos que producen una amplia gama de productos de hidrocarburos, incluyendo los que tienen ramificaciones o puntos de insaturación particulares.

Se identificaron cuatro genes en *Stenotrophomonas maltophilia* que codifican cada uno para una proteína implicada en la biosíntesis de hidrocarburos, tales como olefinas, y productos intermedios de hidrocarburos, tales como cetonas alifáticas. Estos cuatro genes se denominan *oleA*, *oleB*, *oleC* y *oleD*, y las proteínas codificadas por estos genes se denominan OleA, OleB, OleC y OleD, respectivamente. Además, estos genes también se han encontrado en parejas dentro de otros organismos. Este apareamiento incluye una fusión entre *oleB* y *oleC*, en el presente documento denominada *oleBC*, y la proteína codificada por este gen se denomina OleBC. Conjuntamente, estas cinco proteínas representan una familia de proteínas Ole que están implicadas en la biosíntesis de hidrocarburos y productos intermedios de hidrocarburos. Individualmente, OleA se refiere a una familia de proteínas con actividad de OleA, OleB se refiere a una familia de proteínas con actividad de OleB, OleC se refiere a una familia de proteínas con actividad de OleC, OleBC se refiere a una familia de proteínas con actividad de OleBC y OleD se refiere a una familia de proteínas con actividad de OleD. Un experto habitual en la técnica apreciará que usando la información proporcionada en el presente documento referente a la estructura y función de secuencias génicas de *ole* de *S. maltophilia*, pueden obtenerse otras secuencias génicas que codifican para proteínas que tienen actividad similar.

Dadas estas enseñanzas, un experto habitual en la técnica apreciará que pueden clonarse fácilmente secuencias de *oleA*, *oleB*, *oleC*, *oleBC* y *oleD* adicionales y usarse para producir hidrocarburos y productos intermedios de hidrocarburos. Por tanto, a lo largo de toda esta descripción, la referencia a proteínas Ole debe entenderse que significa todas las proteínas que presentan actividad similar a la de cualquiera de las proteínas de la familia Ole, incluyendo OleA, OleB, OleC, OleBC y OleD. De manera similar, la referencia a OleA, OleB, OleC, OleBC u OleD debe entenderse que significa todas las proteínas que presentan la actividad respectiva de cada proteína Ole, incluyendo, por ejemplo, todas las proteínas OleA, OleB, OleC, OleBC y OleD enumeradas en las tablas 1 y 2, así como otras proteínas Ole que pueden identificarse o modificarse por ingeniería genética a través de diversos métodos bioinformáticos o técnicas moleculares, tales como unión de anticuerpos, hibridación de ácidos nucleicos, PCR, y otros métodos adecuados.

Además, a lo largo de toda esta descripción, la referencia a genes *ole* debe entenderse que significa todos los genes que codifican para proteínas que presentan una actividad similar a la de cualquiera de las proteínas de la familia Ole,

incluyendo OleA, OleB, OleC, OleBC y OleD. De manera similar, la referencia a *oleA*, *oleB*, *oleC*, *oleBC* u *oleD* debe entenderse que significa todos los genes que codifican para proteínas que presentan la actividad respectiva de cada proteína Ole, incluyendo, por ejemplo, todas las proteínas OleA, OleB, OleC, OleBC y OleD enumeradas en las tablas 1 y 2, así como otras proteínas Ole que pueden identificarse o modificarse por ingeniería genética a través de diversos métodos bioinformáticos o técnicas moleculares, tales como unión de anticuerpos, hibridación de ácidos nucleicos, PCR, y otros métodos adecuados.

La invención se refiere ácidos nucleicos aislados que codifican para proteínas Ole. El término “aislado” con respecto a ácidos nucleicos se refiere a la retirada de un ácido nucleico de su entorno natural. Además, un “ácido nucleico aislado” pretende incluir fragmentos de ácido nucleico, que no se producen de manera natural como fragmentos y no se encontrarían en la naturaleza. El término “aislado” también se usa en el presente documento para referirse a polipéptidos, que están aislados isolated de otras proteínas celulares y pretende abarcar tanto polipéptidos purificados como recombinantes. El término “aislado” tal como se usa en el presente documento también se refiere a un ácido nucleico o péptido 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 péptido que está sustancialmente libre de precursores químicos u otros productos químicos cuando se sintetiza químicamente.

Los términos “ácido nucleico” y “secuencia de ácido nucleico” pretenden abarcar un polímero de ADN o ARN (es decir, un polinucleótido), que puede ser monocatenario o bicatenario y que puede contener nucleótidos no naturales o alterados. Debe entenderse que los términos incluyen, como equivalentes, análogos de o bien ARN o bien ADN preparado a partir de análogos nucleotídicos y, según sea aplicable a la realización que está describiéndose, polinucleótidos mono (sentido o antisentido) y bicatenarios, EST, cromosomas, ADNc, ARNm y ARNr.

Las tablas 1 y 2 proporcionan una lista a modo de ejemplo de proteínas Ole. En la tabla 1, #/#* indica que el organismo contiene dos homólogos con las identidades respectivas. Los organismos en negrita tenían o bien una actividad génica demostrada o bien una producción de olefinas demostradas. El porcentaje de identidad se determinó en comparación con la secuencia de aminoácidos de OleA, OleB, OleC u OleD (tal como se indica en las tablas 1 y 2) de *S. maltophilia* ATCC17679 tal como se calcula mediante el software BLAST™ fijado a parámetros por defecto. Por ejemplo, puede usarse el software blastn (versión 2.0) para determinar la identidad de secuencia entre dos secuencias de ácido nucleico usando parámetros por defecto (expectativa = 10, matriz = BLOSUM62, filtro = DUST (Tatusov y Lipmann, datos sin publicar; y Hancock y Armstrong, *Comput. Appl. Biosci.*, 10: 67-70 (1994)), coste de existencia de huecos = 11, coste de hueco por residuo = 1 y razón lambda = 0,85). Para la comparación de dos polipéptidos, puede usarse el software blastp (versión 2.0) con parámetros por defecto (expectativa 10, filtro = SEG (Wootton *et al.*, *Computers in Chemistry*, 17: 149-163 (1993)), matriz = BLOSUM62, coste de existencia de huecos = 11, coste de hueco por residuo = 1, lambda = 0,85).

TABLA 1

Organismo	Proteína	SEQ ID NO:	Porcentaje de identidad
<i>Stenotrophomonas maltophilia</i> ATCC17679	OleA	SEQ ID NO: 152	100
<i>Stenotrophomonas maltophilia</i> R551-3	OleA	SEQ ID NO: 150	99
<i>Stenotrophomonas maltophilia</i> K279a	OleA	SEQ ID NO: 151	99
<i>Xanthomonas campestris</i> pv. <i>campestris</i> cepa 8004	OleA	SEQ ID NO: 155	88
<i>Xanthomonas campestris</i> pv. <i>campestris</i> cepa ATCC 33913	OleA	SEQ ID NO: 157	88
<i>Xanthomonas campestris</i> pv. <i>campestris</i> cepa B100	OleA	SEQ ID NO: 154	88
<i>Xanthomonas campestris</i> pv. <i>vesicatoria</i> cepa 85-10	OleA	SEQ ID NO: 153	88
<i>Xanthomonas axonopodis</i> pv. <i>citri</i> cepa 306	OleA	SEQ ID NO: 158	87
<i>Xanthomonas oryzae</i> pv. <i>oryzae</i> KACC10331	OleA	SEQ ID NO: 160	87
<i>Xanthomonas oryzae</i> pv. <i>oryzae</i> MAFF 311018	OleA	SEQ ID NO: 156	87
<i>Xanthomonas oryzae</i> pv. <i>oryzicola</i> BLS256	OleA	SEQ ID NO: 159	87
<i>Xylella fastidiosa</i> Dixon	OleA	SEQ ID NO: 162	78
<i>Xylella fastidiosa</i> 9a5c	OleA	SEQ ID NO: 163	78
<i>Xylella fastidiosa</i> Ann-1	OleA	SEQ ID NO: 478	78
<i>Xylella fastidiosa</i> M12	OleA	SEQ ID NO: 161	78
<i>Xylella fastidiosa</i> M23	OleA	SEQ ID NO: 229	78
<i>Xylella fastidiosa</i> Temecula1	OleA	SEQ ID NO: 479	78
<i>Plesiocystis pacifica</i> SIR-1	OleA	SEQ ID NO: 164	48
<i>Chloroflexus aggregans</i> DSM 9485	OleA	SEQ ID NO: 165	45
<i>Arthrobacter chlorophenolicus</i> A6	OleA	SEQ ID NO: 224	45
<i>Chloroflexus aurantiacus</i> J-10-fl	OleA	SEQ ID NO: 166	44
<i>Clavibacter michiganensis</i> subsp. <i>Sepedonicus</i>	OleA	SEQ ID NO: 199	44
<i>Clavibacter michiganensis</i> subsp. <i>michiganensis</i> NCPPB 382	OleA	SEQ ID NO: 189	44

ES 2 646 815 T3

Arthrobacter aurescens TC1	OleA	SEQ ID NO: 168	44
<i>Brevibacterium linens</i> BL2	OleA	SEQ ID NO: 204	43
<i>Desulfatibacillum alkenivorans</i> AK-01	OleA	SEQ ID NO: 217	42
<i>Congregibacter litoralis</i> KT71	OleA	SEQ ID NO: 206	40
Kineococcus radiotolerans SRS30216	OleA	SEQ ID NO: 192	39
Micrococcus luteus NCTC 2665	OleA	SEQ ID NO: 225	38
<i>Geobacter uraniumreducens</i> Rf4	OleA	SEQ ID NO: 190	36
<i>Pelobacter propionicus</i> DSM 2379	OleA	SEQ ID NO: 182	36
<i>Opitutus terrae</i> PB90-1	OleA	SEQ ID NO: 203	36
<i>Desulfotalea psychrophila</i> LSV54	OleA	SEQ ID NO: 174	36
<i>Geobacter bemidjensis</i> Bem	OleA	SEQ ID NO: 212	35
<i>Geobacter lovleyi</i> SZ	OleA	SEQ ID NO: 211	35
<i>Shewanella benthica</i> KT99	OleA	SEQ ID NO: 218	35
<i>Shewanella loihica</i> PV-4	OleA	SEQ ID NO: 188	35
<i>Photobacterium profundum</i> SS9	OleA	SEQ ID NO: 169	35
<i>Shewanella baltica</i> OS155	OleA	SEQ ID NO: 186	35
<i>Shewanella baltica</i> OS185	OleA	SEQ ID NO: 480	35
<i>Shewanella baltica</i> OS195	OleA	SEQ ID NO: 197	35
<i>Shewanella baltica</i> OS223	OleA	SEQ ID NO: 481	35
<i>Shewanella</i> sp. ANA-3	OleA	SEQ ID NO: 181	34
<i>Desulfococcus oleovorans</i> Hxd3	OleA	SEQ ID NO: 196	34
<i>Shewanella amazonensis</i> SB2B	OleA	SEQ ID NO: 183	34
<i>Shewanella frigidimarina</i> NCIMB 400	OleA	SEQ ID NO: 179	34
<i>Shewanella woodyi</i> ATCC 51908	OleA	SEQ ID NO: 200	34
<i>Photobacterium profundum</i> 3TCK	OleA	SEQ ID NO: 208	34
<i>Moritella</i> sp. PE36	OleA	SEQ ID NO: 215	34
<i>Shewanella denitrificans</i> OS217	OleA	SEQ ID NO: 176	34
<i>Psychromonas ingrahamii</i> 37	OleA	SEQ ID NO: 184	34
<i>Shewanella pealeana</i> ATCC 700345	OleA	SEQ ID NO: 195	34
<i>Shewanella putrefaciens</i> 200	OleA	SEQ ID NO: 482	34
<i>Shewanella putrefaciens</i> CN-32	OleA	SEQ ID NO: 483	34
<i>Shewanella sediminis</i> HAW-EB3	OleA	SEQ ID NO: 194	34
<i>Shewanella</i> sp. W3-18-1	OleA	SEQ ID NO: 185	34
<i>Shewanella halifaxensis</i> HAW-EB4	OleA	SEQ ID NO: 198	34
<i>Shewanella oneidensis</i> MR-1	OleA	SEQ ID NO: 172	34
<i>Shewanella</i> sp. MR-7	OleA	SEQ ID NO: 178	34
<i>Colwellia psychrerythraea</i> 34H	OleA	SEQ ID NO: 175	34
<i>Shewanella</i> sp. MR-4	OleA	SEQ ID NO: 177	34
<i>Geobacter</i> sp. FRC-32	OleA	SEQ ID NO: 210	33
<i>Blastopirellula marina</i> DSM 3645	OleA	SEQ ID NO: 205	33
<i>Planctomyces maris</i> DSM 8797	OleA	SEQ ID NO: 213	33
<i>Rhodopirellula baltica</i> SH 1	OleA	SEQ ID NO: 173	33
<i>Lentisphaera araneosa</i> HTCC2155	OleA	SEQ ID NO: 214	33
<i>Desulfuromonas acetoxidans</i> DSM 684	OleA	SEQ ID NO: 209	32
<i>Gemmata obscuriglobus</i> UQM 2246	OleA	SEQ ID NO: 223	30
<i>Psychromonas</i> sp. CNPT3	OleA	SEQ ID NO: 207	30
<i>Streptomyces ambofaciens</i> ATCC 23877	OleA	SEQ ID NO: 170	29
<i>Opitutaceae bacterium</i> TAV2	OleA	SEQ ID NO: 216	37
<i>Arthrobacter</i> sp. FB24	OleA	SEQ ID NO: 167	27
<i>Burkholderia oklahomensis</i> C6786	OleA	SEQ ID NO: 220	27
<i>Burkholderia oklahomensis</i> EO147	OleA	SEQ ID NO: 219	27
<i>Clostridium botulinum</i> A cepa ATCC 3502	OleA	SEQ ID NO: 191	27
<i>Clostridium botulinum</i> A3 cepa Loch Maree	OleA	SEQ ID NO: 202	27
<i>Clostridium botulinum</i> B1 cepa Okra	OleA	SEQ ID NO: 201	27
<i>Clostridium botulinum</i> F cepa Langeland	OleA	SEQ ID NO: 193	27
<i>Maricaulis maris</i> MCS10	OleA	SEQ ID NO: 180	27
<i>Streptomyces ambofaciens</i> DSM40697	OleA	SEQ ID NO: 171	27
<i>Candidatus Kuenenia stuttgartiensis</i>	OleA	SEQ ID NO: 226	26
<i>Clostridium botulinum</i> Bf	OleA	SEQ ID NO: 222	26
<i>Clostridium botulinum</i> NCTC 2916	OleA	SEQ ID NO: 221	26
<i>Burkholderia pseudomallei</i> 1106a	OleA	SEQ ID NO: 187	25
<i>Neisseria gonorrhoeae</i> FA 1090	OleA	SEQ ID NO: 227	25
Stenotrophomonas maltophilia ATCC17679	OleB	SEQ ID NO: 302	100

ES 2 646 815 T3

<i>Stenotrophomonas maltophilia</i> R551-3	OleB	SEQ ID NO: 300	98
<i>Stenotrophomonas maltophilia</i> K279a	OleB	SEQ ID NO: 301	99
<i>Xanthomonas campestris</i> pv. <i>campestris</i> cepa 8004	OleB	SEQ ID NO: 303	78
<i>Xanthomonas campestris</i> pv. <i>campestris</i> cepa ATCC 33913	OleB	SEQ ID NO: 240	78
<i>Xanthomonas campestris</i> pv. <i>campestris</i> cepa B100	OleB	SEQ ID NO: 236	78
<i>Xanthomonas campestris</i> pv. <i>vesicatoria</i> cepa 85-10	OleB	SEQ ID NO: 253	80
<i>Xanthomonas axonopodis</i> pv. <i>citri</i> cepa 306	OleB	SEQ ID NO: 241	79
<i>Xanthomonas oryzae</i> pv. <i>oryzae</i> KACC10331	OleB	SEQ ID NO: 251	78
<i>Xanthomonas oryzae</i> pv. <i>oryzae</i> MAFF 311018	OleB	SEQ ID NO: 325	78
<i>Xanthomonas oryzae</i> pv. <i>oryzicola</i> BLS256	OleB	SEQ ID NO: 318	78
<i>Xylella fastidiosa</i> Dixon	OleB	SEQ ID NO: 287	72
<i>Xylella fastidiosa</i> 9a5c	OleB	SEQ ID NO: 239	72
<i>Xylella fastidiosa</i> Ann-1	OleB	SEQ ID NO: 288	72
<i>Xylella fastidiosa</i> M12	OleB	SEQ ID NO: 285	72
<i>Xylella fastidiosa</i> M23	OleB	SEQ ID NO: 326	72
<i>Xylella fastidiosa</i> Temecula 1	OleB	SEQ ID NO: 243	72
<i>Plesiocystis pacifica</i> SIR-1	OleB	SEQ ID NO: 314	52
<i>Chloroflexus aggregans</i> DSM 9485	OleB	SEQ ID NO: 297	45
<i>Arthrobacter chlorophenolicus</i> A6	OleB	SEQ ID NO: 323 y 324	46/36*
<i>Chloroflexus aurantiacus</i> J-10-fl	OleB	SEQ ID NO: 282	45
<i>Desulfatibacillum alkenivorans</i> AK-01	OleB	SEQ ID NO: 316	31
<i>Geobacter uraniumreducens</i> Rf4	OleB	SEQ ID NO: 274	53
<i>Pelobacter propionicus</i> DSM 2379	OleB	SEQ ID NO: 265	51
<i>Opitutus terrae</i> PB90-1	OleB	SEQ ID NO: 286	47
<i>Desulfotalea psychrophila</i> LSv54	OleB	SEQ ID NO: 249	39
<i>Geobacter bemidjensis</i> Bem	OleB	SEQ ID NO: 307	52
<i>Geobacter lovleyi</i> SZ	OleB	SEQ ID NO: 298	51
<i>Shewanella benthica</i> KT99	OleB	SEQ ID NO: 317	56
<i>Shewanella loihica</i> PV-4	OleB	SEQ ID NO: 272	53
<i>Photobacterium profundum</i> SS9	OleB	SEQ ID NO: 250	54
<i>Shewanella baltica</i> OS155	OleB	SEQ ID NO: 271	53
<i>Shewanella baltica</i> OS185	OleB	SEQ ID NO: 275	53
<i>Shewanella baltica</i> OS195	OleB	SEQ ID NO: 281	53
<i>Shewanella baltica</i> OS223	OleB	SEQ ID NO: 308	53
<i>Shewanella</i> sp. ANA-3	OleB	SEQ ID NO: 263	53
<i>Desulfococcus oleovorans</i> Hxd3	OleB	SEQ ID NO: 280	42
<i>Shewanella amazonensis</i> SB2B	OleB	SEQ ID NO: 266	53
<i>Shewanella frigidimarina</i> NCIMB 400	OleB	SEQ ID NO: 262	53
<i>Shewanella woodyi</i> ATCC 51908	OleB	SEQ ID NO: 284	54
<i>Photobacterium profundum</i> 3TCK	OleB	SEQ ID NO: 294	54
<i>Moritella</i> sp. PE36	OleB	SEQ ID NO: 313	56
<i>Shewanella denitrificans</i> OS217	OleB	SEQ ID NO: 257	54
<i>Psychromonas ingrahamii</i> 37	OleB	SEQ ID NO: 267	54
<i>Shewanella pealeana</i> ATCC 700345	OleB	SEQ ID NO: 279	54
<i>Shewanella putrefaciens</i> 200	OleB	SEQ ID NO: 306	52
<i>Shewanella putrefaciens</i> CN-32	OleB	SEQ ID NO: 273	52
<i>Shewanella sediminis</i> HAW-EB3	OleB	SEQ ID NO: 277 y 278	54/28*
<i>Shewanella</i> sp. W3-18-1	OleB	SEQ ID NO: 270	52
<i>Shewanella halifaxensis</i> HAW-EB4	OleB	SEQ ID NO: 283	54
<i>Shewanella oneidensis</i> MR-1	OleB	SEQ ID NO: 242	53
<i>Shewanella</i> sp. MR-7	OleB	SEQ ID NO: 261	53
<i>Colwellia psychrerythraea</i> 34H	OleB	SEQ ID NO: 252	55
<i>Shewanella</i> sp. MR-4	OleB	SEQ ID NO: 260	53
<i>Geobacter</i> sp. FRC-32	OleB	SEQ ID NO: 296	51
<i>Blastopirellula marina</i> DSM 3645	OleB	SEQ ID NO: 291	42
<i>Planctomyces maris</i> DSM 8797	OleB	SEQ ID NO: 309	40
<i>Rhodopirellula baltica</i> SH 1	OleB	SEQ ID NO: 244	36
<i>Lentisphaera araneosa</i> HTCC2155	OleB	SEQ ID NO: 310	35
<i>Desulfuromonas acetoxidans</i> DSM 684	OleB	SEQ ID NO: 295	52

ES 2 646 815 T3

<i>Gemmata obscuriglobus</i> UQM 2246	OleB	SEQ ID NO: 320	47
<i>Psychromonas</i> sp. CNPT3	OleB	SEQ ID NO: 293	51
<i>Streptomyces ambofaciens</i> ATCC 23877	OleB	SEQ ID NO: 234	39
<i>Opitutaceae bacterium</i> TAV2	OleB	SEQ ID NO: 315	43
<i>Mycobacterium vanbaalenii</i> PYR-1	OleB	SEQ ID NO: 268	31
<i>Marinobacter algicola</i> DG893	OleB	SEQ ID NO: 311 y 312	30/31*
<i>Hahella chejuensis</i> KCTC 2396	OleB	SEQ ID NO: 255	37
<i>Jannaschia</i> sp. CCS1	OleB	SEQ ID NO: 256	34
<i>Mycobacterium avium</i> 104	OleB	SEQ ID NO: 264	33
<i>Mycobacterium avium</i> subsp. <i>paratuberculosis</i> K-10	OleB	SEQ ID NO: 245	33
<i>Anaeromyxobacter</i> sp. Fw109-5	OleB	SEQ ID NO: 276	32
Proteobacteria gamma marina HTCC2080	OleB	SEQ ID NO: 299	32
<i>Mycobacterium marinum</i> M	OleB	SEQ ID NO: 232 y 233	32
<i>Mycobacterium tuberculosis</i> C	OleB	SEQ ID NO: 289	32
<i>Mycobacterium tuberculosis</i> H37Rv	OleB	SEQ ID NO: 238	32
<i>Mycobacterium tuberculosis</i> cepa Haarlem	OleB	SEQ ID NO: 237 y 319	32
<i>Paenibacillus</i> sp. JDR-2	OleB	SEQ ID NO: 321	32
<i>Burkholderia cenocepacia</i> AU 1054	OleB	SEQ ID NO: 258	31
<i>Burkholderia cenocepacia</i> PC184	OleB	SEQ ID NO: 290	31
<i>Burkholderia</i> sp. 383	OleB	SEQ ID NO: 254	31
<i>Ralstonia pickettii</i> 12J	OleB	SEQ ID NO: 304	31
<i>Burkholderia ambifaria</i> IOP40-10	OleB	SEQ ID NO: 322	30
<i>Marinobacter aquaeolei</i> VT8	OleB	SEQ ID NO: 269	30
<i>Pseudoalteromonas atlantica</i> T6c	OleB	SEQ ID NO: 259	30
<i>Rhodococcus rhodochrous</i>	OleB	SEQ ID NO: 246	30
<i>Rhodococcus</i> sp.	OleB	SEQ ID NO: 230	30
<i>Rhodococcus</i> sp. TDTM0003	OleB	SEQ ID NO: 247	30
<i>Mycobacterium</i> sp. GP1	OleB	SEQ ID NO: 248	29
<i>Microscilla marina</i> ATCC 23134	OleB	SEQ ID NO: 305	28
Organismo marino sin cultivar	OleB	SEQ ID NO: 235	27
<i>Flavobacteriales bacterium</i> HTCC2170	OleB	SEQ ID NO: 292	26
<i>Stenotrophomonas maltophilia</i> ATCC17679	OleC	SEQ ID NO: 385	100
<i>Stenotrophomonas maltophilia</i> R551-3	OleC	SEQ ID NO: 383	98
<i>Stenotrophomonas maltophilia</i> K279a	OleC	SEQ ID NO: 384	96
<i>Xanthomonas campestris</i> pv. <i>campestris</i> cepa 8004	OleC	SEQ ID NO: 399	77
<i>Xanthomonas campestris</i> pv. <i>campestris</i> cepa ATCC 33913	OleC	SEQ ID NO: 330	77
<i>Xanthomonas campestris</i> pv. <i>campestris</i> cepa B100	OleC	SEQ ID NO: 335	77
<i>Xanthomonas campestris</i> pv. <i>vesicatoria</i> cepa 85-10	OleC	SEQ ID NO: 348	77
<i>Xanthomonas axonopodis</i> pv. <i>citri</i> cepa 306	OleC	SEQ ID NO: 329	77
<i>Xanthomonas oryzae</i> pv. <i>oryzae</i> KACC10331	OleC	SEQ ID NO: 332	75
<i>Xanthomonas oryzae</i> pv. <i>oryzae</i> MAFF 311018	OleC	SEQ ID NO: 349	75
<i>Xanthomonas oryzae</i> pv. <i>oryzicola</i> BLS256	OleC	SEQ ID NO: 395	74
<i>Xylella fastidiosa</i> Dixon	OleC	SEQ ID NO: 373	73
<i>Xylella fastidiosa</i> 9a5c	OleC	SEQ ID NO: 327 y 336	71
<i>Xylella fastidiosa</i> Ann-1	OleC	SEQ ID NO: 374	71
<i>Xylella fastidiosa</i> M12	OleC	SEQ ID NO: 400	71
<i>Xylella fastidiosa</i> M23	OleC	SEQ ID NO: 401	71
<i>Xylella fastidiosa</i> Temecula1	OleC	SEQ ID NO: 331 y 340	71
<i>Plesiocystis pacifica</i> SIR-1	OleC	SEQ ID NO: 392	48
<i>Chloroflexus aggregans</i> DSM 9485	OleC	SEQ ID NO: 381	46
<i>Arthrobacter chlorophenolicus</i> A6	OleC	SEQ ID NO: 398	33
<i>Chloroflexus aurantiacus</i> J-10-fl	OleC	SEQ ID NO: 369	46
<i>Desulfatibacillum alkenivorans</i> AK-01	OleC	SEQ ID NO: 402	20
<i>Geobacter uraniumreducens</i> Rf4	OleC	SEQ ID NO: 363	51
<i>Pelobacter propionicus</i> DSM 2379	OleC	SEQ ID NO: 355	50
<i>Opitutus terrae</i> PB90-1	OleC	SEQ ID NO: 372	43

ES 2 646 815 T3

<i>Desulfotalea psychrophila</i> LSv54	OleC	SEQ ID NO: 344	41
<i>Geobacter bermidjiensis</i> Bem	OleC	SEQ ID NO: 387	51
<i>Geobacter lovleyi</i> SZ	OleC	SEQ ID NO: 382	50
<i>Shewanella benthica</i> KT99	OleC	SEQ ID NO: 394	45
<i>Shewanella loihica</i> PV-4	OleC	SEQ ID NO: 361	44
<i>Photobacterium profundum</i> SS9	OleC	SEQ ID NO: 328 y 346	43
<i>Shewanella baltica</i> OS155	OleC	SEQ ID NO: 360	41
<i>Shewanella baltica</i> OS185	OleC	SEQ ID NO: 364	41
<i>Shewanella baltica</i> OS195	OleC	SEQ ID NO: 368	41
<i>Shewanella baltica</i> OS223	OleC	SEQ ID NO: 388	41
<i>Shewanella</i> sp. ANA-3	OleC	SEQ ID NO: 354	50
<i>Desulfococcus oleovorans</i> Hxd3	OleC	SEQ ID NO: 367	46
<i>Shewanella amazonensis</i> SB2B	OleC	SEQ ID NO: 356	45
<i>Shewanella frigidimarina</i> NCIMB 400	OleC	SEQ ID NO: 353	44
<i>Shewanella woodyi</i> ATCC 51908	OleC	SEQ ID NO: 371	44
<i>Photobacterium profundum</i> 3TCK	OleC	SEQ ID NO: 378	43
<i>Moritella</i> sp. PE36	OleC	SEQ ID NO: 391	42
<i>Shewanella denitrificans</i> OS217	OleC	SEQ ID NO: 350	42
<i>Psychromonas ingrahamii</i> 37	OleC	SEQ ID NO: 357	41
<i>Shewanella pealeana</i> ATCC 700345	OleC	SEQ ID NO: 366	41
<i>Shewanella putrefaciens</i> 200	OleC	SEQ ID NO: 386	41
<i>Shewanella putrefaciens</i> CN-32	OleC	SEQ ID NO: 362	41
<i>Shewanella sediminis</i> HAW-EB3	OleC	SEQ ID NO: 365	41
<i>Shewanella</i> sp. W3-18-1	OleC	SEQ ID NO: 359	41
<i>Shewanella halifaxensis</i> HAW-EB4	OleC	SEQ ID NO: 370	40
<i>Shewanella oneidensis</i> MR-1	OleC	SEQ ID NO: 339	40
<i>Shewanella</i> sp. MR-7	OleC	SEQ ID NO: 352	40
<i>Colwellia psychrerythraea</i> 34H	OleC	SEQ ID NO: 347	39
<i>Shewanella</i> sp. MR-4	OleC	SEQ ID NO: 351	39
<i>Geobacter</i> sp. FRC-32	OleC	SEQ ID NO: 380	50
<i>Blastopirellula marina</i> DSM 3645	OleC	SEQ ID NO: 376	44
<i>Planctomyces maris</i> DSM 8797	OleC	SEQ ID NO: 389	44
<i>Rhodopirellula baltica</i> SH 1	OleC	SEQ ID NO: 341	41
<i>Lentisphaera araneosa</i> HTCC2155	OleC	SEQ ID NO: 390	39
<i>Desulfuromonas acetoxidans</i> DSM 684	OleC	SEQ ID NO: 379	48
<i>Gemmata obscuriglobus</i> UQM 2246	OleC	SEQ ID NO: 396 y 397	50/50*
<i>Psychromonas</i> sp. CNPT3	OleC	SEQ ID NO: 377	39
<i>Streptomyces ambofaciens</i> ATCC 23877	OleC	SEQ ID NO: 333	36
<i>Opitutaceae bacterium</i> TAV2	OleC	SEQ ID NO: 393	41
<i>Mycobacterium vanbaalenii</i> PYR-1	OleC	SEQ ID NO: 358	34
<i>Nocardia farcinica</i> IFM 10152	OleC	SEQ ID NO: 345	43
Organismo marino sin cultivar	OleC	SEQ ID NO: 334	31
<i>Bdellovibrio bacteriovorus</i> HD100	OleC	SEQ ID NO: 343	28
<i>Photorhabdus luminescens</i> subsp. <i>laumondii</i> TTO1	OleC	SEQ ID NO: 342	28
<i>Roseovarius nubinhibens</i> ISM	OleC	SEQ ID NO: 375	26
<i>Stenotrophomonas maltophilia</i> ATCC17679	OleD	SEQ ID NO: 456	100
<i>Stenotrophomonas maltophilia</i> R551-3	OleD	SEQ ID NO: 454	98
<i>Stenotrophomonas maltophilia</i> K279a	OleD	SEQ ID NO: 455	96
<i>Xanthomonas campestris</i> pv. <i>campestris</i> cepa 8004	OleD	SEQ ID NO: 404	82
<i>Xanthomonas campestris</i> pv. <i>campestris</i> cepa ATCC 33913	OleD	SEQ ID NO: 408	82
<i>Xanthomonas campestris</i> pv. <i>campestris</i> cepa B100	OleD	SEQ ID NO: 406	80
<i>Xanthomonas campestris</i> pv. <i>vesicatoria</i> cepa 85-10	OleD	SEQ ID NO: 417	80
<i>Xanthomonas axonopodis</i> pv. <i>citri</i> cepa 306	OleD	SEQ ID NO: 409	80
<i>Xanthomonas oryzae</i> pv. <i>oryzae</i> KACC10331	OleD	SEQ ID NO: 415	79
<i>Xanthomonas oryzae</i> pv. <i>oryzae</i> MAFF 311018	OleD	SEQ ID NO: 418	79
<i>Xanthomonas oryzae</i> pv. <i>oryzicola</i> BLS256	OleD	SEQ ID NO: 142	80
<i>Xylella fastidiosa</i> Dixon	OleD	SEQ ID NO: 444	71
<i>Xylella fastidiosa</i> 9a5c	OleD	SEQ ID NO: 407	72
<i>Xylella fastidiosa</i> Ann-1	OleD	SEQ ID NO: 445	72

ES 2 646 815 T3

<i>Xylella fastidiosa</i> M12	OleD	SEQ ID NO: 145	72
<i>Xylella fastidiosa</i> M23	OleD	SEQ ID NO: 146	71
<i>Xylella fastidiosa</i> Temecula 1	OleD	SEQ ID NO: 411	72
<i>Plesiocystis pacifica</i> SIR-1	OleD	SEQ ID NO: 462 y 463	49/38*
<i>Chloroflexus aggregans</i> DSM 9485	OleD	SEQ ID NO: 451 y 452	55/46*
<i>Arthrobacter chlorophenolicus</i> A6	OleD	SEQ ID NO: 144	38
<i>Chloroflexus aurantiacus</i> J-10-fl	OleD	SEQ ID NO: 438	45
<i>Clavibacter michiganensis</i> subsp. <i>Sepedonicus</i>	OleD	SEQ ID NO: 440	41
<i>Clavibacter michiganensis</i> subsp. <i>michiganensis</i> NCPB 382	OleD	SEQ ID NO: 431	41
<i>Arthrobacter aurescens</i> TC1	OleD	SEQ ID NO: 427	37
<i>Brevibacterium linens</i> BL2	OleD	SEQ ID NO: 443	37
<i>Desulfatibacillum alkenivorans</i> AK-01	OleD	SEQ ID NO: 149	
<i>Congregibacter litoralis</i> KT71	OleD	SEQ ID NO: 447	47
<i>Kineococcus radiotolerans</i> SRS30216	OleD	SEQ ID NO: 433	39
<i>Micrococcus luteus</i> NCTC 2665	OleD	SEQ ID NO: 403	39
<i>Geobacter uraniumreducens</i> Rf4	OleD	SEQ ID NO: 432	59
<i>Pelobacter propionicus</i> DSM 2379	OleD	SEQ ID NO: 424	55
<i>Opiritatus terrae</i> PB90-1	OleD	SEQ ID NO: 442	47
<i>Desulfotalea psychrophila</i> LSv54	OleD	SEQ ID NO: 413	46
<i>Geobacter bemidjiensis</i> Bem	OleD	SEQ ID NO: 458	56
<i>Geobacter lovleyi</i> SZ	OleD	SEQ ID NO: 453	56
<i>Shewanella benthica</i> KT99	OleD	SEQ ID NO: 464	49
<i>Shewanella loihica</i> PV-4	OleD	SEQ ID NO: 429	47
<i>Photobacterium profundum</i> SS9	OleD	SEQ ID NO: 414	50
<i>Shewanella baltica</i> OS155	OleD	SEQ ID NO: 148	42
<i>Shewanella baltica</i> OS185	OleD	SEQ ID NO: 484	44
<i>Shewanella baltica</i> OS195	OleD	SEQ ID NO: 485	43
<i>Shewanella baltica</i> OS223	OleD	SEQ ID NO: 486	43
<i>Shewanella</i> sp. ANA-3	OleD	SEQ ID NO: 423	50
<i>Desulfococcus oleovorans</i> Hxd3	OleD	SEQ ID NO: 436 y 437	47/40*
<i>Shewanella amazonensis</i> SB2B	OleD	SEQ ID NO: 425	46
<i>Shewanella frigidimarina</i> NCIMB 400	OleD	SEQ ID NO: 422	50
<i>Shewanella woodyi</i> ATCC 51908	OleD	SEQ ID NO: 441	46
<i>Photobacterium profundum</i> 3TCK	OleD	SEQ ID NO: 449	48
<i>Moritella</i> sp. PE36	OleD	SEQ ID NO: 461	45
<i>Shewanella denitrificans</i> OS217	OleD	SEQ ID NO: 419	49
<i>Psychromonas ingrahamii</i> 37	OleD	SEQ ID NO: 426	48
<i>Shewanella pealeana</i> ATCC 700345	OleD	SEQ ID NO: 435	46
<i>Shewanella putrefaciens</i> 200	OleD	SEQ ID NO: 457	47
<i>Shewanella putrefaciens</i> CN-32	OleD	SEQ ID NO: 430	46
<i>Shewanella sediminis</i> HAW-EB3	OleD	SEQ ID NO: 434	46
<i>Shewanella</i> sp. W3-18-1	OleD	SEQ ID NO: 428	46
<i>Shewanella halifaxensis</i> HAW-EB4	OleD	SEQ ID NO: 439	48
<i>Shewanella oneidensis</i> MR-1	OleD	SEQ ID NO: 410	48
<i>Shewanella</i> sp. MR-7	OleD	SEQ ID NO: 421	49
<i>Colwellia psychrerythraea</i> 34H	OleD	SEQ ID NO: 416	44
<i>Shewanella</i> sp. MR-4	OleD	SEQ ID NO: 420	49
<i>Geobacter</i> sp. FRC-32	OleD	SEQ ID NO: 450	55
<i>Blastopirellula marina</i> DSM 3645	OleD	SEQ ID NO: 446	51
<i>Planctomyces maris</i> DSM 8797	OleD	SEQ ID NO: 459	47
<i>Rhodopirellula baltica</i> SH 1	OleD	SEQ ID NO: 412	40
<i>Lentisphaera araneosa</i> HTCC2155	OleD	SEQ ID NO: 460	40
<i>Desulfuromonas acetoxidans</i> DSM 684	OleD	SEQ ID NO: 487	51
<i>Gemmata obscuriglobus</i> UQM 2246	OleD	SEQ ID NO: 143	52
<i>Psychromonas</i> sp. CNPT3	OleD	SEQ ID NO: 448	48
<i>Streptomyces ambofaciens</i> ATCC 23877	OleD	SEQ ID NO: 405	44
<i>Opiritaceae bacterium</i> TAV2	OleD	SEQ ID NO: 147	28

TABLA 2

Organismo	Proteína	Seq ID No	Porcentaje de identidad de OleB	Porcentaje de identidad de OleC
<i>Clavibacter michiganensis</i> subsp. <i>Sepedonicus</i>	OleBC	SEQ ID NO: 138	33	36
<i>Clavibacter michiganensis</i> subsp. <i>michiganensis</i> NCPPB 382	OleBC	SEQ ID NO: 136	33	35
Arthrobacter aureescens TC1	OleBC	SEQ ID NO: 135	37	33
<i>Brevibacterium linens</i> BL2	OleBC	SEQ ID NO: 139	27	34
Kineococcus radiotolerans SRS30216	OleBC	SEQ ID NO: 137	31	34
Micrococcus luteus NCTC 2665	OleBC	SEQ ID NO: 141	32	32
<i>Congregibacter litoralis</i> KT71	OleBC	SEQ ID NO: 140	32	33

La invención se refiere a cinco familias de proteínas implicadas en la biosíntesis de hidrocarburos, las familias de proteínas OleA, OleB, OleC, OleBC y OleD, que se denominan colectivamente la familia de proteínas Ole. Se usaron programas bioinformáticos, tales como los programas BLAST (proporcionados por los NHI, Bethesda, MD), para identificar secuencias de proteína y nucleótidos que pertenecen a las familias de proteínas OleA, OleB, OleC, OleBC y OleD.

Con el fin de identificar proteínas que lo más probablemente estén implicadas en la biosíntesis de hidrocarburos, se aplicaron restricciones adicionales al análisis bioinformático. La biosíntesis de hidrocarburos no se realiza mediante una única proteína Ole. Por tanto, puede ser más probable que los organismos que contienen las cuatro proteínas Ole produzcan hidrocarburos. Por consiguiente, se realizó el análisis bioinformático en más de 940 genomas bacterianos, así como más de 100 genomas de arqueas y eucariotas. Esta investigación reveló 67 genomas bacterianos que contienen los cuatro genes de síntesis de hidrocarburos *ole*: *Stenotrophomonas maltophilia* R551-3, *Stenotrophomonas maltophilia* K279a, *Arthrobacter aureescens* TC1, *Arthrobacter chlorophenolicus* A6, *Blastopirellula marina* DSM 3645, *Brevibacterium linens* BL2, *Desulfococcus oleovorans* Hxd3, *Chloroflexus aggregans* DSM 9485, *Chloroflexus aurantiacus* J-10-fl, *Clavibacter michiganensis* subsp. *michiganensis* NCPPB 382, *Clavibacter michiganensis* subsp. *Sepedonicus*, *Colwellia psychrerythraea* 34H, *Congregibacter litoralis* KT71, *Desulfotalea psychrophila* LSv54, *Desulfuromonas acetoxidans* DSM 684, *Gemmata obscuriglobus* UQM 2246, *Geobacter bemidjensis* Bem, *Geobacter lovleyi* SZ, *Geobacter* sp. FRC-32, *Geobacter uraniumreducens* Rf4, *Kineococcus radiotolerans* SRS30216, *Lentisphaera araneosa* HTCC2155, *Micrococcus luteus* NCTC 2665, *Moritella* sp. PE36, *Opiritatus terrae* PB90-1, *Pelobacter propionicus* DSM 2379, *Photobacterium profundum* 3TCK, *Photobacterium profundum* SS9, *Planctomyces maris* DSM 8797, *Plesiocystis pacifica* SIR-1, *Psychromonas ingrahamii* 37, *Psychromonas* sp. CNPT3, *Rhodopirellula baltica* SH 1, *Shewanella amazonensis* SB2B, *Shewanella baltica* OS155, *Shewanella baltica* OS185, *Shewanella baltica* OS195, *Shewanella baltica* OS223, *Shewanella benthica* KT99, *Shewanella denitrificans* OS217, *Shewanella frigidimarina* NCIMB 400, *Shewanella halfaxensis* HAW-EB4, *Shewanella loihica* PV-4, *Shewanella oneidensis* MR-1, *Shewanella pealeana* ATCC 700345, *Shewanella putrefaciens* 200, *Shewanella putrefaciens* CN-32, *Shewanella sediminis* HAW-EB3, *Shewanella* sp. ANA-3, *Shewanella* sp. MR-4, *Shewanella* sp. MR-7, *Shewanella* sp. W3-18-1, *Shewanella woodyi* ATCC 51908, *Streptomyces ambofaciens* ATCC 23877, *Xanthomonas axonopodis* pv. *citri* cepa 306, *Xanthomonas campestris* pv. *campestris* cepa 8004, *Xanthomonas campestris* pv. *campestris* cepa B100, *Xanthomonas campestris* pv. *campestris* cepa ATCC 33913, *Xanthomonas campestris* pv. *vesicatoria* cepa 85-10, *Xanthomonas oryzae* pv. *oryzae* KACC10331, *Xanthomonas oryzae* pv. *oryzae* MAFF 311018, *Xanthomonas oryzae* pv. *oryzicola* BLS256, *Xylella fastidiosa* 9a5c, *Xylella fastidiosa* Ann-1, *Xylella fastidiosa* Dixon, *Xylella fastidiosa* M12, *Xylella fastidiosa* M23 y *Xylella fastidiosa* Temecula 1.

Informes previos han intentado caracterizar la capacidad de diversos organismos para producir hidrocarburos, pero estos informes no pudieron confirmarse (véanse, por ejemplo, Jones *et al.*, *J. Gen. Microbiol.*, 59: 145-152 (1969), Ladygina *et al.*, *Process Biochemistry*, 41:1001-1014 (2006)). Aunque se publica muy poco sobre la capacidad de los organismos para producir hidrocarburos no isoprenoides, hay cuatro organismos que se ha mostrado en la bibliografía que producen los tipos de olefinas descritos en el presente documento: *Stenotrophomonas maltophilia*, *Kineococcus radiotolerans*, especies de *Chloroflexus* y diversas especies de *Micrococcus*, incluyendo la secuenciada recientemente *Micrococcus luteus* (Tomabene *et al.*, *Can. J. Microbiol.*, 24: 525-532 (1978); Suen *et al.*, *Journal of Industrial Microbiology*, 2: 337-348 (1988); Morrison *et al.*, *J. Bacteriol.*, 108: 353-358 (1971); van der Meer *et al.*, *Org. Geochem.*, 30: 1585-1587 (1999); Albro *et al.*, *Biochemistry*, 8: 394-404 (1969); Philips *et al.*, *Int. J. Syst. Evol. Microbiol.*, 52:933-938 (2002)).

Búsquedas bioinformáticas identificaron independientemente los cuatro genes *ole* en *Stenotrophomonas maltophilia* (R551-3 y K279a), *Kineococcus radiotolerans*, *Chloroflexus aggregans*, *Chloroflexus aurantiacus* y *Micrococcus luteus*. Además, se ha confirmado la producción de hidrocarburos (por ejemplo, olefinas) a partir de estos organismos. Específicamente, se sometieron a prueba las proteínas Ole de varios de estos organismos para determinar su capacidad para producir hidrocarburos. Se seleccionaron dos organismos que se sabe que producen hidrocarburos y dos organismos que no se ha notificado que produzcan hidrocarburos, pero se identificaron en el análisis bioinformático, y sometieron a prueba para determinar su capacidad para producir hidrocarburos cuando se expresan en *E. coli*.

El ejemplo 6 muestra que las secuencias de proteína OleA, OleC y OleD de *Stenotrophomonas maltophilia*, *Xanthomonas axonopodis* y *Chloroflexus aggregans* funcionan produciendo hidrocarburos cuando se expresan en un huésped, tal como *E. coli* o *Bacillus megaterium*. Además, la secuencia de proteína OleC de *Plesiocystis pacifica* también produce hidrocarburos cuando se expresa con *oleA* y *oleD* en *E. coli*.

- 5 Para confirmar que estos genes confieren capacidad de producción de hidrocarburos en *Stenotrophomonas maltophilia*, se prepararon mutaciones por delección de *oleA*, *oleC* y *oleD* en una cepa productora de hidrocarburos de *Stenotrophomonas maltophilia*. Los mutantes por delección de *Stenotrophomonas maltophilia* que carecen de *oleA* no producen ningún hidrocarburo, los mutantes por delección de *Stenotrophomonas maltophilia* que carecen de o bien *oleC* o bien *oleD* no producen hidrocarburos, pero producen cetonas alifáticas en su lugar.
- 10 Usando bioinformática, pueden diseñarse motivos de aminoácidos de Ole identificando regiones conservadas de cada proteína Ole. Estos motivos de aminoácidos pueden diseñarse mediante alineaciones de proteínas seguido por inspección visual de las secuencias de proteína. Estos motivos de aminoácidos pueden usarse entonces para identificar proteínas que tienen funciones biológicas similares a las proteínas Ole. Varios programas bien conocidos en la técnica pueden usar los motivos de aminoácidos para identificar proteínas que pertenecen a la familia de proteínas funcionales. Por ejemplo, uno de tales programas disponibles públicamente es <http://motif.genome.jp/motif2.html> (véase el ejemplo 7).

Basándose en los datos experimentales, se predijo que los organismos que contienen los cuatro genes *ole* lo más probablemente tienen la capacidad para producir hidrocarburos. Por tanto, los motivos de aminoácidos se crearon compilando todas las secuencias de proteínas OleA, OleB, OleC y OleD de los sesenta y siete organismos (véase, por ejemplo, la tabla 1) que contienen las cuatro proteínas Ole. Por fusiones de proteínas, tales como las proteínas OleBC, se alinearon las secciones de OleB con las secuencias de proteína OleB y se alinearon las secciones de OleC con la secuencia de proteína OleC. Los organismos con secuencias genómicas parciales o posibles recombinaciones genómicas no se incluyeron en la tabla 1 y, por tanto, no se usaron para diseñar estos motivos de aminoácidos. Basándose en las regiones conservadas de estos sesenta y siete organismos que contienen los cuatro genes *ole*, se predice que otras secuencias de polipéptido que contienen estos motivos de aminoácidos pueden ser proteínas Ole funcionales. Los motivos de aminoácidos diseñados a partir de los sesenta y siete organismos que contienen los cuatro genes *ole* se muestran en las tablas 3-6. En todos los motivos representados, la numeración de subíndices indica la posición del aminoácido dentro del motivo. X representa cualquier aminoácido (por ejemplo, cualquier aminoácido que se produzca de manera natural), y los elementos entre corchetes reflejan la elección de residuos de aminoácido para la posición indicada. Por ejemplo, [LF]₁ significa o bien L o bien F en la posición de residuo 1.

TABLA 3

Motivos para la familia de proteína OleA	
SEQ ID NO: 64	[LF] ₁ -X ₂ -X ₃ -[IVLM] ₄ -[ATSV] ₅ -G ₆ -[IV] ₇ -X ₈ -[EAHS] ₉ -R ₁₀ -R ₁₁ -X ₁₂ -W ₁₃
SEQ ID NO: 65	[TAC] ₁ -[SGT] ₂ -V ₃ -X ₄ -[RQ] ₅ -X ₆ -X ₇ -X ₈ -E ₉ -P ₁₀ -[SA] ₁₁ -[TV] ₁₂ -A ₁₃
SEQ ID NO: 66	D ₁ -[VIL] ₂ -X ₃ -N ₄ -A ₅ -C ₆ -L ₇ -[AG] ₈
SEQ ID NO: 67	[LM] ₁ -T ₂ -X ₃ -G ₄ -[SCA] ₅ -G ₆
SEQ ID NO: 68	M ₁ -X ₂ -T ₃ -[DSN] ₄ -X ₅ -X ₆ -X ₇ -X ₈ -X ₉ -X ₁₀ -X ₁₁ -G ₁₂
SEQ ID NO: 69	[LF] ₁ -X ₂ -X ₃ -G ₄ -[DE] ₅ -X ₆ -X ₇ -X ₈ -X ₉ -[LM] ₁₀ -G ₁₁ -[IVM] ₁₂ -G ₁₃ -S ₁₄ -G ₁₅ -[LI] ₁₆ -N ₁₇ -[CATS] ₁₈ -X ₁₉ -[MAV] ₂₀ -X ₂₁ -X ₂₂ -X ₂₃ -X ₂₄ -W ₂₅

TABLA 4

Motivos para la familia de proteína OleD	
SEQ ID NO: 70	[ILVA] ₁ -[LFV] ₂ -V ₃ -T ₄ -G ₅ -[GAC] ₆ -[GSN] ₇ -G ₈ -[FLM] ₉ -[LV] ₁₀ -G ₁₁ -X ₁₂ -X ₁₃ -[LVIT] ₁₄ -[CVA] ₁₅ -X ₁₆ -X ₁₇ -L ₁₈ -X ₁₉ -X ₂₀ -X ₂₁ -G ₂₂
SEQ ID NO: 71	H ₁ -[NVTL] ₂ -[AG] ₃ -[AS] ₄ -[KLVIQR] ₅ -[VAPI] ₆ -[GSD] ₇ -[AVLIM] ₈ -X ₉ -G ₁₀

TABLA 5

35

Motivos para la familia de proteína OleC	
SEQ ID NO: 72	G ₁ -X ₂ -X ₃ -[DE] ₄ -[LM] ₅ -[PACS] ₆ -[TAG] ₇ -F ₈ -P ₉ -[LP] ₁₀ -F ₁₁ -[ASG] ₁₂ -L ₁₃ -F ₁₄ -X ₁₅ -X ₁₆ -A ₁₇ -[LM] ₁₈ -G ₁₉
SEQ ID NO: 73	H ₁ -R ₂ -M ₃ -G ₄ -D ₅ -[VL] ₆ -[GA] ₇ -X ₈ -X ₉ -D ₁₀ -X(5) ₁₁₋₁₅ -W ₁₆ -[FM] ₁₇ -[CY] ₁₈ -G ₁₉ -R ₂₀ -K ₂₁ -X ₂₂ -[HQ] ₂₃ -[RI] ₂₄ -V ₂₅
SEQ ID NO: 74	L ₁ -X ₂ -H ₃ -X ₄ -X ₅ -F ₆ -P ₇ -[VM] ₈ -D ₉ -[IV] ₁₀ -R ₁₁ -H ₁₂ -N ₁₃ -[AS] ₁₄ -K ₁₅ -I ₁₆ -[GF] ₁₇ -R ₁₈ -E ₁₉ -X ₂₀ -L ₂₁ -A ₂₂ -X ₂₃ -W ₂₄ -A ₂₅

TABLA 6

Motivos para la familia de proteína OleB	
SEQ ID NO: 91	[VILMF] ₁ -[VL] ₂ -[MLFC] ₃ -[VLI] ₄ -H ₅ -G ₆ -N ₇ -P ₈ -[ST] ₉ -W ₁₀ -[SCGT] ₁₁ -[FYL] ₁₂ -[YFLAB] ₁₃ -[WY] ₁₄ -R ₁₅
SEQ ID NO: 92	[RQ] ₁ -[CVILT] ₂ -[IVL] ₃ -[VAC] ₄ -[PVML] ₅ -D ₆ -[HNL] ₇ -[IMLPV] ₈ -G ₉ -[MCF] ₁₀ -G ₁₁ -[LFRYKT] ₁₂ -S ₁₃ -[DERS] ₁₄ -[KR] ₁₅ -[PL] ₁₆ -[DGSAR] ₁₇
SEQ ID NO: 93	[TIHDS] ₁ -[LF] ₂ -[AVIMFC] ₃ -[VLA] ₄ -H ₅ -D ₆ -W ₇ -G ₈ -G ₉ -[MAFP] ₁₀ -[IL] ₁₁ -G ₁₂ -[FMTLVC] ₁₃ -[GAS] ₁₄ -[WYFALMC] ₁₅ -[AM] ₁₆
SEQ ID NO: 94	[RKQN] ₁ -[LIVCF] ₂ -[VGI] ₃ -[IVLMCA] ₄ -[LTCMF] ₅ -N ₆ -T ₇ -[AGVS] ₈ -[AVG] ₉ -F ₁₀
SEQ ID NO: 95	[NDHKE] ₁ -R ₂ -[IVR] ₃ -[SA] ₄ -[TIV] ₅ -X ₆ -X ₇ -[FA] ₈ -[VIM] ₉ -[QGHLKR] ₁₀ -[DAT] ₁₁ -I ₁₂ -[PA] ₁₃ -[LCR] ₁₄
SEQ ID NO: 96	[WF] ₁ -G ₂ -X ₃ -[RQKYH] ₄ -D ₅ -[FIWP] ₆ -[VC] ₇ -F ₈ -[DNTR] ₉ -X ₁₀ -X ₁₁ -[FYCL] ₁₂ -[LFY] ₁₃
SEQ ID NO: 97	[GN] ₁ -H ₂ -Y ₃ -[VIL] ₄ -[LVI] ₅ -E ₆ -D ₇

5 Se desarrollaron motivos de aminoácidos adicionales considerando adicionalmente la bibliografía publicada que identificó organismos como productores de hidrocarburos en combinación con datos bioinformáticos que muestran que estos organismos también contenían los cuatro genes *ole*. Además, se desarrollaron estos motivos usando datos experimentales que demostraron que los organismos particulares sometidos a prueba con los cuatro genes *ole* producían hidrocarburos. Estos motivos de aminoácidos se muestran en las tablas 7-9.

TABLA 7

Motivos para OleA	
SEQ ID NO: 98	V ₁ -X ₂ -[RQ] ₃ -[DPA] ₄ -[YHTN] ₅ -L ₆ -E ₇ -P ₈ -[SA] ₉ -[TV] ₁₀ -A ₁₁ .
SEQ ID NO: 99	F ₁ -D ₂ -[VIL] ₃ -X ₄ -N ₅ -A ₆ -C ₇ -L ₈ -X ₉ -[FW] ₁₀ -X ₁₁ -N ₁₂ -G ₁₃ .
SEQ ID NO: 100	I ₁ -[DER] ₂ -Y ₃ -A ₄ -[LV] ₅ -[IV] ₆ -[VL] ₇ -X ₈ -[GA] ₉ -E ₁₀ .
SEQ ID NO: 101	F ₁ -X ₂ -X ₃ -X ₄ -G ₅ -N ₆ -[IV] ₇ -G ₈ -P ₉ -[AI] ₁₀ -X ₁₁ -X ₁₂ -P ₁₃ .
SEQ ID NO: 102	[LM] ₁ -G ₂ -[IV] ₃ -G ₄ -S ₅ -G ₆ -L ₇ -N ₈ -X ₉ -[SAG] ₁₀ -M ₁₁ .

TABLA 8

10

Motivos para OleC	
SEQ ID NO: 103	A ₁ -A ₂ -[IV] ₃ -X ₄ -F ₅ -T ₆ -S ₇ -G ₈ -S ₉ -T ₁₀ -G ₁₁ -X ₁₂ -[PA] ₁₃ -K ₁₄ -G ₁₅ -V ₁₆ .
SEQ ID NO: 104	F ₁ -X ₂ -X ₃ -F ₄ -A ₅ -L ₆ -X ₇ -X ₈ -X ₉ -A ₁₀ -L ₁₁ -G ₁₂ -X ₁₃ -[TV] ₁₄ -[ST] ₁₅ -X ₁₆ -X ₁₇ -P ₁₈ -X ₁₉ -X ₂₀ -D ₂₁ -[PV] ₂₂ .
SEQ ID NO: 105	[TP] ₁ -P ₂ -Y ₃ -G ₄ -X ₅ -T ₆ -E ₇ -X ₈ -L ₉ -P ₁₀ -V ₁₁ .
SEQ ID NO: 106	H ₁ -R ₂ -X ₃ -G ₄ -D ₅ -X ₆ -G ₇ -[YHMW] ₈ -X ₉ -D ₁₀ -X ₁₁ -X ₁₂ -G ₁₃ -R ₁₄ -[LI] ₁₅ -W ₁₆ -[VF] ₁₇ -X ₁₈ -G ₁₉ -R ₂₀ -[KL] ₂₁ .
SEQ ID NO: 107	A ₁ -[LI] ₂ -[VT] ₃ -G ₄ -[VPI] ₅ -[GQ] ₆ -X ₇ -X ₈ -G ₉ -[AQT] ₁₀ -[QV] ₁₁ -X ₁₂ -[PAV] ₁₃ -[VA] ₁₄ .
SEQ ID NO: 108	P ₁ -[VT] ₂ -D ₃ -L ₄ -R ₅ -H ₆ -N ₇ -A ₈ -K ₉ -I ₁₀ -X ₁₁ -R ₁₂ -X ₁₃ -X ₁₄ -L ₁₅ -[AR] ₁₆ -X ₁₇ -W ₁₈ -A ₁₉ -X ₂₀ -X ₂₁ -X ₂₂ -L ₂₃ .

TABLA 9

Motivo para OleD	
SEQ ID NO: 109	L ₁ -V ₂ -T ₃ -G ₄ -[AG] ₅ -[GNS] ₆ -G ₇ -[FLM] ₈ -[LV] ₉ -G ₁₀ -[QRAG] ₁₁ -X ₁₂ -[LVT] ₁₃ -[CVA] ₁₄ -[RE] ₁₅ -X ₁₆ -L ₁₇ .
SEQ ID NO: 110	[VIA] ₁ -X ₂ -[ATG] ₃ -[IV] ₄ -X ₅ -H ₆ -X ₇ -X ₈ -A ₉ -K ₁₀ -[AV] ₁₁ -[GS] ₁₂ -X ₁₃ -X ₁₄ -G ₁₅ .
SEQ ID NO: 111	[YHM] ₁ -[TV] ₂ -S ₃ -[TS] ₄ -P ₅ -S ₆ -V ₇ -[VTA] ₈ -[HI] ₉ .
SEQ ID NO: 112	[AR] ₁ -[TV] ₂ -X ₃ -[AV] ₄ -[LV] ₅ -R ₆ -P ₇ -[RH] ₈ -[LI] ₉ -X ₁₀ -[WL] ₁₁ -G ₁₂ -P ₁₃ -[GR] ₁₄ -D ₁₅ .
SEQ ID NO: 113	G ₁ -[RE] ₂ -A ₃ -[YLFV] ₄ -[FV] ₅ -[IV] ₆ -X ₇ -[NQ] ₈ -[GE] ₉ -X ₁₀ -P ₁₁ .
SEQ ID NO: 114	E ₁ -[PV] ₂ -P ₃ -[LM] ₄ -T ₅ -[RE] ₆ -[FL] ₇ -[LM] ₈ -[AV] ₉ .
SEQ ID NO: 115	P ₁ -X ₂ -[IV] ₃ -[SP] ₄ -[IL] ₅ -[ED] ₆ -E ₇ -G ₈ -[LF] ₉ -[RQHA] ₁₀ -R ₁₁ .

- 5 Finalmente, se desarrollaron motivos de aminoácidos adicionales alineando secuencias de proteínas Ole de organismos cuyos genes *ole* se expresaron de manera heteróloga en *E. coli* y se confirmó que tenían un papel en la producción de hidrocarburos (véase, por ejemplo, el ejemplo 6). Estos motivos de aminoácidos se muestran en las tablas 10-12.

TABLA 10

Motivos para OleA	
SEQ ID NO: 116	D ₁ -Y ₂ -L ₃ -E ₄ -P ₅ -S ₆ -T ₇ -A ₈
SEQ ID NO: 117	F ₁ -D ₂ -V ₃ -X ₄ -N ₅ -A ₆ -C ₇ -L ₈ -X ₉ -F ₁₀ -X ₁₁ -N ₁₂ -G ₁₃ -M ₁₄
SEQ ID NO: 118	Y ₁ -A ₂ -L ₃ -[IV] ₄ -V ₅ -D ₆ -G ₇ -E ₈
SEQ ID NO: 119	F ₁ -X ₂ -X ₃ -X ₄ -G ₅ -N ₆ -I ₇ -G ₈ -P ₉ -A ₁₀ -X ₁₁ -X ₁₂ -P ₁₃ -I ₁₄
SEQ ID NO: 120	L ₁ -G ₂ -I ₃ -G ₄ -S ₅ -G ₆ -L ₇ -N ₈ -C ₉ -S ₁₀ -M ₁₁

TABLA 11

Motivos para OleC	
SEQ ID NO: 121	A ₁ -A ₂ -I ₃ -[LV] ₄ -F ₅ -T ₆ -S ₇ -G ₈ -S ₉ -T ₁₀ -G ₁₁ -[VTP] ₁₂ -P ₁₃ -K ₁₄ -G ₁₅ -V ₁₆ -[VL] ₁₇ -Y ₁₈ -[RT] ₁₉ -H ₂₀ -[RG] ₂₁
SEQ ID NO: 122	P ₁ -[AT] ₂ -F ₃ -P ₄ -[LP] ₅ -F ₆ -A ₇ -L ₈ -F ₉ -[ND] ₁₀ -[VP] ₁₁ -A ₁₂ -L ₁₃ -G ₁₄ -[LMV] ₁₅ -T ₁₆ -[ST] ₁₇ -[AV] ₁₈ -[IL] ₁₉ -P ₂₀
SEQ ID NO: 123	[TP] ₁ -P ₂ -Y ₃ -G ₄ -A ₅ -T ₆ -E ₇ -X ₈ -L ₉ -P ₁₀ -V ₁₁
SEQ ID NO: 124	H ₁ -R ₂ -M ₃ -G ₄ -D ₅ -X ₆ -G ₇ -[YW] ₈ -X ₉ -D ₁₀ -X ₁₁ -X ₁₂ -G ₁₃ -R ₁₄ -[LI] ₁₅ -W ₁₆ -F ₁₇ -[CY] ₁₈ -G ₁₉ -R ₂₀ -K ₂₁
SEQ ID NO: 125	A ₁ -L ₂ -V ₃ -G ₄ -[VP] ₅ -[GQ] ₆ -X ₇ -X ₈ -G ₉ -[AQ] ₁₀ -[QV] ₁₁ -X ₁₂ -[PA] ₁₃ -[VA] ₁₄
SEQ ID NO: 126	F ₁ -P ₂ -V ₃ -D ₄ -I ₅ -R ₆ -H ₇ -N ₈ -A ₉ -K ₁₀ -I ₁₁ -[GF] ₁₂ -R ₁₃ -[EG] ₁₄ -X ₁₅ -L ₁₆ -[AR] ₁₇ -[VA] ₁₈ -W ₁₉ -A ₂₀ -X ₂₁ -X ₂₂ -X ₂₃ -L ₂₄

TABLA 12

Motivos para OleD	
SEQ ID NO: 127	L ₁ -V ₂ -T ₃ -G ₄ -G ₅ -[GN] ₆ -G ₇ -F ₈ -[LV] ₉ -G ₁₀ -X ₁₁ -X ₁₂ -L ₁₃ -X ₁₄ -X ₁₅ -X ₁₆ -L ₁₇
SEQ ID NO: 128	G ₁ -[VI] ₂ -[DT] ₃ -A ₄ -V ₅ -F ₆ -H ₇ -X ₈ -X ₉ -A ₁₀ -K ₁₁ -A ₁₂ -G ₁₃ -X ₁₄ -W ₁₅ -G ₁₆ -X ₁₇ -Y ₁₈ -D ₁₉ -X ₂₀ -X ₂₁ -X ₂₂ -X ₂₃ -A ₂₄ -N ₂₅ -V ₂₆
SEQ ID NO: 129	Y ₁ -T ₂ -S ₃ -T ₄ -P ₅ -S ₆ -V ₇ -[VT] ₈
SEQ ID NO: 130	A ₁ -T ₂ -V ₃ -A ₄ -L ₅ -R ₆ -P ₇ -[RH] ₈ -L ₉ -I ₁₀ -W ₁₁ -G ₁₂ -P ₁₃ -[GR] ₁₄ -D ₁₅
SEQ ID NO: 131	G ₁ -R ₂ -A ₃ -Y ₄ -F ₅ -I ₆ -X ₇ -X ₈ -X ₉ -X ₁₀ -P ₁₁
SEQ ID NO: 132	L ₁ -R ₂ -L ₃ -X ₄ -G ₅ -E ₆ -X ₇ -P ₈ -L ₉ -T ₁₀ -R ₁₁
SEQ ID NO: 133	R ₁ -D ₂ -F ₃ -G ₄ -Y ₅ -X ₆ -P ₇ -X ₈ -X ₉ -S ₁₀ -I ₁₁ -E ₁₂ -E ₁₃ -G ₁₄ -L ₁₅ -[RQ] ₁₆ -R ₁₇

5 La invención se refiere a un ácido nucleico aislado que codifica para un polipéptido que comprende una secuencia de motivo de aminoácido de OleA, OleB, OleC u OleD. Por ejemplo, el ácido nucleico aislado puede codificar para un polipéptido que comprende una secuencia de aminoácidos seleccionada del grupo que consiste en SEQ ID NO: 64-74 y 91-133.

10 El ácido nucleico aislado puede codificar para un polipéptido que comprende una secuencia de motivo de aminoácido de OleA. Por ejemplo, el ácido nucleico aislado puede codificar para un polipéptido que comprende una secuencia de aminoácidos seleccionada del grupo que consiste en SEQ ID NO: 64-69, 98-102 y 116-120. El ácido nucleico aislado puede codificar para un polipéptido que comprende una secuencia de motivo de aminoácido de OleB. Por ejemplo, el ácido nucleico aislado puede codificar para un polipéptido que comprende una secuencia de aminoácidos seleccionada del grupo que consiste en SEQ ID NO: 91-97. El ácido nucleico aislado puede codificar para un polipéptido que comprende una secuencia de motivo de aminoácido de OleC. Por ejemplo, el ácido nucleico aislado puede codificar para un polipéptido que comprende una secuencia de aminoácidos seleccionada del grupo que consiste en SEQ ID NO: 72-74, 103-108 y 121-126. El ácido nucleico aislado puede codificar para un polipéptido que comprende una secuencia de motivo de aminoácido de OleD. Por ejemplo, el ácido nucleico aislado puede codificar para un polipéptido que comprende una secuencia de aminoácidos seleccionada del grupo que consiste en
 20 SEQ ID NO: 70, 71, 109-115 y 127-133.

El ácido nucleico aislado puede codificar para un polipéptido que comprende más de una secuencia de motivo de aminoácido de OleA, OleB, OleC u OleD. Por ejemplo, el ácido nucleico aislado puede codificar para un polipéptido que comprende una, dos, tres o cuatro secuencias de motivo de aminoácido de Ole. Por ejemplo, el ácido nucleico aislado puede codificar para un polipéptido que comprende una secuencia de motivo de aminoácido de OleB y una
 25 secuencia de motivo de aminoácido de OleC.

Alternativamente, el ácido nucleico aislado puede codificar para más de un polipéptido que comprende una secuencia de motivo de aminoácido de Ole. Por ejemplo, el ácido nucleico aislado puede codificar para un primer polipéptido y un segundo polipéptido, en el que cada uno de los polipéptidos primero y segundo comprende una secuencia de motivo de aminoácido de Ole o en el que cada uno de los polipéptidos primero y segundo comprende una secuencia de aminoácidos seleccionada del grupo que consiste en SEQ ID NO: 64-74 y 91-133. El ácido nucleico aislado puede codificar para un primer polipéptido, un segundo polipéptido y un tercer polipéptido, en el que cada uno del primer, segundo y tercer polipéptido comprende una secuencia de motivo de aminoácido de Ole. Por ejemplo, el ácido nucleico aislado puede codificar para un primer polipéptido, un segundo polipéptido y un tercer polipéptido, en el que cada uno del primer, segundo y tercer polipéptido comprende una secuencia de aminoácidos seleccionada del grupo que consiste en SEQ ID NO: 64-74 y 91-133. El ácido nucleico aislado puede codificar para un primer polipéptido, un segundo polipéptido, un tercer polipéptido y un cuarto polipéptido, en el que cada uno del primer, segundo, tercer y cuarto polipéptido comprende una secuencia de motivo de aminoácido de Ole. Por ejemplo, el ácido nucleico aislado puede codificar para un primer polipéptido, un segundo polipéptido, un tercer polipéptido y un cuarto polipéptido, en el que cada uno del primer, segundo, tercer y cuarto polipéptido comprende una secuencia de aminoácidos seleccionada del grupo que consiste en SEQ ID NO: 64-74 y 91-133.

El ácido nucleico aislado puede codificar para un polipéptido de no más de 1500 residuos de aminoácido que comprende una secuencia de motivo de aminoácido de Ole. Por ejemplo, el ácido nucleico aislado puede codificar para un polipéptido de no más de 1500 residuos de aminoácido que comprende una secuencia de aminoácidos seleccionada del grupo que consiste en SEQ ID NO: 64-74 y 91-133. Cada uno de los polipéptidos primero y segundo puede ser de no más de 1500 residuos de aminoácido, cada uno de los polipéptidos primero, segundo y tercero puede ser de no más de 1500 residuos de aminoácido, o cada uno de los polipéptidos primero, segundo, tercer y cuarto es de no más de 1500 residuos de aminoácido.

El análisis bioinformático de los genes *ole* reveló que los genes se encuentran a menudo dentro de una región de pares de bases de 5000 a 10000 transcrita en la misma dirección que si estuviese en un operón. Los genes se han encontrado también en pares dentro de los organismos. Esta formación de pares incluye una fusión de proteínas entre OleB y OleC (en el presente documento denominada OleBC). Por ejemplo, se ha observado la fusión del gen *oleBC* en siete de los 67 organismos descritos anteriormente: *Arthrobacter aureescens*, ambas especies de *Clavibacter michiganensis*, *Kineococcus radiotolerans*, *Brevibacterium linens*, *Congregibacter litoralis* y *Micrococcus luteus*.

Identificar un organismo que contiene uno o más del/de los gen(es) *ole* indicaría que el organismo también produce hidrocarburos de manera natural. Se usaron técnicas bioinformáticas para identificar otros organismos secuenciados que contienen genes que pertenecen a estas familias génicas (véanse, por ejemplo, las tablas 1 y 2). Pueden usarse estos genes para conferir producción de hidrocarburos en su organismo huésped, así como en otros huéspedes cuando se expresan en otros organismos huésped. Un experto habitual en la técnica apreciará que pueden clonarse fácilmente secuencias de *oleA*, *oleB*, *oleC*, *oleD* y *oleBC* adicionales y usarse para elaborar hidrocarburos y productos intermedios de hidrocarburos. Puede encontrarse una lista a modo de ejemplo de proteínas OleA, OleB, OleC, OleD y OleBC adicionales identificadas usando los métodos dados a conocer en el presente documento en las tablas 1 y 2. Un experto en la técnica podría identificar o deducir las secuencias génicas a partir de las secuencias de proteínas OleA, OleB, OleC, OleD y OleBC expuestas en las tablas 1 y 2.

La invención se refiere a un ácido nucleico aislado que codifica para un polipéptido que comprende, consiste esencialmente en o consiste en, una secuencia de aminoácidos seleccionada del grupo que consiste en (a) una secuencia de aminoácidos de OleA, OleB, OleC, OleD u OleBC; (b) un homólogo de una secuencia de aminoácidos de OleA, OleB, OleC, OleD u OleBC; (c) una variante conservativa de una secuencia de aminoácidos de OleA, OleB, OleC, OleD u OleBC que comprende una o más sustituciones de aminoácidos conservativas; y (d) una secuencia de aminoácidos que tiene una identidad de secuencia de al menos el 35% (por ejemplo, al menos el 45%, al menos el 50%, al menos el 60%, al menos el 70%, al menos el 80%, al menos el 85%, al menos el 90%, al menos el 95%, al menos el 96%, al menos el 97%, al menos el 98%, al menos el 99%, o al menos el 99,5%) con una secuencia de aminoácidos de OleA, OleB, OleC, OleD u OleBC.

