

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 600 354**

51 Int. Cl.:

C12N 9/02 (2006.01)

C12N 15/53 (2006.01)

C12P 7/04 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **16.01.2014 PCT/US2014/011859**

87 Fecha y número de publicación internacional: **24.07.2014 WO14113571**

96 Fecha de presentación y número de la solicitud europea: **16.01.2014 E 14716444 (6)**

97 Fecha y número de publicación de la concesión europea: **27.07.2016 EP 2946009**

54 Título: **Acil-ACP reductasa con propiedades mejoradas**

30 Prioridad:

16.01.2013 US 201361753273 P

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
08.02.2017

73 Titular/es:

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

72 Inventor/es:

**RUDE, MATHEW;
TRINH, NA;
SCHIRMER, ANDREAS y
GANO, JACOB**

74 Agente/Representante:

MILTENYI, Peter

ES 2 600 354 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

Acil-ACP reductasa con propiedades mejoradas

Campo

5 La divulgación se refiere a variantes de enzima acil-ACP reductasa (AAR) que dan como resultado la producción mejorada de aldehídos grasos y/o alcoholes grasos cuando se expresan en células huésped recombinantes. La divulgación se refiere además a métodos de preparación y uso de tales variantes de AAR para la producción de composiciones de alcohol graso que tiene características particulares.

Antecedentes

10 Los alcoholes grasos indican una importante categoría de productos bioquímicos industriales. Por ejemplo, las ventas anuales a nivel mundial de alcoholes grasos y sus derivados son superiores a mil millones de US\$. Estas moléculas y sus derivados tienen numerosas aplicaciones, incluyendo su uso como tensioactivos, lubricantes, plastificantes, disolventes, emulsionantes, emolientes, espesantes, sabores, fragancias y combustibles. Debido a su naturaleza anfifila, los alcoholes grasos se comportan como tensioactivos no iónicos, que son útiles en productos para el cuidado personal y de uso doméstico, por ejemplo, detergentes. Los alcoholes grasos de cadena más corta se usan en las industrias cosmética y alimentaria como emulsionantes, emolientes y espesantes.

15 En la naturaleza, los alcoholes grasos los producen enzimas que pueden reducir diversas moléculas de acil-ACP o acil-CoA para dar los alcoholes primarios correspondientes (por ejemplo, las patentes estadounidenses n.^{os} 8.323.924; 8.268.599 y 8.097.439; y las publicaciones de patente estadounidense n.^{os} 20120282663 y 20100105963). Sin embargo, las tecnologías actuales implican la reducción mediada por catalizadores en su mayor parte inorgánicos de ácidos grasos para dar los alcoholes primarios correspondientes. Estos alcoholes grasos se producen mediante hidrogenación catalítica de ácidos grasos producidos a partir fuentes naturales, tales como aceite de coco, aceite de palma, aceite de palmiste, sebo y manteca de cerdo, o mediante hidratación química de alfa-olefinas producidas a partir de materias primas petroquímicas. Los alcoholes grasos derivados de fuentes naturales tienen longitudes de cadena variables, que son relevantes y específicos para aplicaciones particulares. La deshidratación de alcoholes grasos para dar alfa-olefinas puede lograrse mediante catálisis química.

20 El documento WO 2011/127409 se refiere a polipéptidos de acil-ACP reductasa (AAR), incluyendo AAR de *Synechococcus elongatus* PCC7942, y la expresión de los mismos en células huésped para producir alcoholes grasos.

30 Pueden usarse aldehídos grasos para producir productos químicos especializados industriales. Por ejemplo, se usan comúnmente aldehídos para producir polímeros, resinas, colorantes, saborizantes, plastificantes, perfumes y productos farmacéuticos. Pueden usarse aldehídos como disolventes, conservantes, y desinfectantes. Determinados compuestos naturales y sintéticos, tales como vitaminas y hormonas, son aldehídos, y muchos azúcares contienen grupos aldehído. Los aldehídos grasos pueden convertirse en alcoholes grasos mediante reducción química o enzimática.

35 Una alternativa más ecológica y más limpia para la producción de aldehídos grasos y alcoholes grasos es mediante biomasa y/o azúcares fermentables. Sin embargo, para que la producción de aldehídos grasos y alcoholes grasos a partir de biomasa o azúcares fermentables sea comercialmente viable, deben optimizarse procesos industriales para una conversión y recuperación eficientes del producto final. La presente divulgación aborda esta necesidad proporcionando composiciones y métodos para la producción mejorada de aldehídos grasos y alcoholes grasos usando células huésped modificadas por ingeniería como biocatalizadores.

Sumario

45 La presente divulgación se refiere a células huésped fotosintéticas y heterotróficas que producen directamente aldehídos grasos y/o alcoholes grasos de longitudes de cadena específicas de manera que no es necesaria la conversión catalítica de ácidos grasos purificados. Esta ruta biológica proporciona un producto de mayor calidad, una reducción significativa de los costes y un menor impacto sobre el medio ambiente. Más específicamente, la presente divulgación se refiere a variantes de enzima acil-ACP reductasa (AAR) novedosas que producen aldehídos grasos y/o alcoholes grasos y composiciones de los mismos. También se contemplan secuencias de ácido nucleico y proteína de variantes de AAR específicas así como células huésped recombinantes novedosas y cultivos celulares que engloban tales variantes de enzima AAR modificadas por ingeniería. La divulgación también engloba métodos de uso de las células huésped que expresan variantes de AAR recombinantes para producir composiciones de aldehído graso y/o alcohol graso con características particulares.

50 Un aspecto de la divulgación engloba polipéptidos de acil-ACP reductasa (AAR) variantes que catalizan la conversión de una acil-ACP en un aldehído graso, en los que el polipéptido de AAR tiene una identidad de secuencia de al menos el 90% con la secuencia de polipéptido de AAR silvestre correspondiente presentada como SEQ ID NO: 28, SEQ ID NO: 30, SEQ ID NO: 32, SEQ ID NO: 34, SEQ ID NO: 36, SEQ ID NO: 38, SEQ ID NO: 40, SEQ ID NO: 42 o SEQ ID NO: 44, y métodos para expresar los polipéptidos de AAR variantes en una célula

huésped recombinante, lo que da como resultado un mayor título de composición de aldehído graso y/o alcohol graso en comparación con el título de una composición de aldehído graso y/o alcohol graso producida mediante la expresión de un polipéptido de AAR silvestre en una célula huésped correspondiente. El polipéptido de AAR variante modificado por ingeniería genética puede tener una identidad de secuencia de al menos el 90% con la secuencia de polipéptido de AAR silvestre correspondiente presentada como SEQ ID NO: 28 y la expresión del polipéptido de AAR variante en una célula huésped recombinante da como resultado un mayor título de composición de aldehído graso y/o alcohol graso o un mayor título de alcoholes grasos C₁₂, C₁₄ o C₁₆ en comparación con el título producido mediante la expresión de un polipéptido de AAR silvestre en una célula huésped correspondiente.

En un aspecto de la divulgación, el polipéptido de AAR variante tiene una mutación en una o más posiciones de aminoácido de los aminoácidos 18, 24, 31, 34, 35, 43, 50, 63, 86, 112, 113, 116, 118, 120, 135, 148, 153, 155, 157, 159, 168, 172, 187, 188, 191, 209, 210, 211, 236, 277, 283, 285, 291, 324, 328, 335, 337 y 338 de SEQ ID NO: 28. El polipéptido de AAR variante modificado por ingeniería genética puede tener una mutación S18W. En un aspecto preferido, el polipéptido de AAR variante modificado por ingeniería genética tiene una mutación S18W y comprende además una mutación tal como M21L, D24E, D24Y, L31V, W34F, W35F, D43E, A50Q, C63A, C63G, C63Y, S86G, A112R, S113K, Q116G, R118Q, T120S, A135S, T148C, T148E, T148V, I153P, Q155C, Q155L, T157V, A159V, I168V, C172L, T187V, T188H, T188V, Q191A, L209R, E210Y, A211W, T236C, Q277V, E283G, E283S, A285V, M291V, A324T, A328S, Q335N, L337V y/o L338W.

En otro aspecto de la divulgación, el polipéptido de AAR variante que tiene una identidad de secuencia de al menos el 90% con la secuencia de polipéptido de AAR silvestre correspondiente presentada como SEQ ID NO: 34 y la expresión del polipéptido de AAR variante en una célula huésped recombinante da como resultado un mayor título de aldehído graso y/o alcohol graso o un mayor título de alcohol graso C₁₂ en comparación con el título producido mediante la expresión de un polipéptido de AAR silvestre en una célula huésped correspondiente. El polipéptido de AAR variante tiene una mutación en una posición de aminoácido incluyendo el aminoácido 40, 52, 58, 61, 273, 303, 339, 340, 344, 345, 346 y 588 de SEQ ID NO: 34. El polipéptido de AAR variante puede tener una mutación en la posición de aminoácido Q40V, G52V, S58V, D61E, G273E, K303G, K339L, H340P, L344A, L344D, L344S, L344T, L345R, V346P, V346G y/o S588V.

Otro aspecto de la divulgación engloba una célula huésped recombinante que tiene una o más mutaciones tal como se describió anteriormente y en la que cuando se modifica por ingeniería la célula huésped para expresar un polipéptido de AAR variante de SEQ ID NO: 28 o SEQ ID NO: 34, esta célula huésped recombinante produce una composición de aldehído graso y/o alcohol graso con un título que es al menos el 10% mayor, al menos el 15% mayor, al menos el 20% mayor, al menos el 25% mayor, o al menos el 30% mayor que el título de una composición de aldehído graso y/o alcohol graso producida por una célula huésped que expresa el polipéptido de AAR silvestre correspondiente, cuando se cultiva en medio que contiene una fuente de carbono en condiciones eficaces para expresar el polipéptido de AAR variante. La composición de aldehído graso y/o alcohol graso puede producirse a un título de 30 g/l a 250 g/l, por ejemplo, un título de al menos 100 mg/l. La composición de alcohol graso puede producirse de manera extracelular.

La divulgación engloba además un cultivo celular que incluye la célula huésped recombinante tal como se describió anteriormente, en el que la composición de alcohol graso incluye uno o más de un alcohol graso C₆, C₈, C₁₀, C₁₂, C₁₃, C₁₄, C₁₅, C₁₆, C₁₇, y C₁₈, por ejemplo, un alcohol graso insaturado C_{10:1}, C_{12:1}, C_{14:1}, C_{16:1}, o C_{18:1}. La composición de alcohol graso puede comprender un alcohol graso saturado.

La presente invención se define mediante las reivindicaciones.

La presente invención proporciona un polipéptido de acil-ACP reductasa (AAR) variante que tiene una identidad de secuencia de al menos el 90% con SEQ ID NO: 57, en el que el polipéptido de AAR variante comprende una mutación en la posición de aminoácido 18, en el que la mutación es S18W, y en el que el polipéptido de AAR cataliza la conversión de una acil-ACP en un aldehído graso. En una realización, la expresión del polipéptido de AAR variante en una célula huésped recombinante da como resultado un mayor título de una composición de aldehído graso o alcohol graso en comparación con el título de una composición de aldehído graso o alcohol graso producida mediante la expresión de un polipéptido de AAR silvestre en una célula huésped silvestre correspondiente. En otra realización, la expresión del polipéptido de AAR variante en una célula huésped recombinante da como resultado un mayor título de una composición de aldehído graso o alcohol graso que es una composición de alcohol graso C₁₂, C₁₄ y/o C₁₆ en comparación con el título de una composición de aldehído graso o alcohol graso producida mediante la expresión de un polipéptido de AAR silvestre en una célula huésped silvestre correspondiente.

En realizaciones preferidas, el polipéptido de acil-ACP reductasa (AAR) variante que tiene una identidad de secuencia de al menos el 90% con SEQ ID NO: 57 y que tiene una mutación en la posición de aminoácido 18, en el que la mutación es S18W, tiene otra mutación en una posición de aminoácido en el aminoácido 8, 16, 21, 24, 31, 34, 35, 43, 50, 63, 86, 112, 113, 116, 118, 120, 135, 148, 153, 154, 155, 157, 159, 168, 172, 187, 188, 191, 209, 210, 211, 236, 277, 281, 283, 285, 291, 324, 328, 335, 337 y/o 338. En una realización, la mutación se selecciona de L8A, D16L, M21L, D24E, D24Y, D24V, D24P, L31V, L31M, W34F, W35F, D43E, A50Q, C63A, C63G, C63Y, S86G, A112R, S113K, Q116G, R118Q, T120S, A135S, T148C, T148E, T148V, I153P, T154A, Q155C, Q155L, T157V, A159V, I168V, C172L, T187V, T188H, T188V, Q191A, L209R, E210Y, A211W, T236C, Q277V, A281L, E283G,

- E283S, A285V, M291V, A324T, A328S, Q335N, L337V y/o L338W. En una realización preferida, el polipéptido de AAR variante tiene una mutación M21L, una mutación C63G, una mutación S113K, T154A y una mutación A281L (SEQ ID NO: 58). En otra realización preferida, el polipéptido de AAR variante tiene una mutación L8A, una mutación M21L, una mutación C63G, una mutación S113K, una mutación T154A y una mutación A281L (SEQ ID NO: 59). En otra realización preferida, el polipéptido de AAR variante tiene una mutación D16L, una mutación M21L, una mutación C63G, una mutación S113K, T154A y una mutación A281L (SEQ ID NO: 60). En otra realización preferida, el polipéptido de AAR variante tiene una mutación L8A, una mutación D24V, una mutación C63G, una mutación S113K, una mutación Q155L y una mutación A281L (SEQ ID NO: 61). En otra realización preferida, el polipéptido de AAR variante tiene una mutación D24P, una mutación L31M, una mutación C63G, una mutación S113K, una mutación T154A y una mutación A281L (SEQ ID NO: 62). En otra realización preferida, el polipéptido de AAR variante tiene una mutación L8A, una mutación D16L, una mutación D24V, una mutación C63G, una mutación S113K, una mutación T154A y una mutación A281L (SEQ ID NO: 63). En otra realización preferida, el polipéptido de AAR variante tiene una mutación D24E, una mutación C63G, una mutación S113K, una mutación T154A y una mutación A281L (SEQ ID NO: 64).
- Otro aspecto de la presente divulgación engloba una célula huésped recombinante que expresa los polipéptidos de AAR variantes tal como se describió anteriormente (véase antes). La célula huésped recombinante puede producir una composición de aldehído graso o alcohol graso con un título que es al menos el 10% mayor, al menos el 15% mayor, al menos el 20% mayor, al menos el 25% mayor o al menos el 30% mayor que el título de una composición de aldehído graso o alcohol graso producida por una célula huésped que expresa un polipéptido de AAR silvestre correspondiente, cuando se cultiva en medio que contiene una fuente de carbono en condiciones eficaces para expresar el polipéptido de AAR variante. La composición de aldehído graso o alcohol graso producida por la célula huésped recombinante puede producirse a un título de aproximadamente 30 g/l a aproximadamente 250 g/l. La composición de aldehído graso o alcohol graso puede producirse de manera extracelular.
- La divulgación contempla además un cultivo celular que incluye la célula huésped recombinante que expresa los polipéptidos de AAR variantes tal como se describió anteriormente (véase antes). La composición de alcohol graso puede incluir un alcohol graso saturado y/o insaturado. El cultivo celular puede incluir una composición de alcohol graso que incluye uno o más de un alcohol graso C₆, C₈, C₁₀, C₁₂, C₁₃, C₁₄, C₁₅, C₁₆, C₁₇, y C₁₈. La composición de alcohol graso puede incluir uno o más de un alcohol graso insaturado C_{10:1}, C_{12:1}, C_{14:1}, C_{16:1}, y C_{18:1}. La composición de alcohol graso puede incluir un alcohol graso que tiene un doble enlace en la posición 7 en la cadena de carbonos entre C₇ y C₈ desde el extremo reducido del alcohol graso.
- La divulgación engloba además un método de producción de una composición de alcohol graso que tiene un aumento del título, que incluye cultivar la célula huésped que expresa AAR variante (tal como se describió anteriormente) con una fuente de carbono; y recoger una composición de alcohol graso. El título del alcohol graso es de al menos del 20% al 30% mayor que el título de una composición de alcohol graso producida por una célula huésped que expresa AAR silvestre.
- Otro aspecto de la divulgación engloba un polipéptido de acil-ACP reductasa (AAR) variante que tiene una identidad de secuencia de al menos el 90% con SEQ ID NO: 65, en el que el polipéptido cataliza la conversión de una acil-ACP en un aldehído graso. La expresión del polipéptido de AAR variante en una célula huésped recombinante puede dar como resultado un mayor título de una composición de aldehído graso o alcohol graso en comparación con el título de una composición de aldehído graso o alcohol graso producida mediante la expresión de un polipéptido de AAR silvestre en una célula huésped silvestre correspondiente. La expresión del polipéptido de AAR variante en una célula huésped recombinante puede dar como resultado un mayor título de una composición de alcohol graso C₁₂, C₁₄ y/o C₁₆ en comparación con el título de una composición de aldehído graso o alcohol graso producida mediante la expresión de un polipéptido de AAR silvestre en una célula huésped silvestre correspondiente. En un aspecto particular, la divulgación engloba un polipéptido de acil-ACP reductasa (AAR) variante que tiene una identidad de secuencia de al menos el 90% con SEQ ID NO: 65, en el que el polipéptido tiene una mutación en la posición de aminoácido 61. La mutación puede ser D61E.
- La divulgación engloba además un polipéptido de acil-ACP reductasa (AAR) variante que tiene una identidad de secuencia de al menos el 90% con SEQ ID NO: 34, en el que expresión del polipéptido de AAR variante en una célula huésped recombinante da como resultado un mayor título de una composición de aldehído graso o alcohol graso o un mayor título de una composición de alcohol graso C₁₂, C₁₄ y/o C₁₆ en comparación con el título de una composición de alcohol graso producida mediante la expresión de un polipéptido de AAR silvestre en una célula huésped silvestre correspondiente, y en el que el polipéptido de AAR tiene una mutación en la posición de aminoácido 40, 52, 273, 303, 340, 344, 345 o 346. El polipéptido de AAR variante puede tener una mutación seleccionada de Q40V, G52V, G273E, K303G, H340P, L344A, L344D, L344S, L344T, L345R, V346P y V346G. El polipéptido de AAR variante puede tener una mutación en V346P (SEQ ID NO: 66), en Q40V (SEQ ID NO: 67), en A345R (SEQ ID NO: 68), en L344S (SEQ ID NO: 69), en V346G (SEQ ID NO: 70), en L344D (SEQ ID NO: 71), en G52V (SEQ ID NO: 72), en L344T (SEQ ID NO: 73), en K303G (SEQ ID NO: 74), L344A (SEQ ID NO: 75), en H340P (SEQ ID NO: 76) o en G273E (SEQ ID NO: 77).
- Aún otro aspecto de la divulgación engloba una célula huésped recombinante que expresa el polipéptido de AAR variante tal como se describió anteriormente (véase antes). La célula huésped recombinante puede producir una

composición de aldehído graso o alcohol graso con un título que es al menos el 10% mayor, al menos el 15% mayor, al menos el 20% mayor, al menos el 25% mayor, o al menos el 30% mayor que el título de una composición de aldehído o alcohol graso producida por una célula huésped que expresa un polipéptido de AAR silvestre correspondiente, cuando se cultiva en medio que contiene una fuente de carbono en condiciones eficaces para expresar el polipéptido de AAR variante. La composición de alcohol graso puede producirse a un título de aproximadamente 30 g/l a aproximadamente 250 g/l. La composición de alcohol graso puede producirse de manera extracelular.

La divulgación contempla además un cultivo celular con la célula huésped recombinante que expresa el polipéptido de AAR variante tal como se describió anteriormente (véase antes). La composición de alcohol graso puede incluir uno o más de un alcohol graso C₆, C₈, C₁₀, C₁₂, C₁₃, C₁₄, C₁₅, C₁₆, C₁₇, y C₁₈. La composición de alcohol graso puede incluir un alcohol graso insaturado o saturado. La composición de alcohol graso puede incluir uno o más de un alcohol graso C_{10:1}, C_{12:1}, C_{14:1}, C_{16:1}, y uno insaturado C_{18:1}. La composición de alcohol graso puede incluir un alcohol graso que tiene un doble enlace en la posición 7 en la cadena de carbonos entre C₇ y C₈ desde el extremo reducido del alcohol graso.

Otro aspecto de la divulgación engloba un método de producción de una composición de alcohol graso que tiene un aumento del título, que incluye cultivar la célula huésped que expresa la AAR (tal como se describió anteriormente) con una fuente de carbono; y recoger una composición de alcohol graso. El alcohol graso puede ser de al menos del 20% al 30% mayor que el título de una composición de alcohol graso producida por una célula huésped que expresa AAR silvestre.

20 Breve descripción de los dibujos

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

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

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

La figura 3 ilustra la estructura y función del complejo enzimático de acetil-CoA carboxilasa (accABCD). La biotina carboxilasa está codificada por el gen *accC*, mientras que la proteína transportadora de biotina y carboxilo (BCCP) está codificada por el gen *accB*. Las dos subunidades implicadas en la actividad carboxilo transferasa están codificadas por los genes *accA* y *accD*. La biotina unida covalentemente de BCCP transporta el resto carboxilato. El gen *birA* biotinila holo-*accB*.

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

La figura 5 muestra la producción de alcoholes grasos en *E. coli* DV2 que expresa acil-ACP reductasa de *Synechococcus elongatus* (AAR_7942) y que coexpresa diversas proteínas transportadoras de acilo (ACP) cianobacterianas.

La figura 6 muestra la producción de alcoholes grasos en *E. coli* DV2 que expresa acil-ACP reductasa de *Synechococcus elongatus* (AAR_7942) (pLS9-185) y que coexpresa el complejo de acetil carboxilasa de *Corynebacterium glutamicum* (pSL9-185-D+) (se muestran tres cepas individuales, véanse D+1, D+2 y D+3).

La figura 7 presenta resultados que ilustran la producción mejorada de alcoholes grasos de células huésped recombinantes que se basan en la sobreexpresión de AAR mediada por *ifab* e *ifadR*.

La figura 8 presenta resultados que muestran niveles de alcoholes grasos elevados en células huésped recombinantes que expresan variantes de AAR_7942 derivadas de bibliotecas de combinación (AAR_Com 2a-d) que se coexpresan con ACP, AlrA (alcohol deshidrogenasa (ADH)) y un operón acc sintético en la cepa Shu2. Son útiles varios polipéptidos de alcohol deshidrogenasa según la divulgación e incluyen, pero no se limitan a AlrA de *Acinetobacter* sp. M-1 (SEQ ID NO: 52) y un homólogo de AlrA tal como AlrAadp1 (SEQ ID NO: 53)

Las figuras 9A y 9B presentan resultados de una fermentación en tanque de la variante de AAR_7942 S18W que ilustran el título de FALC (9A) y el rendimiento con respecto a glucosa (9B) que muestran que la expresión de la variante de AAR_7942 S18W da como resultado actividad y distribución de longitud de cadena alteradas.

La figura 10 presenta resultados que ilustran un desplazamiento en la distribución de longitud de cadena para FALC desde C₁₆ hasta C₁₄ cuando la variante de D61E de MED4_AAR se expresó en células huésped recombinantes.

Descripción detallada

Visión general

Un modo de eliminar la dependencia de los productos petroquímicos es producir derivados de ácido graso tales como aldehídos grasos y alcoholes grasos a través de microorganismos respetuosos con el medio ambiente que sirven como huéspedes de producción en miniatura. Tales huéspedes celulares (es decir, células huésped recombinantes o cepas de producción) se han modificado por ingeniería para producir aldehídos grasos y/o alcoholes grasos a partir de fuentes renovables tales como materia prima renovable (por ejemplo, azúcares fermentables, hidratos de carbono, biomasa, celulosa, glicerol, CO, CO₂, etc.). Estos aldehídos grasos y alcoholes grasos son los materiales de partida para muchos productos industriales incluyendo detergentes y combustibles.

La presente divulgación se refiere a variantes de enzima acil-ACP reductasa (AAR) que dan como resultado título, rendimiento y/o productividad mejorados de composiciones de aldehído graso y/o alcohol graso cuando se expresan en células huésped recombinantes. En el presente documento, se logra la biosíntesis potenciada de aldehído graso y/o alcohol graso transformando células huésped de manera que expresen una proteína acil-ACP reductasa (AAR) variante, que cataliza la reacción de una acil-ACP para dar un aldehído graso y/o un alcohol graso. La divulgación se refiere además a las células huésped recombinantes o cepas de producción que expresan las variantes de enzima AAR.

Definiciones

Tal como se usa en esta memoria descriptiva y las reivindicaciones adjuntas, las formas en singular “un/o”, “una” y “el/la” incluyen referentes en plural a menos que el contexto dicte claramente lo contrario. Por tanto, por ejemplo, la referencia a “una célula huésped” incluye dos o más de tales células huésped, la referencia a “un éster graso” incluye uno o más ésteres grasos, o mezclas de ésteres, la referencia a “una secuencia de ácido nucleico” incluye una o más secuencias de ácido nucleico, la referencia a “una enzima” incluye una o más enzimas, y similares.

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

Se establecen los números de clasificación de enzimas (EC) por el Comité de Nomenclatura de la Unión Internacional de Bioquímica y Biología Molecular (IUBMB), cuya descripción está disponible en el sitio web de Nomenclatura de Enzimas del IUBMB en Internet. Los números EC clasifican las enzimas según la reacción que catalizan. Por ejemplo, la actividad enzimática acil-ACP reductasa (AAR) se clasifica según E.C. 1.2.1.80 (también conocida como acil-[proteína transportadora de acilo] de cadena larga reductasa) o EC 1.2.1.42. La funcionalidad de AAR se conserva en la mayor parte de procariotas de una especie a la siguiente. Por tanto, diferentes especies microbianas pueden portar la misma actividad enzimática AAR que se clasifica según E.C. 1.2.1.80 o EC 1.2.1.42.

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

El término “polinucleótido” se refiere a un polímero de ribonucleótidos (ARN) o deoxirribonucleótidos (ADN), que puede ser monocatenario o bicatenario y que puede contener nucleótidos no naturales o alterados. Los términos

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

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

Tal como se usa en el presente documento, los términos “homólogo” y “homólogo/a” se refieren a un polinucleótido o un polipéptido que comprende una secuencia que es idéntica en al menos aproximadamente el 50 por ciento (%) a la secuencia de polinucleótido o polipéptido correspondiente. Preferiblemente, los polinucleótidos o polipéptidos homólogos tienen secuencias de polinucleótido o secuencias de aminoácidos que tienen una homología de al menos aproximadamente el 80%, el 81%, el 82%, el 83%, el 84%, el 85%, el 86%, el 87%, el 88%, el 89%, el 90%, el 91%,
 20 el 92%, el 93%, el 94%, el 95%, el 96%, el 97%, el 98% o al menos aproximadamente el 99% con la secuencia de aminoácidos o secuencia de polinucleótido correspondiente. Tal como se usa en el presente documento los términos “homología” de secuencia e “identidad” de secuencia se usan de manera intercambiable. Un experto habitual en la técnica conoce métodos para determinar la homología entre dos o más secuencias. En resumen, pueden realizarse cálculos de “homología” entre dos secuencias de la siguiente manera. Las secuencias se alinean para fines de comparación óptima (por ejemplo, pueden introducirse huecos en una o ambas de una primera y una segunda secuencias de aminoácidos o de ácido nucleico para la alineación óptima y pueden excluirse las secuencias no homólogas para fines de comparación). En una realización preferida, la longitud de una primera secuencia que se alinea para fines de comparación es de al menos aproximadamente el 30%, preferiblemente al menos aproximadamente el 40%, más preferiblemente al menos aproximadamente el 50%, incluso más preferiblemente al menos aproximadamente el 60%, y incluso más preferiblemente al menos aproximadamente el 70%, al menos aproximadamente el 80%, al menos aproximadamente el 85%, al menos aproximadamente el 90%, al menos aproximadamente el 95%, al menos aproximadamente el 98% o aproximadamente el 100% de la longitud de una segunda secuencia. Se comparan entonces los residuos de aminoácido o nucleótidos en posiciones de aminoácido o posiciones de nucleótido correspondientes de las secuencias primera y segunda. Cuando una posición en la primera secuencia está ocupada por el mismo residuo de aminoácido o nucleótido que la posición correspondiente en la segunda secuencia, entonces las moléculas son idénticas en esa posición. El tanto por ciento de homología entre las dos secuencias es una función del número de posiciones idénticas compartidas por las secuencias, teniendo en cuenta el número de huecos y la longitud de cada hueco, que es necesario introducir para la alineación óptima de las dos secuencias. La comparación de secuencias y determinación del tanto por ciento de homología entre dos secuencias puede lograrse usando un algoritmo matemático, tal como BLAST (Altschul *et al.* (1990) J. Mol. Biol. 215(3):403-410). El tanto por ciento de homología entre dos secuencias de aminoácidos también puede determinarse usando el algoritmo de Needleman y Wunsch que se ha incorporado en el programa GAP en el paquete de software de GCG, usando o bien una matriz Blossum 62 o bien una matriz PAM250, y un peso de hueco de 16, 14, 12, 10, 8, 6 ó 4 y un peso de longitud de 1, 2, 3, 4, 5 ó 6 (Needleman y Wunsch (1970) J. Mol. Biol. 48:444-453). El tanto por ciento de homología entre dos secuencias de nucleótidos también puede determinarse usando el programa GAP en el paquete de software de GCG, usando una matriz NWSgapdna.CMP y un peso de hueco de 40, 50, 60, 70 u 80 y un peso de longitud de 1, 2, 3, 4, 5 ó 6. Un experto habitual en la técnica puede realizar cálculos de homología inicial y ajustar los parámetros del algoritmo en consecuencia. Un conjunto preferido de parámetros (y el que debe usarse si un profesional no está seguro de qué parámetros deben aplicarse para determinar si una molécula está dentro de una limitación de homología de las reivindicaciones) son una matriz de puntuación Blossum 62 con una penalización por hueco de 12, penalización por extensión de hueco de 4 y penalización por hueco con desplazamiento del marco de lectura de 5. Se conocen métodos adicionales de alineación de secuencias en las técnicas biotecnológicas (véase, por ejemplo, Rosenberg (2005) BMC Bioinformatics 6:278; Altschul *et al.* (2005) FEBS J. 272(20):5101-5109).

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

60°C; (3) condiciones de hibridación de rigurosidad alta -- 6X SSC a aproximadamente 45°C, seguido por uno o más lavados en 0,2X SSC, SDS al 0,1% a 65°C; y (4) condiciones de hibridación de rigurosidad muy alta -- fosfato de sodio 0,5 M, 7% SDS a 65°C, seguido por uno o más lavados a 0,2X SSC, SDS al 1% a 65°C. Las condiciones de rigurosidad muy alta (4) son las condiciones preferidas a menos que se especifique de otro modo.

5 Un polipéptido “endógeno” se refiere a un polipéptido codificado por el genoma de la célula parental (o célula huésped). Un polipéptido “exógeno” se refiere a un polipéptido que no está codificado por el genoma de la célula parental. Un polipéptido variante o mutante es un ejemplo de un polipéptido exógeno. Por tanto, se considera que una molécula de ácido nucleico que se produce de manera no natural es exógena para una célula una vez introducida en la célula. Una molécula de ácido nucleico que se produce de manera natural también puede ser
10 exógena para una célula particular. Por ejemplo, una secuencia codificante completa aislada de la célula X es un ácido nucleico exógeno con respecto a la célula Y una vez que se introduce la secuencia codificante en la célula Y, aunque X e Y sean el mismo tipo celular.

El término “sobreexpresado” significa que se hace que se transcriba un gen a una tasa elevada en comparación con la tasa de transcripción endógena para ese gen. En algunos ejemplos, la sobreexpresión incluye adicionalmente una
15 tasa de traducción elevada del gen en comparación con la tasa de transcripción endógena para ese gen. Se conocen bien en la técnica métodos de prueba de la sobreexpresión, por ejemplo pueden evaluarse los niveles de ARN transcrito usando rtPCR y pueden evaluarse los niveles de proteína usando análisis en gel mediante SDS-PAGE.

El término “heterólogo” significa derivados de un organismo diferente, tipo celular diferente o especie diferente. Tal como se usa en el presente documento, se refiere a una secuencia de nucleótidos, de polinucleótido, de polipéptido o de proteína, no presente de manera natural en un organismo dado. Por ejemplo, una secuencia de polinucleótido que es nativa para cianobacterias puede introducirse en una célula huésped de *E. coli* mediante métodos recombinantes, y el polinucleótido de cianobacterias es entonces heterólogo para la célula de *E. coli* (por ejemplo, célula recombinante). El término “heterólogo” también puede usarse con referencia a una secuencia de nucleótidos,
20 de polinucleótido, de polipéptido o de proteína que está presente en una célula huésped recombinante en un estado no nativo. Por ejemplo, una secuencia de nucleótido, polinucleótido, polipéptido o proteína “heteróloga” puede modificarse con relación a la secuencia silvestre presente de manera natural en la célula huésped silvestre correspondiente, por ejemplo, una modificación en el nivel de expresión o en la secuencia de un nucleótido, polinucleótido, polipéptido o proteína.

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

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

Una “mutación”, tal como se usa en el presente documento, se refiere a un cambio permanente en una posición de ácido nucleico de un gen o en una posición de aminoácido de un polipéptido o una proteína. Las mutaciones incluyen sustituciones, adiciones, inserciones y deleciones. Por ejemplo, una mutación en una posición de aminoácido puede ser una sustitución de un tipo de aminoácido por otro tipo de aminoácido (por ejemplo, puede sustituirse una serina (S) por una alanina (A); puede sustituirse una lisina (L) por una T (treonina); etc.). Como tal, un polipéptido o una proteína pueden tener una o más mutaciones en las que se sustituye un aminoácido por otro aminoácido.
45

Los términos “variante de acil-ACP reductasa (AAR)” y “acil-ACP reductasa (AAR) variante” se usan de manera intercambiable en el presente documento y significan un polipéptido o una proteína relacionados con AAR que tienen una o más mutaciones en su secuencia de aminoácidos. La AAR se refiere a una enzima que cataliza la reducción de una acil-ACP para dar un aldehído graso y/o un alcohol graso. La variante de AAR puede englobar una mutación en uno o más aminoácidos de su secuencia de polipéptido. Cuando se ha transformado una célula con una variante de AAR es una célula que expresa la variante de AAR (por ejemplo, una célula recombinante). En una realización, el título y/o rendimiento de un alcohol graso producido por una célula que expresa la variante de AAR es al menos el doble del de una célula silvestre correspondiente (es decir, una célula correspondiente que no expresa la variante de AAR). En un huésped heterólogo tal como *Escherichia coli*, pueden convertirse aldehídos grasos en alcoholes grasos por alcohol deshidrogenasas endógenas. En otra realización, el título y/o rendimiento de un alcohol graso producido por una célula que expresa la variante de AAR es al menos aproximadamente 1 vez, al menos aproximadamente 2 veces, al menos aproximadamente 3 veces, al menos aproximadamente 4 veces, al menos aproximadamente 5 veces, al menos aproximadamente 6 veces, al menos aproximadamente 7 veces, al menos aproximadamente 8 veces, al menos aproximadamente 9 veces o al menos aproximadamente 10 veces mayor que el de una célula silvestre correspondiente. En una realización, el título y/o rendimiento de un alcohol graso producido
60

por una célula que expresa la variante de AAR es al menos aproximadamente el 1 por ciento, al menos aproximadamente el 2 por ciento, al menos aproximadamente el 3 por ciento, al menos aproximadamente el 4 por ciento, al menos aproximadamente el 5 por ciento, al menos aproximadamente el 6 por ciento, al menos aproximadamente el 7 por ciento, al menos aproximadamente el 8 por ciento, al menos aproximadamente el 9 por ciento o aproximadamente el 10 por ciento mayor que el de una célula silvestre correspondiente. En otra realización, el título y/o rendimiento de alcoholes grasos producidos en una célula recombinante debido a la expresión de una variante de AAR es de al menos aproximadamente el 20 por ciento a al menos aproximadamente el 100 por ciento mayor que el de una silvestre célula. En realizaciones particulares, el título y/o rendimiento de un alcohol graso producido por una célula es al menos aproximadamente el 20 por ciento, al menos aproximadamente el 25 por ciento, al menos aproximadamente el 30 por ciento, al menos aproximadamente el 35 por ciento, al menos aproximadamente el 40 por ciento, al menos aproximadamente el 45 por ciento al menos aproximadamente el 50 por ciento, al menos aproximadamente el 55 por ciento, al menos aproximadamente el 60 por ciento, al menos aproximadamente el 65 por ciento, al menos aproximadamente el 70 por ciento, al menos aproximadamente el 75 por ciento, al menos aproximadamente el 80 por ciento, al menos aproximadamente el 85 por ciento, al menos aproximadamente el 90 por ciento, al menos aproximadamente el 95 por ciento, al menos aproximadamente el 97 por ciento, al menos aproximadamente el 98 por ciento, o al menos aproximadamente el 100 por ciento mayor que el de la célula silvestre correspondiente.

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

Se conocen en la técnica secuencias de control de la expresión e incluyen, por ejemplo, promotores, potenciadores, señales de poliadenilación, terminadores de la transcripción, sitios internos de entrada al ribosoma (IRES), y similares, que proporcionan la expresión de la secuencia de polinucleótido en una célula huésped. Las secuencias de control de la expresión interactúan específicamente con proteínas celulares implicadas en la transcripción (Maniatis *et al.* (1987) *Science* 236:1237-1245). Se describen secuencias de control de la expresión a modo de ejemplo, por ejemplo, en Goeddel, *Gene Expression Technology: Methods in Enzymology*, vol. 185, Academic Press, San Diego, Calif. (1990). En los métodos de la divulgación, una secuencia de control de la expresión está operativamente unida a una secuencia de polinucleótido. Por "operativamente unida" quiere decirse que una secuencia de polinucleótido y una secuencia de control de la expresión se conectan de tal manera que se permita la expresión génica cuando se unen las moléculas apropiadas (por ejemplo, proteínas activadoras de la transcripción) a la secuencia de control de la expresión. Están ubicados promotores operativamente unidos en el sentido de 5' de la secuencia de polinucleótido seleccionada en cuanto a la dirección de transcripción y traducción. Pueden estar ubicados potenciadores operativamente unidos en el sentido de 5', dentro de o en el sentido de 3' del polinucleótido seleccionado.