Por ejemplo, el ácido nucleico aislado puede codificar para un polipéptido que comprende, consiste esencialmente en o consiste en, una secuencia de aminoácidos seleccionada del grupo que consiste en (a) SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, SEQ ID NO: 10, SEQ ID NO: 12, SEQ ID NO: 14, SEQ ID NO: 16, SEQ ID NO: 18, SEQ ID NO: 20, SEQ ID NO: 22, SEQ ID NO: 88, SEQ ID NO: 135-464 y SEQ ID NO 478-487; (b) un homólogo de SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, SEQ ID NO: 10, SEQ ID NO: 12, SEQ ID NO: 14, SEQ ID NO: 16, SEQ ID NO: 18, SEQ ID NO: 20, SEQ ID NO: 22, SEQ ID NO: 88, SEQ ID NO: 135-464 y SEQ ID NO 478-487; (c) una variante conservativa de SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, SEQ ID NO: 10, SEQ ID NO: 12, SEQ ID NO: 14, SEQ ID NO: 16, SEQ ID NO: 18, SEQ ID NO: 20, SEQ ID NO: 22, SEQ ID NO: 88, SEQ ID NO: 135-464 y SEQ ID NO 478-487 que comprende una o más sustituciones de aminoácidos conservativas; y (d) una secuencia que tiene una identidad de secuencia de al menos el 35% (por ejemplo, al menos el 45%, al menos el 50%, al menos el 60%, al menos el 70%, al menos el 80%, al menos el 85%, al menos el 90%, al menos el 95%, al menos el 96%, al menos el 97%, al menos el 98%, al menos el 99% o al menos el 99,5%) con SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, SEQ ID NO: 10, SEQ ID NO: 12, SEQ ID NO: 14, SEQ ID NO: 16, SEQ ID NO: 18, SEQ ID NO: 20, SEQ ID NO: 22, SEQ ID NO: 88, SEQ ID NO: 135-464 y SEQ ID NO

- 5 El ácido nucleico aislado puede codificar para un polipéptido que comprende, consiste esencialmente en o consiste en, una secuencia de aminoácidos que es un homólogo de OleB. Por ejemplo, el ácido nucleico aislado puede codificar para un polipéptido que comprende, consiste esencialmente en o consiste en, una secuencia de aminoácidos seleccionada del grupo que consiste en un homólogo de SEQ ID NO: 10 y SEQ ID NO: 230-326. El ácido nucleico aislado puede codificar para un polipéptido que comprende, consiste esencialmente en o consiste en, un homólogo de SEQ ID NO: 10.
- 10 El ácido nucleico aislado puede codificar para un polipéptido que comprende, consiste esencialmente en o consiste en, una secuencia de aminoácidos que es un homólogo de OleC. Por ejemplo, el ácido nucleico aislado puede codificar para un polipéptido que comprende, consiste esencialmente en o consiste en, una secuencia de aminoácidos seleccionada del grupo que consiste en un homólogo de SEQ ID NO: 6, SEQ ID NO: 14, SEQ ID NO: 20, SEQ ID NO: 88 y SEQ ID NO: 327-402. El ácido nucleico aislado puede codificar para un polipéptido que comprende, consiste esencialmente en o consiste en, una secuencia de aminoácidos seleccionada del grupo que consiste en un homólogo de SEQ ID NO: 6, SEQ ID NO: 14, SEQ ID NO: 20 y SEQ ID NO: 88.
- 15 El ácido nucleico aislado puede codificar para un polipéptido que comprende, consiste esencialmente en o consiste en, una secuencia de aminoácidos que es un homólogo de OleD. Por ejemplo, el ácido nucleico aislado puede codificar para un polipéptido que comprende, consiste esencialmente en o consiste en, una secuencia de aminoácidos seleccionada del grupo que consiste en un homólogo de SEQ ID NO: 8, SEQ ID NO: 16, SEQ ID NO: 22, SEQ ID NO: 142-149 y SEQ ID NO: 403-464. El ácido nucleico aislado puede codificar para un polipéptido que comprende, consiste esencialmente en o consiste en, una secuencia de aminoácidos seleccionada del grupo que consiste en un homólogo de SEQ ID NO: 8, SEQ ID NO: 16 y SEQ ID NO: 22.
- 20 El ácido nucleico aislado puede codificar para un polipéptido que comprende, consiste esencialmente en o consiste en, una secuencia de aminoácidos que es un homólogo de OleBC. Por ejemplo, el ácido nucleico aislado puede codificar para un polipéptido que comprende, consiste esencialmente en o consiste en, una secuencia de aminoácidos seleccionada del grupo que consiste en un homólogo de SEQ ID NO: 135-141.
- 25 Un homólogo de una proteína OleA, OleB, OleC, OleD u OleBC es uno que se comporta de manera funcional sustancialmente igual que una proteína OleA, OleB, OleC, OleD u OleBC, por ejemplo, en cuanto a tener actividad hidrocarburo sintasa. Por ejemplo, una proteína OleA, OleB, OleC, OleD u OleBC y un homólogo de proteína OleA, OleB, OleC, OleD u OleBC no tienen necesariamente secuencias de aminoácidos similares. Sin embargo, tienen actividades hidrocarburo sintasa similares.
- 30 Los cálculos de "homología" entre dos secuencias pueden realizarse tal como sigue. Se alinean las secuencias con fines de comparación óptima (por ejemplo, pueden introducirse huecos en una o ambas de una primera y una segunda secuencia de ácido nucleico o aminoácidos para lograr una alineación óptima y pueden ignorarse 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 el 30%, preferiblemente al menos el 40%, más preferiblemente al menos el 50%, incluso más preferiblemente al menos el 60%, e incluso más preferiblemente al menos el 70%, al menos el 80%, al menos el 90%, o el 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 en 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 compartidas por las secuencias, teniendo en cuenta el número de huecos y la longitud de cada hueco, que tienen que introducirse para lograr una alineación óptima de las dos secuencias.
- 35
- 40
- 45 El ácido nucleico aislado puede codificar para un polipéptido que comprende, consiste esencialmente en o consiste en, una secuencia de aminoácidos que es una variante conservativa de una secuencia de aminoácidos de OleA, OleB, OleC, OleD u OleBC. Por ejemplo, el ácido nucleico aislado puede codificar para un polipéptido que comprende, consiste esencialmente en o consiste en, una secuencia de aminoácidos seleccionada del grupo que consiste en una variante conservativa de SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, SEQ ID NO: 10, SEQ ID NO: 12, SEQ ID NO: 14, SEQ ID NO: 16, SEQ ID NO: 18, SEQ ID NO: 20, SEQ ID NO: 22, SEQ ID NO: 50 88, SEQ ID NO: 135-464 y SEQ ID NO 478-487. El ácido nucleico aislado puede codificar para un polipéptido que comprende, consiste esencialmente en o consiste en, una secuencia de aminoácidos seleccionada del grupo que consiste en una variante conservativa de SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, SEQ ID NO: 10, SEQ ID NO: 12, SEQ ID NO: 14, SEQ ID NO: 16, SEQ ID NO: 18, SEQ ID NO: 20, SEQ ID NO: 22 y SEQ ID NO: 88.
- 55 El ácido nucleico aislado puede codificar para un polipéptido que comprende, consiste esencialmente en o consiste en, una secuencia de aminoácidos que es una variante conservativa de una secuencia de aminoácidos de OleA. Por ejemplo, el ácido nucleico aislado puede codificar para un polipéptido que comprende, consiste esencialmente en o consiste en, una secuencia de aminoácidos seleccionada del grupo que consiste en una variante conservativa de SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 12, SEQ ID NO: 18 y SEQ ID NO: 150-229. El ácido nucleico aislado puede codificar para un polipéptido que comprende, consiste esencialmente en o consiste en, una secuencia de
- 60

aminoácidos seleccionada del grupo que consiste en una variante conservativa de SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 12 y SEQ ID NO: 18.

El ácido nucleico aislado puede codificar para un polipéptido que comprende, consiste esencialmente en o consiste en, una secuencia de aminoácidos que es una variante conservativa de una secuencia de aminoácidos de OleB. Por ejemplo, el ácido nucleico aislado puede codificar para un polipéptido que comprende, consiste esencialmente en o consiste en, una secuencia de aminoácidos seleccionada del grupo que consiste en una variante conservativa de SEQ ID NO: 10 y SEQ ID NO: 230-326. El ácido nucleico aislado puede codificar para un polipéptido que comprende, consiste esencialmente en o consiste en, una secuencia de aminoácidos que es una variante conservativa de SEQ ID NO: 10.

El ácido nucleico aislado puede codificar para un polipéptido que comprende, consiste esencialmente en o consiste en, una secuencia de aminoácidos que es una variante conservativa de una secuencia de aminoácidos de OleC. Por ejemplo, el ácido nucleico aislado puede codificar para un polipéptido que comprende, consiste esencialmente en o consiste en, una secuencia de aminoácidos seleccionada del grupo que consiste en una variante conservativa de SEQ ID NO: 6, SEQ ID NO: 14, SEQ ID NO: 20, SEQ ID NO: 88 y SEQ ID NO: 327-402. El ácido nucleico aislado puede codificar para un polipéptido que comprende, consiste esencialmente en o consiste en, una secuencia de aminoácidos seleccionada del grupo que consiste en una variante conservativa de SEQ ID NO: 6, SEQ ID NO: 14, SEQ ID NO: 20 y SEQ ID NO: 88.

El ácido nucleico aislado puede codificar para un polipéptido que comprende, consiste esencialmente en o consiste en, una secuencia de aminoácidos que es una variante conservativa de una secuencia de aminoácidos de OleD. Por ejemplo, el ácido nucleico aislado puede codificar para un polipéptido que comprende, consiste esencialmente en o consiste en, una secuencia de aminoácidos seleccionada del grupo que consiste en una variante conservativa de SEQ ID NO: 8, SEQ ID NO: 16, SEQ ID NO: 22, SEQ ID NO: 142-149 y SEQ ID NO: 403-464. El ácido nucleico aislado puede codificar para un polipéptido que comprende, consiste esencialmente en o consiste en, una secuencia de aminoácidos seleccionada del grupo que consiste en una variante conservativa de SEQ ID NO: 8, SEQ ID NO: 16 y SEQ ID NO: 22.

El ácido nucleico aislado puede codificar para un polipéptido que comprende, consiste esencialmente en o consiste en, una secuencia de aminoácidos que es una variante conservativa de una secuencia de aminoácidos de OleBC. Por ejemplo, el ácido nucleico aislado puede codificar para un polipéptido que comprende, consiste esencialmente en o consiste en, una secuencia de aminoácidos seleccionada del grupo que consiste en una variante conservativa de SEQ ID NO: 135-141.

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

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 puede incluir también, por ejemplo, variantes "alélicas", "de corte y empalme", "de especie" o "polimórficas". Una variante de corte y empalme puede tener identidad significativa con un polinucleótido de referencia, pero tendrá generalmente un número mayor o menor de polinucleótidos debido al corte y empalme alternativo de exones durante el procesamiento del ARNm. El polipéptido correspondiente puede presentar dominios funcionales adicionales o no tener dominios. Las variantes de especies son secuencias de polinucleótido que varían de una especie a otra. Los polipéptidos resultantes tendrán generalmente una identidad de aminoácidos significativa entre sí. Una variante polimórfica es una variación en la secuencia de polinucleótido de un gen particular entre individuos de una especie dada.

Las sustituciones de aminoácidos conservativas pueden ser cualquier sustitución de aminoácidos mostrada en la tabla 13. Específicamente, las sustituciones de aminoácidos conservativas pueden seleccionarse del grupo que consiste en alanina a D-Ala, Gly, beta-Ala, L-Cys, D-Cys; arginina a D-Arg, Lys, D-Lys, homo-Arg, D-homo-Arg, Met, Ile, D-Met, D-Ile, Orn, D-Orn; asparagina a D-Asn, Asp, D-Asp, Glu, D-Glu, Gln, D-Gln; ácido aspártico a D-Asp, D-Asn, Asn, Glu, D-Glu, Gln, D-Gln; cisteína a D-Cys, S-Me-Cys, Met, D-Met, Thr, D-Thr; ácido glutámico a D-Glu, D-Asp, Asp, Asn, D-Asn, Gln, D-Gln; glicina a Ala, D-Ala, Pro, D-Pro, b-Ala, Acp; isoleucina a D-Ile, Val, D-Val, Leu, D-Leu, Met, D-Met; leucina a D-Leu, Val, D-Val, Leu, D-Leu, Met, D-Met; lisina a D-Lys, Arg, D-Arg, homo-Arg, D-homo-Arg, Met, D-Met, Ile, D-Ile, Orn, D-Orn; metionina a D-Met, S-Me-Cys, Ile, D-Ile, Leu, D-Leu, Val, D-Val; fenilalanina a D-Phe, Tyr, D-Thr, L-Dopa, His, D-His, Trp, D-Trp, trans-3, 4 ó 5-fenilprolina, cis-3, 4 ó 5-fenilprolina; prolina a D-Pro, ácido L-1-tioazolidin-4-carboxílico, ácido D- o L-1-oxazolidin-4-carboxílico; serina a D-Ser, Thr, D-Thr, alo-Thr, Met, D-Met, Met(O), D-Met(O), L-Cys, D-Cys; treonina a D-Thr, Ser, D-Ser, alo-Thr, Met, D-Met, Met(O), D-Met(O), Val, D-Val; tirosina a D-Tyr, Phe, D-Phe, L-Dopa, His, D-His; y valina a D-Val, Leu, D-Leu, Ile, D-Ile, Met, D-Met.

TABLA 13

Para aminoácido	Código	Sustituir con cualquiera de
Alanina	A	D-Ala, Gly, beta-Ala, L-Cys, D-Cys
Arginina	R	D-Arg, Lys, D-Lys, homo-Arg, D-homo-Arg, Met, Ile, D-Met, D-Ile, Orn, D-Orn
Asparagina	N	D-Asn, Asp, D-Asp, Glu, D-Glu, Gln, D-Gln
Ácido aspártico	D	D-Asp, D-Asn, Asn, Glu, D-Glu, Gln, D-Gln
Cisteína	C	D-Cys, S-Me-Cys, Met, D-Met, Thr, D-Thr
Ácido glutámico	E	D-Glu, D-Asp, Asp, Asn, D-Asn, Gln, D-Gln
Glicina	G	Ala, D-Ala, Pro, D-Pro, b-Ala, Acp
Isoleucina	I	D-Ile, Val, D-Val, Leu, D-Leu, Met, D-Met
Leucina	L	D-Leu, Val, D-Val, Leu, D-Leu, Met, D-Met
Lisina	K	D-Lys, Arg, D-Arg, homo-Arg, D-homo-Arg, Met, D-Met, Ile, D-Ile, Orn, D-Orn
Metionina	M	D-Met, S-Me-Cys, Ile, D-Ile, Leu, D-Leu, Val, D-Val
Fenilalanina	F	D-Phe, Tyr, D-Thr, L-Dopa, His, D-His, Trp, D-Trp, Trans-3, 4 ó 5-fenilprolina, cis-3, 4 ó 5-fenilprolina
Prolina	P	D-Pro, ácido L-1-tioazolidin-4-carboxílico, ácido D- o L-1-oxazolidin-4-carboxílico
Serina	S	D-Ser, Thr, D-Thr, alo-Thr, Met, D-Met, Met(O), D-Met(O), L-Cys, D-Cys
Treonina	T	D-Thr, Ser, D-Ser, alo-Thr, Met, D-Met, Met(O), D-Met(O), Val, D-Val
Tirosina	Y	D-Tyr, Phe, D-Phe, L-Dopa, His, D-His
Valina	V	D-Val, Leu, D-Leu, Ile, D-Ile, Met, D-Met

- Una variante conservativa de una proteína Ole puede comprender una o más sustituciones de aminoácidos conservativas. Por ejemplo, una variante conservativa puede tener 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40 ó 50 sustituciones de aminoácidos conservativas. En una realización preferida, una variante conservativa tiene no más de 50 sustituciones de aminoácidos conservativas. Por ejemplo, una variante conservativa puede tener no más de 3, 5, 10, 15, 20, 25, 30, 35, 40 ó 50 sustituciones de aminoácidos conservativas. Una variante conservativa de SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, SEQ ID NO: 10, SEQ ID NO: 12, SEQ ID NO: 14, SEQ ID NO: 16, SEQ ID NO: 18, SEQ ID NO: 20, SEQ ID NO: 22, SEQ ID NO: 88, SEQ ID NO: 135-464 y SEQ ID NO 478-487 es una que se comporta de manera funcional sustancialmente igual que la secuencia de aminoácidos respectiva sin las sustituciones de aminoácidos. Uno cualquiera de los ensayos proporcionados en el presente documento puede usarse para evaluar la actividad. En algunos ejemplos, una variante conservativa de una proteína OleA puede someterse a ensayo para determinar la actividad hidrocarburo sintasa, tal como actividad de condensación de acilo, actividad cetona alifática sintasa y/o actividad olefina sintasa. Tal como se usa en el presente documento, el término "sintasa" se refiere a una enzima que cataliza un proceso de síntesis. Tal como se usa en el presente documento, el término sintasa incluye sintasas, sintetetasas y ligasas.
- En otros ejemplos, pueden someterse a ensayo variantes conservativas de OleC y OleD para determinar la actividad tal como se describe en el presente documento. La variante conservativa puede tener, por ejemplo, una sustitución de aminoácido conservativa, dos sustituciones de aminoácido, tres sustituciones de aminoácido, cuatro sustituciones de aminoácido o cinco o más sustituciones de aminoácido en una secuencia de aminoácidos, siempre que se mantenga la actividad de la proteína.
- El ácido nucleico aislado puede codificar para un polipéptido que comprende, consiste esencialmente en o consiste en, una secuencia de aminoácidos seleccionada del grupo que consiste en una variante conservativa de SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, SEQ ID NO: 10, SEQ ID NO: 12, SEQ ID NO: 14, SEQ ID NO: 16, SEQ ID NO: 18, SEQ ID NO: 20, SEQ ID NO: 22 y SEQ ID NO: 88 que comprende una o más sustituciones de aminoácidos conservativas seleccionada del grupo que consiste en Ala a Cys, Gly o Ser; Arg a Ile, Lys, Met u Orn; Asn a Asp, Gln, Glu o His; Asp a Asn, Gln o Glu; Cys a Met, Ser o Thr; Gln a Asn, Asp o Glu; Glu a Asn, Asp o Gln; Gly a Acp, Ala o Pro; His a Asn o Gln; Ile a Leu, Met o Val; Leu a Ile, Met o Val; Lys a Arg, Gln, Glu, Ile, Met u Orn; Met a Cys, Ile, Leu o Val; Phe a His, L-Dopa, Leu, Met, Thr, Trp, Tyr, 3-fenilprolina, 4-fenilprolina o 5-fenilprolina; Pro a ácido L-1-tioazolidin-4-carboxílico o ácido D- o L-1-oxazolidin-4-carboxílico; Ser a Cys, Met o Thr; Thr a Met, Ser o Val; Trp a Tyr; Tyr a L-Dopa, His o Phe; y Val a Ile, Leu o Met.
- En algunas realizaciones, una variante conservativa incluye una o más sustituciones de aminoácidos conservativas en comparación con la secuencia de la que se deriva, y aún conserva su actividad respectiva. Por ejemplo, una variante conservativa puede conservar al menos el 10% de la actividad biológica de la proteína original de la que se deriva, o alternativamente, al menos el 20%, al menos el 30%, o al menos el 40% de la actividad biológica de la proteína original. Una variante conservativa puede conservar al menos el 50% de la actividad biológica de la proteína original de la que se deriva. Las sustituciones de aminoácidos conservativas de una variante conservativa pueden producirse en cualquier dominio de la proteína. Las sustituciones de aminoácidos conservativas pueden dar como resultado actividad biológica potenciada en comparación con la proteína original. Por ejemplo, la variante conservativa puede tener una actividad biológica de al menos el 100% de la actividad biológica de la proteína original de la que se deriva, o alternativamente, al menos el 110%, al menos el 120%, al menos el 150%, al menos el 200% o al menos el 1000% de la actividad biológica de la proteína original de la que se deriva.

consiste en, una secuencia de aminoácidos seleccionada del grupo que consiste en una secuencia que tiene una identidad de secuencia de al menos el 35% (por ejemplo, al menos el 45%, al menos el 50%, al menos el 60%, al menos el 70%, al menos el 80%, al menos el 85%, al menos el 90%, al menos el 95%, al menos el 96%, al menos el 97%, al menos el 98%, al menos el 99% o al menos el 99,5%) con SEQ ID NO: 135-141.

- 5 Un ácido nucleico aislado que codifica para un polipéptido que comprende, consiste esencialmente en o consiste en, una secuencia de aminoácidos que tiene una identidad de secuencia de entre el 35% y el 100% con una secuencia de aminoácidos que codifica para una proteína Ole también se proporciona en el presente documento. Por ejemplo, el ácido nucleico aislado puede codificar para un polipéptido que comprende, consiste esencialmente en o consiste en, una secuencia de aminoácidos que tiene una identidad de secuencia de al menos el 35%, al menos el 45%, al menos el 50%, al menos el 60%, al menos el 70%, al menos el 80%, al menos el 85%, al menos el 90%, al menos el 95%, al menos el 96%, al menos el 97%, al menos el 98%, al menos el 99% o al menos el 99,5% con una secuencia de aminoácidos que codifica para una proteína Ole.

15 Aunque se dan a conocer realizaciones particulares de hidrocarburos y productos intermedios de hidrocarburos que forman secuencias de ácido nucleico y aminoácidos, se entenderá que pueden aislarse secuencias que tienen características estructurales similares de otros organismos. Estas secuencias recién aisladas pueden someterse a ensayo para determinar la actividad hidrocarburo sintasa (véanse las tablas 1 y 2 para una lista de ejemplos no limitativos, específicos de secuencias relacionadas). Además, se entenderá que otras formas funcionalmente equivalentes de secuencias de ácido nucleico y aminoácidos dadas a conocer en el presente documento pueden identificarse fácilmente y/o generarse usando técnicas de biología molecular convencionales, incluyendo, por ejemplo, mutagénesis dirigida al sitio, mutagénesis de cebador M13, PCR propensa a error, PCR sexual, síntesis de ADN o intercambio de ADN. Se proporcionan detalles de muchas de estas técnicas en Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual*, 3ª ed., vol. 1-3, Cold Spring Harbor, Nueva York (2000).

20 Por tanto, además de secuencias estructuralmente relacionadas y secuencias homólogas, la divulgación también abarca secuencias de aminoácidos que tienen una identidad de secuencia de al menos el 30%, el 35%, el 40%, el 45%, el 50%, el 55%, el 60%, el 65%, el 70%, el 75%, el 80%, el 85%, el 90% o el 95% con SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, SEQ ID NO: 10, SEQ ID NO: 12, SEQ ID NO: 14, SEQ ID NO: 16, SEQ ID NO: 18, SEQ ID NO: 20, SEQ ID NO: 22, SEQ ID NO: 88 y/o SEQ ID NO: 135-464. También se abarcan secuencias de aminoácidos que tienen una identidad de secuencia de al menos el 96%, el 97%, el 98%, el 99% o el 99,5% con SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, SEQ ID NO: 10, SEQ ID NO: 12, SEQ ID NO: 14, SEQ ID NO: 16, SEQ ID NO: 18, SEQ ID NO: 20, SEQ ID NO: 22, SEQ ID NO: 88 y/o SEQ ID NO: 135-464.

25 Las secuencias que conservan similitud estructural y funcional con respecto a OleA, OleB, OleC, OleBC y OleD pueden identificarse mediante varios métodos conocidos. Uno de tales métodos implica el examen de secuencias genómicas para la alineación de secuencias con la(s) secuencia(s) conocida(s). Los métodos para alinear secuencias para comparación se conocen bien en la técnica. Se describen diversos programas y algoritmos de alineación en: Smith *et al.*, *Adv. Appl. Math.*, 2: 482 (1981); Needleman *et al.*, *J. Mol. Biol.*, 48: 443 (1970); Pearson *et al.*, *Proc. Natl. Acad. Sci. USA*, 85: 2444 (1988); Higgins *et al.*, *Gene*, 73: 237-244 (1988); Higgins & Sharp, *CABIOS*, 5: 151-153 (1989); Corpet *et al.*, *Nucleic Acids Research*, 16: 10881-10890 (1988); Huang *et al.*, *CABIOS*, 8: 155-165 (1992); y Pearson *et al.*, *Methods in Molecular Biology*, 24: 307-331 (1994). Altschul *et al.*, *J. Mol. Biol.*, 215: 403-410 (1990), presentan una descripción detallada de los métodos de alineación de secuencias y cálculos de homología.

35 El porcentaje de homología entre dos secuencias de aminoácidos puede determinarse usando el algoritmo de Needleman, citado anteriormente, que se ha incorporado en el programa GAP en el paquete de software GCG, usando o bien una matriz BLOSUM62 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. El porcentaje de homología entre dos secuencias de nucleótidos puede determinarse también 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 profesional no tiene claro qué parámetros deben aplicarse para determinar si una molécula está dentro de una limitación de homología de las reivindicaciones) es una matriz de puntuación BLOSUM62 con una penalización de hueco de 12, una penalización de extensión de hueco de 4 y una penalización de hueco de desplazamiento del marco de lectura de 5.

40 En el presente documento se da a conocer un ácido nucleico aislado que codifica para más de un polipéptido, en el que cada polipéptido comprende, consiste esencialmente en o consiste en una secuencia de aminoácidos seleccionada del grupo que consiste en (a) una secuencia de aminoácidos de OleA, OleB, OleC, OleD u OleBC; (b) un homólogo de una secuencia de aminoácidos de OleA, OleB, OleC, OleD, u OleBC; (c) una variante conservativa de una secuencia de aminoácidos de OleA, OleB, OleC, OleD u OleBC que comprende una o más sustituciones de aminoácidos conservativas; y (d) una secuencia de aminoácidos que tiene una identidad de secuencia de al menos el 35% (por ejemplo, al menos el 45%, al menos el 50%, al menos el 60%, al menos el 70%, al menos el 80%, al menos el 85%, al menos el 90%, al menos el 95%, al menos el 96%, al menos el 97%, al menos el 98%, al menos el 99% o al menos el 99,5%) con una secuencia de aminoácidos de OleA, OleB, OleC, OleD u OleBC. El ácido nucleico aislado puede codificar para un primer polipéptido y un segundo polipéptido, en el que cada uno del primer y segundo polipéptido comprende, consiste esencialmente en o consiste en una secuencia de aminoácidos

conservativa de SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, SEQ ID NO: 10, SEQ ID NO: 12, SEQ ID NO: 14, SEQ ID NO: 16, SEQ ID NO: 18, SEQ ID NO: 20, SEQ ID NO: 22, SEQ ID NO: 88, SEQ ID NO: 135-464 y SEQ ID NO 478-487 que comprende una o más sustituciones de aminoácidos conservativas; y (d) una secuencia que tiene una identidad de secuencia de al menos el 35% (por ejemplo, al menos el 45%, al menos el 50%, al menos el 60%, al menos el 70%, al menos el 80%, al menos el 85%, al menos el 90%, al menos el 95%, al menos el 96%, al menos el 97%, al menos el 98%, al menos el 99% o al menos el 99,5%) con SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, SEQ ID NO: 10, SEQ ID NO: 12, SEQ ID NO: 14, SEQ ID NO: 16, SEQ ID NO: 18, SEQ ID NO: 20, SEQ ID NO: 22, SEQ ID NO: 88, SEQ ID NO: 135-464 y SEQ ID NO 478-487.

El ácido nucleico aislado puede codificar para un polipéptido que tiene 1500 residuos de aminoácido, 1400 residuos de aminoácido, 1300 residuos de aminoácido, 1200 residuos de aminoácido, 1100 residuos de aminoácido, 1000 residuos de aminoácido, 900 residuos de aminoácido, 800 residuos de aminoácido, 700 residuos de aminoácido, 600 residuos de aminoácido, 500 residuos de aminoácido, 400 residuos de aminoácido o 300 residuos de aminoácido. El ácido nucleico aislado puede codificar para un polipéptido que tiene no más de 1500 (por ejemplo, no más de 1500, 1400, 1300, 1200, 1100, 1000, 900, 800, 700, 600, 500, 400, 300 ó 200) residuos de aminoácido.

El ácido nucleico aislado puede codificar para un polipéptido que tiene al menos 100 residuos de aminoácido. Por ejemplo, el ácido nucleico aislado puede codificar para un polipéptido que tiene al menos 200 residuos de aminoácido, al menos 250 residuos de aminoácido o al menos 300 residuos de aminoácido. Alternativamente, el ácido nucleico aislado puede codificar para un polipéptido que tiene no menos de 100 (por ejemplo, no menos de 100, 150, 200 ó 250) residuos de aminoácido. El ácido nucleico aislado puede codificar para un polipéptido que tiene entre 200 residuos de aminoácido y 1500 residuos de aminoácido, entre 300 residuos de aminoácido y 1000 residuos de aminoácido, entre 500 residuos de aminoácido y 800 residuos de aminoácido, o entre 600 residuos de aminoácido y 1000 residuos de aminoácido.

El ácido nucleico aislado puede aislarse de una bacteria, una planta, un insecto, una levadura, un hongo o un animal (por ejemplo, un mamífero). Cuando el ácido nucleico se aísla de una bacteria, la bacteria puede ser cualquier género de bacteria. Por ejemplo, la bacteria puede ser de un género seleccionado del grupo que consiste en *Anaeromyxobacter*, *Arthrobacter*, *Bdellovibrio*, *Blastopirellula*, *Brevibacterium*, *Burkholderia*, *Candidatus*, *Chloroflexus*, *Clavibacter*, *Clostridium*, *Colwellia*, *Congregibacter*, *Desulfatibacillum*, *Desulfococcus*, *Desulfotalea*, *Desulfuromonas*, *Flavobacteriales*, *Gemmata*, *Geobacter*, *Hahella*, *Jannaschia*, *Kineococcus*, *Lentisphaera*, *Maricaulis*, *Marinobacter*, *Micrococcus*, *Microscilla*, *Moritella*, *Mycobacterium*, *Neisseria*, *Nocardia*, *Opiritaceae*, *Opiritus*, *Paenibacillus*, *Pelobacter*, *Photobacterium*, *Photorhabdus*, *Planctomyces*, *Plesiocystis*, *Pseudoalteromonas*, *Psychromonas*, *Ralstonia*, *Rhodococcus*, *Rhodopirellula*, *Roseovarius*, *Shewanella*, *Stenotrophomonas*, *Streptomyces*, *Xanthomonas* y *Xilella*.

Más específicamente, el ácido nucleico puede aislarse de cualquier bacteria seleccionada del grupo que consiste en *Anaeromyxobacter*, *Arthrobacter aurescens*, *Arthrobacter chlorophenolicus*, *Arthrobacter* sp. FB24, *Bdellovibrio bacteriovorus*, *Blastopirellula marina*, *Brevibacterium linens*, *Burkholderia ambifaria*, *Burkholderia cenocepacia*, *Burkholderia oklahomensis*, *Burkholderia pseudomallei*, *Burkholderia* sp. 383, *Candidatus Kuenenia stuttgartiensis*, *Chloroflexus aggregans*, *Chloroflexus aurantiacus*, *Clavibacter michiganensis* subsp. *michiganensis*, *Clavibacter michiganensis* subsp. *Sepedonicus*, *Clostridium botulinum*, *Colwellia psychrerythraea*, *Congregibacter litoralis*, *Desulfatibacillum alkenivorans*, *Desulfococcus oleovorans*, *Desulfotalea psychrophila*, *Desulfuromonas acetoxidans*, *Flavobacteriales bacterium*, *Gemmata obscuriglobus*, *Geobacter bemidjensis*, *Geobacter lovleyi*, *Geobacter* sp. FRC-32, *Geobacter uraniumreducens*, *Hahella chejuensis*, *Jannaschia* sp. CCS1, *Kineococcus radiotolerans*, *Lentisphaera araneosa*, *Maricaulis maris*, *Marinobacter algicola*, *Marinobacter aquaeolei*, *Micrococcus luteus*, *Microscilla marina*, *Moritella* sp. PE36, *Mycobacterium avium*, *Mycobacterium marinum*, *Mycobacterium* sp. GP1, *Mycobacterium tuberculosis*, *Mycobacterium vanbaalenii*, *Neisseria gonorrhoeae*, *Nocardia farcinica*, *Opiritaceae bacterium*, *Opiritus terrae*, *Paenibacillus* sp. JDR-2, *Pelobacter propionicus*, *Photobacterium profundum*, *Photobacterium profundum*, *Photorhabdus luminescens*, *Planctomyces maris*, *Plesiocystis pacifica*, *Pseudoalteromonas atlantica*, *Psychromonas ingrahamii*, *Psychromonas* sp. CNPT3, *Ralstonia pickettii*, *Rhodococcus rhodochrous*, *Rhodococcus* sp., *Rhodopirellula baltica*, *Roseovarius nubinhibens* ISM, *Shewanella amazonensis* SB2B, *Shewanella baltica* OS155, *Shewanella baltica*, *Shewanella benthica*, *Shewanella denitrificans*, *Shewanella frigidimarina*, *Shewanella halifaxensis*, *Shewanella loihica*, *Shewanella oneidensis*, *Shewanella pealeana*, *Shewanella putrefaciens*, *Shewanella sediminis*, *Shewanella* sp. ANA-3, *Shewanella* sp. MR-4, *Shewanella* sp. MR-7, *Shewanella* sp. W3-18-1, *Shewanella woodyi*, *Stenotrophomonas maltophilia*, *Streptomyces ambofaciens*, *Xanthomonas axonopodis*, *Xanthomonas campestris*, *Xanthomonas oryzae* y *Xilella fastidiosa*. En una realización preferida, el ácido nucleico se aísla de una cepa de *Stenotrophomonas maltophilia* (por ejemplo, *Stenotrophomonas maltophilia* ATCC 17679, *Stenotrophomonas maltophilia* ATCC 17674, *Stenotrophomonas maltophilia* ATCC 17445, *Stenotrophomonas maltophilia* ATCC 17666, *Stenotrophomonas maltophilia* K279a o *Stenotrophomonas maltophilia* R551-3).

Cuando una secuencia genómica no está disponible para una especie de interés particular, pueden amplificarse secuencias relacionadas a partir de ADN genómico usando métodos de PCR convencionales. Brevemente, se extrae ADN genómico de las células de interés mediante uno cualquiera de una variedad de métodos bien conocidos. Sambrook *et al.*, citado anteriormente, y Ausubel *et al.*, *Current Protocols in Molecular Biology*, Greene Publ. Assoc. & Wiley-Intersciences (1989), proporcionan descripciones de métodos para el aislamiento de ADN.

Generalmente, puede usarse cualquier organismo como fuente de tal ADN. El ADN extraído se usa entonces como molde para realizar una reacción en cadena de la polimerasa. Puede requerirse el uso de cebadores degenerados para PCR. Se describen métodos y condiciones para PCR, por ejemplo, en PCR Protocols, A Guide to Methods and Applications, Innis *et al.* (eds.), Academic Press, Inc., San Diego, California, 1990.

- 5 La selección de cebadores de amplificación se realizará según el gen particular que va a amplificarse. En la tabla 14, a continuación, se muestran ejemplos de cebadores específicos de uso. Sin embargo, estos cebadores son ilustrativos solo. Un experto habitual en la técnica apreciará que pueden derivarse muchos cebadores diferentes de las secuencias de ácido nucleico de *oleA*, *oleB*, *oleC*, *oleBC* y *oleD*. Pueden requerirse variaciones en las condiciones de amplificación para acomodar cebadores y amplicones de composición y longitudes diferentes. Tales consideraciones se conocen bien en la técnica y se comentan, por ejemplo, en PCR Protocols, A Guide to Methods and Applications, Innis *et al.* (eds.), Academic Press, Inc., San Diego, California, 1990.

- 15 La secuenciación de los productos de PCR obtenidos mediante estos procedimientos de amplificación puede usarse para facilitar la confirmación de la secuencia amplificada y proporcionar información sobre la variación natural de esta secuencia en diferentes especies. Los oligonucleótidos derivados de las secuencias de *OleA*, *OleB*, *OleC*, *OleBC* y *OleD* proporcionadas pueden usarse en tales métodos de secuenciación. Las secuencias ortólogas estrechamente relacionadas de *OleA*, *OleB*, *OleC*, *OleBC* y *OleD* pueden compartir una identidad de secuencia de al menos el 30%, el 35%, el 40%, el 45%, el 50%, el 55%, el 60%, el 65%, el 70%, el 75%, el 80%, el 85%, el 90%, el 95%, el 98%, el 99% o el 99,5% con las secuencias de *OleA*, *OleB*, *OleC*, *OleBC* y *OleD* dadas a conocer (véanse, por ejemplo, las tablas 1 y 2).

- 20 El ácido nucleico puede seleccionarse del grupo que consiste en SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7, SEQ ID NO: 9, SEQ ID NO: 11, SEQ ID NO: 13, SEQ ID NO: 15, SEQ ID NO: 17, SEQ ID NO: 19, SEQ ID NO: 21, SEQ ID NO: 23 y SEQ ID NO: 87.

- 25 También se dan a conocer en el presente documento constructos de ácido nucleico recombinante que incluyen uno o más ácidos nucleicos aislados que codifican para proteínas *Ole*, homólogos de proteínas *Ole*, variantes conservativas de proteínas *Ole* y/o secuencias que tienen una identidad de secuencia de al menos el 35% (por ejemplo, al menos el 45%, al menos el 50%, al menos el 60%, al menos el 70%, al menos el 80%, al menos el 85%, al menos el 90%, al menos el 95%, al menos el 96%, al menos el 97%, al menos el 98%, al menos el 99% o al menos el 99,5%) con proteínas *Ole*. Los constructos de ácido nucleico recombinante a modo de ejemplo incluyen vectores de clonación, vectores de expresión u operones sintéticos.

- 30 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 replicarse de manera extra cromosómica). Vectores útiles son los que pueden replicarse de manera autónoma y/o expresar ácidos nucleicos a los que están unidos. Los vectores que pueden dirigir la expresión de genes a los que están unidos de manera operativa se denominan en el presente documento "vectores de expresión". En general, los vectores de expresión de utilidad en técnicas de ADN recombinante están a menudo en forma de "plásmidos" que se refieren generalmente a bucles de ADN bicatenarios circulares que, en su forma de vector, no están unidos al cromosoma. Tal como se usa en el presente documento, "plásmido" y "vector" se usan de manera intercambiable, puesto que el plásmido es la forma más frecuentemente usada de vector. Sin embargo, también se incluyen otras formas de vectores de expresión tales que sirven para funciones equivalentes y que se conocieron en la técnica posteriormente a esto.

- 35 Tanto los vectores de clonación como de expresión contienen secuencias de nucleótidos que permiten que los vectores se repliquen en uno o más organismos recombinantes adecuados. En los vectores de clonación, esta secuencia es generalmente una que permite que el vector se replique independientemente de los cromosomas del organismo recombinante y también incluye o bien orígenes de replicación o bien secuencias de replicación autónoma. Diversos orígenes de replicación bacterianos y víricos se conocen bien e incluyen, pero no se limitan a, replicón ColE1 derivado de pBR322, el replicón P15A, el replicón pCloDF13, el replicón pKN402, el replicón pMB1 (pUC), el replicón pSC101, y los orígenes virales de SV40, polioma, adenovirus, VSV y VPB.

- 45 Los ácidos nucleicos dados a conocer en el presente documento pueden usarse para producir proteínas mediante el uso de vectores de expresión recombinantes que comprenden un ácido nucleico aislado. Puede usarse una amplia variedad de vectores de expresión. Por ejemplo, vectores derivados de plásmidos, de cromosomas, de episomas y de virus, incluyendo vectores derivados de plásmidos bacterianos, bacteriófagos, episomas de levadura, elementos cromosómicos de levaduras, virus, tales como baculovirus, papovirus, tales como SV40, virus vaccinia, adenovirus, virus de la viruela aviar, virus de la seudorrabia y retrovirus, y vectores derivados de combinaciones de los mismos, tales como los derivados de elementos genéticos de plásmido y bacteriófago, tales como cósmidos y fagémidos.

- 55 Generalmente, cualquier vector adecuado para mantener, propagar o expresar polinucleótidos para expresar un polipéptido en un organismo recombinante puede usarse para la expresión en este sentido. Por tanto, puede usarse cualquier otro vector que sea replicable y viable en el organismo recombinante. Los ejemplos no limitativos de vectores de expresión bacterianos incluyen pKK223-3 y pTrc 99A, que portan el promotor *trp-lac*; pUC, pTZ, pSK y pGEM, que portan el promotor *lac*; el vector pET y derivados del mismo, que contienen el promotor T7; y la serie de

vectores pHUB, la serie de vectores pPLc, pKC30, pAS1, pRM1/pRM9 y pTrxFus, que contienen todos el promotor λ p_L de bacteriófago. Los vectores a modo de ejemplo adicionales incluyen la serie de vectores pATH, la serie de vectores pBAD, la serie de vectores pBEc, la serie de vectores pCAL, la serie de vectores pCRT7, pGAL, pGEX, y derivados, la serie de vectores pLEX, la serie de vectores pMAL, la serie de vectores pOSEX, la serie de vectores pQE, la serie de vectores pRSET y la serie de vectores pTriEx. Los vectores adecuados para la expresión del ácido nucleico en *S. cerevisiae* incluyen, por ejemplo, pAD-GAL4 y derivados del mismo, pBridge, pCM y derivados del mismo, la serie de vectores pEMBL4, pESC y derivados del mismo, la serie de vectores pFL, pSZ62, la serie de vectores pYC2 y pYC6, y la serie de plásmidos YIP.

La secuencia de ácido nucleico apropiada se inserta en el vector mediante cualquiera de una variedad de técnicas de rutina y bien conocidas. En general, una secuencia de ácido nucleico para la expresión se une a un vector de expresión escindiendo la secuencia de ácido nucleico y el vector de expresión con una o más endonucleasas de restricción y luego uniendo los fragmentos de restricción entre sí usando una ligasa T4-ADN. Los procedimientos para la restricción y ligación se conocen bien en la técnica. Los procedimientos adecuados en este sentido y para la construcción de vectores de expresión usando técnicas alternativas, que también se conocen bien en la técnica, se exponen en más detalle en Sambrook *et al.*, citado anteriormente. Los ejemplos no limitativos de estas técnicas alternativas incluyen, por ejemplo, la incorporación de la secuencia de ácido nucleico mediante recombinasa o topoisomerasa.

Las secuencias de ácido nucleico pueden modificarse o unirse entre sí mediante técnicas convencionales tales como PCR SOE, síntesis de ADN, ligamiento de extremos romos o ligamiento en sitios de enzimas de restricción. Si no están disponibles sitios de restricción adecuados, entonces pueden usarse adaptadores o ligadores de oligonucleótidos sintéticos (véase, por ejemplo, Sambrook *et al.*, citado anteriormente; Ausubel *et al.*, citado anteriormente).

Un experto habitual en la técnica reconocerá que numerosos promotores son funcionales en células y se han descrito en la bibliografía, incluyendo promotores constitutivos, inducibles, regulados por el desarrollo y regulados medioambientalmente. De particular interés es el uso de promotores (también denominados regiones de iniciación transcripcional) que son funcionales en el organismo recombinante apropiado. Por ejemplo, si se usa *E. coli* como organismo recombinante, entonces los promotores a modo de ejemplo que pueden usarse incluyen, pero no se limitan a, el promotor PL de fago lambda, los promotores lac, trp, trc y tac de *E. coli*, los promotores tempranos y tardíos de SV40, promotores de LTR retrovirales y el promotor 35S de CaMV. Si *Saccharomyces cerevisiae* es el huésped, entonces las secuencias de interés están normalmente bajo el control de promotores de levadura. Un ejemplo no limitativo de un promotor de levadura útil incluye el promotor GAL/CYC.

Cualquier promotor adecuado conocido para un experto habitual en la técnica que no se menciona en el presente documento puede emplearse fácilmente en la invención descrita en el presente documento. Por ejemplo, pueden usarse otros promotores conocidos para controlar la expresión de genes en células procariontas o eucariotas. Los vectores de expresión pueden contener también un sitio de unión a ribosoma para el inicio de la traducción y un sitio de terminación de la transcripción. El vector puede contener también secuencias útiles para la amplificación de la expresión génica.

En el presente documento se da a conocer un ácido nucleico aislado que comprende un promotor operativamente unido a un ácido nucleico que codifica para el polipéptido. El promotor puede ser un promotor inducible, un promotor constitutivo o un promotor específico para células. El promotor puede ser un promotor T7. El promotor puede ser también un promotor pTrc, un promotor PxyIA, un promotor Pgrac, un promotor GAL1 o un promotor GAL10.

Tal como se usa en el presente documento, el término "operativamente unido" significa que la 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 del ADN seleccionado. Además, el promotor se ubica en el sentido de 5' de la secuencia de nucleótidos seleccionada en cuanto a la dirección de transcripción y traducción. Por "operativamente unido" quiere decirse que una secuencia de nucleótidos y una(s) secuencia(s) regulatoria(s) se conecta de tal manera que permite la expresión génica cuando las moléculas apropiadas (por ejemplo, proteínas de activador transcripcional) se unen a la(s) secuencia(s) regulatoria(s).

También pueden proporcionarse regiones reguladoras de terminación de la transcripción en constructos de expresión. Las regiones de terminación de la transcripción pueden proporcionarse mediante la secuencia de vector que codifica para las secuencias de proteína Ole o puede usarse una región de terminación de la transcripción que está asociada de manera natural con la región de iniciación de la transcripción. Puede emplearse cualquier región de terminación de la transcripción conveniente que pueda terminar la transcripción en un organismo recombinante en los constructos dados a conocer en el presente documento. Los vectores de expresión y clonación pueden contener, y habitualmente contienen, un gen estructural o marcador de selección que tiene las regiones regulatorias necesarias para la expresión en un organismo recombinante para proporcionar la selección de células transformantes. El gen puede proporcionar resistencia a un agente citotóxico (por ejemplo, un antibiótico, metal pesado, o toxina), complementación que proporciona prototrofia a un huésped auxotrófico, inmunidad vírica, o similar. Dependiendo del número de especies huésped diferentes en las que se introduce el constructo de expresión o componentes del mismo, pueden emplearse uno o más marcadores en los que se usan diferentes condiciones

para la selección para los diferentes huéspedes.

Los ejemplos no limitativos, específicos de marcadores de selección adecuados incluyen genes que confieren resistencia a bleomicina, eritromicina, gentamicina, glifosato, higromicina, kanamicina, metotrexato, ácido nalidíxico, fleomicina, fosfotricina, espectinomicina, estreptomycin, sulfonamida, sulfonilureas, ampicilina/carbenicilina, cloramfenicol, estreptomycin/espectinomicina o tetraciclina. Otro ejemplo de un marcador de selección adecuado son los genes marcadores seleccionables auxotróficos, tales como genes marcadores seleccionables de histidina. Los ejemplos no limitativos, específicos de marcadores incluyen, pero no se limitan a, fosfatasa alcalina (AP), myc, hemaglutinina (HA), 13 glucuronidasa (GUS), luciferasa y proteína fluorescente verde (GFP). Preferiblemente, el ácido nucleico aislado comprende además un marcador de selección acoplado al ácido nucleico que codifica para el polipéptido. El marcador de selección puede ser resistencia a ampicilina/carbenicilina, resistencia a kanamicina, resistencia a cloramfenicol, resistencia a eritromicina, resistencia a estreptomycin/espectinomicina o un gen marcador seleccionable auxotrófico de histidina.

Además, los vectores de expresión pueden contener también secuencias de marcadores operativamente unidas a una secuencia de nucleótidos para una proteína que codifica para una proteína adicional usada como marcador. El resultado es un híbrido o proteína de fusión que comprende dos proteínas diferentes y unidas. La proteína marcadora puede proporcionar, por ejemplo, un marcador enzimático o inmunológico para la proteína recombinante producida mediante el vector de expresión. Adicionalmente, puede modificarse el extremo del polinucleótido mediante la adición de una secuencia que codifica para una secuencia de aminoácidos útil para la purificación de la proteína producida. Por ejemplo, puede incluirse una secuencia de ADN que codifica para una secuencia de aminoácidos que confiere afinidad para un método de cromatografía particular. Se han ideado diversos métodos para la adición de tales restos de purificación por afinidad a proteínas. Pueden encontrarse ejemplos representativos en las patentes estadounidenses n.ºs 4.703.004, 4.782.137, 4.845.341, 5.935.824 y 5.594.115. Puede usarse cualquier método conocido en la técnica para la adición de secuencias de nucleótidos que codifican para restos de purificación (véase, por ejemplo, Sambrook *et al.*, citado anteriormente).

En particular, se dan a conocer en el presente documento constructos recombinantes que incluyen uno o más ácidos nucleicos aislados que codifican para proteínas Ole o variantes y homólogos de las mismas. Los constructos pueden incluir un vector, tal como un vector viral o plasmídico, en el que se ha insertado la secuencia, o bien en la orientación directa o inversa. El constructo recombinante puede incluir además una secuencia regulatoria, incluyendo, por ejemplo, un promotor operativamente unido a la secuencia. Se conoce un gran número de vectores y promotores adecuados y están disponibles comercialmente. En una realización, se usan los vectores pET-21b(+), pCOLADuet-1, pCDFDuet-1, pcDNA3.1(+), pCMV SPORT6.1 (Invitrogen), o cualquiera de los vectores descritos en el presente documento. Sin embargo, puede usarse cualquier plásmido o vector adecuado siempre que sean replicables y viables en el huésped.

En el presente documento se da a conocer un vector que comprende un ácido nucleico aislado. Por ejemplo, el vector puede ser un plásmido. Preferiblemente, el vector es un plásmido seleccionado de pET21b(+), pCOLADuet-1, pCDFDuet-1, pACYCDuet-1, pACYCpTrc, pCL1920pTrc, pESC-HIS, pSUP104, pMM1522, pWH1520 y pHT01. En una realización preferida, el vector es un plásmido seleccionado de pET21b(+), pCOLADuet-1, pCDFDuet-1, pWH1520, pHT01, pESC-HIS, pET-21d(+), pETDuet-1, pACYCDuet-1, pTrcHis2A, pMAL-c2X o pCL1920pTrc.

La tecnología de ADN recombinante que da como resultado la integración de los ácidos nucleicos respectivos que codifican para OleA (por ejemplo, SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, SEQ ID NO: 10, SEQ ID NO: 12, SEQ ID NO: 14, SEQ ID NO: 16, SEQ ID NO: 18, SEQ ID NO: 20, SEQ ID NO: 22, SEQ ID NO: 88, SEQ ID NO: 135-464 y SEQ ID NO 478-487), OleB (por ejemplo, SEQ ID NO: 10 y SEQ ID NO: 230-326), OleC (por ejemplo, SEQ ID NO: 6, SEQ ID NO: 14, SEQ ID NO: 20, SEQ ID NO: 88 y SEQ ID NO: 327-402), OleD (por ejemplo, SEQ ID NO: 8, SEQ ID NO: 16, SEQ ID NO: 22, SEQ ID NO: 142-149, y SEQ ID NO: 403-464), OleBC (por ejemplo, SEQ ID NO: 135-141), y/o variantes y homólogos de estas secuencias en el cromosoma de cualquier organismo vivo puede dar como resultado la expresión y producción de las proteínas respectivas.

El ácido nucleico aislado también puede ser parte de un casete de expresión que, como mínimo, incluye un promotor, uno o más ácidos nucleicos aislados que codifican para OleA, OleB, OleC, OleD u OleBC, y una secuencia de señal de terminación transcripcional funcional en un organismo recombinante. El promotor puede ser cualquiera de los tipos tratados en el presente documento, por ejemplo, un promotor inducible o promotor constitutivo. El casete de expresión puede incluir además una secuencia de direccionamiento operativamente unida o péptido de tránsito o secreción que codifica para una región que puede dirigir el transporte de la proteína producida. El casete de expresión puede incluir además una secuencia de ácido nucleico que codifica para un marcador seleccionable y/o un resto de purificación.

Pueden introducirse o alterarse secuencias regulatorias, secuencias codificantes y combinaciones de las mismas en el genoma de la cepa del huésped. En algunos ejemplos, la integración de la secuencia recombinante deseada en la secuencia genómica del organismo recombinante no requiere el uso de un marcador seleccionable, tal como un antibiótico. En algunos ejemplos, las alteraciones genómicas incluyen cambiar la secuencia de control de los genes diana, tales como *oleA*, *oleB*, *oleC*, *oleBC* u *oleD* sustituyendo el/los promotor(es) nativo(s) con un promotor insensible a la regulación. Existen numerosos enfoques para realizar esto. Por ejemplo, Valle y Flores, *Methods Mol.*

Biol. 267: 113-122 (2006) describen un método basado en PCR para sobreexpresar genes cromosómicos en *E. coli*. Otro enfoque se basa en el uso de oligonucleótidos monocatenarios para crear mutaciones específicas directamente en el cromosoma usando la tecnología desarrollada por Costantino *et al.*, *Proc. Nat. Acad. Sci.* 100: 15748-15753 (2003). Esta tecnología se basa en el uso de la sobreexpresión de la proteína beta del bacteriófago lambda para potenciar la recombinación genética. Las ventajas de este enfoque son que pueden usarse oligonucleótidos sintéticos de 70 o más bases de largo para crear mutaciones puntuales, inserciones y deleciones. Este método elimina las etapas de clonación. Además, el sistema es tan eficaz que no son necesarios marcadores para aislar las mutaciones. Este enfoque es útil para sobreexpresar, entre otras cosas, las secuencias codificantes endógenas, tales como las que codifican para OleA, OleB, OleC, OleD, OleBC, o enzimas de rutas de biosíntesis de ácidos grasos.

En el presente documento se da a conocer un ácido nucleico aislado que comprende además al menos una secuencia de ácido nucleico adicional seleccionada del grupo que consiste en (a) una secuencia regulatoria operativamente acoplada al ácido nucleico que codifica para el polipéptido; (b) un marcador de selección operativamente acoplado al ácido nucleico que codifica para el polipéptido; (c) un resto de purificación operativamente acoplado al ácido nucleico que codifica para el polipéptido; (d) una secuencia de secreción operativamente acoplada al ácido nucleico que codifica para el polipéptido; y (e) una secuencia de direccionamiento operativamente acoplada al ácido nucleico que codifica para el polipéptido.

En el presente documento se da a conocer una célula que comprende un ácido nucleico aislado. En particular, la célula puede comprender un ácido nucleico aislado o un vector que comprende un ácido nucleico aislado. La célula puede transformarse con un ácido nucleico aislado usando cualquier método adecuado conocido en la técnica. Alternativamente, la célula puede transfectarse con un vector usando cualquier método adecuado conocido en la técnica.

Tal como se usa en el presente documento, el término “transfectar” significa la introducción de un ácido nucleico (por ejemplo, por medio de un vector de expresión) en una célula receptora mediante transferencia génica mediada por ácido nucleico. Tal como se usa en el presente documento, “transformar” se refiere a un proceso en el que el genotipo de una célula se cambia como resultado de la captación celular de ADN o ARN 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.

La célula puede ser una célula de levadura, una célula fúngica, una célula de alga, una célula animal, una célula de insecto, una célula bacteriana o una célula vegetal. La célula puede ser una célula de *Archaea*. En una realización, la célula es una célula bacteriana.

La célula puede seleccionarse de cualquier célula del género *Escherichia*, *Bacillus*, *Lactobacillus*, *Rhodococcus*, *Pseudomonas*, *Aspergillus*, *Trichoderma*, *Neurospora*, *Fusarium*, *Humicola*, *Rhizomucor*, *Kluyveromyces*, *Pichia*, *Mucor*, *Myceliophthora*, *Penicillium*, *Phanerochaete*, *Pleurotus*, *Trametes*, *Chrysosporium*, *Saccharomyces*, *Schizosaccharomyces*, *Stenotrophomonas*, *Kineococcus*, *Yarrowia* o *Streptomyces*. Específicamente, la célula puede 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* o una célula de *Bacillus amyloliquefaciens*. Adicionalmente, la célula puede ser 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 fumigatus*, 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* o una célula de *Mucor michei*.

En una realización preferida, la célula es una célula de *Streptomyces lividans*, una célula de *Streptomyces murinus*, una célula de *Actinomycetes* o una célula de *Escherichia coli*. La célula de *Escherichia coli* puede ser una célula de una cepa B, una cepa C, una cepa K o una cepa W de *Escherichia coli*. La célula puede ser también una célula de *Stenotrophomonas maltophilia*, una célula de *Kineococcus radiotolerans*, una célula de *Bacillus megaterium* o una célula de *Saccharomyces cerevisiae*.

Más específicamente, la célula puede ser cualquier cepa de *Anaeromyxobacter sp.* Fw109-5, *Arthrobacter aureus*, *Arthrobacter chlorophenolicus*, *Arthrobacter sp.* FB24, *Bdellovibrio bacteriovorus*, *Blastopirellula marina*, *Brevibacterium linens*, *Burkholderia ambifaria*, *Burkholderia cenocepacia*, *Burkholderia oklahomensis*, *Burkholderia pseudomallei*, *Burkholderia sp.* 383, *Candidatus Kuenenia stuttgartiensis*, *Chloroflexus aggregans*, *Chloroflexus aurantiacus*, *Clavibacter michiganensis* subsp. *michiganensis*, *Clavibacter michiganensis* subsp. *Sepedonicus*, *Clostridium botulinum*, *Colwellia psychrerythraea*, *Congregibacter litoralis*, *Desulfatibacillum alkenivorans*, *Desulfococcus oleovorans*, *Desulfotalea psychrophila*, *Desulfuromonas acetoxidans*, *Flavobacteriales bacterium*, *Gemmata obscuriglobus*, *Geobacter bemidjensis*, *Geobacter lovleyi*, *Geobacter sp.* FRC-32, *Geobacter uraniumreducens*, *Hahella chejuensis*, *Jannaschia sp.* CCS1, *Kineococcus radiotolerans*, *Lentisphaera araneosa*, *Maricaulis maris*, *Marinobacter algicola*, *Marinobacter aquaeolei*, *Micrococcus luteus*, *Microscilla marina*, *Moritella sp.*

PE36, *Mycobacterium avium*, *Mycobacterium avium* subsp. *paratuberculosis*, *Mycobacterium marinum*, *Mycobacterium sp.* GPI, *Mycobacterium tuberculosis*, *Mycobacterium vanbaalenii*, *Neisseria gonorrhoeae*, *Nocardia farcinica*, *Opiritaceae bacterium*, *Opiritus terrae*, *Paenibacillus sp.* JDR-2, *Pelobacter propionicus*, *Photobacterium profundum*, *Photobacterium profundum*, *Photorhabdus luminescens* subsp. *laumondii*, *Planctomyces maris*,
 5 *Plesiocystis pacifica*, *Pseudoalteromonas atlantica*, *Psychromonas ingrahamii*, *Psychromonas sp.* CNPT3, *Ralstonia pickettii*, *Rhodococcus rhodochrous*, *Rhodococcus sp.*, *Rhodopirellula baltica*, *Roseovarius nubinhibens*, *Shewanella amazonensis*, *Shewanella baltica*, *Shewanella benthica*, *Shewanella denitrificans*, *Shewanella frigidimarina*,
 10 *Shewanella halifaxensis*, *Shewanella loihica*, *Shewanella oneidensis*, *Shewanella pealeana*, *Shewanella putrefaciens*, *Shewanella sediminis*, *Shewanella sp.* ANA-3, *Shewanella sp.* MR-4, *Shewanella sp.* MR-7,
Shewanella sp. W3-18-1, *Shewanella woodyi*, *Stenotrophomonas maltophilia*, *Streptomyces ambofaciens*,
Xanthomonas axonopodis, *Xanthomonas campestris*, *Xanthomonas oryzae* y *Xylella fastidiosa*.

Opcionalmente, la célula puede ser una célula animal. Por ejemplo, la célula animal puede seleccionarse del grupo que consiste en 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 y una célula PC12.

15 En el presente documento se da a conocer una célula que comprende una alteración en un gen implicado en la 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 ácidos grasos 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
 20 para producir ácidos grasos que tienen características de cadena de carbono deseadas.

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

25 Por ejemplo, la célula puede sobreexpresar acil-Coa sintasa (EC 6.2.1.3, 2.3.1.86), tioesterasa (EC 3.1.2.-, 3.1.1.15, 3.1.2.14), acetil-CoA carboxilasa (EC 6.4.1.2, 6.3.4.14), una proteína portadora de acilo, piruvato deshidrogenasa (EC 1.2.4.1), aldehído descarboxilasa (EC4.1.99.5), beta-hidroxidecanoil tioéster deshidrasa (EC 4.2.1.60), 3-oxoacil-[proteína portadora de acilo] sintasa I (EC 2.3.1.41), [proteína portadora de acilo] S-maloniltransferasa (EC 2.3.1.39), 3-oxoacil-[proteína portadora de acilo] reductasa (EC 1.1.1.100), 3-oxoacil-[proteína portadora de acilo]
 30 sintasa III (EC 2.3.1.180), enoil-[proteína portadora de acilo] reductasa (EC 1.3.1.9), (3R)-hidroximiristol-proteína portadora de acilo deshidratasa (EC 4.2.1.-), lipasa (EC 3.1.1.3), malonil-CoA descarboxilasa (EC 4.1.1.9, 4.1.1.41), aspartato 1-descarboxilasa (EC 4.1.1.11), pantotenato cinasa (EC 2.7.1.33), piruvato deshidrogenasa (EC 1.2.4.1), nucleótido de piridina transhidrogenasa (EC 1.6.1.1), y combinaciones de las mismas.

35 Además para sobreexpresar uno o más péptidos para producir sustratos que contienen cadenas de acilo graso, la célula puede tener adicionalmente uno o más péptidos delecionados, mutados o atenuados de manera funcional. 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).

40 Por ejemplo, uno o más de los siguientes pueden delecionarse, mutarse o atenuarse: acetato cinasa (EC 2.7.2.1), alcohol deshidrogenasa (EC 1.1.1.1, 1.2.1.10), 3-oxoacil-[proteína portadora de acilo] sintasa II (EC 2.3.1.179), represor transcripcional FabR (registro NP_418398), acil-CoA deshidrogenasa (EC 1.3.99.3, 1.3.99.-), sn-glicerol 3-fosfato deshidrogenasa biosintética (EC 1.1.1.94), lactato deshidrogenasa (EC 1.1.1.28), formiato acetiltransferasa (EC 2.3.1.54), aciltransferasa (EC 2.3.1.15), piruvato oxidasa (EC 1.2.2.2) y fosfotransacetilasa (EC 2.3.1.8).

45 En algunos ejemplos, la célula puede producir productos ramificados, incluyendo sustratos que contienen cadenas de acilo graso; hidrocarburos y productos intermedios de hidrocarburos. Por consiguiente, las células pueden modificarse por ingeniería para aumentar la ramificación sobreexpresando un péptido seleccionado de uno o más componentes del complejo ceto ácido de cadena ramificada deshidrogenasa (EC 1.2.4.4), aminoácido de cadena ramificada aminotransferasa (EC 2.6.1.42), dihidrolipoamida deshidrogenasa (E3) (EC 1.8.1.4), crotonil-CoA reductasa (EC 1.6.5.5, 1.1.1.1), isobutiril-CoA mutasa, subunidad A (EC 5.4.99.2), isobutiril-CoA mutasa, subunidad B (5.4.99.2), beta-cetoacil-ACP sintasa III (EC 2.3.1.180), beta-cetoacil-ACP sintasa II (EC 2.3.1.179), proteína portadora de acilo (NP_823468), enoil-CoA reductasa (EC 1.3.1.34), enoil-CoA isomerasa (EC 4.2.1.-), y combinaciones de los mismos.

55 El nivel de saturación del sustrato que contiene una cadena de acilo graso, hidrocarburo y producto intermedio de hidrocarburo puede alterarse modificando por ingeniería genética la célula para sobreexpresar un péptido seleccionado de 3-oxoacil-[proteína portadora de acilo] sintasa I (EC 2.3.1.41), trans-2-enoil-ACP reductasa II (EC 1.3.1.9), enoil-(proteína portadora de acilo) reductasa (EC 1.3.1.9), trans-2, cis-3-decenoil-ACP isomerasa (4.2.1.17), acil-CoA deshidrogenasa (EC 1.3.99.3, 1.3.99.-) y combinaciones de las mismas.

Además para modificar por ingeniería la célula para producir hidrocarburos o productos intermedios de

- hidrocarburos, pueden regularse también las condiciones medioambientales, tales como temperatura, para cambiar los tipos de hidrocarburos producidos. Por ejemplo pueden usarse temperaturas más bajas para producir mayores números de dobles enlaces en un hidrocarburo y temperaturas más altas pueden conducir a mayores niveles de saturación. En 1962, Marr y Ingraham publicaron un artículo que demostraba que la temperatura influye en el grado de saturación en lípidos producidos por *E. coli* (Marr *et al.*, *J Bacteriol.*, 84: 1260-7 (1962)). Temperaturas bajas dan como resultado la producción de lípidos con un mayor grado de insaturación mientras que temperaturas más altas dan como resultado mayores cantidades de lípidos saturados (véase, por ejemplo, el ejemplo 14). Por tanto, reducir las temperaturas durante la producción puede alterar el producto para conducir a mayores grados de insaturación en hidrocarburos o productos intermedios de hidrocarburos.
- 5 La célula puede comprender una alteración en un gen que codifica para una acil-CoA deshidrogenasa. Esta alteración puede ser una delección, una mutación o una atenuación de un gen que codifica para acil-CoA deshidrogenasa. En otra realización preferida, la célula comprende una alteración en un gen que codifica para una tioesterasa. Esta alteración puede ser una sobreexpresión de un gen que codifica para una tioesterasa.
- 10 En algunos ejemplos se sobreexpresa un péptido que tiene actividad acil-CoA sintasa. Por ejemplo, acil-CoA sintasas tales como fadD (NP_416319), fadK (NP_416216), fadD (YP_045024), fadD (NP_438551), BH3103 (NP_243969), yhfL (NP_388908), Pfl_4354 (YP_350082), fadD1 (NP_251989), fadD2 (NP_251990), fadD (YP_533919), RPC_4074 (YP_533919), fadD1 (NP_520978), fadD35 (NP_217021), fadD22 (NP_217464), y combinaciones de las mismas pueden expresarse, deleccionarse, mutarse o atenuarse. Otro ejemplo de una acil-CoA sintasa es un homólogo fadD de *Stenotrophomonas maltophilia* R551-3 (ZP_01644857.1).
- 15 Además, puede controlarse la expresión de tioesterasa para alterar la cantidad de productos y/o la longitud de cadena de carbono de los productos que incluyen cadenas de acilo graso. Por ejemplo, tioesterasas tales como tesA sin secuencia líder (AAC73596), tesB (AAC73555), Uc fatB (Q41635, AAA34215), Ch fatB2 (Q39513, AAC49269), Ch fatB3 (AAC49269, AAC72881), Ce fatB (Q39473, AAC49151), At fatB [M141T] (CAA85388), At fatA (NP_189147, NP_193041), Ch fatA (AAC72883), Ha fatAI (AAL79361), o combinaciones de las mismas pueden expresarse, deleccionarse, mutarse o atenuarse.
- 20 Opcionalmente, la célula puede comprender una acil-CoA sintasa o una tioesterasa. Por ejemplo, la célula puede transformarse con un ácido nucleico aislado que codifica para una acil-CoA sintasa o una tioesterasa. Alternativamente, la célula puede transfectarse con un vector que comprende un ácido nucleico aislado que codifica para una acil-CoA sintasa o una tioesterasa.
- 25 Pueden modificarse organismos recombinantes por ingeniería genética usando los ácidos nucleicos aislados y las proteínas dados a conocer en el presente documento para producir hidrocarburos y cetonas alifáticas que tienen características estructurales definidas (por ejemplo, grados de ramificación, saturación o longitud de cadena de carbono). Un método de elaboración de hidrocarburos implica aumentar la expresión de, o expresar formas más activas de, una o más enzimas de condensación de acilo (enzimas que condensan más de una acil-CoA, acil-ACP, acil-AMP, acil-éster, ácido graso, o mezclas de los mismos). Un experto habitual en la técnica apreciará que los productos producidos a partir de tales reacciones de condensación varían con la cadena acilo que se condensa. Los productos que pueden producirse incluyen, por ejemplo, hidrocarburos y productos intermedios de hidrocarburos, tales como cetonas alifáticas.
- 30 Un experto habitual en la técnica apreciará que los sustratos que contienen una cadena de acilo graso y productos intermedios de la misma pueden producirse usando reacciones *in vitro*, incluyendo conversiones químicas o enzimáticas, así como mediante reacciones *in vivo*. Adicionalmente, puede utilizarse una combinación de conversiones *in vivo* e *in vitro*. Además, pueden producirse cetonas alifáticas específicas proporcionando de manera selectiva sustratos seleccionados, tales como ácidos grasos, acil-ACP o acil-CoA para la conversión. Alternativamente, pueden producirse hidrocarburos proporcionando de manera selectiva sustratos seleccionados, tales como ácidos grasos, acil-ACP, acil-CoA, cetonas alifáticas, α -alquil- β -cetoácidos o α -alquil- β -cetoésteres para la conversión.
- 35 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 4 y 22 átomos de carbono. Los ácidos grasos pueden ser saturados, monoinsaturados o poliinsaturados. En una realización preferida, el ácido graso se elabora a partir de una ruta de biosíntesis de ácidos grasos.
- 40 Los términos “convertir” o “conversión” se refieren al uso de o bien medios químicos o bien medios biológicos (por ejemplo, polipéptidos en una reacción) que cambian un primer producto intermedio o sustrato a un segundo producto intermedio o producto. El término “conversión química” se refiere a reacciones que no están facilitadas activamente por polipéptidos. El término “conversión biológica” se refiere a reacciones que están facilitadas activamente por polipéptidos. Las conversiones pueden tener lugar *in vivo*, *in vitro*, o ambas. Cuando se usan conversiones biológicas, los péptidos y/o las células pueden inmovilizarse sobre soportes, tal como mediante unión química sobre soportes poliméricos. Las conversiones pueden realizarse usando cualquier reactor conocido por un experto habitual en la técnica, por ejemplo en un reactor discontinuo o continuo.
- 45
- 50
- 55

El organismo recombinante puede convertir varios productos intermedios en productos intermedios posteriores o el organismo recombinante puede alimentarse, o ponerse en contacto con, un producto intermedio que se convierte en un producto. En determinados ejemplos, el organismo recombinante se pone en contacto con un producto intermedio, tal como una molécula de acil-CoA y esa molécula de acil-CoA se convierte luego en un producto.

- 5 Dada la divulgación proporcionada en el presente documento, ahora es posible la producción enzimática a gran escala de las proteínas Ole (por ejemplo, OleA, OleB, OleC, OleD y OleBC) y homólogos de las mismas. Brevemente, las secuencias codificantes de uno cualquiera de estos péptidos u homólogos de estos péptidos (véanse, por ejemplo, las tablas 1 y 2) puede clonarse en un plásmido de alta expresión, tal como pET-21b(+), pCOLADuet-1 (EMD Chemicals, Inc., Alemania), pWH1520 (Mo Bi Tec, Alemania) o pHT01 (Mo Bi Tec, Alemania).
10 El plásmido puede introducirse en una célula huésped para la producción de las enzimas. Los péptidos resultantes pueden purificarse entonces y usarse en producción discontinua.

- 15 Cuando se usan métodos *in vitro*, los péptidos suministrados a la reacción dependerán del material de partida. Por ejemplo, cuando se desea un hidrocarburo, se añadirían los sustratos acil-ACP y/o acil-CoA a una mezcla de reacción *in vitro* que contiene OleA, OleC y OleD. De manera similar, cuando el material de partida es una cetona alifática, α -alquil- β -cetoácidos, o α -alquil- β -ceto ésteres, pueden usarse los péptidos OleC y OleD en la reacción *in vitro*.

- 20 Cuando se usa un primer péptido para convertir un primer producto intermedio en un segundo producto intermedio y luego se usa un segundo péptido para convertir el segundo producto intermedio en un tercer producto intermedio, pueden añadirse los péptidos a la reacción simultáneamente o en serie. En algunos ejemplos, en los que se añaden péptidos en serie, puede retirarse el primer péptido antes de la adición del segundo péptido.

Adicionalmente, puede usarse una combinación de conversiones químicas y conversiones biológicas para producir un producto deseado. Por ejemplo, un experto habitual en la técnica apreciará que dos ácidos grasos pueden condensarse para elaborar una cetona alifática por medio de conversión química, y la cetona alifática resultante puede convertirse entonces en un hidrocarburo usando conversiones biológicas.

- 25 Dada la divulgación proporcionada en el presente documento, pueden producirse cetonas alifáticas, hidrocarburos, y productos intermedios de los mismos en una célula recombinante. La célula recombinante puede producir uno o más péptidos codificados por OleA, OleB, OleC, OleD, OleBC y secuencias relacionadas de los mismos. Un experto habitual en la técnica apreciará que la elección de péptidos para expresar en la célula recombinante dependerá del producto deseado y el material de partida proporcionado a las células. Por ejemplo, si a la célula se le suministran cetonas alifáticas y el producto deseado es un hidrocarburo, entonces la célula recombinante puede modificarse por
30 ingeniería genética con un ácido nucleico que codifica para OleC y OleD.

- 35 Los métodos *in vivo* descritos en el presente documento pueden usarse también en combinación con conversiones químicas y conversiones biológicas *in vitro*. Por ejemplo, un primer producto intermedio puede convertirse en un segundo producto intermedio usando un péptido *in vitro*; el segundo producto intermedio puede alimentarse entonces a una célula que expresa péptidos necesarios para la conversión del segundo producto intermedio en un tercer producto intermedio. En otro ejemplo, un primer producto intermedio puede convertirse en un segundo producto intermedio por medio de conversión química, y entonces el segundo producto intermedio puede alimentarse a una célula recombinante que codifica para los péptidos necesarios para conversiones posteriores.

- 40 Adicionalmente, pueden producirse productos usando dos o más etapas de reacción *in vivo*. Por ejemplo, una primera célula recombinante puede usarse para convertir un primer producto intermedio en un segundo producto intermedio. El segundo producto intermedio puede liberarse de la célula, por ejemplo mediante transporte pasivo, transporte activo o lisis celular, y el segundo producto intermedio puede alimentarse entonces a una segunda célula recombinante donde se convierte en un tercer producto intermedio. En algunos ejemplos, el tercer producto intermedio será el producto deseado.

- 45 La invención permite la producción a gran escala de cetonas alifáticas, hidrocarburos y productos intermedios de hidrocarburos que tienen longitudes de la cadena de carbono, niveles de saturación y puntos de ramificación definidos. La producción de tales moléculas modificadas por ingeniería genética proporciona una diversidad de productos que pueden usarse como combustibles y productos químicos de especialidad.

- 50 En el presente documento se da a conocer una célula transformada con cualquiera de los ácidos nucleicos dados a conocer en el presente documento (por ejemplo, secuencias de ácido nucleico que codifican para cualquiera de las secuencias de polipéptido dadas a conocer en el presente documento) o transfectada con cualquiera de los vectores dados a conocer en el presente documento. La célula puede producir un hidrocarburo o una cetona alifática. El hidrocarburo puede ser una olefina. El hidrocarburo o cetona alifática puede secretarse por la célula. La célula puede comprender un ácido nucleico aislado que codifica para OleA. Específicamente, la célula que comprende un ácido
55 nucleico aislado que codifica para OleA puede ser una célula de *Saccharomyces cerevisiae*.