Tal como se usa en el presente documento, el término "vector" se refiere a una molécula de ácido nucleico que puede transportar otro ácido nucleico, es decir, una secuencia de polinucleótido, a la que se ha unido. Un tipo de vector útil es un episoma (es decir, un ácido nucleico que puede dar replicación extracromosómica). Los vectores útiles son aquellos que pueden dar replicación y/o expresión autónoma de ácidos nucleicos a los que se unen. Los vectores que pueden dirigir la expresión de genes a los que se unen operativamente se denominan en el presente documento "vectores de expresión". En general, los vectores de expresión de utilidad en técnicas de ADN recombinante están a menudo en forma de "plásmidos", que se refieren en general a bucles de ADN bicatenario circular que, en su forma de vector, no se unen al cromosoma. Otros vectores de expresión útiles se proporcionan en forma lineal. También están incluidas otras de tales formas de vectores de expresión que sirven para funciones equivalentes y que se han conocido en la técnica posteriormente al presente documento. En algunas realizaciones, un vector recombinante incluye además un promotor operativamente unido a la secuencia de polinucleótido. En algunas realizaciones, el promotor es un promotor regulado en el desarrollo, un promotor específico de órgano, un promotor específico de tejido, un promotor inducible, un promotor constitutivo o un promotor específico de célula. El vector recombinante comprende normalmente al menos una secuencia seleccionada de una secuencia de control de la expresión acoplada operativamente a la secuencia de polinucleótido; un marcador de selección acoplado operativamente a la secuencia de polinucleótido; una secuencia de marcador acoplada operativamente a la secuencia de polinucleótido; un resto de purificación acoplado operativamente a la secuencia de polinucleótido; una secuencia de secreción acoplada operativamente a la secuencia de polinucleótido; y una secuencia de selección como diana acoplada operativamente a la secuencia de polinucleótido. En determinadas realizaciones, la secuencia de nucleótidos se incorpora de manera estable en el ADN genómico de la célula huésped, y la expresión de la secuencia de nucleótidos está bajo el control de una región promotora regulada. Los vectores de expresión tal como se usan en el presente documento incluyen una secuencia de polinucleótido particular tal como se describe en el presente documento en una forma adecuada para la expresión de la secuencia de polinucleótido en una célula huésped. Los expertos en la técnica apreciarán que el diseño del vector de expresión puede depender de factores tales como la elección de la célula huésped que va transformarse, el nivel de expresión de polipéptido deseado, etc. Los vectores de expresión descritos en el presente documento pueden introducirse en células huésped para producir

polipéptidos, incluyendo polipéptidos de fusión, codificados por las secuencias de polinucleótido tal como se describe en el presente documento. La expresión de genes que codifican para polipéptidos en procariontes, por ejemplo, *E. coli*, se lleva a cabo con la mayor frecuencia con vectores que contienen promotores constitutivos o inducibles que dirigen la expresión de polipéptidos o bien de fusión o bien de no fusión. Los vectores de fusión añaden un número de aminoácidos a un polipéptido codificado en el mismo, habitualmente al extremo amino-terminal o carboxi-terminal del polipéptido recombinante. Tales vectores de fusión sirven normalmente para uno o más de los tres propósitos siguientes, incluyendo aumentar expresión del polipéptido recombinante; aumentar la solubilidad del polipéptido recombinante; y ayudar en la purificación del polipéptido recombinante actuando como ligando en purificación por afinidad. A menudo, en los vectores de expresión de fusión, se introduce un sitio de escisión proteolítica en la unión del resto de fusión y el polipéptido recombinante. Esto permite la separación del polipéptido recombinante del resto de fusión después de la purificación del polipéptido de fusión. En determinadas realizaciones, una secuencia de polinucleótido de la divulgación está operativamente unida a un promotor derivado del bacteriófago T5.

En determinadas realizaciones, la célula huésped es una célula de levadura, y el vector de expresión es un vector de expresión de levadura. Los ejemplos de vectores para la expresión en la levadura *S. cerevisiae* incluyen pYepSec1 (Baldari *et al.* (1987) EMBO J. 6:229-234); pMFa (Kurjan *et al.* (1982) Cell 30:933-943); pJRY88 (Schultz *et al.* (1987) Gene 54: 113-123); pYES2 (Invitrogen Corp., San Diego, CA) y picZ (Invitrogen Corp., San Diego, CA). En otras realizaciones, la célula huésped es una célula de insecto, y el vector de expresión es un vector de expresión de baculovirus. Los vectores de baculovirus disponibles para la expresión de proteínas en células de insecto en cultivo (por ejemplo, células Sf9) incluyen, por ejemplo, la serie de pAc (Smith *et al.* (1983) Mol. Cell Biol. 3:2156-2165) y la serie de pVL (Lucklow *et al.* (1989) Virology 170:31-39). En aún otra realización, las secuencias de polinucleótido descritas en el presente documento pueden expresarse en células de mamífero usando un vector de expresión de mamífero. Se conocen bien en la técnica otros sistemas de expresión adecuados para células tanto procariontes como eucariotas; véase, por ejemplo, Sambrook *et al.*, "Molecular Cloning: A Laboratory Manual", segunda edición, Cold Spring Harbor Laboratory, (1989).

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

El término "ACP" significa proteína transportadora de acilo. ACP es un producto intermedio transportador de acilo altamente conservado durante la biosíntesis de ácidos grasos, en el que la cadena en crecimiento se une durante la síntesis como éster de tiol en el tiol distal de un resto 4'-fosfopanteteína. La proteína existe en dos formas, es decir, apo-ACP (inactiva en la biosíntesis de ácidos grasos) y ACP o holo-ACP (activa en la biosíntesis de ácidos grasos). Los términos "ACP" y "holo-ACP" se usan de manera intercambiable en el presente documento y se refieren a la forma activa de la proteína. Una enzima denominada fosfopanteteiniltransferasa está implicada en la conversión de la apo-ACP inactiva en la holo-ACP activa. Más específicamente, ACP se expresa en la forma apo-ACP inactiva y debe unirse de manera postraduccional un resto 4'-fosfopanteteína a un residuo de serina conservado en la ACP mediante la acción de la proteína transportadora de holo-acilo sintasa (ACPS), una fosfopanteteiniltransferasa, para producir holo-ACP.

Tal como se usa en el presente documento, el término "acil-ACP" se refiere a un acil-tioéster formado entre el carbono de carbonilo de una cadena de alquilo y el grupo sulfhidrilo del resto fosfopanteteinilo de una proteína transportadora de acilo (ACP). En algunas realizaciones, una ACP es un producto intermedio en la síntesis acil-ACP totalmente saturadas. En otras realizaciones, una ACP es un producto intermedio en la síntesis de acil-ACP insaturadas. En algunas realizaciones, la cadena de carbonos tendrá aproximadamente 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25 ó 26 carbonos.

Tal como se usa en el presente documento, el término "derivado de ácido graso" significa un "ácido graso" o un "derivado de ácido graso", al que puede hacerse referencia como "ácido graso o derivado del mismo". El término "ácido graso" significa un ácido carboxílico que tiene la fórmula RCOOH. R representa un grupo alifático, preferiblemente un grupo alquilo. R puede incluir entre aproximadamente 4 y aproximadamente 22 átomos de carbono. Los ácidos grasos pueden ser saturados, monoinsaturados o poliinsaturados. Un "derivado de ácido graso" es un producto producido en parte a partir de la ruta de biosíntesis de ácidos grasos del organismo huésped de producción. Los "derivados de ácido graso" incluyen productos producidos en parte a partir de ACP, acil-ACP o derivados de acil-ACP. Los derivados de ácido graso a modo de ejemplo incluyen, por ejemplo, acil-CoA, ácidos grasos, aldehídos grasos, alcoholes de cadena corta y larga, alcoholes grasos, hidrocarburos, ésteres (por ejemplo, ceras, ésteres de ácidos grasos o ésteres grasos), olefinas terminales, olefinas internas y cetonas.

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 y derivados de los mismos. La ruta de biosíntesis de ácidos grasos puede incluir enzimas adicionales para producir derivados de ácidos grasos que tienen características deseadas.

Tal como se usa en el presente documento, "aldehído graso" significa un aldehído que tiene la fórmula RCHO caracterizada por un grupo carbonilo (C=O). En algunas realizaciones, el aldehído graso es cualquier aldehído

5 producido a partir de un alcohol graso. En determinadas realizaciones, el grupo R tiene al menos 5, al menos 6, al menos 7, al menos 8, al menos 9, al menos 10, al menos 11, al menos 12, al menos 13, al menos 14, al menos 15, al menos 16, al menos 17, al menos 18 o al menos 19, carbonos de longitud. Alternativamente, o además, el grupo R tiene 20 o menos, 19 o menos, 18 o menos, 17 o menos, 16 o menos, 15 o menos, 14 o menos, 13 o menos, 12 o menos, 11 o menos, 10 o menos, 9 o menos, 8 o menos, 7 o menos o 6 o menos carbonos de longitud. Por tanto, el grupo R puede tener un grupo R delimitado por dos cualesquiera de los extremos anteriores. Por ejemplo, el grupo R puede tener 6-16 carbonos de longitud, 10-14 carbonos de longitud o 12-18 carbonos de longitud. En algunas realizaciones, el aldehído graso es un aldehído graso C₆, C₇, C₈, C₉, C₁₀, C₁₁, C₁₂, C₁₃, C₁₄, C₁₅, C₁₆, C₁₇, C₁₈, C₁₉, C₂₀, C₂₁, C₂₂, C₂₃, C₂₄, C₂₅, o uno C₂₆. En determinadas realizaciones, el aldehído graso es un aldehído graso C₆, C₈, C₁₀, C₁₂, C₁₃, C₁₄, C₁₅, C₁₆, C₁₇ o C₁₈.

15 Tal como se usa en el presente documento, "alcohol graso" significa un alcohol que tiene la fórmula ROH. En algunas realizaciones, el grupo R tiene al menos 5, al menos 6, al menos 7, al menos 8, al menos 9, al menos 10, al menos 11, al menos 12, al menos 13, al menos 14, al menos 15, al menos 16, al menos 17, al menos 18 o al menos 19, carbonos de longitud. Alternativamente, o además, el grupo R tiene 20 o menos, 19 o menos, 18 o menos, 17 o menos, 16 o menos, 15 o menos, 14 o menos, 13 o menos, 12 o menos, 11 o menos, 10 o menos, 9 o menos, 8 o menos, 7 o menos o 6 o menos carbonos de longitud. Por tanto, el grupo R puede tener un grupo R delimitado por dos cualesquiera de los extremos anteriores. Por ejemplo, el grupo R puede tener 6-16 carbonos de longitud, 10-14 carbonos de longitud o 12-18 carbonos. En algunas realizaciones, el alcohol graso es un alcohol graso C₆, C₇, C₈, C₉, C₁₀, C₁₁, C₁₂, C₁₃, C₁₄, C₁₅, C₁₆, C₁₇, C₁₈, C₁₉, C₂₀, C₂₁, C₂₂, C₂₃, C₂₄, C₂₅, o uno C₂₆. En determinadas realizaciones, el alcohol graso es un alcohol graso C₆, C₈, C₁₀, C₁₂, C₁₃, C₁₄, C₁₅, C₁₆, C₁₇ o C₁₈.

20 Una "composición de alcohol graso" tal como se hace referencia en el presente documento se produce por una célula huésped recombinante y comprende normalmente una mezcla de alcoholes grasos. En algunos casos, la mezcla incluye más de un tipo de producto (por ejemplo, alcoholes grasos y ácidos grasos). En otros casos, las composiciones de derivado de ácido graso pueden comprender, por ejemplo, una mezcla de alcoholes grasos con diversas longitudes de cadena y características de saturación o ramificación. En todavía otros casos, la composición de alcohol graso comprende una mezcla de tanto más de un tipo de producto como de productos con diversas longitudes de cadena y características de saturación o ramificación.

25 Una célula huésped modificada por ingeniería para producir un aldehído graso convertirá normalmente parte del aldehído graso en un alcohol graso. En una realización a modo de ejemplo, se convierte acil-ACP en un aldehído graso mediante la acción de AAR. La conversión de aldehídos grasos en alcoholes grasos puede verse facilitada además, por ejemplo, mediante un polipéptido de biosíntesis de alcoholes grasos. En algunas realizaciones, un gen que codifica para un polipéptido de biosíntesis de alcoholes grasos se expresa o sobreexpresa en la célula huésped. En determinadas realizaciones, el polipéptido de biosíntesis de alcoholes grasos tiene actividad aldehído reductasa o alcohol deshidrogenasa.

30 Los ejemplos de polipéptidos de alcohol deshidrogenasa útiles según la divulgación incluyen, pero no se limitan a AlrA de *Acinetobacter* sp. M-1 (SEQ ID NO: 52) u homólogos de AlrA, tales como AlrAadp1 (SEQ ID NO:53) y alcohol deshidrogenasas de *E. coli* endógenas tales como YjgB, (AAC77226), DkgA (NP_417485), DkgB (NP_414743), YdjL (AAC74846), YdjJ (NP_416288), AdhP (NP_415995), YhdH (NP_417719), YahK (NP_414859), YphC (AAC75598), YqhD (446856) e YbbO [AAC73595,1]. Se describen ejemplos adicionales en las publicaciones de solicitud de patente internacional n.^{os} WO 2007/136762, WO2008/119082 y WO 2010/062480. En determinadas realizaciones, el polipéptido de biosíntesis de alcoholes grasos tiene actividad aldehído reductasa o alcohol deshidrogenasa (EC 1.1.1.1).

35 El grupo R de un ácido graso, aldehído graso o alcohol graso puede ser una cadena lineal o una cadena ramificada. Las cadenas ramificadas pueden tener más de un punto de ramificación y pueden incluir ramificaciones cíclicas. En algunas realizaciones, el ácido graso ramificado, aldehído graso ramificado o alcohol graso ramificado es un ácido graso ramificado, aldehído graso ramificado o alcohol graso ramificado C₆, C₇, C₈, C₉, C₁₀, C₁₁, C₁₂, C₁₃, C₁₄, C₁₅, C₁₆, C₁₇, C₁₈, C₁₉, C₂₀, C₂₁, C₂₂, C₂₃, C₂₄, C₂₅, o uno C₂₆. En realizaciones particulares, el ácido graso ramificado, aldehído graso ramificado o alcohol graso ramificado es un ácido graso ramificado, aldehído graso ramificado o alcohol graso ramificado C₆, C₈, C₁₀, C₁₂, C₁₃, C₁₄, C₁₅, C₁₆, C₁₇ o C₁₈. En determinadas realizaciones, el grupo hidroxilo del ácido graso ramificado, aldehído graso ramificado o alcohol graso ramificado está en la posición primaria (C₁).

40 En determinadas realizaciones, el ácido graso ramificado, aldehído graso ramificado o alcohol graso ramificado es un ácido graso iso, aldehído graso iso o alcohol graso iso, o un ácido graso anteiso, un aldehído graso anteiso o alcohol graso anteiso. En realizaciones a modo de ejemplo, el ácido graso ramificado, aldehído graso ramificado o alcohol graso ramificado se selecciona de ácido graso ramificado, aldehído graso ramificado o alcohol graso ramificado iso-C_{7:0}, iso-C_{8:0}, iso-C_{9:0}, iso-C_{10:0}, iso-C_{11:0}, iso-C_{12:0}, iso-C_{13:0}, iso-C_{14:0}, iso-C_{15:0}, iso-C_{16:0}, iso-C_{17:0}, iso-C_{18:0}, iso-C_{19:0}, anteiso-C_{7:0}, anteiso-C_{8:0}, anteiso-C_{9:0}, anteiso-C_{10:0}, anteiso-C_{11:0}, anteiso-C_{12:0}, anteiso-C_{13:0}, anteiso-C_{14:0}, anteiso-C_{15:0}, anteiso-C_{16:0}, anteiso-C_{17:0}, anteiso-C_{18:0} y anteiso-C_{19:0}.

45 El grupo R de un ácido graso ramificado o no ramificado, aldehído graso ramificado o no ramificado o alcohol graso ramificado o no ramificado puede ser saturado o insaturado. Si es insaturado, el grupo R puede tener uno o más de

un punto de insaturación. En algunas realizaciones, el ácido graso insaturado, aldehído graso insaturado o alcohol graso insaturado es un ácido graso monoinsaturado, aldehído graso monoinsaturado o alcohol graso monoinsaturado. En determinadas realizaciones, el ácido graso insaturado, aldehído graso insaturado o alcohol graso insaturado es un ácido graso insaturado, aldehído graso insaturado o alcohol graso insaturado C6:1, C7:1, C8:1, C9:1, C10:1, C11:1, C12:1, C13:1, C14:1, C15:1, C16:1, C17:1, C18:1, C19:1, C20:1, C21:1, C22:1, C23:1, C24:1, C25:1, o uno C26:1. En determinadas realizaciones preferidas, el ácido graso insaturado, aldehído graso insaturado o alcohol graso insaturado es C10:1, C12:1, C14:1, C16:1 o C18:1. En aún otras realizaciones, el ácido graso insaturado, aldehído graso insaturado o alcohol graso insaturado es insaturado en la posición omega-7. En determinadas realizaciones, el ácido graso insaturado, aldehído graso insaturado o alcohol graso insaturado comprende un doble enlace cis.

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

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

Tal como se usa en el presente documento, el término “cultivo” se refiere normalmente a un medio líquido que comprende células viables. En una realización, un cultivo comprende células que se reproducen en medios de cultivo predeterminados en condiciones controladas, por ejemplo, un cultivo de células huésped recombinantes hechas crecer en medios líquidos que comprenden una fuente de carbono y/o nitrógeno seleccionada.

Los términos “cultivar” o “cultivo” se refieren al crecimiento de una población de células (por ejemplo, células microbianas) en condiciones adecuadas en un medio líquido o sólido. En realizaciones particulares, cultivar se refiere a la bioconversión fermentativa de un sustrato en un producto final. Se conocen bien medios de cultivo y los componentes individuales de tales medios de cultivo están disponibles de fuentes comerciales, por ejemplo, en los medios DIFCO y medios BBL. En un ejemplo no limitativo, el medio de nutrientes acuoso es un “medio rico” que comprende fuentes complejas de nitrógeno, sales y carbono, tal como el medio YP, que comprende 10 g/l de peptona y 10 g/l de extracto de levadura de un medio de este tipo.

La célula huésped puede modificarse por ingeniería adicionalmente para asimilar el carbono eficazmente y usar materiales celulósicos como fuentes de carbono según los métodos descritos, por ejemplo, en las patentes estadounidenses 5.000.000; 5.028.539; 5.424.202; 5.482.846; 5.602.030 y WO2010127318. Además, la célula huésped puede modificarse por ingeniería para expresar una invertasa de modo que puede usarse sacarosa como fuente de carbono.

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

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

El término “secuencias reguladoras” tal como se usa en el presente documento se refiere normalmente a una secuencia de bases en ADN, operativamente unida a secuencias de ADN que codifican para una proteína que

controla en última instancia la expresión de la proteína. Los ejemplos de secuencias reguladoras incluyen, pero no se limitan a, secuencias promotoras de ARN, secuencias de unión a factor transcripción, secuencias de terminación de la transcripción, moduladores de transcripción (tales como elementos potenciadores), secuencias de nucleótidos que afectan a la estabilidad del ARN y secuencias reguladoras de la traducción (tales como, sitios de unión al ribosoma (por ejemplo, secuencias de Shine-Dalgarno en procariontes o secuencias de Kozak en eucariotas), codones de iniciación, codones de terminación).

Tal como se usa en el presente documento, la frase “la expresión de dicha secuencia de nucleótidos se modifica con relación a la secuencia de nucleótidos silvestre”, significa un aumento o una disminución en el nivel de expresión y/o actividad de una secuencia de nucleótidos endógena o la expresión y/o actividad de una secuencia de nucleótidos que codifica para polipéptido heterólogo o no nativo.

Tal como se usa en el presente documento, el término “expresar” con respecto a un polinucleótido es hacer que funcione. Un polinucleótido que codifica para un polipéptido (o proteína) se transcribirá y traducirá, cuando se exprese, para producir ese polipéptido (o proteína). Tal como se usa en el presente documento, el término “sobrexpresar” significa expresar (o hacer que se exprese) un polinucleótido o polipéptido en una célula a una mayor concentración de lo que se expresa normalmente en una célula silvestre correspondiente en las mismas condiciones.

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

Tal como se usa en el presente documento, el término “título” se refiere a la cantidad de aldehído graso o alcohol graso producido por volumen unitario de célula huésped cultivado. En cualquier aspecto de las composiciones y los métodos descritos en el presente documento, un alcohol graso se produce a un título de aproximadamente 25 mg/l, aproximadamente 50 mg/l, aproximadamente 75 mg/l, aproximadamente 100 mg/l, aproximadamente 125 mg/l, aproximadamente 150 mg/l, aproximadamente 175 mg/l, aproximadamente 200 mg/l, aproximadamente 225 mg/l, aproximadamente 250 mg/l, aproximadamente 275 mg/l, aproximadamente 300 mg/l, aproximadamente 325 mg/l, aproximadamente 350 mg/l, aproximadamente 375 mg/l, aproximadamente 400 mg/l, aproximadamente 425 mg/l, aproximadamente 450 mg/l, aproximadamente 475 mg/l, aproximadamente 500 mg/l, aproximadamente 525 mg/l, aproximadamente 550 mg/l, aproximadamente 575 mg/l, aproximadamente 600 mg/l, aproximadamente 625 mg/l, aproximadamente 650 mg/l, aproximadamente 675 mg/l, aproximadamente 700 mg/l, aproximadamente 725 mg/l, aproximadamente 750 mg/l, aproximadamente 775 mg/l, aproximadamente 800 mg/l, aproximadamente 825 mg/l, aproximadamente 850 mg/l, aproximadamente 875 mg/l, aproximadamente 900 mg/l, aproximadamente 925 mg/l, aproximadamente 950 mg/l, aproximadamente 975 mg/l, aproximadamente 1000 mg/l, aproximadamente 1050 mg/l, aproximadamente 1075 mg/l, aproximadamente 1100 mg/l, aproximadamente 1125 mg/l, aproximadamente 1150 mg/l, aproximadamente 1175 mg/l, aproximadamente 1200 mg/l, aproximadamente 1225 mg/l, aproximadamente 1250 mg/l, aproximadamente 1275 mg/l, aproximadamente 1300 mg/l, aproximadamente 1325 mg/l, aproximadamente 1350 mg/l, aproximadamente 1375 mg/l, aproximadamente 1400 mg/l, aproximadamente 1425 mg/l, aproximadamente 1450 mg/l, aproximadamente 1475 mg/l, aproximadamente 1500 mg/l, aproximadamente 1525 mg/l, aproximadamente 1550 mg/l, aproximadamente 1575 mg/l, aproximadamente 1600 mg/l, aproximadamente 1625 mg/l, aproximadamente 1650 mg/l, aproximadamente 1675 mg/l, aproximadamente 1700 mg/l, aproximadamente 1725 mg/l, aproximadamente 1750 mg/l, aproximadamente 1775 mg/l, aproximadamente 1800 mg/l, aproximadamente 1825 mg/l, aproximadamente 1850 mg/l, aproximadamente 1875 mg/l, aproximadamente 1900 mg/l, aproximadamente 1925 mg/l, aproximadamente 1950 mg/l, aproximadamente 1975 mg/l, aproximadamente 2000 mg/l (2 g/l), 3 g/l, 5 g/l, 10 g/l, 20 g/l, 30 g/l, 40 g/l, 50 g/l, 60 g/l, 70 g/l, 80 g/l, 90 g/l, 100 g/l o un intervalo delimitado por dos cualesquiera de los valores anteriores. En otras realizaciones, se produce un aldehído graso o alcohol graso a un título de más de 100 g/l, más de 200 g/l, más de 300 g/l, o mayor, tal como 500 g/l, 700 g/l, 1000 g/l, 1200 g/l, 1500 g/l o 2000 g/l. El título preferido de aldehído graso o alcohol graso producido por una célula huésped recombinante según los métodos de la divulgación es de desde 5 g/l hasta 200 g/l, de 10 g/l a 150 g/l, de 20 g/l a 120 g/l y de 30 g/l a 100 g/l.

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

ejemplo, el rendimiento del alcohol graso o aldehído graso producido por la célula huésped recombinante según los métodos de la divulgación puede ser del 5% al 15%, del 10% al 25%, del 10% al 22%, del 15% al 27%, del 18% al 22%, del 20% al 28% o del 20% al 30%. El rendimiento preferido de alcohol graso producido por la célula huésped recombinante según los métodos de la divulgación es de desde el 10% hasta el 30%.

5 Tal como se usa en el presente documento, el término “productividad” se refiere a la cantidad de aldehído graso o alcohol graso producido por volumen unitario de célula huésped cultivo por tiempo unitario. En cualquier aspecto de las composiciones y los métodos descritos en el presente documento, la productividad de aldehído graso o alcohol graso producido por una célula huésped recombinante es de al menos 100 mg/l/hora, al menos 200 mg/l/hora, al menos 300 mg/l/hora, al menos 400 mg/l/hora, al menos 500 mg/l/hora, al menos 600 mg/l/hora, al menos 700 mg/l/hora, al menos 800 mg/l/hora, al menos 900 mg/l/hora, al menos 1000 mg/l/hora, al menos 1100 mg/l/hora, al menos 1200 mg/l/hora, al menos 1300 mg/l/hora, al menos 1400 mg/l/hora, al menos 1500 mg/l/hora, al menos 1600 mg/l/hora, al menos 1700 mg/l/hora, al menos 1800 mg/l/hora, al menos 1900 mg/l/hora, al menos 2000 mg/l/hora, al menos 2100 mg/l/hora, al menos 2200 mg/l/hora, al menos 2300 mg/l/hora, al menos 2400 mg/l/hora o al menos 2500 mg/l/hora. Alternativamente, o además, la productividad es de 2500 mg/l/hora o menos, 2000 mg/l/DO₆₀₀ o menos, 1500 mg/l/DO₆₀₀ o menos, 120 mg/l/hora o menos, 1000 mg/l/hora o menos, 800 mg/l/hora o menos o 600 mg/l/hora o menos. Por tanto, la productividad puede estar delimitada por dos cualesquiera de los extremos anteriores. Por ejemplo, la productividad puede ser de 3 a 30 mg/l/hora, de 6 a 20 mg/l/hora o de 15 a 30 mg/l/hora. La productividad preferida de un aldehído graso o alcohol graso producido por una célula huésped recombinante según los métodos de la divulgación se selecciona de 500 mg/l/hora a 2500 mg/l/hora, o desde 700 mg/l/hora hasta 2000 mg/l/hora.

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

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

Tal como se usa en el presente documento, el término “fuente de carbono” se refiere a un sustrato o compuesto adecuado para usarse como fuente de carbono para el crecimiento de células procariotas o eucariotas simples. Las fuentes de carbono pueden estar en diversas formas, incluyendo, pero sin limitarse a polímeros, hidratos de carbono, ácidos, alcoholes, aldehídos, cetonas, aminoácidos, péptidos y gases (por ejemplo, CO y CO₂). Las fuentes de carbono a modo de ejemplo incluyen, pero no se limitan a, monosacáridos, tales como glucosa, fructosa, manosa, galactosa, xilosa y arabinosa; oligosacáridos, tales como fructo-oligosacárido y galacto-oligosacárido; polisacáridos tales como almidón, celulosa, pectina y xilano; disacáridos, tales como sacarosa, maltosa, celobiosa y turanosa; material celulósico y variantes tales como hemicelulosas, metilcelulosa y carboximetilcelulosas sódicas; ácidos grasos saturados o insaturados, succinato, lactato y acetato; alcoholes, tales como etanol, metanol y glicerol, o mezclas de los mismos. La fuente de carbono también puede ser un producto de fotosíntesis, tales como glucosa. En determinadas realizaciones, la fuente de carbono es una mezcla gaseosa que contiene CO procedente de gas de chimenea. En otra realización, la fuente de carbono es una mezcla gaseosa que contiene CO procedente del reformado de un material que contiene carbono, tal como biomasa, carbón o gas natural. En otras realizaciones, la fuente de carbono es gas de síntesis, metano o gas natural. En determinadas realizaciones preferidas, la fuente de carbono es biomasa. En otras realizaciones preferidas, la fuente de carbono es glucosa. En otras realizaciones preferidas, la fuente de carbono es sacarosa. En otras realizaciones, la fuente de carbono es glicerol. En otras realizaciones preferidas, la fuente de carbono es guarapo, jarabe de azúcar de caña o jarabe de maíz. En otras realizaciones preferidas, la fuente de carbono se deriva de materias primas renovables, tales como CO₂, CO, glucosa, sacarosa, xilosa, arabinosa, glicerol, manosa, o mezclas de los mismos. En otras realizaciones, la fuente de carbono se deriva de materias primas renovables incluyendo almidones, biomasa celulósica, melaza, y otras fuentes de hidratos de carbono incluyendo mezclas de hidratos de carbono derivadas de la hidrólisis de biomasa celulósica, o los materiales de desecho derivados del procesamiento de aceites naturales o vegetales.

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

Tal como se usa en el presente documento, el término “aislados”, con respecto a productos (tal como ácidos grasos

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

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

Aumento de la producción de alcoholes grasos

La divulgación proporciona la producción de una composición de alcohol graso que está potenciada como resultado de la expresión modificada de un gen de acil-ACP reductasa (AAR) en una célula huésped. AAR está implicada en una ruta de biosíntesis para la producción de aldehídos grasos y alcoholes grasos. Se usa AAR variante sola o en combinación con la expresión modificada de otro gen implicado en la ruta de biosíntesis que convierte un aldehído graso en un alcohol graso. En el presente documento, la divulgación proporciona células huésped recombinantes, que se han modificado por ingeniería para expresar una AAR variante para proporcionar biosíntesis potenciada de alcoholes grasos con relación a células huésped no modificadas por ingeniería o nativas o silvestres que expresan AAR silvestre u otros polipéptidos de biosíntesis de alcoholes grasos con la misma función que AAR. La divulgación identifica polinucleótidos y polipéptidos relacionados con AAR útiles en las células huésped recombinantes. Sin embargo, se reconocerá que no es necesaria una identidad de secuencia absoluta con polinucleótidos relacionados con AAR. Por ejemplo, pueden realizarse cambios en una secuencia de polinucleótido particular y examinarse el polipéptido codificado para detectar actividad. Tales cambios incluyen normalmente mutaciones conservativas y mutaciones silenciosas tales como, por ejemplo, a través de la optimización de codones. Pueden examinarse polinucleótidos modificados o mutados (es decir, mutantes) y polipéptidos variantes codificados para detectar una función deseada, tal como, una función mejorada en comparación con el polipéptido parental, incluyendo pero sin limitarse a actividad catalítica aumentada, estabilidad aumentada o inhibición disminuida (por ejemplo, retroinhibición disminuida), usando métodos conocidos en la técnica. La divulgación identifica actividades enzimáticas implicadas en diversas etapas (es decir, reacciones) de las rutas de biosíntesis de ácidos grasos descritas en el presente documento según el número de clasificación de enzimas (EC), y proporciona polipéptidos (enzimas) a modo de ejemplo clasificados mediante tales números EC, y polinucleótidos a modo de ejemplo que codifican para tales polipéptidos. Tales polipéptidos y polinucleótidos a modo de ejemplo, que se identifican en el presente documento mediante los números de registro y/o números de identificación de secuencia (SEQ ID NO), son útiles para modificar por ingeniería rutas de ácidos grasos en células huésped parentales para obtener las células huésped recombinantes descritas en el presente documento. Sin embargo, ha de entenderse que los polipéptidos y polinucleótidos descritos en el presente documento son a modo de ejemplo y, por tanto, no limitativos. Las secuencias de homólogos de polipéptidos a modo de ejemplo descritos en el presente documento están disponibles para los expertos en la técnica usando bases de datos tales como, por ejemplo, las bases de datos Entrez proporcionadas por el Centro Nacional para la Información Biotecnológica (NCBI), las bases de datos ExPasy proporcionadas por el Instituto Suizo de Bioinformática, la base de datos BRENDA proporcionada por la Universidad Técnica de Brunswick, y la base de datos KEGG proporcionada por el Centro de Bioinformática de la Universidad de Kioto y la Universidad de Tokio, que están todas disponibles en Internet. Una variedad de diferentes células huésped pueden modificarse para expresar enzimas de biosíntesis de alcoholes grasos AAR variantes tales como las descritas en el presente documento, lo que da como resultado células huésped recombinantes adecuadas para la producción potenciada de composiciones de alcohol graso. Se entiende que una variedad de células pueden proporcionar fuentes de material genético, incluyendo secuencias de polinucleótido que codifican para polipéptidos adecuados para su uso en una célula huésped recombinante tal como se describe en el presente documento.

Polipéptidos de acil-ACP reductasa (AAR) y variantes de los mismos

En un aspecto, la divulgación se refiere a la producción mejorada de derivados de ácido graso tales como aldehídos grasos y/o alcoholes grasos modificando por ingeniería una célula huésped para expresar una proteína acil-ACP reductasa (AAR) nativa o no nativa. La proteína AAR cataliza la reducción de una acil-ACP para dar un aldehído

graso y también puede catalizar la conversión de un aldehído graso en un alcohol graso (véase la publicación de patente estadounidense n.º 20120282663). El polipéptido de AAR o la secuencia de polinucleótido que codifica para el polipéptido de AAR puede ser nativo (por ejemplo, endógeno) o no nativo (por ejemplo, exógeno, heterólogo, etc.), es decir, puede diferir de la secuencia silvestre y la expresión del mismo presente de manera natural en la célula huésped silvestre correspondiente. Los ejemplos incluyen una modificación en la secuencia del polinucleótido, polipéptido o proteína AAR que da como resultado una AAR variante (por ejemplo, mutante) y/o en el nivel de expresión de la misma. La divulgación incluye polipéptidos, homólogos y variantes de AAR.

En un aspecto, un polipéptido de AAR para su uso en la puesta en práctica de la divulgación tiene una identidad de secuencia de al menos el 90% con la secuencia de polipéptido de AAR silvestre de SEQ ID NO: 28, SEQ ID NO: 30, SEQ ID NO: 32, SEQ ID NO: 34, SEQ ID NO: 36, SEQ ID NO: 38, SEQ ID NO: 40, SEQ ID NO: 42 o SEQ ID NO: 44. La AAR puede derivarse de una especie de *Synechococcus* o una especie de *Prochlorococcus*. En otros aspectos, un polipéptido de AAR para su uso en la puesta en práctica de la divulgación tiene una identidad de secuencia de al menos el 75% (por ejemplo, al menos el 76%, el 77%, el 78%, el 79%, el 80%, el 81%, el 82%, el 83%, el 84%, el 85%, el 86%, el 87%, el 88%, el 89%, el 90%, el 91%, el 92%, el 93%, el 94%, el 95%, el 96%, el 97%, el 98% o al menos el 99%) con la secuencia de polipéptido de AAR silvestre de SEQ ID NO: 28, SEQ ID NO: 30, SEQ ID NO: 32, SEQ ID NO: 34, SEQ ID NO: 36, SEQ ID NO: 38, SEQ ID NO: 40, SEQ ID NO: 42 o SEQ ID NO: 44, y también puede incluir una o más sustituciones que dan como resultado características y/o propiedades útiles tal como se describe en el presente documento. En un aspecto, el polipéptido de AAR para su uso en la puesta en práctica de la presente divulgación tiene una identidad de secuencia de al menos el 75% (por ejemplo, al menos el 76%, el 77%, el 78%, el 79%, el 80%, el 81%, el 82%, el 83%, el 84%, el 85%, el 86%, el 87%, el 88%, el 89%, el 90%, el 91%, el 92%, el 93%, el 94%, el 95%, el 96%, el 97%, el 98% o al menos el 99%) con la secuencia de polipéptido de AAR silvestre de SEQ ID NO: 28 o SEQ ID NO: 34. En otros aspectos, un polipéptido de AAR para su uso en la puesta en práctica de la divulgación tiene una identidad de secuencia del 100% con SEQ ID NO: 28, SEQ ID NO: 30, SEQ ID NO: 32, SEQ ID NO: 34, SEQ ID NO: 36, SEQ ID NO: 38, SEQ ID NO: 40, SEQ ID NO: 42 o SEQ ID NO: 44. La secuencia de polipéptido de AAR mejorado o variante puede derivarse de una especie distinta de una especie de *Synechococcus* o una especie de *Prochlorococcus*.

En un aspecto relacionado, la divulgación incluye polipéptidos de AAR que tienen una secuencia de aminoácidos codificada por una secuencia de ácido nucleico que tiene una identidad de secuencia de al menos el 75% (por ejemplo, al menos el 76%, el 77%, el 78%, el 79%, el 80%, el 81%, el 82%, el 83%, el 84%, el 85%, el 86%, el 87%, el 88%, el 89%, el 90%, el 91%, el 92%, el 93%, el 94%, el 95%, el 96%, el 97%, el 98% o/y al menos el 99%) con SEQ ID NO: 27, SEQ ID NO: 29, SEQ ID NO: 31, SEQ ID NO: 33, SEQ ID NO: 35, SEQ ID NO: 37, SEQ ID NO: 39, SEQ ID NO: 41 o SEQ ID NO: 43. En algunas realizaciones, la secuencia de ácido nucleico codifica para una variante de AAR con una o más sustituciones que dan como resultado características y/o propiedades mejoradas tal como se describe en el presente documento. En aún otra realización relacionada, un polipéptido de AAR para su uso en la puesta en práctica de la divulgación está codificada por una secuencia de nucleótidos que tienen una identidad de secuencia del 100% con SEQ ID NO: 27, SEQ ID NO: 29, SEQ ID NO: 31, SEQ ID NO: 33, SEQ ID NO: 35, SEQ ID NO: 37, SEQ ID NO: 39, SEQ ID NO: 41 o SEQ ID NO: 43. En otro aspecto, la divulgación se refiere a polipéptidos de AAR que comprenden una secuencia de aminoácidos codificada por un ácido nucleico que se hibrida en condiciones rigurosas por sustancialmente toda la longitud de un ácido nucleico correspondiente a SEQ ID NO: 27, SEQ ID NO: 29, SEQ ID NO: 31, SEQ ID NO: 33, SEQ ID NO: 35, SEQ ID NO: 37, SEQ ID NO: 39, SEQ ID NO: 41 o SEQ ID NO: 43.

En una realización preferida, la divulgación proporciona un polipéptido de AAR para su uso en la puesta en práctica de la divulgación que tiene una identidad de secuencia de al menos el 90% con la secuencia de polipéptido de AAR variante de una cualquiera de SEQ ID NO: 57 a SEQ ID NO: 78. En algunas realizaciones, la AAR variante se deriva de una especie de *Synechococcus* o una especie *Prochlorococcus*. En otras realizaciones, un polipéptido de AAR para su uso en la puesta en práctica de la divulgación tiene una identidad de secuencia de al menos el 75% (por ejemplo, al menos el 76%, el 77%, el 78%, el 79%, el 80%, el 81%, el 82%, el 83%, el 84%, el 85%, el 86%, el 87%, el 88%, el 89%, el 90%, el 91%, el 92%, el 93%, el 94%, el 95%, el 96%, el 97%, el 98% o al menos el 99%) con la secuencia de polipéptido de AAR variante de una cualquiera de SEQ ID NO: 57 a SEQ ID NO: 78. El polipéptido de AAR variante puede incluir una o más sustituciones que dan como resultado características y/o propiedades útiles tal como se describe en el presente documento. En otra realización preferida, el polipéptido de AAR para su uso en la puesta en práctica de la divulgación tiene una identidad de secuencia de al menos el 75% (por ejemplo, al menos el 76%, el 77%, el 78%, el 79%, el 80%, el 81%, el 82%, el 83%, el 84%, el 85%, el 86%, el 87%, el 88%, el 89%, el 90%, el 91%, el 92%, el 93%, el 94%, el 95%, el 96%, el 97%, el 98% o al menos el 99%) con la secuencia de polipéptido de AAR variante de SEQ ID NO: 57, SEQ ID NO: 58, SEQ ID NO: 59 o SEQ ID NO: 65. En otras realizaciones, un polipéptido de AAR para su uso en poner en práctica la divulgación tiene una identidad de secuencia del 100% con una cualquiera de SEQ ID NO: 57 a SEQ ID NO: 78. En todavía otras realizaciones, la secuencia de polipéptido de AAR mejorado o variante se deriva de una especie distinta de una especie de *Synechococcus* o una especie de *Prochlorococcus*.