En el presente documento se da a conocer un método para producir un hidrocarburo que comprende cultivar cualquier célula que comprenda cualquiera de los ácidos nucleicos aislados dados a conocer en el presente documento (por ejemplo, secuencias de ácido nucleico que codifican para cualquiera de las secuencias de

polipéptido dadas a conocer en el presente documento) con un sustrato en condiciones suficientes para producir un hidrocarburo. Por ejemplo, el sustrato puede ser una fuente de carbono, ácido graso, acil-CoA, acil-AMP, acil-ACP, α -alquil- β -cetoácido, α -alquil- β -cetoéster o una cetona alifática. Preferiblemente, la productividad de la célula es de al menos 3 mg//DO₆₀₀. Por ejemplo, la productividad de la célula puede ser de al menos 5 mg//DO₆₀₀, al menos 8 mg//DO₆₀₀, al menos 15 mg//DO₆₀₀, al menos 20 mg//DO₆₀₀ o al menos 30 mg//DO₆₀₀.

En el presente documento se da a conocer un método para producir un hidrocarburo que comprende cultivar cualquier célula que comprenda un ácido nucleico aislado dado a conocer en el presente documento (por ejemplo, una secuencia de ácido nucleico que codifica para cualquiera de las secuencias de polipéptido dadas a conocer en el presente documento) con un sustrato en condiciones suficientes para producir un hidrocarburo que comprende aislar un hidrocarburo. El hidrocarburo puede aislarse de la célula o del medio en el que se cultiva la célula. El sustrato puede ser, por ejemplo, una fuente de carbono, ácido graso, acil-CoA, acil-AMP, acil-ACP, α -alquil- β -cetoácido, α -alquil- β -cetoéster o una cetona alifática.

En el presente documento se da a conocer un método para producir un hidrocarburo que comprende cultivar cualquier célula que comprenda los ácidos nucleicos aislados dados a conocer en el presente documento (por ejemplo, una secuencia de ácido nucleico que codifica para cualquiera de las secuencias de polipéptido dadas a conocer en el presente documento) con un sustrato (por ejemplo, una fuente de carbono, ácido graso, acil-CoA, acil-AMP, acil-ACP, α -alquil- β -cetoácido, α -alquil- β -cetoéster o una cetona alifática) en condiciones suficientes para producir un hidrocarburo que comprende además someter a craqueo o refinado el hidrocarburo.

El método puede producir un hidrocarburo que es monoinsaturado o poliinsaturado (por ejemplo, diinsaturado, triinsaturado, etc.). El hidrocarburo puede tener una longitud de cadena de carbono de entre 10 y 40 carbonos. Por ejemplo, el hidrocarburo puede tener una longitud de cadena de carbono de entre 15 y 35, 17 y 34, 18 y 33, 19 y 33 carbonos, entre 27 y 33 carbonos, entre 29 y 31 carbonos, o 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32 ó 33 carbonos.

En el presente documento se da a conocer un método para producir una cetona alifática que comprende cultivar cualquier célula que comprenda los ácidos nucleicos aislados dados a conocer en el presente documento (por ejemplo, una secuencia de ácido nucleico que codifica para cualquiera de las secuencias de polipéptido dadas a conocer en el presente documento) con un sustrato en condiciones suficientes para producir una cetona alifática. Por ejemplo, el sustrato puede ser una fuente de carbono, ácido graso, acil-CoA, acil-AMP, acil-ACP, α -alquil- β -cetoácido o α -alquil- β -cetoéster. Preferiblemente, la productividad de la célula es de al menos 0,1 mg//DO₆₀₀. Por ejemplo, la productividad de la célula puede ser de al menos 0,1 mg//DO₆₀₀, al menos 1 mg//DO₆₀₀, al menos 3 mg//DO₆₀₀, al menos 6 mg//DO₆₀₀, al menos 9 mg//DO₆₀₀ o al menos 12 mg//DO₆₀₀.

En el presente documento se da a conocer un método para producir una cetona alifática que comprende cultivar cualquier célula que comprenda los ácidos nucleicos aislados dados a conocer en el presente documento con un sustrato en condiciones suficientes para producir una cetona alifática que comprende aislar una cetona alifática. La cetona alifática puede aislarse de la célula o del medio en el que se cultiva la célula

El método puede producir una cetona alifática que es saturada, monoinsaturada, o poliinsaturada (por ejemplo, diinsaturada, triinsaturada, etc.). La cetona alifática puede tener una longitud de cadena de carbono de 10 a 40 carbonos. Por ejemplo, la cetona alifática puede tener una longitud de cadena de carbono de entre 15 y 35, 17 y 34, 18 y 33, 19 y 33 carbonos, entre 23 y 29 carbonos, entre 25 y 27 carbonos, o 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32 ó 33 carbonos.

En el presente documento se da a conocer un método para producir un polipéptido purificado que comprende cultivar cualquier célula que comprenda un ácido nucleico aislado dado a conocer en el presente documento en condiciones suficientes para producir el polipéptido codificado por el ácido nucleico aislado. Específicamente, el polipéptido es OleA, OleB, OleC, OleD u OleBC.

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

Tal como se usa en el presente documento, los términos "purificar", "purificado" y "purificación" no requieren pureza absoluta. Son términos relativos. Por tanto, por ejemplo, cuando se producen polipéptidos en células, un polipéptido purificado es uno que está sustancialmente separado de otros componentes celulares (por ejemplo, ácidos nucleicos, lípidos, hidratos de carbono o hidrocarburos). En otro ejemplo, la preparación de olefina purificada es una en la que la olefina está sustancialmente libre de contaminantes, tales como los que pueden estar presentes tras la fermentación. En algunas realizaciones, una olefina está purificada cuando al menos el 50% en peso de una

muestra está compuesta por la olefina. En otras realizaciones, una olefina está purificada cuando al menos 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 compuesta por la olefina.

5 En el presente documento se da a conocer un método para producir una cetona alifática que comprende incubar un sustrato con OleA, en condiciones suficientes para producir una cetona alifática. Específicamente, el sustrato puede comprender acil-CoA, acil-AMP o acil-ACP.

10 En el presente documento se da a conocer un método para producir un hidrocarburo que comprende incubar un sustrato con OleA, OleB, OleC, OleD, OleBC, o una combinación de las mismas en condiciones suficientes para producir un hidrocarburo. El sustrato puede comprender acil-CoA, acil-AMP o acil-ACP. Las proteínas OleA, OleB, OleC, OleD u OleBC usadas para producir un hidrocarburo pueden ser proteínas purificadas o sin purificar. Por ejemplo, puede añadirse un sustrato de acil-CoA a un lisado celular de un organismo que expresa los genes *ole* para producir hidrocarburos.

15 El método para producir un hidrocarburo puede comprender incubar un sustrato con OleA y OleD en condiciones suficientes para producir un hidrocarburo. Opcionalmente, el método puede comprender OleB. El método puede comprender incubar un sustrato con OleA, OleC y OleD en condiciones suficientes para producir un hidrocarburo. Opcionalmente, el método puede comprender OleB.

20 Los péptidos de condensación de acilo incluyen péptidos que pueden catalizar la condensación de acil-ACP, acil-CoA, acil-AMP, ácidos grasos, y mezclas de los mismos usando los métodos descritos en el presente documento. En algunas realizaciones, estos péptidos de condensación de acilo tienen alta, media o baja especificidad de sustrato. En algunos ejemplos, los péptidos de condensación de acilo son más específicos de sustrato y solo aceptarán sustratos de una longitud de cadena específica. Adicionalmente, un experto habitual en la técnica apreciará que algunos péptidos de condensación de acilo catalizarán también otras reacciones. Por ejemplo, algunos péptidos de condensación de acilo aceptarán otros sustratos además de acil-CoA, acil-ACP, acil-AMP, acil-éster, ácido graso, o mezclas de los mismos. Por tanto, tales péptidos no específicos de condensación de acilo se incluyen también. Los ejemplos de enzimas de condensación de acilo, además de las secuencias de OleA proporcionadas en la tabla 1, están disponibles públicamente.

25 OleC y OleD pueden usarse para producir olefinas sin OleA. Además, OleC puede usarse para producir olefinas sin OleA y OleD.

30 Los ácidos micólicos son ácidos 2-alquil-3-hidroxi grasos producidos por bacterias (por ejemplo, *Mycobacterium* y *Corynebacterium*). Los ácidos micólicos se incorporan a menudo en paredes celulares bacterianas. Estos ácidos 2-alquil-3-hidroxi grasos se derivan de una condensación Claisen seguida por una reducción del grupo ceto. Esta reacción es similar a las reacciones enzimáticas realizadas por OleA y OleD en el procedimiento de síntesis de hidrocarburos descrito en el presente documento. Por tanto, las primeras etapas en la ruta del ácido micólico pueden usarse para producir los sustratos necesarios para OleC o la combinación de OleC y OleD para producir hidrocarburos. Además, modificaciones genéticas adicionales en la ruta del ácido micólico podrían aumentar los niveles de producción de alquenos.

35 La figura 20 esboza una ruta sintética propuesta para la biosíntesis de ácido micólico (véanse, por ejemplo, Lea-Smith *et al.*, *Journal of Biological Chemistry*, 282: 11000-11008 (2007) y Portevin *et al.*, *Proceedings of the National Academy of Science*, 101: 314-319 (2004)) y para la síntesis de olefinas mediante los genes *ole*. El examen de estas dos rutas revela productos intermedios similares. El producto de la condensación de ácidos grasos en ambas rutas es una 2-alquil-3-cetoacil graso-CoA o 2-alquil-3-cetoacil graso-ACP. Como resultado, la sobreexpresión de *oleC* y *oleD* en un organismo que produce ácido micólico, tal como *C. glutamicum*, puede "secuestrar" la ruta de biosíntesis de ácido micólico conduciendo a la producción de olefinas.

40 En el primer esquema, *oleC* y *oleD* están sobreexpresados en un organismo que produce ácido micólico. Un organismo huésped apropiado podría ser *C. glutamicum*, un huésped industrial bien establecido con herramientas genéticas bien elaboradas que es tolerante a la pérdida de producción de ácido micólico. La especificidad de ácido graso de los genes de ácido micólico de *C. glutamicum* es similar a los perfiles de especificidad observados para *oleA*, *oleB*, *oleC* y *oleD*. La sobreexpresión de *oleC* y *oleD* dirigirá al producto de *pkS13* hacia la síntesis de olefinas tal como se esboza en la figura 21.

45 En el segundo esquema, solo se sobreexpresa *oleC* en un organismo que produce ácido micólico. Este esquema propuesto asume que OleC puede funcionar independientemente sobre un producto intermedio de 2-alquil-3-hidroxiacilo graso para formar olefinas. Esta ruta se destaca en la figura 22.

50 La potenciación de la producción de olefinas en una cepa que produce ácido micólico que expresa *oleC* y *oleD* puede obtenerse completando las siguientes modificaciones genéticas. Para evitar la eliminación de sustratos para OleC y OleD, *cmrA*, o su homólogo funcional, puede desactivarse para evitar la formación de ácido micólico mientras se permite la acumulación de 2-alquil-3-cetoacil graso-ACP, un sustrato para OleD. Adicionalmente, la sobreexpresión de *fadD32*, *accA3*, *accD4*, *accD5* o *pkS13* debe aumentar la producción de productos intermedios de ácido micólico requeridos, dando como resultado una mayor producción de olefinas con la sobreexpresión de *oleC* y

oleD.

Generalmente, existen varios métodos de identificación de péptidos que tienen actividad de condensación de acilo. La formación de producto usando uno o más de estos métodos indica que el péptido tiene actividad de condensación de acilo. Además de los ensayos *in vitro* proporcionados en el ejemplo 3, el péptido puede expresarse a partir de un ácido nucleico exógeno en una célula y luego puede prepararse un lisado celular. Diversos sustratos tales como acil-CoA, acil-ACP, acil-AMP, acil-éster, ácido graso, o mezclas de los mismos pueden añadirse al lisado y los productos pueden detectarse usando los métodos de CG/EM o CG/FID descritos en el presente documento. En otro ejemplo, el péptido puede purificarse e incubarse con lisado celular a partir de una célula que no expresa el péptido (a continuación en el presente documento, lisado silvestre). El péptido purificado, lisado silvestre y diversos sustratos pueden incubarse. Los productos resultantes pueden caracterizarse usando los métodos de CG/EM o CG/FID descritos en el presente documento. En aún otro ejemplo, la actividad de condensación de acilo puede caracterizarse incubando enzima purificada y sustrato en presencia de lisado celular que se ha calentado hasta desnaturalizar las proteínas. En otro ejemplo, el péptido purificado y diversos sustratos pueden incubarse, y el producto resultante puede caracterizarse usando los métodos de CG/EM descritos en el presente documento. Los péptidos que tienen actividad de condensación de acilo se identifican como aquellos que producen cetonas alifáticas. Un experto habitual en la técnica apreciará que cuando se usa un lisado celular que ya contiene cetonas alifáticas, los péptidos que tienen actividad de condensación de acilo se reconocerán por un aumento en cetonas alifáticas en comparación con el lisado sin la adición de sustrato (tal como un aumento de al menos el 10%, al menos el 20%, al menos el 50% o al menos el 90%).

En algunos casos, la condensación puede dar como resultado la producción de moléculas derivadas de uno o más de los sustratos. Por ejemplo, la condensación de dos moléculas de acil-CoA puede producir al menos una molécula de CoA. Puesto que CoA tiene un resto tiol libre (RSH), que es altamente reactivo, esta molécula puede detectarse mediante una variedad de métodos. Un método de este tipo es la reacción con ácido ditionitrobenzoico (reactivo de Ellman) que puede seguirse espectrofotométricamente a 411 nm. Alternativamente, puede hacerse reaccionar CoA con monobromobimano y detectarse mediante HPLC (Fahey, *et al.*, *Methods Enzymol.* 143: 85-96, (1987)).

Pueden usarse métodos bioinformáticos para encontrar péptidos de condensación de acilo. Las condensaciones de acilo se producen mediante una reacción química bien conocida, conocida como "condensación de Claisen". La condensación de Claisen es una reacción que forma enlaces carbono-carbono que se produce entre dos ésteres o un éster y otro compuesto de carbonilo en presencia de una base fuerte que dando como resultado un β -cetoéster o una β -dicetona.

Los péptidos de condensación de acilo contienen normalmente una tríada catalítica compuesta por Cys-His-Asn. Las enzimas de condensación comparten un pliegue tridimensional común, aunque comparten poca similitud al nivel de aminoácidos. Sus sitios activos, sin embargo, presentan similitudes significativas. (Heath *et al.*, *Nat. Prod. Rep.*, 19: 581-596, (2002)).

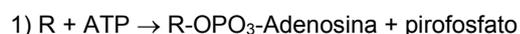
Los péptidos de condensación de acilo a modo de ejemplo incluyen las secuencias de OleA dadas a conocer en el presente documento, por ejemplo, SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, SEQ ID NO: 10, SEQ ID NO: 12, SEQ ID NO: 14, SEQ ID NO: 16, SEQ ID NO: 18, SEQ ID NO: 20, SEQ ID NO: 22, SEQ ID NO: 88, SEQ ID NO: 135-464 y SEQ ID NO 478-487, homólogos de estas secuencias, enzimas que tienen uno o más de los motivos estructurales de OleA proporcionados en el presente documento, y fragmentos activos/variantes de las mismas que presentan actividad de condensación de acilo.

Pueden modificarse organismos recombinantes por ingeniería genética usando los péptidos dados a conocer en el presente documento para producir hidrocarburos y productos intermedios de hidrocarburos que tienen características estructurales definidas (grados de ramificación, saturación y longitud de cadena de carbono). Un método de elaboración de productos intermedios de hidrocarburos implica expresar, aumentar la expresión de, o expresar formas más activas de, una o más enzimas, por ejemplo, actividad hidrocarburo sintasa, péptidos de adenilación, deshidrogenasas, deshidratasas o enzimas de condensación de acilo. Las enzimas a modo de ejemplo que pueden manipularse para aumentar la producción de hidrocarburos incluyen OleA, OleB, OleC, OleBC y OleD, así como otras enzimas que aumentan o modifican la producción de ácidos grasos. Un experto habitual en la técnica apreciará que los productos producidos a partir de tales enzimas varían con la cadena de acilo del sustrato.

Los péptidos de adenilación incluyen péptidos que pueden catalizar la adición de adenosina monofosfato a productos intermedios de hidrocarburos, tales como un β -cetoácido, incluyendo β -cetoácidos α -sustituídos, particularmente los que incluyen un hidrocarburo alifático en la posición α . Tal como se describió anteriormente, el grupo α -alifático en tales productos intermedios es normalmente una cadena de hidrocarburo opcionalmente ramificada que incluye opcionalmente uno o más sitios de insaturación, por ejemplo, uno, dos, o tres sitios de insaturación en la cadena de hidrocarburo. Tales péptidos de adenilación pueden catalizar también la adición de adenosina monofosfato a β -hidroxiocetoácidos para formar un β -cetoéster. Métodos de identificación de tal actividad se proporcionan en el presente documento.

En algunos ejemplos, los péptidos de adenilación son más específicos para sustrato y solo aceptarán, por ejemplo, β -cetoésteres activados por CoA o ACP. Adicionalmente, un experto habitual en la técnica apreciará que algunos

péptidos de adenilación también reaccionan. Por ejemplo, algunos péptidos de adenilación aceptarán otros sustratos además de β -cetoácidos α -sustituídos. Por tanto, tales péptidos de adenilación relativamente no específicos también se incluyen. Los ejemplos de péptidos de adenilación están disponibles públicamente (véanse, por ejemplo, las tablas 1 y 2). A menudo el péptido de adenilación cataliza reacciones adicionales, tales como la transesterificación del compuesto adenilado con otros grupos activadores, tales como CoA. Esta actividad se considera actividad sintasa. Un ejemplo sería el siguiente conjunto de reacciones:



Existen varios métodos de identificación de péptidos que tienen actividad de adenilación. La formación de producto usando uno o más de estos métodos indica que el péptido tiene actividad de adenilación. Además de los ensayos *in vivo* proporcionados en el presente documento, el péptido puede expresarse a partir de un ácido nucleico exógeno en una célula y luego puede prepararse un lisado celular. Pueden añadirse diversos sustratos tales como ATP al lisado, y pueden detectarse los productos usando los métodos descritos en el presente documento. En otro ejemplo, el péptido puede purificarse e incubarse con lisado celular de una célula que no expresa el péptido. El péptido purificado, lisado silvestre y diversos sustratos pueden incubarse. Los productos resultantes pueden caracterizarse usando los métodos descritos en el presente documento. Un experto habitual en la técnica apreciará que cuando se usa un lisado celular que ya contiene productos adenilados, los péptidos que tienen actividad de adenilación se reconocerán por un aumento en o bien PPI libre, β -cetoésteres α -sustituídos de AMP o β -hidroxiésteres α -sustituídos de AMP en comparación con el lisado sin la adición de sustrato. Los péptidos de adenilación a modo de ejemplo incluyen OleC (por ejemplo, SEQ ID NO: 6), las proteínas OleC enumeradas en la tabla 1 y fragmentos activos/variantes de las mismas que presentan actividad de adenilación.

Los péptidos de deshidrogenasa incluyen péptidos que pueden catalizar la reducción de un grupo ceto en una cetona alifática, un β -cetoácido alifático o una molécula de β -cetoéster alifático al grupo hidroxilo correspondiente (la adición de H_2 a través del doble enlace carbono-oxígeno). En el presente documento se proporcionan métodos de identificación de tal actividad. En algunos ejemplos, los péptidos de deshidrogenasa son más específicos de sustrato y solo aceptarán, por ejemplo, ésteres de β -cetoésteres α -alifáticos de CoA o ACP. Adicionalmente, un experto habitual en la técnica apreciará que algunos péptidos de deshidrogenasa catalizarán otras reacciones también. Por ejemplo, algunos péptidos de deshidrogenasa aceptarán otros sustratos además de β -cetoésteres. Por tanto, también se incluyen tales péptidos de deshidrogenasa no específicos. Ejemplos de péptidos de deshidrogenasa son OleD (por ejemplo, SEQ ID NO: 8) y los péptidos deshidrogenasas disponibles públicamente proporcionados en la tabla 1.

Existen varios métodos de identificación de péptidos que tienen actividad deshidrogenasa. La formación de producto usando uno o más de estos métodos indica que el péptido tiene actividad deshidrogenasa. Además de los ensayos *in vivo* proporcionados en el presente documento, el péptido puede expresarse a partir de una secuencia exógena de ácido nucleico en una célula y luego puede prepararse un ensayo *in vitro* que contiene lisado celular o péptido purificado. Pueden añadirse diversos sustratos, tales como NADPH y/o NADH, al ensayo y pueden detectarse los productos usando los métodos de CG/EM descritos en el presente documento.

En otro ejemplo, el péptido puede purificarse e incubarse con lisado celular de una célula que no expresa el péptido. El péptido purificado, lisado silvestre y diversos sustratos (por ejemplo, ácido graso, acil-CoA, acil-AMP, acil-ACP, α -alquil- β -cetoácido, α -alquil- β -cetoéster, o una cetona alifática) pueden incubarse. Los productos resultantes pueden caracterizarse usando los métodos descritos en el presente documento (véase, por ejemplo, el ejemplo 1).

En aún otro ejemplo, puede detectarse la actividad deshidrogenasa monitorizando espectrofotométricamente la oxidación dependiente de deshidrogenasa del NADPH o NADH en presencia del sustrato cetona. La actividad deshidrogenasa se detecta como una disminución en la absorbancia de la disolución de reacción a 340 nm.

En aún otro ejemplo, la actividad deshidrogenasa puede caracterizarse incubando enzima purificada y sustrato (por ejemplo, NAD(P)H, β -cetoésteres α -alifáticos y/o β -cetoácidos α -alifáticos) en presencia de lisado celular que se ha calentado para desnaturalizar las proteínas. Los péptidos que tienen actividad deshidrogenasa se identifican como los que producen moléculas de β -hidroxiácido o éster (particularmente, éster activado) a partir de una o más de las reacciones descritas anteriormente. Un experto habitual en la técnica apreciará que cuando se usa un lisado celular que ya contiene productos de β -hidroxiácido y/o éster, los péptidos que tienen actividad deshidrogenasa se reconocerán por un aumento en o bien NADP, β -hidroxiácido y/o bien moléculas de éster en comparación con el lisado sin la adición de sustrato.

Los péptidos de deshidrogenasa a modo de ejemplo incluyen OleD (por ejemplo, SEQ ID NO: 8) y las enzimas relacionadas mostradas en la tabla 1. Pueden identificarse enzimas OleD adicionales buscando en diversas bases de datos usando los motivos de OleD proporcionados en el presente documento y la metodología descrita en el presente documento.

Además, en el presente documento se dan a conocer ácidos nucleicos que codifican para un péptido que tiene

actividad hidrolasa. En particular, el péptido tendría actividad hidrolítica de β -ceto o β -hidroxiéster. Un péptido de este tipo catalizaría probablemente la hidrólisis de ésteres de todos los sustratos descritos anteriormente para producir el ácido carboxílico correspondiente. Pueden detectarse éster hidrolasas monitorizando la producción de producto, tal como el β -cetoácido o β -hidroxiácido mediante HPLC (u otras técnicas bien conocidas). En otra realización, la disminución resultante en el pH debida a la formación de los ácidos libres puede monitorizarse. Alternativamente, puede monitorizarse la hidrólisis de éster midiendo la acumulación del resto liberado del éster graso, tal como CoASH, AMP o fosfato. Los expertos en la técnica conocen bien los métodos para monitorizar estos compuestos y algunos de estos métodos se describieron anteriormente. Puede monitorizarse fosfato, por ejemplo, mediante la reacción con molibdato y verde malaquita. Pueden obtenerse ensayos adicionales comercialmente (por ejemplo, de BioVision, Inc., Mountain View, CA).

En el presente documento se da a conocer un polipéptido aislado que comprende, consiste esencialmente en o consiste en, una secuencia de aminoácidos que codifica para OleA que comprende una o más (por ejemplo, 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 35, 40, 45 ó 50) sustituciones, adiciones, inserciones o deleciones de aminoácido. Por ejemplo, el polipéptido aislado puede comprender una secuencia de aminoácidos seleccionada del grupo que consiste en SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 12 y SEQ ID NO: 18, en el que la secuencia de aminoácidos comprende una o más (por ejemplo, 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 35, 40, 45 ó 50) sustituciones, adiciones, inserciones o deleciones de aminoácido.

En el presente documento se dan a conocer polipéptidos que comprenden actividad biológica. La actividad biológica del polipéptido puede ser actividad de condensación de Claisen. Específicamente, la actividad biológica del polipéptido puede ser la condensación de dos aciltioésteres.

En el presente documento se da a conocer un polipéptido aislado que comprende, consiste esencialmente en o consiste en una secuencia de aminoácidos que codifica para OleD que comprende una o más (por ejemplo, 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 35, 40, 45 ó 50) sustituciones, adiciones, inserciones o deleciones de aminoácido. Por ejemplo, el polipéptido aislado comprende una secuencia de aminoácidos seleccionada del grupo que consiste en SEQ ID NO: 8, SEQ ID NO: 16 y SEQ ID NO: 22, en el que la secuencia de aminoácidos comprende una o más (por ejemplo, 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 35, 40, 45 ó 50) sustituciones, adiciones, inserciones o deleciones de aminoácido.

En el presente documento se da a conocer un ácido nucleico aislado que codifica para un polipéptido que tiene la misma actividad biológica que una proteína OleA. Por ejemplo, en el presente documento se da a conocer un ácido nucleico aislado que codifica para un polipéptido que tiene la misma actividad biológica que un polipéptido que comprende, consiste esencialmente en o consiste en, la secuencia de aminoácidos seleccionada del grupo que consiste en SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 8, SEQ ID NO: 12, SEQ ID NO: 16, SEQ ID NO: 18 y SEQ ID NO: 22. El ácido nucleico aislado puede comprender (i) una secuencia de ácido nucleico seleccionada del grupo que consiste en SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 7, SEQ ID NO: 11, SEQ ID NO: 15, SEQ ID NO: 17 y SEQ ID NO: 21 o un fragmento de la misma, o (ii) una secuencia de ácido nucleico que se hibrida con un complemento de una secuencia de ácido nucleico seleccionada del grupo que consiste en SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 7, SEQ ID NO: 11, SEQ ID NO: 15, SEQ ID NO: 17 y SEQ ID NO: 21 o a un fragmento de la misma.

Puede encontrarse una guía para realizar reacciones de hibridación en Current Protocols in Molecular Biology, John Wiley & Sons, N.Y. (1989), 6.3.1 – 6.3.6. Se describen métodos acuosos y no acuosos en esa referencia y puede usarse cualquier método. Las condiciones de hibridación específicas a las que se hace referencia en el presente documento son tal como sigue: 1) condiciones de hibridación de baja rigurosidad en cloruro de sodio/citrato de sodio 6X (SSC) a 45°C, seguido por dos lavados en SSC 0,2X, 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 SSC 6X a 45°C, seguido por uno o más lavados en SSC 0,2X, SDS al 0,1% a 60°C; 3) condiciones de hibridación de alta rigurosidad en SSC 6X a 45°C, seguido por uno o más lavados en SSC 0,2X, 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 SSC 0,2X, SDS al 1% a 65°C. Las condiciones de muy alta rigurosidad (4) son las condiciones preferidas a no ser que se especifique otra cosa. Tal como se usa en el presente documento, el término “se hibrida” se refiere generalmente a condiciones de baja rigurosidad, de rigurosidad media, de alta rigurosidad o de muy alta rigurosidad.

En el presente documento se dan a conocer polipéptidos que comprenden actividad biológica. La actividad biológica puede ser actividad oxidorreductasa. Específicamente, la actividad biológica puede ser reducción de un grupo ceto (por ejemplo, una cetona, cetona alifática, α -alquil- β -cetoácido, α -alquil- β -cetoéster).

La actividad hidrocarburo sintasa es la actividad de uno o más péptidos que producen la conversión de un sustrato que contiene una cadena de acilo graso, tal como acil-CoA, acil-ACP, o ácido graso, en un hidrocarburo o producto intermedio de hidrocarburo. Los ejemplos de péptidos que tienen actividad hidrocarburo sintasa incluyen Ole A, Ole B, OleC, OleD y OleBC.

La actividad hidrocarburo sintasa puede someterse a prueba, por ejemplo, usando un ensayo de complementación (véase el ejemplo 6, más adelante). Los organismos que se sabe que elaboran hidrocarburos tras la expresión de

OleA, OleB, OleC y OleD pueden usarse como huésped de prueba. Por ejemplo, si está sometiéndose a prueba una OleC candidata para determinar la actividad hidrocarburo sintasa, entonces se modifica por ingeniería genética el huésped de prueba para expresar solo *oleA* y *oleD*, pero no *oleC*. El *oleC* candidato se expresa entonces en el huésped de prueba que carece de *oleC*. Se considera que la OleC candidata tiene actividad hidrocarburo sintasa si el huésped de prueba produce hidrocarburos.

Usando las secuencias de OleA, OleBC u OleC y OleD proporcionadas en el presente documento y el ensayo de complementación descrito en el ejemplo 6, pueden identificarse genes que forman hidrocarburos y productos intermedios de hidrocarburos adicionales. Pueden formarse hidrocarburos y productos intermedios de los mismos expresando OleA, OleBC u OleC y OleD en *E. coli*. Por tanto, puede usarse *E. coli* modificada por ingeniería genética para elaborar hidrocarburos u otros organismos que producen hidrocarburos de manera natural (por ejemplo, *S. maltophilia*, *C. aggregans*, *X. axonopodis* o *A. aurescens*) para determinar la actividad hidrocarburo sintasa de una proteína o secuencia de ADN específica cuando esa secuencia de ADN específica que va a someterse a prueba no se expresa en la célula huésped.

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, una olefina descrita en el presente documento). Una célula huésped puede modificarse para expresar o sobreexpresar genes seleccionados o para tener una expresión atenuada de genes seleccionados.

Como ejemplo, cuando la secuencia de ADN que va a someterse a prueba codifica para un homólogo de una secuencia de proteína OleA, se expresa en un huésped que ya expresa las secuencias de *oleC* y *oleD*, pero no las secuencias de *oleA*. Se considera que el homólogo de OleA es activo (es decir, tiene actividad hidrocarburo sintasa) si el huésped produce hidrocarburos o productos intermedios de hidrocarburos cuando expresa el homólogo de *oleA*.

En el presente documento se da a conocer un método para identificar una enzima útil para la producción de hidrocarburos que comprende (i) transformar una célula que comprende polipéptidos seleccionados del grupo que consiste en (a) OleA y OleD, (b) OleA y OleC, y (c) OleC y OleD con un ácido nucleico que codifica para una enzima que se sospecha que tiene la capacidad para producir hidrocarburos; y (ii) determinar si la célula produce hidrocarburos, en el que la existencia de producción de hidrocarburos por la célula indica que el ácido nucleico codifica para un polipéptido útil para la producción de hidrocarburos.

Por ejemplo, el método puede comprender (i) transformar una célula que comprende polipéptidos seleccionados del grupo que consiste en (a) OleA y OleD, (b) OleA y OleC, (c) OleC y OleD, (d) OleA y OleBC, y (e) OleBC y OleD con un ácido nucleico que codifica para una enzima que se sospecha que tiene la capacidad para producir hidrocarburos, en el que la OleA comparte una identidad de secuencia de al menos el 35% (por ejemplo, al menos el 45%, al menos el 50%, al menos el 60%, al menos el 70%, al menos el 80%, al menos el 85%, al menos el 90%, al menos el 95%, al menos el 96%, al menos el 97%, al menos el 98%, al menos el 99% o al menos el 99,5%) con SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 12 o SEQ ID NO: 18; en el que la OleC comparte una identidad de secuencia de al menos el 35% (por ejemplo, al menos el 45%, al menos el 50%, al menos el 60%, al menos el 70%, al menos el 80%, al menos el 85%, al menos el 90%, al menos el 95%, al menos el 96%, al menos el 97%, al menos el 98%, al menos el 99% o al menos el 99,5%) con SEQ ID NO: 6, SEQ ID NO: 14, SEQ ID NO: 20 o SEQ ID NO: 88; y en el que la OleD comparte una identidad de secuencia de al menos el 35% (por ejemplo, al menos el 45%, al menos el 50%, al menos el 60%, al menos el 70%, al menos el 80%, al menos el 85%, al menos el 90%, al menos el 95%, al menos el 96%, al menos el 97%, al menos el 98%, al menos el 99% o al menos el 99,5%) con SEQ ID NO: 8, SEQ ID NO: 16 o SEQ ID NO: 22, con un ácido nucleico que codifica para una enzima que se sospecha que tiene la capacidad para producir hidrocarburos; y (ii) determinar si la célula produce hidrocarburos, en el que la existencia de producción de hidrocarburos por la célula indica que el ácido nucleico codifica para un polipéptido útil para la producción de hidrocarburos.

Los organismos que producen hidrocarburos de manera natural (sin modificación por ingeniería genética) pueden modificarse por ingeniería genética para sobreproducir hidrocarburos o producir hidrocarburos con características de cadena de carbono específicas alterando la ruta de biosíntesis de ácidos grasos tal como se describe en el presente documento. Los organismos a modo de ejemplo que se sabe que producen hidrocarburos y que pueden modificarse por ingeniería genética para alterar la producción de hidrocarburos usando las enseñanzas proporcionadas en el presente documento incluyen, pero no se limitan a, *Acinetobacter* spp., *Chloroflexus* spp., *Kineococcus radiotolerans*, *Stenotrophomonas maltophilia*, *Micrococcus* spp., *Arthrobacter* spp., *Vibrio furnissii* y cianobacterias. Estos organismos recombinantes modificados por ingeniería genética son útiles para la producción de hidrocarburos.

Los organismos recombinantes modificados por ingeniería genética también pueden usarse para producir cetonas alifáticas. Por ejemplo, los organismos que tienen *oleA*, *oleC* y *oleD* pueden modificarse por ingeniería genética para producir cetonas alifáticas delecionando o atenuando genes que codifican para *oleC* y *oleD*. El organismo modificado genéticamente resultante produce cetonas como resultado de la expresión de *oleA* endógena cuando *oleC* y *oleD* están delecionados o atenuados. Los organismos que tienen *oleA* y *oleC* pueden modificarse por ingeniería genética para producir cetonas alifáticas delecionando o atenuando el gen que codifica para *oleC*. De manera similar, los organismos que tienen *oleA* y *oleD* pueden modificarse por ingeniería para producir cetonas

alifáticas delecionando o atenuando el gen que codifica para *oleD*. El organismo modificado genéticamente resultante produce cetonas como resultado de la expresión de *oleA* endógeno cuando *oleC* o *oleD* están delecionados o atenuados.

5 En otros ejemplos, los organismos recombinantes que producen hidrocarburos se modifican por ingeniería genética para sobreexpresar uno o más péptidos seleccionados de *OleA*, *OleB*, *OleC*, *OleD*, *OleBC*, y combinaciones de los mismos. Estos genes pueden sobreexpresarse en organismos que producen hidrocarburos de manera natural, tales como los descritos anteriormente, o pueden sobreexpresarse en organismos que no producen hidrocarburos de manera natural.

10 Tal como se usa en el presente documento, "sobreexpresar" significa expresar o hacer que se exprese un ácido nucleico, polipéptido o hidrocarburo en una célula a una concentración mayor que la expresada normalmente en una célula silvestre correspondiente. Por ejemplo, un polipéptido puede "sobreexpresarse" en una célula recombinante cuando el polipéptido está presente en una concentración mayor en la célula recombinante en comparación con su concentración en una célula no recombinante de la misma especie.

15 Los ejemplos de organismos recombinantes que sobreexpresan un péptido incluyen organismos que expresan ácidos nucleicos que codifican para *OleA*, *OleB*, *OleC*, *OleD*, *OleBC*, o combinaciones de las mismas. Otros ejemplos incluyen organismos en los que se han introducido secuencias de promotor exógenas en el sentido de 5' de la secuencia codificante endógena de *OleA*, *OleB*, *OleC*, *OleD*, *OleBC*, o combinaciones de las mismas. En algunos ejemplos, la sobreexpresión de uno o más genes que alteran rutas de biosíntesis de ácidos grasos puede sobreexpresarse en combinación con *OleA*, *OleB*, *OleC*, *OleBC* u *OleD*.

20 Se proporcionan organismos recombinantes (por ejemplo, células bacterianas, fúngicas o eucariotas) que están modificados por ingeniería genética (por ejemplo, transformados, transducidos o transfectados) con una o más moléculas de ácido nucleico que codifican para una o más de *OleA* (por ejemplo, SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 12 y SEQ ID NO: 18), *OleB* (por ejemplo, SEQ ID NO: 10), *OleC* (por ejemplo, SEQ ID NO: 6, SEQ ID NO: 14, SEQ ID NO: 20 y SEQ ID NO: 88), *OleD* (por ejemplo, SEQ ID NO: 8, SEQ ID NO: 16 y SEQ ID NO: 22), o
25 una variante u homólogo de una o más de estas secuencias. Estas secuencias pueden expresarse a partir de constructos de vector, directamente a partir del cromosoma tras la integración génica o a partir de matrices extracromosómicas. Por ejemplo, una proteína *OleA* (por ejemplo, SEQ ID NO: 2), *OleC* (por ejemplo, SEQ ID NO: 6) u *OleD* (por ejemplo, SEQ ID NO: 8) está codificada por un ácido nucleico que está operativamente unido a elementos de control de la expresión génica que son funcionales en el organismo recombinante deseado, por
30 ejemplo un promotor T7 en *E. coli*.

Tal como se usa en el presente documento, "elemento de control" significa un elemento de control transcripcional. Los elementos de control incluyen promotores y potenciadores. El término "elemento promotor", "promotor" o "secuencia de promotor" se refiere a una secuencia de ADN que funciona como un interruptor que activa la
35 expresión de un gen. Si el gen está activado, se dice que se transcribe o participa en la transcripción. La transcripción implica la síntesis de ARNm a partir del gen. Un promotor, por tanto, sirve como un elemento regulatorio transcripcional y también proporciona un sitio para la iniciación 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)).

40 Los métodos de expresión de proteínas en sistemas de expresión heterólogos se conocen bien en la técnica. Normalmente, un organismo recombinante de levadura o bacteriano se transforma mediante transformación natural, electroporación, conjugación o transducción. El constructo de expresión resultante puede ser o bien extracromosómico, como con un plásmido, o bien estar integrado en el cromosoma tras la recombinación. En células eucariotas, normalmente, se transfecta un organismo recombinante con (o se infecta con un virus que contiene) un
45 vector de expresión usando cualquier método adecuado para el organismo recombinante particular. Tales métodos de transfección se conocen bien también en la técnica y se describen métodos a modo de ejemplo no limitativos en el presente documento. El organismo recombinante transformado puede expresar la proteína codificada por el ácido nucleico en el casete de expresión. En otra realización, puede realizarse también transfección transitoria o estable del organismo recombinante con uno o más vectores de expresión.

50 Pueden usarse muchos tipos diferentes de organismos recombinantes para producir las proteínas proporcionadas en el presente documento, tales como bacterias, levaduras, algas, hongos, insectos, células de vertebrados (tales como células de mamífero) y células vegetales, incluyendo (tal como sea apropiado) células primarias y líneas celulares inmortales. Se usan frecuentemente numerosos representantes de cada tipo de célula y están disponibles de una amplia variedad de fuentes comerciales, incluyendo, por ejemplo, ATCC, Pharmacia e Invitrogen.

55 Frecuentemente se usan diversas cepas de levadura y vectores derivados de levaduras para la producción de proteínas heterólogas. Por ejemplo, los ejemplos no limitativos, específicos de células de levadura adecuadas incluyen células de *Saccharomyces cerevisiae*, células de *Aspergillus*, células de *Trichoderma*, células de *Neurospora*, células de *Fusarium* o células de *Chryso sporium*. En un ejemplo específico no limitativo, pueden usarse sistemas de expresión de *Pichia pastoris*, obtenidos de Invitrogen (Carlsbad, California), para producir un péptido de *OleA*, *OleB*, *OleC*, *OleBC* u *OleD*. Tales sistemas incluyen cepas de *Pichia pastoris*, vectores, reactivos,

transformantes, cebadores de secuenciación y medios adecuados. Por ejemplo, las cepas disponibles incluyen, pero no se limitan a, KM71H (una cepa prototrófica), SMD1168H (una cepa prototrófica) y SMD1168 (una cepa mutante de pep4) (Invitrogen).

5 *Saccharomyces cerevisiae* es otra especie de levadura usada frecuentemente como huésped. El plásmido YRp7 (Stinchcomb *et al.*, *Nature*, 282: 39, (1979); Kingsman *et al.*, *Gene*, 7: 141, (1979); Tschemper *et al.*, *Gene*, 10: 157, (1980)) se usa frecuentemente como vector de expresión en un mutante de *Saccharomyces* que no puede producir triptófano. Este plásmido contiene el gen *trp1* que cuando se transforma en la cepa mutante de levadura permite que la cepa mutante de la levadura produzca triptófano y crezca en ausencia de triptófano. Los ejemplos de cepas huésped en las que puede usarse el gen *trp1* como marcador de selección incluyen, pero no se limitan a, ATCC n.º 10 44.076 y PEP4-1 (Jones, *Genetics*, 23: 12, (1977)). La presencia de lesión de *trp1* en el genoma del organismo recombinante de levadura proporciona una característica eficaz para detectar la transformación mediante crecimiento en ausencia de triptófano.

15 Los organismos recombinantes de levadura pueden transformarse usando el método de polietilenglicol, tal como se describe por Hinnen, *Proc. Natl. Acad. Sci. USA*, 75: 1929, (1978). Se exponen protocolos de transformación de levadura en Gietz *et al.*, *Nucl. Acids Res.*, 20(17): 1425, (1992) y Reeves *et al.*, *FEMS*, 99(2-3): 193-197, (1992).

Muchos organismos celulares, tales como levadura, animales y bacterias, producen lípidos como componentes esenciales de sus membranas celulares. Los métodos mediante los que estos organismos producen los grupos acilo grasos usados como lípidos están altamente conservados. No obstante, existen variaciones en las rutas que pueden influir en la disponibilidad de algunos de los productos intermedios de lípidos, tales como acil-coA y acil-ACP. Estos 20 productos intermedios de lípidos son también sustratos clave en la producción de cetonas alifáticas e hidrocarburos por las proteínas Ole. En *Saccharomyces cerevisiae*, acil-coA se une a proteína de unión a acil-coenzima A (ACBP), que protege las acil-coA de la hidrólisis por tioesterasas (véanse, por ejemplo, Rose *et al.*, *PNAS*, 89: 11287-11291 (1992), Feddersen *et al.*, *Biochem. J.*, 407: 219-230 (2007)). Las proteínas Ole competirían con la ACBP para usar acil-coA como sustrato para la síntesis de hidrocarburos o cetonas alifáticas. Por tanto, pueden ser necesarias 25 modificaciones en ACBP para reducir su competición por acil-CoA. Las mutaciones condicionales en la ACBP han mostrado en la bibliografía que liberan acumulaciones de acil-CoA en *S. cerevisiae* (véase, por ejemplo, Gaigg *et al.*, *Mol. Biol. Cell*, 12: 1147-1160 (2001)). Por tanto, la expresión de los genes *ole* en una cepa huésped en la que existe control sobre la cantidad de acil-CoA libre disponible, tal como la cepa de *S. cerevisiae* Y700pGAL1-ACB1 (véase, por ejemplo, Gaigg *et al.*, citado anteriormente), debe conducir a mayores niveles de cetonas alifáticas o 30 hidrocarburos producidos.

En la construcción de vectores de expresión adecuados, las secuencias de terminación asociadas con estos genes están ligadas también en la región en 3' de la secuencia que se desea expresar. Cualquier vector de plásmido que contiene un promotor compatible con levadura que puede transcribir una secuencia de ácido nucleico que codifica para un ARNt procarionta, un origen de replicación y una secuencia de terminación, es adecuado.

35 Otros organismos recombinantes adecuados son células bacterianas. Los ejemplos no limitativos, específicos de filos bacterianos adecuados que podrían ser organismos recombinantes incluyen *Acidobacteria*, *Actinobacteria*, *Aquificae*, *Bacteroidetes*, *Chlamydiae*, *Clorobi*, *Cloroflexi Chrysiogenetes*, *Cianobacteria*, *Deferribacteres*, *Deinococcus*, *Thermus*, *Dictyoglomi*, *Fibrobacteres*, *Firmicutes*, *Fusobacteria*, *Gemmatimonadetes*, *Lentisphaerae*, *Nitrospira*, *Planctomycetes*, *Proteobacteria*, *Spirochaetes*, *Tenericutes*, *Thermodesulfobacteria*, *Thermomicrobia*, 40 *Thermotoga* y *Verrucomicrobia*.

Los ejemplos no limitativos, específicos, de especies bacterianas que podrían usarse como organismos recombinantes incluyen *Escherichia coli*, *Thermus thermophilus*, *Stenotrophomonas maltophilia*, *Kineococcus radiotolerans*, *Bacillus stearothermophilus*, *Methanococcusjannaschii*, *Methanosarcina mazei*, *Methanobacterium thermoautotrophicum*, *Methanococcus maripaludis*, *Methanopyrus kandleri*, *Halobacterium* tal como *Haloferax volcanii* y especies de *Halobacterium NRC-i*, *Archaeoglobusfulgidus*, *Pyrococcus fitriosus*, *Pyrococcus horikoshii*, *Pyrobaculum aerophilum*, *Pyrococcus abyssi*, *Sulfolobus solfataricus*, *Sulfolobus tokodaii*, *Aeuropyrum pernix*, 45 *Thermoplasma acidophilum* y *Thermoplasma volcanium*.

En una realización, el organismo recombinante es una célula de *E. coli*, una célula de *S. maltophilia*, una célula de *Pseudomonas sp.*, una célula de *Bacillus sp.*, una célula de *Actinomycetes* o células que pertenecen al género *Rhodococcus*. La introducción del constructo en el organismo recombinante puede lograrse mediante una variedad de métodos incluyendo, pero sin limitarse a, transfección por fosfato de calcio, transfección mediada por dextrano-DEAE, transfección mediada por polibreno, fusión de protoplastos, transfección mediada por liposomas, conjugación, transformación natural, electroporación, y otros métodos conocidos en la técnica.

55 Todavía otros organismos recombinantes adecuados son células vegetales (por ejemplo, licopodios, helechos, angiospermas o gimnospermas). Los organismos recombinantes adecuados adicionales incluyen, pero no se limitan a, algas, musgos y líquenes. Puede emplearse cualquier método conocido para la transformación, el cultivo y la regeneración de células vegetales. Los métodos para la introducción de ADN foráneo en células vegetales incluyen, pero no se limitan a, transferencia que implica el uso de *Agrobacterium tumefaciens* y vectores Ti apropiados, incluyendo vectores binarios; transferencia inducida químicamente (por ejemplo, con polietilenglicol); biolística; y

microinyección. Véase, por ejemplo, An *et al.*, *Plant Molecular Biology Manual*, A3: 1-19 (1988). Se conocen en la técnica diversos promotores adecuados para la expresión de genes heterólogos en células vegetales, incluyendo promotores constitutivos, por ejemplo el promotor 35S del virus del mosaico de la coliflor (CaMV), que se expresa en muchos tejidos vegetales, promotores específicos de órganos o tejidos y promotores que son inducibles mediante productos químicos, tales como jazmonato de metilo, ácido salicílico o protectores.

Los organismos recombinantes se hacen crecer en condiciones apropiadas hasta una densidad celular adecuada. Si la secuencia de interés está operativamente unida a un promotor inducible, se realiza la alteración medioambiental apropiada para inducir la expresión. Si el producto (por ejemplo, hidrocarburo) se acumula en el organismo recombinante, se recogen las células, por ejemplo, mediante centrifugación o filtración. Pueden realizarse extracciones de células completas para purificar los productos de las células completas. En una realización alternativa, podría realizarse una extracción de cultivo completo, en la que el organismo, el medio y el producto se recogen juntos, para recuperar el producto deseado. Entonces puede purificarse el extracto de cultivo completo para obtener el producto deseado. Si los organismos recombinantes secretan el producto al medio, las células y el medio se separan. El medio se conserva entonces para la purificación del producto deseado.

En el presente documento se da a conocer un organismo modificado genéticamente que comprende una secuencia exógena de ácido nucleico incorporada de manera estable en el genoma de un organismo en el sentido de 5' de una secuencia genómica de ácido nucleico que (a) tiene una identidad de secuencia de al menos el 35% (por ejemplo, al menos el 45%, al menos el 50%, al menos el 60%, al menos el 70%, al menos el 80%, al menos el 85%, al menos el 90%, al menos el 95%, al menos el 96%, al menos el 97%, al menos el 98%, al menos el 99% o al menos el 99,5%) con respecto a la secuencia de ácido nucleico de un gen *ole* y (b) codifica para un polipéptido. Por ejemplo, el organismo modificado genéticamente puede comprender una secuencia exógena de ácido nucleico incorporada de manera estable en el genoma de un organismo en el sentido de 5' de una secuencia genómica de ácido nucleico que tiene una identidad de secuencia de al menos el 35% (por ejemplo, al menos el 45%, al menos el 50%, al menos el 60%, al menos el 70%, al menos el 80%, al menos el 85%, al menos el 90%, al menos el 95%, al menos el 96%, al menos el 97%, al menos el 98%, al menos el 99% o al menos el 99,5%) con respecto a una secuencia de ácido nucleico seleccionada del grupo que consiste en SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7, SEQ ID NO: 9, SEQ ID NO: 11, SEQ ID NO: 13, SEQ ID NO: 15, SEQ ID NO: 17, SEQ ID NO: 19, SEQ ID NO: 21, SEQ ID NO: 87, y una secuencia de ácido nucleico que codifica para cualquiera de SEQ ID NO: 135-464. El organismo modificado genéticamente puede comprender una secuencia exógena de ácido nucleico incorporada de manera estable en el genoma de un organismo en el sentido de 5' de una secuencia genómica de ácido nucleico que tiene una identidad de secuencia de al menos el 35% (por ejemplo, al menos el 45%, al menos el 50%, al menos el 60%, al menos el 70%, al menos el 80%, al menos el 85%, al menos el 90%, al menos el 95%, al menos el 96%, al menos el 97%, al menos el 98%, al menos el 99% o al menos el 99,5%) con respecto a una secuencia de ácido nucleico seleccionada del grupo que consiste en SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7, SEQ ID NO: 9, SEQ ID NO: 11, SEQ ID NO: 13, SEQ ID NO: 15, SEQ ID NO: 17, SEQ ID NO: 19, SEQ ID NO: 21 y SEQ ID NO: 87.

El organismo modificado genéticamente puede comprender una secuencia exógena de ácido nucleico incorporada de manera estable en el genoma de un organismo en el sentido de 5' de una secuencia genómica de ácido nucleico que comprende, consiste esencialmente en o consiste en, una secuencia de ácido nucleico seleccionada del grupo que consiste en SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7, SEQ ID NO: 9, SEQ ID NO: 11, SEQ ID NO: 13, SEQ ID NO: 15, SEQ ID NO: 17, SEQ ID NO: 19, SEQ ID NO: 21 y SEQ ID NO: 87.

En el presente documento se da a conocer un organismo modificado genéticamente preparado (a) proporcionando un organismo que tiene una secuencia de ácido nucleico que tiene una identidad de secuencia de al menos el 35% (por ejemplo, al menos el 45%, al menos el 50%, al menos el 60%, al menos el 70%, al menos el 80%, al menos el 85%, al menos el 90%, al menos el 95%, al menos el 96%, al menos el 97%, al menos el 98%, al menos el 99% o al menos el 99,5%) con respecto a la secuencia de ácido nucleico de un gen *ole* y (b) delecionando o mutando la secuencia de ácido nucleico. Por ejemplo, el organismo modificado genéticamente se prepara delecionando o mutando una secuencia de ácido nucleico que tiene una identidad de secuencia de al menos el 35% (por ejemplo, al menos el 45%, al menos el 50%, al menos el 60%, al menos el 70%, al menos el 80%, al menos el 85%, al menos el 90%, al menos el 95%, al menos el 96%, al menos el 97%, al menos el 98%, al menos el 99% o al menos el 99,5%) con respecto a una secuencia de ácido nucleico seleccionada del grupo que consiste en SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7, SEQ ID NO: 9, SEQ ID NO: 11, SEQ ID NO: 13, SEQ ID NO: 15, SEQ ID NO: 17, SEQ ID NO: 19, SEQ ID NO: 21, SEQ ID NO: 87, y una secuencia de ácido nucleico que codifica para cualquiera de SEQ ID NO: 135-464. El organismo modificado genéticamente puede prepararse delecionando o mutando una secuencia de ácido nucleico que tiene una identidad de secuencia de al menos el 35% (por ejemplo, al menos el 45%, al menos el 50%, al menos el 60%, al menos el 70%, al menos el 80%, al menos el 85%, al menos el 90%, al menos el 95%, al menos el 96%, al menos el 97%, al menos el 98%, al menos el 99%, o al menos el 99,5%) con respecto a una secuencia de ácido nucleico seleccionada del grupo que consiste en SEQ ID NO: 5, SEQ ID NO: 7, SEQ ID NO: 13, SEQ ID NO: 15, SEQ ID NO: 19, SEQ ID NO: 21 y SEQ ID NO: 87.

El organismo modificado genéticamente puede prepararse delecionando o mutando una secuencia de ácido nucleico que comprende, consiste esencialmente en o consiste en, una secuencia de ácido nucleico seleccionada del grupo que consiste en SEQ ID NO: 5, SEQ ID NO: 7, SEQ ID NO: 13, SEQ ID NO: 15, SEQ ID NO: 19, SEQ ID NO: 21 y

SEQ ID NO: 87.

El organismo modificado genéticamente puede ser cualquier organismo susceptible de modificación por ingeniería genética. Preferiblemente, el organismo modificado genéticamente se selecciona del grupo que consiste en

5 *Anaeromyxobacter* sp. Fw109-5, *Arthrobacter aureescens*, *Arthrobacter chlorophenolicus*, *Arthrobacter* sp. FB24, *Bdellovibrio bacteriovorus*, *Blastopirellula marina*, *Brevibacterium linens*, *Burkholderia ambifaria*, *Burkholderia cenocepacia*, *Burkholderia oklahomensis*, *Burkholderia pseudomallei*, *Burkholderia* sp. 383, *Candidatus Kueneria stuttgartiensis*, *Chloroflexus aggregans*, *Chloroflexus aurantiacus*, *Clavibacter michiganensis* subsp. *michiganensis*, *Clavibacter michiganensis* subsp. *Sepedonicus*, *Clostridium botulinum* A, *Clostridium botulinum* A3, *Clostridium botulinum* B1, *Clostridium botulinum* Bf, *Clostridium botulinum* F, *Clostridium botulinum*, *Colwellia psychrerythraea*,

10 *Congregibacter litoralis*, *Desulfatibacillum alkenivorans*, *Desulfococcus oleovorans*, *Desulfotalea psychrophila*, *Desulfuromonas acetoxidans*, *Flavobacteriales bacterium*, *Gemmata obscuriglobus*, *Geobacter bemidjensis*, *Geobacter lovleyi*, *Geobacter* sp. FRC-32, *Geobacter uraniumreducens*, *Hahella chejuensis*, *Jannaschia* sp. CCS1, *Kineococcus radiotolerans*, *Lentisphaera araneosa*, *Maricaulis maris*, *Marinobacter algicola*, *Marinobacter aquaeolei*, *Micrococcus luteus*, *Microscilla marina*, *Moritella* sp. PE36, *Mycobacterium avium*, *Mycobacterium avium* subsp. *paratuberculosis*, *Mycobacterium marinum*, *Mycobacterium* sp. GP1, *Mycobacterium tuberculosis*, *Mycobacterium vanbaalenii*, *Neisseria gonorrhoeae*, *Nocardia farcinica*, *Opiritaceae bacterium*, *Opiritatus terrae*, *Paenibacillus* sp. JDR-2, *Pelobacter propionicus*, *Photobacterium profundum*, *Photobacterium profundum*, *Photorhabdus luminescens* subsp. *laumondii*, *Planctomyces maris*, *Plesiocystis pacifica*, *Pseudoalteromonas atlantica*, *Psychromonas ingrahamii*, *Psychromonas* sp. CNPT3, *Ralstonia pickettii*, *Rhodococcus rhodochrous*, *Rhodococcus* sp.,

20 *Rhodopirellula baltica*, *Roseovarius nubinhibens*, *Shewanella amazonensis*, *Shewanella baltica*, *Shewanella benthica*, *Shewanella denitrificans*, *Shewanella frigidimarina*, *Shewanella halifaxensis*, *Shewanella loihica*, *Shewanella oneidensis*, *Shewanella pealeana*, *Shewanella putrefaciens*, *Shewanella sediminis*, *Shewanella* sp. ANA-3, *Shewanella* sp. MR-4, *Shewanella* sp. MR-7, *Shewanella* sp. W3-18-1, *Shewanella woodyi*, *Stenotrophomonas maltophilia*, *Streptomyces ambofaciens*, *Xanthomonas axonopodis*, *Xanthomonas campestris*,

25 *Xanthomonas oryzae* y *Xylella fastidiosa*.

La secuencia exógena de ácido nucleico incorporada de manera estable en el genoma de un organismo modificado genéticamente puede aumentar la expresión del polipéptido codificado por la secuencia genómica de ácido nucleico en relación con la expresión del mismo polipéptido en el organismo antes de la incorporación de la secuencia exógena de ácido nucleico en las mismas condiciones. El organismo modificado genéticamente puede producir una

30 cantidad aumentada de un hidrocarburo en relación con la producción de un hidrocarburo en un organismo silvestre en las mismas condiciones.

La delección o mutación de la secuencia de ácido nucleico en el organismo modificado genéticamente puede dar como resultado una producción disminuida de un polipéptido codificado por la secuencia de ácido nucleico en relación con la producción del mismo polipéptido en el organismo antes de la delección o mutación de la secuencia de

35 ácido nucleico en las mismas condiciones. El organismo modificado genéticamente puede producir cetonas. Por ejemplo, el organismo modificado genéticamente puede producir una cantidad aumentada de cetonas en relación con la producción de cetonas en el organismo antes de la delección o mutación de la secuencia de ácido nucleico en las mismas condiciones.

En el presente documento se dan a conocer métodos de elaboración de cetonas alifáticas, hidrocarburos y productos intermedios de los mismos. Se proporcionan diversos organismos recombinantes que pueden usarse para producir productos que tienen longitudes de la cadena de carbono, sitio de insaturación y puntos de ramificación modificados por ingeniería genética. También se proporcionan métodos de elaboración de tales productos así como métodos de modificación adicional de los productos, tales como mediante craqueo, para crear biocombustibles y productos químicos de especialidad de alta calidad.

40

Las cetonas alifáticas, los hidrocarburos y los productos intermedios de hidrocarburos pueden modificarse por ingeniería genética para tener características de cadena de carbono específicas expresando diversas enzimas o atenuando la expresión de diversas enzimas en el organismo recombinante. Por ejemplo, la longitud de la cadena de carbono puede controlarse expresando diversas tioesterasas en el organismo recombinante al tiempo que se atenúa la expresión de tioesterasas endógenas. De manera similar, pueden introducirse diversos puntos de ramificación en

50 la cadena de carbono expresando diversos genes α -cetoácido descarboxilasa/deshidrogenasa de cadena ramificada (por ejemplo, genes *bkd*), y el grado de saturación también puede controlarse expresando diversos genes, por ejemplo, sobreexpresando genes de β -ceto-ACP-sintasa (por ejemplo, *fabB*). Se proporciona una descripción detallada de las diversas modificaciones que pueden incluirse en un organismo recombinante para proporcionar características de cadena de carbono específicas y para aumentar la producción de la ruta de biosíntesis de ácidos grasos en la publicación de solicitud de patente internacional WO 2007/136762, que se incorpora en el presente documento en su totalidad por referencia.

55

Se da a conocer en el presente documento un método para producir biocombustibles que comprende cualquiera de los métodos descritos en el presente documento. Tal como se usa en el mismo, el término "biocombustible" se refiere a cualquier combustible derivado de biomasa. Los biocombustibles pueden sustituir a combustibles a base de petróleo. Por ejemplo, los biocombustibles incluyen combustibles de transporte (por ejemplo, gasolina, diesel, combustible de motores de reacción, etc.), combustibles de calefacción y combustibles de generación de

60

electricidad. Los biocombustibles son una fuente de energía renovable. Específicamente, el biocombustible producido puede ser una gasolina, biodiesel o combustible de motores de reacción.

- 5 Los biocombustibles que comprenden hidrocarburos producidos de manera biológica, particularmente hidrocarburos producidos de manera biológica usando la ruta de biosíntesis de ácidos grasos, no se han producido a partir de fuentes renovables y, como tales, son nuevas composiciones de materia. Estos nuevos combustibles pueden distinguirse de los combustibles derivados del carbono petroquímico basándose en la huella de carbono isotópico doble o la datación de ^{14}C . Adicionalmente, la fuente específica de carbono de biofuentes (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).
- 10 La capacidad de distinguir biocombustibles de combustibles a base de petróleo es beneficiosa en el seguimiento de estos materiales en el comercio. Por ejemplo, pueden distinguirse productos químicos o combustibles que comprenden perfiles de isótopos de carbono tanto de base biológica como a base de petróleo de productos químicos y combustibles preparados únicamente a partir de materiales a base de petróleo. Por tanto, puede realizarse un seguimiento de los presentes materiales en el comercio basándose en su perfil de isótopos de carbono único.
- 15 Los biocombustibles pueden distinguirse de los combustibles a base de petróleo comparando la razón de isótopos de carbono estables ($^{13}\text{C}/^{12}\text{C}$) en cada combustible. La razón $^{13}\text{C}/^{12}\text{C}$ en un material de base biológica dado es una consecuencia de la razón $^{13}\text{C}/^{12}\text{C}$ en el dióxido de carbono atmosférico en el momento de fijar el dióxido de carbono. También refleja la ruta metabólica precisa. También se producen variaciones regionales. El petróleo, las plantas C_3 (de hoja ancha), las plantas C_4 (las hierbas) y los carbonatos marinos muestran diferencias significativas en $^{13}\text{C}/^{12}\text{C}$ y los valores de $\delta^{13}\text{C}$ correspondientes. Además, los análisis de materia lipídica de plantas C_3 y C_4 son diferentes de los materiales derivados de componentes de hidratos de carbono de las mismas plantas como consecuencia de la ruta metabólica.

25 Dentro de la precisión de la medición, ^{13}C muestra grandes variaciones debido a efectos de fraccionamiento isotópico, el más significativo de los cuales para biocombustibles es el mecanismo fotosintético. La principal causa de diferencias en la razón de isótopos de carbono en plantas está estrechamente asociada con diferencias en la ruta del metabolismo de carbono fotosintético en las plantas, particularmente la reacción que se produce durante la carboxilación primaria (es decir, la fijación inicial de CO_2 atmosférico). Dos grandes clases de vegetación son las que incorporan el ciclo fotosintético " C_3 " (o de Calvin-Benson) y las que incorporan el ciclo fotosintético " C_4 " (o de Hatch-Slack).

30 En las plantas C_3 , la reacción de carboxilación o fijación de CO_2 principal implica la enzima ribulosa-1,5-difosfato carboxilasa, y el primer producto estable es un compuesto de 3 carbonos. Las plantas C_3 , tales como de madera dura y coníferas, son dominantes en las zonas de clima templado.

35 En las plantas C_4 , una reacción de carboxilación adicional que implica otra enzima, fosfoenol-piruvato carboxilasa, es la reacción de carboxilación principal. El primer compuesto de carbono estable es un ácido de 4 carbonos que posteriormente se descarboxila. El CO_2 así liberado vuelve a fijarse en el ciclo C_3 . Ejemplos de plantas C_4 son hierbas tropicales, maíz y caña de azúcar.

40 Tanto las plantas C_4 como C_3 muestran un intervalo de razones isotópicas $^{13}\text{C}/^{12}\text{C}$, pero los valores normales son de -7 a -13 por mil para plantas C_4 y de -19 a -27 por mil para plantas C_3 (véase, por ejemplo, Stuiver *et al.*, Radiocarbon 19:355, 1977). El carbón y el petróleo se encuentran generalmente en este último intervalo. La escala de medición de ^{13}C se definió originalmente por un ajuste a cero por la caliza Pee Dee Belemnite (PDB), en la que los valores se facilitan en partes por mil desviaciones a partir de este material. Los valores de " $\delta^{13}\text{C}$ " son en partes por mil (por mil), abreviado, ‰, y se calculan de la siguiente manera:

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

45 Dado que el material de referencia (RM) PDB se ha agotado, se ha desarrollado una serie de RM alternativos en colaboración con IAEA, USGS, NIST, y otros laboratorios de isótopos internacionales seleccionados. Las anotaciones para las desviaciones por mil a partir de PDB son de $\delta^{13}\text{C}$. Las mediciones se realizan con CO_2 mediante espectrometría de masas de razón estable de alta precisión (IRMS) con iones moleculares con masas de 44, 45 y 46.

50 Se da a conocer en el presente documento un hidrocarburo o biocombustible producido mediante cualquiera de los métodos dados a conocer en el presente documento. Específicamente, el hidrocarburo o biocombustible puede tener una $\delta^{13}\text{C}$ de -28 o superior, -27 o superior, -20 o superior, -18 o superior, -15 o superior, -13 o superior, -10 o superior y -8 o superior. Por ejemplo, el hidrocarburo puede tener una $\delta^{13}\text{C}$ de -30 a -15, de -27 a -19, de -25 a -21, de -15 a -5, de -13 a -7 o de -13 a -10. También se da a conocer en el presente documento un hidrocarburo o biocombustible con una $\delta^{13}\text{C}$ de -10, -11, -12 o -12,3.

55 Los biocombustibles también pueden distinguirse de combustibles a base de petróleo comparando la cantidad de ^{14}C en cada combustible. Dado que ^{14}C tiene una semivida nuclear de 5730 años, los combustibles a base de

petróleo que contienen carbono “más antiguo” pueden distinguirse de los biocombustibles que contienen carbono “más reciente” (véase, por ejemplo, Currie, “Source Apportionment of Atmospheric Particles”, Characterization of Environmental Particles, J. Buffle y H. P. van Leeuwen, Eds., 1 de vol. I de IUPAC Environmental Analytical Chemistry Series (Lewis Publishers, Inc) (1992) 3-74).

5 La suposición básica en la datación de radiocarbono es que la constancia de la concentración de ^{14}C en la atmósfera conduce a la constancia de ^{14}C en los organismos vivos. Sin embargo, debido a pruebas nucleares atmosféricas desde 1950 y al quemado de combustible fósil desde 1850, ^{14}C ha adquirido una segunda característica temporal geoquímica. Su concentración en el CO_2 atmosférico, y por tanto en la biosfera viva, aproximadamente se duplicó en el máximo de las pruebas nucleares, a mediados de la década de 1960. Desde entonces ha vuelto gradualmente a la tasa de isótopos ($^{14}\text{C}/^{12}\text{C}$) inicial cosmogénica (atmosférica) en estado estacionario de aproximadamente $1,2 \times 10^{-12}$, con una “semivida” de relajación aproximada de 7-10 años. (Esta última semivida no debe interpretarse de manera literal; en vez de eso, debe usarse la función de introducción/degradación nuclear atmosférica detallada para rastrear la variación de ^{14}C atmosférico y biosférico desde el inicio de la era nuclear).

15 Esta última característica temporal de ^{14}C biosférico es la que mantiene la promesa de una datación anual de carbono biosférico reciente. ^{14}C puede medirse mediante espectrometría de masas con acelerador (AMS), con resultados facilitados en unidades de “fracción de carbono moderno” (f_M). f_M se define por el National Institute of Standards and Technology (NIST) Standard Reference Materials (SRM) 4990B y 4990C. Tal como se usa en el presente documento, la “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, conocidos 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 (con referencia a 1950 DC). Esto es aproximadamente equivalente a la madera previa a la revolución industrial corregida para la degradación. Para la biosfera viva actual (material vegetal), f_M es de aproximadamente 1,1.

25 Se da a conocer en el presente documento un hidrocarburo o biocombustible que puede tener una f_M de ^{14}C de al menos 1. Por ejemplo, el hidrocarburo o biocombustible puede tener una f_M de ^{14}C de al menos 1,01, una f_M de ^{14}C de 1 a 1,5, una f_M de ^{14}C de 1,04 a 1,18 o una f_M de ^{14}C de 1,111 a 1,124.

30 Otra medición de ^{14}C se conoce como el porcentaje de carbono moderno, pMC. Para un arqueólogo o geólogo que usa dataciones de ^{14}C , 1950 DC es igual a “cero años de edad”. Esto también representa 100 pMC. El “carbono de bombas” en la atmósfera alcanzó casi el doble del nivel normal en 1963 en el máximo de las armas termonucleares. Su distribución dentro de la atmósfera se ha aproximado desde su aparición, mostrando valores que son superiores a 100 pMC para plantas y animales vivos desde 1950 DC. Ha disminuido gradualmente a lo largo del tiempo, siendo el valor actual próximo a 107,5 pMC. Esto significa que un material de biomasa reciente, tal como maíz, tendrá una firma de ^{14}C próxima a 107,5 pMC. Los compuestos a base de petróleo tendrán un valor de pMC de cero. Combinar carbono fósil con carbono actual dará como resultado una dilución del contenido de pMC actual. Suponiendo que 35 107,5 pMC representa el contenido de ^{14}C de los materiales de biomasa actuales y 0 pMC representa el contenido de ^{14}C de los productos a base de petróleo, el valor de pMC medido para ese material reflejará las proporciones de los dos tipos de componentes. Por ejemplo, un material derivado al 100% de semillas de soja actuales dará una firma de radiocarbono próxima a 107,5 pMC. Si ese material se diluyó al 50% con productos a base de petróleo, dará una firma de radiocarbono de aproximadamente 54 pMC.

40 Un contenido de carbono de base biológica se deriva asignando el 100% igual a 107,5 pMC y el 0% igual a 0 pMC. Por ejemplo, una muestra que mide 99 pMC dará un contenido de carbono de base biológica equivalente del 93%. Este valor se denomina el resultado de carbono de base biológica medio y supone que todos los componentes dentro del material analizado se originaron o bien de material biológico actual o bien de material a base de petróleo.

45 Se da a conocer en el presente documento un hidrocarburo o biocombustible que puede tener un pMC de al menos 50, 60, 70, 75, 80, 85, 90, 95, 96, 97, 98, 99 ó 100. Se da a conocer además en el presente documento un hidrocarburo o combustible que tiene un pMC de entre 50 y 100, 60 y 100, 70 y 100, 80 y 100, 85 y 100 y 87 y 98, 90 y 95. Se da a conocer además en el presente documento un hidrocarburo o biocombustible con un pMC de 90, 91, 92, 93, 94 ó 94,2.

50 El hidrocarburo puede ser una olefina. La olefina puede ser monoinsaturada o poliinsaturada (por ejemplo, diinsaturada, triinsaturada, etc.). La olefina puede tener una longitud de cadena de carbono de entre 10 y 40 carbonos. Por ejemplo, la olefina puede tener una longitud de cadena de carbono de entre 15 y 35, 17 y 34, 18 y 33, 19 y 33 carbonos, entre 27 y 33 hidrocarburos, entre 29 y 31 hidrocarburos, o 27, 28, 29, 30, 31, 32 ó 33 hidrocarburos. El hidrocarburo puede ser un hidrocarburo de cadena lineal o un hidrocarburo de cadena ramificada. El hidrocarburo puede comprender un resto cíclico.

55 Se da a conocer en el presente documento un biocombustible que comprende los hidrocarburos dados a conocer en el presente documento. El biocombustible puede ser gasolina, biodiesel o combustible de motores de reacción. El biocombustible puede derivarse de una fuente de carbono. Tal como se usa en el presente documento, la frase “fuente de carbono” se refiere a un sustrato o compuesto adecuado para usarse como fuente de carbono para el crecimiento de células eucariotas sencillas o procariotas. 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). Estos 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
 5 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 la fotosíntesis, incluyendo, pero sin limitarse a, glucosa.

Una fuente de carbono preferida es biomasa. 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.
 10 Una fuente de biomasa a modo de ejemplo es materia vegetal. Por ejemplo, puede usarse maíz, caña de azúcar o pasto varilla como biomasa. Otro ejemplo no limitativo de biomasa es materia animal, por ejemplo estiércol de vaca. La biomasa también incluye productos de desecho de la industria, agricultura, silvicultura y del hogar. Ejemplos de tales productos de desecho que pueden usarse como biomasa son desechos de fermentación, paja, leña, aguas residuales, basura y sobras de alimentos. La biomasa también incluye fuentes de carbono, tales como hidratos de
 15 carbono (por ejemplo, monosacáridos, disacáridos o polisacáridos). Otra fuente de carbono preferida es glucosa. Preferiblemente, la fuente de carbono es una fuente de energía renovable. La fuente de energía renovable puede ser una biomasa.

Los siguientes ejemplos ilustran adicionalmente la invención, pero no deben interpretarse como limitativos en ningún modo de su alcance.

20 EJEMPLO 1

Este ejemplo describe materiales y métodos usados para llevar a cabo los ejemplos dentro de esta patente. Aunque se describen métodos particulares, un experto habitual en la técnica comprenderá que también pueden usarse otros métodos similares. En general, se usaron prácticas de laboratorio convencionales, a menos que se estipule otra cosa. Por ejemplo, se usaron prácticas de laboratorio convencionales para: clonación; manipulación y secuenciación
 25 de ácidos nucleicos; purificación y análisis de proteínas; y otras técnicas de bioquímica y biología molecular. Tales técnicas se explican en detalle en manuales de laboratorio convencionales, tales como Sambrook *et al.*, citado anteriormente; y Ausubel *et al.*, citado anteriormente.