Los inventores construyeron una biblioteca de propensión a error de la acil-ACP reductasa de *Synechococcus elongatus* PCC7942 (AAR_7942) para examinar variantes que tienen mejoras con respecto a la AAR_7942 silvestre (véase el ejemplo 3, más adelante). Las mejoras se clasifican como o bien que mejoran el título de alcohol graso global o bien que aumentan la fracción de alcoholes grasos C₁₀, C₁₂, C₁₄ o C₁₆ sin afectar significativamente al título.

La biblioteca de propensión a error identificó diversas posiciones de aminoácido incluyendo la posición de aminoácido 18. Se prepararon bibliotecas de saturación y de combinación para someter a prueba adicionalmente estas posiciones. SEQ ID NO: 57 codifica para la secuencia de aminoácidos para una variante de AAR (mutante) que tiene una mutación en el aminoácido 18. La mutación S18W en la que se reemplaza serina por triptófano da como resultado un aumento significativo de la producción de alcoholes grasos cuando se expresan en células (véanse los ejemplos 3 y 4, más adelante). Más específicamente, la mutación S18W conduce a un aumento del 227 por ciento en la producción de alcohol graso total (FALC) y un aumento del 324 por ciento en alcoholes grasos C₁₄ en comparación con AAR silvestre usado como control (véanse las tablas 4A y 4B, más adelante).

Los inventores construyeron bibliotecas de saturación basándose en la mutación S18W para identificar variantes (mutantes) que aumentaron adicionalmente el título de FALC global o la fracción de alcoholes grasos C₁₂ (véanse el ejemplo 4 y la tabla 5, más adelante). Las bibliotecas de combinación que usaron la mutación S18W (SEQ ID NO: 57) como molde produjeron 7 mutantes de combinación que mostraron aumentos significativos adicionales en el título total de FALC y/o la producción de alcoholes grasos C₁₂ (véase la tabla 6B, más adelante). Los 7 mutantes de combinación incluyen AAR con la mutación S18W (SEQ ID NO: 57); AAR con las mutaciones M21L, C63G, S113K, T154A, A281L (SEQ ID NO: 58); AAR con las mutaciones L8A, M21L, C63G, A77A (mutación de codón silenciosa de GCC a GCA), S113K, T154A, A281L (SEQ ID NO: 59); AAR con las mutaciones D16L, M21L, C63G, S113K, T154A, A281L (SEQ ID NO: 60); AAR con las mutaciones L8A, D24V, C63G, S113K, Q155L, A281L (SEQ ID NO: 61); AAR con las mutaciones D24P, L31M, C63G, S113K, T154A, A281L (SEQ ID NO: 62); AAR con las mutaciones L8A, D16L, D24V, C63G, S113K, T154A, A281L (SEQ ID NO: 63); y AAR con las mutaciones D24E, C63G, S113K, T154A, A281L (SEQ ID NO: 64). Particularmente, SEQ ID NO: 58 mostró la mayor fracción de alcoholes grasos C₁₂ mientras que SEQ ID NO: 59 mostró el mayor título de los mutantes de combinación (véase la tabla 6B, más adelante).

Los inventores también construyeron una biblioteca de saturación completa de la acil-ACP reductasa de *Prochlorococcus marinus* MED4_AAR para examinar variantes que muestran mejoras con respecto MED4_AAR silvestre (véase el ejemplo 7, más adelante). Las variantes de AAR se seleccionaron basándose en la producción de más alcoholes grasos que la enzima AAR silvestre o la capacidad para producir alcoholes grasos con un perfil alterado de longitud de cadena, por ejemplo, una fracción aumentada de alcoholes grasos C₁₂, C₁₄ o C₁₆. La tabla 8 (véase el ejemplo 7, más adelante) muestra datos representativos de 16 variantes de AAR que produjeron los mayores títulos de FALC que oscilaron entre 1,4 veces y 2,2 veces con respecto a MED4_AAR silvestre. Los inventores también examinaron variantes de AAR que tienen un perfil alterado de longitud de cadena, en el que es de interés un aumento de la proporción de especies FALC con una longitud de cadena más corta que C₁₆. Dos clones variantes que conducen a un aumento de 2-3 veces en la cantidad de FALC se muestran en la figura 10. La expresión de una de estas variantes de AAR, es decir, el mutante D61E (SEQ ID NO: 65) en una célula huésped recombinante, desvió la distribución de longitud de cadena de especies de alcohol graso hacia cadenas de carbonos más cortas. Todas las variantes condujeron a una mayor cantidad de FALC ya que tienen un título aumentado, pero sólo SEQ ID NO: 65 tenía una fracción aumentada de C₁₄ (y un mayor título). La tabla 9 (véase el ejemplo 7, más adelante) ilustra la distribución de longitud de cadena de FALC producida por células huésped recombinantes que expresan la variante de D61E de MED4_AAR en comparación con MED4_AAR silvestre (WT) y la variante de V346P de MED4_AAR que no produjo productos con longitudes de cadena alteradas.

40 Proteína transportadora de acilo (ACP)

Existen informes contradictorios en la bibliografía en cuanto a los factores que pueden limitar la biosíntesis de ácidos grasos en células huésped, tales como *Escherichia coli* (*E. coli*). Aunque las proteínas transportadoras de acilo (ACP) se conservan en cierta medida en todos los organismos, su secuencia primaria puede diferir. Se ha sugerido que cuando se expresan en *E. coli* enzimas de rutas terminales de fuentes distintas de *E. coli* para convertir las acil graso-ACP en productos, pueden existir limitaciones tales como en el reconocimiento, la afinidad y/o el recambio de la enzima de ruta recombinante hacia las acil graso-ACP (véanse Suh *et al.* (1999) *The Plant Journal* 17(6):679-688; Salas *et al.* (2002) *Archives of Biochemistry and Biophysics* 403:25-34). Una sugerencia es que una limitación de los precursores principales para la biosíntesis de ácidos grasos, por ejemplo, acetil-CoA y malonil-CoA puede dar como resultado una disminución de la síntesis de derivados de ácido graso. Un enfoque para aumentar el flujo a través de la biosíntesis de ácidos grasos es manipular diversas enzimas en la ruta (véanse las figuras 1-3). El suministro de acil-ACP a partir de acetil-CoA mediante el complejo de acetil-CoA carboxilasa (acc) y la ruta de biosíntesis de ácidos grasos (fab) pueden tener un impacto sobre la tasa de producción de derivados de ácido graso (véase la figura 2). Tal como se detalla en los ejemplos (más adelante), se evaluó el efecto de la sobreexpresión de ACP sobre la producción de alcoholes grasos con fines ilustrativos.

Una célula huésped que se modifica por ingeniería para expresar una ACP puede presentar un aumento en el título de una composición de aldehído graso y/o alcohol graso o una composición específica de aldehído graso y/o alcohol graso en el que el aumento es de al menos el 3%, al menos el 4%, al menos el 5%, al menos el 6%, al menos el 7%, al menos el 8%, al menos el 9%, al menos el 10%, al menos el 11%, al menos el 12%, al menos el 13%, al menos el 14%, al menos el 15%, al menos el 16%, al menos el 17%, al menos el 18%, al menos el 19%, al menos el 20 %, al menos el 21%, al menos el 22%, al menos el 23%, al menos el 24%, al menos el 25%, al menos el 26%, al menos el 27%, al menos el 28%, al menos el 29%, o al menos el 30% mayor que el título del composición de aldehído graso y/o alcohol graso producida por una célula huésped correspondiente que no expresa ACP cuando se cultiva en las

mismas condiciones. En un aspecto, la divulgación se refiere a la producción mejorada de una composición de aldehído graso y/o alcohol graso modificando por ingeniería una célula huésped para expresar una proteína ACP nativa (por ejemplo, endógena) o no nativa (por ejemplo, exógena, heteróloga). El polipéptido de ACP o la secuencia de polinucleótido que codifica para el polipéptido de ACP puede ser no nativo, es decir, puede diferir de la secuencia silvestre presente de manera natural en la célula huésped silvestre correspondiente. Los ejemplos incluyen, una modificación en el nivel de expresión o en la secuencia de un nucleótido, polipéptido o una proteína. La divulgación incluye polipéptidos de ACP y homólogos de los mismos.

En un aspecto, un polipéptido de ACP para su uso en la puesta en práctica de la divulgación tiene una identidad de secuencia de al menos el 70% con SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8 o SEQ ID NO: 10. La ACP puede derivarse de *Marinobacter hydrocarbonoclasticus* o *E. coli*. En otros aspectos, un polipéptido de ACP para su uso en la puesta en práctica de la divulgación tiene una identidad de secuencia de al menos el 75% (por ejemplo, al menos el 76%, el 77%, el 78%, el 79%, el 80%, el 81%, el 82%, el 83%, el 84%, el 85%, el 86%, el 87%, el 88%, el 89%, el 90%, el 91%, el 92%, el 93%, el 94%, el 95%, el 96%, el 97%, el 98% o al menos el 99%) con la secuencia de polipéptido de ACP silvestre de SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8 o SEQ ID NO: 10, y también puede incluir una o más sustituciones que dan como resultado características y/o propiedades útiles tal como se describe en el presente documento. En un aspecto, un polipéptido de ACP para su uso en la puesta en práctica de la divulgación tiene una identidad de secuencia del 100% con SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8 o SEQ ID NO: 10. La secuencia de polipéptido de ACP mejorada o variante puede derivarse de una especie distinta de *M. hydrocarbonoclasticus* o *E. coli*. En un aspecto relacionado, un polipéptido de ACP para su uso en la puesta en práctica de la divulgación está codificado por una secuencia de nucleótidos que tiene una identidad de secuencia del 100% con SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7 o SEQ ID NO: 9. En otro aspecto relacionado, la divulgación se refiere a polipéptidos de ACP que comprenden una secuencia de aminoácidos codificada por una secuencia de ácido nucleico que tiene una identidad de secuencia de al menos el 75% (por ejemplo, al menos el 76%, el 77%, el 78%, el 79%, el 80%, el 81%, el 82%, el 83%, el 84%, el 85%, el 86%, el 87%, el 88%, el 89%, el 90%, el 91%, el 92%, el 93%, el 94%, el 95%, el 96%, el 97%, el 98% o/y al menos el 99%) con SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7 o SEQ ID NO: 9. La secuencia de ácido nucleico puede codificar para una variante de ACP con una o más sustituciones que dan como resultado características y/o propiedades mejoradas tal como se describe en el presente documento. En otros aspectos de la divulgación, la secuencia de ácido nucleico de ACP mejorada o variante se deriva de una especie distinta de una especie de *Marinobacter* o *E. coli*. En otro aspecto, la divulgación se refiere a polipéptidos de ACP que tienen una secuencia de aminoácidos codificada por un ácido nucleico que se hibrida en condiciones rigurosas por sustancialmente toda la longitud de un ácido nucleico correspondiente a SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7 o SEQ ID NO: 9.

Variaciones y mutaciones

En algunas realizaciones, el polipéptido de AAR o ACP es un mutante o una variante de cualquiera de los polipéptidos descritos en el presente documento. Un polipéptido variante o mutante tal como se usa en el presente documento se refiere a un polipéptido que tiene una secuencia de aminoácidos que difiere de un polipéptido silvestre en al menos un aminoácido. Por ejemplo, el mutante puede tener una o más de las siguientes sustituciones de aminoácidos conservativas, incluyendo pero sin limitarse a, reemplazo de un aminoácido alifático, tal como alanina, valina, leucina e isoleucina, por otro aminoácido alifático; reemplazo de una serina por una treonina; reemplazo de una treonina por una serina; reemplazo de un residuo ácido, tal como ácido aspártico y ácido glutámico, por otro residuo ácido; reemplazo de un residuo que porta un grupo amida, tal como asparagina y glutamina, por otro residuo que porta un grupo amida; intercambio de un residuo básico, tal como lisina y arginina, por otro residuo básico; y reemplazo de un residuo aromático, tal como fenilalanina y tirosina, por otro residuo aromático. En algunas realizaciones, el polipéptido variante o mutante tiene aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 30, 40, 50, 60, 70, 80, 90, 100 o más sustituciones, adiciones, inserciones o deleciones de aminoácidos. Algunos fragmentos preferidos de un polipéptido que funcionan como variante o mutante conservan parte o la totalidad de la función biológica (por ejemplo, actividad enzimática) del polipéptido silvestre correspondiente. En algunas realizaciones, el fragmento conserva al menos el 75%, al menos el 80%, al menos el 90%, al menos el 95% o al menos el 98% o más de la función biológica del polipéptido silvestre correspondiente. En otras realizaciones, el fragmento o mutante conserva aproximadamente el 100% de la función biológica del polipéptido silvestre correspondiente. Puede encontrarse orientación sobre la determinación de qué residuos de aminoácido pueden sustituirse, insertarse o deleccionarse sin afectar la biológica actividad usando programas informáticos bien conocidos en la técnica, por ejemplo, el software LASERGENE (ADNSTAR, Inc., Madison, WI). En aún otras realizaciones, un fragmento presenta un aumento de la función biológica en comparación con un polipéptido silvestre correspondiente. Por ejemplo, un fragmento puede presentar una mejora de al menos un 10%, al menos un 25%, al menos un 50%, al menos un 75% o al menos un 90% de la actividad enzimática en comparación con el polipéptido silvestre correspondiente. En otras realizaciones, el fragmento presenta una mejora de al menos el 100%, al menos el 200%, o al menos el 500% de la actividad enzimática en comparación con el polipéptido silvestre correspondiente.

Se entiende que los polipéptidos descritos en el presente documento pueden tener sustituciones de aminoácidos conservativas o no esenciales adicionales que no tienen un efecto sustancial sobre la función del polipéptido. Puede determinarse si se tolerará o no una sustitución particular (es decir, no afectará adversamente a la función biológica deseada, tal como actividad acil-ACP reductasa), tal como se conoce en la técnica (véase Bowie *et al.* (1990)

Science, 247:1306-1310). Una sustitución de aminoácidos conservativa es una en la que el residuo de aminoácido se reemplaza por un residuo de aminoácido que tiene una cadena lateral similar. Se han definido en la técnica familias de residuos de aminoácido que tienen cadenas laterales similares. Estas familias incluyen aminoácidos con cadenas laterales básicas (por ejemplo, lisina, arginina, histidina), cadenas laterales ácidas (por ejemplo, ácido aspártico, ácido glutámico), cadenas laterales polares no cargadas (por ejemplo, glicina, asparagina, glutamina, serina, treonina, tirosina, cisteína), cadenas laterales apolares (por ejemplo, alanina, valina, leucina, isoleucina, prolina, fenilalanina, metionina, triptófano), cadenas laterales con ramificación beta (por ejemplo, treonina, valina, isoleucina) y cadenas laterales aromáticas (por ejemplo, tirosina, fenilalanina, triptófano, histidina).

Las variantes pueden producirse de manera natural o crearse *in vitro*. En particular, tales variantes pueden crearse usando técnicas de ingeniería genética, tales como mutagénesis dirigida al sitio, mutagénesis química al azar, procedimientos de delección con exonucleasa III o técnicas de clonación convencionales. Alternativamente, tales variantes, mutantes, fragmentos, análogos o derivados pueden crearse usando procedimientos de modificación o síntesis química. Se conocen bien en la técnica métodos de producción de variantes. Por ejemplo, pueden prepararse variantes usando mutagénesis al azar y dirigida al sitio. La mutagénesis al azar y dirigida al sitio se conocen generalmente en la técnica (véase, por ejemplo, Arnold (1993) *Curr. Opin. Biotech.* 4:450-455). Puede lograrse la mutagénesis al azar usando PCR propensa a error (véanse, por ejemplo, Leung *et al.* (1989) *Technique* 1:11-15; y Caldwell *et al.* (1992) *PCR Methods Applic.* 2: 28-33). En la PCR propensa a error, la propia PCR se realiza en condiciones en las que la fidelidad de copiado de la ADN polimerasa es baja, de manera que se obtiene una alta tasa de mutaciones puntuales a lo largo de toda la longitud del producto de PCR. En resumen, en tales procedimientos, se mezclan ácidos nucleicos que van a mutagenizarse (por ejemplo, una secuencia de polinucleótido que codifica para una enzima AAR) con cebadores de PCR, tampón de reacción, MgCl₂, MnCl₂, Taq polimerasa, y una concentración apropiada de dNTP para lograr una alta tasa de mutaciones puntuales a lo largo de toda la longitud del producto de PCR. Por ejemplo, la reacción puede realizarse usando 20 fmoles de ácido nucleico que va a mutagenizarse, 30 pmoles de cada cebador de PCR, un tampón de reacción que comprende KCl 50 mM, Tris HCl 10 mM (pH 8,3), gelatina al 0,01%, MgCl₂ 7 mM, MnCl₂ 0,5 mM, 5 unidades de Taq polimerasa, dGTP 0,2 mM, dATP 0,2 mM, dCTP 1 mM y dTTP 1 mM. Puede realizarse PCR durante 30 ciclos de 94°C durante 1 min, 45°C durante 1 min, y 72°C durante 1 min. Sin embargo, los expertos en la técnica apreciarán que estos parámetros pueden variarse según sea apropiado. Los ácidos nucleicos mutagenizados se clonan entonces en un vector apropiado, y se evalúan las actividades de los polipéptidos codificados por los ácidos nucleicos mutagenizados. Puede lograrse la mutagénesis dirigida al sitio usando mutagénesis dirigida por oligonucleótidos para generar mutaciones específicas de sitio en cualquier ADN clonado de interés. La mutagénesis por oligonucleótidos se describe en la técnica (véase, por ejemplo, Reidhaar-Olson *et al.* (1988) *Science* 241:53-57). En resumen, en tales procedimientos, se sintetizan una pluralidad de oligonucleótidos bicatenarios que portan una o más mutaciones que van a introducirse en el ADN clonado, y se insertan en el ADN clonado que va a mutagenizarse (por ejemplo, una secuencia de polinucleótido que codifica para un polipéptido de AAR). Se recuperan los clones que contienen el ADN mutagenizado, y se evalúan las actividades de los polipéptidos para los que codifican.

Otro método para generar variantes es la PCR de ensamblaje. La PCR de ensamblaje implica el ensamblaje de un producto de PCR de una mezcla de fragmentos de ADN pequeños. Un gran número de diferentes reacciones PCR se producen en paralelo en el mismo vial, cebando con los productos de una reacción los productos de otra reacción (véase la patente estadounidense 5.965.408). Todavía otro método de generación de variantes es la mutagénesis por PCR sexual. En la mutagénesis por PCR sexual, se produce recombinación homóloga forzada entre moléculas de ADN de diferentes secuencias de ADN, pero altamente relacionadas, *in vitro* como resultado de fragmentación al azar de la molécula de ADN basándose en la homología de secuencia. Esto va seguido por fijación del cruzamiento mediante extensión de cebadores en una reacción PCR. La mutagénesis por PCR sexual se describe en publicaciones conocidas en la técnica (véase, por ejemplo, Stemmer (1994) *Proc. Natl. Acad. Sci. U.S.A.* 91:10747-10751). También pueden crearse variantes mediante mutagénesis *in vivo*. En algunas realizaciones, se generan mutaciones al azar en una secuencia de ácido nucleico propagando la secuencia en una cepa bacteriana, tal como una cepa de *E. coli*, que porta mutaciones en una o más de las rutas de reparación de ADN. Tales cepas mutadoras tienen una mayor tasa de mutación al azar que la de una cepa silvestre. La propagación de una secuencia de ADN (por ejemplo, una secuencia de polinucleótido que codifica para un polipéptido de AAR) en una de estas cepas generará eventualmente mutaciones al azar dentro del ADN. Se describen cepas mutadoras adecuadas para su uso para mutagénesis *in vivo* en la publicación en la técnica (véase, por ejemplo, la publicación de solicitud de patente internacional n.º WO1991/016427). También pueden generarse variantes usando mutagénesis de casete. En la mutagénesis de casete, una pequeña región de una molécula de ADN bicatenario se reemplaza por un casete de oligonucleótido sintético que difiere de la secuencia nativa. El oligonucleótido contiene a menudo una secuencia nativa completa y/o parcialmente aleatorizada. También puede usarse mutagénesis de conjunto recursivo para generar variantes. La mutagénesis de conjunto recursivo es un algoritmo para la modificación por ingeniería de proteínas (es decir, mutagénesis de proteínas) desarrollado para producir diversas poblaciones de mutantes relacionados fenotípicamente cuyos miembros difieren en la secuencia de aminoácidos. Este método usa un mecanismo de realimentación para controlar tandas sucesivas de mutagénesis de casete combinatoria (véase, por ejemplo, Arkin *et al.* (1992) *Proc. Natl. Acad. Sci., U.S.A.* 89:7811-7815). En algunas realizaciones, se crean variantes usando mutagénesis de conjunto exponencial. La mutagénesis de conjunto exponencial es un procedimiento para generar bibliotecas combinatorias con un alto porcentaje de mutantes únicos y funcionales, en los que se aleatorizan pequeños grupos de residuos en paralelo para identificar, en cada posición alterada,

aminoácidos que conducen a proteínas funcionales (véase, por ejemplo, Delegrave *et al.* (1993) *Biotech. Res.* 11:1548-1552). En algunas realizaciones, se crean variantes usando procedimientos de intercambio en los se fusionan entre sí partes de una pluralidad de ácidos nucleicos que codifican para distintos polipéptidos para crear secuencias de ácido nucleico quiméricas que codifican para polipéptidos quiméricos (tal como se describe, por ejemplo, en las patentes estadounidenses n.ºs 5.965.408 y 5.939.250).

Producción de aldehídos grasos y alcoholes grasos

Una célula huésped nativa o recombinante puede comprender un polinucleótido que codifica para una enzima que tiene actividad de biosíntesis de aldehídos grasos (también denominada en el presente documento polipéptido de biosíntesis de aldehídos grasos o enzima o polipéptido de biosíntesis de aldehídos grasos). Se produce un aldehído graso cuando la enzima de biosíntesis de aldehídos grasos se expresa o sobreexpresa en la célula huésped. La expresión o sobreexpresión de un polipéptido de acil-ACP reductasa (AAR) en una célula huésped recombinante puede dar como resultado la producción de un aldehído graso por la célula huésped recombinante. En una realización, la célula huésped recombinante produce un aldehído graso. En algunas realizaciones, el aldehído graso producido por la célula huésped recombinante se convierte en un alcohol graso. En algunas realizaciones, polipéptidos nativos (endógenos) de biosíntesis de aldehídos grasos, tales como aldehído reductasas, están presentes en la célula huésped (por ejemplo, *E. coli*) y son eficaces para convertir aldehídos grasos en alcoholes grasos. En otras realizaciones, se sobreexpresa un polipéptido nativo (endógeno) de biosíntesis de aldehídos grasos. En todavía otras realizaciones, se introduce un polipéptido exógeno de biosíntesis de aldehídos grasos en una célula huésped recombinante y se expresa o sobreexpresa. Un aldehído graso puede producirse expresando o sobreexpresando en la célula huésped recombinante un polinucleótido que codifica para un polipéptido de biosíntesis de aldehídos grasos, tal como un polipéptido que tiene actividad acil-ACP reductasa (AAR). La expresión de AAR en una célula huésped recombinante da como resultado la producción de aldehídos grasos y alcoholes grasos (figura 4). Se describen polipéptidos de AAR a modo de ejemplo en el presente documento y en las publicaciones PCT n.ºs WO2009/140695 y WO/2009/140696.

Se produce una composición que comprende aldehídos grasos (una composición de aldehído graso) cultivando una célula huésped en presencia de una fuente de carbono en condiciones eficaces para expresar la enzima de biosíntesis de aldehídos grasos, por ejemplo, AAR. Una célula huésped recombinante modificada por ingeniería para producir un aldehído graso convertirá normalmente parte del aldehído graso en un alcohol graso. En algunas realizaciones, la composición de aldehído graso comprende aldehídos grasos y alcoholes grasos. Normalmente, la composición de aldehído graso se recupera del entorno extracelular de la célula huésped recombinante, es decir, el medio de cultivo celular. En algunas realizaciones, la célula huésped recombinante incluye un polinucleótido que codifica para un polipéptido (una enzima) que tiene actividad de biosíntesis de alcoholes grasos (también denominado en el presente documento polipéptido de biosíntesis de alcoholes grasos o enzima de biosíntesis de alcoholes grasos), y se produce un alcohol graso por la célula huésped recombinante. Una composición que incluye los alcoholes grasos (es decir, una composición de alcohol graso) puede producirse cultivando la célula huésped recombinante en presencia de una fuente de carbono en condiciones eficaces para expresar un enzima de biosíntesis de alcoholes grasos. Las aldehído reductasas nativas (por ejemplo, endógenas) presentes en una célula huésped recombinante (por ejemplo, *E. coli*), convertirán aldehídos grasos en alcoholes grasos. En algunas realizaciones, polipéptidos nativos (por ejemplo, endógenos) de biosíntesis de aldehídos grasos, tales como aldehído reductasas presentes en la célula huésped son suficientes para convertir aldehídos grasos en alcoholes grasos. Sin embargo, en otras realizaciones, se produce un alcohol graso expresando o sobreexpresando en la célula huésped recombinante un polinucleótido que codifica para un polipéptido que tiene actividad de biosíntesis de alcoholes grasos que convierte un aldehído graso en un alcohol graso. Por ejemplo, una alcohol deshidrogenasa (también denominada en el presente documento aldehído reductasa, por ejemplo, EC 1.1.1.1), puede ser útil en la puesta en práctica de la divulgación. Tal como se usa en el presente documento, el término alcohol deshidrogenasa se refiere a un polipéptido que puede catalizar la conversión de un aldehído graso en un alcohol graso. Un experto habitual en la técnica apreciará que determinadas alcohol deshidrogenasas también pueden catalizar otras reacciones, y estas alcohol deshidrogenasas no específicas también están englobadas por el término alcohol deshidrogenasa. Los ejemplos de polipéptidos de alcohol deshidrogenasa útiles según la divulgación incluyen, pero no se limitan a, AlrA de *Acinetobacter* sp. M-1 (SEQ ID NO: 52) o homólogos de AlrA tales como AlrAadp1 (SEQ ID NO: 53) y alcohol deshidrogenasas endógenas de *E. coli* tales como YjgB, (AAC77226) (SEQ ID NO: 5), DkgA (NP_417485), DkgB (NP_414743), YdjL (AAC74846), YdjJ (NP_416288), AdhP (NP_415995), YhdH (NP_417719), YahK (NP_414859), YphC (AAC75598), YqhD (446856) y YbbO [AAC73595,1]. Se describen ejemplos adicionales en las publicaciones de solicitud de patente internacional n.ºs WO2007/136762, WO2008/119082 y WO2010/062480. En determinadas realizaciones, el polipéptido de biosíntesis de alcoholes grasos tiene actividad aldehído reductasa o alcohol deshidrogenasa (EC 1.1.1.1). En algunas realizaciones, un polipéptido nativo (por ejemplo, endógeno) de biosíntesis de alcoholes grasos se sobreexpresa y en otras realizaciones, un polipéptido exógeno de biosíntesis de alcoholes grasos se introduce en una célula huésped recombinante y se expresa o sobreexpresa.

Pueden producirse alcoholes grasos mediante una ruta dependiente de acil-CoA que utiliza productos intermedios de acil graso-ACP y acil graso-CoA y una ruta independiente de acil-CoA que utiliza productos intermedios de acil graso-ACP pero no un producto intermedio de acil graso-CoA. En realizaciones particulares, se selecciona la enzima codificada por el gen sobreexpresado de una ácido graso sintasa, una acil-ACP tioesterasa, una acil graso-CoA sintasa y una acetil-CoA carboxilasa. También se producen alcoholes grasos en la naturaleza por enzimas que

5 puede reducir diversas moléculas de acil-ACP o acil-CoA para dar los alcoholes primarios correspondientes (véanse las publicaciones de patente estadounidense n.ºs 20100105963 y 20110206630, y la patente estadounidense n.º 8097439). Una composición de alcohol graso incluye a menudo alcoholes grasos junto con otros derivados de ácido graso, por ejemplo, aldehídos grasos y/o ácidos grasos. Normalmente, la composición de alcohol graso se recupera del entorno extracelular de la célula huésped recombinante, es decir, el medio de cultivo celular. En determinadas realizaciones, se modula la expresión de un polipéptido, por ejemplo, una enzima directa o indirectamente implicada en la biosíntesis de ácidos grasos (por ejemplo, se expresa, sobreexpresa o atenúa), en la que tal modulación da como resultado un mayor rendimiento, mayor título o mayor productividad de un derivado de ácido graso de interés, tal como un alcohol graso. La enzima puede estar codificada por un polinucleótido de biosíntesis de ácidos grasos que es exógeno o heterólogo (por ejemplo, un polipéptido que se origina a partir de un organismo distinto de la célula huésped parental, o, una variante de un polipéptido nativo para la célula microbiana parental) o un polipéptido endógeno (por ejemplo, un polipéptido nativo para la célula huésped parental) en el que el polipéptido endógeno se sobreexpresa en la célula huésped recombinante. La tabla 1 proporciona un listado de proteínas a modo de ejemplo que pueden expresarse en células huésped recombinantes para facilitar la producción de composiciones de alcohol graso particulares.

Tabla 1: Designaciones de genes

Designación de gen	Organismo fuente	Nombre de enzima	N.º de registro	Número EC	Uso a modo de ejemplo
Aumento de la producción de ácido graso / aumento de la producción de producto					
<i>accA</i>	<i>E. coli</i> , lactococos	acetil-CoA carboxilasa, subunidad A (carboxiltransferasa alfa)	AAC73296, NP_414727	6.4.1.2	aumento de la producción de malonil-CoA
<i>accB</i>	<i>E. coli</i> , lactococos	acetil-CoA carboxilasa, subunidad B (BCCP: proteína transportadora de biotina carboxilasa)	NP_417721	6.4.1.2	aumento de la producción de malonil-CoA
<i>accC</i>	<i>E. coli</i> , lactococos	acetil-CoA carboxilasa, subunidad C (biotina carboxilasa)	NP_417722	6.4.1.2, 6.3.4.14	aumento de la producción de malonil-CoA
<i>accD</i>	<i>E. coli</i> , lactococos	acetil-CoA carboxilasa, subunidad D (carboxiltransferasa beta)	NP_416819	6.4.1.2	aumento de la producción de malonil-CoA
<i>fadD</i>	<i>E. coli W3110</i>	acil-CoA sintasa	AP_002424	2.3.1.86, 6.2.1.3	aumento de la producción de ácidos grasos
<i>fabA</i>	<i>E. coli K12</i>	β -hidroxidecanoiltoéster deshidratasa/ isomerasa	NP_415474	4.2.1.60	aumento de la producción de acil graso-ACP/CoA
<i>fabB</i>	<i>E. coli</i>	3-oxoacil-[acil-portador-proteína] sintasa I	BAA16180	2.3.1.41	aumento de la producción de acil graso-ACP/CoA
<i>fabD</i>	<i>E. coli K12</i>	[proteína transportadora de acilo] S-maloniltransferasa	AAC74176	2.3.1.39	aumento de la producción de acil graso-ACP/CoA
<i>fabF</i>	<i>E. coli K12</i>	3-oxoacil-[proteína transportadora de acilo] sintasa II	AAC74179	2.3.1.179	aumento de la producción de acil graso-ACP/CoA
<i>fabG</i>	<i>E. coli K12</i>	3-oxoacil-[proteína transportadora de acilo] reductasa	AAC74177	1.1.1.100	aumento de la producción de acil graso-ACP/CoA
<i>fabH</i>	<i>E. coli K12</i>	3-oxoacil-[proteína transportadora de acilo] sintasa III	AAC74175	2.3.1.180	aumento de la producción de acil graso-ACP/CoA
<i>fabI</i>	<i>E. coli K12</i>	enoil-[proteína transportadora de acilo] reductasa	NP_415804	1.3.1.9	aumento de la producción de acil graso-ACP/CoA
<i>fabR</i>	<i>E. coli K12</i>	represor transcripcional	NP_418398	ninguna	modular la producción de ácidos grasos insaturados
<i>fabV</i>	<i>Vibrio cholerae</i>	enoil-[proteína transportadora de acilo] reductasa	YP_001217283	1.3.1.9	aumento de la producción de acil graso-ACP/CoA
<i>fabZ</i>	<i>E. coli K12</i>	(3R)-hidroximristol-proteína transportadora de acilo deshidratasa	NP_414722	4.2.1.-	aumento de la producción de acil graso-ACP/CoA
<i>fadE</i>	<i>E. coli K13</i>	acil-CoA deshidrogenasa	AAC73325	1.3.99.3, 1.3.99.-	reducir la degradación de ácidos grasos
<i>fadR</i>	<i>E. coli</i>	proteína reguladora transcripcional	NP_415705	ninguna	Bloquear o revertir la degradación de ácidos grasos

ES 2 600 354 T3

Designación de gen	Organismo fuente	Nombre de enzima	N.º de registro	Número EC	Uso a modo de ejemplo
Control de la longitud de cadena					
<i>tesA</i> (con o sin secuencia líder)	<i>E. coli</i>	tioesterasa – la secuencia líder son los aminoácidos 1-26	POADA1	3.1.2.-, 3.1.1.5	Longitud de cadena C18
<i>tesA</i> (sin secuencia líder)	<i>E. coli</i>	tioesterasa	AAC73596, NP_415027	3.1.2.-, 3.1.1.5	Longitud de cadena C18:1
<i>tesA</i> (mutante de <i>E. coli</i> tioesterasa I complejo con ácido octanoico)	<i>E. coli</i>	tioesterasa	L109P	3.1.2.-, 3.1.1.5	Longitud de cadena <C18
<i>fatB1</i>	<i>Umbellularia californica</i>	tioesterasa	Q41635	3.1.2.14	Longitud de cadena C12:0
<i>fatB2</i>	<i>Cuphea hookeriana</i>	tioesterasa	AAC49269	3.1.2.14	Longitud de cadena C8:0 - C10:0
<i>fatB3</i>	<i>Cuphea hookeriana</i>	tioesterasa	AAC72881	3.1.2.14	Longitud de cadena C14:0 - C16:0
<i>fatB</i>	<i>Cinnamomum camphora</i>	tioesterasa	Q39473	3.1.2.14	Longitud de cadena C14:0
<i>fatB</i>	<i>Arabidopsis thaliana</i>	tioesterasa	CAA85388	3.1.2.14	Longitud de cadena C16:1
<i>fatA1</i>	<i>Helianthus annuus</i>	tioesterasa	AAL79361	3.1.2.14	Longitud de cadena C18:1
<i>atfata</i>	<i>Arabidopsis thaliana</i>	tioesterasa	NP_189147, NP_193041	3.1.2.14	Longitud de cadena C18:1
<i>fatA</i>	<i>Brassica juncea</i>	tioesterasa	CAC39106	3.1.2.14	Longitud de cadena C18:1
<i>fatA</i>	<i>Cuphea hookeriana</i>	tioesterasa	AAC72883	3.1.2.14	Longitud de cadena C18:1
<i>tesA</i>	<i>Photobacterium profundum</i>	tioesterasa	YP_130990	3.1.2.14	Longitud de cadena
<i>tesB</i>	<i>E. coli</i>	tioesterasa	NP_414986	3.1.2.14	Longitud de cadena
<i>fadM</i>	<i>E. coli</i>	tioesterasa	NP_414977	3.1.2.14	Longitud de cadena
<i>yciA</i>	<i>E. coli</i>	tioesterasa	NP_415769	3.1.2.14	Longitud de cadena
<i>ybgC</i>	<i>E. coli</i>	tioesterasa	NP_415264	3.1.2.14	Longitud de cadena
Control del nivel de saturación*					
<i>Sfa</i>	<i>E. coli</i>	supresor de <i>fabA</i>	AAN79592, AAC44390	ninguno	aumento de ácidos grasos monoinsaturados
<i>fabA</i>	<i>E. coli K12</i>	β-hidroxidecanoil-tioéster deshidratasa/isomerasa	NP_415474	4.2.1.60	producir ácidos grasos insaturados
<i>GnsA</i>	<i>E. coli</i>	supresores de la mutación nula de <i>secG</i>	ABD 18647.1	ninguno	aumento de ésteres de ácidos grasos insaturados
<i>GnsB</i>	<i>E. coli</i>	supresores de la mutación nula de <i>secG</i>	AAC74076.1	ninguno	aumento de ésteres de ácidos grasos insaturados
<i>fabB</i>	<i>E. coli</i>	3-oxoacil-[proteína transportadora de acilo] sintasa I	BAA16180	2.3.1.41	modular la producción de ácidos grasos insaturados
<i>des</i>	<i>Bacillus subtilis</i>	D5 acil graso desaturasa	O34653	1.14.19	modular la producción de ácidos grasos insaturados
Salida de productos: Producción de ésteres					
<i>AT3G51970</i>	<i>Arabidopsis thaliana</i>	alcohol de cadena larga O-acil graso transferasa	NP_190765	2.3.1.26	producción de cera
<i>ELO1</i>	<i>Pichia angusta</i>	ácido graso elongasa	BAD98251	2.3.1.-	producir ácidos grasos de longitud de cadena de muy larga
<i>plsC</i>	<i>Saccharomyces cerevisiae</i>	aciltransferasa	AAA16514	2.3.1.51	producción de cera
<i>DAGAT/DG</i>	<i>Arabidopsis</i>	diacilglicerolaciltransferasa	AAF19262	2.3.1.20	producción de cera

ES 2 600 354 T3

Designación de gen	Organismo fuente	Nombre de enzima	N.º de registro	Número EC	Uso a modo de ejemplo
<i>AT</i>	<i>thaliana</i>				
<i>hWS</i>	Homo sapiens	acil-CoA cera alcohol aciltransferasa	AAX48018	2.3.1.20	producción de cera
<i>aft1</i>	<i>Acinetobacter</i> sp. ADP1	éster de cera bifuncional sintasa/acil-CoA:diacilglicerolacil transferasa	AAO17391	2.3.1.20	producción de cera
<i>ES9</i>	<i>Marinobacter hydrocarbono-clasticus</i>	éster de cera sintasa	ABO21021	2.3.1.20	producción de cera
<i>mWS</i>	<i>Simmondsiachi nensis</i>	éster de cera sintasa	AAD38041	2.3.1.-	producción de cera
<i>acr1</i>	<i>Acinetobacter</i> sp. ADP1	acil-CoA reductasa	YP_047869	1.2.1.42	modificar salida
<i>yqhD</i>	<i>E. coli</i> K12	alcohol deshidrogenasa	AP_003562	1.1.-.-	modificar salida
<i>AAT</i>	<i>Fragaria x ananassa</i>	alcohol O-acetiltransferasa	AAG13130	2.3.1.84	modificar salida
Salida de productos: Salida de alcohol graso					
		tioesterasas (véase anteriormente)			aumento de la producción de ácidos grasos/ alcoholes grasos
<i>BmFAR</i>	<i>Bombyx mori</i>	FAR (acil-CoA reductasa de formación de alcohol graso)	BAC79425	1.1.1.-	convertir acil-CoA en alcohol graso
<i>acr1</i>	<i>Acinetobacter</i> sp. ADP1	acil-CoA reductasa	YP_047869	1.2.1.42	reducir acil graso-CoA para dar aldehídos grasos
<i>yqhD</i>	<i>E. coli</i> W3110	alcohol deshidrogenasa	AP_003562	1.1.-.-	reducir aldehídos grasos para dar alcoholes grasos; aumento de la producción de alcoholes grasos
<i>alrA</i>	<i>Acinetobacter</i> sp. ADP1	alcohol deshidrogenasa	CAG70252	1.1.-.-	reducir aldehídos grasos para dar alcoholes grasos
<i>BmFAR</i>	<i>Bombyx mori</i>	FAR (acil-CoA reductasa de formación de alcohol graso)	BAC79425	1.1.1.-	reducir acil graso-CoA para dar alcohol graso
<i>GTNG_1865</i>	<i>Geobacillus thermodenitrificans</i> NG80-2	aldehído de cadena larga deshidrogenasa	YP_001125970	1.2.1.3	reducir aldehídos grasos para dar alcoholes grasos
<i>AAR</i>	<i>Synechococcus elongatus</i>	acil-ACP reductasa	YP_400611	1.2.1.80 1.2.1.42	reducir acil graso-ACP/CoA para dar aldehídos grasos
<i>carB</i>	<i>Mycobacterium smegmatis</i>	proteína ácido carboxílico reductasa (CAR)	YP_889972	6.2.1.3, 1.2.1.42	reducir ácidos grasos para dar aldehído graso
<i>FadD</i>	<i>E. coli</i> K12	acil-CoA sintetasa	NP_416319	6.2.1.3	activa ácidos grasos para dar acil graso-CoA
<i>atoB</i>	<i>Erwinia carotovora</i>	acetil-CoA acetiltransferasa	YP_049388	2.3.1.9	producción de butanol
<i>hbd</i>	<i>Butyrivibrio fibrisolvens</i>	beta-hidroxibutiril-CoA deshidrogenasa	BAD51424	1.1.1.157	producción de butanol
<i>CPE0095</i>	<i>Clostridium perfringens</i>	crotonasebutiril-CoA deshidrogenasa	BAB79801	4.2.1.55	producción de butanol
<i>bcd</i>	<i>Clostridium beijerinckii</i>	butiril-CoA deshidrogenasa	AAM14583	1.3.99.2	producción de butanol
<i>ALDH</i>	<i>Clostridium beijerinckii</i>	aldehído deshidrogenasa de acilación de coenzima A	AAT66436	1.2.1.3	producción de butanol
<i>AdhE</i>	<i>E. coli</i> CFT073	aldehído-alcohol deshidrogenasa	AAN80172	1.1.1.1 1.2.1.10	producción de butanol
Exportación de producto					
<i>AtMRP5</i>	<i>Arabidopsis thaliana</i>	<i>Arabidopsis thaliana</i> asociada a resistencia a múltiples fármacos	NP_171908	ninguno	modificar producto exportar cantidad
<i>AmiS2</i>	<i>Rhodococcus</i> sp.	Transportador de ABC AmiS2	JC5491	ninguno	modificar producto exportar cantidad

Designación de gen	Organismo fuente	Nombre de enzima	N.º de registro	Número EC	Uso a modo de ejemplo
<i>AtPGP1</i>	<i>Arabidopsis thaliana</i>	Glicoproteína 1 de <i>Arabidopsis thaliana</i> p	NP_181228	ninguno	modificar producto exportar cantidad
<i>AcrA</i>	<i>Candidatus Protochlamydia amoebophila UWE25</i>	supuesta proteína transportadora de flujo de entrada de múltiples fármacos <i>acrA</i>	CAF23274	ninguno	modificar producto exportar cantidad
<i>AcrB</i>	<i>Candidatus Protochlamydia amoebophila UWE25</i>	probable proteína transportadora de flujo de entrada de múltiples fármacos, <i>acrB</i>	CAF23275	ninguno	modificar producto exportar cantidad
<i>TolC</i>	<i>Francisella tularensis subsp. novicida</i>	proteína de la membrana externa [biogénesis de envuelta celular,	ABD59001	ninguno	modificar producto exportar cantidad
<i>AcrE</i>	<i>Shigella sonnei Ss046</i>	proteína transmembrana afecta a la formación del tabique y la permeabilidad de la membrana celular	YP_312213	ninguno	modificar producto exportar cantidad
<i>AcrF</i>	<i>E. coli</i>	proteína F con resistencia a acriflavina	P24181	ninguno	modificar producto exportar cantidad
<i>tII1619</i>	<i>Thermosynechococcus elongatus [BP-1]</i>	transportador de flujo de entrada de múltiples fármacos	NP_682409.1	ninguno	modificar producto exportar cantidad
<i>tI10139</i>	<i>Thermosynechococcus elongatus [BP-1]</i>	transportador de flujo de entrada de múltiples fármacos	NP_680930.1	ninguno	modificar producto exportar cantidad
Fermentación					
replicación checkpoint genes					aumentar eficiencia de salida
<i>umuD</i>	<i>Shigella sonnei Ss046</i>	ADN polimerasa V, subunidad	YP_310132	3.4.21.-	aumentar eficiencia de salida
<i>umuC</i>	<i>E. coli</i>	ADN polimerasa V, subunidad	ABC42261	2.7.7.7	aumentar eficiencia de salida
<i>pntA, pntB</i>	<i>Shigella flexneri</i>	NADH:NADPH transhidrogenasa (subunidades alfa y beta)	P07001, P0AB70	1.6.1.2	aumentar eficiencia de salida
Otros					
<i>fabK</i>	<i>Streptococcus pneumoniae</i>	trans-2-enoil-ACP reductasa II	AAF98273	1.3.1.9	Contribuye a la biosíntesis de ácidos grasos
<i>fabL</i>	<i>Bacillus licheniformis DSM 13</i>	enoiil-(proteína transportadora de acilo) reductasa	AAU39821	1.3.1.9	Contribuye a la biosíntesis de ácidos grasos
<i>fabM</i>	<i>Streptococcus mutans</i>	trans-2, cis-3-decenoil-ACP isomerasa	DAA05501	4.2.1.17	Contribuye a la biosíntesis de ácidos grasos

Células huésped recombinantes y cultivos celulares

5 Las estrategias para aumentar la producción de composiciones de aldehído graso o de alcohol graso por células huésped recombinantes incluyen el aumento del flujo a través de la ruta de biosíntesis de ácidos grasos mediante la sobreexpresión de genes nativos de biosíntesis de aldehídos grasos o alcoholes grasos y la expresión de genes exógenos de biosíntesis de aldehídos grasos o alcoholes grasos de diferentes organismos en el huésped de producción. Tal como se usa en el presente documento, el término célula huésped recombinante o célula huésped modificada por ingeniería se refiere a una célula huésped cuya composición genética se ha alterado con relación a la célula huésped silvestre correspondiente, por ejemplo, mediante introducción deliberada de nuevos elementos genéticos y/o la modificación deliberada de elementos genéticos presentes de manera natural en la célula huésped.

10 La descendencia de tales células huésped recombinantes también contiene estos elementos genéticos nuevos y/o modificados. En cualquiera de los aspectos de la divulgación descritos en el presente documento, la célula huésped

puede seleccionarse de una célula vegetal, célula de insecto, célula fúngica (por ejemplo, de un hongo filamentoso, tal como *Candida* sp., o una levadura en gemación, tal como *Saccharomyces* sp.), una célula de alga y una célula bacteriana. En una realización preferida, las células huésped recombinantes son microorganismos recombinantes. Los ejemplos de células huésped que son microorganismos incluyen, pero no se limitan a, células del género

5 *Escherichia*, *Bacillus*, *Lactobacillus*, *Zymomonas*, *Rhodococcus*, *Pseudomonas*, *Aspergillus*, *Trichoderma*, *Neurospora*, *Fusarium*, *Humicola*, *Rhizomucor*, *Kluyveromyces*, *Pichia*, *Mucor*, *Myceliophthora*, *Penicillium*, *Phanerochaete*, *Pleurotus*, *Trametes*, *Chrysosporium*, *Saccharomyces*, *Stenotrophomonas*, *Schizosaccharomyces*, *Yarrowia* o *Streptomyces*. En algunas realizaciones, la célula huésped es una célula bacteriana Gram-positiva. En otras realizaciones, la célula huésped es una célula bacteriana Gram-negativa. En algunas realizaciones, la célula huésped es una célula de *E. coli*. En otras realizaciones, la célula huésped es una célula de *Bacillus lentus*, una célula de *Bacillus brevis*, una célula de *Bacillus stearothermophilus*, una célula de *Bacillus licheniformis*, una célula de *Bacillus alkalophilus*, una célula de *Bacillus coagulans*, una célula de *Bacillus circulans*, una célula de *Bacillus pumilis*, una célula de *Bacillus thuringiensis*, una célula de *Bacillus clausii*, una célula de *Bacillus megaterium*, una célula de *Bacillus subtilis* o una célula de *Bacillus amyloliquefaciens*. En otras realizaciones, la célula huésped es una célula de *Trichoderma koningii*, una célula de *Trichoderma viride*, una célula de *Trichoderma reesei*, una célula de *Trichoderma longibrachiatum*, una célula de *Aspergillus awamori*, una célula de *Aspergillus fumigates*, una célula de *Aspergillus foetidus*, una célula de *Aspergillus nidulans*, una célula de *Aspergillus niger*, una célula de *Aspergillus oryzae*, una célula de *Humicola insolens*, una célula de *Humicola lanuginosa*, una célula de *Rhodococcus opacus*, una célula de *Rhizomucor miehei* o una célula *Mucor miehei*. En aún otras realizaciones, la célula huésped es una célula de *Streptomyces lividans* o una célula *Streptomyces murinus*. En aún otras realizaciones, la célula huésped es una célula *Actinomycetes*. En algunas realizaciones, la célula huésped es una célula *Saccharomyces cerevisiae*.

En otras realizaciones, la célula huésped es una célula vegetal eucariota, una célula de alga, una célula de cianobacteria, una célula de bacteria verde de azufre, una célula de bacteria verde no de azufre, una célula de bacteria púrpura de azufre, una célula de bacteria púrpura no de azufre, una célula extremófila, una célula de levadura, una célula fúngica, una célula modificada por ingeniería de cualquiera de las especies descritas en el presente documento, o un organismo sintético. En algunas realizaciones, la célula huésped depende la luz o fija carbono. En algunas realizaciones, la célula huésped tiene actividad autótrofa. En algunas realizaciones, la célula huésped tiene actividad fotoautótrofa, tal como en presencia de luz. En algunas realizaciones, la célula huésped es heterótrofa o mixótrofa en ausencia de luz. En determinadas realizaciones, la célula huésped es una célula de

25 *Arabidopsis thaliana*, *Panicum virgatum*, *Miscanthus giganteus*, *Zea mays*, *Botryococcuse braunii*, *Chlamydomonas reinhardtii*, *Dunaliella salina*, *Synechococcus* Sp. PCC 7002, *Synechococcus* Sp. PCC 7942, *Synechocystis* Sp. PCC 6803, *Thermosynechococcus elongates BP-1*, *Chlorobium tepidum*, *Chlorojlexus auranticus*, *Chromatium vinosum*, *Rhodospirillum rubrum*, *Rhodobacter capsulatus*, *Rhodopseudomonas palustris*, *Clostridium ljungdahlii*, *Clostridium thermocellum*, *Penicillium chrysogenum*, *Pichia pastoris*, *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Pseudomonas fluorescens* o *Zymomonas mobilis*.

Modificación por ingeniería de células huésped

En algunas realizaciones, se proporciona una secuencia de polinucleótido (o gen) a la célula huésped por medio de un vector recombinante, que incluye un promotor operativamente unido a la secuencia de polinucleótido. En determinadas realizaciones, el promotor es un promotor regulado en el desarrollo, específico de orgánulo, específico de tejido, inducible, constitutivo o uno específico de célula. En algunas realizaciones, el vector recombinante incluye al menos una secuencia seleccionada de una secuencia de control de la expresión acoplada operativamente a la secuencia de polinucleótido; un marcador de selección acoplado operativamente a la secuencia de polinucleótido; una secuencia de marcador acoplada operativamente a la secuencia de polinucleótido; un resto de purificación acoplado operativamente a la secuencia de polinucleótido; una secuencia de secreción acoplada operativamente a la secuencia de polinucleótido; y una secuencia de selección como diana acoplada operativamente a la secuencia de polinucleótido. Los vectores de expresión descritos en el presente documento incluyen una secuencia de polinucleótido en una forma adecuada para la expresión de la secuencia de polinucleótido en una célula huésped. Los expertos en la técnica apreciarán que el diseño del vector de expresión puede depender de factores tales como la elección de la célula huésped que va transformarse, el nivel de expresión de polipéptido deseado, y similares. Los vectores de expresión descritos en el presente documento pueden introducirse en células huésped para producir polipéptidos, incluyendo polipéptidos de fusión, codificados por las secuencias de polinucleótido tal como se describió anteriormente (véase antes). La expresión de genes que codifican para polipéptidos en procariotas, por ejemplo, *E. coli*, se lleva a cabo con la mayor frecuencia con vectores que contienen promotores constitutivos o inducibles que dirigen la expresión de polipéptidos o bien de fusión o bien de no fusión. Los vectores de fusión añaden un número de aminoácidos a un polipéptido codificado en los mismos, habitualmente al extremo amino-terminal o carboxi-terminal del polipéptido recombinante. Tales vectores de fusión sirven normalmente para uno o más de los tres propósitos siguientes incluyendo aumentar expresión del polipéptido recombinante; aumentar la solubilidad del polipéptido recombinante; y ayudar en la purificación del polipéptido recombinante actuando como ligando en purificación por afinidad. A menudo, en los vectores de expresión de fusión, se introduce un sitio de escisión proteolítica en la unión del resto de fusión y el polipéptido recombinante. Esto permite la separación del polipéptido recombinante del resto de fusión después de la purificación del polipéptido de fusión. Los ejemplos de tales enzimas, y sus secuencias de reconocimiento relacionadas, incluyen factor Xa, trombina y enterocinasa. Los vectores de expresión de fusión a modo de ejemplo incluyen el vector pGEX (Pharmacia Biotech, Inc., Piscataway,

NJ; Smith *et al.* (1988) Gene 67:31-40), el vector pMAL (New England Biolabs, Beverly, MA) y el vector pRITS (Pharmacia Biotech, Inc., Piscataway, N.J.), que fusionan glutatión S-transferasa (GST), proteína de unión a maltosa E, o proteína A, respectivamente, con el polipéptido recombinante diana.

- Los ejemplos de vectores de expresión de *E. coli* inducibles, no de fusión incluyen el vector pTrc (Aman *et al.* (1988) Gene 69:301-315) y el vector pET 11d (Studier *et al.*, Gene Expression Technology: Methods in Enzymology 185, Academic Press, San Diego, Calif. (1990) 60-89). La expresión de genes diana a partir del vector pTrc se basa en la transcripción de ARN polimerasa del huésped a partir de un promotor de fusión trp-lac híbrido. La expresión de genes diana a partir del vector pET 11d se basa en la transcripción a partir de un promotor de fusión gn10-lac de T7 mediado por una ARN polimerasa viral coexpresada (gn1 de T7). Esta polimerasa viral la suministran cepas huésped tales como BL21(DE3) o HMS174(DE3) de un profago λ residente que alberga un gen de gn1 de T7 bajo el control transcripcional del promotor lacUV 5. Se conocen bien en la técnica sistemas de expresión adecuados para células tanto procariontas como eucariotas (véase, por ejemplo, Sambrook *et al.* (1989) Molecular Cloning: A Laboratory Manual, segunda edición, Cold Spring Harbor Laboratory). Los ejemplos de vectores de expresión de *E. coli* inducibles, no de fusión incluyen el vector pTrc (Aman *et al.* (1988) Gene 69:301-315) y el vector PET 11d (Studier *et al.* (1990) Gene Expression Technology: Methods in Enzymology 185, Academic Press, San Diego, CA, pp. 60-89). En determinadas realizaciones, una secuencia de polinucleótido de la divulgación está operativamente unida a un promotor derivado del bacteriófago T5. En una realización, la célula huésped es una célula de levadura. En esta realización, el vector de expresión es un vector de expresión de levadura. Pueden introducirse vectores en células procariontas o eucariotas mediante una variedad de técnicas reconocidas en la técnica para introducir ácido nucleico foráneo (por ejemplo, ADN) en una célula huésped. Pueden encontrarse métodos adecuados para transformar o transfectar células huésped, por ejemplo, en Sambrook *et al.* (véase antes). Para la transformación estable de células bacterianas, se sabe que (dependiendo del vector de expresión y la técnica de transformación usados) una determinada fracción de células captarán y replicarán el vector de expresión. Para identificar y seleccionar estos transformantes, puede introducirse un gen que codifica para un marcador seleccionable (por ejemplo, resistencia a un antibiótico) en las células huésped junto con el gen de interés. Los marcadores seleccionables incluyen los que confieren resistencia a fármacos tales como, pero sin limitarse a, ampicilina, kanamicina, cloranfenicol o tetraciclina. Pueden introducirse ácidos nucleicos que codifican para un marcador seleccionable en una célula huésped en el mismo vector que el que codifica para un polipéptido descrito en el presente documento o pueden introducirse en un vector independiente. Pueden identificarse células transformadas de manera estable con el ácido nucleico introducido mediante crecimiento en presencia de un fármaco de selección apropiado. La célula huésped modificada por ingeniería o recombinante tal como se describe en el presente documento (véase antes) es una célula usada para producir una composición de derivado de ácido graso tal como un aldehído graso o un alcohol graso. En cualquiera de los aspectos de la divulgación descritos en el presente documento, la célula huésped puede seleccionarse de una planta eucariota, bacteria, alga, cianobacteria, bacteria verde de azufre, bacteria verde no de azufre, bacteria púrpura de azufre, bacteria púrpura no de azufre, extremófila, levadura, hongo, organismos modificados por ingeniería de los mismos, o un organismo sintético. En algunas realizaciones, la célula huésped depende de la luz o fija carbono. En algunas realizaciones, la célula huésped tiene actividad autótrofa. Diversas células huésped pueden usarse para producir derivados de ácido graso, tal como se describe en el presente documento.
- Las células huésped o microorganismos de la divulgación incluyen cepas huésped o células huésped que se modifican por ingeniería genética o se modifican para contener alteraciones para someter a prueba la eficacia de mutaciones específicas sobre actividades enzimáticas (es decir, células o microorganismos recombinantes). Pueden usarse de manera intercambiable diversas manipulaciones y alteraciones genéticas opcionales de una célula huésped a otra, dependiendo de qué rutas enzimáticas nativas estén presentes en la célula huésped original. En una realización, puede usarse una cepa huésped para someter a prueba la expresión de un polipéptido de AAR en combinación con otros polipéptidos de biosíntesis (por ejemplo, enzimas). Una cepa huésped puede englobar varias alteraciones genéticas para someter a prueba variables específicas, incluyendo pero sin limitarse a, condiciones de cultivo incluyendo componentes de fermentación, fuente de carbono (por ejemplo, materia prima), temperatura, presión, condiciones de contaminación reducida del cultivo y niveles de oxígeno.
- En una realización, una cepa huésped engloba una delección de *fadE* y *fhuA* opcional. La acil-CoA deshidrogenasa (FadE) es una enzima que es importante para metabolizar ácidos grasos. Cataliza la segunda etapa en la utilización de ácidos grasos (beta-oxidación), que es el proceso de ruptura de largas cadenas de ácidos grasos (acil-CoA) para dar moléculas de acetil-CoA. Más específicamente, la segunda etapa del ciclo de β -oxidación de degradación de ácidos grasos en bacterias es la oxidación de acil-CoA para dar 2-enoil-CoA, que está catalizada por FadE. Cuando *E. coli* carece de FadE, no puede crecer con ácidos grasos como fuente de carbono pero puede crecer con acetato. La incapacidad para utilizar ácidos grasos de cualquier longitud de cadena concuerda con el fenotipo notificado de cepas de *fadE*, es decir, cepas mutantes de *fadE* en las que se perturba la función de FadE. El gen *fadE* puede desactivarse opcionalmente o atenuarse para garantizar que las acil-CoA, que pueden ser productos intermedios en una ruta de derivados de ácido graso, pueden acumularse en la célula de manera que todas las acil-CoA pueden convertirse eficazmente en derivados de ácido graso. Sin embargo, la atenuación de *fadE* es opcional cuando se usa azúcar como fuente de carbono puesto que en tales condiciones, es probable que esté reprimida la expresión de FadE y por tanto FadE sólo puede estar presente en pequeñas cantidades y no puede competir eficazmente con éster sintasa u otras enzimas por los sustratos de acil-CoA. FadE se reprime debido a represión por catabolitos. *E.*

coli y muchos otros microbios prefieren consumir azúcar con respecto a ácidos grasos, así que cuando están disponibles ambas fuentes, se consume en primer lugar el azúcar mediante la represión del regulón *fad* (véase D. Clark, J Bacteriol. (1981) 148(2):521-6)). Además, la ausencia de azúcares y la presencia de ácidos grasos induce la expresión de *FadE*. Podrían perderse productos intermedios de acil-CoA en la ruta de beta-oxidación puesto que las proteínas expresadas por el regulón *fad* (incluyendo *FadE*) están reguladas por incremento y competirán eficazmente por las acil-CoA. Por tanto, puede ser beneficioso tener el gen *fadE* desactivado o atenuado. Puesto que la mayor parte de las fuentes de carbono se basan principalmente en azúcar, es opcional atenuar *FadE*. El gen *fhuA* codifica para la proteína TonA, que es un transportador con acoplamiento de energía y receptor en la membrana externa de *E. coli* (V. Braun (2009) J Bacteriol. 191(11):3431-3436). Su delección es opcional. La delección de *fhuA* permite que la célula se vuelva más resistente al ataque por fagos, lo que puede ser beneficioso en determinadas condiciones de fermentación. Por tanto, puede ser deseable deleccionar *fhuA* en una célula huésped que es probable que esté sujeta a posible contaminación durante series de fermentación.

En otra realización, la cepa huésped (véase antes) también engloba la sobreexpresión opcional de uno o más de los siguientes genes incluyendo *fadR*, *fabA*, *fabD*, *fabG*, *fabH*, *fabV* y/o *fabF*. Ejemplos de tales genes son *fadR* de *Escherichia coli*, *fabA* de *Salmonella typhimurium* (NP_460041), *fabD* de *Salmonella typhimurium* (NP_460164), *fabG* de *Salmonella typhimurium* (NP_460165), *fabH* de *Salmonella typhimurium* (NP_460163), *fabV* de *Vibrio cholera* (YP_001217283) y *fabF* de *Clostridium acetobutylicum* (NP_350156). La sobreexpresión de uno o más de estos genes, que codifican para enzimas y reguladores en la biosíntesis de ácidos grasos, puede servir para aumentar el título de compuestos derivados de ácido graso incluyendo aldehídos grasos y alcoholes grasos en diversas condiciones de cultivo.

En otra realización, se usan cepas de *E. coli* como células huésped para la producción de derivados de ácido graso tales como aldehídos grasos y/o alcoholes grasos. De manera similar, estas células huésped proporcionan la sobreexpresión opcional de uno o más genes de biosíntesis (es decir, genes que codifican para enzimas y reguladores de la biosíntesis de ácidos grasos) que pueden aumentar o potenciar adicionalmente el título de compuestos derivados de ácido graso tales como derivados de ácido graso (por ejemplo, alcoholes grasos, aldehídos grasos, etc.) en diversas condiciones de cultivo incluyendo, pero sin limitarse a, *fadR*, *fabA*, *fabD*, *fabG*, *fabH*, *fabV* y/o *fabF*. Los ejemplos de alteraciones genéticas incluyen *fadR* de *Escherichia coli*, *fabA* de *Salmonella typhimurium* (NP_460041), *fabD* de *Salmonella typhimurium* (NP_460164), *fabG* de *Salmonella typhimurium* (NP_460165), *fabH* de *Salmonella typhimurium* (NP_460163), *fabV* de *Vibrio cholera* (YP_001217283) y *fabF* de *Clostridium acetobutylicum* (NP_350156). En algunas realizaciones, modificarse por ingeniería operones sintéticos que portan estos genes de biosíntesis y expresarse en células para someter a prueba la sobreexpresión de aldehídos grasos y/o alcoholes grasos en diversas condiciones de cultivo y/o potenciar adicionalmente la producción de alcoholes grasos y/o aldehídos grasos. Tales operones sintéticos contienen uno o más genes de biosíntesis. El operón *ifab138*, por ejemplo, es un operón modificado por ingeniería que contiene genes de biosíntesis de ácidos grasos opcionales, incluyendo *fabV* de *Vibrio cholera*, *fabH* de *Salmonella typhimurium*, *fabD* de *S. typhimurium*, *fabG* de *S. typhimurium*, *fabA* de *S. typhimurium* y/o *fabF* de *Clostridium acetobutylicum* que pueden usarse para facilitar la sobreexpresión de derivados de ácido graso para someter a prueba condiciones de cultivo específicas. Una ventaja de tales operones sintéticos es que la tasa de producción de derivados de ácido graso aumentarse o potenciarse adicionalmente.

En algunas realizaciones, las células huésped o microorganismos que se usan para expresar ACP y enzimas de biosíntesis (por ejemplo, TE, ES, CAR, AAR, ADC, etc.) expresarán además genes que engloban determinadas actividades enzimáticas que pueden aumentar la producción para uno o más derivados de ácido graso particulares tales como alcoholes grasos, aldehídos grasos, ésteres grasos, aminas grasas, derivados bifuncionales de ácido graso, diácidos y similares. En una realización, la célula huésped tiene actividad tioesterasa (E.C.3.1.2.* o E.C.3.1.2.14 o E.C.3.1.1.5) para la producción de ácidos grasos que puede aumentarse mediante la sobreexpresión del gen. En otra realización, la célula huésped tiene actividad éster sintasa (E.C.2.3.1.75) para la producción de ésteres grasos. En otra realización, la célula huésped tiene actividad acil-ACP reductasa (AAR) (E.C.1.2.1.80) y/o actividad alcohol deshidrogenasa (E.C.1.1.1.1.) y/o actividad alcohol graso acil-CoA reductasa (FAR) (E.C.1.1.1.*) y/o actividad ácido carboxílico reductasa (CAR) (E.C.1.2.99.6) para la producción de alcoholes grasos. En otra realización, la célula huésped tiene actividad acil-ACP reductasa (AAR) (E.C.1.2.1.80) para la producción de aldehídos grasos. En otra realización, la célula huésped tiene actividad acil-ACP reductasa (AAR) (E.C.1.2.1.80) y actividad descarboxilasa (ADC) para la producción de alcanos y alquenos. En otra realización, la célula huésped tiene actividad acil-CoA reductasa (E.C.1.2.1.50), actividad acil-CoA sintasa (*FadD*) (E.C.2.3.1.86) y actividad tioesterasa (E.C.3.1.2.* o E.C.3.1.2.14 o E.C.3.1.1.5) para la producción de alcoholes grasos. En otra realización, la célula huésped tiene actividad éster sintasa (E.C.2.3.1.75), actividad acil-CoA sintasa (*FadD*) (E.C.2.3.1.86) y actividad tioesterasa (E.C.3.1.2.* o E.C.3.1.2.14 o E.C.3.1.1.5) para la producción de ésteres grasos. En otra realización, la célula huésped tiene actividad *OleA* para la producción de cetonas. En otra realización, la célula huésped tiene actividad *OleBCD* para la producción de olefinas internas. En otra realización, la célula huésped tiene actividad acil-ACP reductasa (AAR) (E.C.1.2.1.80) y actividad alcohol deshidrogenasa (E.C.1.1.1.1.) para la producción de alcoholes grasos. En otra realización, la célula huésped tiene actividad tioesterasa (E.C.3.1.2.* o E.C.3.1.2.14 o E.C.3.1.1.5) y actividad descarboxilasa para producir olefinas terminales. La expresión de actividades enzimáticas en microorganismos y células microbianas se enseña en las patentes estadounidenses número 8.097.439; 8.110.093; 8.110.670; 8.183.028; 8.268.599; 8.283.143; 8.232.924; 8.372.610; y 8.530.221. En otras

realizaciones, las células huésped o microorganismos que se usan para expresar ACP y otras enzimas de biosíntesis incluirán determinadas actividades enzimáticas nativas que están reguladas por incremento o sobreexpresadas para producir uno o más derivados de ácido graso particulares tales como aldehídos grasos y/o alcoholes grasos. En una realización, la célula huésped tiene una actividad tioesterasa nativa (E.C.3.1.2.* o E.C.3.1.2.14 o E.C.3.1.1.5) para la producción de ácidos grasos que puede aumentarse mediante la sobreexpresión del gen de tioesterasa.

La presente divulgación incluye cepas huésped o microorganismos que expresan genes que codifican para AAR y otras enzimas de biosíntesis (véase antes). Las células huésped recombinantes producen derivados de ácido graso tales como aldehídos grasos y alcoholes grasos y composiciones y mezclas de los mismos. Los derivados de ácido graso se recuperan normalmente del medio de cultivo y/o se aíslan de las células huésped. En una realización, los derivados de ácido graso tales como aldehídos grasos y alcoholes grasos se recuperan del medio de cultivo (extracelular). En otra realización, los derivados de ácido graso tales como aldehídos grasos y alcoholes grasos se aíslan de las células huésped (intracelular). En otra realización, los derivados de ácido graso tales como aldehídos grasos y alcoholes grasos se recuperan del medio de cultivo y se aíslan de las células huésped. La composición de derivados de ácido graso producida por una célula huésped puede analizarse usando métodos conocidos en la técnica, por ejemplo, CG-FID, para determinar la distribución de derivados de ácido graso particulares así como longitudes de cadena y grado de saturación de los componentes de la composición de derivados de ácido graso tales como composiciones de aldehído graso y alcohol graso.

Los ejemplos de células huésped que funcionan como microorganismos (por ejemplo, células microbianas), incluyen pero no se limitan a células del género *Escherichia*, *Bacillus*, *Lactobacillus*, *Zymomonas*, *Rhodococcus*, *Pseudomonas*, *Aspergillus*, *Trichoderma*, *Neurospora*, *Fusarium*, *Humicola*, *Rhizomucor*, *Kluyveromyces*, *Pichia*, *Mucor*, *Myceliophthora*, *Penicillium*, *Phanerochaete*, *Pleurotus*, *Trametes*, *Chrysosporium*, *Saccharomyces*, *Stenotrophomonas*, *Schizosaccharomyces*, *Yarrowia* o *Streptomyces*. En algunas realizaciones, la célula huésped es una célula bacteriana Gram-positiva. En otras realizaciones, la célula huésped es una célula bacteriana Gram-negativa. En algunas realizaciones, la célula huésped es una célula de *E. coli*. En alguna realización, la célula huésped es una célula B de *E. coli*, una célula C de *E. coli*, una célula K de *E. coli* o una célula W de *E. coli*. En otras realizaciones, la célula huésped es una célula de *Bacillus lentus*, una célula de *Bacillus brevis*, una célula de *Bacillus stearothermophilus*, una célula de *Bacillus licheniformis*, una célula de *Bacillus alkalophilus*, una célula de *Bacillus coagulans*, una célula de *Bacillus circulans*, una célula de *Bacillus pumilus*, una célula de *Bacillus thuringiensis*, una célula de *Bacillus clausii*, una célula de *Bacillus megaterium*, una célula de *Bacillus subtilis* o una célula de *Bacillus amyloliquefaciens*. En todavía otras realizaciones, la célula huésped es una célula de *Trichoderma koningii*, una célula de *Trichoderma viride*, una célula de *Trichoderma reesei*, una célula de *Trichoderma longibrachiatum*, una célula de *Aspergillus awamori*, una célula de *Aspergillus fumigates*, una célula de *Aspergillus foetidus*, una célula de *Aspergillus nidulans*, una célula de *Aspergillus niger*, una célula de *Aspergillus oryzae*, una célula de *Humicola insolens*, una célula de *Humicola lanuginosa*, una célula de *Rhodococcus opacus*, una célula de *Rhizomucor miehei* o una célula de *Mucor miehei*. En aún otras realizaciones, la célula huésped es una célula de *Streptomyces lividans* o una célula de *Streptomyces murinus*. En aún otras realizaciones, la célula huésped es una célula de *Actinomycetes*. En algunas realizaciones, la célula huésped es una célula de *Saccharomyces cerevisiae*. En otras realizaciones, la célula huésped es una célula de una planta eucariota, alga, cianobacteria, bacteria verde de azufre, bacteria verde no de azufre, bacteria púrpura de azufre, bacteria púrpura no de azufre, extremófila, levadura, hongo, un organismo modificado por ingeniería de los mismos, o un organismo sintético. En algunas realizaciones, la célula huésped depende de la luz o fija carbono. En algunas realizaciones, la célula huésped tiene actividad autótrofa. En algunas realizaciones, la célula huésped tiene actividad fotoautótrofa, tal como en presencia de luz. En algunas realizaciones, la célula huésped es heterótrofa o mixótrofa en ausencia de luz. En determinadas realizaciones, la célula huésped es una célula de *Arabidopsis thaliana*, *Panicum virgatum*, *Miscanthus giganteus*, *Zea mays*, *Botryococcus braunii*, *Chlamydomonas reinhardtii*, *Dunaliella salina*, *Synechococcus* Sp. PCC 7002, *Synechococcus* Sp. PCC 7942, *Synechocystis* Sp. PCC 6803, *Thermosynechococcus elongatus* BP-1, *Chlorobium tepidum*, *Chlorojlexus auranticus*, *Chromatium vinosum*, *Rhodospirillum rubrum*, *Rhodobacter capsulatus*, *Rhodopseudomonas palustris*, *Clostridium ljungdahlii*, *Clostridium thermocellum*, *Penicillium chrysogenum*, *Pichia pastoris*, *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Pseudomonas fluorescens* o *Zymomonas mobilis*. En una realización particular, la célula microbiana es de una cianobacteria incluyendo, pero sin limitarse a, *Prochlorococcus*, *Synechococcus*, *Synechocystis*, *Cianothece* y *Nostoc punctiforme*. En otra realización, la célula microbiana es de una especie cianobacteriana específica incluyendo, pero sin limitarse a, *Synechococcus elongatus* PCC7942, *Synechocystis* sp. PCC6803 y *Synechococcus* sp. PCC7001.

Cultivo de células huésped recombinantes y fermentación

Tal como se usa en el presente documento, el término fermentación se refiere ampliamente a la conversión de materiales orgánicos en sustancias diana por células huésped, por ejemplo, la conversión de una fuente de carbono por células huésped recombinantes en ácidos grasos o derivados de los mismos mediante la propagación de un cultivo de las células huésped recombinantes en un medio que comprende la fuente de carbono. Las condiciones permisivas para la producción se refieren a cualesquiera condiciones que permiten que una célula huésped produzca un producto deseado, tal como un aldehído graso o un alcohol graso. De manera similar, la condición o condiciones en las que se expresa la secuencia de polinucleótido de un vector significan cualesquiera condiciones que permiten que una célula huésped sintetice un polipéptido. Las condiciones adecuadas incluyen, por ejemplo,

condiciones de fermentación. Las condiciones de fermentación pueden incluir muchos parámetros incluyendo, pero sin limitarse a, intervalos de temperatura, niveles de aireación, tasas de alimentación y composición de medios. Cada una de estas condiciones, individualmente y en combinación, permiten que crezca la célula huésped. La fermentación puede ser aerobia, anaerobia, o variaciones de las mismas (tales como microaerobias). Los medios de cultivo a modo de ejemplo incluyen caldos o geles. Generalmente, el medio incluye una fuente de carbono que puede metabolizarse por una célula huésped directamente. Además, pueden usarse enzimas en el medio para facilitar la movilización (por ejemplo, la despolimerización de almidón o celulosa para dar azúcares fermentables) y el metabolismo posterior de la fuente de carbono.

Para la producción a pequeña escala, las células huésped modificadas por ingeniería pueden hacerse crecer en lotes de, por ejemplo, aproximadamente 100 μ l, 200 μ l, 300 μ l, 400 μ l, 500 μ l, 1 ml, 5 ml, 10 ml, 15 ml, 25 ml, 50 ml, 75 ml, 100 ml, 500 ml, 1 l, 2 l, 5 l, o 10 l; fermentarse; e inducirse para expresar una secuencia de polinucleótido deseada, tal como una secuencia de polinucleótido que codifica para una ACP y/o polipéptido de biosíntesis. Para la producción a gran escala, las células huésped modificadas por ingeniería pueden hacerse crecer en lotes de aproximadamente 10 l, 100 l, 1000 l, 10.000 l, 100.000 l y 1.000.000 l o más; fermentarse; e inducirse para expresar una secuencia de polinucleótido deseada. Alternativamente, puede llevarse a cabo fermentación semicontinua a gran escala. Las composiciones de derivados de ácido graso descritas en el presente documento se encuentran en el entorno extracelular del cultivo de célula huésped recombinante y pueden aislarse fácilmente del medio de cultivo. Un derivado de ácido graso tal como un aldehído graso o alcohol graso puede secretarlo la célula huésped recombinante, transportarse al entorno extracelular o transferirse de manera pasiva al entorno extracelular del cultivo de célula huésped recombinante. El derivado de ácido graso se aísla de un cultivo de célula huésped recombinante usando métodos de rutina conocidos en la técnica.