Secuencias genómicas: La secuencia genómica completa de *Stenotrophomonas maltophilia* está disponible para el público para dos cepas diferentes. La secuencia genómica completa de *S. maltophilia* R551-3 puede encontrarse en
 30 <http://genome.ornl.gov/microbial/smal/> (último acceso el 16 de mayo de 2007). La secuencia genómica completa de la cepa de *S. maltophilia* K279a puede encontrarse en http://www.sanger.ac.uk/Projects/S_maltophilia/ (último acceso del 16 de mayo de 2007). Las secuencias de ácido nucleico que codifican para las proteínas descritas en el presente documento se encuentran en ambos genomas secuenciados y se confirmaron experimentalmente en la
 35 cepa de *S. maltophilia* de ATCC 17679. Además, algunas de las actividades proteicas se confirmaron en *E. coli*, *B. subtilis*, *B. megaterium* y *S. cerevisiae* usando secuencias de ácido nucleico y secuencias de ácido nucleico con codones optimizados tal como se describe en el presente documento.

Cepas microbianas: Las cepas microbianas usadas en el presente documento fueron las siguientes:

S. maltophilia (números de cepas de la ATCC: 17674, 17679, 17445, 17666);

S. maltophilia ATCC 17679 Δ *oleA*

40 *S. maltophilia* ATCC 17679 Δ *oleC*

S. maltophilia ATCC 17679 Δ *oleD*

E. coli T7 Express lysY/I^q (New England Biolabs, Ipswich, MA 01938-2723)

E. coli C41(DE3) (Lucigen Corporation, Middleton, WI 53562)

45 *E. coli* C41(DE3) Δ *fadE* (la cepa de *E. coli* C41(DE3) de Lucigen Corporation, Middleton, WI 53562 con una delección del gen *fadE* EC 1.3.99.3, una acil-CoA deshidrogenasa, Klein. K. *et al.*, *Eur. J. Biochem. II* 19: 442-450 (1971)).

E. coli C41(DE3) Δ *fadE*; pET-21b(+)_OleA, pCOLADuet-1_OleC, pCDFDuet-1_OleD

E. coli C41(DE3); pET-21b(+)_OleA

E. coli C41(DE3); pET-21b(+)_OleB

E. coli C41(DE3); pET-21b(+)_OleD

50 *E. coli* C41(DE3); pETDuet-1_OleAB; pCOLADuet-1_OleCD

ES 2 646 815 T3

E. coli MG1655 Δ *fadE* (la *E. coli* MG1655 con una delección del gen *fadE*)

E. coli MG1655 Δ *fadE*, *fadD* (+) (la cepa de *E. coli* MG1655 con una delección del gen *fadE* y una secuencia de promotor T5 en el sentido de 5' de *fadD*)

Saccharomyces cerevisiae Hansen, teleomorfo BY4741 (ATCC 201388)

5 *S. cerevisiae* BY4741; pESC-HIS_OleA

Bacillus megaterium WH320 (cepa de Mo Bi Tec, Alemania)

B. megaterium WH320; pWH1520_OleCDAB

B. megaterium WH320; pWH1520

10 *Bacillus subtilis* IHA01 *lacA::spec leuB8 metB5 r(-)m(+)* Sp (cepa de Bacillus Genetic Stock Center, Columbus, OH número de cepa BGSC IA785)

B. subtilis IHA01, pHT01_OleA

Arthrobacter aureescens TC1 (cepa de ATCC BAA-1386)

15 Marcadores de resistencia: En los ejemplos descritos en el presente documento, se usaron los marcadores de resistencia AmpR, ampicilina/carbenicilina (100 μ g/ml); KanR, kanamicina (30 μ g/ml); CamR, cloramfenicol (34 μ g/ml); SmR, estreptomina/espectinomicina (50 μ g/ml); y tetraciclina (*E. coli*, *B. megaterium* 15 μ g/ml, *S. maltophilia* 50 μ g/ml).

20 Reacción en cadena de la polimerasa (PCR): Se usó PCR para amplificar las secuencias de ácido nucleico especificadas a partir de ADN para crear muchos de los constructos de expresión descritos en el presente documento. Los cebadores usados para las reacciones de PCR descritas en el presente documento se enumeran en la tabla 14. Los plásmidos se enumeran en la tabla 15.

Nombre	Secuencia 5' a 3'
Gen 1: /locus_eti.="OleA"	
LB118	GATACATATGCTCTTCAAGAATGTCTCG (SEQ ID NO: 29)
LB119	TCAGCTCGAGCCAGACCACTTCAGCCATCGAG (SEQ ID NO: 30)
Gen 2: /locus_eti.="OleC"	
LB155	GATACATATGAACCGACCCTGCAATATTGC (SEQ ID NO: 31)
LB159	TCAGCTCGAGTCATGCGCGCTTCTCCAGTTCGGCGCTGGC (SEQ ID NO: 32)
Gen 3: /locus_eti.="OleD"	
LB157	GATACATATGAAGATCCTGGTCACCGGTGGTGG (SEQ ID NO: 33)
LB158	TCAGCTCGAGCTATGCGGCAGATGAAGCCTTCAG (SEQ ID NO: 34)
Gen 4: /locus_eti.="OleB"	
LB151	GATACATATGTCCCAGCTTCCCGGTTACC (SEQ ID NO: 35)
LB152	TCAGCTCGAGTCAGATCGGGTCTTGTCCAGG (SEQ ID NO: 36)
Gen 1 y 4 como operón	
LB118	GATACATATGCTCTTCAAGAATGTCTCG (SEQ ID NO: 29)
LB152	TCAGCTCGAGTCAGATCGGGTCTTGTCCAGG (SEQ ID NO: 36)
Gen 2 y 3 como operón	
LB155	GATACATATGAACCGACCCTGCAATATTGC (SEQ ID NO: 31)
LB158	TCAGCTCGAGCTATGCGGCAGATGAAGCCTTCAG (SEQ ID NO: 34)
Cebadores usados para clonar genes que producen hidrocarburos adicionales	
LF305	GGATACATGTTATTCAAAAATGTATCTATC (SEQ ID NO: 37)
LF306	CTCGAGAAGCTTACCACACAACCTCAGCC (SEQ ID NO: 38)
LF307	GGATACATGTTATTTCAGAATGTTTCTATCGC (SEQ ID NO: 39)
LF308	CTCGAGAAGCTTACCAAACCACTTCGGCCATGCTG (SEQ ID NO: 40)
LF313	GGATACATGTTATTCAAGCACGTCATGATCG (SEQ ID NO: 41)
LF314	CTCGAGAAGCTTACCACGTAACGGACATCATAG (SEQ ID NO: 42)
Cebadores para clonar <i>oleA</i> en el vector de expresión de <i>S. cerevisiae</i> pESC-HIS	
LB217	CAATGGGCCCAAACAATGCTCTTCAAGAATGTCTCG (SEQ ID NO: 43)
LF304	TCATCTCGAGTTACCAGACCACTTCAGCCATCGAG (SEQ ID NO: 44)
Cebadores para clonar <i>oleA</i> en el vector pCL1920pTrc	
LB189	GATCACATGTTATTCAAGAATGTCTCGATCG (SEQ ID NO: 45)
Cebadores para clonar <i>fadD</i> en los vectores Duet, pCDFDuet-1 o pETDuet-1	
fadD_F	CATGCCATGGTGAAGAAGGTTTGGCTTAA (SEQ ID NO: 46)
fadD_R	CCCAAGCTTTCAGGCTTTATTGTCCAC (SEQ ID NO: 47)
Cebadores para reemplazar el promotor de <i>fadD</i> por el promotor T5	
LB284	TTTTTAAAGAAAAGAAACAGCGGCTGGTCCGCTGTTTCTGCATTC

	TTACGGATGCATATGGCGGCCGC (SEQ ID NO: 48)
LB285	AACCTTCTTCAATTCTTCACCTCTAAAATGCGTGTTTCGTCGTCATCG CAATTGAATCTATTATAATTGTTATCCGCTCACAAAGCAAATAAAT TTTTTATGATTTGTATCGATAAGCTGGATCCA (SEQ ID NO: 49)
Cebadores usados para clonar genes de tioesterasa	
TesA_F	CATATGGCGGACACGTTATTGATT (SEQ ID NO: 50)
TesA_R	CCTAGGTTATGAGTCATGATTTACTAAAG (SEQ ID NO: 51)
CcFatB_F	TCTAGAGTCTCGAGACATATGCTCCCAGATTGGAGTATGTT (SEQ ID NO: 52)
CcFatB_R	TCTAGACGCTCTAAGCTTCTAGGTTAAACGCTACTCT (SEQ ID NO: 53)
ChFatB2_ F	TCTAGAGTCTCGAGACATATGCTCCCCGATTGGTCCCG (SEQ ID NO: 54)
ChFatB2_ R	TCTAGACGCTCTAAGCTTCTAGGTTAAGAGACCGAATT (SEQ ID NO: 55)
UcFatB1_ F	TCTAGAGTCTCGAGACATATGCTCCCTGATTGGAGTATGTTA (SEQ ID NO: 56)
UcFatB1_ R	TCTAGACGCTCTAAGCTTCTAGGTTAAACGCGCG (SEQ ID NO: 57)
Cebadores usados para preparar constructos de vector usados para delecionar el marco de lectura abierto de <i>oleA</i> en <i>S. maltophilia</i>	
LB200	GTCAGCGGTGACTTTGTTACCAGACGAAGAGCATCGAGGTTAGCGC CAAG (SEQ ID NO: 58)
LB201	CTTGGCGCTAACCTCGATGCTCTTCGTCTGGTAACAAAGTCACCGC TGAC (SEQ ID NO: 59)
LB202	GCTCTAGAGGTCAACCGCATCCGCAACATCC (SEQ ID NO: 60)
LB203	CTTCACCTCGCGCCTGAACATGAAC (SEQ ID NO: 61)
LB204	TGACAAGCTTAGCGCGGCCTGGAAGCCCTTCAGG (SEQ ID NO: 62)
LB205	AAGGCCAGCCCGCAGTGCGACAAG (SEQ ID NO: 63)
Cebadores usados para preparar el vector pCL1920pTrc	
LF302	ATATGACGTCCGCATCCGCTTACAGACA (SEQ ID NO: 76)
LF303	AATTCTTAAGTCAGGAGAGCGTTCACCGACAA (SEQ ID NO: 77)
Cebadores usados para preparar pCL1920pTrcOleCDAB	
LB270	AACAATAATTCAGACATACATC (SEQ ID NO: 79)
LB271	CTTCCGCTAGCTGCCTGACGCCAGAAGCATTGGTGC (SEQ ID NO: 80)
LB272	ATAGGATCCTAAGGAGGAATAAACCATGTTGTTCAAAAATGTATC (SEQ ID NO: 81)

LB275	TTAAGATCTAGATCGGGTTTTTATCCAGGAACGCACG (SEQ ID NO: 82)
Cebadores usados para clonar el casete loxPcat	
LB284	TTTTTAAAGAAAAAGAAACAGCGGCTGGTCCGCTGTTTCTGCATTC TTACGGATGCATATGGCGGCCGC (SEQ ID NO: 83)
LB285	AACCTTCTTCAATTCTTCACCTCTAAAATGCGTGTTTCGTCGTCATCG CAATTGAATCTATTATAATTGTTATCCGCTCACAAAGCAAATAAAT TTTTTATGATTTGTATCGATAAGCTGGATCCA (SEQ ID NO: 84)
Cebadores usados para subclonar oleC a partir de Plesiocystis pacifica	
LB223	GCTGAACATGTCTACTGAAGGTGCT (SEQ ID NO: 85)
LB224	TGCAAGCTTACGCAGACGTGCTTCTGC (SEQ ID NO: 86)
Cebadores usados para preparar constructos de vector usados para delecionar el marco de lectura abierto de oleC en S. maltophilia	
LB110	CCGGTGACCAGGATCTTCATGCGCGTCGGTTCATCCGTTTCATTGTCG CCG (SEQ ID NO: 465)
LB111	CGGCGACAATGAACGGATGAACCGACGCGCATGAAGATCCTGGTC ACCGG (SEQ ID NO: 466)
LB112	TGTCGCACCACGACCAGGTcAAG (SEQ ID NO: 467)
LB113	CATCTCTAGAACTGGAAGATCGGCGAGTGGATC (SEQ ID NO: 468)
LB114	GCGTAATGCAGCATGGCGCTC (SEQ ID NO: 469)
LB115	TGACAAGCTTTCAGCCTGCGCAGCCCTTCTTC (SEQ ID NO: 470)
Cebadores usados para preparar constructos de vector usados para delecionar el marco de lectura abierto de oleD en S. maltophilia	
LB122	AGCCAGTGACGGCGATCTATGCGGCGATCTTCATGCGCGCTTCTCC AGTT (SEQ ID NO: 472)
LB123	AACTGGAGAAGCGCGCATGAAGATCGCCGCATAGATCGCCGTCAC TGGCT (SEQ ID NO: (473)
LB124	AGGACATGGCCGCGATCCTG (SEQ ID NO: 474)
LB125	CATCTCTAGATGGGCAGCGCCTTCGGCATGG (SEQ ID NO: 475)
LB127	TGACAAGCTTCGGCCAACCAGGGTTCGTCGG (SEQ ID NO: 476)
Cebadores usados para clonar oleA en el vector de expresión de B. megaterium pWH1520	
LB402	ATATACTAGTATGAATCGTCCGTGC (SEQ ID NO: 488)
LB387	TATATCTAGACCCGGGAGATCTAGATCGGGTTTTTATC (SEQ ID NO: 489)
Cebadores usados para clonar oleA en el vector de expresión de B. subtilis pHT01	
LB386	TATATCTAGACCCGGGTACCACACAACCTCAG (SEQ ID NO: 490)
LB388	TATAGGATCCATGTTGTTCAAAA (SEQ ID NO: 491)

TABLA 15

Vectores			Fuente
pET-21b(+)	Expresión de T7		Novagen Brand
pET-21d(+)	Vectores		EMD Chemicals, Inc. Gibbstown, NJ 08027
pCOLADuet-1			
pCDFDuet-1			
pETDuet-1			
pACYCDuet-1			

pJ201	Estructura principal de vector usada por DNA2.0		DNA2.0 Menlo Park, CA
pEX18-Tc	Vector suicida para <i>S. maltophilia</i>		Hoang <i>et al.</i> 1998
pESC-HIS	Vector de expresión usado en <i>S. cerevisiae</i>		Stratagene, La Jolla, CA
pTrcHis2A	Vector de expresión		Invitrogen, Carlsbad, CA
pCL1920 (NCCB#3176)	Plásmido de bajo número de copias		The Netherlands Culture Collection of Bacteria, Utrecht, Países bajos
pCL1920pTrc	Plásmido de bajo número de copias para expresión		Descrito en el presente documento
pLoxCat2 pJW168			Palmeros <i>et al.</i> 2000
pKD46			Datsenko y Wanner 2000
pMAL-c2X			NEB, Ipswich, MA
pWH1520	Vector de expresión usado en <i>B. megaterium</i>		Mo Bi Tec, Alemania
pHT01	Vector de expresión usado en <i>B. subtilis</i>		Mo Bi Tec, Alemania
Constructos de expresión	Vector	Inserto	
OleA de <i>S. maltophilia</i> : pET-21b(+)_OleA pET-21d(+)_OleA pCOLADuet-1_OleA pCDFDuet-1_OleA pETDuet-1_OleA pACYCDuet-1_OleA: ATCC17679	pET-21b(+), o pCOLADuet-1 cortado con NdeI y XhoI	Producto de PCR derivado de los cebadores LB118 y LB119 cortado con NdeI y XhoI	Descrito en el presente documento
OleC de <i>S. maltophilia</i> : pET-21b(+)_OleC pET-21d(+)_OleC pCOLADuet-1_OleC pCDFDuet-1_OleC pETDuet-1_OleC pACYDuet-1_OleC	vectores anteriores cortados con NdeI y XhoI	Producto de PCR derivado de los cebadores LB155 y LB159 cortado con NdeI y XhoI	Descrito en el presente documento
OleD de <i>S. maltophilia</i> : pET-21b(+)_OleD pET-21d(+)_OleD pCOLADuet-1_OleD pCDFDuet-1_OleD pETDuet-1_OleD pACYDuet-1_OleD	vectores anteriores cortados con NdeI y XhoI	Producto de PCR derivado de los cebadores LB157 y LB158 cortado con NdeI y XhoI	Descrito en el presente documento
OleB de <i>S. maltophilia</i> : pET-21b(+)_OleB pET-21d(+)_OleB pCOLADuet-1_OleB pCDFDuet-1_OleB pETDuet-1_OleB pACYDuet-1_OleB	vectores anteriores cortados con NdeI y XhoI	Producto de PCR derivado de los cebadores LB151 y LB152 cortado con NdeI y XhoI	Descrito en el presente documento
OleA y OleB de <i>S. maltophilia</i> : pET-21b(+)_OleAB pET-21d(+)_OleAB pCOLADuet-1_OleAB pCDFDuet-1_OleAB pETDuet-1_OleAB pACYDuet-1_OleAB	vectores anteriores cortados con NdeI y XhoI	Producto de PCR derivado de los cebadores LB118 y LB152 cortado con NdeI y XhoI	Descrito en el presente documento
OleC y OleD de <i>S. maltophilia</i> : pET-21b(+)_OleCD pET-21d(+)_OleCD pCOLADuet-1_OleCD	vectores anteriores cortados con NdeI y XhoI	Producto de PCR derivado de los cebadores LB155 y LB158 cortado con NdeI y XhoI	Descrito en el presente documento

pCDFDuet-1_OleCD pETDuet-1_OleCD pACYDuet-1_OleCD			
OleA con codones optimizados (CO) para la expresión en <i>E. coli</i> , secuencia de aminoácidos basada en <i>S. maltophilia</i> : pET-21d_OleA(CO) pCOLADuet-1_OleA(CO)	pET-21d, o pCOLADuet-1 cortado con NcoI y HindIII	Producto de PCR derivado de los cebadores LF305 y LF306 cortado con PciI y HindIII	Descrito en el presente documento
OleA <i>Xanthomonas axonopodis</i> pET-21d_OleA(Xan) pCOLADuet-1_OleA(Xan)	vectores anteriores cortados con NcoI y HindIII	Producto de PCR derivado de los cebadores LF307 y LF308 cortado con PciI y HindIII	Descrito en el presente documento
OleA <i>Chloroflexus aggregans</i> pET-21d_OleA(Chl) pCOLADuet-1_OleA(Chl)	vectores anteriores cortados con NcoI y HindIII	Producto de PCR derivado de los cebadores LF314 y LF315 cortado con PciI y HindIII	Descrito en el presente documento
OleC <i>Chloroflexus aggregans</i> pET-21d_OleC(chl) pCOLADuet-1_OleC(chl)	vectores anteriores cortados con NcoI y HindIII	fragmento de restricción de ADN sintético cortado con NcoI y HindIII que codifica para la proteína OleC de <i>Chloroflexus aggregans</i>	Descrito en el presente documento
OleC <i>Plesiocystis pacifica</i> pET-21d_OleC(Ple) pCOLADuet-1_OleC(Ple)	vectores anteriores cortados con NcoI y HindIII	fragmento de restricción de ADN sintético cortado con NcoI y HindIII que codifica para la proteína OleC de <i>Plesiocystis pacifica</i>	Descrito en el presente documento
OleC <i>Xanthomonas axonopodis</i> pET-21d_OleC(Xan) pCOLADuet-1_OleC(Xan)	vectores anteriores cortados con NcoI y HindIII	fragmento de restricción de ADN sintético cortado con NcoI y HindIII que codifica para la proteína OleC de <i>Xanthomonas axonopodis</i>	Descrito en el presente documento
OleD <i>Chloroflexus aggregans</i> pET-21d_OleD(Chl) pCOLADuet-1_OleD(chl)	vectores anteriores cortados con NcoI y HindIII	fragmento de restricción de ADN sintético cortado con NcoI y HindIII que codifica para la proteína OleD de <i>Chloroflexus aggregans</i>	
OleD <i>Xanthomonas axonopodis</i> pET-21d_OleD(Xan) pCOLADuet-1_OleD(Xan)	vectores anteriores cortados con NcoI y HindIII	fragmento de restricción de ADN sintético cortado con NcoI y HindIII que codifica para la proteína OleD de <i>Xanthomonas axonopodis</i>	Descrito en el presente documento
pCDFDuet-1_FadD y pETDuet-1_FadD: FadD una acil-CoA sintasa de la cepa de <i>E. coli</i> K12 sucepta W3110, n.º de registro de Genbank BAA15609	pCDFDuet-1 o pETDuet-1 cortado con NcoI y HindIII	Producto de PCR de <i>E. coli</i> derivado de los cebadores fadD_F y fadD_R cortado con NcoI y HindIII	Descrito en el presente documento
pETDuet-1_’TesA:	vector pETDuet-1 cortado	Producto de PCR de <i>E.</i>	Descrito en el presente

gen de TesA (gen de tioesterasa A, n.º de registro de Genbank AAA24664 sin secuencia líder (Cho y Cronan, <i>J. Biol. Chem.</i> , 270:4216-9 (1995)) a partir de <i>E. coli</i>)	con NdeI y AvrII	<i>coli</i> derivado de los cebadores TesA_F y TesA_R cortado con NdeI y AvrII	documento
pESC-HIS_OleA: OleA de <i>S. maltophilia</i>	pESC-HIS cortado con ApaI y XhoI	Producto de PCR derivado de los cebadores LB217 y LF304 y cortado con ApaI y XhoI	Descrito en el presente documento
pCL1920pTrc_OleA: OleA de <i>S. maltophilia</i>	pCL1920-pTrc cortado con NcoI y XhoI	Producto de PCR derivado de los cebadores LB189 y LF304 cortado con PciI y XhoI	Descrito en el presente documento
pHT01_OleA: OleA de <i>S. maltophilia</i> con codones optimizados para la expresión en <i>E. coli</i>	pHT01 cortado con BamHI y XbaI	Producto de PCR derivado de los cebadores LB388 y LB386 cortado con BamHI y XbaI	Descrito en el presente documento
Uc FatB1	vector pMAL-c2X	Producto de PCR derivado de los cebadores UcFatB1_F y UcFatB1_R	Descrito en el presente documento
Cc FatB	vector pETDuet-1 cortado con NdeI y AvrII	Producto de PCR derivado de los cebadores CcFatB_F y CcFatB_R cortado con NdeI y AvrII	Descrito en el presente documento
ChFatB2	vector pMAL-c2X	Producto de PCR derivado de los cebadores ChFatB2_F y ChFatB2_R	Descrito en el presente documento

Métodos de clonación: Se usaron procedimientos de clonación de biología molecular convencionales para clonar ADN en los vectores descritos en la tabla 15 (véase, por ejemplo, Sambrook *et al.*, citado anteriormente). Las enzimas de restricción AatII, AflII, AvrII, BamHI, BglII, HindIII, NcoI, NdeI, NheI, NruI, PciI, ScaI, SfoI, SpeI, XbaI, XhoI, y ZraI se adquirieron de New England Biolabs (Ipswich, MA 01938).

- 5 Protocolo de expresión: Se realizaron diversas escalas de fermentaciones para someter a prueba la expresión o producción de producto. Estos protocolos se describen a continuación y se les hace referencia en cada ejemplo como o bien la fermentación de 5 ml, la fermentación de 25 ml o bien se describe una técnica de fermentación alternativa.

10 Fermentación de 5 ml: Estas fermentaciones se llevaron a cabo en tubos de ensayo de 15 ml con 5 ml de caldo de Luria Miller (LB) (EMD, Chemicals, Inc., San Diego, CA). Se hicieron crecer los cultivos hasta una DO_{600} de entre 0,1 y 1 y se indujeron con IPTG a una concentración final de 1 mM. Se extrajeron los cultivos en cualquier momento desde las 6 hasta 48 horas tras la inducción dependiendo del experimento. Se incubaron las fermentaciones a 25°C, 30°C o 37°C con agitación.

15 Fermentación de 25 ml: Se llevaron a cabo estas fermentaciones en frascos de 125 ml con un volumen final de 25 ml de medio. Se prepararon cultivos simiente inoculando 5 ml de medio LB en un tubo de ensayo de 15 ml con células de un raspado de una reserva de congelador. Se prepararon las reservas de congelador añadiendo glicerol hasta una concentración final del 20% en medio LB y almacenando los cultivos a -80°C. Se incubó la simiente con agitación a 37°C hasta que la DO_{600} del cultivo alcanzó entre 0,15 y 0,6. Entonces se usaron los cultivos para inocular la fermentación. Se inocularon 22 ml de medio M9 (Na_2HPO_4 6 g/l, KH_2PO_4 3 g/l, NaCl 0,5 g/l, NH_4Cl 1 g/l, tiamina 1 mg/l, $MgSO_4$ 1 mM, $CaCl_2$ 0,1 mM) con el 2% de glucosa con 2 ml de medio LB con una densidad celular de 0,15 unidades de DO_{600} . Entonces se incubaron los cultivos con agitación hasta que la DO_{600} era de entre 0,3 y 0,6 unidades, momento en el que se indujeron los cultivos con IPTG hasta una concentración final de 1 mM. En experimentos en los que se alimentaron ácidos grasos, se añadieron en el mismo punto en el tiempo cuando se indujeron los cultivos. Se extrajeron los cultivos en cualquier momento desde las 2 hasta 48 horas tras la inducción, dependiendo del experimento. Se incubaron las fermentaciones a 25°C, 30°C o 37°C con agitación.

Protocolo de lisado celular: Se usaron protocolos de lisis celular convencionales. En resumen, se rompieron las

células mediante sonicación y/o mediante el uso del kit de reactivo de nucleasa Benzonase® más BugBuster® (n.º de catálogo 70750 Novagen de EMD Chemicals, Inc., San Diego, CA). Por ejemplo, se centrifugó un cultivo de 10 ml a 3500 rpm durante 15 minutos (centrífuga Allegra X-15R con rotor SX-4750A, Beckman Coulter, Fullerton, CA), y se resuspendió el sedimento resultante en 2 ml de BugBuster® y 2 µl de nucleasa Benzonase®.

- 5 Protocolo de purificación de proteínas: Se purificaron proteínas etiquetadas con His usando procedimientos convencionales. Se purificaron las proteínas según las instrucciones encontradas en el protocolo de usuario TB054 Rev. F0106 (Novagen de EMD Chemicals, Inc., San Diego, CA).

Métodos de extracción de hidrocarburos:

Método de extracción 1: Extracciones de sedimentos de 5 ml

- 10 Se extrajeron compuestos orgánicos (olefinas, cetonas alifáticas e hidrocarburos) de sedimentos de células bacterianas usando un protocolo de extracción con metanol:hexano. En resumen, se centrifugaron 5 ml de caldo de cultivo/fermentación en un tubo de ensayo de vidrio a 3500 rpm durante 15 minutos (centrífuga Allegra X-15R con rotor SX-4750A, Beckman Coulter, Fullerton, CA), se decantó el sobrenadante, se resuspendió el sedimento resultante en 100 µl de agua destilada estéril, y se mezcló en una mezcladora de vórtex hasta que era homogénea.
- 15 A continuación, se añadió 1 ml de metanol y se mezcló la muestra usando una mezcladora de vórtex. Entonces se sonicó la muestra entre 15 minutos y 1,5 horas en un baño de agua de sonicación (Branson®[®], Tabletop Ultrasonic Cleaners, modelo 5510, Danbury, CT). Tras la sonicación, se añadieron 4 ml de hexano, y se mezcló la muestra en una mezcladora de vórtex. Entonces se centrifugaron las muestras a 3500 rpm durante 15 minutos (centrífuga Allegra X-15R con rotor SX-4750A, Beckman Coulter, Fullerton, CA). Se retiró la fase superior (fase de hexano) y se
- 20 añadió a un tubo de vidrio limpio. Entonces se secó la muestra a vacío en una centrífuga (Vacufuge 5301, Eppendorf, Westbury, NY) durante aproximadamente 30 minutos hasta que no estaba presente esencialmente disolvente. Entonces se resuspendió la muestra en 100 µl de acetato de etilo o cloroformo. A continuación, se analizó 1 µl de la muestra en la CG/EM según los métodos de detección descritos más adelante.

Método de extracción 2: Método para ensayos *in vitro*

- 25 Para ensayos *in vitro*, se extrajeron compuestos orgánicos (olefinas, cetonas alifáticas e hidrocarburos) de muestras *in vitro* usando un protocolo de extracción de acetato de etilo/ácido acético al 1%. Se extrajeron muestras de ensayo *in vitro* mediante la adición de 500 µl de acetato de etilo que contenía ácido acético al 1%. Se mezcló la muestra en una mezcladora de vórtex, seguido por centrifugación a 3500 rpm durante 5 minutos para separar las fases acuosa y orgánica (centrífuga 5424 con rotor FA-45-24-11, Eppendorf, Westbury, NY). Se transfirió la fase superior (fase de acetato de etilo) a un tubo limpio. Entonces se secó la muestra a vacío en una centrífuga (Vacufuge 5301, Eppendorf, Westbury, NY) durante aproximadamente 30 minutos hasta que no estaba presente esencialmente disolvente. Se resuspendió la muestra en 50 µl de acetato de etilo y se analizó mediante CG/EM y/o CG/FID. Se analizaron entre 1 y 10 µl en la CG/EM o CG/FID para determinar el contenido de hidrocarburos según los métodos de detección descritos más adelante.

- 35 Método de extracción 3: Extracciones de cultivo de células completas a pequeña escala

- Se añadieron entre 0,5 y 1 ml del cultivo de fermentación a un tubo Eppendorf de 1,7 ml. Entonces se extrajo este cultivo con entre 0,25 y 0,5 ml de acetato de etilo que contenía la adición conocida de hidrocarburos apropiada (por ejemplo, cis-9-tricoseno 10 mg/l o hexacosano 10 mg/l para olefinas, 14-heptacosanona 10 mg/l para cetonas alifáticas, etc.). Se mezcló la mezcla de cultivo-acetato de etilo en una mezcladora de vórtex a alta velocidad durante
- 40 10 minutos. Entonces se centrifugó la muestra a 13.000 rpm en una centrífuga Eppendorf durante 5 minutos (centrífuga 5424 con rotor FA-45-24-11, Eppendorf, Westbury, NY). Se retiró la fase orgánica (fase superior) para su análisis usando los métodos de detección de CG/EM o CG/FID descritos más adelante.

Métodos de detección de hidrocarburos:

Método de detección 1: CG/EM de 20 minutos

- 45 Para la detección por CG/EM, se observaron hidrocarburos y cetonas alifáticas y se verificaron usando el siguiente protocolo:

Tiempo de ejecución: 20 minutos

Columna: HP-5-MS (5% de difenilsiloxano, 95% de dimetilsiloxano) n.º de parte 19091S-433E, longitud: (metros) 30, diámetro interno: orificio estrecho de (mm) 0,25, película: (µM) 0,25

- 50 Intervalo de barrido de MSD: 50 - 800 M/Z

Inyector: entrada Agilent 6850 de 1 µl

Entrada: 300°C sin desdoblamiento

ES 2 646 815 T3

Gas portador: Helio

Temp. del horno: 5 minutos mantenida a 100°C; 25°C/minuto hasta 320°C; 5 minutos mantenida a 320°C

Det.: Agilent 5975B VL MSD

Temp. de det.: 300°C

5 Método de detección 2: CG/EM de 5 minutos

Para la detección por CG/EM, se detectaron hidrocarburos y cetonas alifáticas y se verificaron usando el siguiente protocolo:

Tiempo de ejecución: 4,9 minutos

10 Columna: HP-5-MS (5% de difenilsiloxano, 95% de dimetilsiloxano) n.º de parte 19091S-433E, longitud: (metros) 30, diámetro interno: orificio estrecho de (mm) 0,25, película: (μM) 0,25

Intervalo de barrido de MSD: 50 - 140 M/Z

Inyector: entrada Agilent 6850 de 1 μl

Entrada: 300°C sin desdoblamiento

Gas portador: Helio

15 Temp. del horno: no mantenida a 225°C; 25°C/minuto hasta 320°C; 1,1 minutos mantenida a 320°C

Det: Agilent 5975B VL MSD

Temp. de det.: 300°C

Método de detección 3: CG/FID

20 Para la detección por CG/FID, se detectaron hidrocarburos y cetonas alifáticas y se verificaron usando el siguiente protocolo:

Tiempo de ejecución: de 2,75 a 4 minutos dependiendo de la longitud de los hidrocarburos que interesa observar. Se usaron tiempos más largos para hidrocarburos más largos.

25 Columna: Thermo Electron Corporation: (n.º de parte UFMC 002 000 000 00): Ph5, grosor de película 0,4 mm, longitud: (metros) 5, diámetro interno: 0,1 mm, volumen de muestra inyectado: de 1 a 10 μl dependiendo del experimento

Entrada: sin desdoblamiento

Gas portador: Helio

Temp. del horno: 1 min mantenida a 100°C; 200°C/min a 350°C; 0,50 min mantenida a 350°C

Det.: detector Thermo FID

30 Método de detección 4: CG/EM de 5,8 minutos

Para la detección por CG/EM, se observaron hidrocarburos y cetonas alifáticas y se verificaron usando el siguiente protocolo:

Tiempo de ejecución: 5,83 minutos

35 Columna: DB-1HT (100% de dimetilpolisiloxano) n.º de parte 122-1111 (Agilent Technologies, Inc), longitud: (metros) 15, diámetro interno: orificio estrecho de (mm) 0,25, película: (μM) 0,1

Intervalo de barrido de MSD: 50 - 550 m/z

Inyector: entrada Agilent 6890 N de 1 μl

Entrada: 300°C sin desdoblamiento

Gas portador: Helio (velocidad de flujo: 1,3 ml/min)

40 Temp. del horno: 0,3 minuto mantenida a 160°C; 30°C/minuto a 320°C; 0,2 minuto mantenida a 320°C

Det: Agilent 5975B XL EI/CI MSD

Temp. de det.: 250°C

Construcción de plásmidos: Se construyeron varios plásmidos para la expresión de *oleA*, *oleB*, *oleC* y *oleD* en cualquier célula de interés, incluyendo *E. coli*, *B. subtilis*, *B. megaterium* y *S. cerevisiae*. Estos constructos pueden dividirse en plásmidos que contienen genes individuales y plásmidos que contienen combinaciones de genes en operones.

Plásmidos que contienen genes individuales *oleA*, *oleB*, *oleC* y *oleD*: Se derivaron las secuencias génicas de *oleA*, *oleB*, *oleC* y *oleD* o bien mediante amplificación por PCR a partir de ADN genómico de una cepa huésped que contiene estas secuencias génicas o bien mediante el diseño y la producción de genes sintéticos con un contrato (DNA2.0 Inc., Menlo Park, CA 94025). En resumen, se amplificaron *oleA*, *oleB*, *oleC* y *oleD* a partir de ADN genómico aislado de *S. maltophilia* ATCC 17679 tal como sigue: se amplificó *oleA* (SEQ ID NO: 1) usando los cebadores LB118 (SEQ ID NO: 29) y LB119 (SEQ ID NO: 30); se amplificó *oleB* (SEQ ID NO: 9) usando los cebadores LB151 (SEQ ID NO: 35) y LB152 (SEQ ID NO: 36); se amplificó *oleC* (SEQ ID NO: 5) usando los cebadores LB155 (SEQ ID NO: 31) y LB159 (SEQ ID NO: 32); se amplificó *oleD* (SEQ ID NO: 7) usando los cebadores LB157 (SEQ ID NO: 33) y LB158 (SEQ ID NO: 34). Se insertaron los productos de amplificación de *oleA* (SEQ ID NO: 1), *oleB* (SEQ ID NO: 9), *oleC* (SEQ ID NO: 5) y *oleD* (SEQ ID NO: 7) en pET-21b(+), pCDFDuet-1, pCOLADuet-1, pETDuet-1 o pACYCDuet-1 usando las enzimas de restricción NdeI y XhoI. Se seleccionaron los clones de plásmido correcto a partir de los transformantes resultantes y se confirmaron mediante digestión y secuenciación de ADN. Pueden usarse métodos de clonación similares para insertar *oleA* (SEQ ID NO: 1), *oleB* (SEQ ID NO: 9), *oleC* (SEQ ID NO: 5) y *oleD* (SEQ ID NO: 7) en cualquier vector de interés.

Se construyó el plásmido pCL1920pTrcOleA clonando la secuencia de nucleótidos de *oleA* a partir del plásmido pET21b(+)_OleA que contiene la secuencia de ADN de *oleA* nativa de *S. maltophilia* ATCC17679 en el vector de expresión bacteriano pCL1920pTrc usando técnicas de clonación convencionales tal como se describió anteriormente. Se construyó la estructura principal de vector pCL1920pTrc (SEQ ID NO: 24) en LS9, Inc. En resumen, se amplificó un producto de PCR que contenía la secuencia de lacIq, el promotor pTrc, y un sitio de clonación múltiple usando LF302 (SEQ ID NO: 76) y LF303 (SEQ ID NO: 77) a partir del plásmido pTrcHis2A (Invitrogen, Carlsbad, CA) ADN. Se dirigió el producto de PCR resultante con AflIII y ZraI y se clonó mediante ligamiento de ADN convencional en el plásmido pCL1920 (NCCBNr3176 *Nuecl. Acids Res.*, 18: 4631, (1990)) cortado con AflIII y SfoI. Los clones resultantes contenían el inserto en la orientación incorrecta y el sitio de restricción AflIII no se conservó. Se designó el plásmido resultante pCL1920pTrc (SEQ ID NO: 24).

Para construir el plásmido pCL1920pTrcOleA, se amplificó por PCR *oleA* (SEQ ID NO: 1) usando los cebadores LB189 (SEQ ID NO: 45) y LF304 (SEQ ID NO: 44). Se clonó el producto de amplificación de OleA en pCL1920pTrc usando las enzimas de restricción PciI y XhoI para crear el plásmido pCL1920pTrcOleA.

Para construir el plásmido pHT01_OleA para la expresión de *oleA* en *B. subtilis*, se amplificó por PCR *oleA* (SEQ ID NO: 3) usando los cebadores LB388 (SEQ ID NO: 491) y LB386 (SEQ ID NO: 490) a partir del molde de ADN pCL1920pTrcOleCDAB, que contiene la secuencia de nucleótidos basada en la secuencia de aminoácidos de *S. maltophilia*, cuyos codones se han optimizado para la expresión en *E. coli*. Se clonó el producto de amplificación de OleA en pHT01 usando las enzimas de restricción BamHI y XbaI para crear el plásmido pHT01_OleA.

Plásmidos que contienen el operón de olefina sintasa:

Se construyeron múltiples plásmidos que contenían los cuatro genes: *oleA*, *oleB*, *oleC* y *oleD* en diversos órdenes. Todas las fermentaciones sometidas a prueba con cepas que expresan los genes a partir de estos operones produjeron olefinas. El orden de los genes en un operón sintético no influye en la capacidad del organismo para producir hidrocarburos. Muchos de estos plásmidos pueden usarse de manera intercambiable en los ejemplos descritos a continuación. Los plásmidos reales usados en cada ejemplo se describen en el ejemplo pertinente.

En plásmidos que contienen genes *ole* basados en las secuencias de aminoácidos de *S. maltophilia*, los plásmidos se nombraron dependiendo del orden de los genes en la secuencia de ADN, por ejemplo, con OleA, OleB, OleC, OleD, OleAB, OleBC, OleCD, OleABCD, etc.

Cuando el gen *ole* es de un genoma de otro organismo, el nombre del gen va seguido por paréntesis con las primeras tres letras del organismo del que se derivó el gen. Por ejemplo, OleA(Xan) indica la secuencia génica que codifica para la secuencia de aminoácidos de OleA encontrada en la secuencia del genoma de *Xanthomonas axonopodis*.

Se creó un operón sintético que contenía los genes *oleA*, *oleB*, *oleC* y *oleD* que codifican para las secuencias de proteínas basadas en *S. maltophilia*. Se optimizaron los codones de los genes *ole* para la expresión en *E. coli* mediante modificación del programa en DNA2.0, Inc. (Menlo Park, CA 94025).

Se usó la secuencia génica sintetizada (SEQ ID NO: 23) para construir el plásmido pCL1920pTrcOleABpTrcOleCD. Se subclonó el operón de olefina sintético (SEQ ID NO: 23) proporcionado por DNA2.0 en el plásmido pCL1920pTrc.

Se aisló el fragmento de operón digiriendo con PciI y HindIII, y se ligó en el plásmido pCL1920pTrc cortado con NcoI y HindIII. Se designó el clon resultante pCL1920pTrcOleABpTrcLOleCD (SEQ ID NO: 78).

Se construyó otro plásmido, pCL1920pTrcOleCDAB, que contenía los cuatro genes de olefina sintasa (*oleA*, *oleB*, *oleC* y *oleD*) en un orden diferente y se usó en algunos de los siguientes ejemplos. Se construyó este plásmido tal como sigue: en primer lugar, se construyó el plásmido pCL1920pTrcOleCD clonando un producto de PCR que contenía la secuencia de *lacI^q* en la estructura principal de vector digerida con NheI y SpeI de pCL1920pTrcOleABpTrcLOleCD. El plásmido resultante contiene la secuencia de *lacI^q* en el sentido de 5' de los genes *oleC* y *oleD*. Se amplificó el producto de PCR que contiene el represor lac usando LB270 (SEQ ID NO: 79) y LB271 (SEQ ID NO: 80) de pCL1920pTrc, se prepararon el producto de PCR y el vector para clonación mediante digestión con NheI y SpeI.

En segundo lugar, se construyó entonces el plásmido de operón, pCL1920pTrcOleCDAB, clonando un fragmento de PCR que contenía *oleA* y *oleB* en el sentido de 3' de *oleC* y *oleD* en el plásmido pCL1920pTrcOleCD. Se amplificó el producto de PCR de *oleAB* usando LB272 (SEQ ID NO: 81) y LB275 (SEQ ID NO: 82) a partir de pCL1920pTrcOleABpTrcLOleCD. Se digirió el producto de PCR resultante con BamHI y BglII y se clonó mediante ligamiento de ADN convencional en el pCL1920pTrcOleCD digerido con BamHI y BglII y tratado con fosfatasa Antarctic (New England Biolabs, Ipswich, MA 01938). El plásmido resultante se designó pCL1920pTrcOleCDAB. Pueden usarse métodos similares para insertar un operón de olefina sintético en cualquier vector de interés.

Para construir el plásmido pWH1520_OleCDAB para la expresión en *B. megaterium*, se amplificaron por PCR *oleC*, *oleD*, *oleA* y *oleB* usando los cebadores LB402 (SEQ ID NO: 488) y LB387 (SEQ ID NO: 489) a partir del molde de ADN pCL1920pTrcOleCDAB que contiene la secuencia de nucleótidos basada en la secuencia de aminoácidos de *S. maltophilia* cuyos codones se han optimizado para la expresión en *E. coli*. El producto de PCR que contiene *oleC*, *oleD*, *oleA* y *oleB* se digirió con SpeI y BglII y se clonó en pWH1520 usando las enzimas de restricción SpeI y BamHI para crear el plásmido pWH1520_OleCDAB.

Plásmidos que contienen genes de tioesterasa: Se construyeron múltiples plásmidos que contenían la región codificante para diferentes tioesterasas, incluyendo (Uc FatB1, Ce FatB1, Ch FatB2 y *tesA*). Se derivó la secuencia génica para *tesA* de *E. coli* mediante amplificación por PCR a partir de ADN genómico de *E. coli* MG1655. Las secuencias génicas para FatB1 de *Umbellularia californica* (Uc FatB1), FatB1 de *Cinnamomum camphora* (Cc FatB1) y FatB2 de *Cuphea hookeriana* (Ch FatB2) se derivaron mediante diseño y producción de genes sintéticos con un contrato (Codon Devices, Cambridge, MA 02139). En resumen, se amplificó *tesA* (SEQ ID NO: 25) usando los cebadores TesA_F (SEQ ID NO: 50) y TesA_R (SEQ ID NO: 51). Volvieron a clonarse los genes sintéticos en plásmidos de expresión mediante amplificación por PCR y protocolos de digestión y ligamiento convencionales. Se amplificó Cc FatB1 (SEQ ID NO: 26) usando los cebadores CcFatB_F (SEQ ID NO: 52) y CcFatB_R (SEQ ID NO: 53). Se amplificó Ch FatB2 (SEQ ID NO: 27) usando los cebadores ChFatB2_F (SEQ ID NO: 54) y ChFatB2_R (SEQ ID NO: 55). Se amplificó Uc FatB1 (SEQ ID NO: 28) usando los cebadores UcFatB1_F (SEQ ID NO: 56) y UcFatB1_R (SEQ ID NO: 57). Se clonaron los productos de amplificación de *tesA* (SEQ ID NO: 25) y Cc FatB1 (SEQ ID NO: 26) en pETDuet-1 usando las enzimas de restricción NdeI y AvrII. Se clonaron los productos de amplificación de Ch FatB2 (SEQ ID NO: 27) y Uc FatB1 (SEQ ID NO: 28) en pMAL-c2X (NEB, Ipswich, MA) en marco con el gen *malE* usando las enzimas de restricción XbaI y HindIII.

Plásmidos que contienen la acil-CoA sintasa: Se expresó el gen *fadD* a partir de varios constructos diferentes, incluyendo dos vectores distintos, pETDuet-1 y pACYCpTrc, así como a partir de un promotor constitutivo integrado directamente en la cepa huésped de *E. coli*. Estos constructos se describen a continuación.

pETDuet-1_fadD: Se clonó el gen *fadD* que codifica para acil-CoA sintasa a partir de *E. coli* en un vector pETDuet-1 digerido con NcoI/HindIII. La secuencia de ácido nucleico de *fadD* (SEQ ID NO: 492) se amplificó por PCR a partir de ADN genómico de *E. coli* C41(DE3) usando los cebadores fadD_F (SEQ ID NO: 46) y fadD_R (SEQ ID NO: 47). Se clonó el producto de amplificación de *fadD* en pETDuet-1 usando las enzimas de restricción NcoI y HindIII para crear pETDuet-1_fadD.

pACYCpTrcfadD: Para construir pACYCpTrcfadD, se construyó en primer lugar el plásmido de vector pACYCpTrc. Las regiones de promotor y terminador *lacI^q* y pTrc se amplificaron por PCR usando los cebadores pTrc_F (SEQ ID NO: 493) y pTrc_R (SEQ ID NO: 494) a partir de pTrcHis2A (Invitrogen, Calrsbad, CA). Se digirió entonces el producto de PCR con AatII y NruI. Entonces se clonó el producto de PCR en pACYC177 digerido con AatII y Scal. La secuencia de ácido nucleico de *fadD* se amplificó por PCR usando los cebadores fadD_F (SEQ ID NO: 46) y fadD_R (SEQ ID NO: 47) a partir del ADN genómico de *E. coli* C41(DE3) y se clonó en pACYCpTrc mediante digestión con NcoI y EcoRI para crear el plásmido pACYCpTrcfadD.

Expresión constitutiva de acil-CoA sintasa

Se expresó de manera constitutiva FadD mediante sustitución de la región flanqueante en 5' en el sentido de 5' del gen *fadD* (SEQ ID NO: 495) con una secuencia de ADN sintético que contiene la secuencia de promotor T5 (SEQ ID NO: 496) usando recombinación homóloga (intercambio alélico) seguido por eliminación del marcador de resistencia antimicrobiana Cre lox (véanse, por ejemplo, Valle, Fernando y Noemí Flores, Overexpression of Chromosomal

Genes in *Escherichia coli*. vol. 267, Recombinant Gene Expression: Reviews and Protocols. Totowa: Humana Press, 2004 y Palmeros *et al.*, *Gene*, 247: 255-64 (2000) para protocolos detallados). En resumen, se usaron los cebadores LB284 (SEQ ID NO: 83) y LB285 (SEQ ID NO: 84) para amplificar por PCR el casete loxPcat a partir de pLoxCat2 (SEQ ID NO: 89; n.º de registro de Genebank AJ401047). Entonces se transformó el producto de amplificación en MG1655 $\Delta fadE$ que contenía el plásmido pKD46 (n.º de registro de Genebank AY048746). Tras la selección para detectar la integración del casete loxPcat, se curó la cepa de pKD46, seguido por la eliminación de la resistencia a cloramfenicol usando el plásmido pJW168 (Palmeros *et al.*, *Gene*, 247: 255-64 (2000)). Tras la confirmación del reemplazo del promotor de *fadD* por el promotor de T5, se curó la cepa de pJW168. La cepa resultante es *E. coli* MG1655 $\Delta fadE$, *fadD*(+).

10 EJEMPLO 2

Este ejemplo demuestra que la expresión de OleA, OleC y OleD en bacterias da como resultado la producción de olefinas.

Se amplificaron las secuencias de ácido nucleico de *oleA*, *oleC* y *oleD* de *S. maltophilia* usando PCR. Se insertaron estas secuencias en vectores de expresión bacteriana usando técnicas de clonación convencionales, tal como se describe en el ejemplo 1. Se usaron entonces los plásmidos que contenían *oleA*, *oleC* y *oleD* para transformar *E. coli* C41(DE3) $\Delta fadE$. La cepa bacteriana resultante con el genotipo *E. coli* C41(DE3) $\Delta fadE$; pET-21b(+)-*OleA*, pCOLADuet-1-*OleC*, pCDFDuet-1-*OleD* se sometió a prueba usando el protocolo de fermentación de 5 ml y se extrajo usando el método de extracción 1. Entonces se analizó el extracto mediante CG/EM usando el método de detección 1, tal como se describe en el ejemplo 1, para la detección de hidrocarburos (por ejemplo, olefinas) y cetonas alifáticas. Los hidrocarburos observados mediante CG/EM eran olefinas mono-, di- y triinsaturadas que oscilaban en la longitud de la cadena de carbono entre C₂₇ y C₃₁ (véanse las figuras 1 y 2A-J). Esto demostró que la expresión de *oleA*, *oleC* y *oleD* en *E. coli* daba como resultado la producción de olefinas. Pueden usarse métodos similares para expresar *oleA*, *oleC* y *oleD* en cualquier célula de interés. Además, pueden usarse métodos similares para analizar la producción de cetonas alifáticas o hidrocarburos.

Para demostrar que la expresión de *oleA*, *oleB*, *oleC* y *oleD* conduce a la producción de hidrocarburos en una amplia gama de bacterias, se preparó un plásmido pWH1520-*OleCDAB* para la expresión de *oleA*, *oleB*, *oleC* y *oleD* en *Bacillus megaterium* (véase el ejemplo 1). Se transformaron protoplastos preparados previamente de *B. megaterium* WH320 (de Mo Bi Tec, Alemania) con pWH1520-*OleCDAB* para crear la cepa *B. megaterium* WH320; pWH1520-*OleCDAB*. Se transformaron los protoplastos preparados previamente de *B. megaterium* WH320 (de Mo Bi Tec, Alemania) con pWH1520 para preparar *B. megaterium* WH320; pWH1520 va a usarse como control de vector vacío.

Las secuencias de ácido nucleico se amplificaron por PCR usando un plásmido que contiene las secuencias con codones optimizados para *E. coli* de *oleC*, *oleD*, *oleA* y *oleB* de *S. maltophilia*. Se insertaron estas secuencias en vectores de expresión bacteriana usando técnicas de clonación convencionales para *E. coli* (véase el ejemplo 1). Se usaron entonces el plásmido que contenía *oleA*, *oleB*, *oleC* y *oleD* (pWH1520-*OleCDAB*) y un plásmido de control de vector vacío (pWH1520) para transformar *B. megaterium* WH320.

Se transformaron protoplastos de *B. megaterium* WH320 (Mo Bi Tec, Alemania) con los plásmidos mencionados anteriormente usando el protocolo proporcionado con los protoplastos (véase, por ejemplo, el manual del sistema de expresión de proteínas en *Bacillus megaterium* de 2007), que es una modificación del método de Puyet *et al.*, *FEMS Microbiol. Lett.*, 40: 1-5. (1987).

Se sometieron a prueba las cepas bacterianas resultantes con el genotipo *B. megaterium* WH320; pWH1520-*OleCDAB* y *B. megaterium* WH320; pWH1520 usando el protocolo de fermentación de 5 ml (se usó tetraciclina 15 mg/l para seleccionar células transformadas con el plásmido) que se modificó para usar xilosa al 0,5% para inducir los cultivos en lugar de IPTG. El método de extracción usado fue el método de extracción 1.

Se realizó el análisis para determinar la presencia de hidrocarburos mediante análisis por CG/EM usando el método de detección 4 tal como se describe en el ejemplo 1. Los datos de CG/EM fueron el ión extraído para distintos iones parentales 350, 364, 378, 392 y 406, que muestran la producción de olefinas monoinsaturadas que contenían 25, 26, 27, 28 y 29 carbonos, respectivamente (véanse las figuras 17A-J).

Estos compuestos se eluyen ligeramente más pronto que la olefina de cadena lineal del mismo número de carbonos analizada en las mismas condiciones. Esto sugiere que *B. megaterium* producía olefinas de cadena ramificada. Debido a que *B. megaterium* producía de manera natural sustratos de acilo graso de cadena ramificada, es probable que las olefinas observadas sean olefinas de cadena ramificada. No se detectaron hidrocarburos en los extractos de control a partir de *B. megaterium* WH320; pWH1520.

Los resultados de los experimentos reflejados en este ejemplo demuestran que pueden modificarse por ingeniería genética una variedad de bacterias para expresar *oleA*, *oleB*, *oleC* y *oleD* y producir hidrocarburos. Además, este ejemplo demuestra que pueden modificarse por ingeniería genética bacterias Gram-positivas para expresar *oleA*, *oleB*, *oleC* y *oleD* y producir hidrocarburos. Además, este ejemplo demuestra que las bacterias del género *Bacillus*

pueden modificarse por ingeniería genética para expresar *oleA*, *oleB*, *oleC* y *oleD* y producir hidrocarburos. Además, este ejemplo demuestra que los tipos de cadenas de acilo graso producidas de manera natural por el organismo huésped influyen en los tipos de hidrocarburos producidos.

5 También se extrajeron cetonas alifáticas e hidrocarburos (por ejemplo, olefinas) de los sedimentos de células bacterianas de *S. maltophilia* y se analizaron usando CG/EM usando el método de detección 1. Las olefinas producidas por *S. maltophilia* (véanse las figuras 3, 4A y 4C) difieren de las producidas por *E. coli oleA*, *oleC* y *oleD* (véase la figura 1). Por tanto, las estructuras de olefinas (por ejemplo, el grado de saturación, longitud de cadena y presencia de cadenas ramificadas o no ramificadas) dependen del huésped. Estas diferencias son un reflejo directo de la capacidad del organismo para producir diversos tipos de cadenas de acilo graso. Esto demuestra que si se altera la ruta de biosíntesis de ácidos grasos, se producen diferentes tipos de olefinas.

10 Para determinar si se requiere *oleA* para la producción de olefinas en *S. maltophilia*, se generó una cepa deficiente en *oleA* de *S. maltophilia* usando un método basado en vector suicida de recombinación homóloga (intercambio alélico) que conduce a deleciones de genes. En resumen, la deleción de la región del gen *oleA* se produjo usando los métodos de reemplazo de genes descritos por Hoang *et al.*, *Gene*, 212: 77-86, (1998). Más específicamente, se amplificaron las regiones en el sentido de 5' y en el sentido de 3' de la región del gen *oleA* y se colocaron juntas usando la técnica de SOE PCR (véase, por ejemplo, Horton *et al.*, *Gene*, 77: 61-8 (1989)). Se cortó el producto de PCR de SOE con XbaI y HindIII y se clonó en un vector suicida pEX18Tc (véase Hoang *et al.*, citado anteriormente) digerido con XbaI y HindIII. La región flanqueante en 5' usó los cebadores de PCR LB200 (SEQ ID NO: 58) y LB203 (SEQ ID NO: 61). La región flanqueante en 3' usó los cebadores de PCR LB201 (SEQ ID NO: 59) y LB205 (SEQ ID NO: 63). El producto de deleción de PCR combinado (SEQ ID NO: 75), que se clonó en pEX18Tc, se produjo mediante amplificación por PCR de la combinación de los productos de PCR en 5' y 3' usando los cebadores de PCR LB202 (SEQ ID NO: 60) y LB204 (SEQ ID NO: 62).

15 Se analizaron las extracciones a partir de *S. maltophilia* silvestre y *S. maltophilia* $\square\Delta oleA$ mutante (*S. maltophilia* que carece de *oleA*) mediante CG/EM, usando el método de detección 2. Se observaron olefinas en la *S. maltophilia* silvestre (véase la figura 4A). No se observaron olefinas en *S. maltophilia* $\square\Delta oleA$ (véase la figura 4B). Estos resultados demuestran que la deleción del gen *oleA* dio como resultado la pérdida de la producción de olefinas en *S. maltophilia*.

20 Para determinar si se requiere *oleC* para la producción de olefinas en *S. maltophilia*, se generó una cepa deficiente en *oleC* de *S. maltophilia* usando un método basado en vector suicida de recombinación homóloga (intercambio alélico) que conduce a deleciones de genes. En resumen, la deleción de la región del gen *oleC* se produjo usando los métodos de reemplazo de genes descritos por Hoang *et al.*, *Gene*, 212: 77-86, (1998). Más específicamente, se amplificaron las regiones en el sentido de 5' y en el sentido de 3' de la región del gen *oleC* y se colocaron juntas usando la técnica de SOE PCR (véase, por ejemplo, Horton *et al.*, *Gene*, 77: 61-8 (1989)). Se cortó el producto de PCR de SOE con XbaI y HindIII y se clonó en un vector suicida pEX18Tc (véase Hoang *et al.*, citado anteriormente) digerido con XbaI y HindIII.

25 La región flanqueante en 5' usó los cebadores de PCR LB110 (SEQ ID NO: 465) y LB112 (SEQ ID NO: 467). La región flanqueante en 3' usó los cebadores de PCR LB111 (SEQ ID NO: 466) y LB114 (SEQ ID NO: 469). El producto de deleción de PCR combinado (SEQ ID NO: 471), que se clonó en pEX18Tc, se produjo mediante amplificación por PCR de la combinación de los productos de PCR en 5' y 3' usando los cebadores de PCR LB113 (SEQ ID NO: 468) y LB115 (SEQ ID NO: 470).

30 Se analizaron las extracciones de *S. maltophilia* silvestre y *S. maltophilia* $\Delta oleC$ mutante (*S. maltophilia* que carece de *oleC*) mediante CG/EM, usando el método de detección 2. Se observaron olefinas en la *S. maltophilia* silvestre (véase la figura 4A). No se observaron olefinas en *S. maltophilia* $\Delta oleC$ (véase la figura 4D). Se observaron cetonas alifáticas en la *S. maltophilia* $\Delta oleC$ (véase la figura 4D). Estos resultados demuestran que la deleción del gen *oleC* dio como resultado la pérdida de la producción de olefinas en *S. maltophilia*.

35 Para determinar si se requiere *oleD* para la producción de olefinas en *S. maltophilia*, se generó una cepa deficiente en *oleD* de *S. maltophilia* usando un método basado en vector suicida de recombinación homóloga (intercambio alélico) que conduce a deleciones de genes. En resumen, se produjo la deleción de la región del gen *oleD* usando los métodos de reemplazo de genes descritos por Hoang *et al.*, *Gene*, 212: 77-86, (1998). Más específicamente, se amplificaron las regiones en el sentido de 5' y en el sentido de 3' de la región del gen *oleD* y se colocaron juntas usando la técnica de SOE PCR (véase, por ejemplo, Horton *et al.*, *Gene*, 77: 61-8 (1989)). Se cortó el producto de PCR de SOE con XbaI y HindIII y se clonó en un vector suicida pEX18Tc (véase Hoang *et al.*, citado anteriormente) digerido con XbaI y HindIII.

40 La región flanqueante en 5' usó los cebadores de PCR LB122 (SEQ ID NO: 472) y LB124 (SEQ ID NO: 474). La región flanqueante en 3' usó los cebadores de PCR LB123 (SEQ ID NO: 473) y LB127 (SEQ ID NO: 476). El producto de deleción de PCR combinado (SEQ ID NO: 477), que se clonó en pEX18Tc, se produjo mediante amplificación por PCR de la combinación de los productos de PCR 5' y 3' usando los cebadores de PCR LB125 (SEQ ID NO: 475) y LB127 (SEQ ID NO: 476).

Se analizaron las extracciones a partir de *S. maltophilia* silvestre y *S. maltophilia* $\Delta oleD$ mutante (*S. maltophilia* que carece de *oleD*) mediante CG/EM, usando el método de detección 2. Se observaron olefinas en la *S. maltophilia* silvestre (véase la figura 4C). No se observaron olefinas en *S. maltophilia* $\Delta oleD$. Se observaron cetonas alifáticas en la *S. maltophilia* $\Delta oleD$ (véase la figura 4E). Estos resultados demuestran que la delección del gen *oleD* dio como resultado la pérdida de la producción de olefinas en *S. maltophilia*.

Los resultados de los experimentos reflejados en este ejemplo demuestran que la expresión de *oleA*, *oleC* y *oleD* en bacterias da como resultado la producción de olefinas. Las olefinas producidas pueden diferir en estructura dependiendo del huésped. Adicionalmente, los resultados de los experimentos reflejados en este ejemplo demuestran que se requieren *OleA*, *OleC* y *OleD* para la producción de olefinas en *S. maltophilia*.

10 EJEMPLO 3

Este ejemplo demuestra que la expresión de *oleA* en bacterias daba como resultado la producción de cetonas alifáticas.

Se expresó *oleA* en *E. coli* tal como se describe en los ejemplos 1 y 2. Se usó el plásmido que contenía *oleA* para transformar *E. coli* C41(DE3). La cepa bacteriana resultante con el genotipo *E. coli* C41(DE3); pET-21b(+)_{OleA} se sometió a prueba usando el protocolo de fermentación de 5 ml. Se realizó el análisis para la producción de cetonas alifáticas e hidrocarburos (por ejemplo, olefinas) usando el método de extracción 1 seguido por el análisis de CG/EM, usando el método de detección 1 tal como se describe en el ejemplo 1. Las cetonas alifáticas observadas mediante CG/EM eran cetonas alifáticas saturadas, monoinsaturadas y diinsaturadas. Las cetonas alifáticas oscilaban en la longitud de la cadena de carbono entre C₂₇ y C₃₁ (véanse las figuras 5A-C y 6A-H).

Para demostrar que la expresión de *oleA* conduce a la producción de cetonas alifáticas en una amplia gama de bacterias, se preparó el plásmido pHT01_OleA para la expresión de *oleA* en *Bacillus subtilis* (véase el ejemplo 1). Se transformó *B. subtilis* IHA01 con pHT01_OleA para crear la cepa *B. subtilis* IHA01; se transformó pHT01_OleA. *B. subtilis* IHA01 con pHT01 para preparar *B. subtilis* IHA01; pHT01 que actuó como control.

Se amplificaron las secuencias de ácido nucleico mediante PCR usando un plásmido que contenía la secuencia de codones optimizados para *E. coli* de *oleA* de *S. maltophilia*. Se insertaron estas secuencias en vectores de expresión bacteriana usando técnicas de clonación convencionales para *E. coli* (véase, por ejemplo, el ejemplo 1).

Entonces se usaron el plásmido que contenía *oleA* (pHT01_OleA) y un plásmido de control de vector vacío (pHT01) para transformar *B. subtilis* IHA01. Se transformó *B. subtilis* mediante transformación natural según los protocolos de Spizizen, *Proc. Natl. Acad. Sci. USA*, 44: 1072-1078 (1958).

Se sometieron a prueba las cepas bacterianas resultantes con el genotipo *B. subtilis* IHA01; pHT01_OleA y *B. subtilis* IHA01; pHT01 usando el protocolo de fermentación de 5 ml. El método de extracción usado fue el método de extracción 1. Se realizó el análisis para determinar la presencia de cetonas alifáticas e hidrocarburos (por ejemplo, olefinas) mediante CG/EM usando el método de detección 4 tal como se describe en el ejemplo 1. Se observaron cetonas alifáticas saturadas que contenían 25, 27 y 29 carbonos. Los múltiples picos para cada cetona alifática C25, C27 y C29 respectiva reflejan que estaban presentes varios isómeros ramificados diferentes de las cetonas alifáticas (véanse las figuras 18A-E).

Los resultados de los experimentos reflejados en este ejemplo demuestran que pueden modificarse por ingeniería genética una variedad de bacterias para expresar *oleA* y producir cetonas alifáticas. Además, este ejemplo demuestra que pueden modificarse por ingeniería genética bacterias Gram-positivas para expresar *oleA* y producir cetonas alifáticas. Además, este ejemplo demuestra que pueden modificarse por ingeniería genética bacterias del género *Bacillus* para expresar *oleA* y producir cetonas alifáticas. Además, este ejemplo demuestra que los tipos de cadenas de acilo graso producidas de manera natural por el organismo huésped influyen en los tipos de hidrocarburos producidos.

EJEMPLO 4

Este ejemplo demuestra la capacidad de observar producción de cetonas alifáticas usando un ensayo *in vitro* que combina lisado de células de *E. coli* que expresan *oleA* (SEQ ID NO: 2) con sustratos de acil-CoA.

Se expresó *oleA* en *E. coli* tal como se describe en los ejemplos 1 y 3. Las bacterias recombinantes resultantes se cultivaron, se indujeron, se sedimentaron y se usaron para preparar un lisado celular que contenía *OleA* (lisado celular de *OleA*) tal como se describe en el ejemplo 1. La mezcla de ensayo *in vitro* consistía en 10 μ l de una disolución madre 10 mM de sal de litio de miristoil-CoA (el sustrato) (M4414 de Sigma-Aldrich St. Louis, MO) suspendida en tampón fosfato 0,1 M pH 7,0, 1 μ l de agente reductor de tris(hidroxipropil)fosfina (THP) 500 mM, 50 μ l de lisado celular de *OleA* y 39 μ l de tampón fosfato 0,1 M pH 7,0. También se prepararon muestras de control que contenían combinaciones con y sin sustrato o extracto celular de *OleA*. Se incubaron las muestras a 37°C durante 1 hora. Tras el periodo de incubación, se añadieron 10 μ l de disolución de hexacosano 0,1 mg/ml como adición conocida de control a cada reacción antes de la extracción. Se extrajeron las muestras usando 500 μ l de acetato de etilo que contenía ácido acético al 1%. Se mezcló la mezcla en una mezcladora de vórtex seguido por

centrifugación. Se transfirió la fase superior (fase de acetato de etilo) a un tubo de vidrio limpio y se secó a vacío en una centrifuga (Vacufuge 5301, Eppendorf, Westbury, NY). Se resuspendió la muestra en 50 µl de acetato de etilo. Se analizaron entre 1 y 10 µl mediante o bien CG/EM o bien CG/FID según los métodos descritos en el ejemplo 1. Pueden usarse métodos similares para incubar cualquier lisado celular de OleA con un sustrato en condiciones suficientes para producir cetonas alifáticas y posteriormente analizar la producción de cetonas alifáticas.

Las cetonas alifáticas resultantes observadas mediante CG/EM y CG/FID eran saturadas, monoinsaturadas y diinsaturadas. Las cetonas alifáticas oscilaban en longitud de cadena de carbono entre C₂₇ y C₃₁ (véanse las figuras 7A y 7B). Se comparó la muestra que contenía sustrato y extracto de OleA con todas las otras combinaciones de muestras de control. La producción de una cantidad aumentada de cetona alifática C₂₇ en relación con todas las muestras de control era indicativa de la actividad de condensación de acilos encontrada en OleA (véase la figura 7C).

Los resultados de los experimentos reflejados en este ejemplo demuestran que acil coenzima A es un sustrato para la producción de cetona alifática en un ensayo *in vitro* y que se requiere la expresión de *oleA* para la producción de cetonas alifáticas.

EJEMPLO 5

Este ejemplo demuestra la capacidad para observar cetonas alifáticas usando un ensayo *in vitro* que combina enzima purificada (por ejemplo, proteína OleA) con sustrato purificado (por ejemplo, acil-coenzima A, acil-ACP, etc.) en un tampón fosfato 0,1 M pH 7,0.

Se expresó *oleA* en *E. coli* tal como se describe en los ejemplos 1 y 3. Las bacterias recombinantes resultantes se cultivaron, se indujeron, se sedimentaron y se usaron para preparar proteína OleA purificada tal como se describe en el ejemplo 1. La mezcla de ensayo *in vitro* consistía en un sustrato (por ejemplo, acil-coenzima A, acil-ACP, o una mezcla de acil-coenzima A y acil-ACP) diluido hasta una concentración final de 0,1 mM a 1 mM, 20 µl de una disolución 0,6 mg/ml de proteína OleA purificada en un tampón fosfato 0,1 M pH 7,0 con cloruro de magnesio 500 mM. Se incubó cada mezcla de ensayo a 37°C durante 1 hora. Tras el periodo de incubación, se añadieron 250 µl de acetato de etilo a cada mezcla de ensayo y se mezcló cada mezcla de ensayo en una mezcladora de vórtex durante 10 minutos. Se separó la fracción de acetato de etilo de la mezcla de ensayo de la fase acuosa mediante centrifugación en una microcentrifuga a 13000 rpm durante 5 minutos. Se transfirieron 15 µl de la fracción de acetato de etilo (la fase superior) a un vial de CG/EM al que se le añadieron 1,5 µl de una disolución 0,1 mg/ml de disolución de hexacosano como adición conocida de control a cada fracción de acetato de etilo antes de que se analizara cada fracción de acetato de etilo en la CG/EM usando métodos de detección de ionización química que se conocen bien en la técnica (véase la figura 8A-B).

Se observaron cetonas alifáticas saturadas, monoinsaturadas y diinsaturadas con longitudes de cadena de carbono que oscilan entre C₁₉ y C₃₁. No se observó cetona alifática en ninguna de las mezclas de ensayo de control (por ejemplo, mezclas de ensayo que contienen una enzima purificada sin un sustrato o mezclas de ensayo que contienen un sustrato purificado sin una enzima purificada).

Las tablas 16-18 ilustran las combinaciones de sustratos que se sometieron a prueba y los tipos de cetonas alifáticas que se produjeron. Se formaron cetonas alifáticas mediante las siguientes combinaciones de sustratos: acil-CoA con acil-CoA, acil-ACP con acil-ACP y acil-CoA con acil-ACP (los datos de cada combinación se muestran en las tablas 16, 17 y 18, respectivamente). Los sustratos se muestran a lo largo del lado superior e izquierdo de la tabla y cada entrada muestra la longitud de la cadena de carbono seguido por el número de dobles enlaces para cada cetona alifática que se formó (por ejemplo, C27:1 se refiere a una cetona alifática con 27 carbonos y un único doble enlace, C23 se refiere a una cetona alifática completamente saturada con 23 carbonos, etc.).

TABLA 16

acil-CoA/acil-CoA	C10	C12	C14	C16	C16:1
C10	C19	N/A	C23	N/A	N/A
C12	-	C23	C25		C27:1
C14	-	-	C27	C29	C29:1
C16	-	-	-	C31	C31:1
C16:1	-	-	-	-	C31:2

N/A = no se ha sometido a prueba.

TABLA 17

acil-ACP/acil-ACP	C10	C12	C14	C14:1	C16:1
C10	C19	N/A	C23	N/A	N/A
C12	-	C23	C25	N/A	C27:1
C14	-	-	C27	C27:1	C29:1
C14:1	-	-	-	C27:2	N/A
C16:1	-	-	-	-	C31:2

N/A = no se ha sometido a prueba.

TABLA 18

Sustrato	C14-ACP
C16:1-CoA	C27, C29:1, C31:2

Los resultados de los experimentos reflejados en este ejemplo demuestran que puede utilizarse un ensayo *in vitro* para detectar la actividad de OleA mediante la presencia de cetonas alifáticas. Además, los resultados de los experimentos reflejados en este ejemplo demuestran que pueden usarse tanto acil-coenzima A como acil-ACP como sustratos mediante OleA solo o en combinación para producir cetonas alifáticas. La longitud de cadena y el grado de saturación del sustrato pueden variar para producir una amplia gama de productos de cetona alifática.

EJEMPLO 6

Este ejemplo describe la identificación de secuencias de aminoácidos y ácido nucleico de *oleA*, *oleC* y *oleD* adicionales usando las secuencias de aminoácidos y ácido nucleico *S. maltophilia* descritas en el presente documento.

Las secuencias de aminoácidos de proteínas relacionadas con las secuencias de OleA, OleC y OleD de *S. maltophilia* se determinaron usando el programa de alineación de proteínas NCBI BLAST para realizar búsquedas en la base de datos nr database así como varias de otras bases de datos accesibles públicamente. Para demostrar cómo identificar y someter a prueba la actividad de miembros adicionales de la familia de genes de *oleA*, *oleC* y *oleD*, se clonaron y sometieron a prueba genes de un organismo estrechamente relacionado, *Xanthomonas axonopodis* (*oleA*, *oleC* y *oleD*) y genes de dos organismos relacionados de manera distante, *Chloroflexus aggregans* (*oleA*, *oleC* y *oleD*) y *Plesiocystis pacifica* (*oleC*), tal como sigue.

Los plásmidos usados en este ejemplo se prepararon usando las mismas cepas bacterianas, marcadores de resistencia y técnicas de PCR descritas en el ejemplo 1. Para una descripción más detallada de los plásmidos, véase la tabla 15. De manera similar, se usaron el protocolo de fermentación de 5 ml, el método de extracción de hidrocarburos 1 y el método de detección de hidrocarburos por CG/EM 1 descritos en el ejemplo 1 para identificar hidrocarburos y productos intermedios de hidrocarburos.

Se diseñaron los genes de interés para que estuviesen optimizados para la expresión en *E. coli* usando el software Protein-2-DNA (véase, por ejemplo, Gustafsson *et al.*, *Trends Biotechnol.* 22: 346-353 (2004)) para seleccionar una distribución de codones que imite a las proteínas de *E. coli* altamente expresadas naturales (véase, por ejemplo, Henaut *et al.*, Analysis and predictions from *E. coli* sequences, en *E. coli* and *Salmonella typhimurium* Cellular and Molecular Biology, volumen 2, editado por: Neidhardt *et al.*, Washington DC, ASM press, págs. 2047-2066 (1996)). Se sintetizó/sintetizaron el/los gene(s) mediante PCR sin molde similar a lo que se ha descrito previamente en la bibliografía (véase, por ejemplo, Dillon *et al.*, *Biotechniques*, 9: 298-300 (1990)). Se realizó la síntesis de genes mediante DNA2.0 (Menlo Park, CA).