Productos derivados de células huésped recombinantes

Tal como se usa en el presente documento, la fracción de carbono moderno o fM tiene el mismo significado definido por los materiales de referencia patrón del Instituto Nacional de Normas y Tecnología (NIST) (SRMs4990B y 4990C, conocidos como patrones de ácidos oxálicos HOxI y HOxII, respectivamente). La definición fundamental se refiere a 0,95 veces la razón isotópica $^{14}\text{C}/^{12}\text{C}$ HOxI (a la que se hace referencia como AD 1950). Esto es aproximadamente equivalente a la madera previa a la Revolución Industrial corregida para la desintegración. Para la biosfera viva actual (material vegetal), fM es de aproximadamente 1,1. Los bioproductos (por ejemplo, los derivados de ácido graso incluyendo aldehídos grasos y/o alcoholes grasos producidos según la presente divulgación) incluyen compuestos orgánicos producidos biológicamente. En particular, los derivados de ácido graso producidos usando la ruta de biosíntesis de ácidos grasos en el presente documento, no se han producido a partir de fuentes renovables y, como tal, son nuevas composiciones de materia. Estos nuevos bioproductos pueden distinguirse de compuestos orgánicos derivados de carbono petroquímico basándose en la obtención de huella isotópica de carbono o la datación mediante ^{14}C . adicionalmente, la fuente específica de carbono de fuente biológica (por ejemplo, glucosa frente a glicerol) puede determinarse mediante obtención de huella isotópica de carbono dual (véase, por ejemplo, la patente estadounidense n.º 7.169.588). La capacidad para distinguir bioproductos de compuestos orgánicos basados en el petróleo es beneficiosa en la trazabilidad de estos materiales en el comercio. Por ejemplo, pueden distinguirse compuestos orgánicos o químicos que incluyen perfiles isotópicos de carbono tanto de base biológica como de base en el petróleo de compuestos orgánicos y químicos que se componen sólo de materiales basados en el petróleo. Así, los bioproductos en el presente documento pueden someterse a seguimiento o trazabilidad en el comercio basándose en su perfil isotópico de carbono único. Los bioproductos pueden distinguirse de compuestos orgánicos basados en el petróleo comparando la razón isotópica de carbono estable ($^{13}\text{C}/^{12}\text{C}$) en cada muestra. La razón $^{13}\text{C}/^{12}\text{C}$ en un bioproducto dado es consecuencia de la razón $^{13}\text{C}/^{12}\text{C}$ en el dióxido de carbono atmosférico en el momento en que se fija el dióxido de carbono. También refleja la ruta metabólica precisa. También se producen variaciones regionales. El petróleo, las plantas C3 (las de hoja caduca), las plantas C4 (las hierbas) y carbonatos marinos muestran todas diferencias significativas en $^{13}\text{C}/^{12}\text{C}$ y los valores de $\delta^{13}\text{C}$ correspondientes. Además, la materia lipídica de plantas C3 y C4 se analiza de manera diferente que en materiales derivados de los componentes de hidrato de carbono de las mismas plantas como consecuencia de la ruta metabólica. Dentro de la precisión de medición, ^{13}C muestra grandes variaciones debido a efectos de fraccionamiento isotópico, siendo el más significativo de ellos para bioproductos el mecanismo de fotosíntesis. La principal causa de las diferencias en la razón isotópica de carbono en plantas se asocia estrechamente con diferencias en la ruta de metabolismo de carbono por fotosíntesis en las plantas, particularmente la reacción que se produce para la carboxilación primaria (es decir, la fijación inicial de CO_2 atmosférico). Dos grandes clases de vegetación son las que incorporan el ciclo de fotosíntesis C3 (o de Calvin-Benson) y las que incorporan el ciclo de fotosíntesis C4 (o de Hatch-Slack). En plantas C3, la fijación primaria de CO_2 o reacción de carboxilación implica la enzima ribulosa-1,5-difosfato carboxilasa, y el primer producto estable es un compuesto de 3 carbonos. Las plantas C3, tales como frondosas y coníferas, son predominantes en las zonas de climas templados. En plantas C4, una reacción de carboxilación adicional que implica otra enzima, fosfoenol-piruvato carboxilasa, es la reacción de carboxilación primaria. El primer compuesto carbonado estable es un ácido de 4 carbonos que se descarboxila posteriormente. El CO_2 así liberado vuelve a fijarse mediante el ciclo C3. Ejemplos de plantas C4 son hierbas tropicales, maíz y caña de azúcar. Las plantas tanto C4 como C3 muestran un intervalo de razones isotópicas $^{13}\text{C}/^{12}\text{C}$, pero valores típicos son de aproximadamente -7 a aproximadamente -13 por cada mil para plantas C4 y de aproximadamente -19 a aproximadamente -27 por cada mil para plantas C3 (véase, por ejemplo, Stuiver *et al.* (1977) Radiocarbon 19:355). El carbón y el petróleo se

encuentran generalmente en este último intervalo. La escala de medición de ^{13}C se definió originariamente por un cero establecido por la piedra caliza belemnita de Pee Dee (PDB), facilitándose los valores en partes por mil desviaciones con respecto a este material. Los valores de $\delta^{13}\text{C}$ se expresan en partes por mil (por mil), abreviado como, ‰, y se calculan de la siguiente manera:

$$5 \quad \delta^{13}\text{C} (\text{‰}) = \left[\left(\frac{^{13}\text{C}}{^{12}\text{C}} \right)_{\text{muestra}} - \left(\frac{^{13}\text{C}}{^{12}\text{C}} \right)_{\text{patrón}} \right] / \left(\frac{^{13}\text{C}}{^{12}\text{C}} \right)_{\text{patrón}} \times 1000$$

Puesto que el material de referencia (RM) de PDB se ha agotado, una serie de RM alternativos se han desarrollado en cooperación con IAEA, USGS, NIST, y otros laboratorios de isótopos internacionales seleccionados. Las notaciones para las desviaciones por mil con respecto a PDB es $\delta^{13}\text{C}$. Se realizan las mediciones con CO_2 mediante espectrometría de masas de razón estable de alta precisión (IRMS) con iones moleculares de masas 44, 45 y 46. Las composiciones descritas en el presente documento incluyen bioproductos producidos mediante cualquiera de los métodos descritos en el presente documento, incluyendo, por ejemplo, productos derivados de ácido graso. Específicamente, el bioproducto puede tener un $\delta^{13}\text{C}$ de aproximadamente -28 o mayor, aproximadamente -27 o mayor, -20 o mayor, -18 o mayor, -15 o mayor, -13 o mayor, -10 o mayor, o -8 o mayor. Por ejemplo, el bioproducto puede tener un $\delta^{13}\text{C}$ de aproximadamente -30 a aproximadamente -15, de aproximadamente -27 a aproximadamente -19, de aproximadamente -25 a aproximadamente -21, de aproximadamente -15 a aproximadamente -5, de aproximadamente -13 a aproximadamente -7 o de aproximadamente -13 a aproximadamente -10. En otros casos, el bioproducto puede tener un $\delta^{13}\text{C}$ de aproximadamente -10, -11, -12 o -12,3. Los bioproductos producidos según la divulgación en el presente documento, también pueden distinguirse de compuestos orgánicos basados en el petróleo comparando la cantidad de ^{14}C en cada compuesto. Dado que el ^{14}C tiene una semivida nuclear de 5730 años, los combustibles basados en el petróleo que contienen carbono más antiguo pueden distinguirse de bioproductos que contienen carbono más nuevo (véanse, por ejemplo, Currie, Source Apportionment of Atmospheric Particles, Characterization of Environmental Particles, J. Buffle y H. P. van Leeuwen, eds., 1 de vol. I de IUPAC Environmental Analytical Chemical Series (Lewis Publishers, Inc.) 3-74, (1992)). La suposición básica en la datación por radiocarbono es que la constancia de la concentración de ^{14}C en la atmósfera conduce a la constancia de ^{14}C en organismos vivos. Sin embargo, debido a las pruebas nucleares atmosféricas desde 1950 y el quemado de combustibles fósiles desde 1850, el ^{14}C ha adquirido una segunda característica temporal geoquímica. Su concentración en el CO_2 atmosférico, y así en la biosfera viva, se duplicó aproximadamente en el apogeo de las pruebas nucleares, a mediados de los años 1960. Desde entonces se ha vuelto gradualmente a la tasa isotópica ($^{14}\text{C}/^{12}\text{C}$) inicial cosmogénica (atmosférica) de estado estacionario de aproximadamente $1,2 \times 10^{-12}$, con una "semivida" de relajación aproximada de 7-10 años. Esta última semivida no ha de tomarse literalmente; más bien, debe usarse la función de entrada/desintegración nuclear atmosférica detallada para rastrear la variación de ^{14}C atmosférico y de la biosfera desde el comienzo de la era nuclear. Es esta última característica temporal de ^{14}C de la biosfera la que mantiene la promesa de datación anual de carbono reciente de la biosfera. Puede medirse el ^{14}C mediante espectrometría de masas con acelerador (AMS), facilitándose los resultados en unidades de fracción de carbono moderno (fM). fM se define por los materiales de referencia patrón 4990B y 4990C del Instituto Nacional de Normas y Tecnología (NIST). Tal como se usa en el presente documento, la fracción de carbono moderno o fM tiene el mismo significado definido por los materiales de referencia patrón del Instituto Nacional de Normas y Tecnología (NIST) (SRMs4990B y 4990C, conocidos como patrones de ácidos oxálicos HOxI y HOxII, respectivamente). La definición fundamental se refiere a 0,95 veces la razón isotópica $^{14}\text{C}/^{12}\text{C}$ HOxI (a la que se hace referencia como AD 1950). Esto es aproximadamente equivalente a la madera previa a la Revolución Industrial corregida para la desintegración. Para la biosfera viva actual (material vegetal), fM es de aproximadamente 1,1. Las composiciones descritas en el presente documento incluyen bioproductos que pueden tener un fM ^{14}C de al menos aproximadamente 1. Por ejemplo, el bioproducto de la divulgación puede tener un fM ^{14}C de al menos aproximadamente 1,01, un fM ^{14}C de aproximadamente 1 a aproximadamente 1,5, un fM ^{14}C de aproximadamente 1,04 a aproximadamente 1,18, o un fM ^{14}C de aproximadamente 1,111 a aproximadamente 1,124.

Otra medición de ^{14}C se conoce como el tanto por ciento de carbono moderno (pMC). Para un arqueólogo o geólogo que usa dataciones por ^{14}C , AD 1950 es igual a una edad de cero años. Esto también representa 100 pMC. El carbono de bombas en la atmósfera alcanzó casi el doble del nivel normal en 1963 en el apogeo de las armas termonucleares. Su distribución dentro de la atmósfera se ha aproximado desde su aparición, mostrando valores que son mayores de 100 pMC para plantas y animales que viven desde AD 1950. Ha disminuido gradualmente a lo largo del tiempo, siendo el valor de hoy en día próximo a 107,5 pMC. Esto significa que un material de biomasa reciente, tal como maíz, proporcionaría una firma de ^{14}C próxima a 107,5 pMC. Los compuestos basados en el petróleo tendrán un valor de pMC de cero. La combinación de carbono fósil con carbono de hoy en día dará como resultado una dilución del contenido de pMC de hoy en día. Suponiendo que 107,5 pMC representa el contenido de ^{14}C de materiales de biomasa de hoy en día y 0 pMC representa el contenido de ^{14}C de productos basados en el petróleo, el valor de pMC medido para ese material reflejará las proporciones de los dos tipos de componente. Por ejemplo, un material derivado al 100% de soja de hoy en día proporcionará una firma de radiocarbono próxima a 107,5 pMC. Si ese material se diluyera al 50% con productos basados en el petróleo, proporcionaría una firma de radiocarbono de aproximadamente 54 pMC. Se deriva un contenido de carbono de base biológica asignando el 100% igual a 107,5 pMC y el 0% igual a 0 pMC. Por ejemplo, una muestra que mide 99 pMC proporcionará un contenido de carbono de base biológica equivalente del 93%. Este valor se denomina el resultado de carbono de base biológica medio y supone que todos los componentes dentro del material analizado se originan o bien a partir de material biológico de hoy en día o bien a partir de material basado en el petróleo. Un bioproducto que comprende uno o más

derivados de ácido graso tal como se describe en el presente documento puede tener un pMC de al menos aproximadamente 50, 60, 70, 75, 80, 85, 90, 95, 96, 97, 98, 99 ó 100. En otros casos, un derivado de ácido graso descritos en el presente documento puede tener un pMC de entre aproximadamente 50 y aproximadamente 100; entre aproximadamente 60 y aproximadamente 100; entre aproximadamente 70 y aproximadamente 100; entre aproximadamente 80 y aproximadamente 100; entre aproximadamente 85 y aproximadamente 100; entre aproximadamente 87 y aproximadamente 98; o entre aproximadamente 90 y aproximadamente 95. En aún otros casos, un derivado de ácido graso descrito en el presente documento puede tener un pMC de aproximadamente 90, 91, 92, 93, 94 ó 94,2.

Composiciones de aldehído graso y alcohol graso y su uso

10 Se usan aldehídos para producir muchos productos químicos especializados. Por ejemplo, se usan aldehídos para producir polímeros, resinas (por ejemplo, resina BAKELITE), colorantes, saborizantes, plastificantes, perfumes, productos farmacéuticos, y otros productos químicos, algunos de los cuales pueden usarse como disolventes, conservantes o desinfectantes. Además, determinados compuestos naturales y sintéticos, tales como vitaminas y hormonas, son aldehídos, y muchos azúcares contienen grupos aldehído. Los aldehídos grasos pueden convertirse
15 en alcoholes grasos mediante reducción química o enzimática. Los alcoholes grasos también tienen múltiples usos comerciales. Las ventas anuales a nivel mundial de alcoholes grasos y sus derivados son superiores a mil millones de USD. Los alcoholes grasos de cadena más corta se usan en las industrias cosmética y alimentaria como emulsionantes, emolientes y espesantes. Debido a su naturaleza anfífila, los alcoholes grasos se comportan como tensioactivos no iónicos, que son útiles en productos para el cuidado personal y de uso doméstico, tales como, por
20 ejemplo, detergentes. Además, se usan alcoholes grasos en ceras, gomas, resinas, pomadas y lociones farmacéuticas, aditivos de aceites lubricantes, agentes textiles antiestáticos y de acabado, plastificantes, cosméticos, disolventes industriales y disolventes para grasas.

La divulgación también proporciona una composición de tensioactivo o una composición de detergente que incluye un alcohol graso producido mediante cualquiera de los métodos descritos en el presente documento. Un experto
25 habitual en la técnica apreciará que, dependiendo del fin previsto de la composición de tensioactivo o detergente, pueden producirse y usarse diferentes alcoholes grasos. Por ejemplo, cuando los alcoholes grasos descritos en el presente documento se usan como materia prima para la producción de tensioactivos o detergentes, un experto habitual en la técnica apreciará que las características de la materia prima de alcohol graso afectarán a las características de la composición de tensioactivo o detergente producida. Así, las características de la composición
30 de tensioactivo o detergente pueden seleccionarse para producir alcoholes grasos particulares para su uso como materia prima. Una composición de detergente y/o tensioactivo basado en alcohol graso descrita en el presente documento puede mezclarse con otros tensioactivos y/o detergentes bien conocidos en la técnica. En algunas realizaciones, la mezcla puede incluir al menos aproximadamente el 10%, al menos aproximadamente el 15%, al menos aproximadamente el 20%, al menos aproximadamente el 30%, al menos aproximadamente el 40%, al menos
35 aproximadamente el 50%, al menos aproximadamente el 60%, o un intervalo delimitado por dos cualesquiera de los valores anteriores, en peso del alcohol graso. En otros ejemplos, puede producirse una composición de tensioactivo o detergente que incluye al menos aproximadamente el 5%, al menos aproximadamente el 10%, al menos aproximadamente el 20%, al menos aproximadamente el 30%, al menos aproximadamente el 40%, al menos aproximadamente el 50%, al menos aproximadamente el 60%, al menos aproximadamente el 70%, al menos
40 aproximadamente el 80%, al menos aproximadamente el 85%, al menos aproximadamente el 90%, al menos aproximadamente el 95%, o un intervalo delimitado por dos cualesquiera de los valores anteriores, en peso de un alcohol graso que incluye una cadena de carbonos que tiene 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21 ó 22 carbonos de longitud. Tales composiciones de tensioactivo o detergente también pueden incluir al menos un aditivo,
45 tal como una microemulsión o un tensioactivo o un detergente de fuentes no microbianas tales como aceites vegetales o petróleo, que pueden estar presentes en la cantidad de al menos aproximadamente el 5%, al menos aproximadamente el 10%, al menos aproximadamente el 15%, al menos aproximadamente el 20%, al menos aproximadamente el 30%, al menos aproximadamente el 40%, al menos aproximadamente el 50%, al menos aproximadamente el 60%, al menos aproximadamente el 70%, al menos aproximadamente el 80%, al menos
50 aproximadamente el 85%, al menos aproximadamente el 90%, al menos aproximadamente el 95%, o un intervalo delimitado por dos cualesquiera de los valores anteriores, en peso del alcohol graso.

Ejemplos

Los siguientes ejemplos específicos pretenden ilustrar la divulgación y no deben interpretarse como limitativos del alcance de las reivindicaciones.

Protocolos y métodos

55 *Examen de una biblioteca*

Todos los protocolos descritos en el presente documento se basan en un sistema de 2 ml de bloque maestro de placa de 96 pocillos (Greiner Bio-One, Monroe, NC o Corning, Ámsterdam, Países Bajos) para cultivos de crecimiento, y placas (Costar, Inc.) para extraer especies de ácido graso del caldo de cultivo. Los protocolos proporcionados a continuación son ejemplos de condiciones de fermentación. Pueden usarse protocolos alternativos

para evaluar la producción de especies de ácido graso.

Protocolo de cultivo a 32°C (4NBT)

Se usaron 20 µl de cultivo LB (de un cultivo LB que se hizo crecer en una placa de 96 pocillos) para inocular 400 µl de medios 2NBT (tabla 2), lo que se incubó luego durante aproximadamente 16 horas con agitación a 32°C. Se usaron 20 µl de la simiente durante la noche para inocular 400 µl de 4NBT con nitrógeno o bien 1 o bien 2 g/l (NBT_1N o NBT_2N). Después de crecimiento a 32°C durante 6 horas, se indujeron los cultivos con IPTG (concentración final de 1 mM) (tabla 2). Entonces se incubaron los cultivos a 32°C con agitación durante 18 horas, después de lo cual se extrajeron siguiendo el protocolo de extracción convencional detallado a continuación.

Tabla 2: Nombres y formulaciones de medios

Nombre de medio	Formulación		
2NBT	1	X	5x disol. salina con NH ₄ Cl
	1	g/l	NH ₄ Cl 100 g/l
	1	mg/l	Tiamina 10 mg/ml
	1	mM	MgSO ₄ 1 M
	0,1	mM	CaCl ₂ 1 M
	20	g/l	glucosa 500 g/l
	1	X	1000x TM2
	10	mg/l	citrato de Fe 10 g/l
	100	µg/ml	espectinomicina 100 mg/ml
4NBT_1N	100	mM	BisTris 2 M (pH 7,0)
	1	X	disol. salina con NH ₄ Cl 5x
	1	mg/l	Tiamina 10 mg/ml
	1	mM	MgSO ₄ 1 M
	0,1	mM	CaCl ₂ 1 M
	40	g/l	glucosa 500 g/l
	1	X	1000x TM2
	10	mg/l	citrato de Fe 10 g/l
4NBT_2N	100	µg/ml	espectinomicina 100 mg/ml
	100	mM	BisTris 2 M (pH 7,0)
	1	X	5x disol. salina con NH ₄ Cl
	1	g/l	NH ₄ Cl 100 g/l
	1	mg/l	Tiamina 10 mg/ml
	1	mM	MgSO ₄ 1 M
	0,1	mM	CaCl ₂ 1 M
	40	g/l	glucosa 500 g/l
	1	X	1000x TM2
10	mg/l	citrato de Fe 10 g/l	
100	µg/ml	espectinomicina 100 mg/ml	
100	mM	BisTris 2 M (pH 7,0)	

10 *Protocolo de extracción convencional de especies de ácido graso*

A cada pocillo que iba a extraerse, se le añadieron 40 µl de HCl 1 M, seguido por 300 µl de acetato de butilo (con 500 mg/l de C11-FAME como patrón interno). Entonces se termosellaron las placas de 96 pocillos usando un sellador de placas (calefactor ALPS-300; Abgene, ThermoScientific, Rockford, IL), y se agitaron durante 15 minutos a 2000 rpm usando el mezclador MIXMATE (Eppendorf, Hamburgo, Alemania). Después de la agitación, se centrifugaron las placas durante 10 minutos a 4500 rpm a temperatura ambiente (Allegra X-15R, rotor SX4750A, Beckman Coulter, Brea, CA) para separar las fases acuosa y orgánica. Se transfirieron 50 µl de la fase orgánica a una placa de 96 pocillos (polipropileno, Corning, Ámsterdam, Países Bajos), que posteriormente se termoselló y se almacenó a -20°C hasta que se evaluó mediante CG-FID usando el método FALC_Broth.met. Se llevó a cabo el método FALC_Broth.met de la siguiente manera: se inyectó 1 µl de muestra sobre una columna analítica (DB-1, 10 m x 180 µm x 0,2 µm de grosor de película, disponible de JW 121-101A) en un dispositivo Agilent 7890A CG Ultra (Agilent, Santa Clara, CA) con un detector de ionización de llama (FID). Se configuró el instrumento para detectar y cuantificar alcoholes grasos C₆ a C₁₈. El protocolo detallado anteriormente representa condiciones convencionales, que pueden modificarse según sea necesario para optimizar los resultados analíticos.

Especies de ácido graso - Protocolo de ensayo con rojo Nilo convencional

25 Después de 24 horas de fermentación, se realizó un ensayo con rojo Nilo añadiendo 70 µl de caldo de fermentación a 130 µl de rojo Nilo 1,54 µg/ml en disolución del 84,6% de agua y el 15,4% de acetonitrilo para una concentración

de ensayo final de rojo Nilo 1 µg/ml en una placa Greiner MicroLonFluotrac 200, y se mezcló mediante aspiración y liberación con pipeta. Se midieron las unidades relativas de fluorescencia a excitación de 540 nm y emisión de 630 nm usando la unidad SpectraMax M2.

Construcción de bibliotecas de propensión a error

5 Se usaron técnicas convencionales que conocen los expertos en la técnica para preparar bibliotecas de propensión a error. En un ejemplo, se preparó la estructura principal de vector usando endonucleasas de restricción en el vector, mientras que se generó la creación de diversidad en el inserto de ADN mediante amplificación por PCR a partir de un molde de ADN en condiciones que favorecerían la incorporación de nucleótidos de apareamiento incorrecto. En un enfoque, se realizó la clonación de la estructura principal de vector y un inserto de ADN usando el sistema de clonación INFUSION (Clontech Laboratories, Inc., Mountain Vista, CA), según el protocolo del fabricante.

Construcción de bibliotecas de saturación

15 Se usaron técnicas convencionales que conocen los expertos en la técnica para preparar bibliotecas de saturación. En un ejemplo, se preparó la estructura principal de vector usando endonucleasas de restricción en el vector, mientras que se generó la creación de diversidad en el inserto de ADN usando cebadores degenerados. En un enfoque, se realizó la clonación de la estructura principal de vector y un inserto de ADN usando el sistema de clonación INFUSION (Clontech Laboratories, Inc., Mountain Vista, CA) según el protocolo del fabricante.

Construcción de bibliotecas de combinación

20 Se combinaron mutaciones identificadas como beneficiosas para proporcionar variantes de AAR con mejoras adicionales en la producción de especies de alcohol graso. Se usaron técnicas convencionales que conocen los expertos en la técnica para preparar las bibliotecas de combinación. En un ejemplo, se preparó la estructura principal de vector usando endonucleasas de restricción en el vector, mientras que se generó la creación de diversidad en el inserto de ADN usando cebadores para introducir las mutaciones deseadas. Tal como se describió anteriormente, en un enfoque, se realizó la clonación de la estructura principal de vector y un inserto de ADN usando el sistema de clonación INFUSION (Clontech Laboratories, Inc., Mountain Vista, CA), según el protocolo del fabricante. Pueden generarse bibliotecas de combinación usando el protocolo de PCR de transferencia (tPCR) (Erijman *et al.* (2011) *J. Structural Bio.*175:171-177).

Examen de bibliotecas

30 Una vez que se generó la diversidad de bibliotecas en una biblioteca de saturación o biblioteca de combinación, de propensión a error, se examinó usando uno de los métodos descritos anteriormente. Se identificaron dos tipos de coincidencias: (1) aumento de la cantidad de alcohol graso (título de FALC); y/o (2) aumento de la cantidad de FALC de cadena media tal como dodecanol (C12) o tetradecanol (C14). También se identificaron hexadecanol (C16) y octadecanol (C18). Se identificaron las mutaciones en las variantes de AAR dentro de cada coincidencia mediante secuenciación usando técnicas convencionales empleadas de manera rutinaria por los expertos en la técnica. Las tablas 4, 5 y 6 enumeran las mutaciones (coincidencias) identificadas como beneficiosas en bibliotecas de saturación y bibliotecas de combinación.

EJEMPLO 1: Producción mejorada de alcoholes grasos usando AAR_7942 mediante un aumento del flujo mediado por la proteína transportadora de acilo (ACP) a través de la ruta de síntesis de ácidos grasos

40 Cuando se expresan en *E. coli* enzimas de rutas terminales de fuentes distintas de *E. coli* como el huésped heterólogo para convertir las acil graso-ACP en productos, pueden existir limitaciones en el reconocimiento, la afinidad y/o el recambio de la enzima de ruta recombinante hacia las acil graso-ACP de *E. coli*. Aunque las proteínas ACP se conservan en cierta medida en todos los organismos, su secuencia primaria puede diferir incluso dentro de una especie dada. Para someter a prueba esta hipótesis, se clonaron los genes *acp* de varias cianobacterias en el sentido de 3' de la acil-ACP reductasa de *Synechococcus elongatus* PCC7942 (AAR_7942) presente en el plásmido pLS9-185, que es un derivado de pCL1920 (3-5 copias/célula). Además, se clonó el gen *sfp* (n.º de registro X63158; SEQ ID NO: 11) de *Bacillus subtilis*, que codifica para una fosfopanteteinil transferasa con amplia especificidad de sustrato, en el sentido de 3' de los genes *acp* respectivos. Esta enzima está implicada en la conversión de la apo-ACP inactiva en la holo-ACP activa. Los plásmidos construidos se describen en la tabla 3.

Tabla 3: Plásmidos que coexpresan ACP de cianobacterias con y sin *sfp* de *B. subtilis* en el sentido de 3' con respecto a AAR de *S. elongatus* PCC7942

Plásmido base	Fuente de ACP	ACP - SEQ ID N.º (NA/AA*)	Sin <i>sfp</i>	Con <i>sfp</i>
pLS9-185	<i>Synechococcus elongatus</i> 7942	7/8	pDS168	pDS168S
pLS9-185	<i>Synechocystis</i> sp. 6803	3/4	pDS169	no disponible
pLS9-185	<i>Prochlorococcus marinus</i> MED4	5/6	pDS170	SDS170S

pLS9-185	<i>Nostoc punctiforme</i> 73102	1/2	pDS171	pDS171S
pLS9-185	<i>Nostoc sp.</i> 7120	9/10	pDS172	pDS172S

*NA = secuencia de ácido nucleico; AA = secuencia de aminoácidos/secuencia de polipéptido

Se clonaron todos los genes *acp* con un RBS sintético en el sitio EcoRI inmediatamente en el sentido de 3' del gen *aar* en pLS9-185 usando la tecnología INFUSION. Se reconstruyó el sitio EcoRI en el sentido de 3' del gene *acp*. De manera similar, se clonó mediante la tecnología INFUSION el gen *sfp* de *B. subtilis* en este sitio EcoRI junto con un con RBS sintético. Se transformaron todos los plásmidos en MG1655 DV2 de *E. coli*. El control durante estos experimentos fue la expresión de AAR sola (pLS9-185).

Los resultados de experimentos de fermentación en matraz de agitación convencionales se muestran en la figura 5. Se observaron mejoras significativas en los títulos de alcoholes grasos en cepas que contenían los plásmidos pDS171S, pDS172S, pDS168 y pDS169, demostrando que la sobreexpresión de ACP puede ser beneficiosa durante la producción de alcoholes grasos. Aunque no desea limitarse por la teoría, se plantea la hipótesis de que la sobreexpresión de ACP puede ser beneficiosa durante la producción de alcoholes grasos ayudando en el reconocimiento, la afinidad y/o el recambio de acil-ACP mediante la enzima de la ruta terminal heteróloga (véase la tabla 3 (anteriormente) para la fuente de las ACP y la presencia o ausencia de *sfp*.)

EJEMPLO 2: Producción mejorada de alcoholes grasos usando AAR_7942 mediante el aumento del flujo mediado por acetil-CoA carboxilasa (ACC) a través de la síntesis de ácidos grasos

Los principales precursores durante la biosíntesis de ácidos grasos son malonil-CoA y acetil-CoA. Se ha sugerido que estos precursores limitan la tasa de la biosíntesis de ácidos grasos en *E. coli*. En este ejemplo, se expresó un operón sintético de acetil-CoA carboxilasa, acc [accDCAB+birA de *Corynebacterium glutamicum*; SEQ ID NO: 45 ó 46, 48 y 50 también denominado D+] junto a acil-ACP reductasa (AAR) de *Synechococcus elongatus* PCC7942 (AAR_7942). Se clonó el operón accD+ en el sentido de 3' del gen de AAR_7942 en el plásmido pLS9-185. Se transformaron el plásmido resultante y el plásmido de control pLS9-185 en DV2 de *E. coli*. Se evaluaron las cepas durante la producción de alcoholes grasos en un protocolo en matraz de agitación convencional. Tal como se muestra en la figura 6, la coexpresión del operón acc sintético de *Corynebacterium glutamicum* condujo al aumento de la producción de alcoholes grasos.

EJEMPLO 3: Biblioteca de propensión a error, bibliotecas de combinación y de saturación limitada preparadas usando AAR_7942 como molde

A. Biblioteca de propensión a error

Se construyó una biblioteca de propensión a error de la acil-ACP reductasa de *Synechococcus elongatus* PCC7942 (AAR_7942) y se examinó para detectar variantes que mostrasen mejoras con respecto a la AAR_7942 silvestre. El plásmido usado para producir la biblioteca de propensión a error se designó como "pDS171S" (véase la tabla 3). Se examinó la biblioteca de propensión a error usando uno de los protocolos convencionales descritos anteriormente. Las mejoras se clasificaron como o bien que mejoran el título o bien que aumentan la fracción de alcoholes grasos C₁₀-C₁₄ producidos sin afectar significativamente al título (resultados no mostrados).

B. Bibliotecas de combinación y de saturación limitada

Se usaron técnicas convencionales que conocen los expertos en la técnica para preparar bibliotecas de combinación y bibliotecas de saturación basándose en las posiciones 17, 18 y 19. Las mutaciones sometidas a prueba en las bibliotecas de combinación y bibliotecas de saturación (tablas 4A y 4B) se identificaron originariamente en la biblioteca de propensión a error de AAR_7942 descrita anteriormente. Los plásmidos, las cepas y los protocolos de examen usados fueron los mismos descritos en el ejemplo 1. Los resultados del examen de la biblioteca de propensión a error se muestran en las tablas 4A y 4B. La tabla 4A muestra mutaciones de AAR_7942 que condujeron a un aumento de los títulos de alcoholes grasos y la tabla 4B muestra mutaciones que condujeron a un aumento de la fracción de alcohol graso C₁₄ sin afectar significativamente al título global de alcoholes grasos.

Tabla 4A: Mutaciones de bibliotecas de combinación y de saturación limitada de AAR_7942 correlacionadas con un título mejorado de alcoholes grasos

Mutaciones	Combinación de bibliotecas α			
	Título de FALC (total*)	FALC C ₁₄ (% de fracción)	FALC normalizado (% con respecto al control)	C ₁₄ normalizado (% con respecto al control)
S18T, C63Y, S96T, L177H, A281V	2473	48%	251%	294%
S18T, C63Y, S96T, L177H, A281V E284K	2202	50%	200%	229%
S18T, C63Y, S96T, L177H	2088	43%	189%	197%
S18T, C63Y	1798	39%	182%	241%

ES 2 600 354 T3

C63H, S967T, L177H, A281V	2045	53%	180%	237%
S18T, C63Y, A281V, E283K	1696	45%	172%	278%
S18T, C63Y	1640	43%	166%	260%
S18T, L177H, A281V, E283K	1618	44%	164%	270%
S18T, L177H, A281V, E283K	1617	43%	164%	264%
S18T, S96T, L177H, E283K	1814	43%	160%	194%
S18T, C63H, L177H	1811	39%	160%	173%
S18T, L177H, A281V, E283K	1503	39%	152%	237%
L65F, S96T, A281V, E283K	1638	43%	144%	191%
C63Y, A281V, A282T	1622	38%	143%	168%
L11F, C63Y, S96T, L177H, A281V A282T, E283K	1579	41%	143%	186%
L65F, S96T, L177H, A281V, E283K	1585	42%	140%	189%
G22S, C63H, S96T, L177H, E283K	1229	39%	125%	237%
C63H, L65F, S96T, A281V	1178	41%	119%	252%
S18T, L65F, S96T, L177H, A281V	1136	33%	115%	203%
C63R, L177H, A281V, E283K	1275	44%	112%	196%

Bibliotecas de saturación				
Mutaciones	Título de FALC (total*)	FALC C ₁₄ (% de fracción)	FALC normalizado (% con respecto al control)	C ₁₄ normalizado (% con respecto al control)
S18M	2809	40%	243%	212%
S18F	2564	54%	222%	285%
S18Y	2550	45%	221%	236%
S18W	2530	55%	227%	314%
S18M	2495	41%	224%	231%
S18Y	2039	44%	188%	252%
S18Y	1997	44%	179%	251%
S18T	1729	31%	155%	177%
S18T	1690	31%	152%	176%
S18T	1683	31%	151%	176%
S18T	1599	30%	143%	172%
S18L	1568	48%	141%	271%
S18C	1381	28%	120%	149%
S18C	1372	26%	123%	147%
S18C	1368	28%	118%	148%
S18L	1359	47%	122%	266%
S18C	1355	26%	121%	148%
S18V	1344	30%	121%	170%
S18C	1340	26%	120%	147%
S18L	1326	47%	119%	268%
S18L	1315	47%	118%	266%
S18C	1313	26%	118%	149%
S18V	1306	30%	121%	170%
S18C	1305	26%	117%	148%
S18L	1302	47%	117%	267%
V17L	1188	23%	110%	130%
V17L	1185	33%	103%	173%
V17L	1175	32%	102%	171%
	Combinación de bibliotecas β			
	Título de FALC (total*)	FALC C ₁₄ (% de fracción)	FALC normalizado (% con respecto al control)	C ₁₄ normalizado (% con respecto al control)
S18W, S96T, E283K, R286Q, A324V	3437	49%	246%	255%
S18W	3286	53%	235%	275%
S18W, S96T, E283K	3199	46%	229%	237%
S18W, C63Y, S96T	3186	46%	228%	241%
S18W, C68Y, E2834K	2949	50%	211%	260%
C63Y, E283K	2680	36%	192%	189%
S18W, C63Y	2675	54%	191%	280%
C63Y, S96T	2583	38%	185%	196%
C63Y, E283K, Q316K	2420	36%	173%	188%

ES 2 600 354 T3

S96T	1605	23%	115%	121%
E283K	1532	25%	110%	131%

Tabla 4B: Mutaciones de bibliotecas de combinación y de saturación limitada de AAR_7942 correlacionadas con un aumento de la fracción de alcohol graso C14

	Combinación de bibliotecas α			
	Título de FALC (total*)	FALC C ₁₄ (% de fracción)	FALC normalizado (% con respecto al control)	C ₁₄ normalizado (% con respecto al control)
C63H, S96T, L177H, A281V	2045,36	53%	180%	237%
G22S, C63H, F64V, S96T, L177H A281V	152,47	51%	14%	233%
S18T, C63H, S96T, L177H, A281V E283K	2202,46	50%	200%	229%
S18T, C63Y, S96T, L177H, A281V	2472,80	48%	251%	294%
S18T, C63Y, A281V, E283K	1695,59	45%	172%	278%
S18T, L177H, A281V, E283K	1618,20	44%	164%	270%
C63R, L177H, A281V, E283K	1275,38	44%	112%	196%
S18T, C63H, L65F, S96T, L177F A281V, E283K	912,50	44%	83%	199%
S18T, S96T, L177H, E283K	1814,34	43%	160%	194%
S18T, L179H, A281V, E283K	1616,51	43%	164%	264%
S18T, C63H, S96T, L177H	2088,12	43%	189%	197%
L65F, S96T, A281V, E283K	1638,28	43%	144%	191%
C63H, L65F, A281V, E283K	632,66	43%	56%	190%
S18T, C63Y	1640,33	43%	166%	260%
L65F, S96T, L177H, A281V, E283K	1585,32	42%	140%	189%
C63Y, L65F, L177H	639,46	41%	56%	184%
C63H, L65F, S96T, A281V	1177,76	41%	119%	252%
L11F, C63Y, S96T, L177H, A281V A282T, E283K	1578,54	41%	143%	186%
S18T, C63Y	1797,99	39%	182%	241%
S18T, L178H, A282V, E284K	1502,59	39%	152%	237%
G23S, C63H, S96T, L177H, E283K	1229,34	39%	125%	237%
S18T, C63H, L177H	1811,40	39%	160%	173%
C63Y, A281V, A282T	1622,23	38%	143%	168%
S18T, C63Y, L65F, S96T, L177H A281V, E283K, M284I	454,42	35%	41%	162%
S18T, L65F, S96T, L177H, A281V	1135,95	33%	115%	203%
L65F, L178H, G180S, E283K	151,40	33%	13%	146%
L66F, L177H, G180S, E283K	143,83	31%	13%	138%
S18T, C63H, L65F, L177H	103,21	30%	9%	133%
	Bibliotecas de saturación			
	Título de FALC (total*)	FALC C ₁₄ (% de fracción)	FALC normalizado (% con respecto al control)	C ₁₄ normalizado (% con respecto al control)
S18W	2530	55%	227%	314%
S18F	2564	54%	222%	285%
V17N	571	49%	49%	260%
S18L	1568	48%	141%	271%
S18L	1024	47%	92%	268%
S18L	1326	47%	119%	268%
S18L	1302	47%	117%	267%
S18L	1106	47%	99%	267%
S18L	1315	47%	118%	266%
S18L	1359	47%	122%	266%
S18Y	2550	45%	221%	236%
S18Y	2039	44%	188%	252%
S18Y	1997	44%	179%	251%
S18M	2495	41%	224%	231%
S18M	2809	40%	243%	212%
V17L	1127	33%	97%	175%

ES 2 600 354 T3

V17L	1185	33%	103%	173%
V17L	1091	32%	94%	171%
V17L	1175	32%	102%	171%
S18T	1729	31%	155%	177%
S18T	1690	31%	152%	176%
S18T	1683	31%	151%	176%
S18T	1599	30%	143%	172%
S18V	1344	30%	121%	170%
S18V	1306	30%	121%	170%
S18V	1015	30%	91%	169%
S18C	1381	28%	120%	149%
S18C	1368	28%	118%	148%
R19H, R58S	873	27%	81%	152%
V17C	927	27%	80%	140%
V17C	882	26%	76%	139%
S18C	1313	26%	118%	149%
S18C	1355	26%	121%	148%
S18C	1305	26%	117%	148%
S18C	1372	26%	123%	147%
S18C	1340	26%	120%	147%
V17W	849	25%	73%	134%
V17W	815	25%	71%	133%
V17W	831	25%	72%	132%
V17W	840	25%	73%	131%
R19V	844	24%	78%	139%
R19V	813	24%	75%	136%
R19V	821	24%	76%	135%
R19V	862	24%	80%	135%
R19V	814	24%	75%	135%
R19K, V117L	1188	23%	110%	130%
R19S	803	22%	74%	128%
R19T	735	22%	68%	127%
R19S	783	22%	72%	127%
R19S	764	22%	75%	129%
R19I	862	22%	80%	123%
R19A	901	21%	83%	122%
R19A	705	21%	65%	119%
R19M	937	20%	86%	114%
	Combinación de bibliotecas β			
	Título de FALC (total*)	FALC C14 (% de fracción)	FALC normalizado (% con respecto al control)	C14 normalizado (% con respecto al control)
S18W	3184	57%	228%	296%
S18W, C63Y	2675	54%	191%	280%
S18W, C63Y, S95T	2832	54%	203%	279%
S18W, S96T, E283K	2874	54%	206%	278%
S18W, C63Y, E283K	2915	51%	209%	263%
S18W, S96T, E283K, R286Q, A324V	3437	49%	246%	255%
C63Y, S96T	2583	38%	185%	196%
C63Y, E283K	2176	38%	156%	195%
C63Y, E283K, Q316K	2420	36%	173%	188%
E283K	1474	25%	105%	132%
S96T	1605	23%	115%	121%
	Combinación de bibliotecas 3			
	Título de FALC (total*)	FALC C14 (% de fracción)	FALC normalizado (% con respecto al control)	C14 normalizado (% con respecto al control)
S18W	2853	217%	51%	274%
Y23N	954	72%	35%	191%
Y23N, Q238H	892	68%	35%	190%
G150D, I274N	766	62%	30%	162%
I274N	1359	107%	29%	161%
I274N, A285V	1317	103%	28%	155%

P38T, I274N, A285V	587	46%	26%	143%
Q238H, M291V	1383	109%	22%	119%
M291V	1618	123%	22%	118%
G151D, M291V	1302	105%	21%	114%
T135M, M291V	1265	100%	18%	101%

EJEMPLO 4: Biblioteca de saturación preparada usando AAR (S18W) como molde

Se construyó una biblioteca de saturación completa de una variante de acil-ACP reductasa de *Synechococcus elongatus* PCC7942 (“AAR(S18W)_7942”), y se examinó para detectar variantes que mostrasen mejoras con respecto a AAR (S18W), identificada como variante de AAR significativamente mejorada en la primera tanda de examen (ejemplos 2 y 3). Los criterios de selección fueron un aumento en el título de FALC o un aumento en el tanto por ciento de fracción de C₁₂. Los esfuerzos de modificación por ingeniería se centraron en mitigar la dependencia de AAR de la sobreexpresión de ACP con el fin de producir altos títulos. Aunque no desea limitarse por la teoría, se planteó la hipótesis de que la ventaja observada en cepas que sobreexpresan ACP resultaba de una mayor concentración de productos intermedios de biosíntesis de ácidos grasos disponibles durante la escisión por AAR. Mediante el examen de bibliotecas de saturación y de combinación construidas con AAR en una cepa que carecía de sobreexpresión de ACP (que tenía una menor concentración de productos intermedios de FAS), pudieron seleccionarse variantes con mayores afinidades por los productos intermedios de FAS.

El plásmido que se usó para producir la biblioteca de saturación completa se designó como pAAR-1. Es un derivado de pLS9-185 que alberga el gen de AAR que codifica para la variante (S18W) seguido por el gen de aldehído reductasa, AlrA, de *Acinetobacter baileyi* (SEQ ID NO: 54). Se añadió AlrA para reducir totalmente los productos intermedios de aldehído graso generados por AAR y no reducidos totalmente por las aldehído graso endógenas de *E. coli*. Se examinó la biblioteca de saturación completa en la cepa Shu.002. La cepa Shu.002 es DV2 PT5-ifab138 PT5_ifadR (tabla 7, más adelante). Para ifab138 véase SEQ ID NO: 55 en la tabla 10. Se examinaron las bibliotecas usando uno de los protocolos convencionales descritos anteriormente. Las mejoras se clasificaron como o bien que mejoran el título o bien que aumentan la fracción de alcoholes grasos C₁₂ producidos por la acil-ACP reductasa sin afectar significativamente al título. Los resultados del examen de bibliotecas de saturación se muestran en la tabla 5 a continuación.

Tabla 5: Mutaciones de una biblioteca de saturación completa de AAR (S18W) 7942 correlacionadas con un aumento del título de alcoholes grasos y/o un aumento de la fracción de alcoholes grasos C₁₂

Mutaciones de ARR (además de S18W)	Título de FALC		C12	
	Total*	FIOC**	Fracción*	FIOC**
T148V	747	1,21	7,9%	0,99
A159V	702	1,14	7,0%	0,87
T157V	674	1,10	7,2%	0,90
A135S	803	1,31	7,8%	0,98
A328S	712	1,16	8,1%	1,02
Q191A	780	1,27	7,7%	0,96
A285V, M291V	837	1,36	8,0%	1,00
Q277V	854	1,39	7,9%	0,99
Q155C	626	1,02	8,3%	1,04
E210Y	676	1,10	8,5%	1,06
T120S	623	1,01	8,1%	1,02
T236C	691	1,12	7,6%	0,95
Q335N	763	1,24	6,8%	0,86
C172L	700	1,14	7,7%	0,97
E283S	858	1,39	8,0%	1,00
L209R	837	1,36	7,6%	0,96
I153P	606	0,99	8,0%	1,00
A211W	797	1,30	6,9%	0,86
A324T	909	1,48	4,0%	0,50
W34F	817	1,22	7,5%	0,98
T187V	825	1,23	6,9%	0,91
D24E	737	1,10	7,1%	0,93
T188H	862	1,29	8,6%	1,13
V18A	798	1,19	6,4%	0,84
V18A, L338W	732	1,27	7,6%	0,97
T188V	775	1,34	7,4%	0,95
I168V	731	1,27	8,1%	1,03
W35F	666	1,15	7,5%	0,96
T148V	836	1,45	6,2%	0,80

ES 2 600 354 T3

Q155L	740	1,28	8,5%	1,08
T148C	731	1,26	8,6%	1,10
T148E	694	1,20	8,3%	1,06
A50Q, L337V	800	1,38	7,2%	0,92
R118Q	724	1,25	7,3%	0,94
L31V	901	1,56	6,8%	0,87
A135S	775	1,34	8,7%	1,12
D24Y	890	0,93	19,1%	2,05
C63A	895	0,94	16,1%	1,73
S113K	812	0,85	13,3%	1,43
L31V	645	0,68	15,7%	1,69
E283G	917	0,96	14,3%	1,54
A112R	918	0,96	10,7%	1,15
D43E	1002	1,05	13,0%	1,39
Q116G	894	0,94	14,0%	1,50
S86G	849	0,89	13,4%	1,44

* Promedio de 4 réplicas

** FIOC = veces de aumento con respecto al control AAR (S18W)

Ejemplo 5: Bibliotecas de combinación preparadas usando AAR (S18W) como molde

- 5 Se usaron técnicas convencionales que conocen los expertos en la técnica para preparar bibliotecas de combinación. Las mutaciones sometidas a prueba en bibliotecas de combinación (tablas 6A, 6B y 6C) se identificaron originariamente en la biblioteca de saturación completa (ejemplo 4). Se construyeron las bibliotecas de combinación en el mismo plásmido y se examinaron en la misma cepa tal como se describe en el ejemplo 3. Se examinaron las bibliotecas usando uno de los protocolos convencionales descritos anteriormente. Las mejoras se clasificaron como o bien que mejoran el título o bien que aumentan la fracción de alcoholes grasos C₁₂ producidos por la acil-ACP reductasa sin afectar significativamente al título. Los resultados del examen de bibliotecas de combinación de AAR se muestran en las tablas 6A, 6B y 6C a continuación.
- 10

Tabla 6A: Mutaciones de la 1ª biblioteca de combinación de AAR(S18W)_7942 correlacionadas con un aumento del título de alcoholes grasos

Mutaciones de ARR (además de S18W)	Título* de FALC		C12	
	mg/l*	FIOC**	Fracción*	FIOC**
A281L	1478	1,29	14,1%	1,11
A281Y	1527	1,28	12,1%	0,96
T154A, A281L	1550	1,28	14,3%	1,13
T154A, A281Y	1384	1,19	13,3%	1,05
C63G, A281F	1472	1,24	12,9%	1,02
C63G, A281L	1432	1,19	18,2%	1,44
N83G, A281Y	1428	1,23	10,7%	0,84
C63G, A281Y	1419	1,21	15,6%	1,24
S10G, A281L	1419	1,20	12,6%	1,00
S113K, Q155G	1305	1,15	12,3%	0,97
D43E, A50Q, A281L	1517	1,32	9,1%	0,72
M21L, N83G, A281Y	1513	1,29	10,2%	0,81
M21L, S113K, A281L	1392	1,20	20,4%	1,61
C63G, T154A, A281F	1369	1,24	12,4%	0,98
C63G, N83G, A281Y	1330	1,32	14,6%	1,16
Q155G, A281L, E283G	1327	1,16	22,2%	1,76
A50Q, C63G, A281F	1286	1,15	17,9%	1,41
S18W, D43E, A50Q, A281L	1635	1,32	9,4%	0,77
M21L, C63G, S113K, T154A A281L	1518	1,24	28,3%	2,23
M21L, L31V, N83G, Q116G A281L	1353	1,18	15,3%	1,21

* Promedio de 4 réplicas

- 15 ** FIOC = veces de aumento con respecto al control (S18W)

Tabla 6B: Mutaciones de la 2ª biblioteca de combinación de AAR(S18W)_7942 correlacionadas con un aumento del título de alcoholes grasos

Mutaciones de ARR (además de S18W)	Título de FALC*		C12	
	mg/l*	FIOC**	Fracción*	FIOC**
S18W (control)	1034	1,00	10,0%	1,00
M21L, C63G, S113K, T154A, A281L	1179	1,14	29,4%	2,94
D16L, M21L, C63G, S113K, T154A, A281L	1221	1,18	19,4%	1,94
L8A, D24V, C63G, S113K, Q155L, A281L	1414	1,37	10,4%	1,04
L8A, M21L, C63G, A77A***, S113K, T154A A281L	1310	1,27	14,2%	1,42
D24P, L31M, C63G, S113K, T154A, A281L	1437	1,39	13,0%	1,30
L8A, D16L, D24V, C63G, S113K, T154A A281L	1342	1,30	12,0%	1,20
D24E, C63G, S113K, T154A, A281L	1425	1,38	14,3%	1,43

* Promedio de 4 réplicas

** FIOC = veces de aumento con respecto al control (S18W)

*** La mutación A77A es una mutación de codón silenciosa de gcc a gca

5 Tabla 6C: Mutaciones de las bibliotecas de combinación de AAR (S18W)_7942 correlacionadas con un aumento de la fracción de alcoholes grasos C₁₂

Mutaciones de AAR (además de S18W)	C ₁₂		Título de FALC*
	Fracción*	FIOC**	FIOC**
S18W (control)	11,0%	1,00	1,00
M21L, C63G, S113K, T154A, A281L	28,3%	2,57	1,24
M21L, Q155G, A281L	24,8%	2,25	1,10
Q155G, A281L, E283G	22,2%	2,02	1,16
M21L, S113K, A281L	20,4%	1,85	1,20
C63G, A281L	18,1%	1,65	1,18
A50Q, C63G, A281F	17,9%	1,63	1,15
C63G	17,7%	1,61	1,00
S10G, C63G, Q155G, S253N, A281F	16,1%	1,46	1,03
D43E, C63G, S113K, A281L	16,0%	1,46	0,98
A281L	15,6%	1,42	1,20
M21L, L31V, N83G, Q116G, A281L	15,3%	1,39	1,18
C63G, N83G, A281Y	14,6%	1,33	1,32
A281L	14,1%	1,28	1,29

* Promedio de 4 réplicas

** FIOC = veces de aumento con respecto al control (S18W)

Ejemplo 6: Aumento del flujo a través de la ruta de síntesis de ácidos grasos - iFab e iFadR Mediados La producción de alcoholes grasos Usando AAR

10 En este ejemplo, se mostró la producción mejorada de alcoholes grasos usando AAR aumentando el flujo a través de la ruta de biosíntesis de ácidos grasos mediada por la sobreexpresión de un operón sintético que comprende varias proteínas FAB (ifab138) y/o la sobreexpresión de la proteína FadR (ifad), un regulador del metabolismo de ácidos grasos. iFAB 138 (SEQ ID NO:55) incluye, en el siguiente orden, los genes fabV de *Vibrio cholerae*, FabH, fabD, fabG y fabA de *Salmonella typhimurium*, y FabF de *Clostridium acetobutlicum*, y se integra como operón
15 sintético controlado por el promotor o bien PlacUV5 o bien PT5. iFadR incluye el gen fadR de *Escherichia coli* (SEQ ID NO:56) controlado por el promotor T5. Los componentes presentes en las cepas de *E. coli* evaluadas en este ejemplo se muestran en la tabla 7, a continuación.

Tabla 7: Cepas de *E. coli* con iFAB138 o fadR

Componentes	DV2	BD061	BD064	Shu002
i(PlacUV5-fab138)	-	+	+	-
i(PTS-fab138)	-	-	-	+
i(PT5-fadR)	-	-	+	+

20 Se expresó AAR (S18W) a partir del plásmido pDS311, una variante de plásmido pDS171S, en la que el codón 18 de AAR especificaba un triptófano en vez de una serina (pCL-AAR(S18W)+ACP-sfp) y se clonó el gen de aldehído reductasa, alrA, de *Acinetobacter baileyi* (SEQ ID NO: 54) en el sentido de 3' del gen sfp de *B. subtilis*. Se transformó

pDS311 en las cepas DV2, BD061, BD064 y Shu002. Se evaluaron las cepas en el protocolo 4NBT_2N (véase anteriormente). Tal como se muestra en la figura 7, la producción de alcoholes grasos aumentó significativamente como resultado de la presencia de ifab138 e ifadR, mostrando la cepa Shu002 el mayor título de alcoholes grasos.

Ejemplo 7: Variantes de AAR mejoradas en una cepa con aumento del flujo a través de la ruta de síntesis de ácidos grasos

En este ejemplo, se demostró la producción mejorada de alcoholes grasos en células huésped recombinantes transformadas con variantes de AAR de bibliotecas de combinación en Shu2, una cepa de *E. coli* con aumento del flujo a través de la ruta de biosíntesis de ácidos grasos mediada por la sobreexpresión de ifab138 y la sobreexpresión de la proteína FadR (véase antes). Se evaluaron cuatro variantes de AAR de la segunda biblioteca de combinación (tabla 6B). Estas variantes albergan las siguientes mutaciones: Com2a: S18W, D24P, L31M, C63G, S113K, T154A, A281L; Comb2b: S18W, D16L, M21L, C63G, S113K, T154A, A281L; Com2c: S18W, L8A, M21L, C63G, A77A, S113K, T154A, A281L; Com2d: D24E, C63G, S113K, T154A, A281L (véase también la figura 8). Se clonaron estas variantes en la estructura principal del plásmido pJL104. Se creó pJL104 clonando el operón sintético accD+ de *C. glutamicum* (tal como se describe en el ejemplo 2, véase antes) en el sentido de 3' del gen alrA en pDS311. Se transformaron los plásmidos resultantes en la cepa Shu002 y se evaluaron las cepas en el protocolo 4NBT_1N (véase antes). Tal como se muestra en la figura 8, se observó una producción elevada de alcoholes grasos en todas las cepas. La cepa BD064 que alberga el plásmido pDS311, que expresa la variante de S18W de AAR_7942, también se evaluó en fermentaciones en tanque. La cepa produjo alcoholes grasos con un título máximo de 42 g/l, un rendimiento del 12,6% (con respecto a glucosa) y una productividad de 0,62 g/l/h tal como se muestra en la figura 9. La distribución de longitud de cadena de los alcoholes grasos fue de la siguiente manera: el 1,2% de C₈, el 7,7% de C₁₀, el 26,9% de C₁₂, el 44,7% de C₁₄, el 16,0% de C₁₆ y el 1,9% de C₁₈. La fracción de alcoholes grasos saturados e insaturados fue del 72,6% y el 27,4%, respectivamente.

Ejemplo 8: Aumento de la actividad de acil aldehído reductasa (AAR) de MED4 y redistribución de la selectividad de longitud de cadena para la producción de alcoholes grasos C14

AAR es uno de los componentes esenciales para la biosíntesis de alcanos de cianobacterias, otro componente esencial es una aldehído descarboxilasa (ADC). Los inventores descubrieron que la MED4_AAR de *Prochlorococcus marinus* es inactiva catalíticamente sin la presencia de MED4_ADC, es decir, cuando sólo se expresa MED4_AAR en *E. coli* no se detectan productos, y cuando se coexpresan MED4_AAR y ADC en *E. coli* los únicos productos detectados con alcanos. Los inventores también descubrieron que cuando se coexpresa MED4_AAR en *E. coli* con una variante de MED4_ADC aparentemente inactiva catalíticamente en la que se reemplaza la histidina 156 por arginina (denominada posteriormente MED4_ADC (H165R)), se detectan alcoholes grasos, y no alcanos (véase la figura 10). Se concluyó a partir de estos datos que MED4_AAR requiere la interacción física con MED4_ADC para ser activa catalíticamente. Los inventores usaron este sistema para identificar variantes de MED4_AAR con aumento de la actividad y/o alteración de la especificidad de sustrato para fines de producción de FALC.

Se expresó MED4_ADC(H165R) junto con una biblioteca de saturación completa de MED4_AAR. Se preparó la biblioteca de saturación de MED4_AAR en un derivado del plásmido pCL1920 (pLS9-195) y se introdujo en una cepa de producción que portaba el plásmido pGLAK-043 (que es el plásmido pACYC-Ptrc-MED4_ADC que alberga la mutación H156R en el gen de ADC). Se indujeron los clones y se seleccionaron variantes de AAR basándose en la producción de más alcohol graso que la enzima AAR silvestre o la capacidad para producir alcoholes grasos con un perfil alterado de longitud de cadena, por ejemplo, un aumento de la fracción de alcohol graso C₁₄. Entonces volvieron a someterse a prueba clones seleccionados en una tanda de validación. Se hicieron crecer de nuevo todas las variantes que mostraron títulos constantes de FALC entre las fermentaciones primaria y secundaria, se aisló el ADN de plásmido, se secuenció y se introdujo de nuevo en la cepa de producción parental para pruebas adicionales. Entonces se sometieron estos nuevos transformantes a otra fermentación confirmatoria y análisis. La tabla 8 a continuación muestra datos representativos de 16 variantes de AAR que produjeron los mayores títulos de FALC (ordenadas de arriba abajo en actividad decreciente). Las variantes oscilaron entre 1,4 veces y 2,2 veces con respecto a la silvestre. Estas variantes mostraron la capacidad para aumentar la actividad de MED4_AAR usando técnicas de evolución dirigida y forma además la base para mejoras adicionales.

Tabla 8: Productividad de FALC de mutantes de MED4_AAR con relación a MED4_AAR silvestre (WT)

Mutación de AAR	Veces de aumento de FALC con respecto a silvestre
V346P	2,2
Q40V	2,2
A345R	2,1
L344S	2,1
D61E	2,1
V346G	2,0
L344D	1,9
G52V	1,9
A345*	1,9

L344T	1,8
K303G	1,8
L344A	1,8
H340P	1,6
S588V	1,5
K339L	1,4
G273E	1,4

* Variante truncada a la que le faltan los dos últimos aminoácidos.

5 También se exploró el conjunto de datos para detectar variantes de AAR que presentaban perfiles alterados de longitud de cadena. La especie más común producida por la AAR silvestre tiene una longitud de cadena de C₁₆. Un aumento de la proporción de especies FALC con longitudes de cadena más cortas que C₁₆ es de interés. Se identificaron dos clones variantes que mostraron un aumento de aproximadamente 3 veces en la cantidad de FALC C₁₄ (figura 10). La secuenciación de estos clones reveló que eran mutantes D61E idénticos con la misma secuencia de codones de nucleótidos. Se introdujo de nuevo el ADN de plásmido para la variante de D61E en la cepa parental que contiene H156R ADC. Los resultados muestran que la expresión de la variante de D61E de AAR en una célula huésped recombinante desvía la distribución de longitud de cadena de especies FALC hacia cadenas de carbonos más cortas. La tabla 9 a continuación ilustra la distribución de longitud de cadena de FALC producida por células huésped recombinantes que expresan la variante de D61E de MED4_AAR en comparación con MED4_AAR silvestre (WT) y la variante de V346P de MED4_AAR que no produjo ningún producto con longitudes de cadena alteradas.

Tabla 9: Distribución de longitud de cadena para variantes de AAR y AAR silvestre

MED4_AAR/ADC	Alcohol graso		
	% de C14	% de C16	% de C18
AAR WT/ADC(H156R)	6	92	2
AAR(D61E)/ADC(H156R)	14*	85	0
AAR(V346P)/ADC(H156R)	4	92	4
AAR WT/ADC WT	ND	ND	ND
AAR(D61E)/ADC	ND	ND	ND
AAR(V346P)/ADC WT	ND	ND	ND

15 ND = no detectado * 1% de varianza atribuible a desviaciones debidas al promediado

20 Estas variantes pueden además recombinarse y examinarse para detectar mejoras en el fondo de MED4_ADC (H156R). La variante de MED4_AAR (D61E) y progenie mutada adicional pueden ser útiles para disminuir la longitud de cadena promedio tanto de alcoholes grasos como de alcanos. La variante de MED4_AAR (D61E) demuestra que la especificidad de longitud de cadena de MED4_AAR es moldeable y presenta la posibilidad para una mejora adicional de esta actividad a través de esfuerzos de modificación por ingeniería de proteínas adicionales. Se secuenciaron todas las variantes descritas a partir de la progenie de pLS9-195 y contenían mutaciones de codones correspondientes a las sustituciones de aminoácidos enumeradas.

Tabla 10: Nombres relacionados con la lista de secuencias

SEQ ID NO	Tipo	Nombre
1	sec. de ácido nucleico	PCC 73102_a cp de <i>Nostoc punctiforme</i> N.º de registro YP_001867863
2	sec. de aminoácidos	PCC 73102_a cp de <i>Nostoc punctiforme</i> N.º de registro YP_001867863
3	sec. de ácido nucleico	PCC 6803_a cp de <i>Synechocystis</i> sp. N.º de registro NP_440632.1
4	sec. de aminoácidos	PCC 6803_a cp de <i>Synechocystis</i> sp. N.º de registro NP_440632.1
5	sec. de ácido nucleico	CCMP1986_a cp de <i>Prochlorococcus marinus</i> subsp. <i>pastoris</i> str. N.º de registro NP_893725.1
6	sec. de aminoácidos	CCMP1986_a cp de <i>Prochlorococcus marinus</i> subsp. <i>pastoris</i> str. N.º de registro NP_893725.1
7	sec. de ácido nucleico	PCC 7942_a cp de <i>Synechococcus elongatus</i> N.º de registro YP_399555
8	sec. de aminoácidos	PCC 7942_a cp de <i>Synechococcus elongatus</i> N.º de registro YP_399555
9	sec. de ácido nucleico	PCC 7120_a cp de <i>Nostoc</i> sp. N.º de registro NP_487382.1
10	sec. de aminoácidos	PCC 7120_a cp de <i>Nostoc</i> sp. N.º de registro NP_487382.1
11	sec. de ácido nucleico	<i>sfp</i> de <i>B. subtilis</i> (sintetizado) como en n.º de registro X63158.1
12	sec. de aminoácidos	<i>sfp</i> de <i>B. subtilis</i> (sintetizado) como en n.º de registro X63158.1
13	sec. de cebador	168IFF
14	sec. de cebador	168IFR
15	sec. de cebador	169IFF

ES 2 600 354 T3

16	sec. de cebador	169IFR
17	sec. de cebador	170IFF
18	sec. de cebador	170IFR
19	sec. de cebador	171IFF
20	sec. de cebador	171IFR
21	sec. de cebador	172IFF
22	sec. de cebador	172IFR
23	sec. de cebador	168SIFF
24	sec. de cebador	170S1FF
25	sec. de cebador	171SIFF
26	sec. de cebador	168SIFR
27	sec. de ácido nucleico	Acil-CoA reductasa (AAR) de <i>Synechococcus elongatus</i> PCC7942 YP_400611 (Synpcc7942_1594)
28	sec. de aminoácidos	Acil-CoA reductasa (AAR) de <i>Synechococcus elongatus</i> PCC7942 YP_400611 (Synpcc7942_1594)
29	sec. de ácido nucleico	AAR de <i>Synechocystis sp.</i> PCC6803 sll0209 (NP_442146)
30	sec. de aminoácidos	AAR de <i>Synechocystis sp.</i> PCC6803 sll0209 (NP_442146)
31	sec. de ácido nucleico	AAR de <i>Cyanothece sp.</i> ATCC51142 cce_1430 (YP_001802846)
32	sec. de aminoácidos	AAR de <i>Cyanothece sp.</i> ATCC51142 cce_1430 (YP_001802846)
33	sec. de ácido nucleico	AAR de <i>Prochlorococcus marinus</i> CCMP1986 PMM0533 (NP_892651)
34	sec. de aminoácidos	AAR de <i>Prochlorococcus marinus</i> CCMP1986 PMM0533 (NP_892651)
35	sec. de ácido nucleico	AAR de <i>Gloeobacter violaceus</i> PCC7421 NP_96091 (gll3145)
36	sec. de aminoácidos	AAR de <i>Gloeobacter violaceus</i> PCC7421 NP_96091 (gll3145)
37	sec. de ácido nucleico	AAR de <i>Nostoc punctiforme</i> PCC73102 ZP_00108837 (Npun02004176)
38	sec. de aminoácidos	AAR de <i>Nostoc punctiforme</i> PCC73102 ZP_00108837 (Npun02004176)
39	sec. de ácido nucleico	AAR de <i>Anabaena variabilis</i> ATCC29413 YP_323044 (Ava_2534)
40	sec. de aminoácidos	AAR de <i>Anabaena variabilis</i> ATCC29413 YP_323044 (Ava_2534)
41	sec. de ácido nucleico	AAR de <i>Synechococcus elongatus</i> PCC6301 YP_170761 (syc0051_d)
42	sec. de aminoácidos	AAR de <i>Synechococcus elongatus</i> PCC6301 YP_170761 (syc0051_d)
43	sec. de ácido nucleico	AAR de <i>Nostoc sp.</i> PCC7120 alr5284 (NP_489324)
44	sec. de aminoácidos	AAR de <i>Nostoc sp.</i> PCC7120 alr5284 (NP_489324)
45	sec. de ácido nucleico	<i>birA</i> de <i>Corynebacterium glutamicum</i> (YP_224991)
46	ADN sintético	<i>birA</i> de <i>Corynebacterium glutamicum</i> (YP_224991)
47	sec. de aminoácidos	<i>birA</i> de <i>Corynebacterium glutamicum</i> (YP_224991)
48	sec. de ácido nucleico	<i>accDA1 (dtsR)</i> de <i>Corynebacterium glutamicum</i> (YP_224991)
49	sec. de aminoácidos	<i>accDA1 (dtsR)</i> de <i>Corynebacterium glutamicum</i> (YP_224991)
50	sec. de ácido nucleico	<i>accCB</i> de <i>Corynebacterium glutamicum</i> (YP_224991)
51	sec. de aminoácidos	<i>accCB</i> de <i>Corynebacterium glutamicum</i> (YP_224991)
52	sec. de aminoácidos	<i>AlrA Acinetobacter sp.</i> M-1
53	sec. de aminoácidos	<i>AlrAadp1</i>
54	sec. de aminoácidos	proteína <i>alrAadp1 Acinetobacter bailyi</i> ADP1-WT
55	sec. de ácido nucleico	<i>iFAB138</i>
56	sec. de ácido nucleico	<i>FadR</i> de <i>E. coli</i> MG1655 (NP_415705)
57	sec. de aminoácidos	Mutante de AAR con la mutación S18W (producido a partir de Acil-CoA reductasa (AAR) de <i>Synechococcus elongatus</i> PCC7942 YP_400611 (Synpcc7942_1594))

Lista de secuencias

<110> REG Life Sciences, LLC

<120> ACIL-ACP REDUCTASA CON PROPIEDADES MEJORADAS

5 <130> EP100911FZRP

<140> Documento EP14716444.6

<141> 16-01-2014

<150> Documento PCT/US2014/011859

<151> 16-01-2014

<150> Documento 61/753.273

<151> 16-01-2013

5 <160> 78

<170> PatentIn versión 3.5

<210> 1

<211> 255

<212> ADN

10 <213> *Nostoc punctiforme*

<400> 1

atgagccaaa	cggaactttt	tgaaaaggtc	aagaaaatcg	tcatcgaaca	actgagtgtt	60
gaagatgctt	ccaaaatcac	tccacaagct	aagtttatgg	aagatttagg	agctgattcc	120
ctggatactg	ttgaactcgt	gatggctttg	gaagaagaat	ttgatatcga	aattccccgac	180
gaagctgccg	agcagattgt	atcggttcaa	gacgcagtag	attacatcaa	taacaaagtt	240
gctgcatcag	cttaa					255

<210> 2

<211> 84

15 <212> PRT

<213> *Nostoc punctiforme*

<400> 2

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

<210> 3

20 <211> 234

<212> ADN

ES 2 600 354 T3

<213> *Synechocystis* sp.

<400> 3

atgaatcagg aaatTTTTga aaaagtaaaa aaaatcgtcg tggaacagtt ggaagtggat	60
cctgacaaag tgacccccga tgccaccttt gccgaagatt taggggctga ttcctcgtat	120
acagtggaaat tggatcatggc cctggaagaa gagtttgata ttgaaattcc cgatgaagtg	180
gCGGaaacca ttgataccgt gggcaaagcc gttgagcata tcgaaagtaa ataa	234

<210> 4

5 <211> 77

<212> PRT

<213> *Synechocystis* sp.

<400> 4

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

10 <210> 5

<211> 243

<212> ADN

<213> *Prochlorococcus marinus*

<400> 5

atgtcacaag aagaaatcct tcaaaaagta tgctctattg tttctgagca actaagtgtt	60
gaatcagccg aagtaaaatc tgattcaaac tttcaaaatg atttaggtgc agactcccta	120
gacaccgtag agctagttat ggctcttgaa gaagcatttg atatcgagat acctgatgaa	180
gcagctgaag gtatcgcaac agtaggagat gctgttaaata tcatcgaaga aaaaaaaggt	240
taa	243

15

<210> 6

<211> 80

<212> PRT

<213> *Prochlorococcus marinus*

20 <400> 6

ES 2 600 354 T3

Met Ser Gln Glu Glu Ile Leu Gln Lys Val Cys Ser Ile Val Ser Glu
1 5 10 15

Gln Leu Ser Val Glu Ser Ala Glu Val Lys Ser Asp Ser Asn Phe Gln
20 25 30

Asn Asp Leu Gly Ala Asp Ser Leu Asp Thr Val Glu Leu Val Met Ala
35 40 45

Leu Glu Glu Ala Phe Asp Ile Glu Ile Pro Asp Glu Ala Ala Glu Gly
50 55 60

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

<210> 7

<211> 243

5 <212> ADN

<213> *Synechococcus elongatus*

<400> 7

atgagccaag aagacatctt cagcaaagtc aaagacattg tggctgagca gctgagtgtg	60
gatgtggctg aagtcaagcc agaatccagc ttccaaaacg atctgggagc ggactcgctg	120
gacaccgtgg aactggtgat ggctctggaa gaggctttcg atatcgaaat ccccgatgaa	180
gccgctgaag gcattgcgac cgttcaagac gccgtcgatt tcatcgctag caaagctgcc	240
tag	243

<210> 8

10 <211> 80

<212> PRT

<213> *Synechococcus elongatus*

<400> 8

Met Ser Gln Glu Asp Ile Phe Ser Lys Val Lys Asp Ile Val Ala Glu
1 5 10 15

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

Asn Asp Leu Gly Ala Asp Ser Leu Asp Thr Val Glu Leu Val Met Ala
35 40 45

Leu Glu Glu Ala Phe Asp Ile Glu Ile Pro Asp Glu Ala Ala Glu Gly
50 55 60

Ile Ala Thr Val Gln Asp Ala Val Asp Phe Ile Ala Ser Lys Ala Ala
65 70 75 80

ES 2 600 354 T3

<210> 9

<211> 255

<212> ADN

<213> *Nostoc* sp.

5 <400> 9

atgagccaat cagaaacttt tgaaaaagtc aaaaaaattg ttatcgaaca actaagtgtg 60
 gagaaccctg acacagtaac tccagaagct agttttgccca acgatttaca ggctgattcc 120
 ctcgatacag tagaactagt aatggctttg gaagaagaat ttgatatcga aattcccgat 180
 gaagccgcag agaaaattac cactgttcaa gaagcgggtgg attacatcaa taaccaagtt 240
 gccgcatcag cttaa 255

<210> 10

<211> 84

10 <212> PRT

<213> *Nostoc* sp.

<400> 10

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

<210> 11

15 <211> 674

<212> ADN

<213> Secuencia artificial

<220>

<221> fuente

20 <223> /nota="Descripción de secuencia artificial: Polinucleótido sintético"

<400> 11

ES 2 600 354 T3

atgaagattt acggaattta tatggaccgc ccgctttcac aggaagaaaa tgaacggttc 60
atgactttca taccacctga aaaacgggag aaatgccgga gattttatca taaagaagat 120
gctcaccgca ccctgctggg agatgtgctc gttcgtcag tcataagcag gcagtatcag 180
ttggacaaat ccgatatccg ctttagcacg caggaatacg ggaagccgtg catccctgat 240
cttcccgcag ctcatctcaa cttttctcac tccggccgct gggtcattgg tgcgtttgat 300
tcacagccga tcggcataga tatcgaaaaa acgaaaccga tcagccttga gatcgccaag 360
cgcttctttt caaaaacaga gtacagcgac ctttttagcaa aagacaagga cgagcagaca 420
gactatTTTT atcatctatg gtcaatgaaa gaaagcttta tcaaacagga aggcaaaggc 480
ttatcgcttc cgcttgattc cttttcagtg cgcttgcac aggacggaca agtatccatt 540
gagcttccgg acagccattc cccatgctat atcaaaacgt atgaggtcga tcccggctac 600
aaaatggctg tatgcccgc acaccctgtt tccccgagga tatcacaatg gtctcgtacg 660
aagagctttt ataa 674

<210> 12

<211> 224

5 <212> PRT

<213> Secuencia artificial

<220>

<221> fuente

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

10 <400> 12

ES 2 600 354 T3

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

<210> 13

<211> 52

5 <212> ADN

<213> Secuencia artificial

<220>

<221> fuente

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

10 <400> 13

	ggcaattga gaatttaagg aggaaaacaa aatgagccaa gaagacatct tc	52
	<210> 14	
	<211> 33	
	<212> ADN	
5	<213> Secuencia artificial	
	<220>	
	<221> fuente	
	<223> /nota="Descripción de secuencia artificial: Cebador sintético"	
	<400> 14	
10	ccaagcttc gaattcctag gcagcttgc tag	33
	<210> 15	
	<211> 56	
	<212> ADN	
	<213> Secuencia artificial	
15	<220>	
	<221> fuente	
	<223> /nota="Descripción de secuencia artificial: Cebador sintético"	
	<400> 15	
	ggcaattga gaatttaagg aggaaaacaa aatgaatcag gaaattttg aaaaag	56
20	<210> 16	
	<211> 41	
	<212> ADN	
	<213> Secuencia artificial	
	<220>	
25	<221> fuente	
	<223> /nota="Descripción de secuencia artificial: Cebador sintético"	
	<400> 16	
	ccaagcttc gaattcttat ttacttgcga tatgctcaac g	41
	<210> 17	
30	<211> 53	
	<212> ADN	
	<213> Secuencia artificial	
	<220>	
	<221> fuente	
35	<223> /nota="Descripción de secuencia artificial: Cebador sintético"	
	<400> 17	
	ggcaattga gaatttaagg aggaaaacaa aatgcacaa gaagaaatcc ttc	53

	<210> 18	
	<211> 46	
	<212> ADN	
	<213> Secuencia artificial	
5	<220>	
	<221> fuente	
	<223> /nota="Descripción de secuencia artificial: Cebador sintético"	
	<400> 18	
	cccaagcttc gaattcttaa ccttttttt ctcgatgaa ttaac	46
10	<210> 19	
	<211> 53	
	<212> ADN	
	<213> Secuencia artificial	
	<220>	
15	<221> fuente	
	<223> /nota="Descripción de secuencia artificial: Cebador sintético"	
	<400> 19	
	ggcaattga gaatttaagg aggaaaacaa aatgagccaa acggaacttt ttg	53
	<210> 20	
20	<211> 38	
	<212> ADN	
	<213> Secuencia artificial	
	<220>	
	<221> fuente	
25	<223> /nota="Descripción de secuencia artificial: Cebador sintético"	
	<400> 20	
	cccaagcttc gaattcttaa gctgatgcag caacttg	38
	<210> 21	
	<211> 53	
30	<212> ADN	
	<213> Secuencia artificial	
	<220>	
	<221> fuente	
	<223> /nota="Descripción de secuencia artificial: Cebador sintético"	
35	<400> 21	
	ggcaattga gaatttaagg aggaaaacaa aatgagccaa tcagaaactt ttg	53
	<210> 22	

<211> 34
 <212> ADN
 <213> Secuencia artificial
 <220>
 5 <221> fuente
 <223> /nota="Descripción de secuencia artificial: cebador sintético"
 <400> 22
 cccaagcttc gaattcttaa gctgatgcgg caac 34
 <210> 23
 10 <211> 43
 <212> ADN
 <213> Secuencia artificial
 <220>
 <221> fuente
 15 <223> /nota="Descripción de secuencia artificial: Cebador sintético"
 <400> 23
 agctgcctag gaatttaagg aggaataaac catgaagatt tac 43
 <210> 24
 <211> 43
 20 <212> ADN
 <213> Secuencia artificial
 <220>
 <221> fuente
 <223> /nota="Descripción de secuencia artificial: Cebador sintético"
 25 <400> 24
 aaaagggttaa gaatttaagg aggaataaac catgaagatt tac 43
 <210> 25
 <211> 43
 <212> ADN
 30 <213> Secuencia artificial
 <220>
 <221> fuente
 <223> /nota="Descripción de secuencia artificial: cebador sintético"
 <400> 25
 35 atcagcttaa gaatttaagg aggaataaac catgaagatt tac 43
 <210> 26
 <211> 41

ES 2 600 354 T3

<212> ADN

<213> Secuencia artificial

<220>

<221> fuente

5 <223> /nota="Descripción de secuencia artificial: Cebador sintético"

<400> 26

cccaagcttc gaattcttat aaaagctctt cgtaagagac c

41

<210> 27

<211> 1029

10 <212> ADN

<213> *Synechococcus elongatus*

<400> 27

atggcattcg	gtcttatcgg	tcatctcacc	agtttggagc	aggcccgcga	cgtttctcgc	60
aggatgggct	acgacgaata	cgccgatcaa	ggattggagt	tttggagtag	cgctcctcct	120
caaatcgttg	atgaaatcac	agtcaccagt	gccacaggca	aggtgattca	cggtcgtctac	180
atcgaatcgt	gtttcttgcc	ggaaatgctg	gcggcgcgcc	gcttcaaac	agccacgcgc	240
aaagtctca	atgccatgtc	ccatgcccaa	aaacacggca	tcgacatctc	ggccttgggg	300
ggctttacct	cgattatfff	cgagaatttc	gatttggcca	gtttgcggca	agtgcgcgac	360
actaccttgg	agtttgaacg	gttcaccacc	ggcaatactc	acacggccta	cgtaatctgt	420
agacaggtgg	aagccgctgc	taaaacgctg	ggcatcgaca	ttaccaagc	gacagtagcg	480
gttgtcggcg	cgactggcga	tatcggtagc	gctgtctgcc	gctggctcga	cctcaaacctg	540
ggtgtcggtg	atftgatcct	gacggcgcgc	aatcaggagc	gtttggataa	cctgcaggct	600
gaactcggcc	ggggcaagat	tctgcccttg	gaagccgctc	tgccggaagc	tgactttatc	660
gtgtgggctg	ccagtatgcc	tcagggcgta	gtgatcgacc	cagcaaccct	gaagcaaccc	720
tgcgctcctaa	tcgacggggg	ctacccaaa	aacttgggca	gcaaagtcca	aggtaggggc	780
atctatgtcc	tcaatggcgg	ggtagttaa	cattgcttcg	acatcgactg	gcagatcatg	840
tccgctgcag	agatggcgcg	gcccgagcgc	cagatgtttg	cctgctttgc	cgaggcgatg	900
ctcttggaat	ttgaaggctg	gcatactaac	ttctctggg	gccgcaacca	aatcacgatc	960
gagaagatgg	aagcgatcgg	tgaggcatcg	gtgcgccacg	gcttccaacc	cttggcattg	1020
gcaatttga						1029

<210> 28

15 <211> 342

<212> PRT

<213> *Synechococcus elongatus*

<400> 28

ES 2 600 354 T3

Met Ala Phe Gly Leu Ile Gly His Leu Thr Ser Leu Glu Gln Ala Arg
1 5 10 15

Asp Val Ser Arg Arg Met Gly Tyr Asp Glu Tyr Ala Asp Gln Gly Leu
20 25 30

Glu Phe Trp Ser Ser Ala Pro Pro Gln Ile Val Asp Glu Ile Thr Val
35 40 45

Thr Ser Ala Thr Gly Lys Val Ile His Gly Arg Tyr Ile Glu Ser Cys
50 55 60

Phe Leu Pro Glu Met Leu Ala Ala Arg Arg Phe Lys Thr Ala Thr Arg
65 70 75 80

Lys Val Leu Asn Ala Met Ser His Ala Gln Lys His Gly Ile Asp Ile
85 90 95

ES 2 600 354 T3

Ser Ala Leu Gly Gly Phe Thr Ser Ile Ile Phe Glu Asn Phe Asp Leu
100 105 110

Ala Ser Leu Arg Gln Val Arg Asp Thr Thr Leu Glu Phe Glu Arg Phe
115 120 125

Thr Thr Gly Asn Thr His Thr Ala Tyr Val Ile Cys Arg Gln Val Glu
130 135 140

Ala Ala Ala Lys Thr Leu Gly Ile Asp Ile Thr Gln Ala Thr Val Ala
145 150 155 160

Val Val Gly Ala Thr Gly Asp Ile Gly Ser Ala Val Cys Arg Trp Leu
165 170 175

Asp Leu Lys Leu Gly Val Gly Asp Leu Ile Leu Thr Ala Arg Asn Gln
180 185 190

Glu Arg Leu Asp Asn Leu Gln Ala Glu Leu Gly Arg Gly Lys Ile Leu
195 200 205

Pro Leu Glu Ala Ala Leu Pro Glu Ala Asp Phe Ile Val Trp Val Ala
210 215 220

Ser Met Pro Gln Gly Val Val Ile Asp Pro Ala Thr Leu Lys Gln Pro
225 230 235 240

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

Gln Gly Glu Gly Ile Tyr Val Leu Asn Gly Gly Val Val Glu His Cys
260 265 270

Phe Asp Ile Asp Trp Gln Ile Met Ser Ala Ala Glu Met Ala Arg Pro
275 280 285

Glu Arg Gln Met Phe Ala Cys Phe Ala Glu Ala Met Leu Leu Glu Phe
290 295 300

Glu Gly Trp His Thr Asn Phe Ser Trp Gly Arg Asn Gln Ile Thr Ile
305 310 315 320

Glu Lys Met Glu Ala Ile Gly Glu Ala Ser Val Arg His Gly Phe Gln
325 330 335

Pro Leu Ala Leu Ala Ile
340

<210> 29

<211> 1023

<212> ADN

5 <213> *Synechocystis* sp.