Se clonaron los marcos de lectura abiertos sintéticos en el plásmido pJ201 (DNA 2.0, Menlo Park, CA). Se subclonaron estos genes en pET21d en el sentido de 5' del promotor T7 en el sitio de clonación múltiple entre los sitios NcoI y HindIII. Se diseñaron los cebadores para que conservaran los extremos 5' y 3' de las secuencias de proteína. Se amplificó por PCR la versión con codones optimizados de *oleA* basándose en la secuencia de aminoácidos de *S. maltophilia* (SEQ ID NO: 3) a partir del plásmido de DNA 2.0 usando los cebadores LF305 (SEQ ID NO: 37) y LF306 (SEQ ID NO: 38) (véase la tabla 14 para secuencias cebador). Se digirió el producto de PCR con PciI y HindIII y se clonó en el vector pET21d. Se amplificó por PCR el marco de lectura abierto de *oleA* (SEQ ID NO: 11) basado en la secuencia de aminoácidos de *Xanthomonas axonopodis*, n.º de registro de GenBank NP_640589.1 GI:21241007 (SEQ ID NO: 12) a partir del plásmido de DNA 2.0 usando los cebadores LF307 (SEQ ID NO: 39) y LF308 (SEQ ID NO: 40). Se digirió el producto de PCR con PciI y HindIII y se clonó en el vector pET21d. Se amplificó por PCR el marco de lectura abierto de *oleA* (SEQ ID NO: 17) basado en la secuencia de aminoácidos de *Chloroflexus aggregans* DSM 9485 NCBI, n.º de registro de GenBank ZP_01515932.1 GI:118047293 (SEQ ID NO: 18) a partir del plásmido de DNA 2.0 usando los cebadores LF313 (SEQ ID NO: 41) y LF314 (SEQ ID NO: 42). Se digirió el producto de PCR con PciI y HindIII y se clonó en el vector pET21d. Las versiones sintéticas de los genes

oleC (SEQ ID NO: 13) y *oleD* (SEQ ID NO: 15) de *Xanthomonas axonopodis*, los genes *oleC* (SEQ ID NO: 19) y *oleD* (SEQ ID NO: 21) de *Chloroflexus aggregans*, y el gen *oleC* de *Plesiocystis pacifica* (SEQ ID NO: 87) se subclonaron directamente a partir de los vectores de DNA 2.0 pJ201 en frente del promotor T7 en los vectores pCOLADuet y pET21d usando NcoI y HindIII.

5 Se evaluó la actividad de producción de cetonas alifáticas de OleA a partir de *S. maltophilia*, *Xanthomonas axonopodis* y *Chloroflexus aggregans* detectando las cetonas alifáticas cuando se expresó la OleA respectiva en *E. coli*. Se transformaron células de *E. coli* C41(DE3) con el plásmido de interés, y se indujeron usando el protocolo de expresión de T7 descrito. A continuación, se extrajeron los sedimentos usando el método de extracción 1 y se observaron cetonas alifáticas mediante el Método de detección por CG/EM 1. Pueden usarse métodos similares para expresar OleA a partir de cualquier organismo de interés y para detectar posteriormente cetonas alifáticas mediante CG/EM. Las cetonas alifáticas observadas mediante CG/EM eran saturadas, monoinsaturadas y diinsaturadas que oscilaban en longitud de cadena de carbono entre C27 y C33 (véase la figura 9 y la tabla 19).

15 Se evaluaron las actividades de hidrocarburo sintasa de OleA, OleC y OleD a partir de *Xanthomonas axonopodis* y *Chloroflexus aggregans* y OleC a partir de *Plesiocystis pacifica* usando un ensayo de complementación que somete a prueba la producción de olefinas. Se evaluó la actividad de OleA a partir de *X. axonopodis* y *C. aggregans* en una cepa de *E. coli* que también expresó OleC y OleD de *S. maltophilia*. Se evaluó la actividad de OleC en una cepa de *E. coli* que también expresaba OleA y OleD a partir de *S. maltophilia*. Más específicamente, para someter a prueba la actividad de hidrocarburo sintasa de OleA, se transformaron células con tres plásmidos que portaban *oleA* a partir del organismo de interés, *oleC* (*S. maltophilia*) y *oleD* (*S. maltophilia*). Entonces se sometieron las células transformadas a fermentación, extracción y métodos de detección por CG/EM.

20 Para someter a prueba la actividad de hidrocarburo sintasa de OleC, se transformaron células con tres plásmidos que portaban *oleAB* (*S. maltophilia*), *oleC* a partir del organismo de interés y *oleD* (*S. maltophilia*). La adición de OleB a la *E. coli* es opcional. Entonces se sometieron las células transformadas a fermentación, extracción y métodos de detección por CG/EM.

25 Para someter a prueba la actividad hidrocarburo sintasa de OleD, se transformaron células con tres plásmidos que portaban *oleAB* (*S. maltophilia*), *oleC* (*S. maltophilia*) y *oleD* a partir del organismo de interés. La adición de OleB a la *E. coli* es opcional. Entonces se sometieron las células transformadas a fermentación, extracción y métodos de detección por CG/EM.

30 Por ejemplo, las cepas usadas para evaluar la actividad de *oleC* a partir de *C. aggregans*, *X. axonopodis* y *P. pacifica* eran de los siguientes genotipos: *E. coli* C41(DE3) Δ *fadE*; pCOLADuet_OleAB, pCDFDuet_OleD, pET21d_OleC(ChI). *E. coli* C41(DE3) Δ *fadE*; pCOLADuet_OleAB, pCDFDuet_OleD, pET21d_OleC(Xan). *E. coli* C41(DE3) Δ *fadE*; pCOLADuet_OleAB, pCDFDuet_OleD, pET21d_OleC(Ple). Las cepas usadas para evaluar la actividad de *oleD* a partir de *C. aggregans* y *X. axonopodis* fueron de los siguientes genotipos: *E. coli* C41(DE3) Δ *fadE*; pCOLADuet_OleAB, pCDFDuet_OleC, pET21d_OleD(ChI). *E. coli* C41(DE3) Δ *fadE*; pCOLADuet_OleAB, pCDFDuet_OleC, pET21d_OleD(Xan). Las cepas usadas para evaluar la actividad de *oleA* a partir de *C. aggregans* y *X. axonopodis* fueron del siguiente genotipo: *E. coli* C41(DE3); Δ *fadE*; pET21d_OleA(ChI); pACYCDuet_OleCD; pCDFDuet_fadD y *E. coli* C41(DE3); Δ *fadE*; pET21d_OleA(Xan); pACYCDuet_OleCD; pCDFDuet_fadD.

35 Para demostrar que los genes *ole* basados únicamente en las secuencias de aminoácidos de *S. maltophilia* no se requieren para la producción de hidrocarburos en *E. coli*, se usó una cepa con los genes *oleA*, *oleC* y *oleD* que codifican para las tres secuencias de proteína de *X. axonopodis* para someter a prueba la producción de hidrocarburos en *E. coli*. Se preparó *E. coli* C41(DE3) Δ *fadE*; pET21d_OleA(Xan), pCOLADuet_OleD(Xan), pCDFDuet_OleC(Xan) y se evaluó para determinar la capacidad para producir hidrocarburos. Entonces se extrajeron los sedimentos y se analizaron para determinar la producción de hidrocarburos mediante CG/EM. Los hidrocarburos observados mediante CG/EM eran olefinas mono-, di- y triinsaturadas que oscilaban en longitud de cadena de carbono entre C₂₇ y C₃₁ (tabla 19).

40 Los resultados de los experimentos reflejados en este ejemplo demuestran que las tres secuencias de proteína OleA relacionadas a partir de *S. maltophilia*, *Xanthomonas axonopodis* y *Chloroflexus aggregans* funcionan todas produciendo cetonas alifáticas cuando se expresan en *E. coli* C41(DE3) Δ *fadE*. Los resultados de los experimentos reflejados en este ejemplo también demuestran que las secuencias de proteína OleA a partir de *S. maltophilia*, *Xanthomonas axonopodis* y *Chloroflexus aggregans* funcionan todas produciendo hidrocarburos cuando se expresan en *E. coli* C41(DE3) Δ *fadE* que expresa genes *oleC* y *oleD* funcionales conocidos. Adicionalmente, los resultados de los experimentos reflejados en este ejemplo demuestran que las tres secuencias de proteína OleC relacionadas a partir de *Xanthomonas axonopodis*, *Chloroflexus aggregans* y *Plesiocystis pacifica* funcionan todas produciendo olefinas cuando se expresan en *E. coli* C41(DE3) Δ *fadE* que expresa genes *oleA* y *oleD* funcionales conocidos. De manera similar, las dos secuencias de proteína OleD relacionadas a partir de *Xanthomonas axonopodis* y *Chloroflexus aggregans* funcionan ambas produciendo olefinas cuando se expresan en *E. coli* C41(DE3) Δ *fadE* que expresa genes *oleA* y *oleC* funcionales conocidos. Además, los genes *ole* que muestran actividad en combinación con otros genes *ole* funcionales producen hidrocarburos, ya se deriven o no del mismo organismo. Por ejemplo, genes *ole* de *Xanthomonas axonopodis* funcionan junto con genes *ole* derivados de *S.*

maltophilia produciendo hidrocarburos.

TABLA 19

Proteína	Organismo	ADN	Cetona alifática							
			C27	C27:1	C27:2	C29	C29:1	C29:2	C31:1	C31:2
OleA	<i>S. maltophilia</i> ATCC17679	nativo	+	+	+	+	+	+	+	+
OleA	<i>S. maltophilia</i> R551-3	sintético	+	+	+	+	+	+	+	+
OleA	<i>Xanthomonas</i> <i>axonopodis</i>	sintético	+	+	+	+	+	+	+	+
OleA	<i>Chloroflexus</i> <i>aggregans</i>	sintético	+	+	+	+	+	+	+	+
Olefina*										
			C27:1	C27:2	C27:3	C29:1	C29:2	C29:3	C31:2	C31:3
OleA	<i>S. maltophilia</i> R551-3	sintético	+	+	+	+	+	+	+	+
OleA	<i>Xanthomonas</i> <i>axonopodis</i>	sintético	+	+	+	+	+	+	+	+
OleA	<i>Chloroflexus</i> <i>aggregans</i>	sintético	+	+	+	+	+	+	+	+
OleC	<i>S. maltophilia</i> ATCC17679	nativo	+	+	+	+	+	+	+	+
OleC	<i>Xanthomonas</i> <i>axonopodis</i>	sintético	+	+	+	+	+	+	+	+
OleC	<i>Chloroflexus</i> <i>aggregans</i>	sintético	+	+	+	+	+	+	+	+
OleC	<i>Plesiocystis</i> <i>pacifica</i>	sintético	+	+	+	+	+	+	+	+
OleD	<i>S. maltophilia</i> ATCC17679	nativo	+	+	+	+	+	+	+	+
OleD	<i>Xanthomonas</i> <i>axonopodis</i>	sintético	+	+	+	+	+	+	+	+
OleD	<i>Chloroflexus</i> <i>aggregans</i>	sintético	+	+	+	+	+	+	+	+
OleACD	<i>Xanthomonas</i> <i>axonopodis</i>	sintético	+	+	+	+	+	+	+	+

* Se sometió a prueba OleC en presencia de OleA y OleD a partir de *S. maltophilia* ATCC17679. Se sometió a prueba OleD en presencia de OleA y OleC a partir de *S. maltophilia* ATCC17679. Se sometió a prueba OleA en presencia de OleC y OleD a partir de *S. maltophilia* ATCC17679.

Este ejemplo demuestra un método de ensayo *in vivo* para identificar proteínas OleA con actividad de cetona alifática. Además, este experimento demuestra un método de ensayo *in vivo* para identificar proteínas OleA, OleC y OleD con actividad hidrocarburo sintasa.

EJEMPLO 7

10 Este ejemplo demuestra que pueden usarse motivos de aminoácidos para identificar secuencias de proteína OleA, OleB, OleC y OleD adicionales en bases de datos disponibles públicamente.

En resumen, en Internet están disponibles programas para realizar búsquedas en bases de datos de proteínas de patrones de aminoácidos específicos (motivos). Uno de tales programas es el ofrecido por el servicio GenomeNet a partir del Kyoto University Bioinformatics Center. El sitio web, con fecha de 1 de agosto de 2007, era <http://motif.genome.jp/MOTIF2.html>. Este programa de búsqueda de motivos ofrece al usuario la capacidad para realizar búsquedas en las siguientes bases de datos: Swiss-Prot, PDBSTR, PIR, PRF, GENES y NR-AA. El usuario introduce el patrón de aminoácidos específico en el formato PROSITE (véase, por ejemplo, Hofmann *et al.*, *Nucleic Acids Res.* 27: 215-219 (1999)). Por ejemplo, cada residuo debe estar separado por un - (menos); x representa cualquier aminoácido; [DE] significa o bien D o bien E; {FWY} significa cualquier aminoácido excepto P, W y Y; A(2,3) significa que A aparece de 2 a 3 veces consecutivamente; la sucesión del patrón debe estar terminada con un periodo. El usuario selecciona "buscar en bases de datos de secuencias un patrón dado", introduce un patrón de aminoácidos específico tal como se describió anteriormente en recuadro de patrón y selecciona la base de datos en

la que va a realizarse la búsqueda.

Los motivos proporcionados en las tablas 3, 7 y 10 pueden usarse para identificar proteínas OleA adicionales. Por ejemplo, se introdujo el motivo [LF]-X-X-[IVLM]-[ATSV]-G-[IV]-X-[EAHS]-R-R-X-W (SEQ ID NO: 64), que es un motivo que define la agrupación de proteína OleA, en el programa de búsqueda de motivos tal como se describió anteriormente. En la tabla 20 se muestran resultados de búsqueda a modo de ejemplo para esta consulta.

5

TABLA 20

Patrón de consulta: [LF]-X-X-[IVLM]-[ATSV]-G-[IV]-X-[EAHS]-R-R-X-W			
Buscado en la base de datos NR-AA			
	Nombre de entrada	Posición	Descripción
1.	pir:H82615	43..55	[H82615] 3-oxoacil-[ACP] sintasa III XF1970 [importada] - <i>Xilella fastidiosa</i> (cepa 9a5c)>tr:Q3RAR4_XYLFA [Q3RAR4] 3-oxoacil-(ACP) sintasa III.>tr:Q87D54_XILFT [Q87D54] 3-oxoacil-[ACP] sintasa III
2.	prf:2816338E	50..62	3-oxoacil-proteína portadora de acilo sintasa - <i>Photobacterium profundum</i> >tr:Q93CH0_PHOPR [Q93CH0] 3-oxoacil-proteína portadora de acilo sintasa III.>gp:AF409100_5 [AF409100] 3-oxoacil-proteína portadora de acilo sintasa III [<i>Photobacterium profundum</i> SS9]
3.	prf:3108482HQQ	43..55	3-oxoacil sintasa - <i>Xanthomonas oryzae oryzae</i> >tr:Q5GV10_XANOR [Q5GV10] 3-oxoacil-sintasa III.>tr:Q2NY94_XANOM [Q2NY94] 3-oxoacil-[ACP] sintasa III.
4.	prf:3117429DGF	43..55	proteína de tipo 3-oxoacil-proteína portadora de acilo sintasa III - <i>Colwellia psychrethyraea</i> >tr:Q482Y9_COLP3 [Q482Y9] 3-oxoacil-(proteína portadora de acilo) sintasa III, homólogo.>gp:CP000083_2096 [CP000083] 3-oxoacil-(proteína portadora de acilo) sintasa III, homólogo [<i>Colwellia psychrethyraea</i> 34H]
5.	prf:3122281XR	43..55	3-oxoacil-proteína portadora de acilo sintasa - <i>Xanthomonas campestris vesicatoria</i> >tr:Q3BZ43_XANC5 [Q3BZ43] 3-oxoacil-[proteína portadora de acilo] sintasa III (EC 2.3.1.41).>gpu:AM039952_239 [AM039952] 3-oxoacil-[proteína portadora de acilo] sintasa III [<i>Xanthomonas campestris</i> pv. <i>vesicatoria</i> cepa 85-10]
6.	tr:A0H249_9CHLR	43..55	[A0H249] 3-oxoacil-(proteína portadora de acilo) sintasa (EC 2.3.1.41).
7.	tr:A1R7R4_ARTAT	81..93	[A1R7R4] Regulador transcripcional supuesto, familia TetR.>gp:CP000474_2443 [CP000474] regulador transcripcional supuesto, familia TetR [<i>Arthrobacter aurescens</i> TC1]
8.	tr:A1S4T3_SHEAM	43..55	[A1S4T3] 3-oxoacil-(proteína portadora de acilo) sintasa III supuesta.>gp:CP000507_1177 [CP000507] 3-oxoacil-(proteína portadora de acilo) sintasa III supuesta [<i>Shewanella amazonensis</i> SB2B]
9.	tr:A2V0V9_SHEPU	43..55	[A2V0V9] 3-Oxoacil-(proteína portadora de acilo (ACP)) sintasa III C terminal.>tr:A1RLD2_SHESW [A1RLD2] 3-Oxoacil-(proteína portadora de acilo (ACP)) sintasa III C, proteína de dominio terminal.>tr:A4Y5D7_SHEPU [A4Y5D7] 3-Oxoacil-(proteína portadora de acilo (ACP)) sintasa III C, proteína de dominio terminal.
10.	tr:Q0HTF5_SHESR	43..55	[Q0HTF5] 3-Oxoacil-(proteína portadora de acilo (ACP)) sintasa III C, proteína de dominio terminal.>gp:CP000444_2594 [CP000444] 3-Oxoacil-(proteína portadora de acilo (ACP)) sintasa III C, proteína de dominio terminal [<i>Shewanella</i> sp. MR-7]
11.	tr:Q0KY22_9GAMM	43..55	[Q0KY22] 3-oxoacil-(proteína portadora de acilo) sintasa III supuesta.
12.	tr:Q1K1K3_DESAC	43..55	[Q1K1K3] Factor de von Willebrand, tipo A.
13.	tr:Q3R6U7_XYLFA	43..55	[Q3R6U7] 3-oxoacil-(ACP) sintasa III.>tr:Q3REG9_XYLFA [Q3REG9] 3-oxoacil-(proteína portadora de acilo) sintasa (EC 2.3.1.41).
14.	tr:A1R681_ARTAT	49..61	[A1R681] 3-oxoacil-[ACP] sintasa III supuesta.>gp:CP000474_1910 [CP000474] 3-oxoacil-[ACP] sintasa III supuesta [<i>Arthrobacter aurescens</i> TC1]

15.	tr:A3D2V1_9GAMM	43..55	[A3D2V1] 3-Oxoacil-(proteína portadora de acilo (ACP)) sintasa III C, proteína de dominio terminal.>tr:A4MM52_9GAMM [A4MM52] 3-Oxoacil-(proteína portadora de acilo (ACP)) sintasa III C, proteína de dominio terminal.>tr:A5NBF7_9GAMM [A5NBF7] 3-Oxoacil-(Proteína portadora de acilo (ACP)) sintasa III C, proteína de dominio terminal.
16.	tr:Q084U7_SHEFN	43..55	[Q084U7] 3-Oxoacil-(proteína portadora de acilo (ACP)) sintasa III C, proteína de dominio terminal.>gp:CP000447_1355 [CP000447] 3-Oxoacil-(proteína portadora de acilo (ACP)) sintasa III C, proteína de dominio terminal [<i>Shewanella frigidimarina</i> NCIMB 400]
17.	tr:Q0HH48_SHESM	43..55	[Q0HH48] 3-Oxoacil-(proteína portadora de acilo (ACP)) sintasa III C, proteína de dominio terminal.>gp:CP000446_2537 [CP000446] 3-Oxoacil-(proteína portadora de acilo (ACP)) sintasa III C, proteína de dominio terminal [<i>Shewanella sp.</i> MR-4]
18.	tr:A0X3V3_9GAMM	43..55	[A0X3V3] 3-Oxoacil-(proteína portadora de acilo (ACP)) sintasa III C, proteína de dominio terminal.
19.	tr:A1IES4_9DELT	43..55	[A1IES4] 3-oxoacil-(proteína portadora de acilo) sintasa III supuesta.
20.	tr:Q1ZDM4_9GAMM	43..55	[Q1ZDM4] 3-oxoacil-(proteína portadora de acilo) sintasa (EC 2.3.1.41).
21.	tr:A0J5A0_9GAMM	43..55	[A0J5A0] 3-oxoacil-(Proteína portadora de acilo (ACP)) sintasa III C terminal.
22.	tr:A0KYS3_SHESA	43..55	[A0KYS3] 3-Oxoacil-(proteína portadora de acilo (ACP)) sintasa III C, proteína de dominio terminal.>gp:CP000469_2697 [CP000469] 3-Oxoacil-(proteína portadora de acilo (ACP)) sintasa III C, proteína de dominio terminal [<i>Shewanella sp.</i> ANA-3]
23.	tr:A1SW97_PSYIN	43..55	[A1SW97] 3-Oxoacil-(proteína portadora de acilo (ACP)) sintasa III C, proteína de dominio terminal.>gp:CP000510_1891 [CP000510] 3-Oxoacil-(proteína portadora de acilo (ACP)) sintasa III C, proteína de dominio terminal [<i>Psychromonas ingrahamii</i> 37]
24.	tr:A5CT00_CLAM3	51..63	[A5CT00] 3-oxoacil-[proteína portadora de acilo] sintasa supuesta (EC 2.3.1.41).>gp:AM711867_2161 [AM711867] 3-oxoacil-[proteína portadora de acilo] sintasa supuesta [<i>Clavibacter michiganensis subsp. michiganensis</i> NCPPB 382]
25.	tr:Q12PH1_SHEDO	43..55	[Q12PH1] 3-oxoacil-(proteína portadora de acilo) sintasa III supuesta.>gp:CP000302_1358 [CP000302] 3-oxoacil-(proteína portadora de acilo) sintasa III supuesta [<i>Shewanella denitrificans</i> OS217]
26.	tr:Q1YZ92_PHOPR	43..55	[Q1YZ92] 3-oxoacil-(proteína portadora de acilo) sintasa (EC 2.3.1.41).
27.	tr:A1ATM6_PELPD	43..55	[A1ATM6] 3-Oxoacil-(proteína portadora de acilo (ACP)) sintasa III C, proteína de dominio terminal.>gp:CP000482_3055 [CP000482] 3-Oxoacil-(proteína portadora de acilo (ACP)) sintasa III C, proteína de dominio terminal [<i>Pelobacter propionicus</i> DSM 2379]
28.	tr:A4A9J6_9GAMM	49..61	[A4A9J6] 3-oxoacil-[proteína portadora de acilo] sintasa III.
29.	tr:A5GFJ5_9DELT	43..55	[A5GFJ5] 3-Oxoacil-(proteína portadora de acilo (ACP)) sintasa III C, proteína de dominio terminal.>gp:CP000698_1982 [CP000698] 3-Oxoacil-(proteína portadora de acilo (ACP)) sintasa III C, proteína de dominio terminal [<i>Geobacter uraniumreducens</i> Rf4]
30.	tr:A0W5R8_9DELT	43..55	[A0W5R8] 3-oxoacil-(Proteína portadora de acilo (ACP)) sintasa III C terminal.
31.	tr:A3QDN2_SHELP	43..55	[A3QDN2] 3-Oxoacil-(proteína portadora de acilo (ACP)) sintasa III C, proteína de dominio terminal.>gp:CP000606_1707 [CP000606] 3-Oxoacil-(proteína portadora de acilo (ACP)) sintasa III C, proteína de dominio terminal [<i>Shewanella loihica</i> PV-4]
32.	tr:A3ZXB5_9PLAN	43..55	[A3ZXB5] 3-oxoacil-(proteína portadora de acilo) sintasa (EC 2.3.1.41).
33.	tr:Q3DYP6_CHLAU	43..55	[Q3DYP6] 3-oxoacil-(ACP) sintasa III.

34.	tr:A1FUB4_XANMA	43..55	[A1FUB4] 3-oxoacil-(proteína portadora de acilo) sintasa (EC 2.3.1.41).
35.	tr:A4LZ02_9DELT	43..55	[A4LZ02] 3-Oxoacil-(proteína portadora de acilo (ACP)) sintasa III C, proteína de dominio terminal.
36.	tr:Q0YLU4_9DELT	43..55	[Q0YLU4] 3-oxoacil-(proteína portadora de acilo) sintasa III supuesta.
37.	tr:Q40YV3_KINRA	87..99	[Q40YV3] Similar a 3-oxoacil-(proteína portadora de acilo).
38.	tr:A6G0Q4_9DELT	43..55	[A6G0Q4] 3-oxoacil-(proteína portadora de acilo) sintasa (EC 2.3.1.41).
39.	tr:Q8EG66_SHEON	43..55	[Q8EG66] Proteína sin caracterizar supuesta.>gp:AE014299_1709 [AE014299] proteína hipotética conservada [<i>Shewanella oneidensis</i> MR-1]
40.	tr:Q7UI20_RHOBA	79..91	[Q7UI20] 3-oxoacil-[ACP] sintasa III (EC 2.3.1.41).>gp:BX294155_248 [BX294155] 3-oxoacil-[ACP] sintasa III [<i>Rhodopirellula baltica</i> SH 1]
41.	tr:Q6LS53_PHOPR	38..50	[Q6LS53] 3-oxoacil-proteína portadora de acilo sintasa III.>gp:CR378667_231 [CR378667] 3-oxoacil-proteína portadora de acilo sintasa III [<i>Photobacterium profundum</i> SS9]
42.	tr:Q4V064_XANC8	43..55	[Q4V064] 3-oxoacil-[ACP] sintasa III.>tr:Q8PDX2_XANCP [Q8PDX2] 3-oxoacil-[ACP] sintasa III.>gp:CP000050_215 [CP000050] 3-oxoacil-[ACP] sintasa III [<i>Xanthomonas campestris</i> pv. <i>campestris</i> cepa 8004]
43.	tr:Q8PQT8_XANAC	43..55	[Q8PQT8] 3-oxoacil-[ACP] sintasa III.>gp:AE011648_4 [AE011648] 3-oxoacil-[ACP] sintasa III [<i>Xanthomonas axonopodis</i> pv. <i>citri</i> cepa 306]
44.	tr:Q6AQ3_DESPS	44..56	[Q6AQ3] 3-oxoacil-[proteína portadora de acilo] sintasa III probable.>gp:CR522870_741 [CR522870] 3-oxoacil-[proteína portadora de acilo] sintasa III probable [<i>Desulfotalea psychrophila</i> LSv54]
45.	gpu:CP000750_1676	65..77	[CP000750] 3-Oxoacil-(proteína portadora de acilo (ACP)) sintasa III, proteína de dominio [<i>Kineococcus radiotolerans</i> SRS30216]

De manera similar, los motivos proporcionados en las tablas 4, 9 y 12 pueden usarse para identificar enzimas OleD adicionales que tienen actividad hidrocarburo sintasa. Estos motivos pueden usarse también para identificar enzimas OleD que tienen actividad deshidrogenasa. Los motivos proporcionados en las tablas 5, 8 y 11 pueden usarse para identificar enzimas OleC adicionales que tienen actividad hidrocarburo sintasa. Los motivos proporcionados en la

5

Los resultados de este ejemplo demuestran que los motivos de aminoácido de OleA, OleB, OleC y OleD pueden usarse para identificar secuencias de aminoácidos de OleA, OleB, OleC y OleD adicionales. Un experto habitual en la técnica, usando las secuencias de aminoácidos de OleA, OleB, OleC y OleD, también podría identificar los genes *oleA*, *oleB*, *oleC* y *oleD* correspondientes que codifican para la secuencia de aminoácidos correspondiente. Más

10

EJEMPLO 8

Este ejemplo demuestra que la expresión de OleA en células en las que la ruta de biosíntesis de ácidos grasos está alterada da como resultado una producción potenciada de cetonas alifáticas.

15

Se expresó OleA en una variedad de células de *E. coli* tal como se describe en los ejemplos 1 y 3. Las bacterias recombinantes resultantes se cultivaron, se indujeron, se sedimentaron y se extrajeron usando el método de fermentación 2, el método de extracción 1 y el método de detección 1. Las cetonas alifáticas observadas mediante CG/EM eran saturadas, monoinsaturadas y diinsaturadas que oscilaban en longitud de cadena de carbono entre C₂₇ y C₃₁.

20

Se sometieron a prueba ocho huéspedes de *E. coli* diferentes:

(1) *E. coli* C41(DE3) silvestre;

(2) *E. coli* C41(DE3) $\Delta fadE$, que tiene una delección completa de la acil-CoA deshidrogenasa *fadE*;

(3) *E. coli* C41(DE3) con un plásmido que expresa *fadD* tras la inducción con IPTG;

(4) *E. coli* C41(DE3) $\Delta fadE$ con un plásmido que expresa *fadD* tras la inducción con IPTG;

(5) *E. coli* C41(DE3) con un plásmido que expresa *'tesA*, una versión truncada del gen de tioesterasa A *tesA*, tras la inducción con IPTG;

(6) *E. coli* C41(DE3) Δ *fadE* con un plásmido que expresa *'tesA*, una versión truncada del gen de tioesterasa A *tesA*, tras la inducción con IPTG;

5 (7) *E. coli* C41(DE3) con tanto *'tesA* como *fadD* contenidos en plásmidos individuales que se expresan tras la inducción con IPTG; y

(8) *E. coli* C41(DE3) Δ *fadE* con tanto *'tesA* como *fadD* contenidos en plásmidos individuales que se expresan tras la inducción con IPTG.

10 Los plásmidos (por ejemplo, *fadD* y *'tesA*) que alteran la producción de productos intermedios de la ruta de ácidos grasos se prepararon usando métodos de biología molecular convencionales. Todos los genes clonados se pusieron bajo el control de promotores inducibles por IPTG (por ejemplo, promotores T7, tac o lac). El gen *'tesA* (SEQ ID NO: 25; gen de tioesterasa A, registro NP 415027 sin secuencia líder (Cho *et al.*, *J.Biol. Chem.*, 270: 4216-9 (1995), EC: 3,1,1,5,3.1.2,-) de *E. coli* se clonó en pETDuet-1 digerido con NdeI/AvrII (pETDuet-1, descrito en el presente documento, está disponible de EMD Chemicals, Inc., San Diego, CA).

15 Se clonó el gen *fadD* (SEQ ID NO: 492) que codifica para acil-CoA sintasa de *E. coli* en un pCDFDuet-1 digerido con NcoI/HindIII. La tabla 15 proporciona un resumen de los plásmidos generados para preparar varias cepas de producción a modo de ejemplo. Un experto habitual en la técnica apreciará que pueden usarse diferentes plásmidos y modificaciones genómicas para lograr cepas similares.

20 La producción de cetonas alifáticas aumentó cuando se expresaba *oleA* en cada una de las cepas de *E. coli* con diferentes alteraciones en *fadE*, *fadD* y/o *tesA* descritas anteriormente (véase la figura 10). Por ejemplo, la sobreexpresión de *'tesA* de *E. coli* (pETDuet-1-*'tesA*) en un sistema logró un aumento de 2,6 veces en la producción de cetonas alifáticas en comparación con *E. coli* C41(DE3) que expresa proteína OleA. En otro ejemplo, la expresión de *oleA* en la cepa que combina la sobreexpresión de *'tesA* y *fadD* de *E. coli* con una delección de *fadE* logró un aumento de 5 veces en la producción de cetonas alifáticas en comparación con *E. coli* C41(DE3) que expresa proteína OleA.

25 Los resultados de este ejemplo demuestran la capacidad para observar aumentos en cetonas alifáticas producidas por células bacterianas que expresan *oleA* en combinación con alteraciones en genes implicados en la ruta de biosíntesis de ácidos grasos de la célula.

EJEMPLO 9

30 Este ejemplo demuestra que la expresión de *oleA*, *oleC* y *oleD* en células en las que la ruta de biosíntesis de ácidos grasos se ha alterado da como resultado una producción de olefinas potenciada.

35 Se expresaron *oleA*, *oleC* y *oleD* en las cepas de *E. coli* descritas en el ejemplo 8. Las bacterias recombinantes resultantes se cultivaron, se indujeron, se sedimentaron y se extrajeron usando el método de fermentación 2, el método de extracción 1 y el método de detección 1. Las olefinas resultantes observadas mediante CG/EM eran monoinsaturadas, diinsaturadas y triinsaturadas que oscilaban en longitud de cadena de carbono entre C₂₇ y C₃₁.

La expresión de *oleA*, *oleC* y *oleD* en cepas de *E. coli* que combinan la sobreexpresión de *'tesA* y *fadD* con una delección de *fadE* dio como resultado un aumento de cuatro veces en la cantidad de olefinas observadas en comparación con la cepa de *E. coli* que expresa *oleA*, *oleC* y *oleD* en los antecedentes de *E. coli* silvestre (véase la figura 11).

40 Los resultados de los experimentos reflejados en este ejemplo demuestran que se observan cantidades significativamente aumentadas de olefinas en células en las que se expresan *oleA*, *oleC* y *oleD* en combinación con alteraciones en la ruta de biosíntesis de ácidos grasos.

EJEMPLO 10

45 Este ejemplo demuestra que se observan aldehídos, cetonas alifáticas y olefinas en un ensayo *in vitro* que contiene OleA y OleD.

50 Se expresaron proteínas OleA y OleD etiquetadas con His en *E. coli* C41(DE3) Δ *fadE* usando medio TB instantáneo de expresión durante la noche según el protocolo del fabricante (Novagen, CA). Se purificaron las proteínas usando cromatografía en columna His Bind usando NaCl 500 mM, Tris-HCl 20 mM, imidazol 5 mM pH 7,9 y Tris(hidroxipropil)fosfina (THP) 1 mM (tampón de unión 1X). A esto le siguió NaCl 500 mM, imidazol 60 mM, Tris-HCl 20 mM pH 7,9 y THP 1 mM (tampón de lavado 1X) según las instrucciones en el protocolo de usuario TB054 Rev. F0106 (Novagen de EMD Chemicals, Inc., San Diego, CA). Se eluyó la proteína OleA de las columnas His-bind usando imidazol 1 M, NaCl 0,5 M y Tris-HCl 20 mM pH 7,9. Se eluyó la proteína OleD de la columna His-bind usando NaCl 0,5 M, EDTA 100 mM y Tris-HCl 20 mM pH 7,9. Se intercambió el tampón de todas las proteínas a Tris

50 mM pH 7,5, NaCl 100 mM, THP 1 mM y glicerol al 10% usando columnas PD-10 según el protocolo del fabricante (GE Healthcare, NJ).

Se usaron las proteínas OleA y OleD purificadas, etiquetadas con His en un ensayo *in vitro*. La reacción de ensayo *in vitro* contenía miristoil coenzima A 1 mM, MgCl₂ 10 mM, OleA 2,3 μM, OleD 2,4 μM, NADPH 10 mM y tampón fosfato 100 mM (pH 7,0) en un volumen total de 100 μl. La NADPH usada en este ensayo se obtuvo de Sigma (N7505, St Louis MO) y se preparó como disolución madre 100 mM en Tris pH 7,5 50 mM. Se incubaron muestras a 37°C durante 1 hora. Se extinguió la reacción con 500 μl de acetato de etilo que contenía ácido acético al 1%. Se mezcló la mezcla agitando con vórtex seguido por centrifugación. Se transfirió la fase superior (fase de acetato de etilo) a un tubo de vidrio limpio y se secó a vacío en una centrifuga (Vacufuge 5301, Eppendorf, Westbury, NY). Se resuspendió la muestra en 40 μl de acetato de etilo y 10 μl de disolución 0,1 mg/ml de hexacosano (preparada en acetato de etilo), que actuó como adición conocida de control, y se analizó mediante CG/EM (tiempo de ejecución: 37,33 minutos; columna: HP-5-MS (5% de difenilsiloxano, 95% de dimetilsiloxano) n.º de parte 19091S-433E, longitud: (metros) 30, diámetro interno: orificio estrecho de (mm) 0,25, película: (μM) 0,25; intervalo de barrido de MSD: 50-550 m/z; inyector: entrada Agilent 6890 N de 1 μl; entrada: 300°C sin desdoblamiento; gas portador: Helio (velocidad de flujo: 1,3 ml/min); temp. del horno: 3 minutos mantenida a 60°C; 15°C/minuto hasta 320°C; 17 minutos mantenida a 320°C; det.: Agilent 5975B XL EI/CI MSD; temp. de det.: 300°C). Se analizaron entre 1 y 10 μl par determinar el contenido en hidrocarburos.

Los aldehídos, cetonas y olefinas resultantes detectados mediante CG/EM fueron tetradecanal, 14-heptacosanona e isómeros de heptacosano (véase la figura 12A-E). Los compuestos se identificaron basándose en el tiempo de retención y los perfiles espectrales de EM.

Los resultados del experimento reflejado en este ejemplo demuestran la capacidad para observar aldehídos, cetonas alifáticas y olefinas usando OleA y OleD en un ensayo *in vitro*. Específicamente, estos resultados demuestran que se observan aldehídos, cetonas alifáticas y olefinas tras la incubación de OleA y OleD en presencia de miristoil coenzima A, MgCl₂ y NADPH en un ensayo *in vitro*.

25 EJEMPLO 11

Este ejemplo demuestra que la cantidad de aldehídos y olefinas observadas en un ensayo *in vitro* que comprende un sustrato de acil coenzima A, OleA y OleD se potencia significativamente cuando se añade proteína OleB purificada al ensayo *in vitro*.

Se expresaron proteínas OleA y OleD etiquetadas con His y se purificaron usando cromatografía en columna His Bind tal como se describe en el ejemplo 10. Se expresaron proteínas OleB etiquetadas con His y se purificaron usando el método descrito en el ejemplo 10. Se eluyeron las proteínas OleA y OleB de las columnas His-bind usando imidazol 1 M, NaCl 0,5 M y Tris-HCl 20 mM pH 7,9. Se eluyó la proteína OleD de la columna His-bind usando NaCl 0,5 M, EDTA 100 mM y Tris-HCl 20 mM pH 7,9. Se intercambió el tampón de todas las proteínas a Tris 50 mM pH 7,5, NaCl 100 mM, THP 1 mM y glicerol al 10% usando columnas PD-10 según el protocolo del fabricante (GE Healthcare, NJ).

Se usaron las proteínas OleA, OleD y OleB purificadas, etiquetadas con His en un ensayo *in vitro*. La reacción de ensayo *in vitro* contenía miristoil coenzima A 1 mM, MgCl₂ 10 mM, OleA 2,3 μM, OleD 2,4 μM, OleB 2,3 μM, NADPH 10 mM y tampón fosfato 100 mM (pH 7,0) en un volumen total de 100 μl. Se incubaron las muestras a 37°C durante 1 hora. Se extinguió la reacción con 500 μl de acetato de etilo que contenía ácido acético al 1%. Se mezcló la mezcla agitando con vórtex seguido por centrifugación. Se transfirió la fase superior (fase de acetato de etilo) a un tubo de vidrio limpio y se secó a vacío en una centrifuga (Vacufuge 5301, Eppendorf, Westbury, NY). Se resuspendió la muestra en 40 μl de acetato de etilo y 10 μl de disolución 0,1 mg/ml de hexacosano (preparada en acetato de etilo), que actuó como adición conocida de control, y se analizó mediante CG/EM. Se analizaron entre 1 y 10 μl para determinar la cantidad de aldehídos, cetonas alifáticas y olefinas presentes.

Los aldehídos, cetonas alifáticas y olefinas resultantes detectados mediante CG/EM fueron tetradecanal, 14-heptacosanona e isómeros de heptacosano (véanse las figuras 13A-C). Se identificaron los compuestos basándose en el tiempo de retención y perfiles espectrales de EM. Se evaluó área bajo el pico (AP) para cada compuesto usando el software MSD ChemStation (Agilent technologies). Se calcularon los valores relativos normalizando cada uno de los productos de reacción producidos por OleA/OleD a 1,0 (por ejemplo, aldehído producido por OleA+OleD+OleB = AP (reacción de OleA+OleD+OleB) / AP (reacción de OleA+OleD)). En la tabla 21 se muestran los valores relativos de aldehído, olefinas y cetonas alifáticas producidos por OleA+OleD y OleA+OleD+OleB.

TABLA 21

Reacción	Aldehído	Olefinas		Cetona alifática
		C27:1 Olefina 1	C27:1 Olefina 2	
OleA+ OleD	1	1	1	1
OleA + OleD + OleB	8,3	8,2	8,2	0,2

Los resultados del experimento reflejado en este ejemplo demuestran que la cantidad de aldehído u olefina observada en un ensayo *in vitro* que comprende OleA, OleD y OleB es mayor que la cantidad de aldehído u olefina observada en un ensayo *in vitro* que comprende OleA y OleD.

EJEMPLO 12

- 5 Este ejemplo demuestra que pueden convertirse ácidos grasos en olefinas por bacterias que expresan un gen de acil-CoA sintasa en combinación con los genes de olefina sintasa (*oleA*, *oleB*, *oleC* y *oleD*).

El gen de acil-CoA sintasa convierte ácidos grasos libres en acil graso-CoA activados, uno de los sustratos para los genes de olefina sintasa. Se transformaron varias cepas huésped diferentes con un gen de acil-CoA sintasa. Se usaron los genes de olefina sintasa en una serie de experimentos de bioconversión. Todas las cepas huésped
 10 dieron resultados similares y se produjeron usando técnicas de biología molecular convencionales con los plásmidos descritos en el ejemplo 1. Se usaron las siguientes cepas huésped para generar los datos mostrados en este ejemplo: Huésped 1: *E. coli* MG1655 Δ *fadE* transformada con el operón de olefina sintasa pCL1920pTrcOleABpTrcOleCD (SEQ ID NO: 78) y el plásmido de sobreexpresión de acil-CoA sintasa FadD pACYCpTrcFadD.

- 15 Huésped 2: *E. coli* C41 (DE3) Δ *fadE* transformada con el operón de olefina sintasa pCL1920pTrcOleABpTrcOleCD (SEQ ID NO: 78) y el plásmido de sobreexpresión de acil-CoA sintasa FadD pETDuetFadD.

Se realizaron experimentos de bioconversión usando el procedimiento de fermentación de 25 ml descrito en el ejemplo 1. En el momento de la inducción, se añadieron ácidos grasos libres al medio de cultivo hasta una
 20 concentración final de entre el 0,01% y el 0,05% en peso por volumen de ácido graso. Se usó el método de cultivo de células completas de 1 ml para extraer los hidrocarburos. Se analizaron los productos usando CG/EM tal como se describe en el ejemplo 1.

Los hidrocarburos detectados mediante CG/EM eran cadenas mono-, di- y triinsaturadas que oscilaban entre C₁₉ y C₃₃. En la tabla 22 se muestran el tipo sustratos de ácido graso añadidos a las fermentaciones y los productos de olefinas que se detectaron. La cantidad y el tipo de olefina detectada variaban dependiendo del tipo de sustrato de
 25 ácido graso que se añadiera al medio de cultivo. Esto es lo más notable cuando las cepas producen olefinas que no se detectan en la cepa de control, que no se alimentó. Por ejemplo, la tabla 22 muestra que un ácido graso C10 se convierte en una olefina C19:1, lo que reflejaría la condensación de dos moléculas de acil graso-CoA C10. También se forman olefinas C23:1 y C23:2. Esto refleja la condensación del acil graso CoA C10 con un éster activado con acil graso C14:0 o C14:1, que se produce de manera natural por la cepa huésped de *E. coli*. Se forman olefinas de
 30 cadena más larga tales como C31:3 cuando la célula convierte ácidos grasos C16:1 en olefinas. Cuando se alimentaron sustratos de ácidos grasos no nativos, tales como los ácidos grasos de cadena impar C13, C15 y C17, se detectaron olefinas de cadena par únicas de C26, C28 y C30. Esto sugiere que los ácidos grasos exógenos están convirtiéndose en olefinas.

TABLA 22

Olefina producida (mg/l/DO)	Alimentación de ácidos grasos de longitud de cadena diferente									
	Ninguno	C10	C12	C13	C14	C15	C17	C14:1	C16:1*	C18:1
019:1		0,32								
023:2		0,60								
023:1		0,44	0,82							
025:2		0,73	0,82	0,01	0,02				0,11	0,06
025:1		0,30	0,50	1,37	0,10			0,34		0,04
026:2				1,05						
026:1				0,72						
027:3	0,60	0,15	0,17	0,33	0,18	0,44	0,40	2,20	0,31	0,85
027:2	0,97	0,29	0,84	0,46	2,33	0,40	0,43	13,72	0,62	1,50
027:1	0,28	0,14	0,44	0,35	5,29	0,35	0,10		0,12	0,51
028:2				0,76		0,60				
028:1				0,55		0,26				
029:3	0,46	0,17	0,23	0,42	0,14	0,78	0,50	1,32	2,00	0,85
029:2	0,58	0,24	0,34	0,54	1,01	0,56	0,41	5,70	1,39	1,10
029:1	0,15	0,10	0,13	0,24	0,80	0,40	0,17	0,82		0,44
030:2				0,03		0,40				
030:1				0,06		0,14				
031:3	0,04		0,02	0,16	0,04	0,27	0,11			0,15
031:2	0,03		0,01	0,14	0,05	0,12	0,08	0,34	2,00	0,20
033:3									1,39	

*cultivo realizado usando *E. coli* C41 (DE3) Δ *fadE*, en lugar de *E. coli* MG 1655 Δ *fadE*

La bioconversión de ácidos grasos potenció la producción total de olefinas en la mayoría de los experimentos. Por ejemplo, la cantidad total de olefina producida tras la alimentación de ácidos grasos C13, C14, C14:1 y C16:1 aumentó más de dos veces en comparación con la cantidad de olefina producida en la cepa de control que no se alimentó (véase la tabla 23).

5 TABLA 23

Bioconversión de ácidos grasos en olefinas		
Ácido graso	DO ₆₀₀	Olefina total (mg/l/DO)
Ninguno	1,51	3,13
C10	1,62	3,47
C12	1,89	4,3
C13	3,53	7,18
C14	2,78	9,96
C15	3,45	4,71
C17	2,55	2,18
C14:1	2,4	24,44
C16:1	3,79	7,95
C18:1	2,24	5,7

Los resultados de los experimentos reflejados en este ejemplo demuestran que la expresión de OleA, OleB, OleC, OleD y FadD en *E. coli* da como resultado la producción de olefinas derivadas de ácidos grasos exógenos y endógenos. Además, los resultados de los experimentos reflejados en este ejemplo demuestran que la expresión de OleA, OleB, OleC, OleD y FadD en bacterias alimentadas con ácidos grasos da como resultado la bioconversión de ácidos grasos en olefinas y un aumento significativo en la producción total de olefinas. Además, este ejemplo demuestra que la bioconversión puede usarse para controlar los tipos de olefinas producidas por el huésped.

EJEMPLO 13

Este ejemplo demuestra que pueden convertirse ácidos grasos en cetonas alifáticas por bacterias que expresan un gen de acil-CoA sintasa en combinación con la enzima de condensación de acilos OleA.

15 El gen de acil-CoA sintasa convierte ácidos grasos libres en acil graso-CoA activados, uno de los sustratos para la reacción de condensación de acil-CoA de OleA.

Se usó la siguiente cepa huésped para generar los datos mostrados en este ejemplo: *E. coli* MG1655 Δ *fadE*, con un gen *fadD* expresado de manera constitutiva generado reemplazando el promotor nativo por un promotor T5 constitutivo (véase el ejemplo 1), transformada con el plásmido pCL1920pTrcOleA para la expresión de OleA (véase el ejemplo 1 para una descripción de este plásmido). Se cultivó la cepa huésped, se indujo con IPTG y se alimentaron ácidos grasos de diferente longitud de cadena según los procedimientos descritos en el ejemplo 12. Se usó el método de cultivo de células completas de 1 ml para extraer los hidrocarburos. Se analizaron los productos usando CG/EM tal como se describe en el ejemplo 1.

25 Los hidrocarburos detectados mediante CG/EM eran cetonas alifáticas saturadas así como mono- y diinsaturadas que oscilaban entre C₁₉ y C₃₁ (véase la tabla 24). En la tabla 18 se muestran el tipo de ácido graso añadido a las fermentaciones y las cetonas alifáticas que se observaron. Se produjeron cetonas alifáticas de cadena de carbono de número par tras la alimentación de ácidos grasos no nativos tales como C13, C15 y C17.

TABLA 24

Cetona alifática producida (mg/l/DO)	Alimentación de ácidos grasos de longitud de cadena diferente									
	Ninguno	C10	C12	C13	C14	C15	C17	C14:1	C16:1	C18:1
25:1			0,10	0,01				0,02	0,05	0,04
25:0										
26:1				0,05						
26:0				0,02						
27:2	0,06	0,39	0,72	0,19	0,26	0,97	1,46	0,85	0,85	0,53
27:1	0,07	0,43	0,70	0,25	0,50	1,07	1,41	3,77	1,87	1,17
27:0	0,01	0,05	0,11	0,04	0,14	0,14	0,21	2,58	0,42	0,36
28:1				0,03		0,19	0,02			
28:0				0,02		0,07	0,01			
29:2	0,02	0,19	0,37	0,09	0,14	0,52	0,71	0,33	0,80	0,16
29:1	0,02	0,13	0,27	0,08	0,21	0,60	0,70	1,37	1,41	0,16

29:0					0,04	0,06	0,07	0,99	0,17	
31:2						0,03			0,12	
31:1									0,14	

La bioconversión de ácidos grasos potenció la producción de cetonas alifáticas en todas las muestras sometidas a prueba con una potenciación mayor de 50 veces cuando se alimentó un sustrato preferido tal como C14:1 (véase la tabla 25).

TABLA 25

Bioconversión de ácidos grasos en cetonas alifáticas		
Ácido graso	DO ₆₀₀	Cetonas alifáticas totales (mg/l/DO)
Ninguno	3,53	0,19
C10	3,43	1,20
C12	4,26	2,28
C13	4,29	0,78
C14	5,48	1,29
C15	4,09	3,65
C17	4,04	4,59
C14:1	4,63	9,91
C16:1	4,15	5,83
C18:1	4,18	2,42

- 5 Los resultados de los experimentos reflejados en este ejemplo demuestran que la expresión de OleA y FadD en *E. coli* dio como resultado la producción de cetonas alifáticas derivadas de ácidos grasos exógenos y endógenos. Además, los resultados de los experimentos reflejados en este ejemplo demuestran que la expresión de OleA y FadD en bacterias alimentadas con ácidos grasos dio como resultado la bioconversión de ácidos grasos en cetonas alifáticas y un aumento significativo en la producción total de cetonas alifáticas. Este ejemplo demuestra que la bioconversión puede usarse para controlar los tipos de cetonas alifáticas producidas por la cepa huésped así como para potenciar las cantidades de cetonas alifáticas producidas.

EJEMPLO 14

Este ejemplo demuestra modos para controlar los tipos de olefinas que pueden producirse en bacterias.

- 15 Con el fin de producir acil-CoA (un sustrato para la síntesis de cetonas alifáticas y olefinas) de diferente longitud de cadena en bacterias, se produjeron diferentes tioesterasas, incluyendo Uc FatB1 (Voelker *et al.*, *Science*, 257: 72-74 (1992)), Ch FatB2 (Dehesh *et al.*, *Plant J.*, 9: 167-72 (1996)), 'Tesa (Cho *et al.*, citado anteriormente) o Cc FatB1 (Yuan *et al.*, *Proc. Natl. Acad. Sci. U S A*, 92: 10639-43 (1995)) en combinación con una acil-CoA sintasa (FadD; EC 6.2.1.3). Con el fin de producir olefinas con diferentes longitudes de cadena, se combinaron variaciones en el tipo de sustratos de acil-CoA producidos (es decir, bacterias que expresan diferentes tioesterasas) con los genes de olefina sintasa *oleA*, *oleB*, *oleC* y *oleD*.

- 20 Se generaron varias cepas que demuestran la capacidad para producir una gama de olefinas con longitudes de cadena variables (véase la tabla 26). Todas las cepas se prepararon en el huésped de *E. coli* C41(DE3) Δ *fadE* usando métodos de transformación convencionales, tales como electroporación o transformación química. Para mantener cada plásmido dentro de la cepa, se aplicó la selección con antibióticos apropiados a lo largo de todo el experimento. Se generó una cepa de control (OS333), que no contiene el plásmido de tioesterasa. Además, se generaron cepas que expresan la combinación de una tioesterasa, una acil graso-CoA sintasa, y los genes de olefina. En el ejemplo 1 puede encontrarse una descripción detallada de la construcción de cada plásmido. En la tabla 26 se muestra una lista de las cepas usadas en este ejemplo.

TABLA 26

Cepa n.º	Olefina sintasa	Acil-CoA sintasa	Tioesterasa
OS333	pCL1920pTrcOleABpTrcLOleCD	pACYCDuet-fadD	Ninguna
OS335	pCL1920pTrcOleABpTrcLOleCD	pACYCDuet-fadD	pLS9-80 (Uc FatB1)
OS336	pCL1920pTrcOleABpTrcLOleCD	pACYCDuet-fadD	pLS9-82 (Ch FatB2)
OS338	pCL1920pTrcOleABpTrcLOleCD	pACYCDuet-fadD	pLS9-85 ('Tesa)
OS342	pCL1920pTrcOleABpTrcLOleCD	pACYCDuet-fadD	pLS9-77 (Cc FatB1)

- 30 Se realizaron fermentaciones de 25 ml convencionales y extracciones de cultivo de células completas de 1 ml convencionales tal como se describe en el ejemplo 1. Se analizaron las muestras mediante el método de CG/EM descrito en el ejemplo 1. Se usó un patrón interno de hexacosano para determinar las cantidades de cada olefina producida (véase la tabla 27).

En 1962 Marr y Ingraham publicaron un artículo demostrando que la temperatura influye en el grado de saturación en los lípidos producidos por *E. coli* (Marr *et al.*, *J Bacteriol.*, 84: 1260-7 (1962)). Bajas temperaturas dan como resultado la producción de lípidos con un mayor grado de insaturación mientras que temperaturas más altas dan como resultado cantidades superiores de lípidos saturados. Para determinar si la temperatura influye en el grado de saturación, se realizaron las fermentaciones a dos temperaturas diferentes, 37°C y 25°C. La parte superior de la tabla 27 muestra los resultados para fermentaciones realizadas a 37°C, y la mitad inferior de la tabla 27 muestra los resultados para fermentaciones realizadas a 25°C.

TABLA 27

Olefina ((mg/l)/unidad de DO600)															
37°C	C23:2	C23:1	C25:2	C25:1	C27:3	C27:2	C27:1	C29:3	C29:2	C29:1	C31:3	C31:2	C33:1	Total	Unidades de DO600
no TE					0,354	0,253	0,037	0,314	0,423	0,050	0,168	0,012		1,611	3,105
UcFatB1		0,150	0,591	0,106	0,633	0,488	0,114	0,459	0,383	0,014	0,168			3,106	1,342
ChFatB2					0,936	0,729	0,095	0,793	1,093	0,136	0,273			4,055	1,754
tesA					0,767	0,650	0,104	0,708	0,583	0,091	0,173	0,076		3,153	3,136
CcFatB1					0,668	0,654	0,130	0,453	0,644	0,106	0,222	0,042		2,919	3,567
25°C															
no TE					0,787	0,378		1,737	0,863		1,621	0,305	0,300	5,991	1,542
UcFatB1			0,981		1,955	0,298		1,640	0,403		0,466			5,742	0,402
ChFatB2	0,225		0,236		0,364	0,191		0,821	0,381		0,675	0,129	0,135	2,933	2,030
tesA			0,069		1,175	1,201	0,263	2,599	1,589	0,173	1,255		0,094	8,417	2,192
CcFatB1			0,335	0,118	2,168	2,384	0,533	2,300	1,424	0,179	1,818	0,684	0,191	12,135	3,168

La longitud de la olefina y el grado de insaturación se indican como CX:Y, en el que X representa el número de carbonos en la olefina e Y indica el número de dobles enlaces.

Los resultados reflejados en la tabla 27 demuestran que la expresión de una tioesterasa en combinación con una acil graso-CoA sintasa y genes de olefina sintasa produce olefinas con diferentes longitudes de cadena dependiendo de la tioesterasa usada. Pueden usarse tioesterasas tales como Uc FatB1, que se conoce como proteína portadora de lauroil-acilo tioesterasa con especificidad por acil graso-ACP C12:0, y Ch FatB2 para producir olefinas de longitud C23 y C25, que no se observan en la cepa de control. Estos resultados también demuestran que la cantidad de una olefina específica, tal como olefina C27, puede potenciarse mediante la adición de una tioesterasa con especificidad por acil graso-ACP C14, tales como Cc FatB1. Alternativamente, la cantidad de C27 y C29 puede potenciarse mediante la adición de 'TesA, que tiene especificidad por acil graso-ACP de longitud de cadena C14 y C16.

Los resultados expuestos en la tabla 27 también demuestran que la temperatura a la que se realiza la fermentación influirá en los tipos de olefinas que producen las bacterias. En particular, la temperatura influye en el nivel de saturación de las olefinas producidas. Se forman olefinas con un único doble enlace a partir de la condensación de dos ácidos grasos saturados. Las fermentaciones realizadas a 37°C dan como resultado la producción de olefinas con un grado de saturación superior en comparación con fermentaciones que se realizaron a 25°C.

Los resultados de los experimentos reflejados en este ejemplo demuestran que combinando genes de olefina sintasa con genes que producen acil graso-CoA de longitud de cadena diferente, puede controlarse la longitud de la olefina producida. Además, los resultados de los experimentos reflejados en este ejemplo demuestran que el grado de saturación en las olefinas producidas por las bacterias puede controlarse alterando las condiciones medioambientales, tales como la temperatura de las fermentaciones.

EJEMPLO 15

Este ejemplo demuestra que la expresión funcional de *oleA* en *Saccharomyces cerevisiae* da como resultado la producción *in vitro* de cetonas alifáticas.

Se amplificó por PCR la secuencia de ácido nucleico de *oleA* mediante métodos convencionales a partir de un plásmido que contiene la secuencia del gen *oleA* de *Stenotrophomonas maltophilia* ATCC 17679 (SEQ ID NO: 1) y se clonó en vectores de expresión de levaduras usando técnicas convencionales, tal como se describe en el presente documento. En resumen, se amplificó *oleA* (SEQ ID NO: 1) usando los cebadores LB217 (SEQ ID NO: 43) y LF304 (SEQ ID NO: 44). Se clonó el producto de amplificación de *oleA* en pESC-HIS usando las enzimas de restricción Apal y XhoI creando el plásmido pESC-His-*oleA* (SEQ ID NO: 134).

Se transformaron entonces células de levadura de *S. cerevisiae* (BY4741) con o bien un plásmido que contiene *oleA* (SEQ ID NO: 134) o bien un plásmido que no contiene *oleA*. Se cultivaron las células de levadura transformadas para permitir la expresión de *oleA*. Se sedimentaron las células y luego se lisaron usando YeastBuster™ (Novagen, Madison, WI). Se añadió miristoil-CoA al lisado celular. Se extrajeron las cetonas alifáticas de la reacción usando ácido acético al 1% en acetato de etilo y se analizaron usando CG/EM, tal como se describe en el ejemplo 1.

Se identificó la cetona alifática 14-heptacosanona, una cetona alifática C27, mediante CG/EM en lisados celulares a partir de células de levadura transformadas con el plásmido de *oleA* pESC-His-*oleA* (véase la figura 14A). Esto demuestra que *oleA* puede expresarse en *S. cerevisiae* para producir cetonas alifáticas. Pueden usarse métodos similares para expresar *oleA* en cualquier célula de interés, preparar un lisado celular y posteriormente analizar la producción de cetonas alifáticas.

Los resultados del experimento reflejado en este ejemplo demuestran que puede expresarse *oleA* en *S. cerevisiae* conduciendo a la producción *in vitro* de cetonas alifáticas.

EJEMPLO 16

Este ejemplo demuestra que se observan olefinas y aldehídos deuterados en un ensayo *in vitro* que contiene OleA purificada, lisado celular que contiene OleD y NADPH deuterada.

Se produjo proteína OleA para estos experimentos mediante fermentación de *E. coli* C41(DE3) transformada con pET-21b(+)_OleA. Se purificó la OleA etiquetada con His tal como se describe en el ejemplo 10. Se produjo lisado celular de OleD para estos experimentos mediante fermentación de *E. coli* C41(DE3) transformada con pET-21b(+)_OleD usando medio TB instantáneo de expresión durante la noche según el protocolo del fabricante (Novagen, CA). Se usaron 10 ml de cultivo durante la noche para preparar los lisados celulares. Se centrifugó el cultivo para concentrar las células para dar un sedimento. Se retiró el sobrenadante. Se resuspendió el sedimento en 2 ml de tampón fosfato de sodio 50 mM (pH 7,0). Entonces se lisó la suspensión celular usando sonicación (5 veces en pulsos de 5 segundos 0,5 de potencia). Se añadieron 2 μ l de Benzonase® (Novagen, CA) al lisado celular y se mantuvo la muestra a temperatura ambiente durante 20 minutos. Se usó este lisado celular como fuente de proteína OleD.

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 marcada mediante la adición de 5 unidades de glucosa deshidrogenasa a partir 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 minutos a 37°C, se filtró con centrifuga usando un filtro de centrifuga Amicon Ultra de MWCO de 10 KDa (Millipore), se congeló inmediatamente en hielo seco y se almacenó a -80°C.

La reacción de ensayo *in vitro* contenía miristoil coenzima A 0,1 mM, MgCl₂ 10 mM, 10 μ l de dilución 1/10 de lisado de OleD, 50 μ l de deuterio-NADPH (preparada tal como se describió anteriormente), OleA 2,1 μ M y tampón fosfato 100 mM (pH 7,0) en un volumen total de 0,1 ml. La miristoil coenzima A usada se obtuvo de Sigma (M4414, St Louis MO) y se preparó como disolución madre 10 mM en tampón fosfato 0,1 M pH 7,0. Se incubaron las muestras a 37°C durante 1 hora. Se extinguió la reacción con 500 μ l de acetato de etilo que contenía ácido acético al 1%/ácido fórmico al 0,15%. Se mezcló la mezcla agitando con vórtex seguido por centrifugación. Se transfirió la fase superior (fase de acetato de etilo) a un tubo de vidrio limpio y se secó a vacío en una centrifuga (Vacufuge 5301, Eppendorf, Westbury, NY). Se resuspendió la muestra en 40 μ l de acetato de etilo y 10 μ l de disolución 0,1 mg/ml de hexacosano (preparada en acetato de etilo), que actuó como adición conocida de control, y se analizó mediante CG/EM. Se analizaron entre 1 y 10 μ l para detectar hidrocarburos específicos.

Los aldehídos y olefinas resultantes detectados mediante CG/EM eran tetradecanal e isómeros de heptacosano (véase la figura 15A-H). Debido a que la transferencia de hidruro a partir de NADPH es estereoespecífica, se sintetizaron tanto R-(4-²H)NADPH como S-(4-²H)NADPH. Se observaron aldehído y olefinas con una masa de más una unidad usando solo la R-(4-²H)NADPH. El hecho de que el aldehído y las olefinas estuvieran marcados indica que el hidrógeno deuterado se ha transferido desde la NADPH marcada hasta la olefina o el aldehído marcado. Esto demuestra que se usa NADPH en la reacción enzimática. El hecho de que la transferencia de hidruro sea estereoespecífica es indicativo de una reacción enzimática.

Los resultados del experimento reflejado en este ejemplo demuestran que *oleD* codifica para un nucleótido de piridina oxidorreductasa. Además, estos resultados demuestran que OleD cataliza la transferencia de hidruro desde NADPH hasta productos intermedios en biosíntesis de olefina, de manera que el hidruro transferido permanece en el producto de olefina. La transferencia de hidruro que se produce en presencia de OleA y miristoil CoA dio como resultado la detección de tanto C1-deuterio-tetradecanal como C14-deuterio, delta-13, heptacosano.

EJEMPLO 17

Este ejemplo demuestra la capacidad para sintetizar olefinas alifáticas *in vitro*.

Protocolo para la expresión de proteínas y la preparación de lisado para evaluar la actividad de proteínas OleA/B/C/D

Se usó *E. coli* C41(DE3) transformada con pETDuet_OleAB y pCOLADuet_OleCD para crear un lisado bacteriano que contenía las cuatro proteínas Ole (OleA, OleB, OleC y OleD).

Se hicieron crecer células de *E. coli* C41(DE3) transformadas con pETDuet_OleAB y pCOLADuet_OleCD en caldo

Luria que contenía kanamicina (concentración final 50 µg/ml) y carbenecilina (concentración final 100 µg/ml) hasta una DO₆₀₀ de 0,5 - 1,0 a 37°C. Entonces se indujeron los cultivos con IPTG 1 mM y se hicieron crecer durante la noche a 37°C. Se centrifugaron los cultivos que se hicieron crecer durante la noche a 3000 rpm durante 20 minutos. Se desechó el sobrenadante. Se congelaron los sedimentos a -80°C para usarse más tarde. Se resuspendieron los sedimentos celulares obtenidos a partir de 10 ml de cultivos durante la noche inducidos en 2 ml de tampón y se lisaron usando sonicación (5 veces en pulsos de 5 s a 0,5 de potencia). En el lisado sonificado, se añadieron 2 µl de Benzonase® (Novagen, CA), y se mantuvo la muestra a temperatura ambiente durante 20 minutos. Se usó este lisado como fuente de proteínas OleA, OleB, OleC y OleD.

Protocolo para la preparación de miristoil-ACP

Se sintetizó miristoil-ACP tal como se describe en Rock *et al.*, *Methods Enzymol.*, 72: 397-403 (1981), con algunas modificaciones. En resumen, una mezcla de reacción que contenía ATP 5 mM, DTT 2 mM, Triton X-100 al 2%, LiCl 10 mM, sal de sodio de ácido mirístico 160 µM, ACP-SH 65 µM, MgCl₂ 10 mM, Tris-HCl 0,1 M pH 8,0 y acil-ACP sintasa 1,5-3 µg/ml en un volumen final de 1-4 ml se incubó a 37°C durante 3 horas y luego se incubó durante la noche a 30°C. La acil-ACP sintasa usada se adquirió de Invitrogen. Se diluyó la mezcla de reacción con tres volúmenes de agua y se valoró el pH hasta 6,0 con ácido acético. Se aplicó la disolución a una columna HiTrap DEAE FF de 5 ml (GE Healthcare). Se lavó la columna con tres volúmenes de columna de bis-Tris-HCl 10 mM, pH 6,0, para eliminar el volumen del Triton X-100. Se eliminaron el Triton X-100 y el ácido graso libre residuales eluyendo la columna con tres volúmenes de columna de 2-propanol al 80%. Se aclaró el 2-propanol con tres volúmenes de columna de bis-Tris-HCl 10 mM, pH 6,0. Se eluyeron el miristoil-ACP y el ACP-SH sin reaccionar de la columna con 10 ml de LiCl 0,6 M en bis-Tris-HCl 10 mM, pH 6,0. Se intercambié el tampón del eluato que contenía miristoil-ACP a tampón fosfato de sodio 50 mM pH 7,0 y se concentró hasta 1 ml usando concentradores de 3 kDa de MWCO (Millipore). Se determinó la concentración de miristoil-ACP usando el ensayo de Bradford (BioRad) y análisis de densitometría en gel de Bis-Tris SDS-PAGE al 12% NuPAGE (Invitrogen) en dos etapas: en primer lugar, se determinó la concentración de proteína total de la fracción de C14-ACP/ACP-SH mediante un ensayo de Bradford (ensayo de proteína de BioRad); en segundo lugar, se calculó la concentración de C14-ACP como porcentaje de la concentración de proteína total basándose en SDS-PAGE seguido por un análisis de densitometría. Se almacenó C14-ACP en tampón fosfato de Na 50 mM, pH 7,0 a -20°C y era estable durante hasta 3 meses.

Un método para producir olefinas

Se usó *E. coli* C41(DE3) transformada con pETDuetOleAOleB y pCOLADuetOleCOleD para crear un lisado celular que contenía proteínas OleA, OleB, OleC y OleD tal como se describió anteriormente. La reacción de ensayo *in vitro* contenía miristoil coenzima A 0,1 mM o miristoil-ACP 0,1 mM, MgCl₂ 10 mM, 360 µl de lisado celular a partir de *E. coli* C41(DE3) transformada con pETDuetOleAOleB y pCOLADuetOleCOleD, NADPH 1 mM (Sigma, MO), ATP 1 mM (Sigma, MO), HSCoA 1 mM (Sigma, MO) y tampón fosfato 100 mM (pH 7,0) en un volumen total de 1 ml. La miristoil coenzima A usada se obtuvo de Sigma (M4414, St Louis MO) y se preparó como disolución madre 10 mM en tampón fosfato 0,1 M pH 7,0. Se incubaron las muestras a 37°C durante 1 hora. Se extinguió la reacción con 5 ml de acetato de etilo que contenía ácido acético al 1%/ácido fórmico al 0,1%. Se mezcló la mezcla agitando con vórtex seguido por centrifugación. Se transfirieron 4 ml de la fase superior (fase de acetato de etilo) a un tubo de vidrio limpio y se secó a vacío en una centrífuga (Vacufuge 5301, Eppendorf, Westbury, NY). Se resuspendió la muestra en 40 µl de acetato de etilo y 10 µl de disolución 0,1 mg/ml de hexacosano (preparada en acetato de etilo), que actuó como adición conocida de control, y se analizó mediante CG/EM. Se analizó 1 µl para determinar hidrocarburos específicos.

Las olefinas resultantes detectadas mediante CG/EM fueron C27:1, C27:2 y C27:3 (véanse las figuras 16A-C). La tabla 28 muestra el porcentaje de olefina C27:1 producida por lisados celulares de *E. coli* C41(DE3) transformada con pETDuetOleAOleB y pCOLADuetOleCOleD en presencia de C14-CoA o C14-ACP. Se calculó el porcentaje de olefina C27:1 mediante: $[\text{área de pico de C27:1}/(\text{área de pico de C27:1} + \text{área de pico de C27:2} + \text{área de pico de C27:3})] \times 100$. Los resultados mostrados en la tabla 28 indican que la adición de C14-CoA o C14-ACP a lisado celular de *E. coli* C41(DE3) transformada con pETDuetOleAOleB y pCOLADuetOleCOleD conduce a un aumento en la cantidad de olefina C27:1. Esto indica que *E. coli* C41(DE3) transformada con pETDuetOleAOleB y pCOLADuetOleCOleD presentes en el lisado puede usar C14-CoA o C14-ACP para formar olefina C27:1.

50 TABLA 28

	% de C27:1 en alquenos C27
Lisado + C14-CoA	47
Lisado + C14-ACP	41
Lisado solo	24

Los resultados de los experimentos reflejados en este ejemplo demuestran que puede usarse un ensayo *in vitro* que comprende miristoil-ACP o miristoil coenzima A en combinación con OleA, OleB, OleC y OleD para sintetizar olefinas.

EJEMPLO 18

Este ejemplo demuestra que las secuencias de proteínas OleA, OleB, OleC y OleD pueden usarse para identificar organismos que producen de manera natural hidrocarburos.

Se usó el programa bioinformático pBLAST para identificar secuencias de proteínas que pertenecen a las familias de proteínas OleA, OleB, OleC y OleD. Se identificó un pequeño número de organismos que contienen las cuatro secuencias de proteínas Ole agrupadas juntas (véase, por ejemplo, la tabla 1). De estos organismos, los siguientes cuatro organismos se ha mostrado en la bibliografía que producen los tipos de olefinas descritos en el presente documento: *Stenotrophomonas maltophilia*, *Kineococcus radiotolerans*, especies de *Chloroflexus* y diversas especies de *Micrococcus*, incluyendo la *Micrococcus luteus* recientemente secuenciada (Tornabene *et al.*, citado anteriormente, Suen *et al.*, citado anteriormente, Morrison *et al.*, citado anteriormente; van der Meer *et al.*, citado anteriormente; Albro *et al.*, citado anteriormente; Philips *et al.*, citado anteriormente). Para confirmar que otros organismos que contienen OleA, OleB, OleC y OleD producirán de manera natural olefinas, se seleccionó *Arthrobacter aureescens* TC1 (ATCC BAA-1386) de la tabla 1 para someter a prueba la producción de olefinas. No se ha notificado en la bibliografía que *Arthrobacter aureescens* TC1 produzca olefinas.

Se hizo crecer *Arthrobacter aureescens* TC1 en 100 ml de caldo de *Corynebacterium* que se compone de 10 g de peptona de caseína (digesto trípico), 5 g de glucosa, 5 g de extracto de levadura y 5 g de NaCl por litro en un frasco de 500 ml con agitación a 30°C durante 48 horas. Tras 48 horas, se extrajeron 5 ml del caldo según un método de extracción 1 modificado. Específicamente, tras la sedimentación inicial del cultivo celular, se resuspendió el sedimento en 1 ml de agua y volvió a centrifugarse para eliminar el caldo de cultivo residual. Se resuspendió el sedimento final en 100 µl de agua. Se siguió el resto del método de extracción 1 (véase el ejemplo 1). Se analizó el extracto mediante CG/EM usando el método de detección 1. Las olefinas detectadas mediante CG/EM fueron olefina monoinsaturada C29 (véase la figura 19A-B).

Los resultados reflejados en este ejemplo demuestran que organismos con secuencias de proteína para OleA, OleB, OleC y OleD producen olefinas. Además, este ejemplo demuestra que puede usarse bioinformática para identificar organismos que producen olefinas identificando organismos con OleA, OleB, OleC y OleD en su secuencia de proteínas. Más específicamente, estos ejemplos demuestran que *Arthrobacter aureescens* TC1, que tiene las secuencias de proteínas OleA, OleB, OleC y OleD, produce olefina monoinsaturada C29.

EJEMPLO 19

Este ejemplo demuestra que las olefinas de base biológica pueden distinguirse de las olefinas a base de petróleo. En particular, este ejemplo demuestra que las olefinas producidas por organismos, tales como *E. coli* que expresa los genes *oleA*, *oleB*, *oleC* y *oleD*, pueden distinguirse de las olefinas a base de petróleo.

Se compararon olefinas de base biológica purificadas producidas mediante fermentación de *E. coli* C41(DE3); pETDuet-1_OleAB; pCOLADuet-1_OleCD con una olefina a base de petróleo, (Z)-9-tricoseno (97%, Sigma-Aldrich, número CAS: 27519-02-4), usando técnicas de datación de carbono convencionales. Se enviaron las muestras a Beta Analytic, Inc. (Miami, FL, EE.UU.) para las pruebas. Se midió el contenido de carbono de base biológica de las muestras usando espectrometría de masas con acelerador (AMS) basándose en la norma ASTM 06866. La AMS mide la razón de isótopos de carbono de ¹³C con respecto a ¹²C y ¹⁴C en grafito derivado de los polvos de muestra. Entonces se calculó el contenido de carbono de base biológica a partir de la razón de isótopos de carbono de ¹³C con respecto a ¹²C y la cantidad de ¹⁴C presente.