ES 2 600 354 T3

<400> 29

atgtttggtc ttattggtca tctcacgagt ttagaacacg cccaagcggg tgctgaagat 60
 ttaggctatc ctgagtacgc caaccaaggc ctggattttt ggtggttcggc tcctccccc 120
 gtggttgata attttcaggt gaaaagtgtg acggggcagg tgattgaagg caaatatgtg 180
 gagtcttgct ttttgccgga aatgttaacc caacggcgga tcaaagcggc cattcgtaaa 240
 atcctcaatg ctatggccct ggcccaaaaag gtgggcttgg atattacggc cctgggaggc 300
 ttttcttcaa tcgtatttga agaatttaac ctcaagcaaa ataatcaagt ccgcaatgtg 360
 gaactagatt ttcagcgggt caccactggt aatacccaca ccgcttatgt gatctgccgt 420
 caggtcgagt ctggagctaa acagttgggt attgatctaa gtcaggcaac ggtagcgggt 480
 tgtggcgcca cgggagatat tggtagcgcc gtatgtcggt ggtagatag caaacatcaa 540
 gttaaggaat tattgctaatt tgcggtaac cgccaaagat tggaaaatct ccaagaggaa 600
 ttgggctggg gcaaaattat ggatttggaa acagccctgc cccaggcaga tattattggt 660
 tgggtggcta gtatgcccaa ggggtagaa attgcggggg aaatgctgaa aaagccctgt 720
 ttgattgtgg atgggggcta tccaagaat ttagacacca gggtgaaagc ggatgggggtg 780
 catattctca agggggggat ttagaacat tcccttgata ttacctgga aattatgaag 840
 attgtggaga tggatattcc ctcccggcaa atgttcgct gttttgcgga ggccattttg 900
 cttagagttt agggctggcg cactaatttt tcctggggcc gcaacaaat ttccgttaat 960
 aaaatggagg cgattggtga agcttctgtc aagcatggct tttgcccttt agtagctctt 1020
 tag 1023

<210> 30

<211> 340

5 <212> PRT

<213> *Synechocystis* sp.

<400> 30

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

ES 2 600 354 T3

85 - 90 95

Ala Leu Gly Gly Phe Ser Ser Ile Val Phe Glu Glu Phe Asn Leu Lys
100 105 110

Gln Asn Asn Gln Val Arg Asn Val Glu Leu Asp Phe Gln Arg Phe Thr
115 120 125

Thr Gly Asn Thr His Thr Ala Tyr Val Ile Cys Arg Gln Val Glu Ser
130 135 140

Gly Ala Lys Gln Leu Gly Ile Asp Leu Ser Gln Ala Thr Val Ala Val
145 150 155 160

Cys Gly Ala Thr Gly Asp Ile Gly Ser Ala Val Cys Arg Trp Leu Asp
165 170 175

Ser Lys His Gln Val Lys Glu Leu Leu Leu Ile Ala Arg Asn Arg Gln
180 185 190

Arg Leu Glu Asn Leu Gln Glu Glu Leu Gly Arg Gly Lys Ile Met Asp
195 200 205

Leu Glu Thr Ala Leu Pro Gln Ala Asp Ile Ile Val Trp Val Ala Ser
210 215 220

Met Pro Lys Gly Val Glu Ile Ala Gly Glu Met Leu Lys Lys Pro Cys
225 230 235 240

Leu Ile Val Asp Gly Gly Tyr Pro Lys Asn Leu Asp Thr Arg Val Lys
245 250 255

Ala Asp Gly Val His Ile Leu Lys Gly Gly Ile Val Glu His Ser Leu
260 265 270

Asp Ile Thr Trp Glu Ile Met Lys Ile Val Glu Met Asp Ile Pro Ser
275 280 285

Arg Gln Met Phe Ala Cys Phe Ala Glu Ala Ile Leu Leu Glu Phe Glu
290 295 300

Gly Trp Arg Thr Asn Phe Ser Trp Gly Arg Asn Gln Ile Ser Val Asn
305 310 315 320

Lys Met Glu Ala Ile Gly Glu Ala Ser Val Lys His Gly Phe Cys Pro
325 330 335

Leu Val Ala Leu
340

<210> 31

<211> 1023

ES 2 600 354 T3

<212> ADN

<213> *Cyanothece* sp.

<400> 31

```

atgtttgggtt taattgggtca tcttacaagt ttagaacacg cccactccgt tgctgatgcc      60
tttggtctatg gcccatacgc cactcagggga cttgatttgt ggtgttctgc tccaccccaa      120
ttcgtcgagc attttcatgt tactagcatc acaggacaaa ccatcgaagg aaagtatata      180
gaatccgctt tcttaccaga aatgctgata aagcgacgga ttaaagcagc aattcgcaaa      240
atactgaatg cgatggcctt tgctcagaaa aataacctta acatcacagc attagggggc      300
ttttcttcga ttatTTTTga agaatttaat ctcaaagaga atagacaagt tcgtaatgtc      360
tctttagagt ttgatcgctt caccaccgga aacaccata ctgcttatat catttgcgt      420
caagttgaac aggcattccgc taaactaggg attgacttat cccaagcaac ggttgctatt      480
tgcggggcaa ccggagatat tggcagtgca gtgtgtcggt ggtagatag aaaaaccgat      540
accaggaac tattcttaat tgctcgaat aaagaacgat tacaacgact gcaagatgag      600
ttgggacggg gtaaaattat gggattggag gaggctttac ccgaagcaga tattatcggt      660
tgggtggcga gtatgcccga aggagtggaa attaatgccg aaactctcaa aaaaccctgt      720
ttaattatcg atggtgggta tcctaagaat ttagacacaa aaattaaaca tcctgatgtc      780
catatcctga aagggggaat tgtagaacat tctctagata ttgactggaa gattatggaa      840
actgtcaata tggatgttcc ttctcgtcaa atgtttgctt gttttgccga agccatttta      900
ttagagtttg aacaatggca cactaatttt tcttggggac gcaatcaaat tacagtgact      960
aaaatggaac aaataggaga agcttctgtc aaacatgggt tacaaccgtt gttgagttgg      1020
taa                                                                                   1023
    
```

5 <210> 32

<211> 340

<212> PRT

<213> *Cyanothece* sp.

<400> 32

```

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

ES 2 600 354 T3

Ile Leu Asn Ala Met 85 Ala Phe Ala Gln Lys 90 Asn Asn Leu Asn Ile Thr 95
 Ala Leu Gly Gly 100 Phe Ser Ser Ile Ile 105 Phe Glu Glu Phe Asn 110 Leu Lys
 Glu Asn Arg 115 Gln Val Arg Asn Val 120 Ser Leu Glu Phe Asp 125 Arg Phe Thr
 Thr Gly 130 Asn Thr His Thr Ala 135 Tyr Ile Ile Cys Arg 140 Gln Val Glu Gln
 Ala Ser Ala Lys Leu Gly 150 Ile Asp Leu Ser Gln 155 Ala Thr Val Ala Ile 160
 Cys Gly Ala Thr Gly 165 Asp Ile Gly Ser Ala 170 Val Cys Arg Trp Leu Asp 175
 Arg Lys Thr Asp 180 Thr Gln Glu Leu Phe 185 Leu Ile Ala Arg Asn 190 Lys Glu
 Arg Leu Gln 195 Arg Leu Gln Asp Glu 200 Leu Gly Arg Gly Lys 205 Ile Met Gly
 Leu Glu 210 Glu Ala Leu Pro Glu 215 Ala Asp Ile Ile Val 220 Trp Val Ala Ser
 Met Pro Lys Gly Val Glu 230 Ile Asn Ala Glu Thr 235 Leu Lys Lys Pro Cys 240
 Leu Ile Ile Asp Gly 245 Gly Tyr Pro Lys Asn 250 Leu Asp Thr Lys Ile Lys 255
 His Pro Asp Val 260 His Ile Leu Lys Gly 265 Gly Ile Val Glu His 270 Ser Leu
 Asp Ile Asp 275 Trp Lys Ile Met Glu 280 Thr Val Asn Met Asp 285 Val Pro Ser
 Arg Gln 290 Met Phe Ala Cys Phe 295 Ala Glu Ala Ile Leu 300 Leu Glu Phe Glu
 Gln Trp His Thr Asn Phe 310 Ser Trp Gly Arg Asn 315 Gln Ile Thr Val Thr 320
 Lys Met Glu Gln Ile 325 Gly Glu Ala Ser Val 330 Lys His Gly Leu Gln Pro 335
 Leu Leu Ser Trp 340

<210> 33

<211> 1044

ES 2 600 354 T3

<212> ADN

<213> *Prochlorococcus marinus*

<400> 33

```

atggcatttg ggcttatagg tcattcaact agttttgaag atgcaaaaag aaaggcttca      60
ttattgggct ttgatcatat tgcggatggt gatttagatg tttgggtgcac agctccacct      120
caactagttg aaaatgtaga ggttaaaagt gctataggta tatcaattga aggttcttat      180
attgattcat gtttcgttcc tgaaatgctt tcaagattta aaacggcaag aagaaaagta      240
ttaaatgcaa tggaattagc tcaaaaaaaaa ggtattaata ttaccgcttt ggggggggttc      300
acttctatca tctttgaaaa ttttaatctc cttcaacata agcagattag aaacacttca      360
ctagagtggg aaaggtttac aactggtaat actcatactg cgtgggttat ttgcaggcaa      420
ttagagatga atgctcctaa aataggtatt gatcttaaaa gcgcaacagt tgctgtagtt      480
ggtgctactg gagatatagg cagtgcgtgt tgtcgatggt taatcaataa aacaggattt      540
ggggaacttc ttttggtagc taggcaaaag gaacccttgg attccttgca aaaggaatta      600
gatggtggaa ctatcaaaaa tctagatgaa gcattgcctg aagcagatat tgttgtatgg      660
gtagcaagta tgccaaagac aatggaaatc gatgctaata atcttaaaca accatgttta      720
atgattgatg gaggttatcc aaagaatcta gatgaaaaat ttcaaggaaa taatatacat      780
gttgtaaaag gaggtatagt aagattcttc aatgatatag gttggaatat gatggaacta      840
gctgaaatgc aaaatcccca gagagaaatg tttgcatgct ttgcagaagc aatgatttta      900
gaatttgaaa aatgtcatac aaactttagc tggggaagaa ataatatatc tctcgagaaa      960
atggagttta ttggagctgc ttctgtaaag catggcttct ctgcaattgg cctagataag     1020
catccaaaag tactagcagt ttga                                             1044
    
```

5 <210> 34

<211> 347

<212> PRT

<213> *Prochlorococcus marinus*

<400> 34

```

Met Ala Phe Gly Leu Ile Gly His Ser Thr Ser Phe Glu Asp Ala Lys
 1           5           10           15

Arg Lys Ala Ser Leu Leu Gly Phe Asp His Ile Ala Asp Gly Asp Leu
          20           25           30

Asp Val Trp Cys Thr Ala Pro Pro Gln Leu Val Glu Asn Val Glu Val
          35           40           45

Lys Ser Ala Ile Gly Ile Ser Ile Glu Gly Ser Tyr Ile Asp Ser Cys
          50           55           60

Phe Val Pro Glu Met Leu Ser Arg Phe Lys Thr Ala Arg Arg Lys Val
          65           70           75           80
    
```

ES 2 600 354 T3

Leu Asn Ala Met Glu₈₅ Leu Ala Gln Lys Lys₉₀ Gly Ile Asn Ile Thr Ala
 Leu Gly Gly Phe₁₀₀ Thr Ser Ile Ile Phe₁₀₅ Glu Asn Phe Asn Leu₁₁₀ Leu Gln
 His Lys Gln₁₁₅ Ile Arg Asn Thr Ser₁₂₀ Leu Glu Trp Glu Arg₁₂₅ Phe Thr Thr
 Gly Asn₁₃₀ Thr His Thr Ala Trp₁₃₅ Val Ile Cys Arg Gln₁₄₀ Leu Glu Met Asn
 Ala₁₄₅ Pro Lys Ile Gly Ile₁₅₀ Asp Leu Lys Ser Ala₁₅₅ Thr Val Ala Val Val₁₆₀
 Gly Ala Thr Gly Asp₁₆₅ Ile Gly Ser Ala Val₁₇₀ Cys Arg Trp Leu Ile₁₇₅ Asn
 Lys Thr Gly Ile₁₈₀ Gly Glu Leu Leu Leu₁₈₅ Val Ala Arg Gln Lys₁₉₀ Glu Pro
 Leu Asp Ser₁₉₅ Leu Gln Lys Glu Leu₂₀₀ Asp Gly Gly Thr Ile₂₀₅ Lys Asn Leu
 Asp Glu₂₁₀ Ala Leu Pro Glu Ala₂₁₅ Asp Ile Val Val Trp₂₂₀ Val Ala Ser Met
 Pro Lys Thr Met Glu Ile₂₃₀ Asp Ala Asn Asn Leu₂₃₅ Lys Gln Pro Cys Leu₂₄₀
 Met Ile Asp Gly Gly₂₄₅ Tyr Pro Lys Asn Leu₂₅₀ Asp Glu Lys Phe Gln₂₅₅ Gly
 Asn Asn Ile His₂₆₀ Val Val Lys Gly Gly₂₆₅ Ile Val Arg Phe Phe₂₇₀ Asn Asp
 Ile Gly Trp₂₇₅ Asn Met Met Glu Leu₂₈₀ Ala Glu Met Gln Asn₂₈₅ Pro Gln Arg
 Glu Met₂₉₀ Phe Ala Cys Phe Ala₂₉₅ Glu Ala Met Ile Leu₃₀₀ Glu Phe Glu Lys
 Cys₃₀₅ His Thr Asn Phe Ser₃₁₀ Trp Gly Arg Asn Asn₃₁₅ Ile Ser Leu Glu Lys₃₂₀
 Met Glu Phe Ile Gly₃₂₅ Ala Ala Ser Val Lys₃₃₀ His Gly Phe Ser Ala₃₃₅ Ile
 Gly Leu Asp Lys₃₄₀ His Pro Lys Val Leu Ala Val

<210> 35

<211> 1053

ES 2 600 354 T3

<212> ADN

<213> Gloeobacter violaceus

<400> 35

```

atgtttggcc tgatcggaca cttgaccaat ctttcccatg cccagcgggt cgccccgcgac      60
ctgggctacg acgagtatgc aagccacgac ctcgaattct ggtgcatggc cctccccag      120
gcggtcgatg aaatcacgat caccagcgtc accggtcagg tgatccacgg tcagtacgtc      180
gaatcgtgct ttctgccgga gatgctcgcc cagggccgct tcaagaccgc catgcgcaag      240
atcctcaatg ccatggccct ggtccagaag cgcggcatcg acattacggc cctgggaggc      300
ttctcgtcga tcatcttcga gaatttcagc ctcgataaat tgctcaacgt ccgcgacatc      360
accctcgaca tccagcgctt caccaccggc aacaccccaca cggcctacat cttttgtcag      420
caggtcgcgagc aggggtgcgggt acgctacggc atcgatccgg ccaaagcgac cgtggcggta      480
gtcggggcca ccggcgacat cggtagcgcc gtctgccgat ggctcaccga ccgcgccggc      540
atccacgaac tcttgctggt ggccccgcgac gccgaaaggc tcgaccggct gcagcaggaa      600
ctcggcaccg gtcggatcct gccggtcgaa gaagcacttc ccaaagccga catcgtcgtc      660
tgggtcgctt cgatgaacca gggcatggcc atcgaccccg ccggcctgcg caccctctgc      720
ctgctcatcg acggcggcta cccaagaac atggccggca ccctgcagcg cccgggcatc      780
catatcctcg acggcggcat ggtcgcgac tcgctcgaca tcgactggca gatcatgctg      840
tttctaaatg tgcccaacct cgccccccag ttcttcgctt gcttcgccga gtcgatgctg      900
ctggaattcg aagggcttca cttcaatttt tcctggggcc gcaaccacat caccgtcgcg      960
aagatggccc agatcggctc gctgtctaaa aaacatggct ttcgtcccct gcttgaacct      1020
agtcagcgca gcggcgaact cgtacacgga taa      1053
    
```

5 <210> 36

<211> 350

<212> PRT

<213> Gloeobacter violaceus

<400> 36

```

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

10

ES 2 600 354 T3

- . ..

Ile Leu Asn Ala Met Ala Leu Val Gln Lys Arg Gly Ile Asp Ile Thr
 85 90 95

Ala Leu Gly Gly Phe Ser Ser Ile Ile Phe Glu Asn Phe Ser Leu Asp
 100 105 110

Lys Leu Leu Asn Val Arg Asp Ile Thr Leu Asp Ile Gln Arg Phe Thr
 115 120 125

Thr Gly Asn Thr His Thr Ala Tyr Ile Leu Cys Gln Gln Val Glu Gln
 130 135 140

Gly Ala Val Arg Tyr Gly Ile Asp Pro Ala Lys Ala Thr Val Ala Val
 145 150 155 160

Val Gly Ala Thr Gly Asp Ile Gly Ser Ala Val Cys Arg Trp Leu Thr
 165 170 175

Asp Arg Ala Gly Ile His Glu Leu Leu Leu Val Ala Arg Asp Ala Glu
 180 185 190

Arg Leu Asp Arg Leu Gln Gln Glu Leu Gly Thr Gly Arg Ile Leu Pro
 195 200 205

Val Glu Glu Ala Leu Pro Lys Ala Asp Ile Val Val Trp Val Ala Ser
 210 215 220

Met Asn Gln Gly Met Ala Ile Asp Pro Ala Gly Leu Arg Thr Pro Cys
 225 230 235 240

Leu Leu Ile Asp Gly Gly Tyr Pro Lys Asn Met Ala Gly Thr Leu Gln
 245 250 255

Arg Pro Gly Ile His Ile Leu Asp Gly Gly Met Val Glu His Ser Leu
 260 265 270

Asp Ile Asp Trp Gln Ile Met Ser Phe Leu Asn Val Pro Asn Pro Ala
 275 280 285

Arg Gln Phe Phe Ala Cys Phe Ala Glu Ser Met Leu Leu Glu Phe Glu
 290 295 300

Gly Leu His Phe Asn Phe Ser Trp Gly Arg Asn His Ile Thr Val Glu
 305 310 315 320

Lys Met Ala Gln Ile Gly Ser Leu Ser Lys Lys His Gly Phe Arg Pro
 325 330 335

Leu Leu Glu Pro Ser Gln Arg Ser Gly Glu Leu Val His Gly
 340 345 350

<210> 37

ES 2 600 354 T3

<211> 1020

<212> ADN

<213> *Nostoc punctiforme*

<400> 37

```

atgtttggtc taattggaca tctgactagt ttagaacacg ctcaagccgt agcccaagaa      60
ttgggatacc cagaatatgc cgatcaaggg ctagactttt ggtgcagcgc cccgccgcaa      120
attgtcgata gtattattgt caccagtgtt actgggcaac aaattgaagg acgatatgta      180
gaatcttgct ttttgccgga aatgctagct agtcgccgca tcaaagccgc aacacggaaa      240
atcctcaacg ctatggccca tgcacagaag cacggcatta acatcacagc tttaggcgga      300
ttttcctcga ttatTTTTga aaactttaag ttagagcagt ttagccaagt ccgaaatattc      360
aagctagagt ttgaacgctt caccacagga aacacgcata ctgcctacat tatttGtaag      420
caggtggaag aagcatccaa acaactggga attaattctat caaacgcgac tgttgcggta      480
tgtggagcaa ctggggatat tggtagtgcc gttacacgct ggctagatgc gagaacagat      540
gtccaagaac tcctgctaatt cgcccgcgat caagaacgtc tcaaagagtt gcaaggcgaa      600
ctggggcggg ggaaaatcat gggtttgaca gaagcactac cccaagccga tgttgtagtt      660
tgggttgcta gtatgcccag aggcgtggaa attgacccca ccactttgaa acaaccctgt      720
ttgttgattg atgggtggcta tcctaaaaac ttagcaacaa aaattcaata tcctggcgta      780
cacgtgttaa atgggtggat ttagagcat tcctggata ttgactggaa aattatgaaa      840
atagtcaata tggacgtgcc agcccgtcag ttgtttgcct gttttgccga atcaatgcta      900
ctggaatttg agaagttata cacgaacttt tcgtggggac ggaatcagat taccgtagat      960
aaaatggagc agattggccg ggtgtcagta aaacatggat ttagaccggt gttggtttag      1020

```

5

<210> 38

<211> 339

<212> PRT

<213> *Nostoc punctiforme*

10 <400> 38

```

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

```

ES 2 600 354 T3

Ile Leu Asn Ala Met Ala His Ala Gln Lys His Gly Ile Asn Ile Thr
 85 90 95
 Ala Leu Gly Gly Phe Ser Ser Ile Ile Phe Glu Asn Phe Lys Leu Glu
 100 105 110
 Gln Phe Ser Gln Val Arg Asn Ile Lys Leu Glu Phe Glu Arg Phe Thr
 115 120 125
 Thr Gly Asn Thr His Thr Ala Tyr Ile Ile Cys Lys Gln Val Glu Glu
 130 135 140
 Ala Ser Lys Gln Leu Gly Ile Asn Leu Ser Asn Ala Thr Val Ala Val
 145 150 155 160
 Cys Gly Ala Thr Gly Asp Ile Gly Ser Ala Val Thr Arg Trp Leu Asp
 165 170 175
 Ala Arg Thr Asp Val Gln Glu Leu Leu Leu Ile Ala Arg Asp Gln Glu
 180 185 190
 Arg Leu Lys Glu Leu Gln Gly Glu Leu Gly Arg Gly Lys Ile Met Gly
 195 200 205
 Leu Thr Glu Ala Leu Pro Gln Ala Asp Val Val Val Trp Val Ala Ser
 210 215 220
 Met Pro Arg Gly Val Glu Ile Asp Pro Thr Thr Leu Lys Gln Pro Cys
 225 230 235 240
 Leu Leu Ile Asp Gly Gly Tyr Pro Lys Asn Leu Ala Thr Lys Ile Gln
 245 250 255
 Tyr Pro Gly Val His Val Leu Asn Gly Gly Ile Val Glu His Ser Leu
 260 265 270
 Asp Ile Asp Trp Lys Ile Met Lys Ile Val Asn Met Asp Val Pro Ala
 275 280 285
 Arg Gln Leu Phe Ala Cys Phe Ala Glu Ser Met Leu Leu Glu Phe Glu
 290 295 300
 Lys Leu Tyr Thr Asn Phe Ser Trp Gly Arg Asn Gln Ile Thr Val Asp
 305 310 315 320
 Lys Met Glu Gln Ile Gly Arg Val Ser Val Lys His Gly Phe Arg Pro
 325 330 335
 Leu Leu Val

<210> 39

<211> 1020

ES 2 600 354 T3

<212> ADN

<213> *Anabaena variabilis*

<400> 39

```

atgtttggtc taattggaca tctgacaagt ttagaacacg ctcaagcggg agctcaagaa      60
ctgggatacc cagaatacgc cgaccaaggg ctagatTTTT ggtgcagcgc tccaccgcaa      120
atagttgacc acattaaagt tactagcatt actggtgaaa taattgaagg gaggtatgta      180
gaatcttgct ttttaccaga aatgctagcc agccgtagga ttaaagccgc aaccgcgaaa      240
gtcctcaatg ctatggctca tgctcaaaaa catggcattg acatcaccgc tttgggtggt      300
ttctcctcca ttatTTTTga aaacttcaaa ttggaacagt ttagccaagt tcgtaatgtc      360
acactagagt ttgaacgctt cactacaggc aacactcaca cagcttatat catttgctcg      420
caggtagaac aagcatcaca acaactcggc attgaactct cccaagcaac agtagctata      480
tgtggggcta ctggtgacat tggtagtgca gttactcgct ggctggatgc caaaacagac      540
gtaaaagaat tactgttaat cgcccgtaat caagaacgtc tccaagagtt gcaaagcgag      600
ttgggacgcg gtaaaatcat gagcctagat gaagcattgc ctcaagctga tattgtagtt      660
tgggtagcta gtatgcctaa aggcgtggaa attaatcctc aagttttgaa acaaccctgt      720
ttattgattg atggtggtta tccgaaaaac ttgggtacaa aagttcagta tcctggtggt      780
tatgtactga acggaggat cgtcgaacat tccctagata ttgactggaa aatcatgaaa      840
atagtcaata tggatgtacc tgcacgcaa ttatTTTgctt gTTTTcgga atctatgctc      900
ttggaatttg agaagttgta cacgaacttt tcttgggggc gcaatcagat taccgtagac      960
aaaatggagc agattggtca agcatcagtg aaacatgggt ttagaccact gctggtttag     1020

```

5 <210> 40

<211> 339

<212> PRT

<213> *Anabaena variabilis*

<400> 40

```

Met Phe Gly Leu Ile Gly His Leu Thr Ser Leu Glu His Ala Gln Ala
 1                               5                               10                               15

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

Phe Trp Cys Ser Ala Pro Pro Gln Ile Val Asp His Ile Lys Val Thr
 35                               40                               45

Ser Ile Thr Gly Glu Ile Ile Glu Gly Arg Tyr Val Glu Ser Cys Phe
 50                               55                               60

Leu Pro Glu Met Leu Ala Ser Arg Arg Ile Lys Ala Ala Thr Arg Lys
 65                               70                               75                               80

```

10

ES 2 600 354 T3

Val Leu Asn Ala Met Ala His Ala Gln Lys His Gly Ile Asp Ile Thr
 85 90 95
 Ala Leu Gly Gly Phe Ser Ser Ile Ile Phe Glu Asn Phe Lys Leu Glu
 100 105 110
 Gln Phe Ser Gln Val Arg Asn Val Thr Leu Glu Phe Glu Arg Phe Thr
 115 120 125
 Thr Gly Asn Thr His Thr Ala Tyr Ile Ile Cys Arg Gln Val Glu Gln
 130 135 140
 Ala Ser Gln Gln Leu Gly Ile Glu Leu Ser Gln Ala Thr Val Ala Ile
 145 150 155 160
 Cys Gly Ala Thr Gly Asp Ile Gly Ser Ala Val Thr Arg Trp Leu Asp
 165 170 175
 Ala Lys Thr Asp Val Lys Glu Leu Leu Leu Ile Ala Arg Asn Gln Glu
 180 185 190
 Arg Leu Gln Glu Leu Gln Ser Glu Leu Gly Arg Gly Lys Ile Met Ser
 195 200 205
 Leu Asp Glu Ala Leu Pro Gln Ala Asp Ile Val Val Trp Val Ala Ser
 210 215 220
 Met Pro Lys Gly Val Glu Ile Asn Pro Gln Val Leu Lys Gln Pro Cys
 225 230 235 240
 Leu Leu Ile Asp Gly Gly Tyr Pro Lys Asn Leu Gly Thr Lys Val Gln
 245 250 255
 Tyr Pro Gly Val Tyr Val Leu Asn Gly Gly Ile Val Glu His Ser Leu
 260 265 270
 Asp Ile Asp Trp Lys Ile Met Lys Ile Val Asn Met Asp Val Pro Ala
 275 280 285
 Arg Gln Leu Phe Ala Cys Phe Ala Glu Ser Met Leu Leu Glu Phe Glu
 290 295 300
 Lys Leu Tyr Thr Asn Phe Ser Trp Gly Arg Asn Gln Ile Thr Val Asp
 305 310 315 320
 Lys Met Glu Gln Ile Gly Gln Ala Ser Val Lys His Gly Phe Arg Pro
 325 330 335
 Leu Leu Val

<210> 41

<211> 1026

ES 2 600 354 T3

<212> ADN

<213> *Synechococcus elongatus*

<400> 41

```

atgttcggtc ttatcgggtca tctcaccagt ttggagcagg cccgcgacgt ttctcgcagg      60
atgggctacg acgaatacgc cgatcaagga ttggagtttt ggagtagcgc tcctcctcaa      120
atcgttgatg aaatcacagt caccagtgcc acaggcaagg tgattcacgg tcgctacatc      180
gaatcgtggt tcttgccgga aatgctggcg gcgcccgtct tcaaacagc cacgcgcaaa      240
gttctcaatg ccatgtccca tgcccaaaaa cacggcatcg acatctcggc cttggggggc      300
tttacctcga ttattttcga gaatttcgat ttggccagtt tgcggcaagt gcgcgacact      360
accttggagt ttgaacgggt caccaccggc aatactcaca cggcctacgt aatctgtaga      420
cagggtggaag ccgctgctaa aacgctgggc atcgacatta cccaagcgac agtagcggtt      480
gtcggcgcga ctggcgatat cggtagcgtt gtctgccgct ggctcgacct caaactgggt      540
gtcgggtgatt tgatcctgac ggcgcgcaat caggagcgtt tggataacct gcaggctgaa      600
ctcggccggg gcaagattct gcccttggaa gccgctctgc cggaagctga ctttatcgtg      660
tgggtcgcca gtatgcctca gggcgtagt atcgaccag caaccctgaa gcaaccctgc      720
gtcctaatcg acgggggcta ccccaaaaac ttgggcagca aagtccaagg tgagggcatc      780
tatgtcctca atggcggggt agttgaacat tgcttcgaca tcgactggca gatcatgtcc      840
gctgcagaga tggcgcggcc cgagcgccag atgtttgcct gctttgccga ggcgatgctc      900
ttggaatttg aaggctggca tactaacttc tcctggggcc gcaaccaa at cacgatcgag      960
aagatggaag cgatcgggtg ggcacgggt cgccacggct tccaaccctt ggcattggca     1020
atttga                                             1026
    
```

5 <210> 42

<211> 340

<212> PRT

<213> *Synechococcus elongatus*

<400> 42

```

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

10

ES 2 600 354 T3

Val Leu Asn Ala Met Ser His Ala Gln Lys His Gly Ile Asp Ile Ser
85 90 95

Ala Leu Gly Gly Phe Thr Ser Ile Ile Phe Glu Asn Phe Asp Leu Ala
100 105 110

Ser Leu Arg Gln Val Arg Asp Thr Thr Leu Glu Phe Glu Arg Phe Thr
115 120 125

Thr Gly Asn Thr His Thr Ala Tyr Val Ile Cys Arg Gln Val Glu Ala
130 135 140

Ala Ala Lys Thr Leu Gly Ile Asp Ile Thr Gln Ala Thr Val Ala Val
145 150 155 160

Val Gly Ala Thr Gly Asp Ile Gly Ser Ala Val Cys Arg Trp Leu Asp
165 170 175

Leu Lys Leu Gly Val Gly Asp Leu Ile Leu Thr Ala Arg Asn Gln Glu
180 185 190

Arg Leu Asp Asn Leu Gln Ala Glu Leu Gly Arg Gly Lys Ile Leu Pro
195 200 205

Leu Glu Ala Ala Leu Pro Glu Ala Asp Phe Ile Val Trp Val Ala Ser
210 215 220

Met Pro Gln Gly Val Val Ile Asp Pro Ala Thr Leu Lys Gln Pro Cys
225 230 235 240

Val Leu Ile Asp Gly Gly Tyr Pro Lys Asn Leu Gly Ser Lys Val Gln
245 250 255

Gly Glu Gly Ile Tyr Val Leu Asn Gly Gly Val Val Glu His Cys Phe
260 265 270

Asp Ile Asp Trp Gln Ile Met Ser Ala Ala Glu Met Ala Arg Pro Glu
275 280 285

Arg Gln Met Phe Ala Cys Phe Ala Glu Ala Met Leu Leu Glu Phe Glu
290 295 300

Gly Trp His Thr Asn Phe Ser Trp Gly Arg Asn Gln Ile Thr Ile Glu
305 310 315 320

Lys Met Glu Ala Ile Gly Glu Ala Ser Val Arg His Gly Phe Gln Pro
325 330 335

Leu Ala Leu Ala
340

<210> 43

<211> 1020

ES 2 600 354 T3

<212> ADN

<213> *Nostoc* sp.

<400> 43

```

atgtttggtc taattggaca tctgacaagt ttagaacacg ctcaagcggg agctcaagaa      60
ctgggatacc cagaatacgc cgaccaaggg ctagatTTTT ggtgtagcgc tccaccgcaa      120
atagttgacc acattaaagt tactagtatt actggtgaaa taattgaagg gaggtatgta      180
gaatcttgct ttttaccgga gatgctagcc agtcgctcga ttaaagccgc aaccgcgaaa      240
gtcctcaatg ctatggctca tgctcaaaag aatggcattg atatcacagc tttgggtggt      300
ttctcctcca ttatTTTTga aaactTTTaaa ttggagcagt ttagccaagt tcgtaatgtg      360
acactagagt ttgaacgctt cactacaggc aacactcaca cagcatatat tatttgtcgg      420
caggtagaac aagcatcaca acaactcggc attgaactct cccaagcaac agtagctata      480
tgtggggcta ctggtgatat tggtagtgca gttactcgtt ggctggatgc taaaacagac      540
gtgaaagaat tgctgttaat cgcccgtaat caagaacgct tccaagagtt gcaaagcgag      600
ctgggacgcg gtaaaatcat gagccttgat gaagcactgc cccaagctga tatcgtagtt      660
tgggtagcca gtatgcctaa aggtgtggaa attaatcctc aagTTTTgaa gcaaccctgt      720
ttgctgattg atggggggtta tccgaaaaac ttgggtacaa aagttcagta tcttgggtgt      780
tatgtactga acggcgggat cgtcgaacat tcgctggata ttgactggaa aatcatgaaa      840
atagtcaata tggatgtacc tgcacgccaa ttatTTTgctt gTTTTgcgga atctatgctc      900
ttggaatttg agaagttgta cacgaacttt tcttggggggc gcaatcagat taccgtagac      960
aaaatggagc agattggtca agcatcagtg aaacatgggt ttagaccact gctggtttag     1020
    