Se produjeron las muestras de olefina analizadas y se purificaron usando el siguiente método. Se hizo crecer *E. coli* C41(DE3); pETDuet-1_OleAB; pCOLADuet-1_OleCD en 5 ml de medio LB complementado con carbenicilina 100 mg/l y kanamicina 50 mg/l durante 12 horas. Se usó este cultivo de 12 horas para inocular un cultivo mayor compuesto por 200 ml de medio de frasco de agitación F1 que contenía: KH₂PO₄ 3 g/l, K₂HPO₄ 6,62 g/l, (NH₄)₂SO₄ 4 g/l, MgSO₄ 0,15 g/l, glucosa 5 g/l (dextrosa DX0145-5 EMD Chemicals, Inc. NJ), 1,25 ml/l de disolución de oligoelementos y 1,25 ml/l de disolución de vitaminas traza complementado con carbenicilina 100 mg/l y kanamicina 50 mg/l. La disolución de oligoelementos contenía por litro: 27 g de FeCl₃·6H₂O, 2 g de ZnCl₂·4H₂O, 2 g de CaCl₂·6H₂O, 2 g de Na₂MoO₄·2H₂O, 1,9 g de CuSO₄·5H₂O, 0,5 g de H₃BO₃ y 100 ml de HCl concentrado. La disolución de vitaminas traza contenía por litro: 0,42 g de riboflavina, 5,4 g de ácido pantoténico, 6 g de niacina, 1,4 g de piridoxina, 0,06 g de biotina y 0,04 g de ácido fólico. Se usaron 50 ml de cultivo simiente para inocular un biorreactor Biostat Aplus de 2 l (Sartorius BBI) que contenía inicialmente 1 l de medio de fermentación F1 esterilizado. El medio de fermentación F1 esterilizado contenía: KH₂PO₄ 1,5 g/l, K₂HPO₄ trihidratado 4,34 g/l, (NH₄)₂SO₄ 4 g/l, MgSO₄ heptahidratado 0,150 g/l, glucosa filtrada estéril 5 g/l (dextrosa EMD Chemicals, Inc. NJ), 1,25 ml/l de disolución de oligoelementos, 1,25 ml/l de disolución de vitaminas traza y antibióticos a la misma concentración que se utilizó en el frasco de agitación. Se mantuvo el pH del cultivo a 7,2 usando H₂SO₄ 1 M y NH₄OH al 30% p/v. Se mantuvo la temperatura a 37°C, la velocidad de aireación a 2 lpm (2 v/v/m) y la tensión de oxígeno disuelto al 30% de saturación, utilizando un bucle de agitación en cascada al controlador de DO. Se controló la formación de espuma mediante la adición automatizada de una disolución esterilizada en autoclave de Antifoam 204 (Sigma-Aldrich St. Louis, MO). Se proporcionó una alimentación de nutrientes compuesta por (NH₄)₂SO₄ 60 g/l, MgSO₄ heptahidratado 3,9 g/l, glucosa 430 g/l, 10 ml/l de la disolución de oligoelementos y 10 ml/l de la disolución de vitaminas en condiciones de fermentación aerobias durante 48 horas en el fermentador. La fuente de azúcar de glucosa para las fermentaciones se derivó de maíz.

Se proporcionó una alimentación de nutrientes compuesta por $(\text{NH}_4)_2\text{SO}_4$ 60 g/l, MgSO_4 heptahidratado 3,9 g/l, glucosa 430 g/l, 10 ml/l de la disolución de oligoelementos y 10 ml/l de la disolución de vitaminas cuando se agotó el medio inicial (aproximadamente 6 horas tras la inoculación). Se aumentó o disminuyó gradualmente la alimentación de nutrientes durante la fermentación para mantener un nivel de glucosa residual en el medio de menos de 10 g/l.

5 Se indujo la producción de olefinas en el biorreactor cuando el cultivo logró una DO_{600} de 30 UA (aproximadamente 8-9 horas tras la inoculación) mediante la adición de una disolución madre de IPTG 1 M hasta una concentración final de 1 mM. También se disminuyó la temperatura de la fermentación en la inducción hasta 30°C. Se cosechó el biorreactor aproximadamente 48 horas tras la inducción.

10 Tras 48 horas, se centrifugaron los cultivos de fermentación a 3500 rpm durante 20 minutos (centrífuga Allegra X-15R con rotor SX-4750A, Beckman Coulter, Fullerton, CA). Se resuspendió el sedimento celular en 50 ml de agua destilada estéril. Se añadieron 100 ml de metanol a la muestra que entonces se sonicó en un baño de agua de sonicación durante 60 minutos. Se transfirió la mezcla a un embudo de separación de 1 l en donde se añadió n-hexano hasta un volumen final de 750 ml. Se mezcló la muestra bien mediante agitación suave, tras lo cual la mezcla se asentó para separar la fase de hexano de la fase acuosa. Una vez que se formó la bicapa, se extrajo mediante sifón la fase acuosa inferior del embudo de separación. Se añadió Na_2SO_4 a la fase orgánica restante para eliminar cualquier H_2O en exceso en la fase orgánica. Entonces se filtró la fase orgánica dos veces a través de un filtro de 150 mm Whatman n.º 4 sin aplicar vacío. Se transfirió la fase orgánica filtrada dos veces a un matraz de fondo redondo conectado a una columna de destilación. Se eliminaron el hexano y la acetona mediante destilación con la temperatura fijada a 56°C seguido por 66-68°C. La muestra restante era de color negro. Para purificar adicionalmente la muestra, se filtró la muestra a través de una columna gel de sílice 60 de 200 ml (tamaño de partícula 0,063 - 0,300 mm, 70-230 de malla ASTM). Se retiró la muestra de la columna mediante elución con hexano. Se eliminó el hexano del eluyente que contenía olefina mediante un evaporador rotatorio a temperatura ambiental. Se usó Magnesol D-sol (Dallas Group of American, Inc) en el procedimiento de purificación final. Se resuspendió la muestra en metil terc-butil éter (MTBE) hasta un volumen final de 8 ml. Se añadió magnesol a la muestra en una razón 10:1 de magnesol frente a muestra (p/v). Se mezcló la mezcla de muestra-magnesol durante 1 hora a 37°C en un agitador rotatorio. Se filtró la mezcla a través de un filtro de PTFE de 0,2 μm al interior de un vial de centelleo de 20 ml. Se lavó el sólido restante con 2,5 ml de MTBE. Se combinaron el lavado y la mezcla filtrada en un vial de centelleo. Se evaporó el MTBE en la campana extractora para productos químicos a temperatura ambiental.

30 Se determinó ^{14}C por AMS en Beta Analytic, Inc., usando la norma ASTM D6866 método B. Se obtuvo el contenido de carbono de base biológica derivando la razón de ^{14}C en la muestra con respecto al de un patrón de referencia moderno. La razón se notifica como porcentaje con las unidades 'pMC' (porcentaje de carbono moderno).

35 El patrón de referencia moderno usado en la datación de radiocarbono es un patrón de NIST (National Institute of Standards and Technology) con un contenido en radiocarbono conocido equivalente a aproximadamente el año 1950 DC. Se eligió 1950 DC puesto que representa un tiempo antes de que las pruebas con armas termonucleares introdujeran grandes cantidades de ^{14}C en la atmósfera con cada prueba (el ^{14}C en exceso se conoce como "carbono de bombas"). Para un arqueólogo o geólogo que usa dataciones de ^{14}C , 1950 DC es igual "cero años de edad". Esto también representa 100 pMC. El "carbono de bombas" en la atmósfera alcanzó casi el doble del nivel normal en 1963 en el máximo de las armas termonucleares. Su distribución dentro de la atmósfera se ha aproximado desde su aparición, mostrando valores que son superiores a 100 pMC para plantas y animales vivos desde 1950 DC. Ha disminuido gradualmente a lo largo del tiempo, siendo el valor actual próximo a 107,5 pMC. Esto significa que un material de biomasa reciente, tal como maíz, tendrá una firma de ^{14}C próxima a 107,5 pMC. Los compuestos a base de petróleo tendrán un valor de pMC de cero. Combinar carbono fósil con carbono actual dará como resultado una dilución del contenido de pMC actual. Suponiendo que 107,5 pMC representa el contenido de ^{14}C de los materiales de biomasa actuales y 0 pMC representa el contenido de ^{14}C de los productos a base de petróleo, el valor de pMC medido para ese material reflejará las proporciones de los dos tipos de componentes. Por ejemplo, un material derivado al 100% de semillas de soja actuales dará una firma de radiocarbono próxima a 107,5 pMC. Si ese material se diluyó al 50% con productos a base de petróleo, daría una firma de radiocarbono de aproximadamente 54 pMC.

50 Un contenido de carbono de base biológica se deriva asignando el 100% igual a 107,5 pMC y el 0% igual a 0 pMC. Por ejemplo, una muestra que mide 99 pMC dará un contenido de carbono de base biológica equivalente del 93%. Este valor se denomina el resultado de carbono de base biológica medio y supone que todos los componentes dentro del material analizado se originaron o bien de material biológico actual o bien de material a base de petróleo.

55 La razón de isótopos de carbono estables ($^{13}\text{C}/^{12}\text{C}$) en un material derivado biológicamente dado es una consecuencia de la razón de $^{13}\text{C}/^{12}\text{C}$ en el dióxido de carbono atmosférico en el momento en el que se fija el dióxido de carbono. También refleja la ruta metabólica precisa de la planta que fija dióxido de carbono. La razón de $^{13}\text{C}/^{12}\text{C}$ se expresa normalmente como $\delta^{13}\text{C}$ que se calcula tal como sigue:

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

60 La $\delta^{13}\text{C}$ de petróleo y plantas C_4 (maíz, sorgo, etc.) tienen un intervalo similar (desde -7 hasta -13 ppt), mientras que $^{13}\text{C}/^{12}\text{C}$ de plantas C_3 (trigo, avenas, arroz, etc.) disminuye en un intervalo de desde -19 hasta -27 ppt (véanse, por

ejemplo, Stuiver *et al.*, citado anteriormente, y Gupta, *et al.*, Radiocarbon dating practices at ANU. Handbook, Radiocarbon Dating Laboratory, Research School of Pacific Studies, ANU, Canberra (1985)). La diferencia del valor de $\delta^{13}\text{C}$ para plantas C_4 y C_3 se debe al ciclo fosfosintético diferente de cada planta.

5 En la tabla 29 se enumeran los valores de pMC y $\delta^{13}\text{C}$ del producto de olefina de base biológica y el 9-tricoseno a base de petróleo.

TABLA 29

	pMC	$\delta^{13}\text{C}$ (‰)	Contenido de base biológica medio, %
Producto de olefina	94,2 ± 0,4	-12,3	88
Cis-9-tricoseno	0 ± 0	-30,5	0

10 Basándose en los resultados de ^{14}C , la olefina de fase biológica producida por *E. coli* tiene un contenido de carbono que se deriva al menos al 88% de carbono de base biológica. Es posible que el MTBE residual, un producto a base de petróleo usado durante la purificación, disminuya el contenido de carbono de base biológica en la muestra de olefina. Además, el valor de $\delta^{13}\text{C}$ del producto de olefina de base biológica proporciona pruebas adicionales de que el producto de olefina, y el carbono contenido en la misma, se deriva principalmente de una fuente biológica.

15 Los resultados de los experimentos reflejados en este ejemplo demuestran que los hidrocarburos de base biológica pueden distinguirse de los hidrocarburos a base de petróleo. En particular, este ejemplo demuestra que las olefinas producidas por organismos, tales como *E. coli* que expresa los genes *oleA*, *oleB*, *oleC* y *oleD*, pueden distinguirse de las olefinas a base de petróleo.

20 El uso de los términos “un” y “una” y “el/la” y referentes similares en el contexto de describir la invención (especialmente en el contexto de las siguientes reivindicaciones) debe interpretarse que cubre tanto el singular como el plural, a menos que se indique lo contrario en el presente documento o el contexto lo contradiga claramente. Los términos “que comprende”, “que tiene”, “que incluye” y “que contiene” han de interpretarse como términos de extremos abiertos (es decir, que significan “que incluye, pero no se limita a”) a menos que se indique lo contrario. La mención de intervalos de valores en el presente documento pretende meramente servir como método de abreviatura de la referencia individual a cada valor separado que se encuentra dentro del intervalo, a menos que se indique lo contrario en el presente documento, y cada valor separado se incorpora en la memoria descriptiva como si se mencionara individualmente en el presente documento. Todos los métodos descritos en el presente documento pueden realizarse en cualquier orden adecuado a menos que se indique lo contrario en el presente documento o el contexto lo contradiga claramente de otra forma. El uso todos y cada uno de los ejemplos, o expresiones a modo de ejemplo (por ejemplo, “tal como”) proporcionados en el presente documento, pretende meramente iluminar mejor la invención y no plantea una limitación sobre el alcance de la invención a menos que se reivindique lo contrario. Ninguna expresión en la memoria descriptiva debe interpretarse como que indica que cualquier elemento no reivindicado es esencial para la práctica de la invención.

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

Los números de registro a los que se hace referencia en el presente documento se derivan de la base de datos del NCBI (National Center for Biotechnology Information) mantenido por el National Institute of Health, EE.UU. Los números de registro son tal como se proporcionan en la base de datos el 15 de abril de 2007.

40 Los números de EC los establece el Comité de Nomenclatura de la Unión Internacional de Bioquímica y Biología Molecular (NC-IUBMB) (disponible en <http://www.chem.qmul.ac.uk/iubmb/enzyme/>). Los números de 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, patrocinada en parte por la Universidad de Tokyo. Los números de EC son tal como se proporcionan en la base de datos en abril de 2007.

45 En el presente documento se describen realizaciones preferidas de esta invención, incluyendo el mejor modo conocido por los inventores para llevar a cabo la invención. Pueden resultar evidentes variaciones de las realizaciones preferidas para los expertos habituales en la técnica tras la lectura de la descripción anterior.

REIVINDICACIONES

1. Célula modificada genéticamente para expresar una secuencia de ácido nucleico aislada que codifica para un polipéptido de OleA que comprende una secuencia de aminoácidos que tiene una identidad de secuencia de al menos el 85% con SEQ ID NO: 2, en la que la célula es una célula de levadura, una célula fúngica, una célula de animal no humano, una célula de insecto, una célula bacteriana o una célula de algas, y en la que el polipéptido de OleA está implicado en la biosíntesis de cetonas alifáticas u olefinas.
2. Célula según la reivindicación 1, en la que la secuencia de ácido nucleico codifica además para un polipéptido de OleC que comprende una secuencia de aminoácidos que tiene una identidad de secuencia de al menos el 85% con SEQ ID NO: 6, en la que el polipéptido de OleC está implicado en la biosíntesis de olefinas.
3. Célula según la reivindicación 1 ó 2, en la que la secuencia de ácido nucleico codifica además para un polipéptido de OleD que comprende una secuencia de aminoácidos que tiene una identidad de secuencia de al menos el 85% con SEQ ID NO: 8, en la que el polipéptido de OleD está implicado en la biosíntesis de olefinas.
4. Célula según una cualquiera de las reivindicaciones 1 a 3, en la que la secuencia de ácido nucleico codifica además para un polipéptido de OleB que comprende una secuencia de aminoácidos que tiene una identidad de secuencia de al menos el 85% con SEQ ID NO: 10, en la que el polipéptido de OleB está implicado en la biosíntesis de olefinas.
5. Célula según una cualquiera de las reivindicaciones 1 a 4, en la que la célula 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*, *Schizosaccharomyces*, *Stenotrophomonas*, *Kineococcus*, *Yarrowia* o *Streptomyces*.
6. Célula según una cualquiera de las reivindicaciones 1 a 5, en la que la célula comprende una mutación en un gen que codifica para una acil-CoA deshidrogenasa.
7. Célula según una cualquiera de las reivindicaciones 1 a 6, en la que la célula comprende adicionalmente una acil-CoA sintasa.
8. Célula según una cualquiera de las reivindicaciones 1 a 7, en la que la célula comprende adicionalmente una tioesterasa.
9. Método para producir una olefina que comprende cultivar la célula según la reivindicación 3 ó 4 con un sustrato en condiciones suficientes para producir dicho hidrocarburo.
10. Método para producir una cetona alifática que comprende cultivar la célula según la reivindicación 1 con un sustrato en condiciones suficientes para producir dicha cetona alifática.
11. Método para producir una cetona alifática que comprende incubar un sustrato con un polipéptido de OleA que comprende una secuencia de aminoácidos que tiene una identidad de secuencia de al menos el 85% con SEQ ID NO: 2 en condiciones suficientes para producir una cetona alifática.
12. Célula según la reivindicación 1, en la que los genes que codifican para OleC y/o OleD se delecionan o atenúan.
13. Método para producir una cetona alifática que comprende cultivar la célula según la reivindicación 12.
14. Método para identificar una enzima útil para la producción de olefinas que comprende
 - (a) transformar una célula que comprende polipéptidos seleccionados del grupo que consiste en
 - (i) OleA y OleD,
 - (ii) OleA y OleC,
 - (iii) OleA, OleB y OleD, y
 - (iv) OleA, OleC y OleD,
 con un ácido nucleico que codifica para una enzima que se sospecha que tiene la capacidad para producir olefinas; y
 - (b) determinar si la célula produce olefinas, en el que la existencia de producción de olefinas por la célula indica que el ácido nucleico codifica para un polipéptido útil para la producción de olefinas;

en el que el polipéptido de OleA comparte una identidad de secuencia de al menos el 85% con SEO ID NO: 2, el polipéptido de OleC comparte una identidad de secuencia de al menos el 85% con SEO ID NO: 6 y el polipéptido de OleD comparte una identidad de secuencia de al menos el 85% con SEO ID NO: 8, en el que el polipéptido de OleA, el polipéptido de OleC y el polipéptido de OleD están implicados en la biosíntesis de olefinas.

- 5
15. Polipéptido de OleA aislado que comprende la secuencia de aminoácidos de SEO ID NO: 2.
 16. Uso de un polipéptido de OleA aislado que comprende la secuencia de aminoácidos de SEO ID NO: 2 para producir cetonas alifáticas.

Figura 1

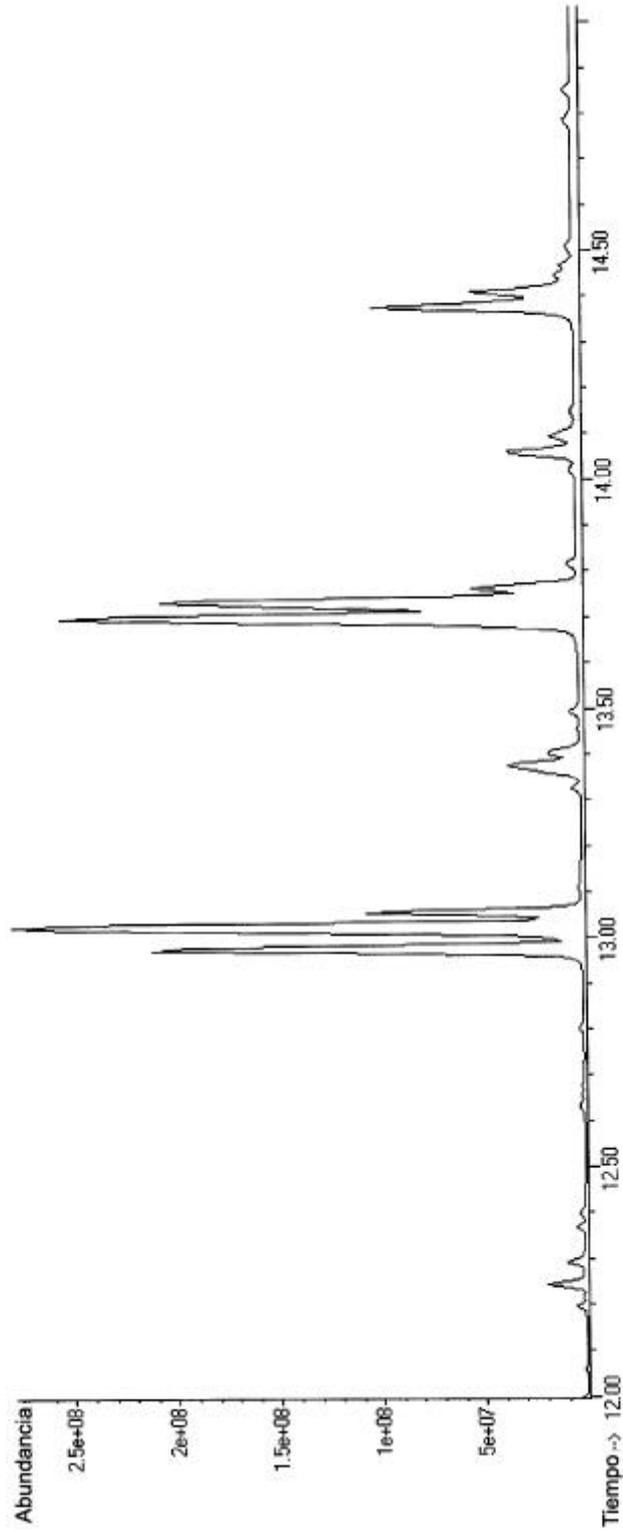


Figura 2A

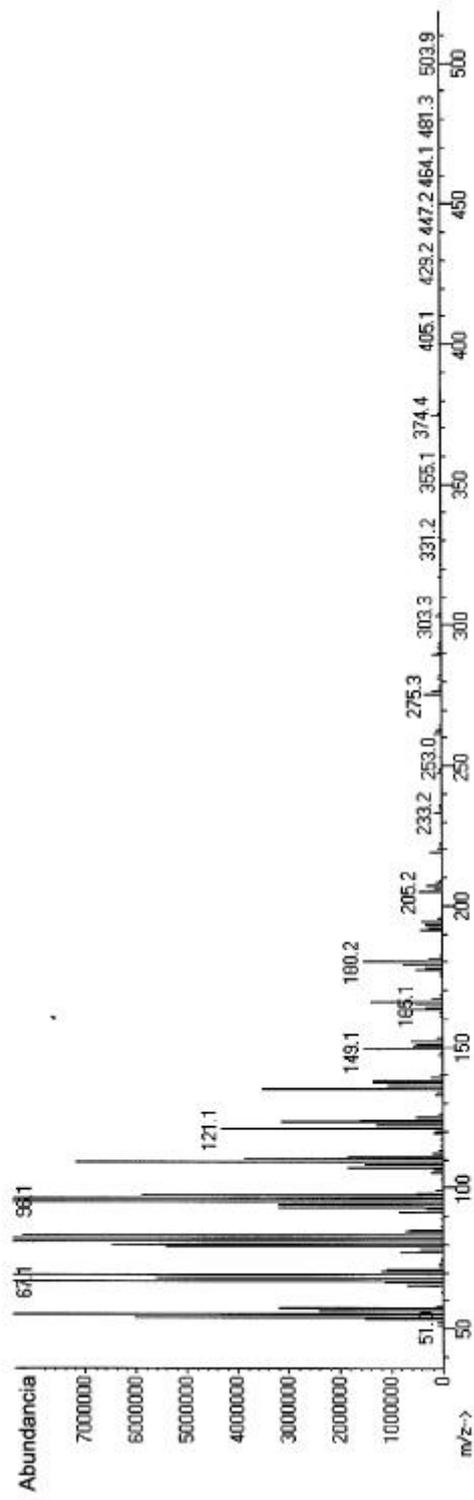


Figura 2B

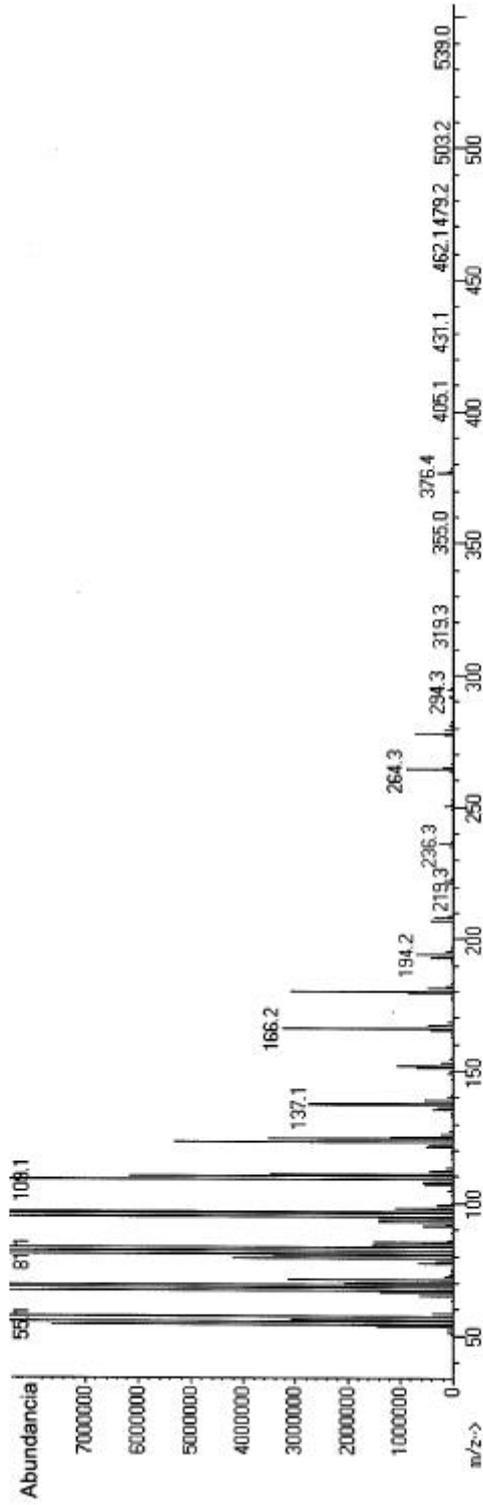


Figura 2C

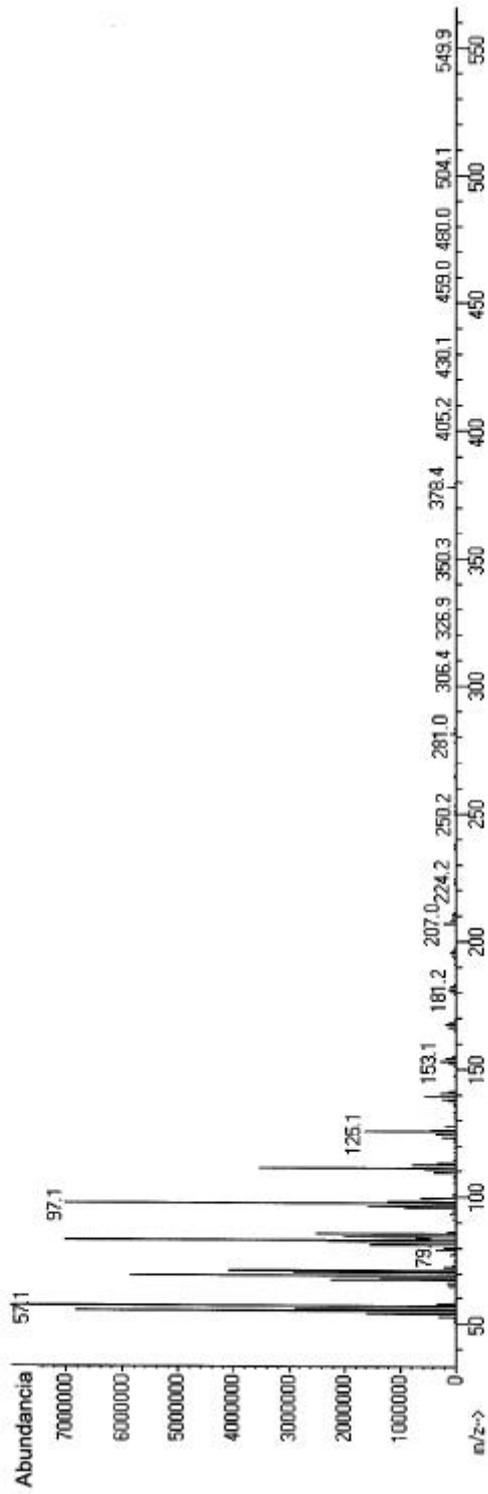


Figura 2D

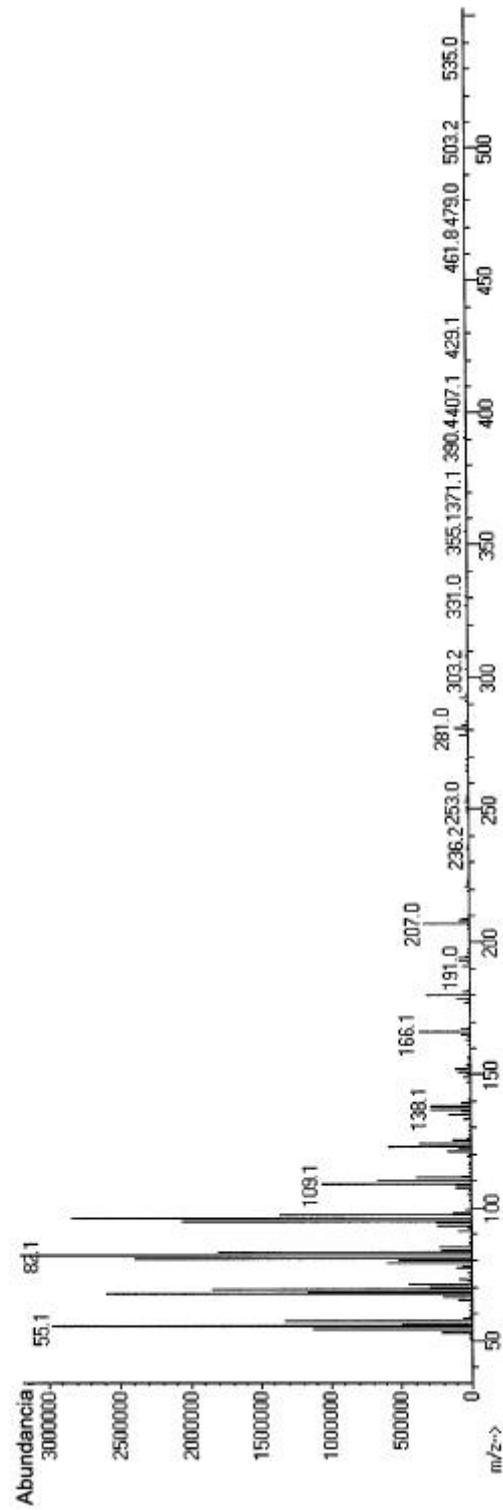


Figura 2E

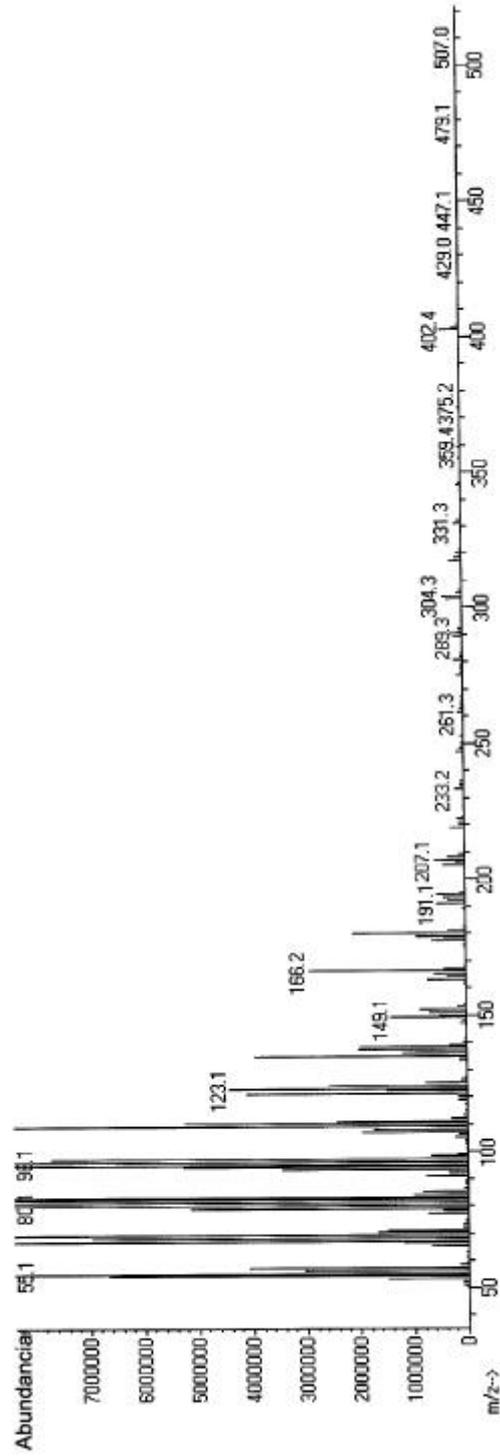


Figura 2F

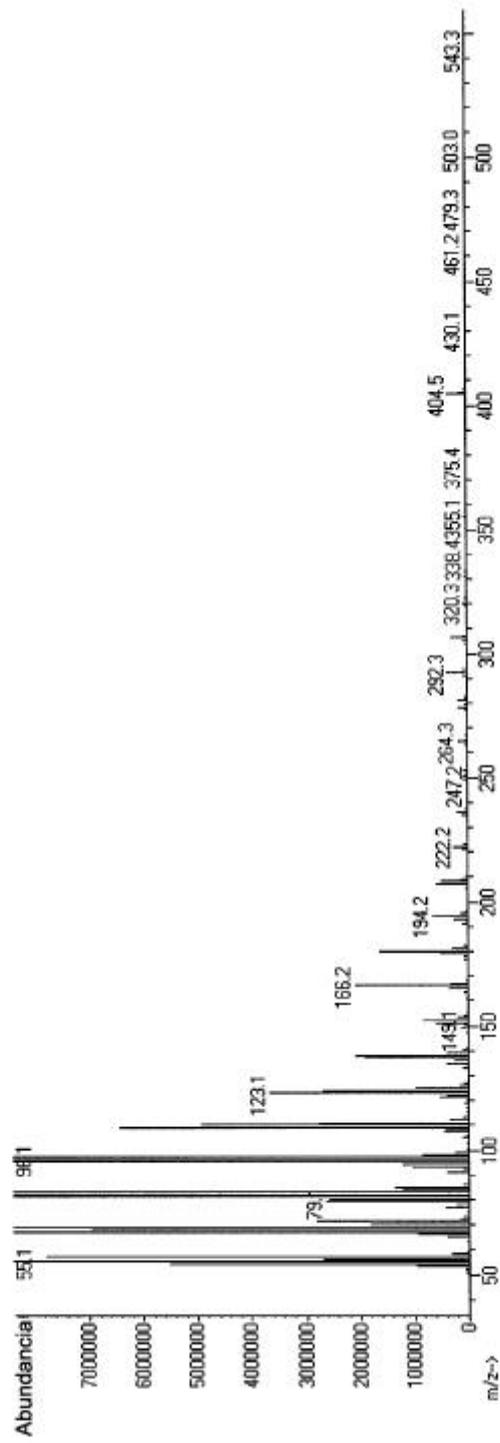


Figura 2G

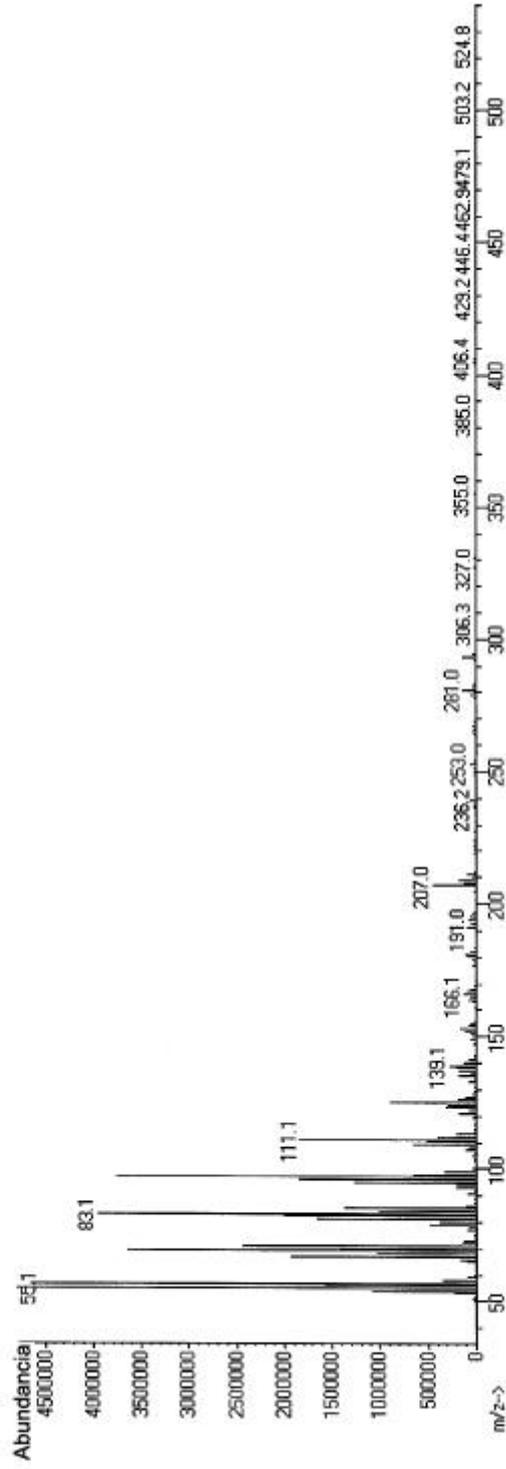


Figura 2H

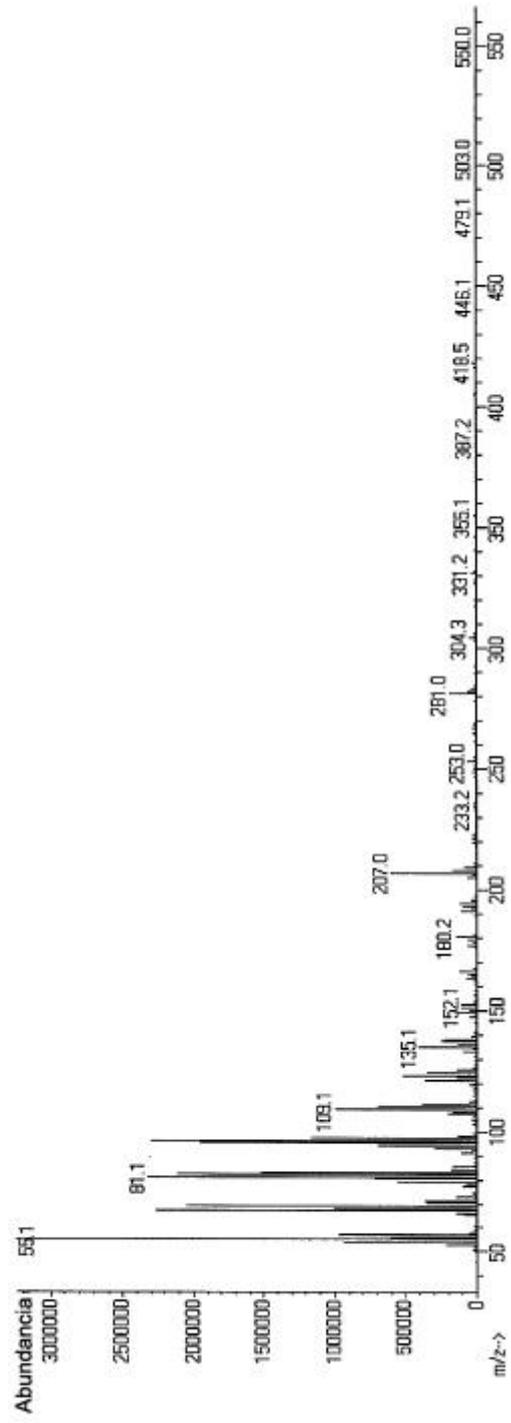


Figura 2I

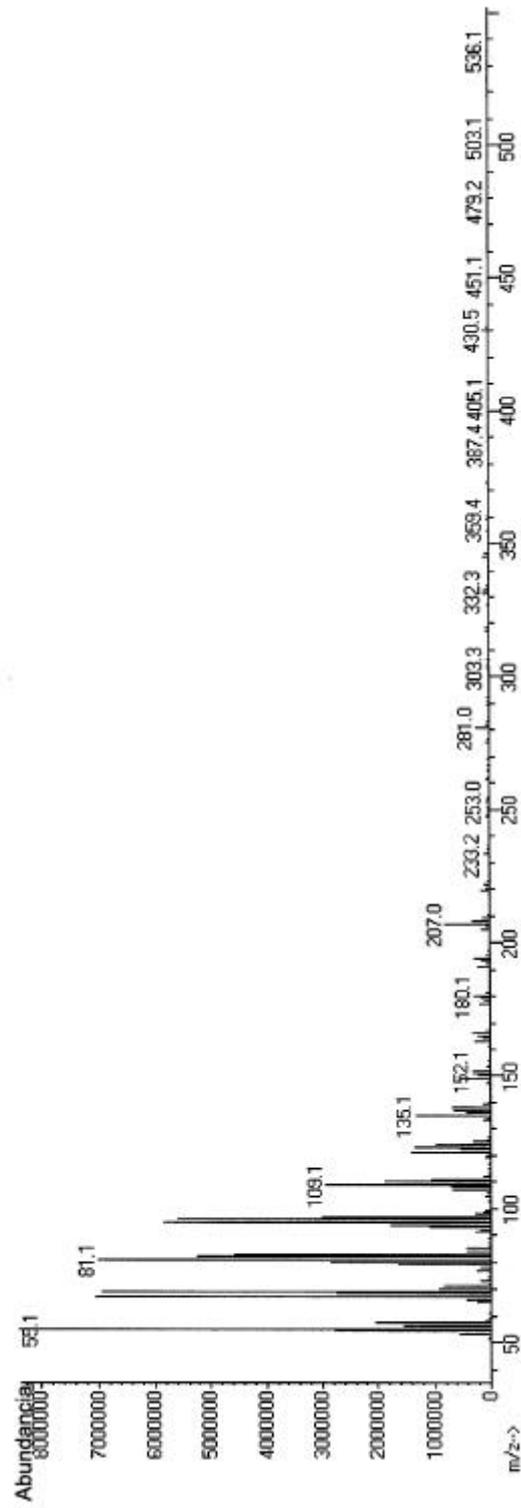


Figura 2J

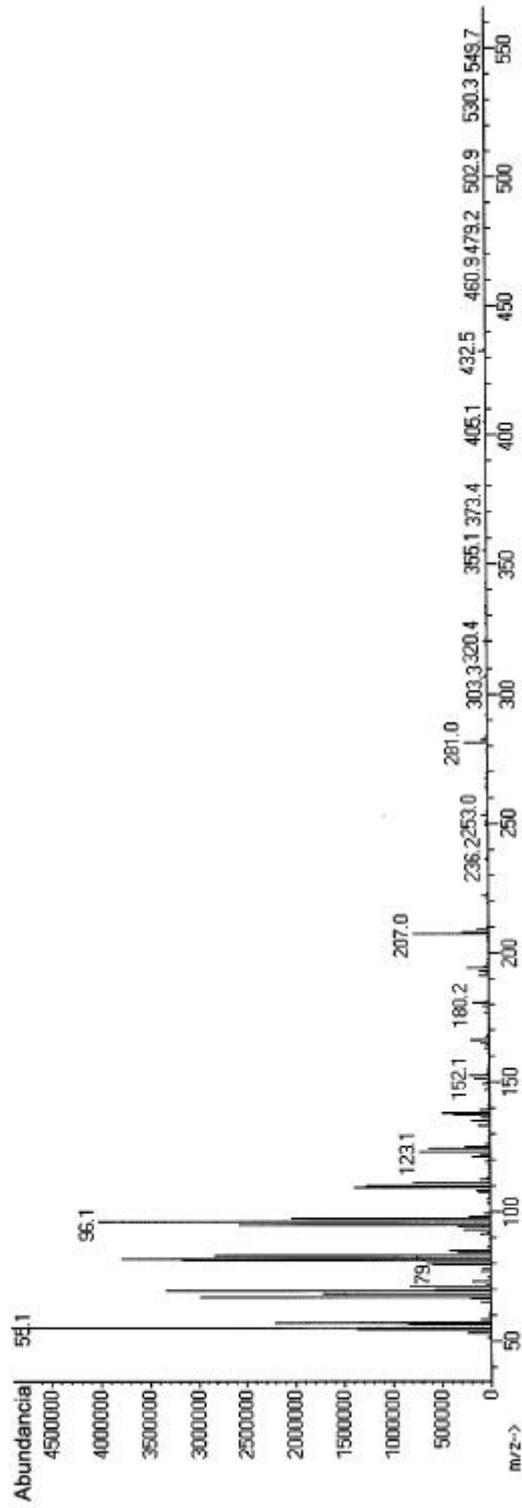


Figura 3

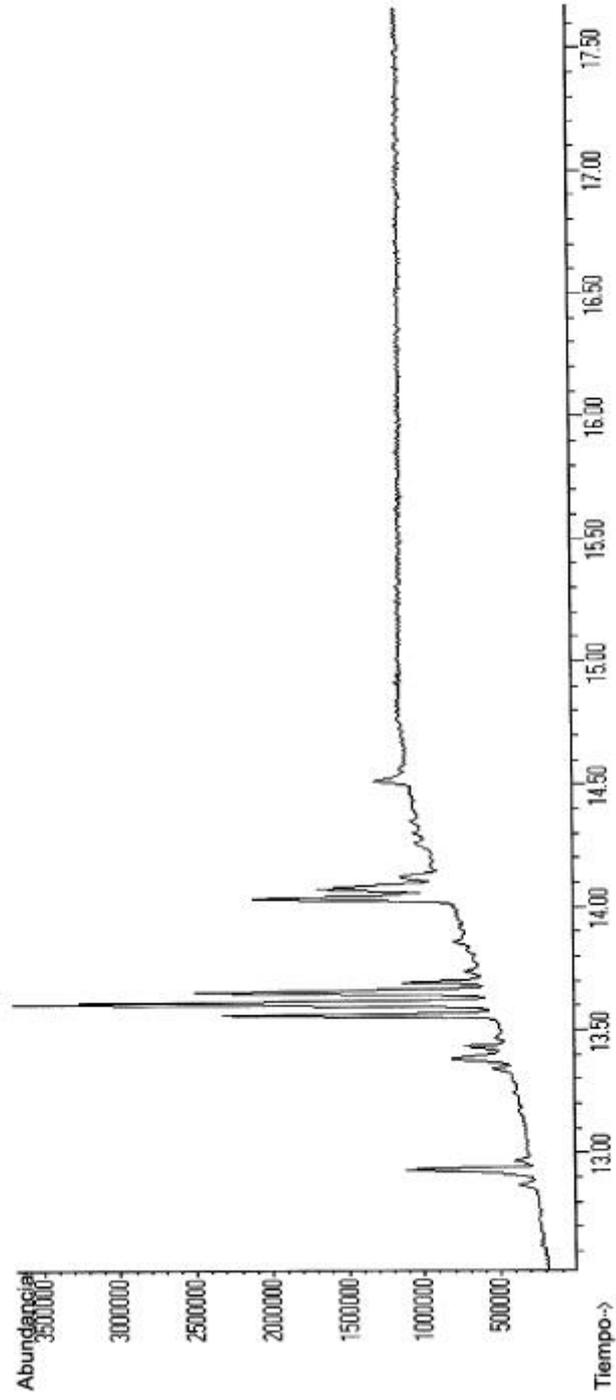


Figura 4A

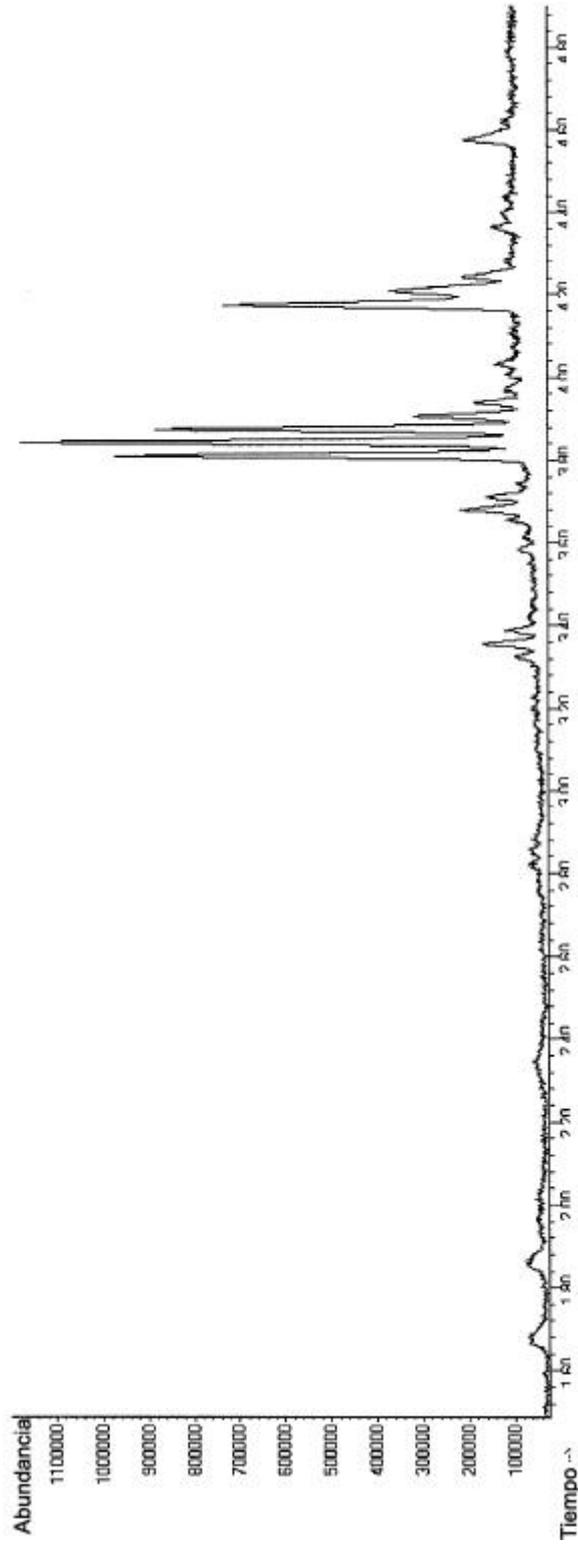


Figura 4B

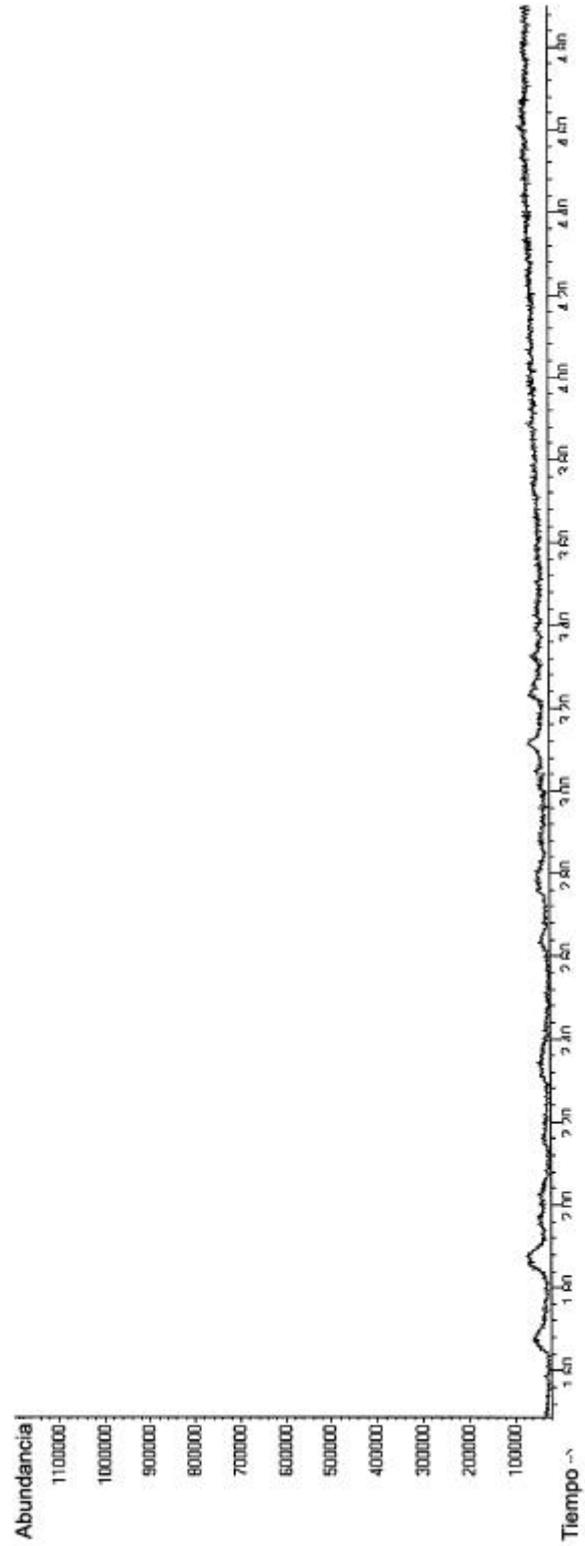


Figura 4C

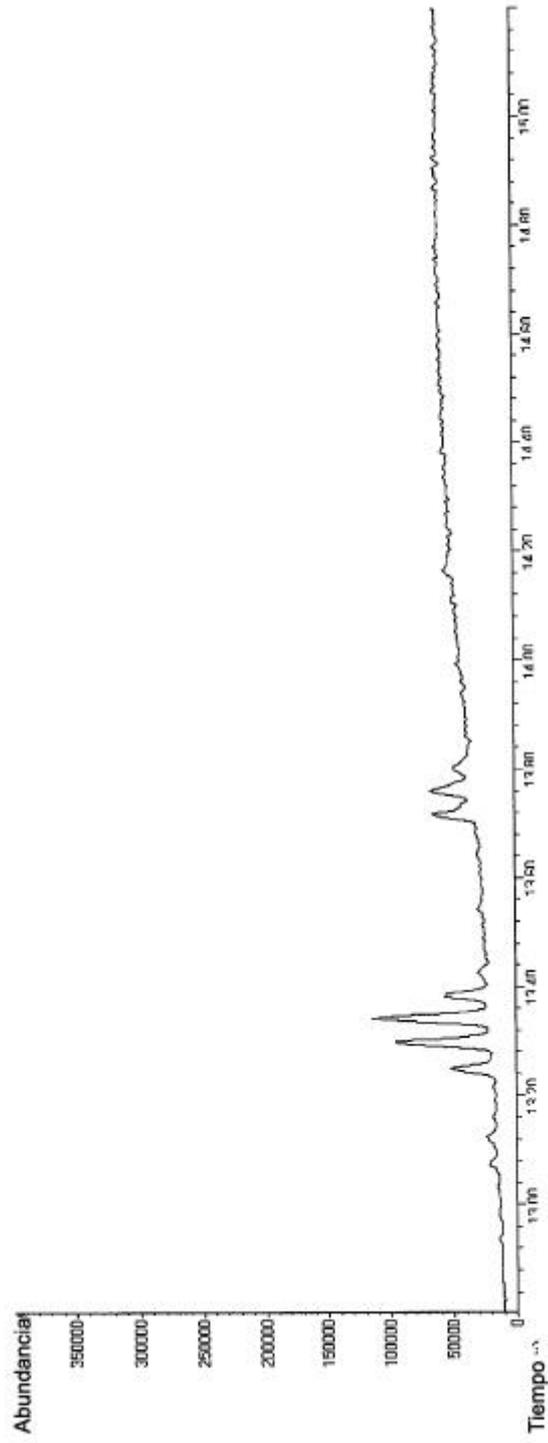


Figura 4D

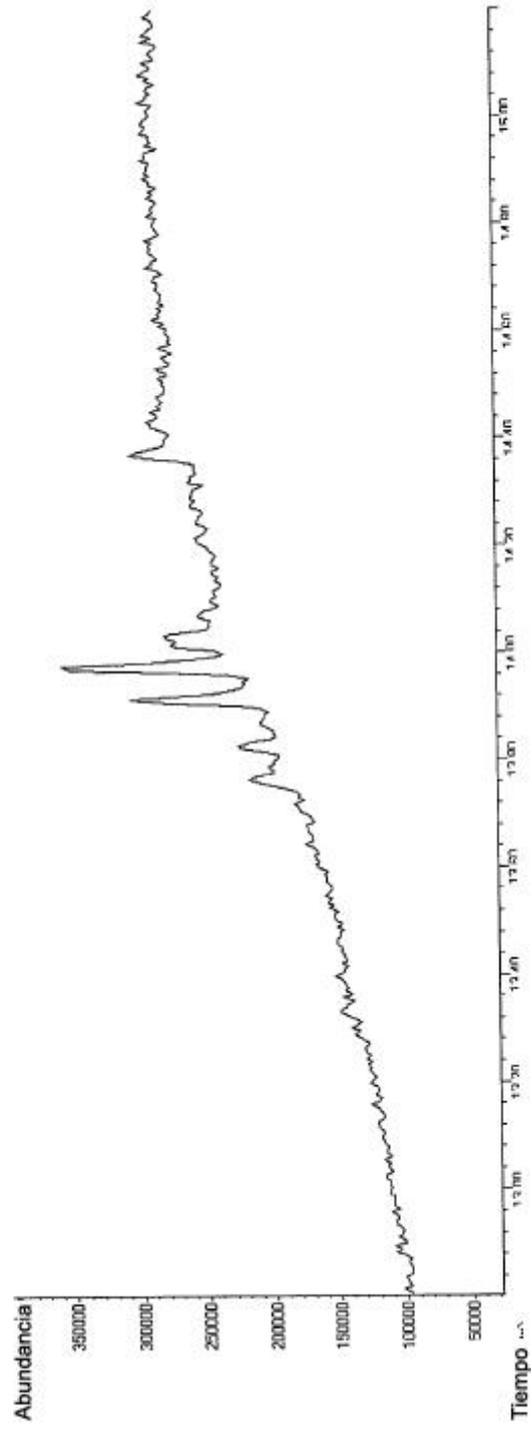


Figura 4E

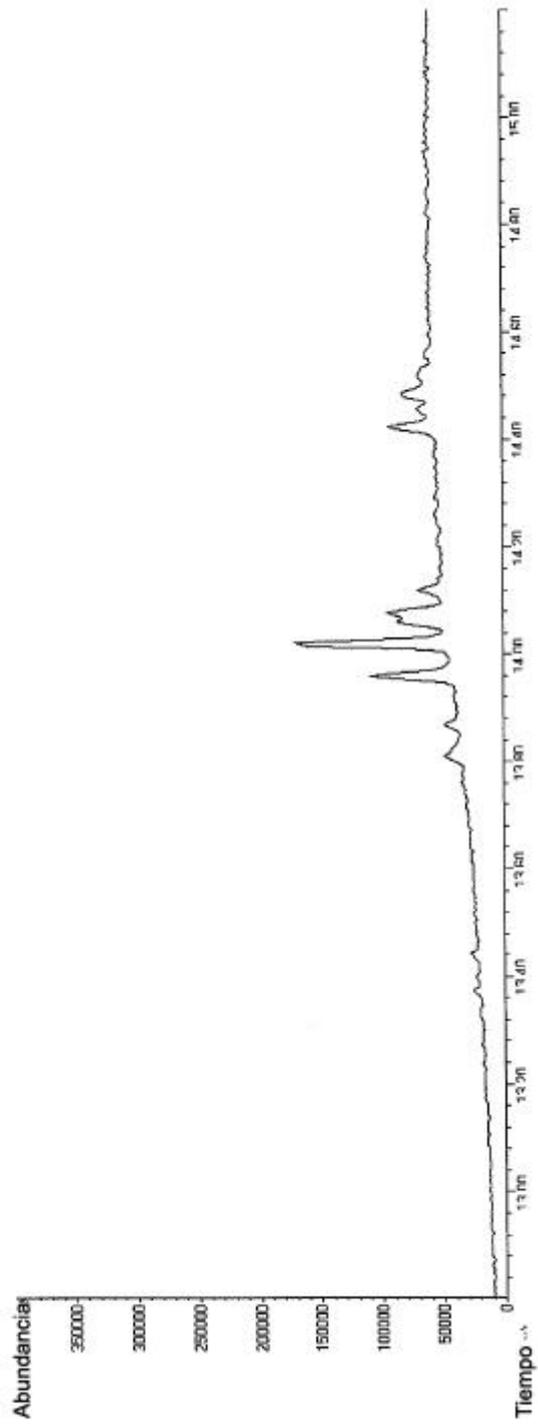


Figura 5A

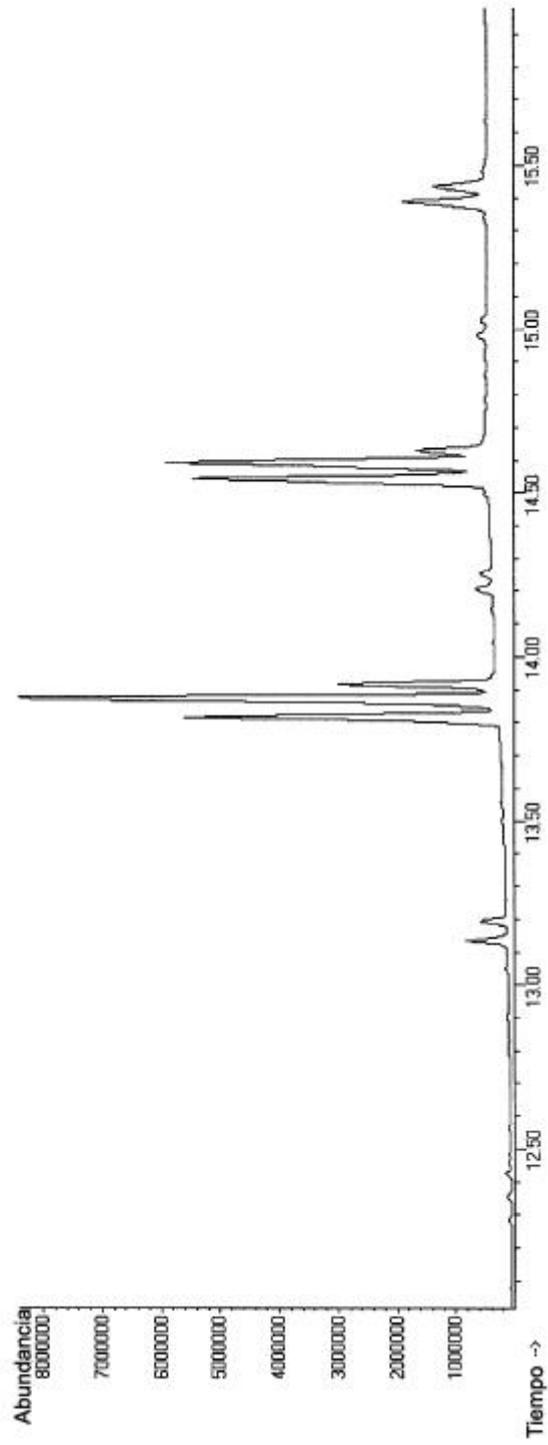


Figura 5B

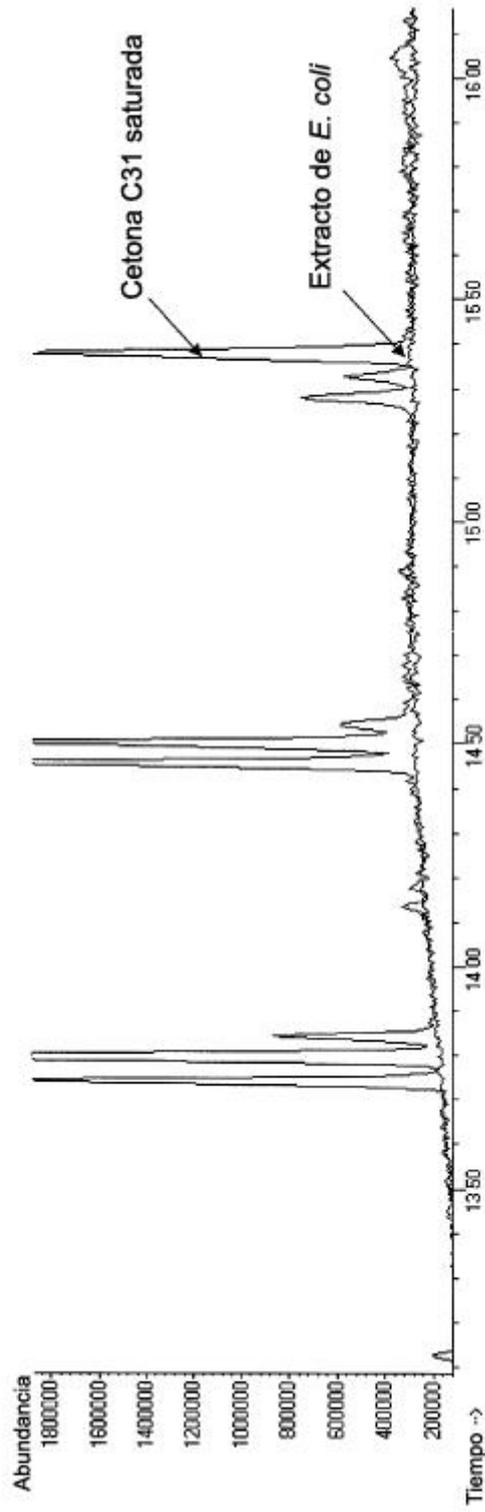


Figura 5C

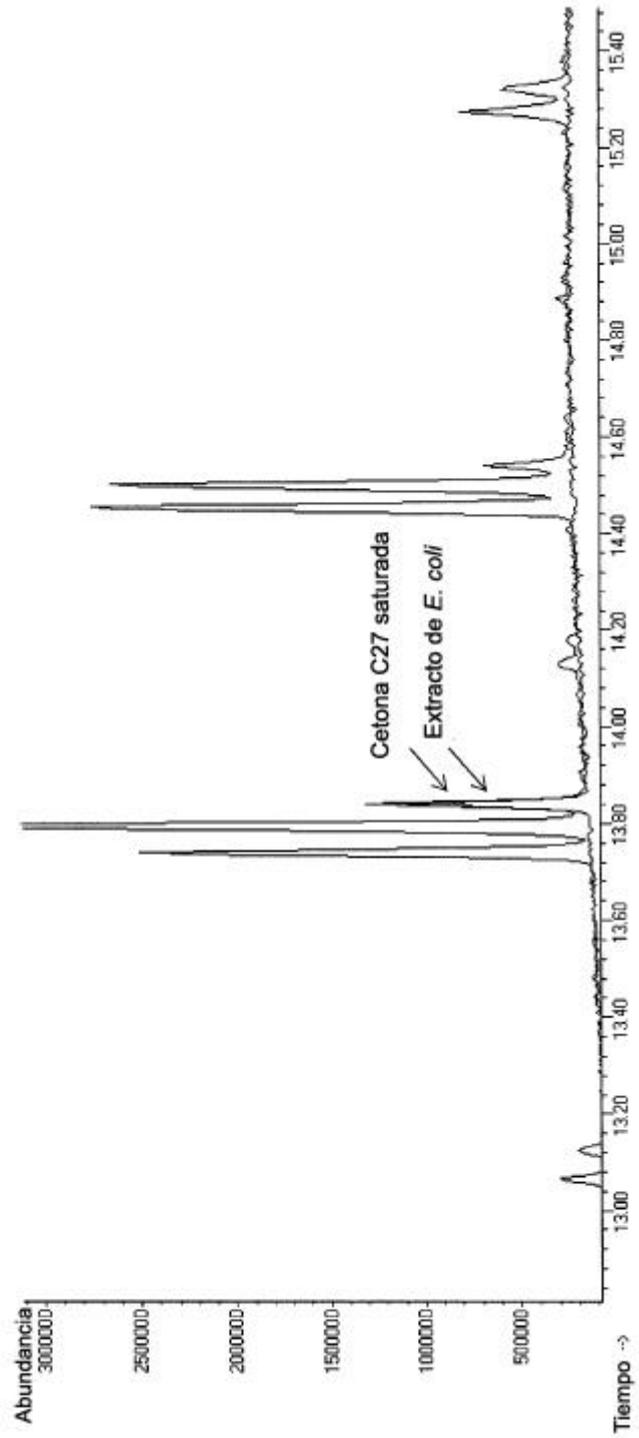


Figura 6A

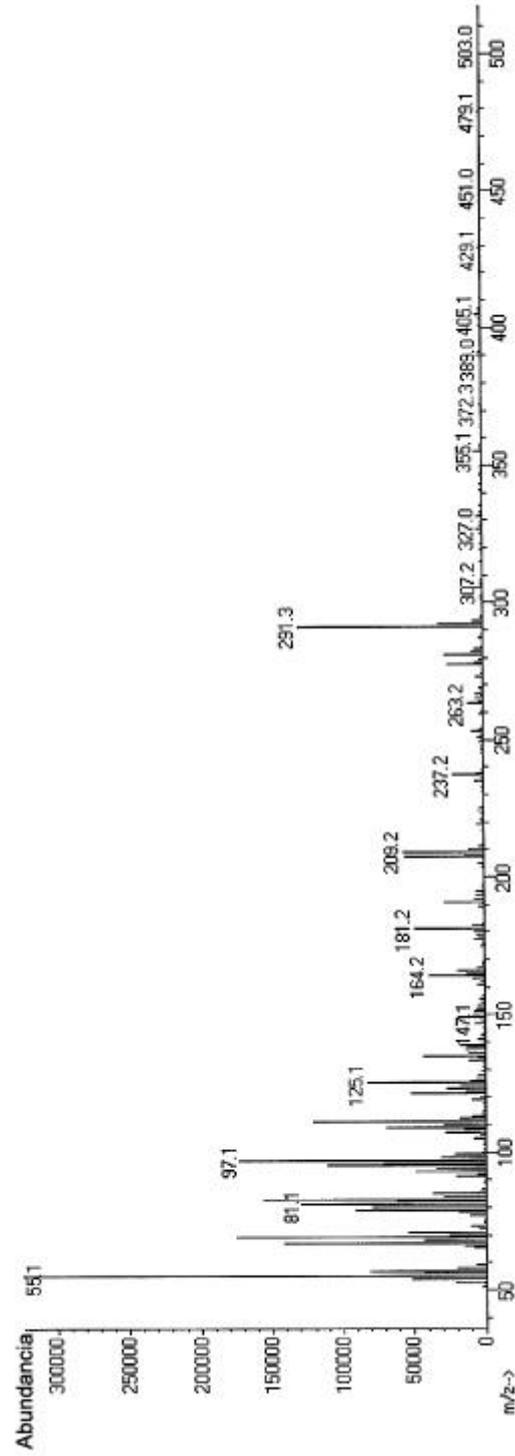


Figura 6B

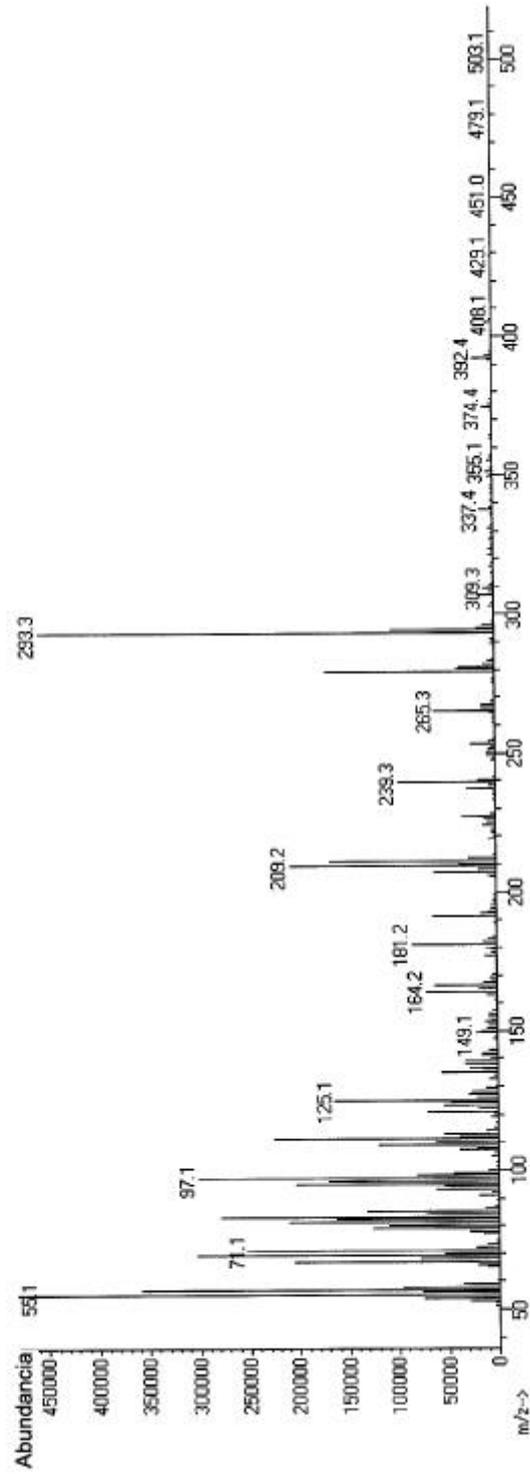


Figura 6C

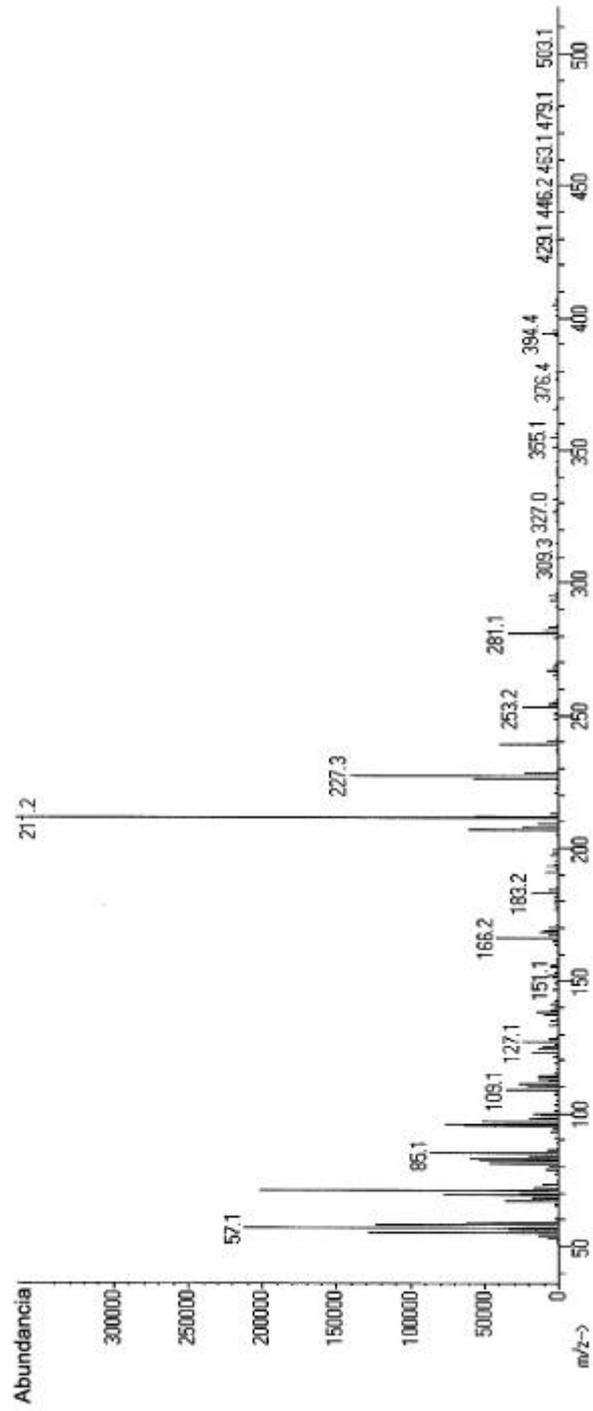


Figura 6D

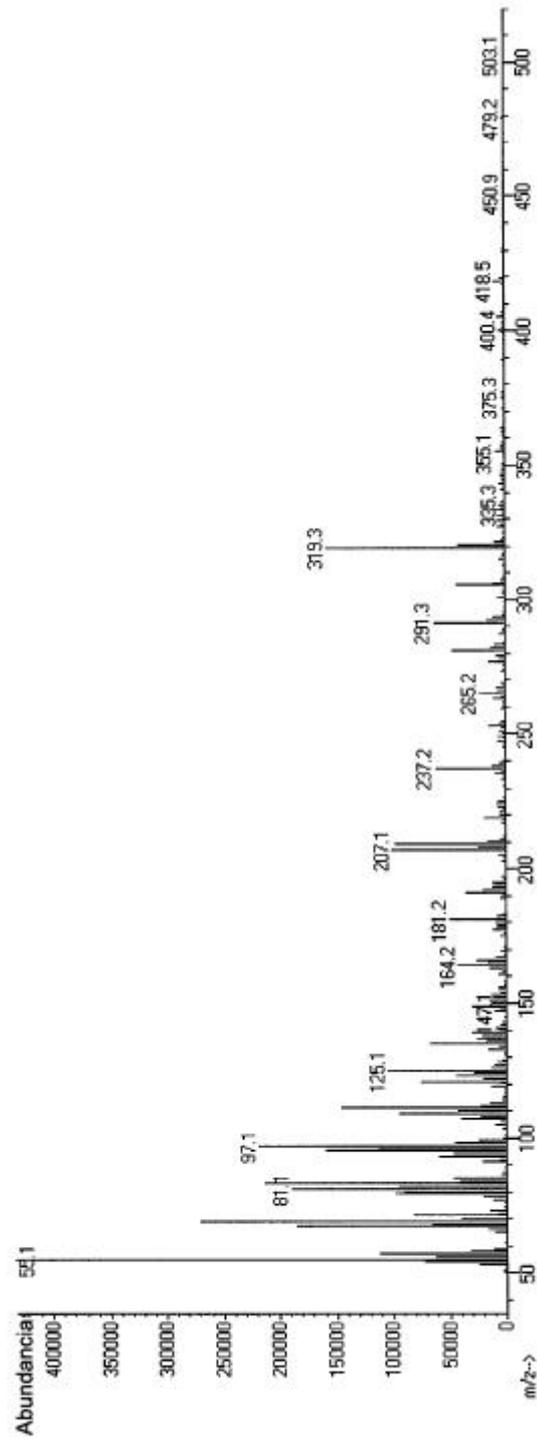


Figura 6E

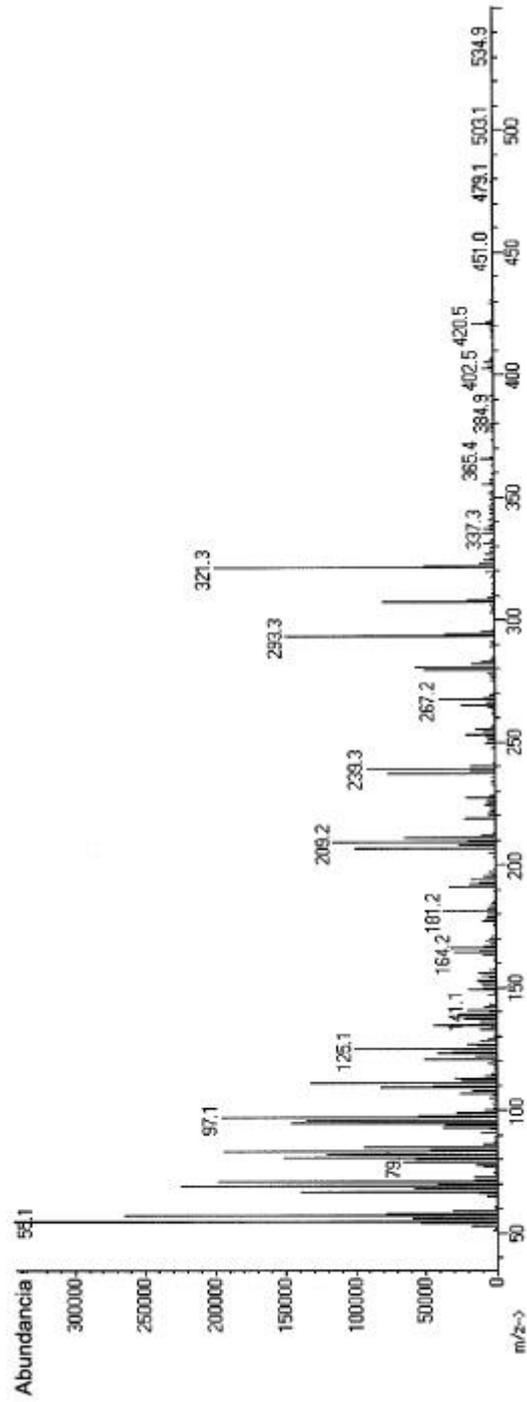


Figura 6F

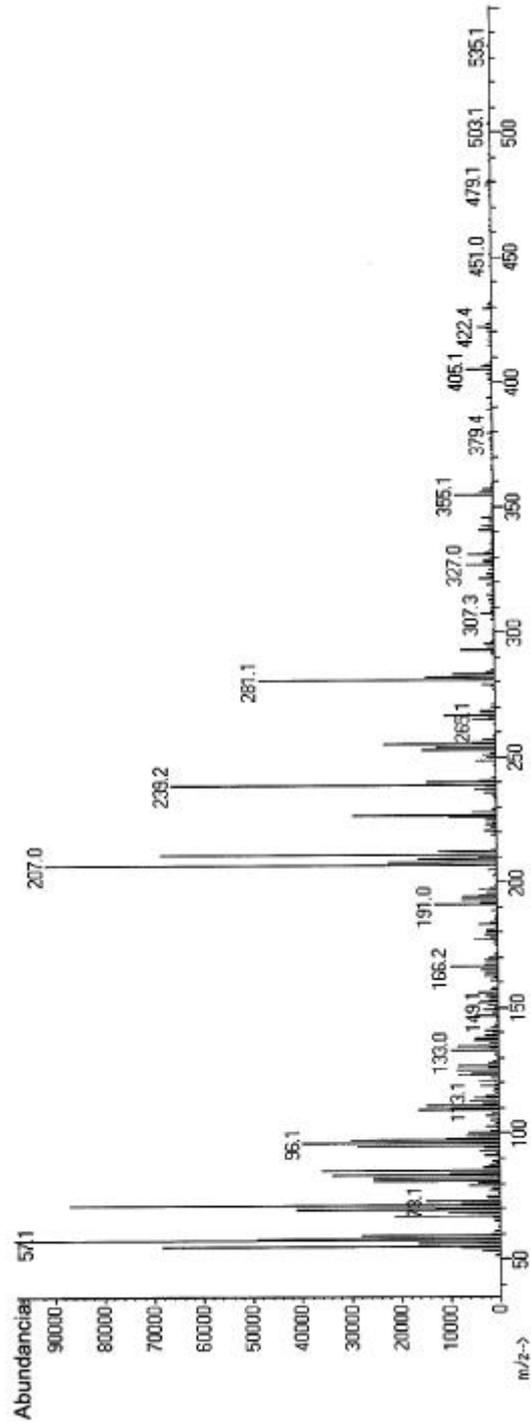


Figura 6G

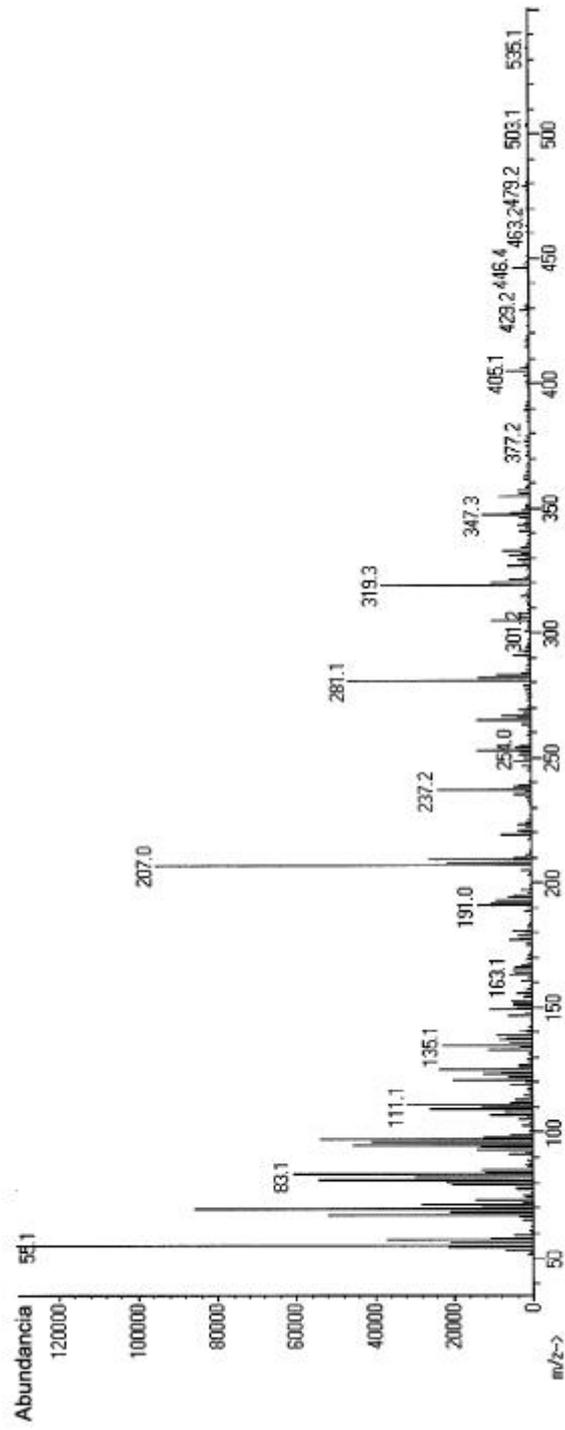


Figura 6H

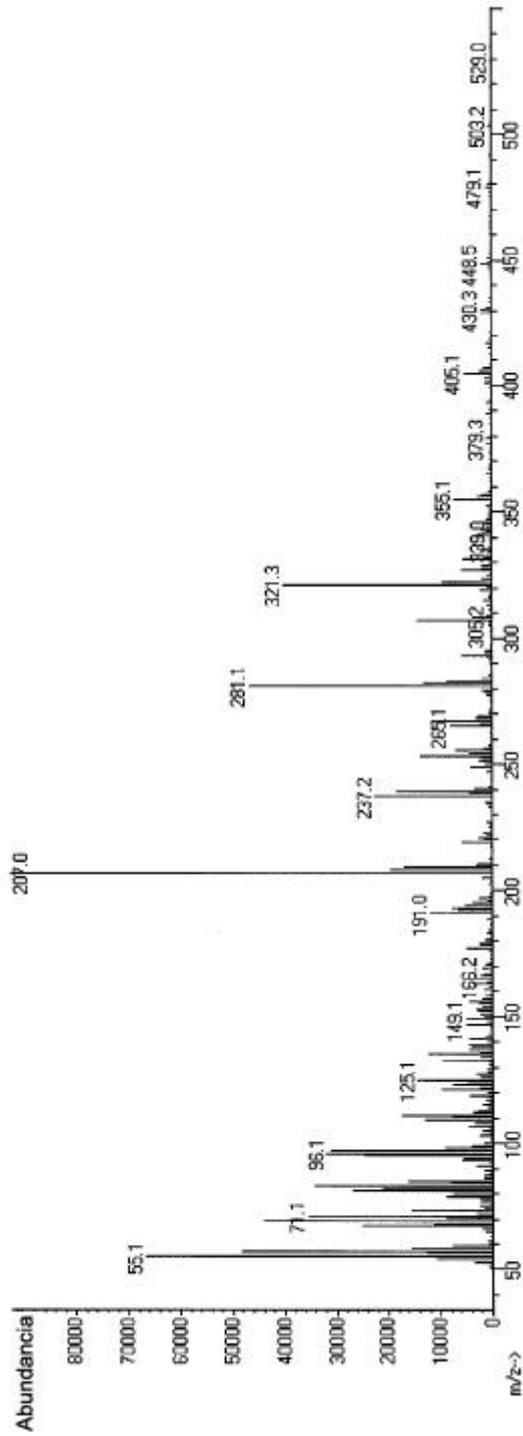


Figura 7A

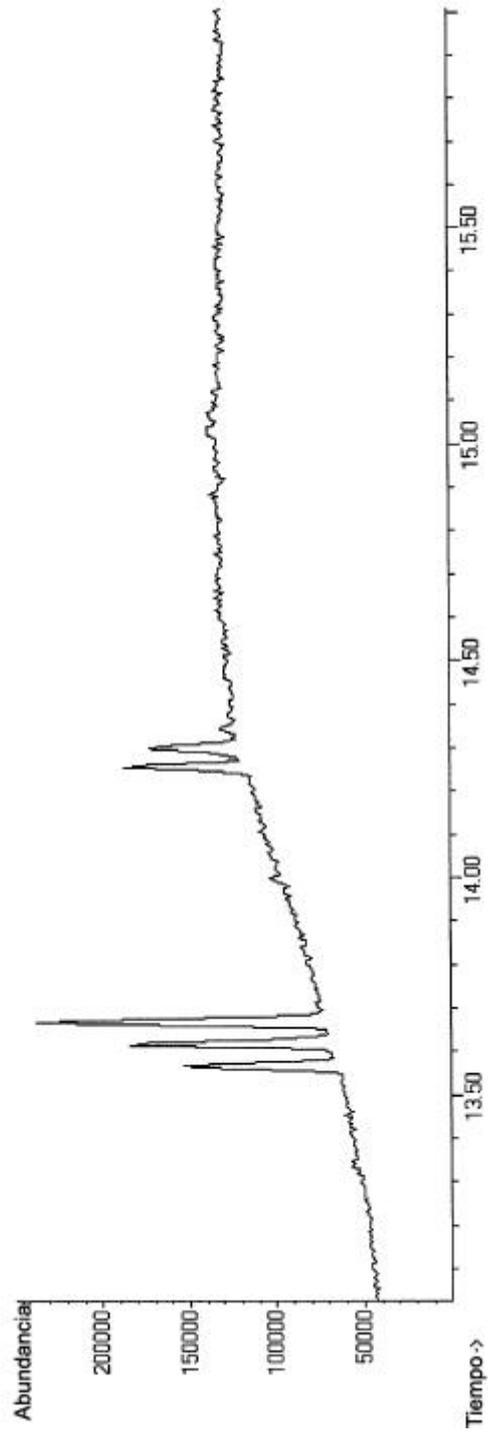


Figura 7B

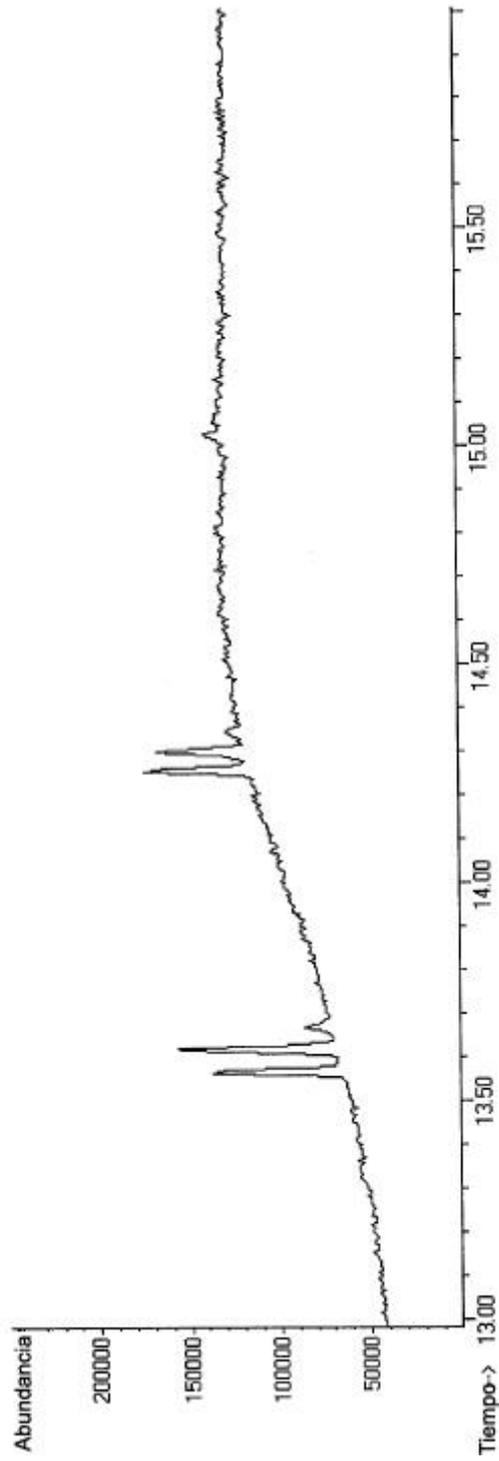


Figura 7C

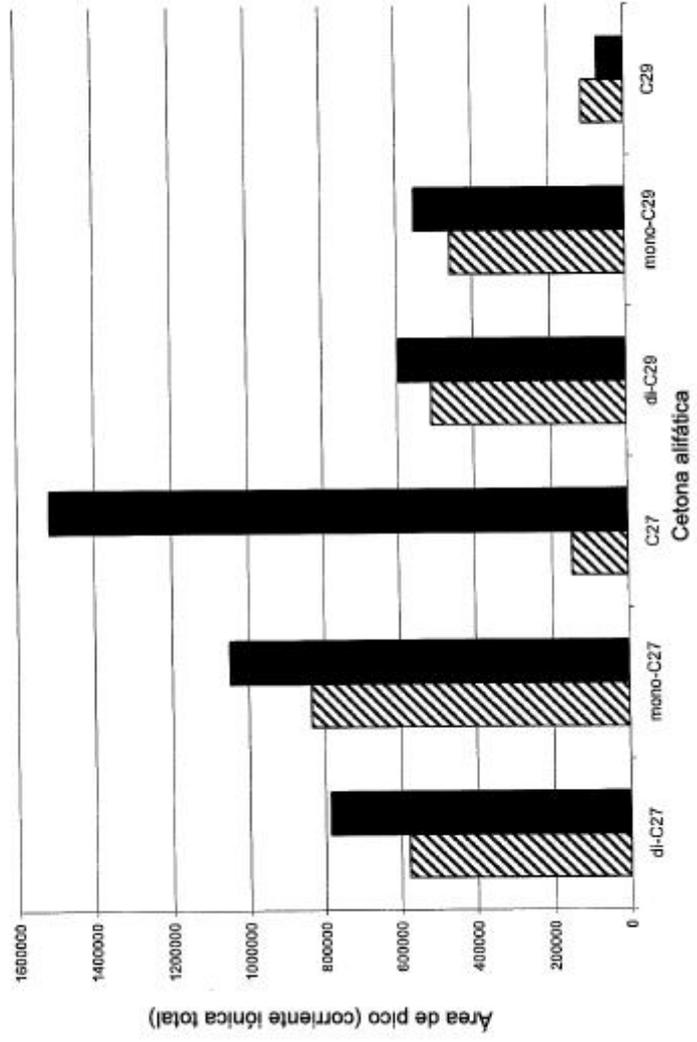


Figura 8A

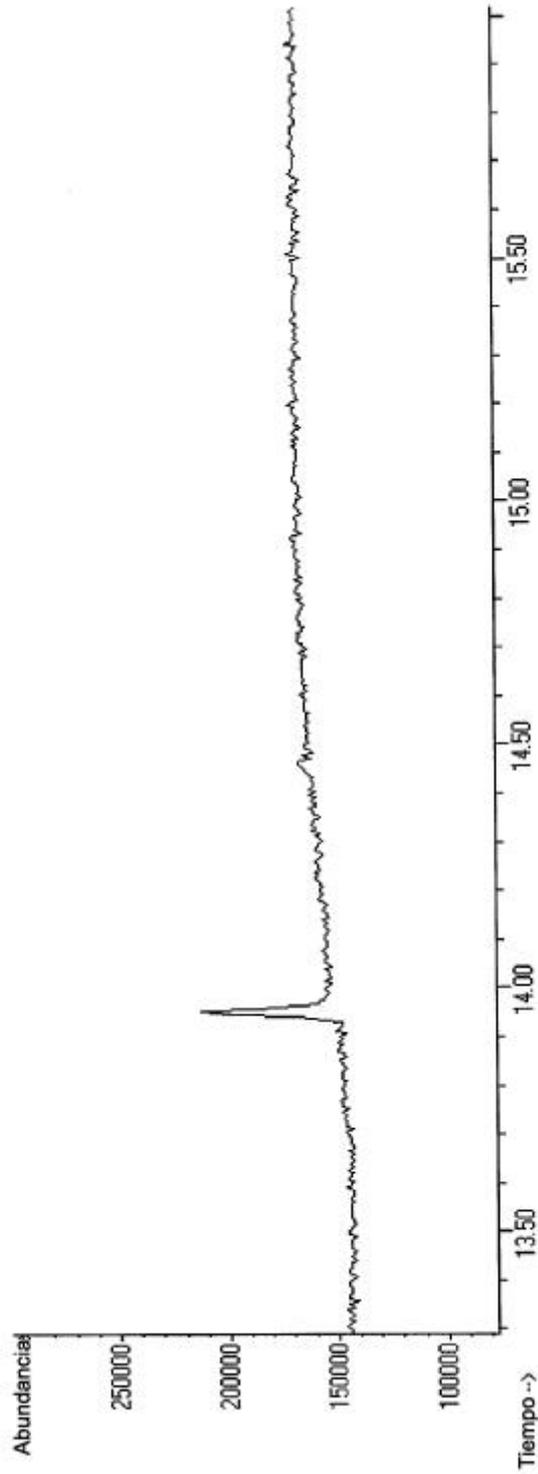


Figura 8B

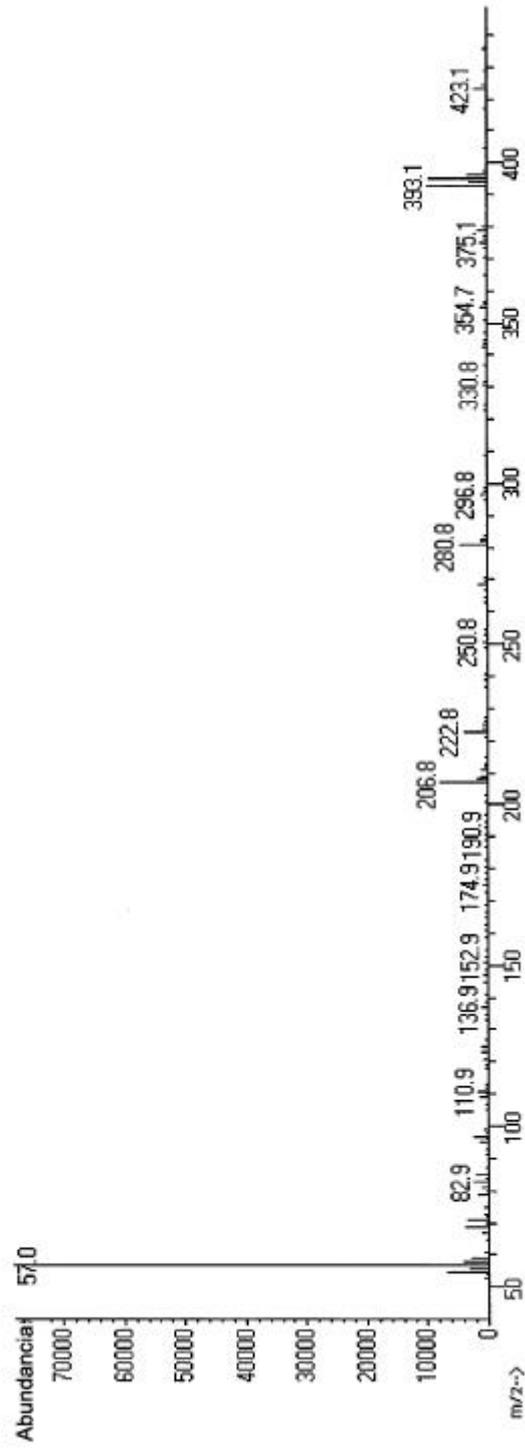


Figura 9

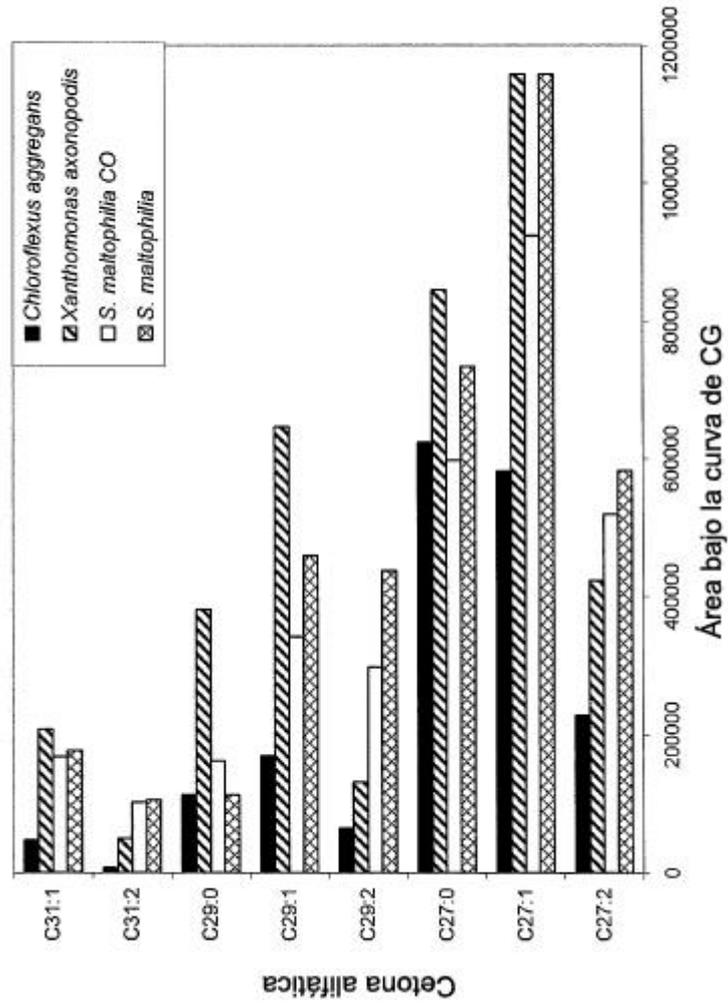


Figura 10

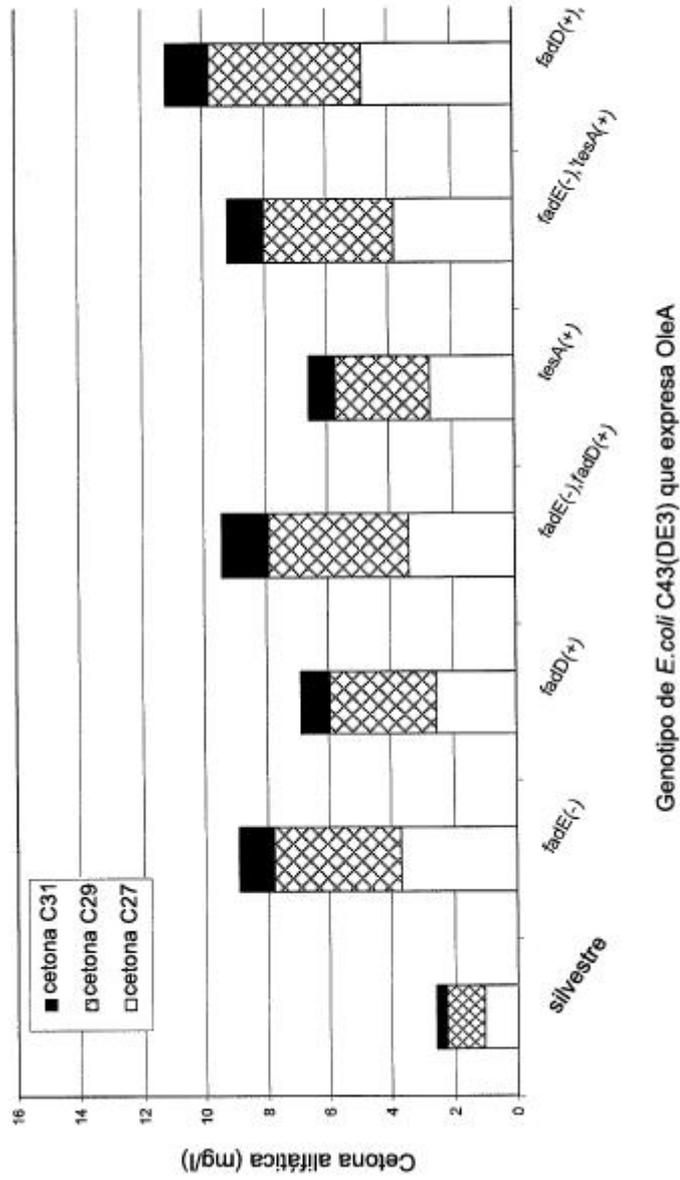


Figura 11

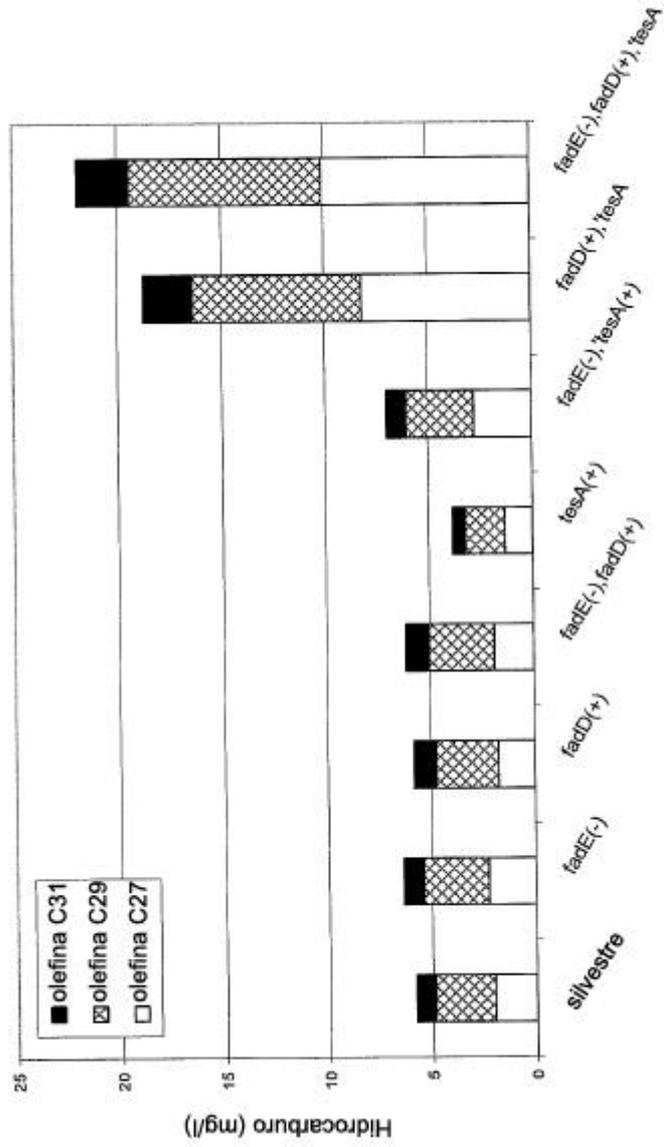


Figura 12A

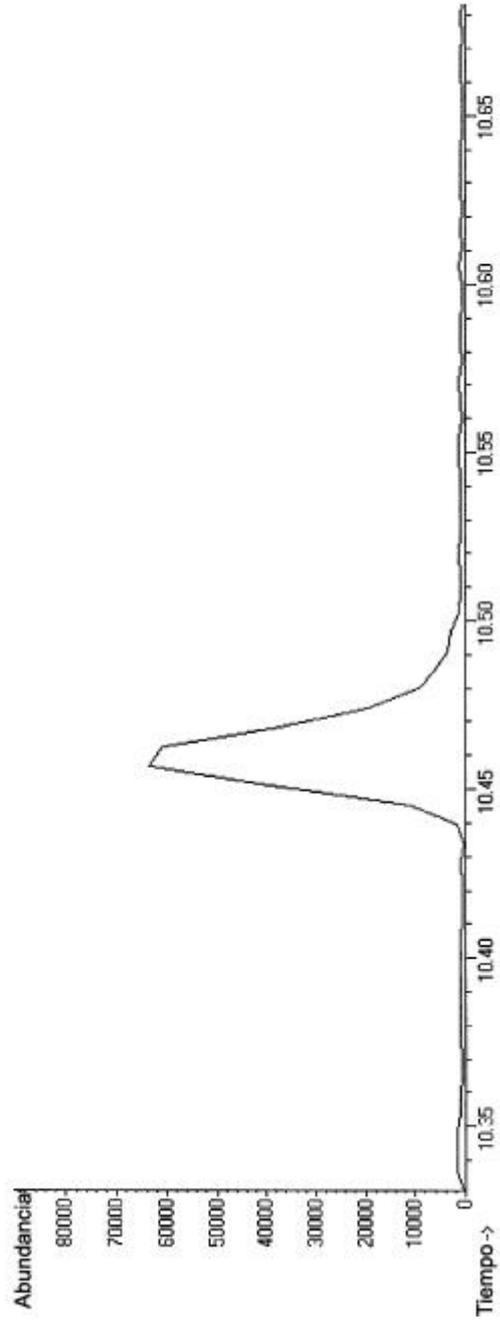


Figura 12B

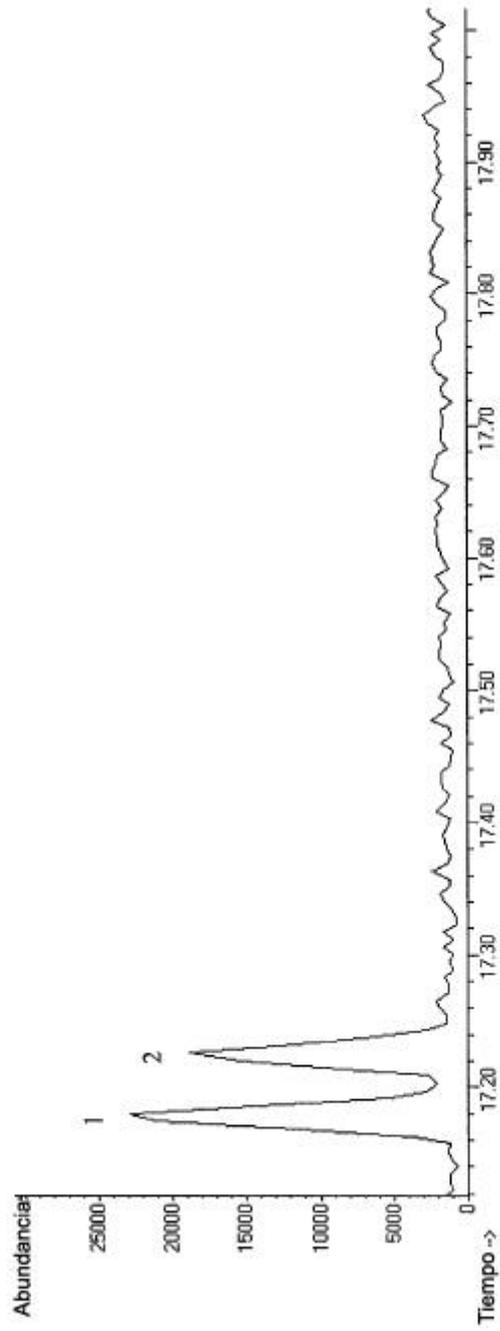


Figura 12C

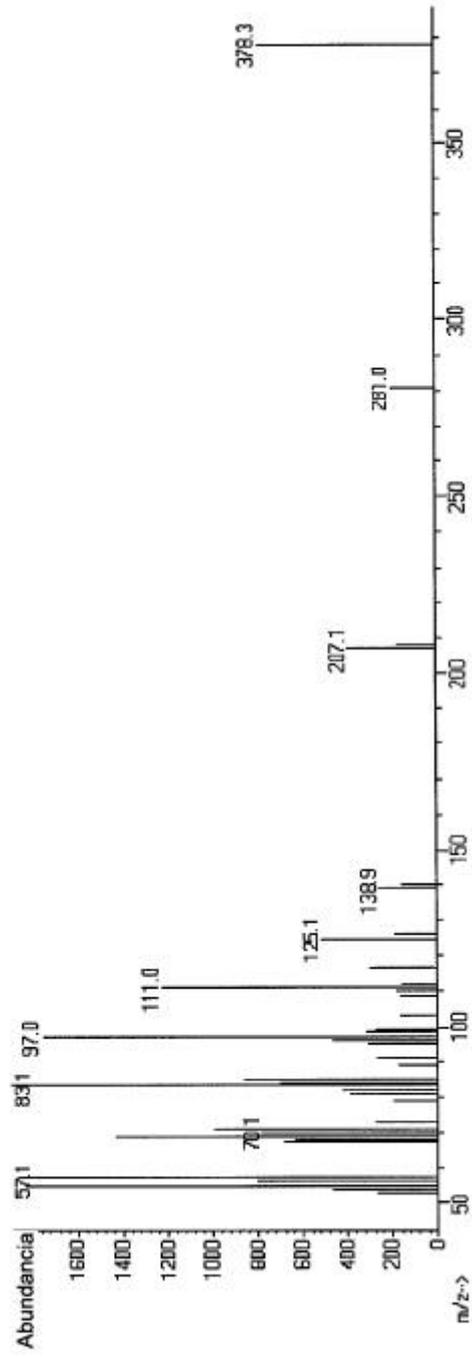


Figura 12D

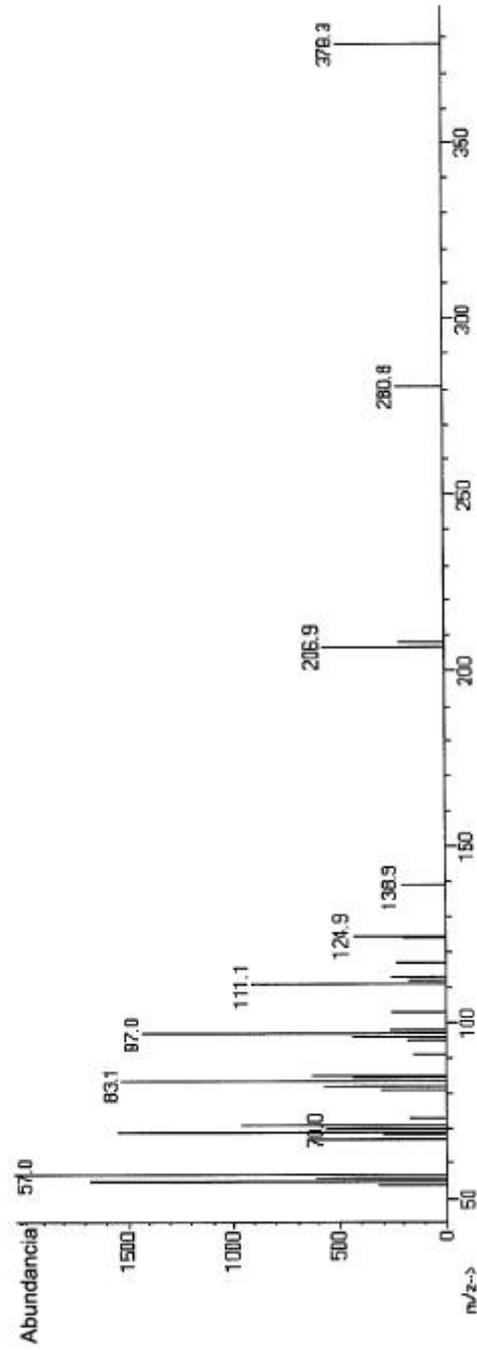


Figura 12E

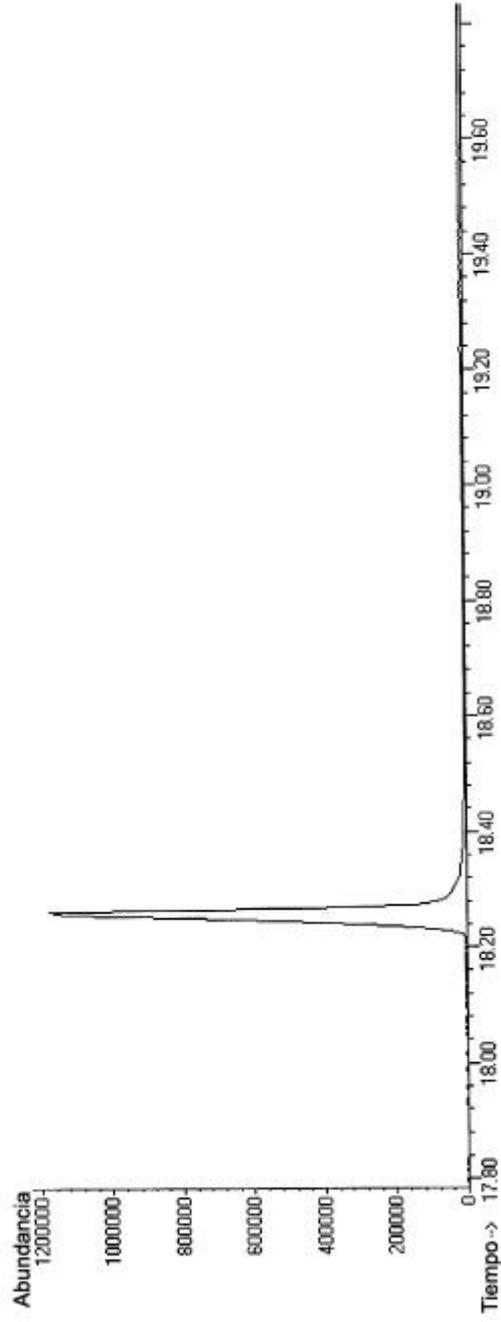


Figura 13A

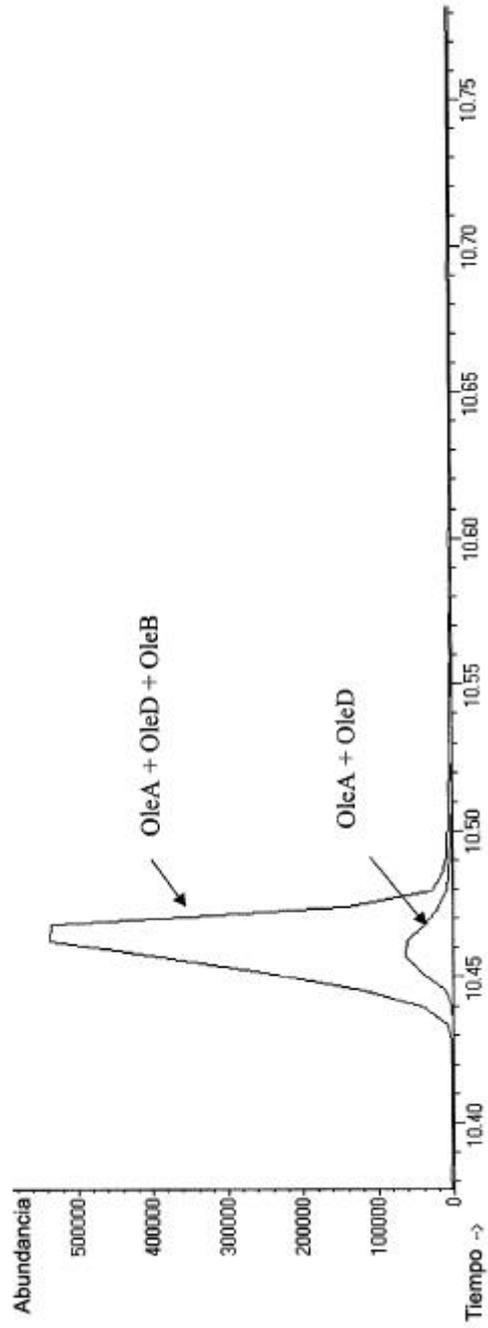


Figura 13B

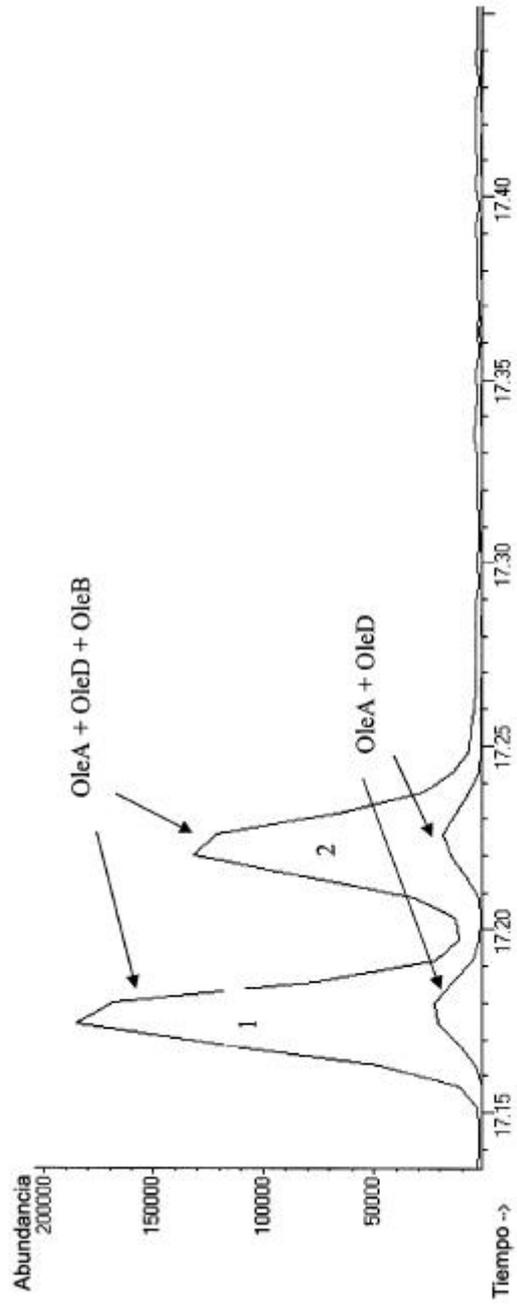


Figura 13C

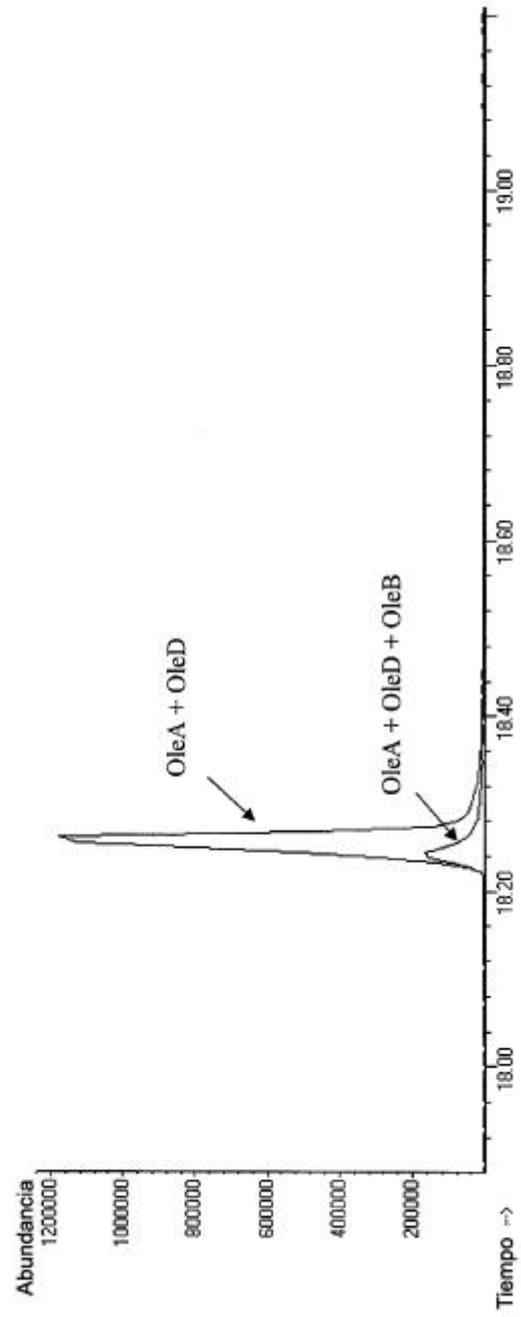


Figura 14

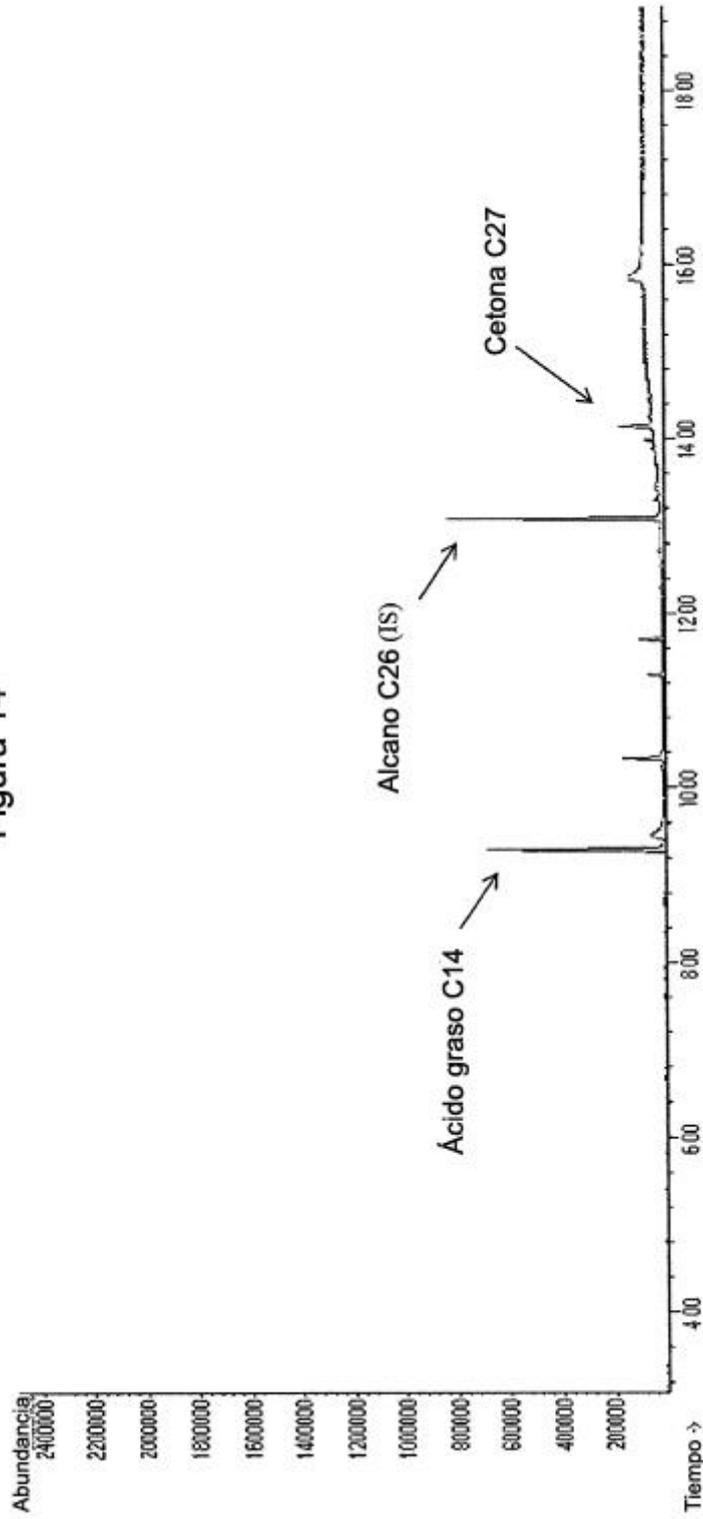


Figura 15A

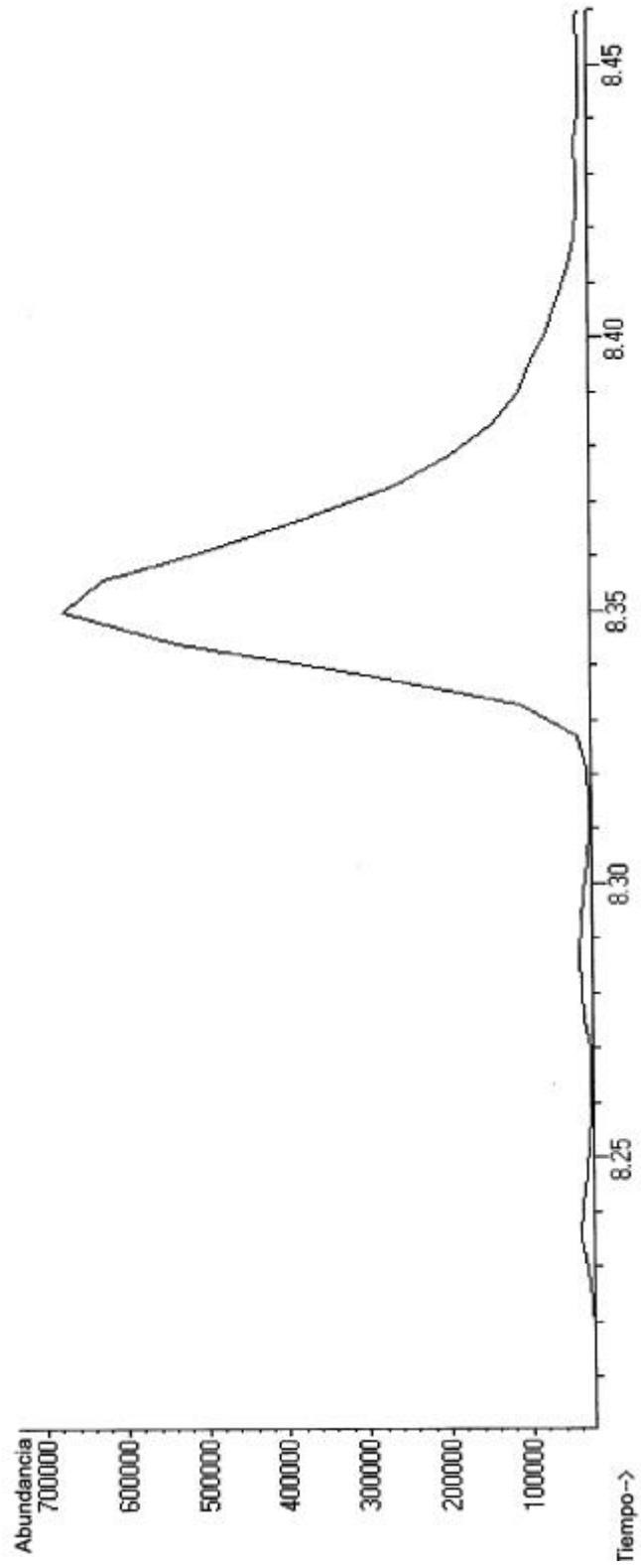


Figura 15B

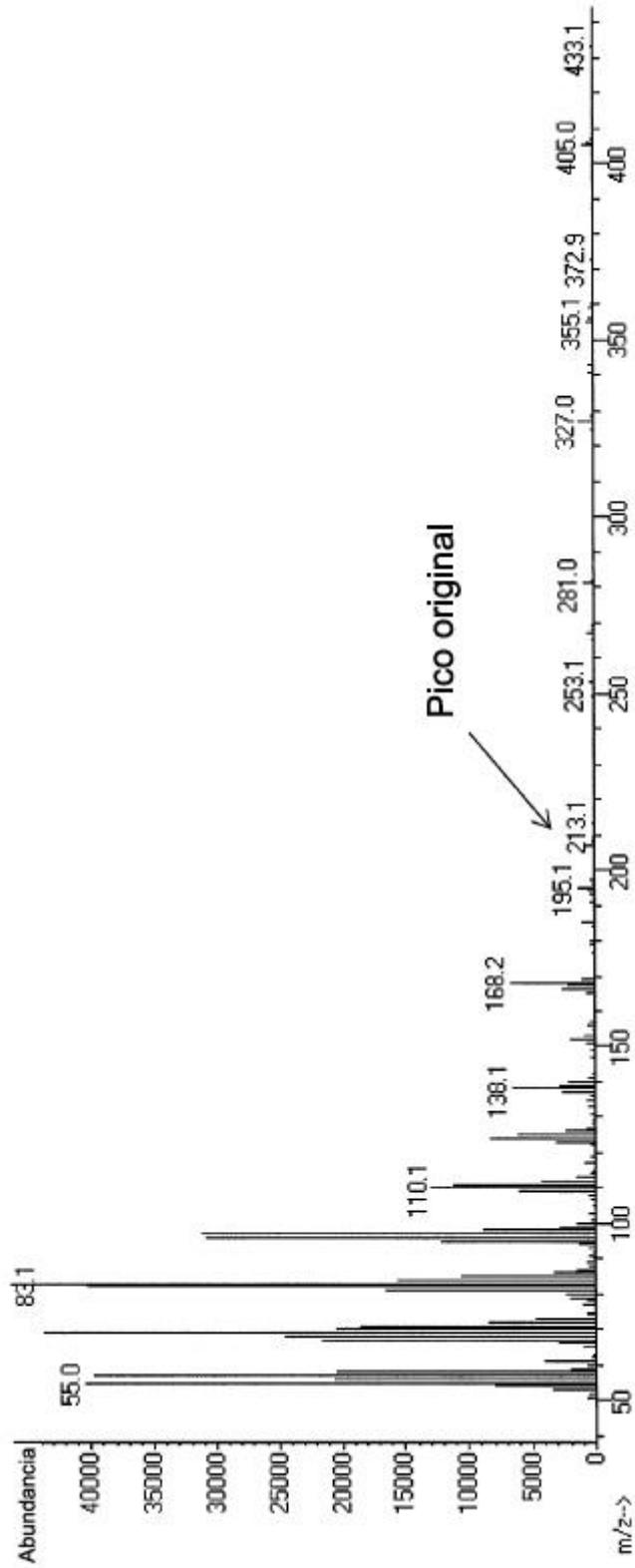


Figura 15C

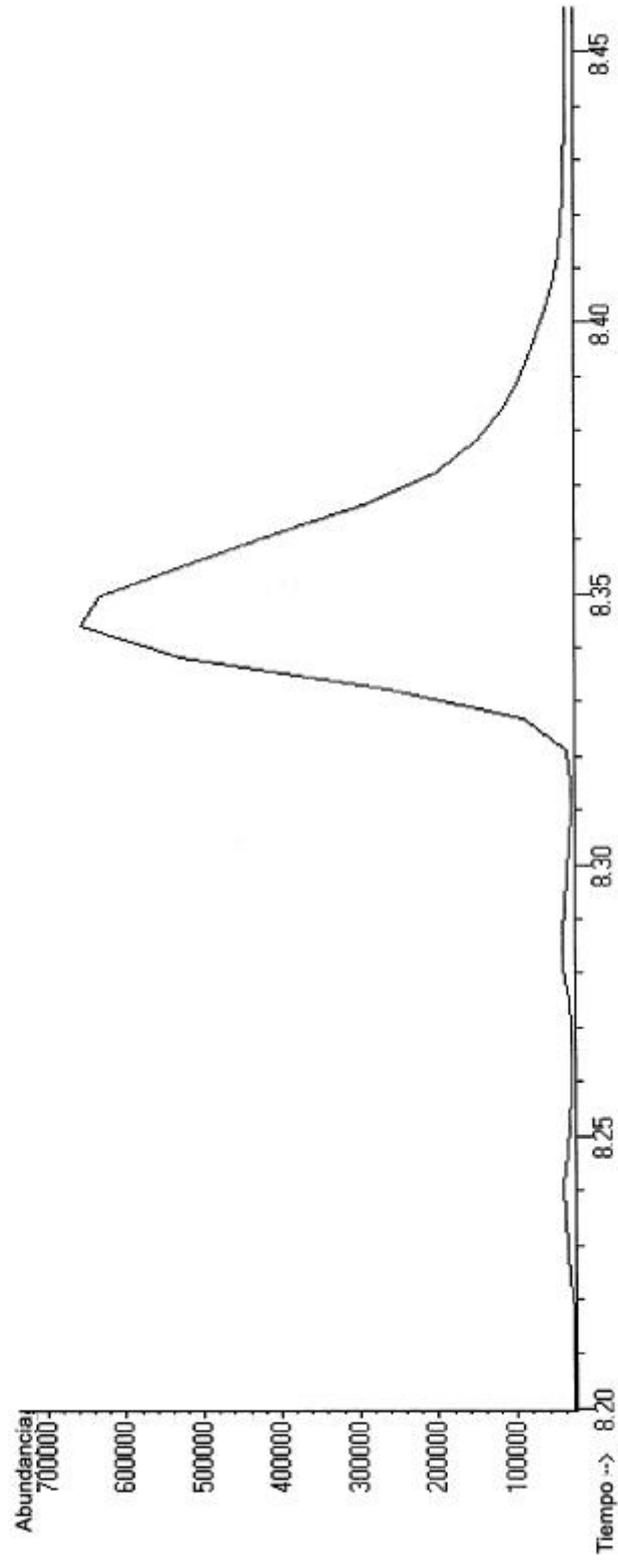


Figura 15D

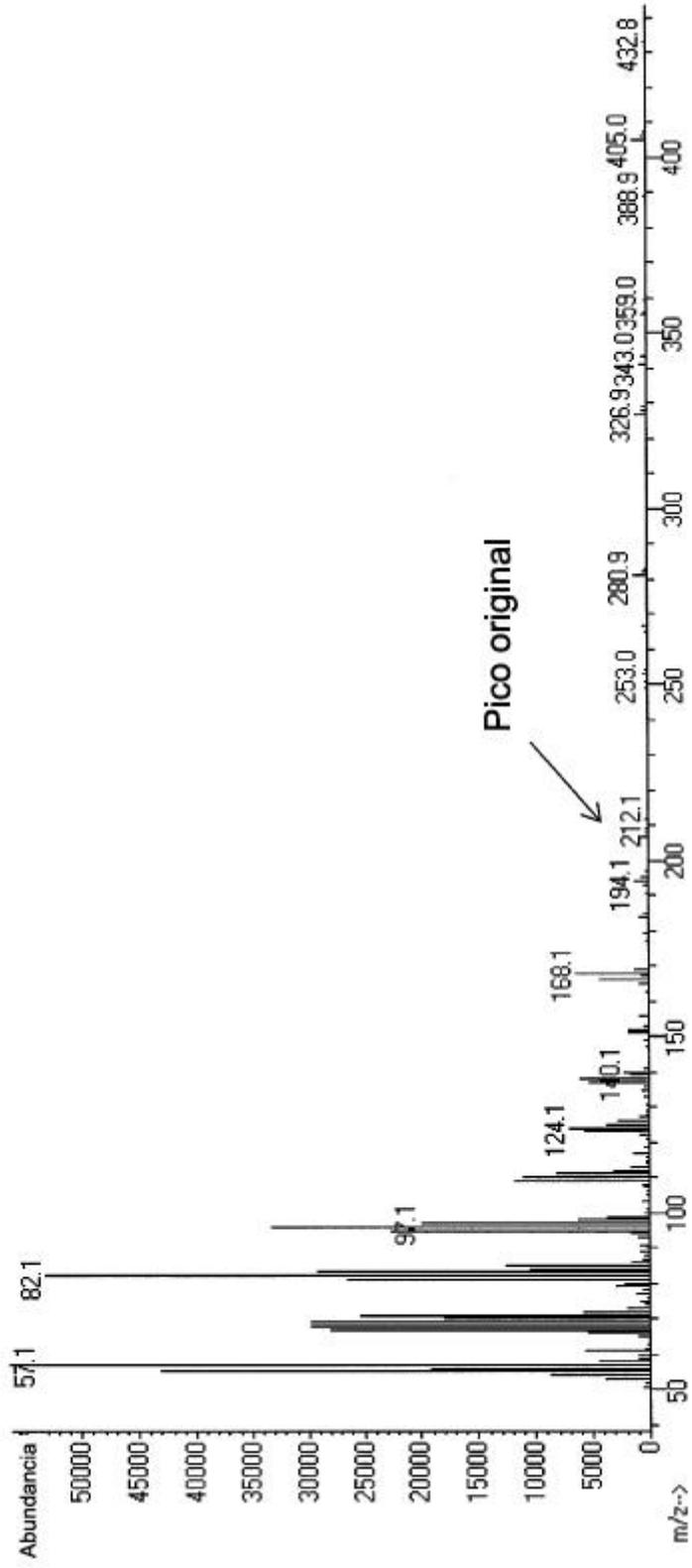


Figura 15E

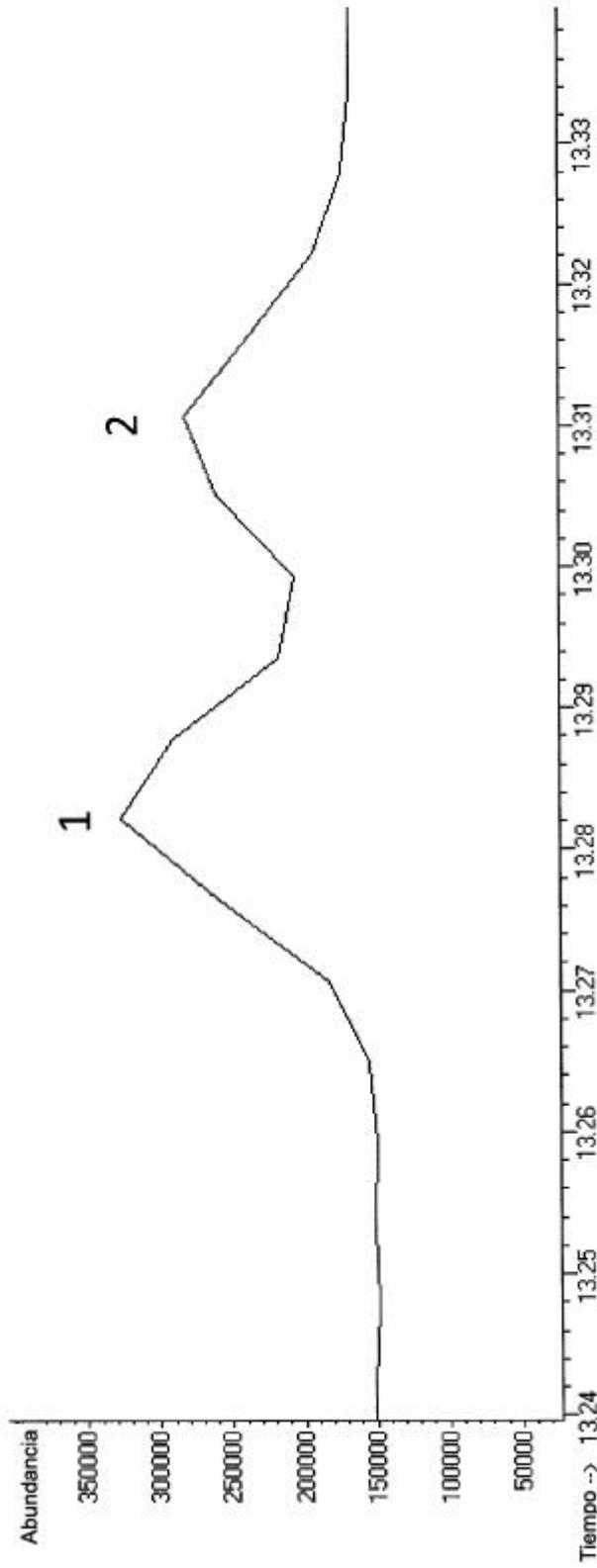


Figura 15F

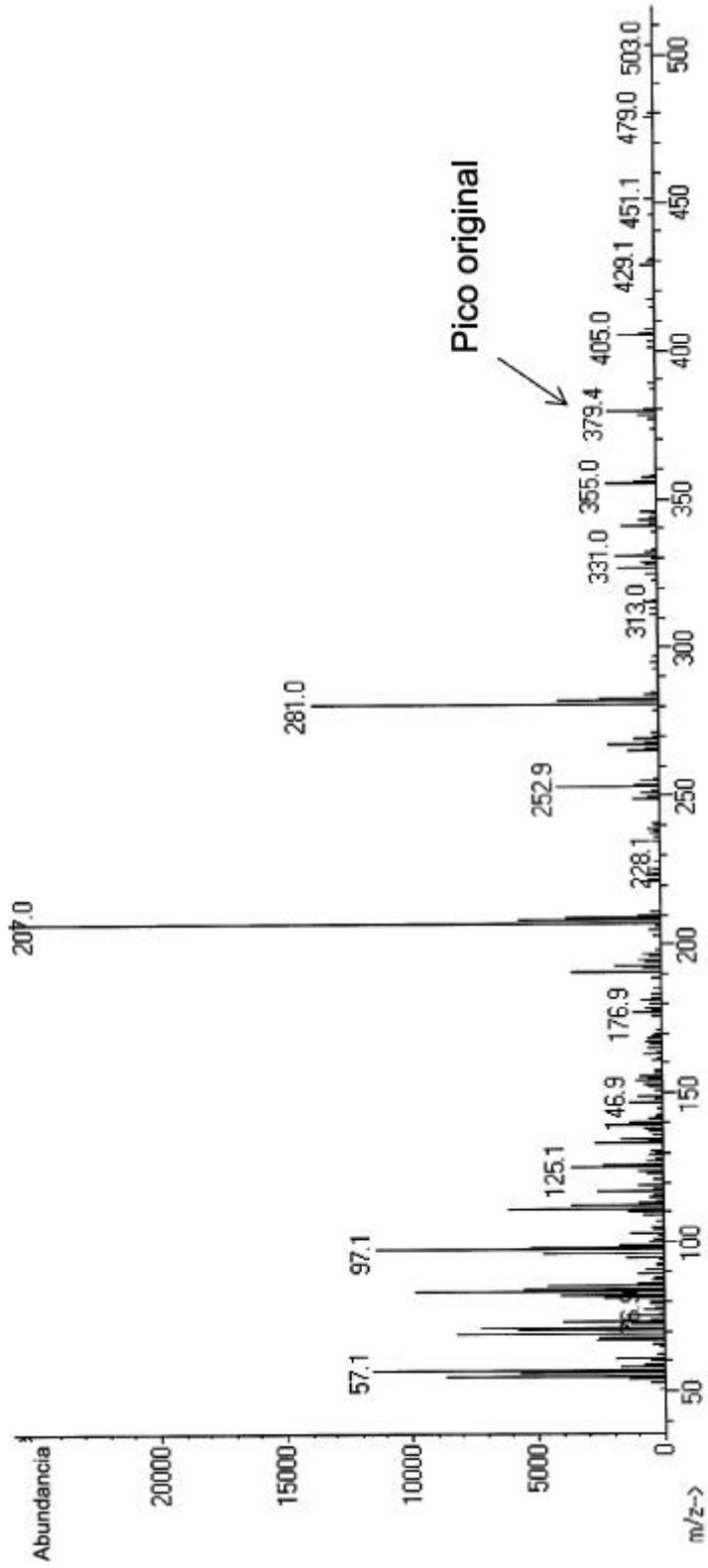


Figura 15G

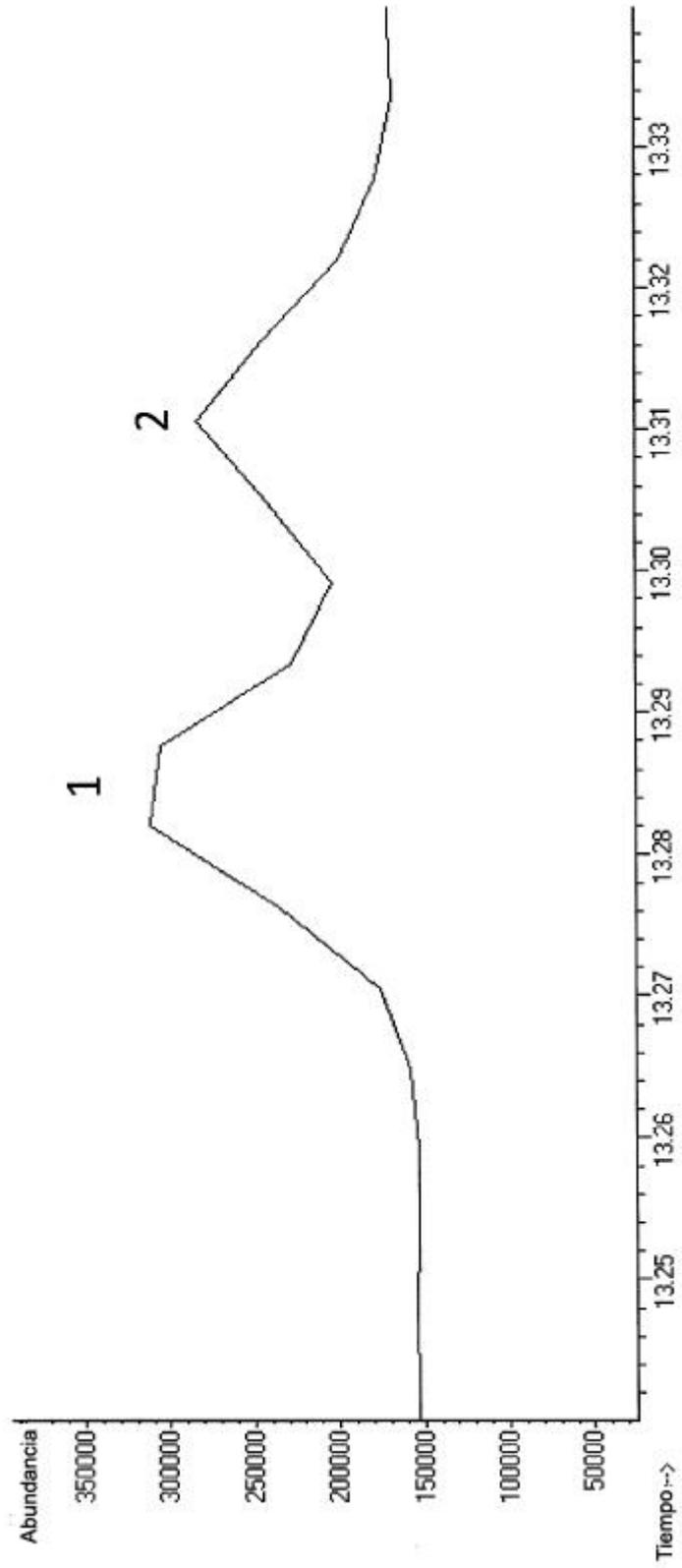


Figura 15H

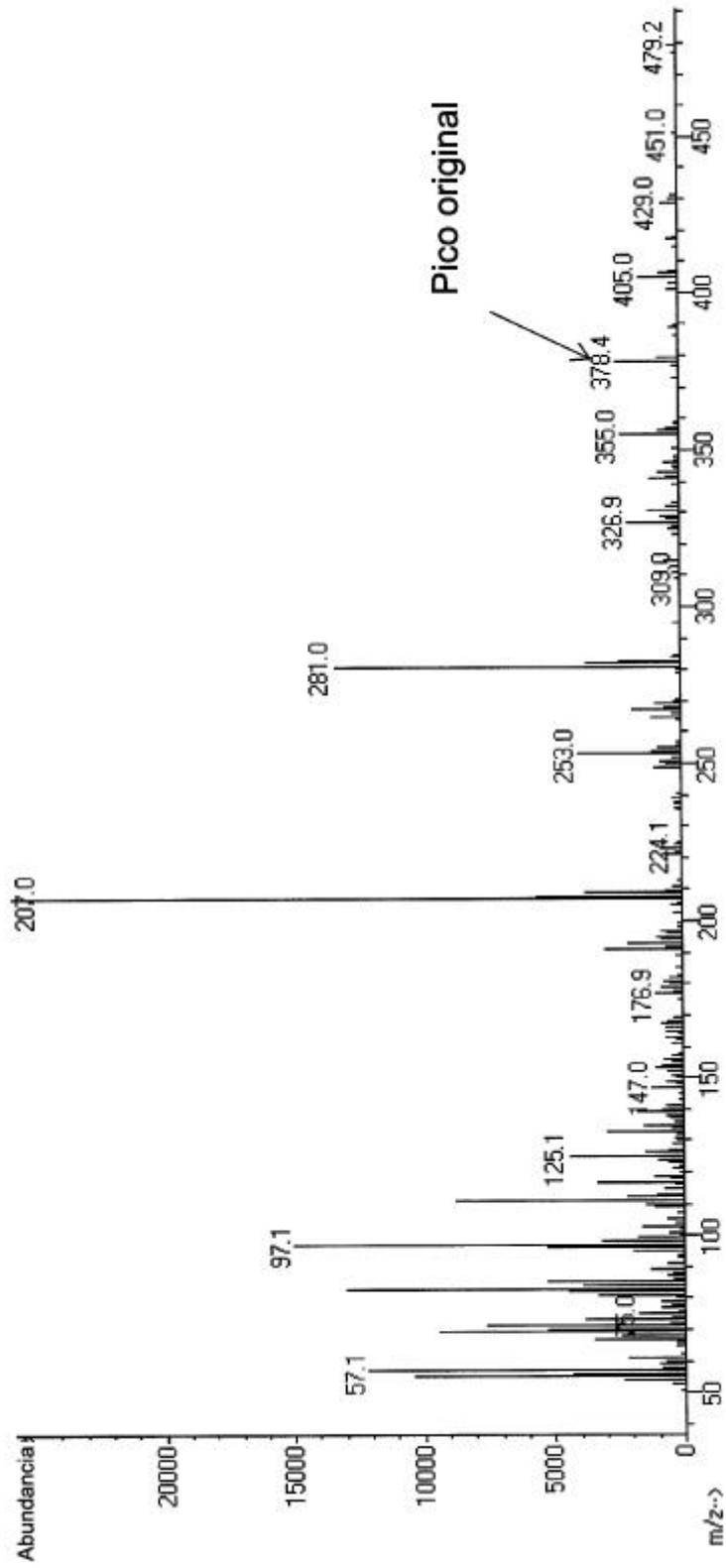


Figura 16A

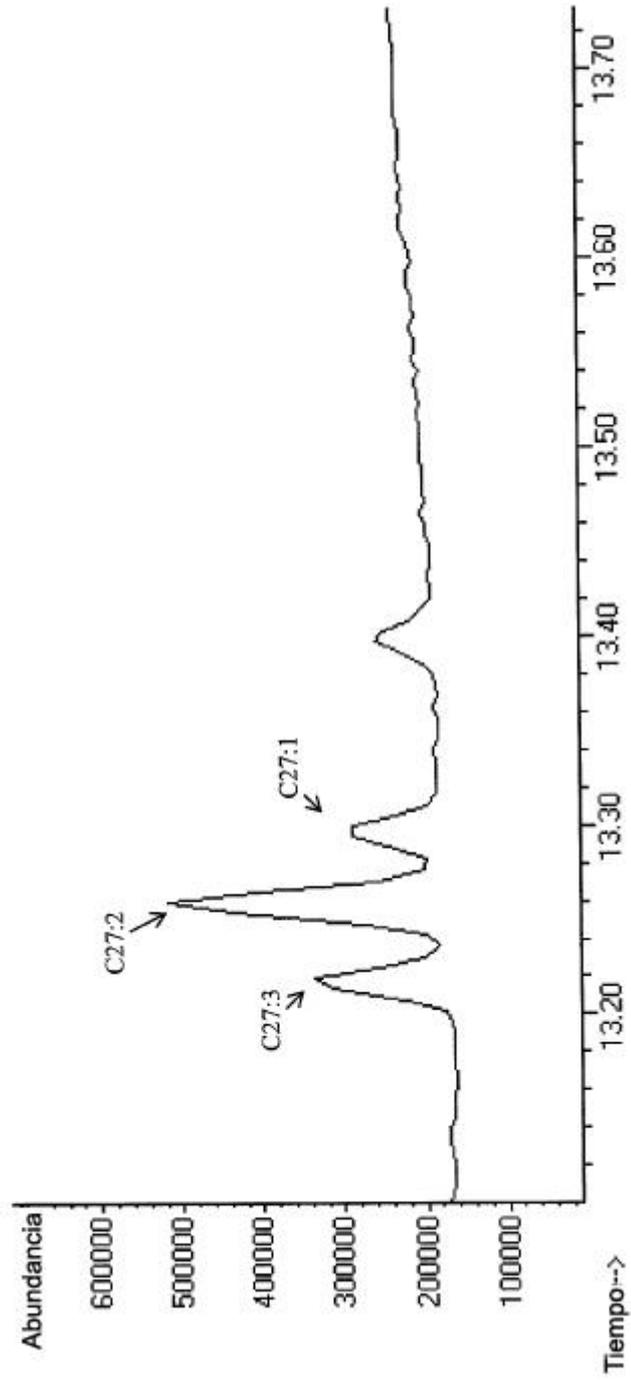


Figura 16B

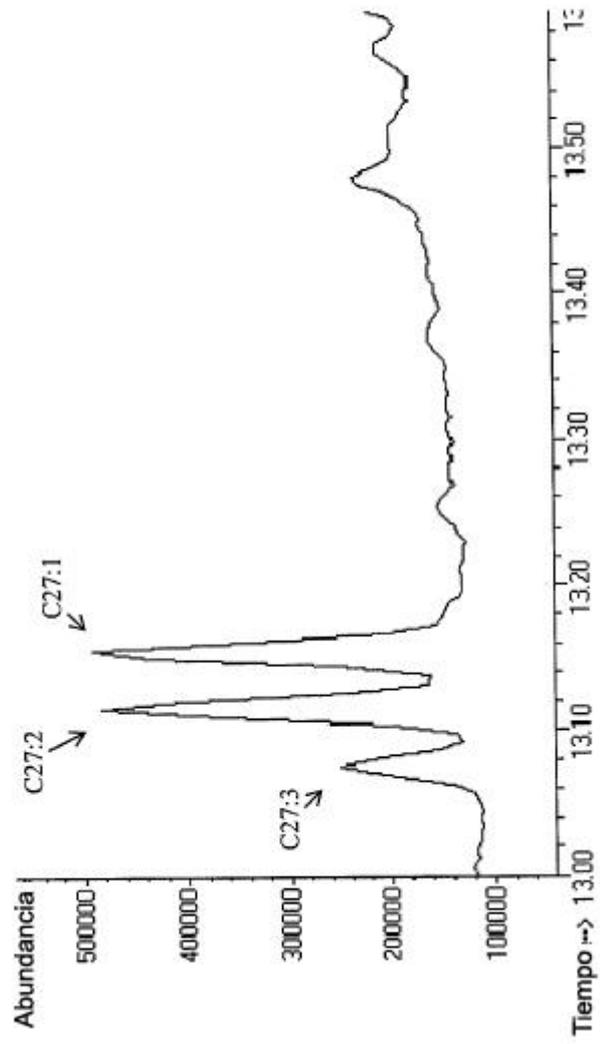


Figura 16C

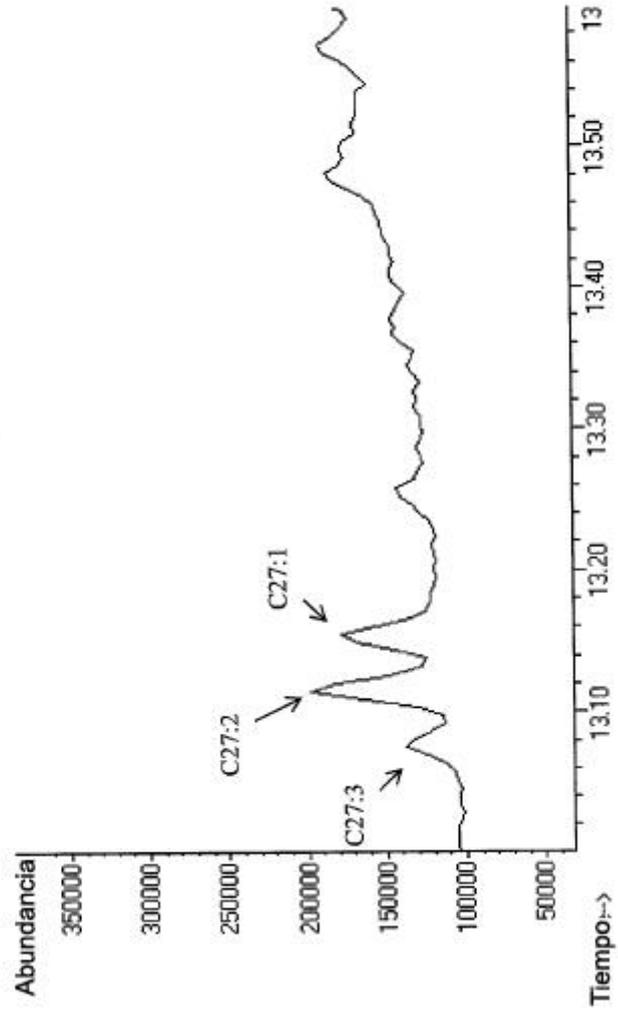


Figura 17A

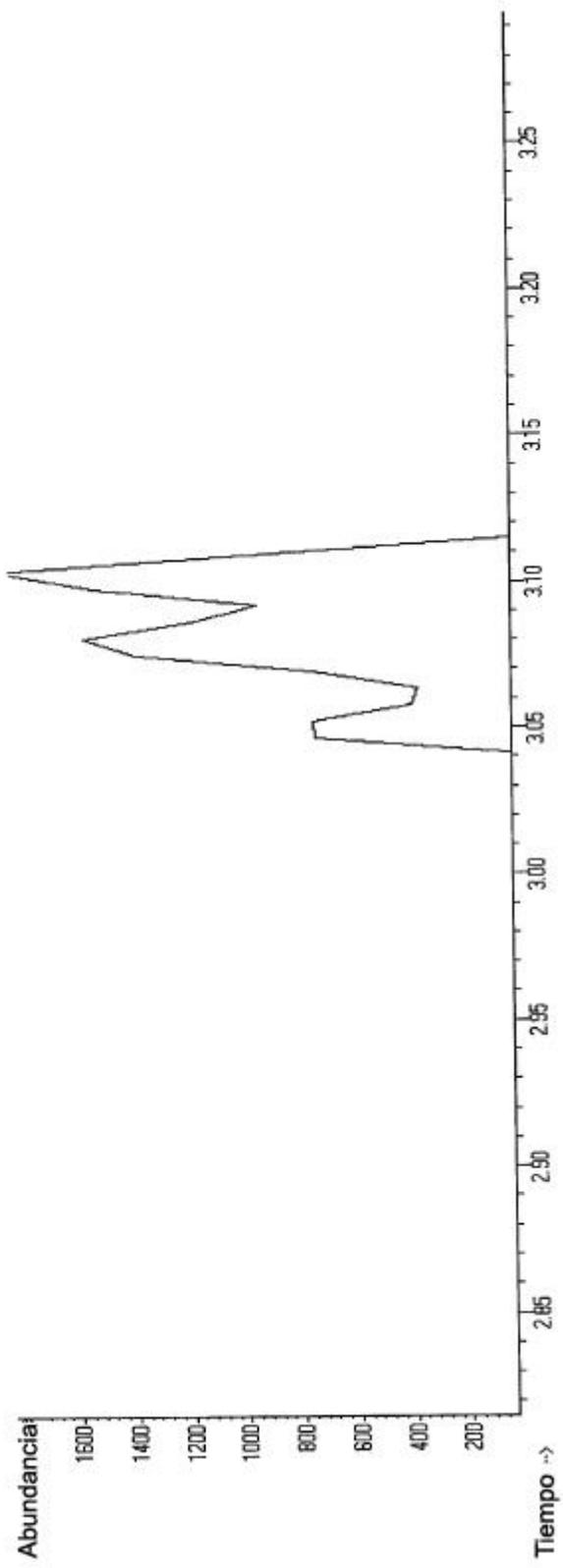


Figura 17B

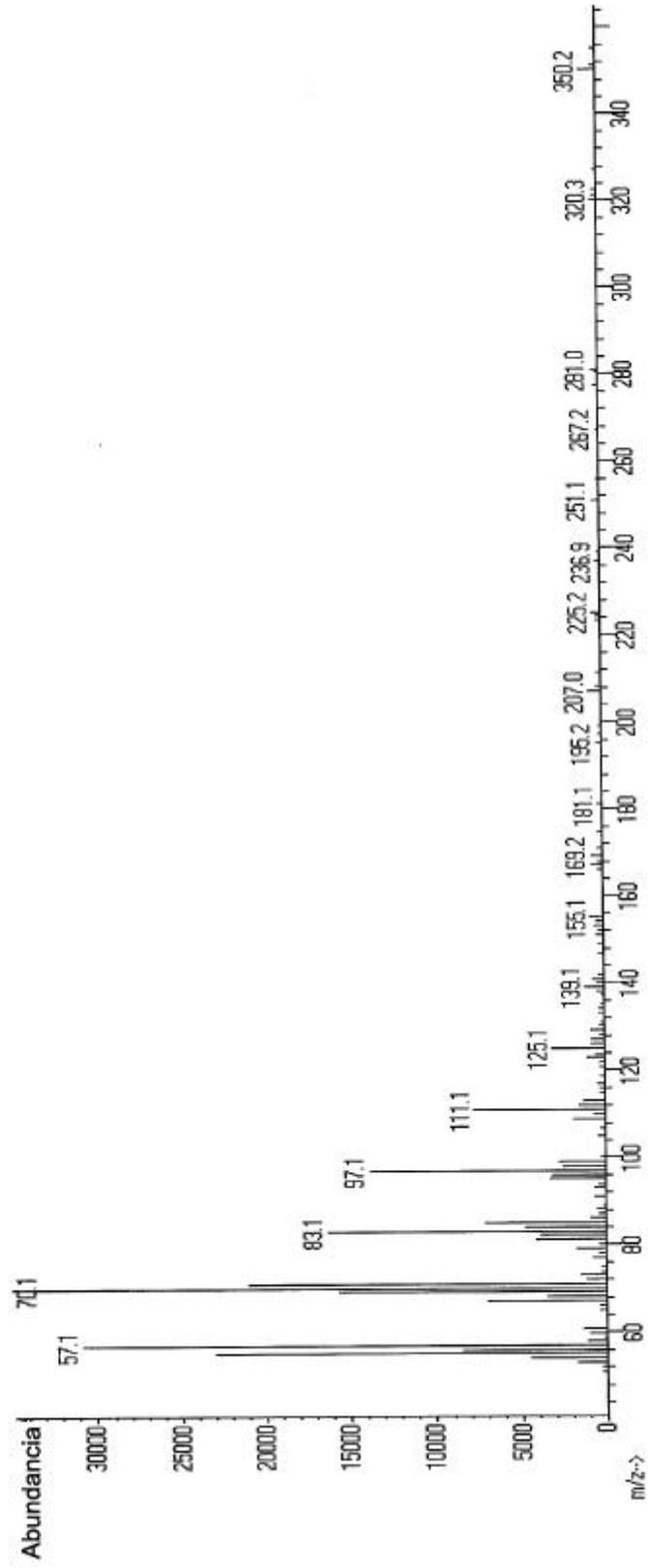


Figura 17C

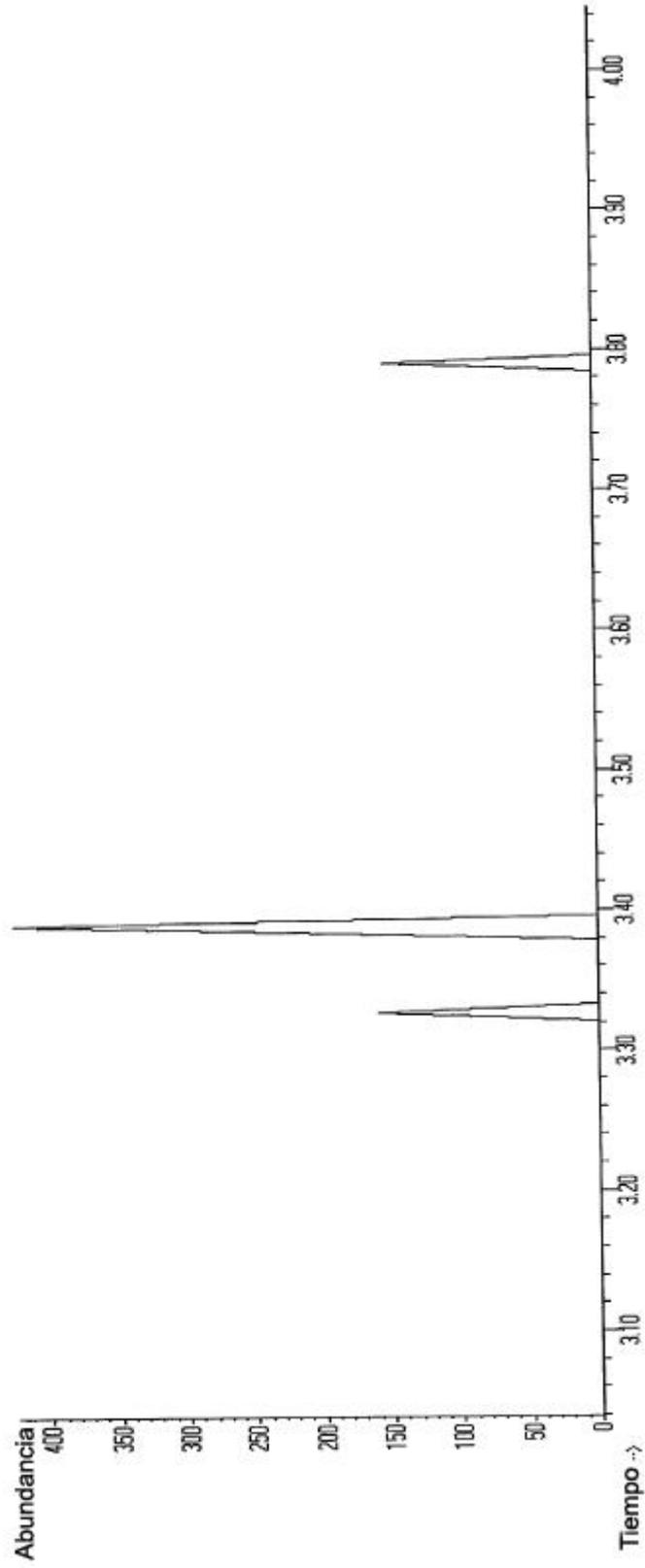


Figura 17D

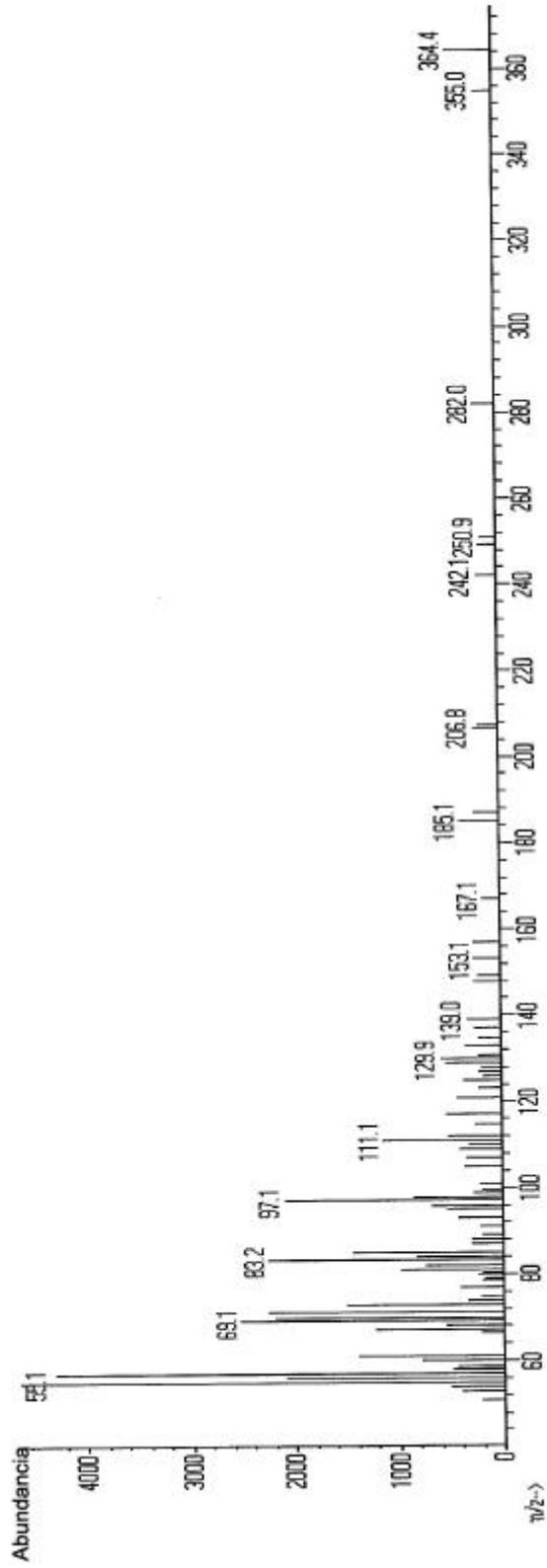


Figura 17E

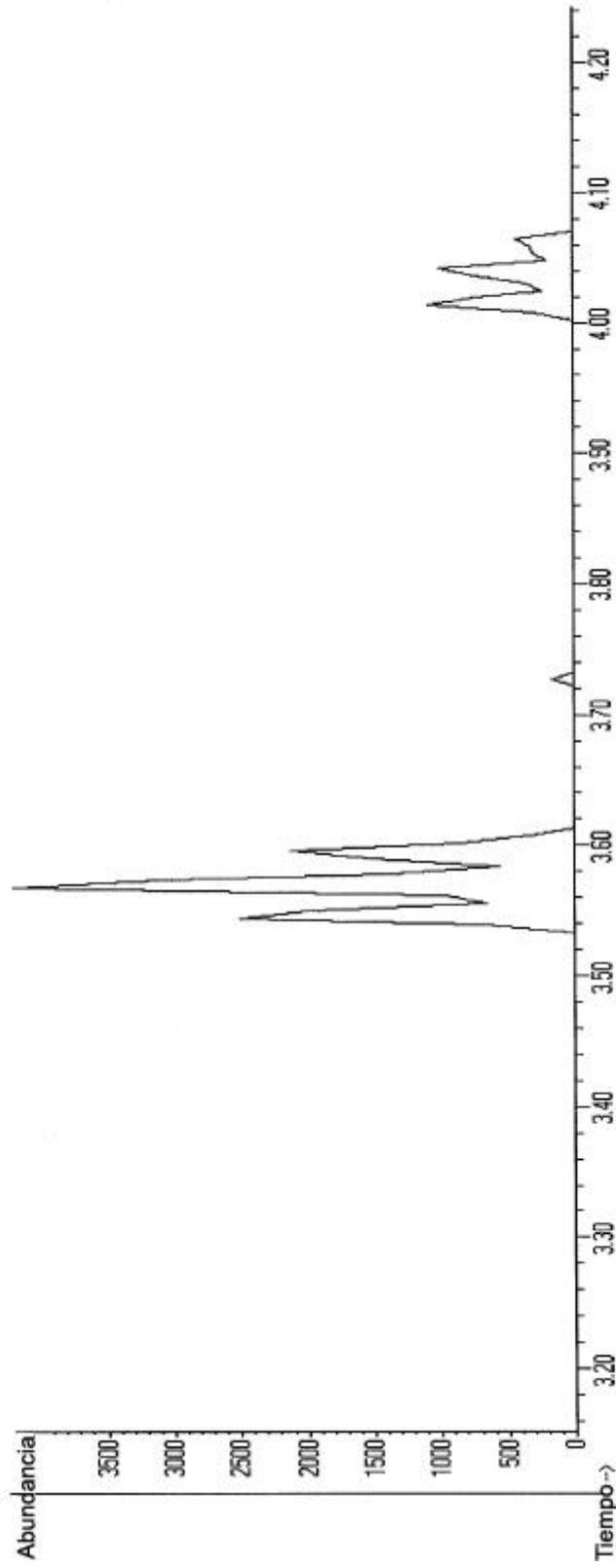


Figura 17F

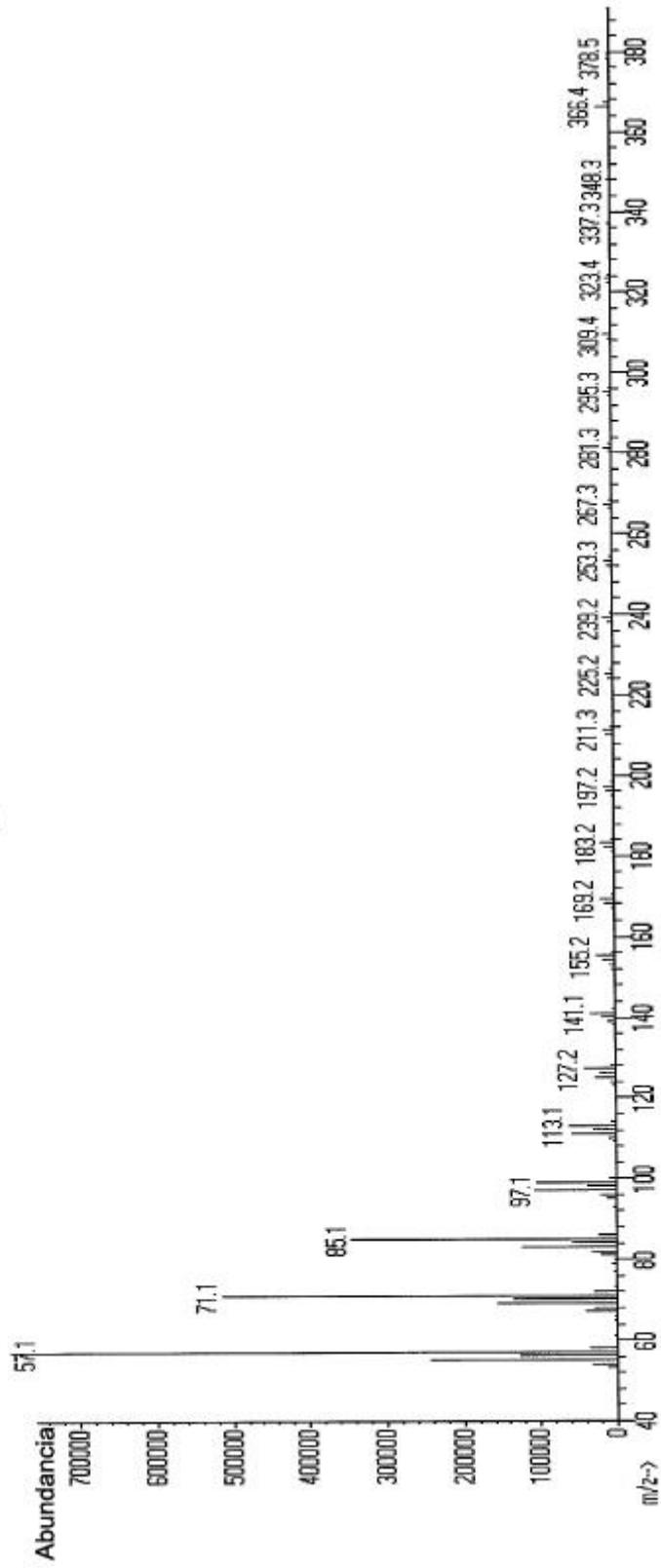


Figura 17G

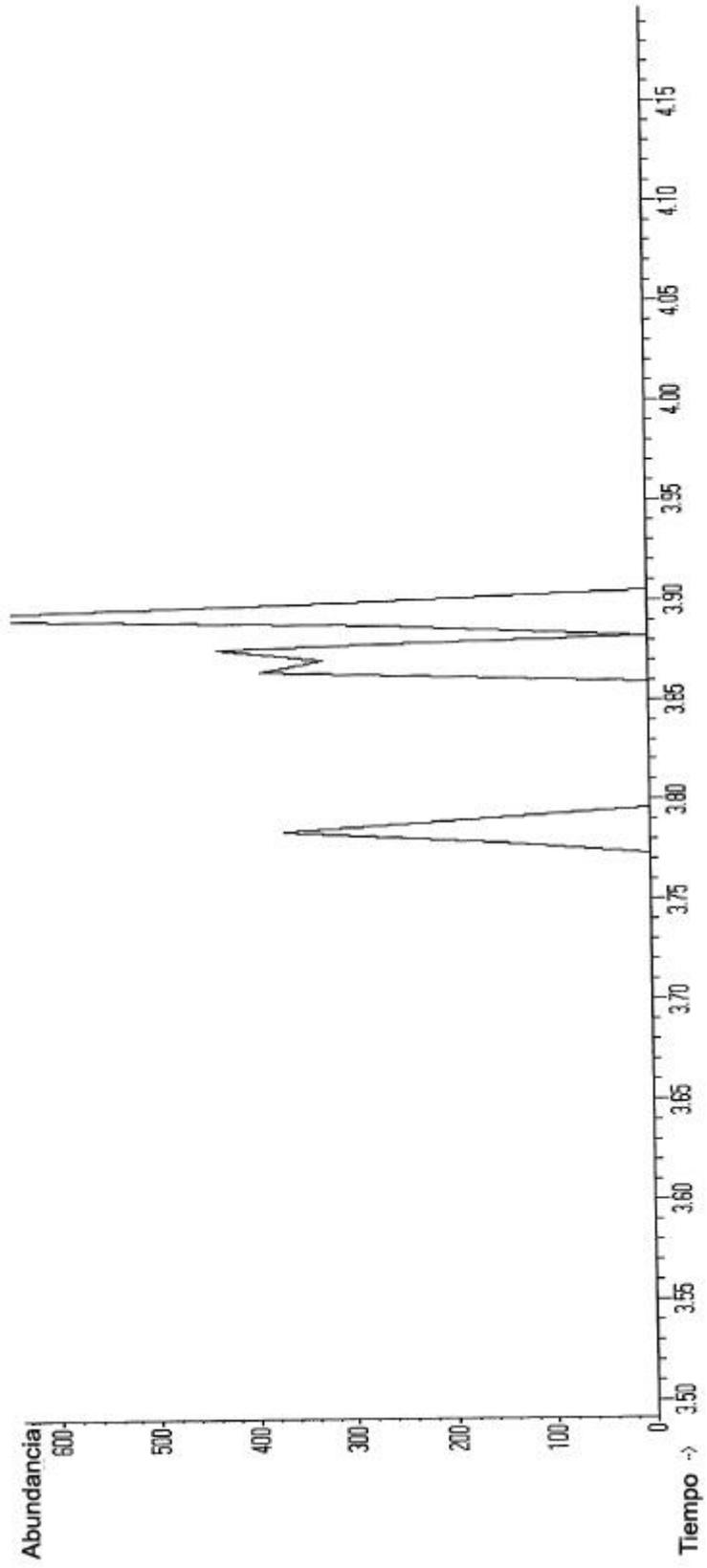


Figura 17H

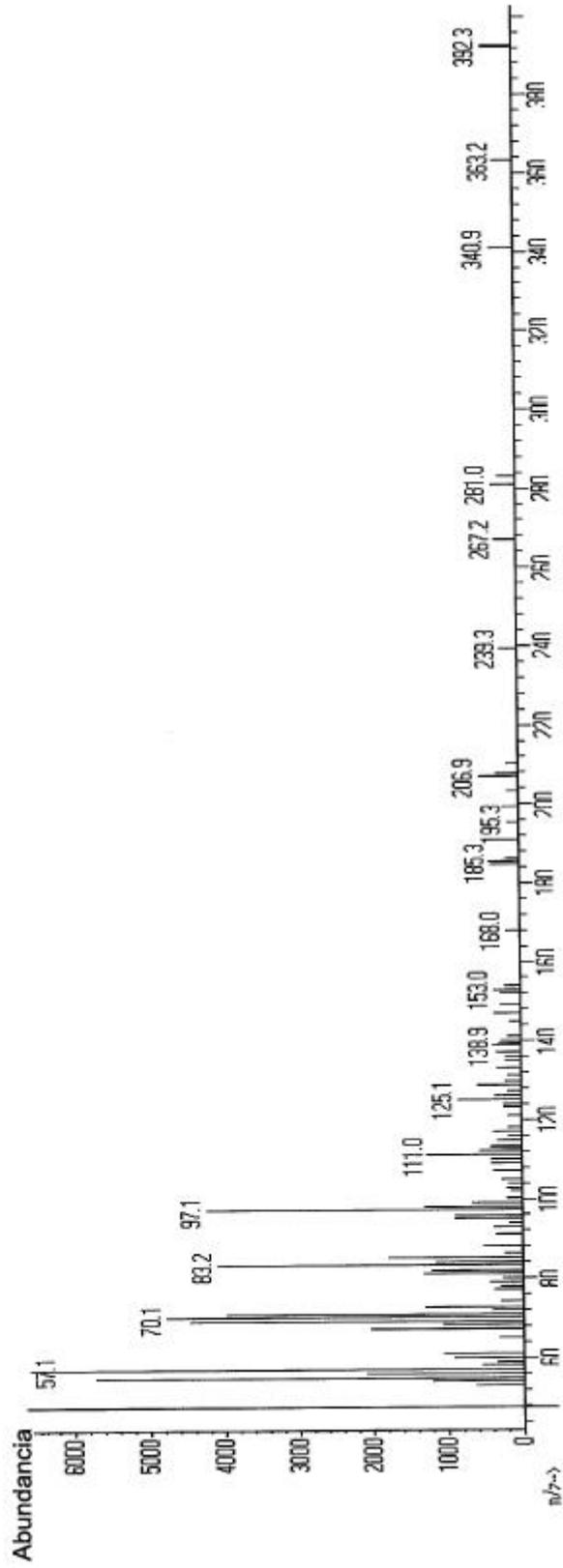


Figura 171

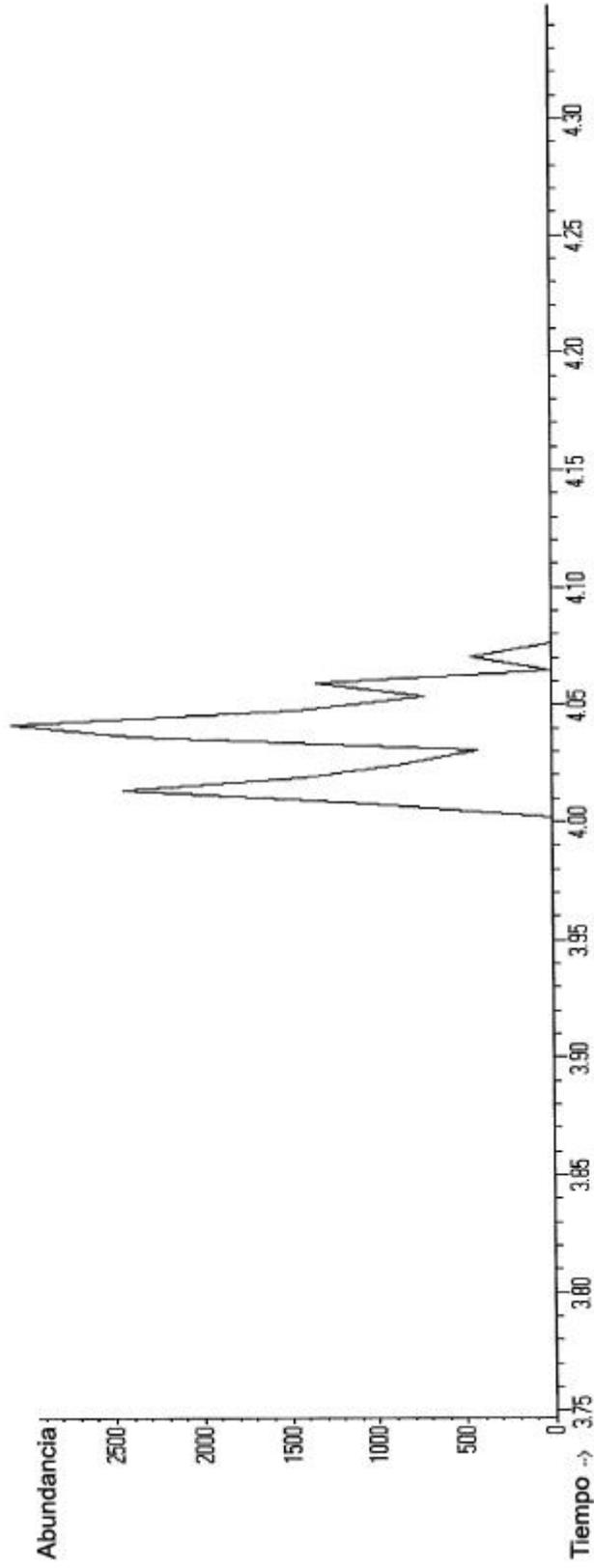


Figura 17J

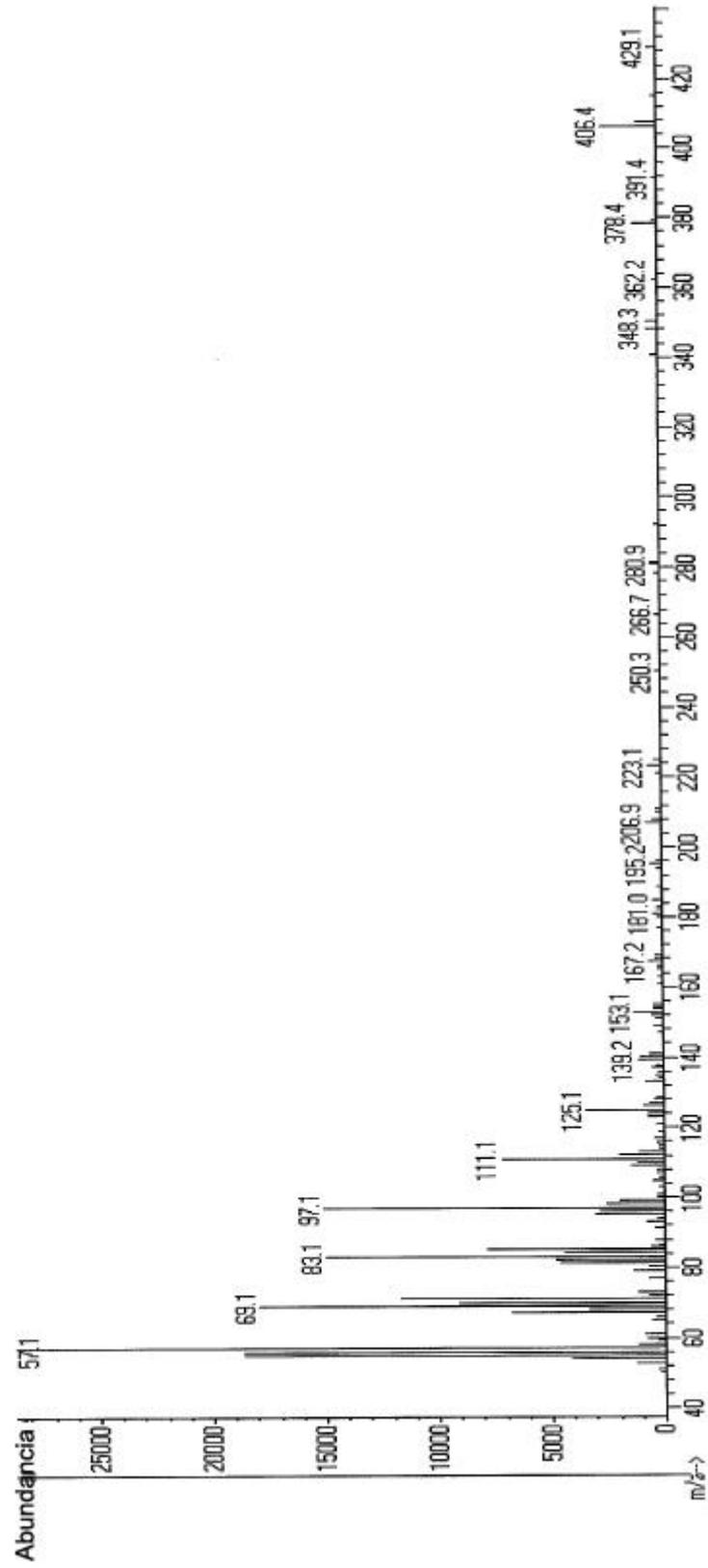


Figura 18A

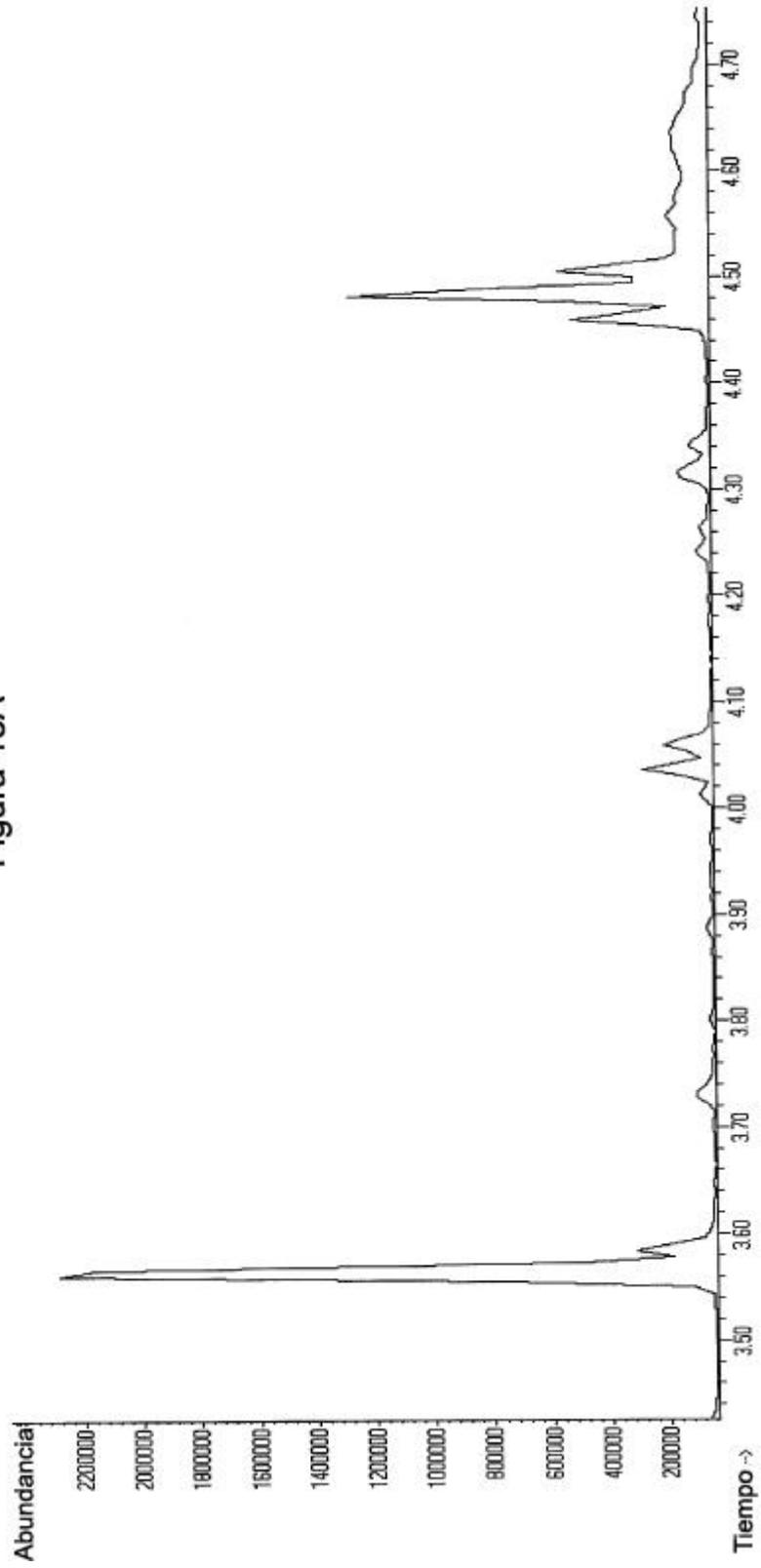


Figura 18B

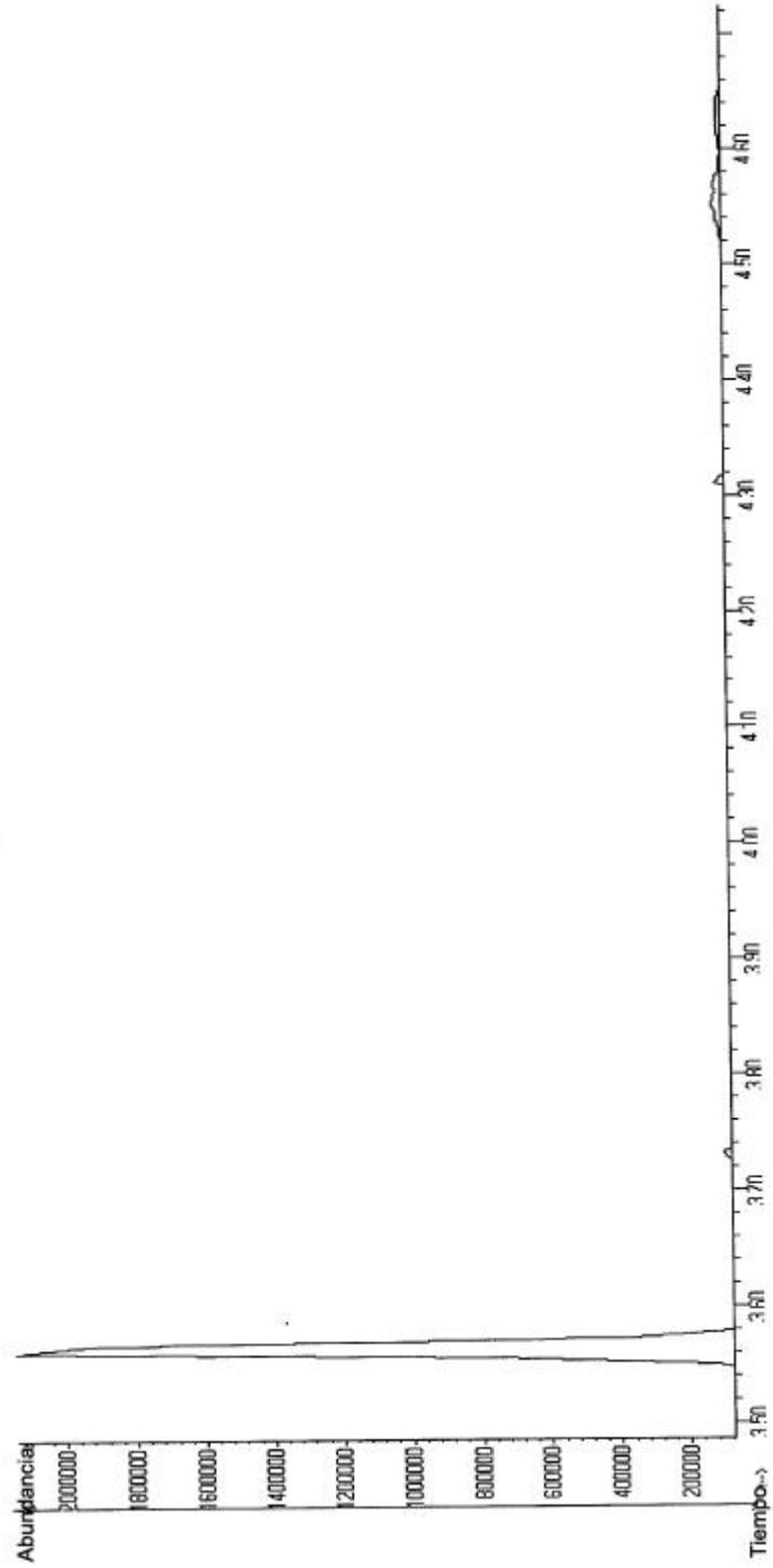


Figura 18C

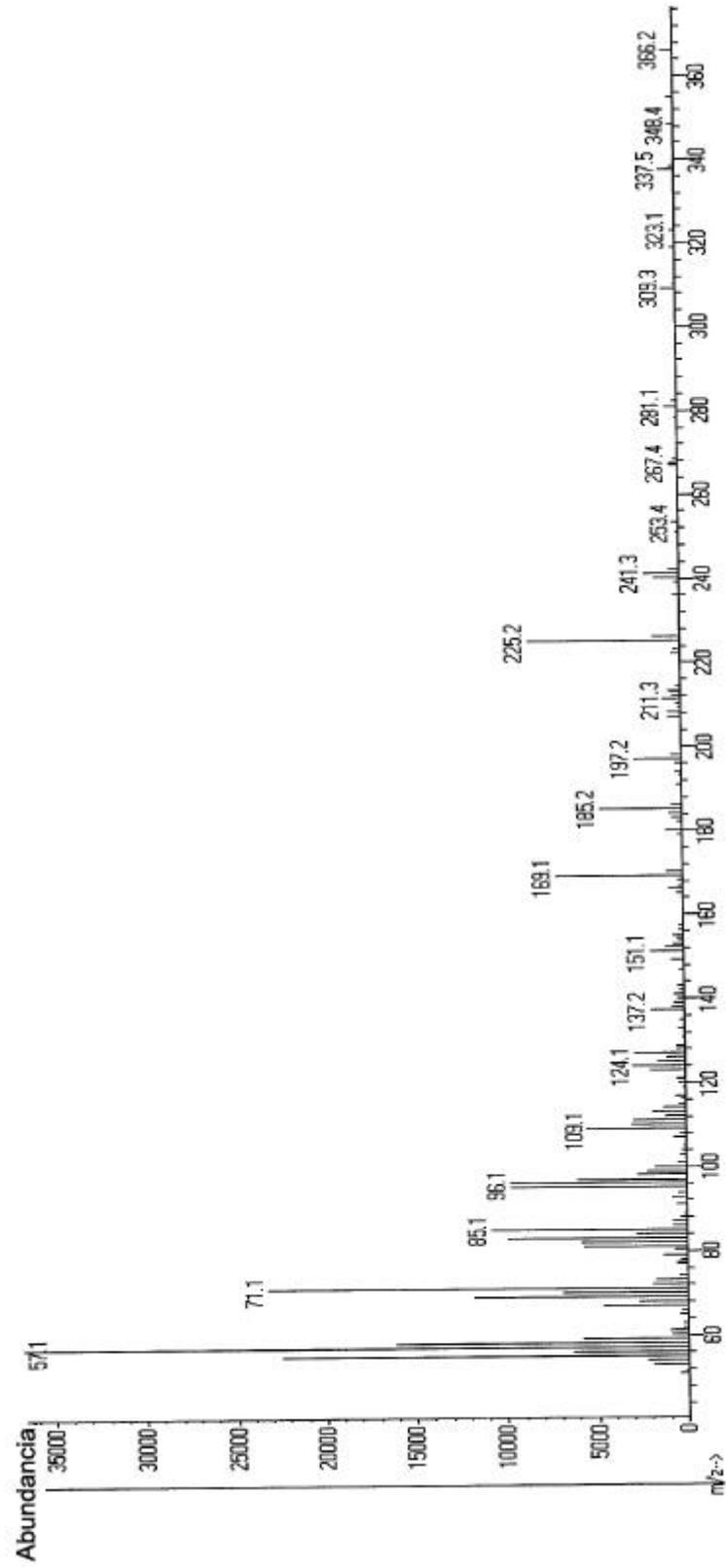


Figura 18D

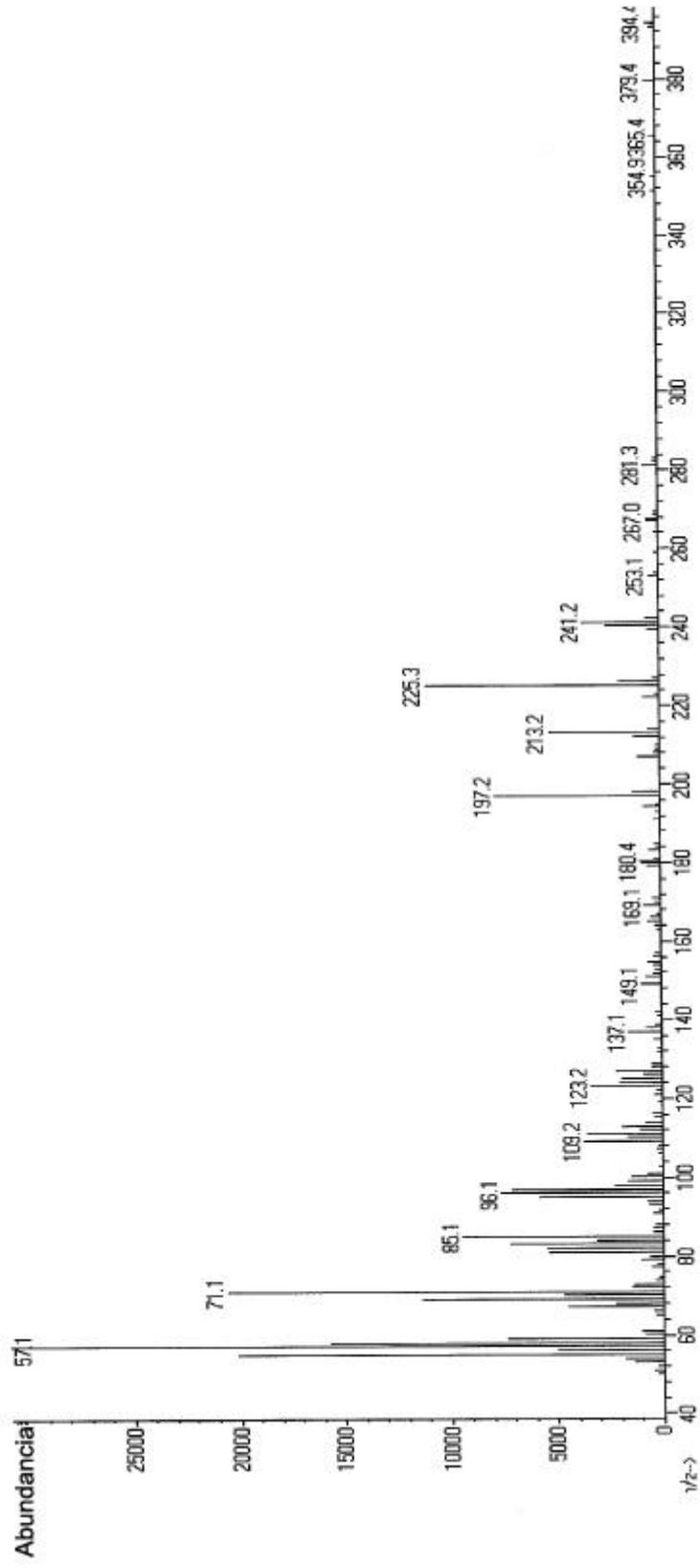


Figura 18E

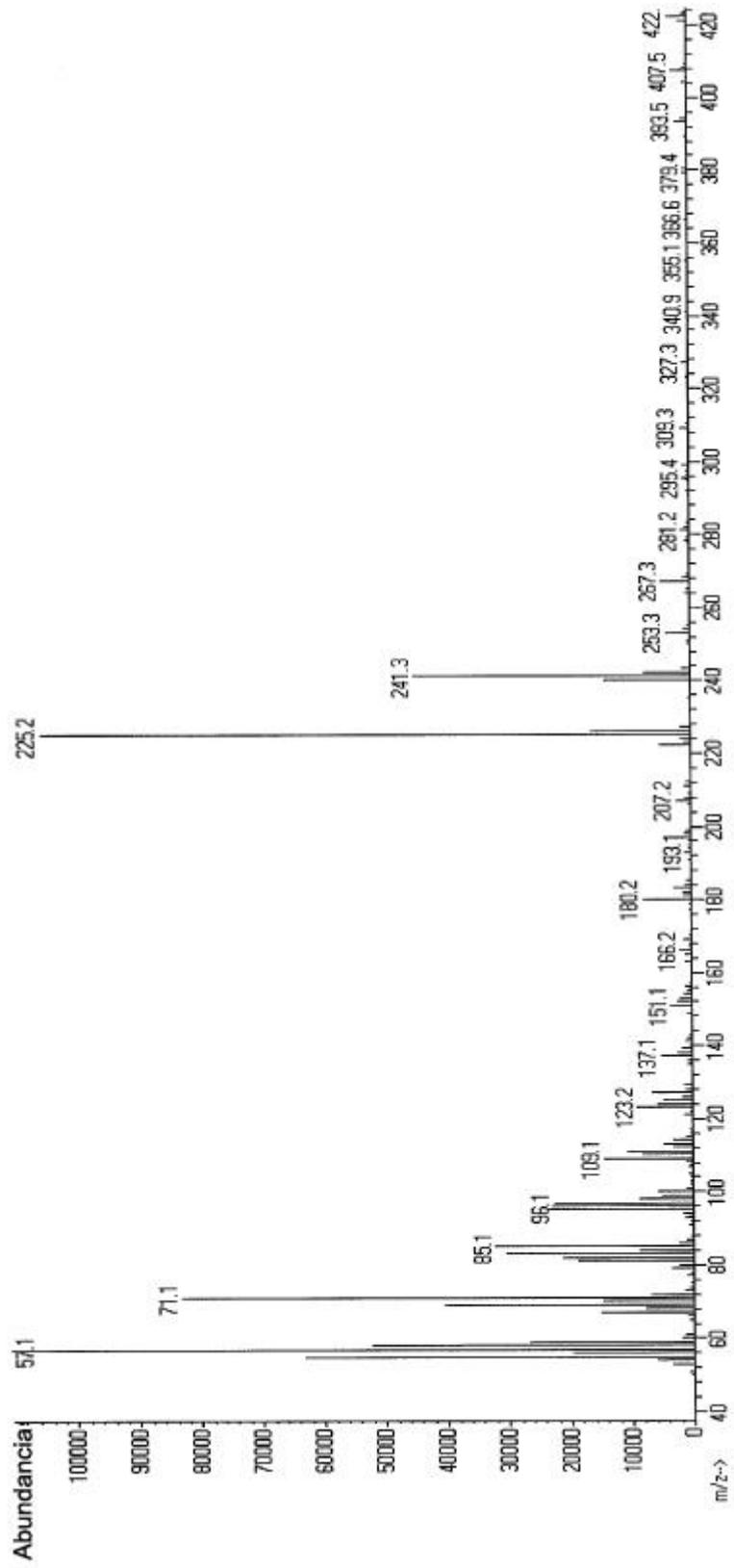


Figura 19A

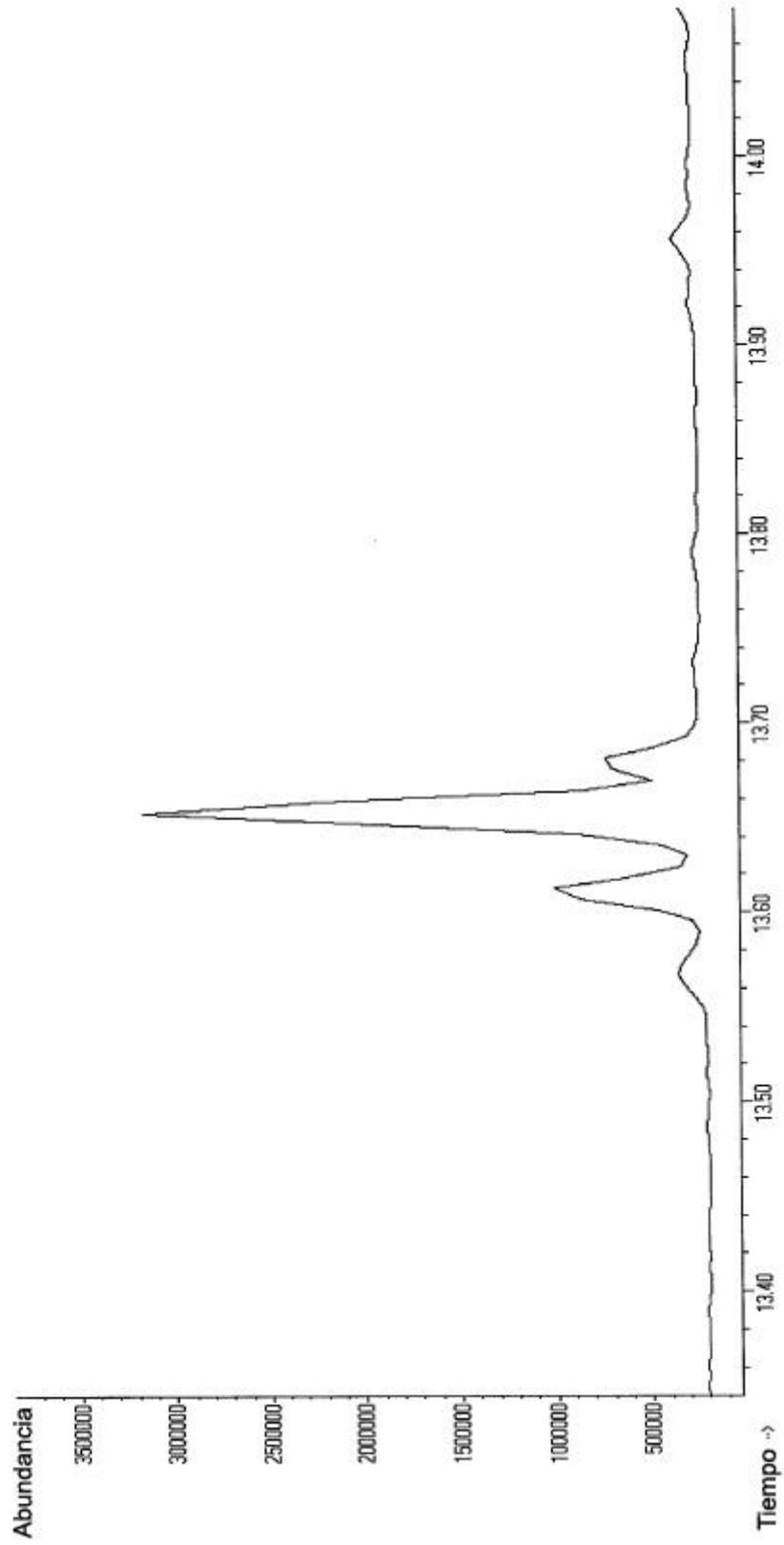


Figura 19B

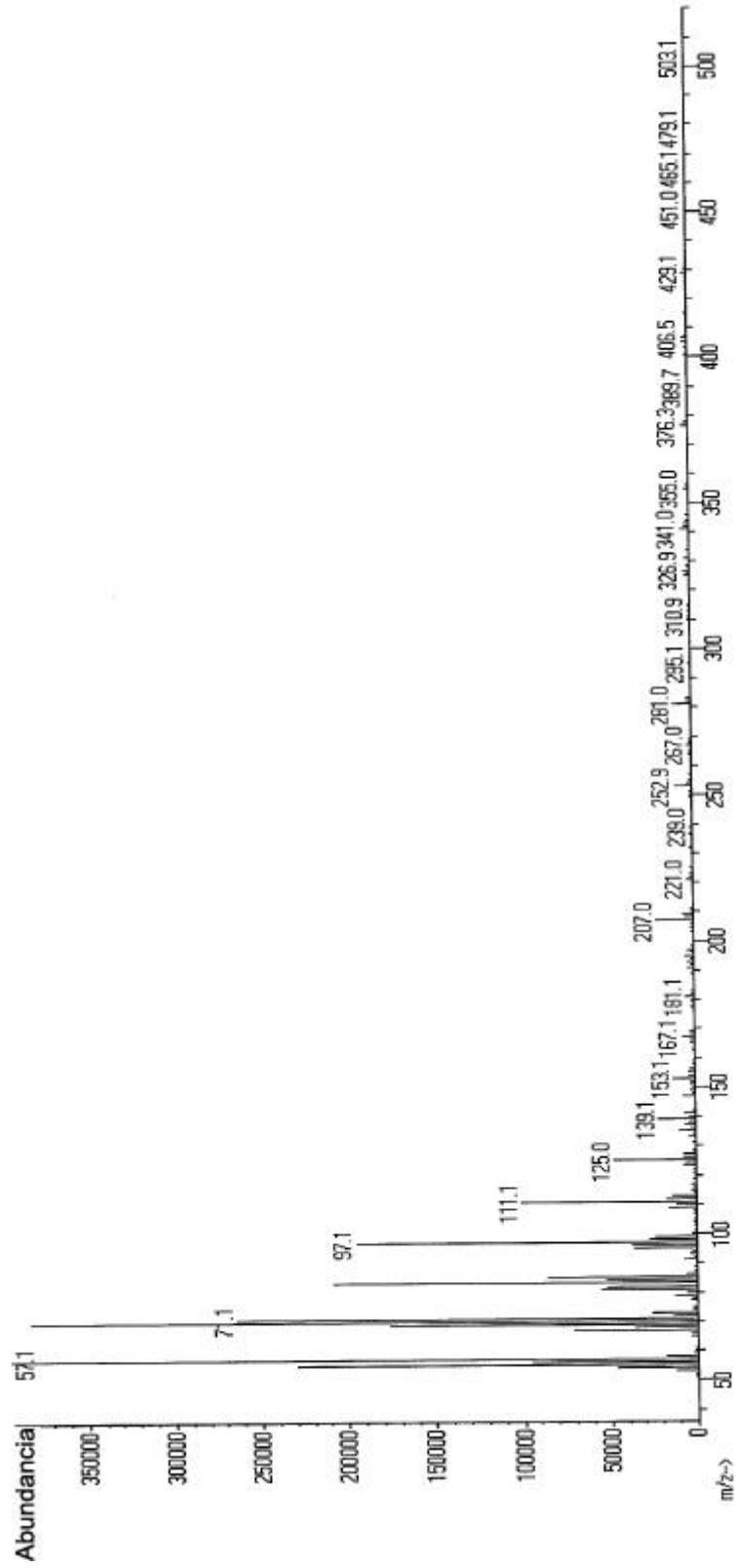


Figura 20

Ruta de biosíntesis de ácido micólico propuesta Ruta de biosíntesis de olefina propuesta

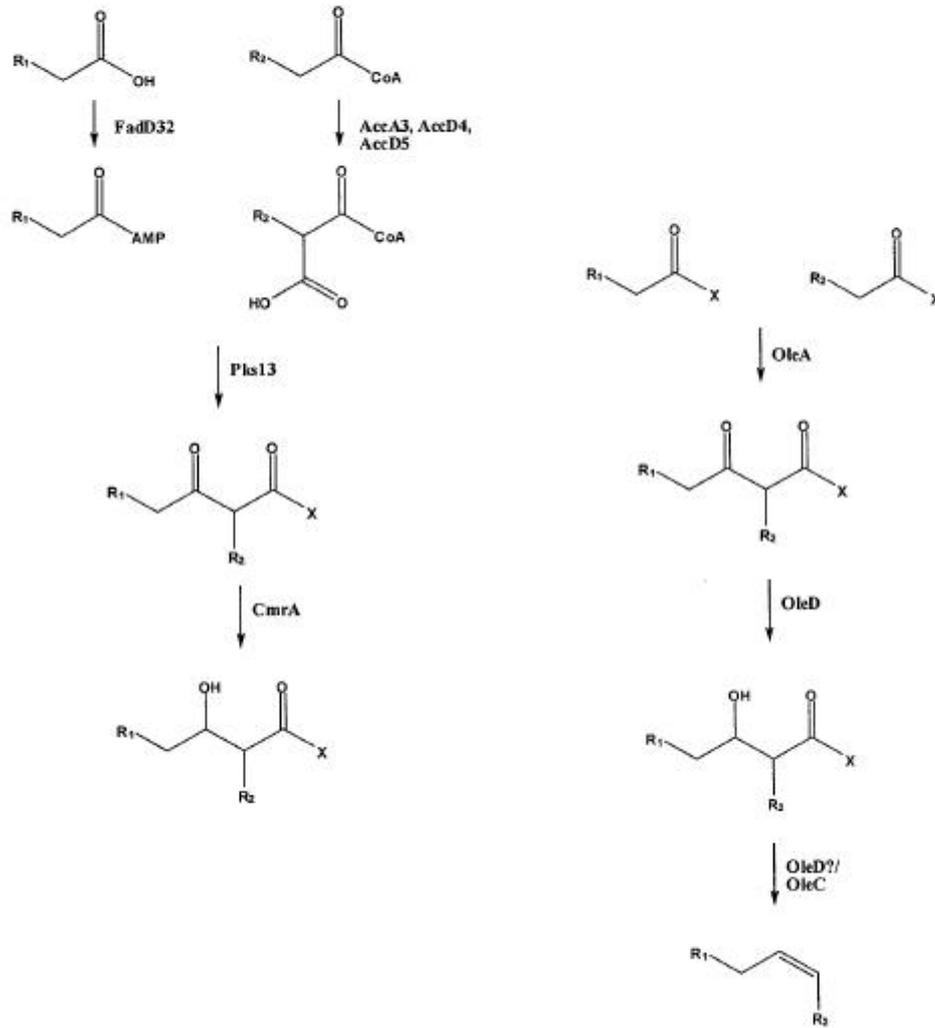


Figura 21

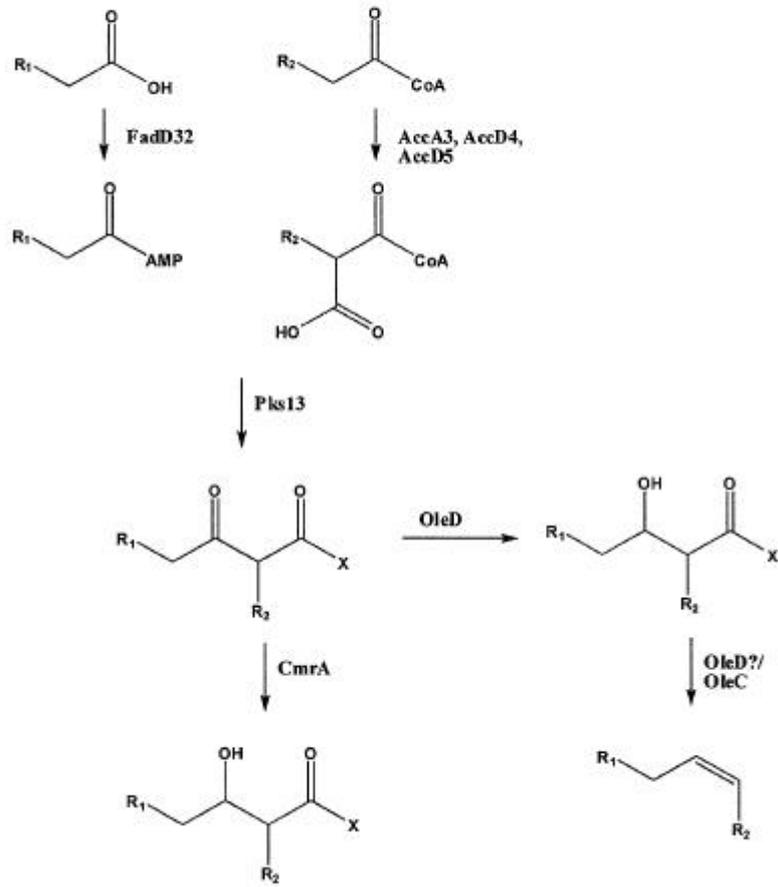


Figura 22