```

5 <210> 44

<211> 339

<212> PRT

<213> *Nostoc* sp.

<400> 44

```

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

ES 2 600 354 T3

Val Leu Asn Ala Met 85 Ala His Ala Gln Lys 90 Asn Gly Ile Asp Ile Thr
 Ala Leu Gly Gly 100 Phe Ser Ser Ile Ile 105 Phe Glu Asn Phe Lys 110 Leu Glu
 Gln Phe Ser 115 Gln Val Arg Asn Val 120 Thr Leu Glu Phe Glu 125 Arg Phe Thr
 Thr Gly 130 Asn Thr His Thr Ala 135 Tyr Ile Ile Cys Arg 140 Gln Val Glu Gln
 Ala Ser Gln Gln Leu Gly 150 Ile Glu Leu Ser Gln 155 Ala Thr Val Ala Ile 160
 Cys Gly Ala Thr Gly 165 Asp Ile Gly Ser Ala 170 Val Thr Arg Trp Leu Asp 175
 Ala Lys Thr Asp 180 Val Lys Glu Leu Leu 185 Leu Ile Ala Arg Asn 190 Gln Glu
 Arg Leu Gln Glu Leu Gln Ser Glu 200 Leu Gly Arg Gly Lys 205 Ile Met Ser
 Leu Asp 210 Glu Ala Leu Pro Gln 215 Ala Asp Ile Val Val 220 Trp Val Ala Ser
 Met Pro Lys Gly Val Glu 230 Ile Asn Pro Gln Val 235 Leu Lys Gln Pro Cys 240
 Leu Leu Ile Asp Gly 245 Gly Tyr Pro Lys Asn 250 Leu Gly Thr Lys Val 255 Gln
 Tyr Pro Gly Val 260 Tyr Val Leu Asn Gly 265 Gly Ile Val Glu His 270 Ser Leu
 Asp Ile Asp 275 Trp Lys Ile Met Lys 280 Ile Val Asn Met Asp 285 Val Pro Ala
 Arg Gln Leu Phe Ala Cys Phe 295 Ala Glu Ser Met Leu 300 Leu Glu Phe Glu
 Lys Leu Tyr Thr Asn Phe 310 Ser Trp Gly Arg Asn 315 Gln Ile Thr Val Asp 320
 Lys Met Glu Gln Ile 325 Gly Gln Ala Ser Val 330 Lys His Gly Phe Arg 335 Pro
 Leu Leu Val

<210> 45

<211> 867

ES 2 600 354 T3

<212> ADN

<213> *Corynebacterium glutamicum*

<400> 45

```

ttgggcgtgt cgcccttaaa gcgcgctttt cgacgcgacc ccactacatt ggcttccatg      60
aacgttgaca tttcacgatc cagagagccg ctaaacgttg agctcctgaa ggaaaaattg      120
ctccaaaacg gtgacttttg ccaggtcatt tacgaaaaag tgacaggctc cactaatgct      180
gacttgctgg cacttgcaagg ttctggcgct ccaaactgga cggtgaaaac tgtcgagttt      240
caagatcatg cgcgtgggcg actcggccgc ccgtggtctg ccctgaggg ttcccaaaca      300
atcgtgtctg tgctcgttca actatctatt gatcaagtgg accggattgg cactattcca      360
ctcgcggcgg gactcgtgtg catggatgcg ttgaatgacc tcggtgtgga aggtgccgga      420
ctgaaatggc ccaacgatgt tcaaattccac ggcaagaaac tctgcggcat cctggtggaa      480
gccaccggct ttgattccac cccaacagtt gtcacggtt ggggactaa tatcagcctg      540
actaaagagg agcttcctgt tcctcatgca acttcctcgc cattggaagg tgttgaagtc      600
gacagaacca cattccttat taatatgctc acacatctgc atactcgact ggaccagtgg      660
cagggccaag gtgtggattg gctcgaatgac taccgtgcgg tatgttccag tattggccaa      720
gatgttcgag tgcttctacc tggggataaa gaactcttag gtgaagcgat cgggtgtcgcg      780
actggcggag aaattcgtgt tcgcgatgct tcgggcaccg ttcacaccct caacgccggt      840
gaaattacgc accttcgcct gcagtaa                                           867

```

5 <210> 46

<211> 810

<212> ADN

<213> Secuencia artificial

<220>

10 <221> fuente

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

<400> 46

```

atgaatggtg acattagccg ctctcgtgaa ccggtgaacg tggaaactggt gaaagaaaaa      60
ctgctgcaga acggtgattt cggtaagtg atctacgaga aggtcaccgg ctctaccaat      120
gcggacctgc tggctctggc gggcagcggc gctccaaact ggaccgtcaa gactggtgaa      180
tttcaggacc acgcccgtgg ccgtctgggt cgtccgtgga gcgcaccgga gggttcccaa      240
accatcgtca gcgttctggt ccaactgagc attgatcagg tggaccgtat tggtagatc      300
ccgctggccg caggcttggc tgttatggat gcgctgaatg atctgggcgt ggaggggtgca      360
ggcctgaaat ggccgaacga tgttcagatc cacgtaaga agttgtgcgg tattctggtt      420
gaagcaaccg gcttcgactc cactccgacc gtggttatcg gttggggtac gaatatctcg      480
ttgacgaaag aagagctgcc ggtcccgcac gcgaccagcc tggccctgga ggggtgtgaa      540
gttgaccgta cgacgttcct gattaacatg ctgacccatc tgcatacccg tctggatcag      600

```

ES 2 600 354 T3

tggcaggggc cgtctgtgga ctggctggat gactatcgcg cggttttag cagcattggc 660
 caagatgtgc gtgtcctgct gcctggtgac aaagagctgc tgggcgaggc gattggcgtg 720
 gcgaccggtg gtgagatccg tgtgcgcgac gccagcggca cgggccacac gctgaatgcg 780
 ggtgaaatca cgcattctgcg tttgcaataa 810

<210> 47

<211> 269

<212> PRT

5 <213> *Corynebacterium glutamicum*

<400> 47

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

ES 2 600 354 T3

atgaccattt cctcaccttt gattgacgtc gccaaccttc cagacatcaa caccactgcc 60
 ggcaagatcg ccgaccttaa ggctcgccgc gcggaagccc atttccccat gggtgaaaag 120
 gcagtagaga aggtccacgc tgctggacgc ctactgccc gtgagcgctt ggattactta 180
 ctcgatgagg gctccttcat cgagaccgat cagctggctc gccaccgcac caccgctttc 240
 ggcttgggcg ctaagcgtcc tgcaaccgac ggcatcgtga ccggctgggg caccattgat 300
 ggacgcgaag tctgcatctt ctgcgaggac ggaccggtat tcggtggcgc gcttgggtgag 360
 gtgtacggcg aaaagatgat caagatcatg gagctggcaa tcgacaccgg ccgcccattg 420
 atcggctttt acgaaggcgc tggcgctcgt attcaggacg gcgctgtctc cctggacttc 480
 atttcccaga ctttctacca aaacattcag gcttctggcg ttatcccaca gatctccgctc 540
 atcatgggcg catgtgcagg tggcaacgct tacggcccag ctctgaccga cttcgtggtc 600
 atggtggaca agacctcaa gatgttcggt accggcccag acgtgatcaa gaccgtcacc 660
 ggcgaggaaa tcaccagga agagcttggc ggagcaacca cccacatggt gaccgctggt 720
 aactcccact acaccgctgc gaccgatgag gaagcactgg attgggtaca ggacctggtg 780
 tccttctcc catccaaca tcgctcctac gcaccgatgg aagacttca cgaggaagaa 840
 ggcgcggtt aagaaaacat caccgctgac gatctgaagc tcgacgagat catcccagat 900
 tccgcgaccg ttccttacga cgtcccgcat gtcacgaat gcctcaccga cgatggcgaa 960
 tacctggaaa tccaggcaga ccgcgcagaa aacgttgtta ttgcattcgg ccgcatcgaa 1020
 ggccagtcg ttggctttgt tgccaaccag ccaaccagt tcgctggctg cctggacatc 1080
 gactcctctg agaaggcagc tcgcttcgtc cgcacctgcg acgcgttcaa catcccattc 1140
 gtcatgcttg tcgacgtccc cggcttcctc ccaggcgag gccaggagta cggtggcatt 1200
 ctgctcgtg gcgcaaagct gctctacgca tacggcgaag caaccgttcc aaagatcacc 1260
 gtcaccatgc gtaaggctta cggcggagcg tactgctgta tgggttcaa gggcttgggc 1320
 tctgacatca accttgcag gccaacgca cagatcgccg tcatgggcgc tgctggcgca 1380
 gttggattca tctaccgcaa ggagctcatg gcagctgat ccaagggcct cgataccgta 1440

 gctctggcta agtccttca gcgcgagtat gaagaccaca tgctcaacc gtaccacgct 1500
 gcagaacgtg gcctgatcga cgccgtgat ctgccaagcg aaaccgcgg acagatttcc 1560
 cgcaaccttc gcctgctcaa gcacaagaac gtcactcgcc ctgctcgaa gcacggcaac 1620
 atgccactgt aa 1632

<210> 49

<211> 543

5 <212> PRT

<213> *Corynebacterium glutamicum*

<400> 49

ES 2 600 354 T3

Met Thr Ile Ser Ser Pro Leu Ile Asp Val Ala Asn Leu Pro Asp Ile
 1 5 10 15

Asn Thr Thr Ala Gly Lys Ile Ala Asp Leu Lys Ala Arg Arg Ala Glu
 20 25 30

Ala His Phe Pro Met Gly Glu Lys Ala Val Glu Lys Val His Ala Ala
 35 40 45

Gly Arg Leu Thr Ala Arg Glu Arg Leu Asp Tyr Leu Leu Asp Glu Gly
 50 55 60

Ser Phe Ile Glu Thr Asp Gln Leu Ala Arg His Arg Thr Thr Ala Phe
 65 70 75 80

Gly Leu Gly Ala Lys Arg Pro Ala Thr Asp Gly Ile Val Thr Gly Trp
 85 90 95

Gly Thr Ile Asp Gly Arg Glu Val Cys Ile Phe Ser Gln Asp Gly Thr
 100 105 110

Val Phe Gly Gly Ala Leu Gly Glu Val Tyr Gly Glu Lys Met Ile Lys
 115 120 125

Ile Met Glu Leu Ala Ile Asp Thr Gly Arg Pro Leu Ile Gly Leu Tyr
 130 135 140

Glu Gly Ala Gly Ala Arg Ile Gln Asp Gly Ala Val Ser Leu Asp Phe
 145 150 155 160

Ile Ser Gln Thr Phe Tyr Gln Asn Ile Gln Ala Ser Gly Val Ile Pro
 165 170 175

Gln Ile Ser Val Ile Met Gly Ala Cys Ala Gly Gly Asn Ala Tyr Gly
 180 185 190

Pro Ala Leu Thr Asp Phe Val Val Met Val Asp Lys Thr Ser Lys Met
 195 200 205

ES 2 600 354 T3

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

ES 2 600 354 T3

Ala Leu Ala Lys Ser Phe Glu Arg Glu Tyr Glu Asp His Met Leu Asn
 485 490 495

Pro Tyr His Ala Ala Glu Arg Gly Leu Ile Asp Ala Val Ile Leu Pro
 500 505 510

Ser Glu Thr Arg Gly Gln Ile Ser Arg Asn Leu Arg Leu Leu Lys His
 515 520 525

Lys Asn Val Thr Arg Pro Ala Arg Lys His Gly Asn Met Pro Leu
 530 535 540

<210> 50

<211> 1776

<212> ADN

5 <213> *Corynebacterium glutamicum*

<400> 50

atgtcagtcg agactcgcaa gatcaccaag gttcttgtcg ctaaccgtgg tgagattgca	60
atccgcgtgt tccgtgcagc tcgagatgaa ggcacatcgat ctgtcgccgt ctacgcagag	120
ccagatgcag atgcaccatt cgtgtcatat gcagacgagg cttttgccct cggtgcccaa	180
acatccgctg agtcctacct tgtcattgac aagatcatcg atgcggcccc caagtccggc	240
gccgacgcca tccaccccgg ctacggcttc ctcgcagaaa acgctgactt cgcagaagca	300
gtcatcaacg aaggcctgat ctggattgga ccttcacctg agtccatccg ctccctcggc	360
gacaaggcca ccgctcgcca catcgcagat accgccaagg ctccaatggc tcctggcacc	420
aaggaaccag taaaagacgc agcagaagtt gtggctttcg ctgaagaatt cggctctcca	480
atcgccatca aggcagcttt cggtggcggc ggacgtggca tgaaggttgc ctacaagatg	540
gaagaagtgc ctgaccttt cgagtcgca acccgtgaag caaccgcagc gttcggccgc	600
ggcgagtgtc tcgtggagcg ctacctggac aaggcacgcc acgttgaggc tcaggtcac	660
gccgataagc acggcaacgt tgttgcgcc ggaacccgtg actgctccct gcagcgccgt	720
ttccagaagc tcgtcgaaga agcaccagca ccattcctca ccgatgacca gcgcgagcgt	780
ctccactcct ccgcgaaggc tatctgtaag gaagctggct actacggtgc aggcaccgtt	840
gagtacctcg ttggctccga cggcctgatc tccttctcag aggtcaacac ccgcctccag	900
gtggaacacc cagtcaccga agagaccacc ggcacgcacc tggccgcga aatgttccgc	960
atcgcagaag gccacgagct ctccatcaag gaagatccag ctccacgcgg ccacgcattc	1020
gagttccgca tcaacggcga agacgctggc tccaacttca tgccctgcacc aggcaagatc	1080
accagctacc gcgagccaca gggcccaggc gtccgcatgg actccggtgt cgttgaaggt	1140
tccgaaatct ccggacagtt cgactccatg ctggcaaagc tgatcgtttg gggcgacacc	1200
cgcgagcagg ctctccagcg ctcccgccgt gcacttgca agtacgttgt cgagggcatg	1260
ccaaccgtta tcccattcca ccagcacatc gtggaaaacc cagcattcgt gggcaacgac	1320
gaaggcttcg agatctacac caagtggatc gaagaggttt gggataaccc aatcgcacct	1380

ES 2 600 354 T3

tacgttgacg cttccgagct cgacgaagat gaggacaaga ccccagcaca gaaggttggt 1440
 gtggagatca acggccgctg cgttgagggt gactcccag gcatctggc actcgggtggc 1500
 accgctggtc ctaagaagaa ggccaagaag cgtcgcgag gtggtgcaaa ggctggcgta 1560
 tccggcgatg cagtggcagc tccaatgcag ggcactgtca tcaaggtaa cgtcgaagaa 1620
 ggcgctgaag tcaacgaagg cgacaccgtt gttgtcctcg aggctatgaa gatggaaaac 1680
 cctgtgaagg ctcataagtc cggaaccgta accggcctta ctgctgctgc aggcgagggt 1740
 gtcaacaagg gcgttgttct cctcgagatc aagtaa 1776

<210> 51

<211> 591

<212> PRT

5 <213> *Corynebacterium glutamicum*

<400> 51

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

ES 2 600 354 T3

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

ES 2 600 354 T3

Ser Glu Leu Asp Glu Asp Glu Asp Lys Thr Pro Ala Gln Lys Val Val
465 470 475 480

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

Ala Leu Gly Gly Thr Ala Gly Pro Lys Lys Lys Ala Lys Lys Arg Arg
500 505 510

Ala Gly Gly Ala Lys Ala Gly Val Ser Gly Asp Ala Val Ala Ala Pro
515 520 525

Met Gln Gly Thr Val Ile Lys Val Asn Val Glu Glu Gly Ala Glu Val
530 535 540

Asn Glu Gly Asp Thr Val Val Val Leu Glu Ala Met Lys Met Glu Asn
545 550 555 560

Pro Val Lys Ala His Lys Ser Gly Thr Val Thr Gly Leu Thr Val Ala
565 570 575

Ala Gly Glu Gly Val Asn Lys Gly Val Val Leu Leu Glu Ile Lys
580 585 590

<210> 52

<211> 340

<212> PRT

5 <213> *Acinetobacter* sp.

<400> 52

Met Ser Asn His Gln Ile Arg Ala Tyr Ala Ala Met Gln Ala Gly Glu
1 5 10 15

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

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

Val Ile Asn Asn Glu Trp Gln Ser Ser Val Tyr Pro Ala Val Ala Gly
50 55 60

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

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

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

Glu Lys Lys Ala Thr Ile Ile Gly His Ala Gly Gly Phe Ala Asp Lys

ES 2 600 354 T3

Glu Ala Leu Val₂₀ Pro Tyr Ser Phe Asp₂₅ Ala Gly Glu Leu Gln₃₀ Pro His
 Gln Val₃₅ Glu Val Lys Val Glu Tyr₄₀ Cys Gly Leu Cys His₄₅ Ser Asp Val
 Ser Val₅₀ Leu Asn Asn Glu Trp₅₅ His Ser Ser Val Tyr₆₀ Pro Val Val Ala
 Gly His Glu Val Ile Gly₇₀ Thr Ile Thr Gln Leu₇₅ Gly Ser Glu Ala Lys₈₀
 Gly Leu Lys Ile Gly₈₅ Gln Arg Val Gly Ile₉₀ Gly Trp Thr Ala Glu Ser
 Cys Gln Ala Cys₁₀₀ Asp Gln Cys Ile Ser₁₀₅ Gly Gln Gln Val Leu Cys Thr
 Gly Glu Asn₁₁₅ Thr Ala Thr Ile Ile₁₂₀ Gly His Ala Gly Gly₁₂₅ Phe Ala Asp
 Lys Val₁₃₀ Arg Ala Gly Trp Gln₁₃₅ Trp Val Ile Pro Leu₁₄₀ Pro Asp Glu Leu
 Asp₁₄₅ Pro Thr Ser Ala Gly₁₅₀ Pro Leu Leu Cys Gly₁₅₅ Gly Ile Thr Val Phe₁₆₀
 Asp Pro Ile Leu Lys₁₆₅ His Gln Ile Gln Ala₁₇₀ Ile His His Val Ala₁₇₅ Val
 Ile Gly Ile Gly₁₈₀ Gly Leu Gly His Met₁₈₅ Ala Ile Lys Leu Leu₁₉₀ Lys Ala
 Trp Gly Cys₁₉₅ Glu Ile Thr Ala Phe₂₀₀ Ser Ser Asn Pro Asn₂₀₅ Lys Thr Asp
 Glu Leu Lys Ala Met Gly Ala₂₁₅ Asp His Val Val Asn₂₂₀ Ser Arg Asp Asp
 Ala Glu Ile Lys Ser Gln₂₃₀ Gln Gly Lys Phe Asp₂₃₅ Leu Leu Leu Ser Thr₂₄₀
 Val Asn Val Pro Leu₂₄₅ Asn Trp Asn Ala Tyr₂₅₀ Leu Asn Thr Leu Ala₂₅₅ Pro
 Asn Gly Thr Phe₂₆₀ His Phe Leu Gly Val₂₆₅ Val Met Glu Pro Ile₂₇₀ Pro Val
 Pro Val Gly₂₇₅ Ala Leu Leu Gly Gly₂₈₀ Ala Lys Ser Leu Thr Ala Ser Pro

ES 2 600 354 T3

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

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

<210> 54

<211> 341

<212> PRT

5 <213> *Acinetobacter baileyi*

<400> 54

ES 2 600 354 T3

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

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

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

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

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

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

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

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

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

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

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

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

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

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

ES 2 600 354 T3

gctgcatcc ttatcgaacg ggatgctggc gcgcttcggg tcctcagatt tggtcacggc 720
cttcatcgag gcaaaacccg ccaggctcaa cggggtgata cctgcttcgc taccaccaga 780
gatcataacg tcgctataac caaacttaat gttacggaag gactcaccaa tgctgttggt 840
cgcgctcgca catgcggtga caatggctgt gcaaatacct ttagcgccat aacgaatcgc 900
cagattaccg cttgccatat tcgcaatgat catcgggaata gtcatagggc tcacacgacc 960
cggacctttg gtaatcagct tttcatcctg cttctcaatg gtgccgatgc cgccaatgcc 1020
gctaccaaca atgacgccga aacgattctt atcaatcgac tccaggcca gtttgctgtc 1080
cttgattgcc tcatccgccg caacgatcgc aaactggcta aaacggcca tacggttcgc 1140
ctcacgcttg tcgataaagt cctccggggg gaagtcctc acttcggcag ccagcttaac 1200
tttgaaatcg gttgctcaa acgctttgat cttgtcaatg ccacatttac cctctttgat 1260
gctgcaccag aagctatcag cgttgttacc caccggcgtc actgcaccaa taccgtaat 1320
gacaacgcgg cgattcattt tgttgctcc ttttagaacg cggaagtatc ctggaacaaa 1380
ccgactttca aatcgtgtgc ggtatagatc aggcgaccat ccaccagaac ctcaccgtcc 1440
gccaggccca tgatcaggcg acggtttacg atacgtttga aatgaatacg ataggtgact 1500
ttcctggctg tcggcagaac ctggccggtg aatttactt cgcccacgcc cagagcgcg 1560
cctttgcctt cgccgcccaa ccagcccagg tagaatcca ccaattgcca catagcatcc 1620
agaccagac aaccgggcat caccggatcg ccgataaagt ggcattccga gaaccataga 1680
tccggattga tatccagctc ggettcgaca tagcctttgt cgaaattgcc gcccgtttcg 1740
gtcatcttaa cgacgcggtc catcatcagc atgttcggtg caggaggtg cggccctta 1800
gcgcaaaca gttcaccacg accagaggca agaaggctt cttttgtata ggattcgcgt 1860
ttatctacca tgttttatgt aaacctaaa attaaacct gtacattccg ccggtgacgt 1920
gcagagtctc accagtgatg taactcgctt cgtcagaggc taaaaatgca accgcactgg 1980
cgatttctg agcgccgccg aggcgacccg caggcacctg cgccaggata cccgcacgct 2040
gatcgtcaga cagcgcacgc gtcattgctg tttcaataaa acccgagcc acaacattga 2100
cagtaatacc acgggacgca acttcacgcg ccagtgattt actgaaaccg atcaggcccc 2160
ctttcgccgc agcgtagttt gcctgacctg catttccat ggtaccaacc acagaaccaa 2220
tagtgataat gcgaccacaa cgcttttca tcatagcgcg cattaccgct tttgacaggc 2280
ggaaaacgga tgataagttg gtttcgataa tatcgttcca ctcatcatct ttcattcgca 2340
tcaacagatt atcacgagtg ataccggcat tattaaccag gatatccact tcaccaaatt 2400
ctgcgcaat attttccaga acagattcaa tagatgcagg atcggtcaca ttcaacatca 2460
aacctttccc gttagcacct aaatagtcgc taatgttctt cgcaccattt tctggtcg 2520
cagtcccgat aactttcgcg ccgcgggcaa cgagagtctc tgcaattgcy cggcctatgc 2580
cacggcttgc accagtcacc agcgcaatct ttccttcaa gctcatggtt ttcctcttt 2640
attgcgtaag tgccgcagac agcgccgccg gctcgttcag cgccgacgct gtcagggtgt 2700

ES 2 600 354 T3

cgacaatacgt tttcgtcaga ccagtgagga ctttacctgg acccacttca taaagatggt 2760
 caacgccctg cgccgcgata aattccacgc tcttcgtcca ctgtaccgga ttgtacaact 2820
 ggcgaccag cgcacgcgg atagcggcgg catcggtttc acatttcacg tcaacgttgt 2880
 tcaactaccg caccgttggc gcgctaaagg taattttggc taattcaacc gccagcttat 2940
 ctgccgctgg tttcatcagc gcgcagtgcg acggtacgct caccggcagc ggcagcgcgc 3000
 gtttcgcgcc agcggcttta caggctgcgc ccgcacgttc taccgcctct ttatgcccgg 3060
 cgataaccac ctgtcccggc gagttaaagt taaccggcga aacaacctgc ctttcggcag 3120
 attcttcaca ggcttttagca atagaggcat catccagccc gatgatcgca gacatgccgc 3180
 cagtgccctc cggaccgct tcctgcatga atttaccgcg catttccacc agacgaacgg 3240
 catcagcaaa gttgatgacg ccagcgcgca ccagcgcgga atattcgccc aggctgtgac 3300
 ctgccattaa cgcaggcatt ttaccgccct gctgctgcca aacgcgcca agcgcgacgg 3360
 aagcggttaa taacgccggc tgcgtctgcc aggttttatt cagttcttcc gctggacctt 3420
 gctgggtgag cgcccacaga tcatatcca gagccgcaga agcttcagca aacgtttctt 3480
 ctacgatagg gtaatttgcc gccatctcgg ccaacatccc aacgctctga gaaccctgac 3540
 cggggaacac aaatgcaaat tgcgtcatgt ttaaactcct atactagaaa cgaatcagcg 3600
 cggagcccca ggtgaatcca cccccgaagg cttcaagcaa taccagctga ccggctttta 3660
 ttcgcccgtc acgcacggct tcatccagcg cgcacggcac agaagccgcg gaggtattgc 3720
 cgtgcctgtc cagcgtgacg acgacattgt ccatcgacat gccgagtttt ttcgctgtcg 3780
 cgctaatgat acgcaggtta gcctgatgcg gcaccagcca atcgagttct gagcgatcca 3840
 ggttattagc cgccagcgtc tcatcgacaa tatgcgccag ttcagtgacc gccactttaa 3900
 agacttcatt gcccgccatt gtcaggtaaa tcgggttatc cggatttacg cgatcggcat 3960
 tcggcagggt cagtaattca ccgtaacggc catcggcatg aagatgagtg gagataatac 4020
 ccggttcttc agaagcgctc agtacggccg cgcttgcgcc atcgccgaaa ataatgatcg 4080
 taccgcgatc gccaggatcg caagtgcggg ctaatacatc ggaaccgacc accagcgcgt 4140
 gtttaaccgc gccggattta acgtactggt cggcgatgct taacgcgtag gtgaaacctg 4200
 cgcacgctgc cgcgacatca aacgccggc aacctttaat accgagcata ctttgaatct 4260
 gacatgccgc gcttggaaat gcatgcgctg ctgatgtggt agccaccaca atcaagccaa 4320
 tttggtcttt atcgatcccc gccatctcaa tcgcgcgatt cgcagcggta aagcccatcg 4380
 tcgcgacagt ttcattcggc gcggcgatat ggcggtttacg aataacctgta cgagtgacaa 4440
 tccactcgtc agaggctctca accatTTTTT ccagatcggc gttagtccgc acttgttcgg 4500
 gcagatagct gccagtacca ataactctcg tatacatgta cgctcagtca ctaaattact 4560
 cgatatcaat cacatcaaat tcgacttctg gattgacgct agcatcgtaa tcaatgcctt 4620
 caatgccaaa gccaaacagc ttgatgaact cttctttgta catgtcgtaa tcggtcagct 4680
 cacgcagggt ctctgtggtg atttgtggcc acagatcacg gcagtgctgc tgaatgtcat 4740
 cacgcagttc ccagtcattc aaacgcagac gattgtgatc atccacttcc ggcgctgaac 4800
 catct 4805

ES 2 600 354 T3

<210> 56

<211> 239

<212> PRT

<213> *Escherichia coli*

5 <400> 56

Met Val Ile Lys Ala Gln Ser Pro Ala Gly Phe Ala Glu Glu Tyr Ile
1 5 10 15

Ile Glu Ser Ile Trp Asn Asn Arg Phe Pro Pro Gly Thr Ile Leu Pro
20 25 30

Ala Glu Arg Glu Leu Ser Glu Leu Ile Gly Val Thr Arg Thr Thr Leu
35 40 45

Arg Glu Val Leu Gln Arg Leu Ala Arg Asp Gly Trp Leu Thr Ile Gln
50 55 60

His Gly Lys Pro Thr Lys Val Asn Asn Phe Trp Glu Thr Ser Gly Leu
65 70 75 80

Asn Ile Leu Glu Thr Leu Ala Arg Leu Asp His Glu Ser Val Pro Gln
85 90 95

Leu Ile Asp Asn Leu Leu Ser Val Arg Thr Asn Ile Ser Thr Ile Phe
100 105 110

Ile Arg Thr Ala Phe Arg Gln His Pro Asp Lys Ala Gln Glu Val Leu
115 120 125

Ala Thr Ala Asn Glu Val Ala Asp His Ala Asp Ala Phe Ala Glu Leu
130 135 140

Asp Tyr Asn Ile Phe Arg Gly Leu Ala Phe Ala Ser Gly Asn Pro Ile
145 150 155 160

Tyr Gly Leu Ile Leu Asn Gly Met Lys Gly Leu Tyr Thr Arg Ile Gly
165 170 175

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

Tyr His Lys Leu Ser Ala Leu Cys Ser Glu Gly Ala His Asp Gln Val
195 200 205

Tyr Glu Thr Val Arg Arg Tyr Gly His Glu Ser Gly Glu Ile Trp His
210 215 220

Arg Met Gln Lys Asn Leu Pro Gly Asp Leu Ala Ile Gln Gly Arg
225 230 235

<210> 57

ES 2 600 354 T3

<211> 342

<212> PRT

<213> *Synechococcus elongatus*

<400> 57

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

5

ES 2 600 354 T3

Cys Val Leu Ile Asp Gly Gly Tyr Pro Lys Asn Leu Gly Ser Lys Val
 240 245 250 255
 Gln Gly Glu Gly Ile Tyr Val Leu Asn Gly Gly Val Val Glu His Cys
 260 265 270
 Phe Asp Ile Asp Trp Gln Ile Met Ser Ala Ala Glu Met Ala Arg Pro
 275 280 285
 Glu Arg Gln Met Phe Ala Cys Phe Ala Glu Ala Met Leu Leu Glu Phe
 290 295 300
 Glu Gly Trp His Thr Asn Phe Ser Trp Gly Arg Asn Gln Ile Thr Ile
 305 310 315
 Glu Lys Met Glu Ala Ile Gly Glu Ala Ser Val Arg His Gly Phe Gln
 320 325 330 335
 Pro Leu Ala Leu Ala Ile
 340

<210> 58

<211> 342

<212> PRT

5 <213> *Synechococcus elongatus*

<400> 58

ES 2 600 354 T3

Met Ala Phe Gly Leu Ile Gly His Leu Thr Ser Leu Glu Gln Ala Arg
 1 5 10 15

Asp Val Trp Arg Arg Leu Gly Tyr Asp Glu Tyr Ala Asp Gln Gly Leu
 20 25 30

Glu Phe Trp Ser Ser Ala Pro Pro Gln Ile Val Asp Glu Ile Thr Val
 35 40 45

Thr Ser Ala Thr Gly Lys Val Ile His Gly Arg Tyr Ile Glu Ser Gly
 50 55 60

Phe Leu Pro Glu Met Leu Ala Ala Arg Arg Phe Lys Thr Ala Thr Arg
 65 70 75 80

Lys Val Leu Asn Ala Met Ser His Ala Gln Lys His Gly Ile Asp Ile
 85 90 95

Ser Ala Leu Gly Gly Phe Thr Ser Ile Ile Phe Glu Asn Phe Asp Leu
 100 105 110

Ala Lys Leu Arg Gln Val Arg Asp Thr Thr Leu Glu Phe Glu Arg Phe
 115 120 125

Thr Thr Gly Asn Thr His Thr Ala Tyr Val Ile Cys Arg Gln Val Glu

ES 2 600 354 T3

Glu Phe Trp Ser Ser Ala Pro Pro Gln Ile Val Asp Glu Ile Thr Val
 35 40 45
 Thr Ser Ala Thr Gly Lys Val Ile His Gly Arg Tyr Ile Glu Ser Gly
 50 55 60
 Phe Leu Pro Glu Met Leu Ala Ala Arg Arg Phe Lys Thr Ala Thr Arg
 65 70 75 80
 Lys Val Leu Asn Ala Met Ser His Ala Gln Lys His Gly Ile Asp Ile
 85 90 95
 Ser Ala Leu Gly Gly Phe Thr Ser Ile Ile Phe Glu Asn Phe Asp Leu
 100 105 110
 Ala Lys Leu Arg Gln Val Arg Asp Thr Thr Leu Glu Phe Glu Arg Phe
 115 120 125
 Thr Thr Gly Asn Thr His Thr Ala Tyr Val Ile Cys Arg Gln Val Glu
 130 135 140
 Ala Ala Ala Lys Thr Leu Gly Ile Asp Ile Ala Gln Ala Thr Val Ala
 145 150 155 160
 Val Val Gly Ala Thr Gly Asp Ile Gly Ser Ala Val Cys Arg Trp Leu
 165 170 175
 Asp Leu Lys Leu Gly Val Gly Asp Leu Ile Leu Thr Ala Arg Asn Gln
 180 185 190
 Glu Arg Leu Asp Asn Leu Gln Ala Glu Leu Gly Arg Gly Lys Ile Leu
 195 200 205
 Pro Leu Glu Ala Ala Leu Pro Glu Ala Asp Phe Ile Val Trp Val Ala
 210 215 220
 Ser Met Pro Gln Gly Val Val Ile Asp Pro Ala Thr Leu Lys Gln Pro
 225 230 235 240
 Cys Val Leu Ile Asp Gly Gly Tyr Pro Lys Asn Leu Gly Ser Lys Val
 245 250 255
 Gln Gly Glu Gly Ile Tyr Val Leu Asn Gly Gly Val Val Glu His Cys
 260 265 270
 Phe Asp Ile Asp Trp Gln Ile Met Ser Leu Ala Glu Met Ala Arg Pro
 275 280 285
 Glu Arg Gln Met Phe Ala Cys Phe Ala Glu Ala Met Leu Leu Glu Phe
 290 295 300

ES 2 600 354 T3

— . ..
Glu Gly Trp His Thr Asn Phe Ser Trp Gly Arg Asn Gln Ile Thr Ile
305 310 315 320

Glu Lys Met Glu Ala Ile Gly Glu Ala Ser Val Arg His Gly Phe Gln
325 330 335

Pro Leu Ala Leu Ala Ile
340

<210> 60

<211> 342

<212> PRT

5 <213> *Synechococcus elongatus*

<400> 60

ES 2 600 354 T3

Met Ala Phe Gly Leu Ile Gly His Leu Thr Ser Leu Glu Gln Ala Arg
1 5 10 15

Leu Val Trp Arg Arg Leu Gly Tyr Asp Glu Tyr Ala Asp Gln Gly Leu
20 25 30

Glu Phe Trp Ser Ser Ala Pro Pro Gln Ile Val Asp Glu Ile Thr Val
35 40 45

Thr Ser Ala Thr Gly Lys Val Ile His Gly Arg Tyr Ile Glu Ser Gly
50 55 60

Phe Leu Pro Glu Met Leu Ala Ala Arg Arg Phe Lys Thr Ala Thr Arg
65 70 75 80

Lys Val Leu Asn Ala Met Ser His Ala Gln Lys His Gly Ile Asp Ile
85 90 95

Ser Ala Leu Gly Gly Phe Thr Ser Ile Ile Phe Glu Asn Phe Asp Leu
100 105 110

Ala Lys Leu Arg Gln Val Arg Asp Thr Thr Leu Glu Phe Glu Arg Phe
115 120 125

Thr Thr Gly Asn Thr His Thr Ala Tyr Val Ile Cys Arg Gln Val Glu
130 135 140

Ala Ala Ala Lys Thr Leu Gly Ile Asp Ile Ala Gln Ala Thr Val Ala
145 150 155 160

Val Val Gly Ala Thr Gly Asp Ile Gly Ser Ala Val Cys Arg Trp Leu
165 170 175

Asp Leu Lys Leu Gly Val Gly Asp Leu Ile Leu Thr Ala Arg Asn Gln
180 185 190

Glu Arg Leu Asp Asn Leu Gln Ala Glu Leu Gly Arg Gly Lys Ile Leu

ES 2 600 354 T3

Ser Ala Leu Gly Gly Phe Thr Ser Ile Ile Phe Glu Asn Phe Asp Leu
100 105 110

Ala Lys Leu Arg Gln Val Arg Asp Thr Thr Leu Glu Phe Glu Arg Phe
115 120 125

Thr Thr Gly Asn Thr His Thr Ala Tyr Val Ile Cys Arg Gln Val Glu
130 135 140

Ala Ala Ala Lys Thr Leu Gly Ile Asp Ile Thr Leu Ala Thr Val Ala
145 150 155 160

Val Val Gly Ala Thr Gly Asp Ile Gly Ser Ala Val Cys Arg Trp Leu
165 170 175

Asp Leu Lys Leu Gly Val Gly Asp Leu Ile Leu Thr Ala Arg Asn Gln
180 185 190

Glu Arg Leu Asp Asn Leu Gln Ala Glu Leu Gly Arg Gly Lys Ile Leu
195 200 205

Pro Leu Glu Ala Ala Leu Pro Glu Ala Asp Phe Ile Val Trp Val Ala
210 215 220

Ser Met Pro Gln Gly Val Val Ile Asp Pro Ala Thr Leu Lys Gln Pro
225 230 235 240

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

Gln Gly Glu Gly Ile Tyr Val Leu Asn Gly Gly Val Val Glu His Cys
260 265 270

Phe Asp Ile Asp Trp Gln Ile Met Ser Leu Ala Glu Met Ala Arg Pro
275 280 285

Glu Arg Gln Met Phe Ala Cys Phe Ala Glu Ala Met Leu Leu Glu Phe
290 295 300

Glu Gly Trp His Thr Asn Phe Ser Trp Gly Arg Asn Gln Ile Thr Ile
305 310 315 320

Glu Lys Met Glu Ala Ile Gly Glu Ala Ser Val Arg His Gly Phe Gln
325 330 335

Pro Leu Ala Leu Ala Ile
340

<210> 62

<211> 342

<212> PRT

5 <213> *Synechococcus elongatus*

ES 2 600 354 T3

<400> 62

```

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

```

ES 2 600 354 T3

260 265 270

Phe Asp Ile Asp Trp Gln Ile Met Ser Leu Ala Glu Met Ala Arg Pro
 275 280 285

Glu Arg Gln Met Phe Ala Cys Phe Ala Glu Ala Met Leu Leu Glu Phe
 290 295 300

Glu Gly Trp His Thr Asn Phe Ser Trp Gly Arg Asn Gln Ile Thr Ile
 305 310 315 320

Glu Lys Met Glu Ala Ile Gly Glu Ala Ser Val Arg His Gly Phe Gln
 325 330 335

Pro Leu Ala Leu Ala Ile
 340

<210> 63

<211> 342

<212> PRT

5 <213> *Synechococcus elongatus*

<400> 63

Met Ala Phe Gly Leu Ile Gly His Ala Thr Ser Leu Glu Gln Ala Arg
 1 5 10 15

Leu Val Trp Arg Arg Met Gly Tyr Val Glu Tyr Ala Asp Gln Gly Leu
 20 25 30

Glu Phe Trp Ser Ser Ala Pro Pro Gln Ile Val Asp Glu Ile Thr Val
 35 40 45

Thr Ser Ala Thr Gly Lys Val Ile His Gly Arg Tyr Ile Glu Ser Gly
 50 55 60

Phe Leu Pro Glu Met Leu Ala Ala Arg Arg Phe Lys Thr Ala Thr Arg
 65 70 75 80

Lys Val Leu Asn Ala Met Ser His Ala Gln Lys His Gly Ile Asp Ile
 85 90 95

Ser Ala Leu Gly Gly Phe Thr Ser Ile Ile Phe Glu Asn Phe Asp Leu
 100 105 110

Ala Lys Leu Arg Gln Val Arg Asp Thr Thr Leu Glu Phe Glu Arg Phe
 115 120 125

Thr Thr Gly Asn Thr His Thr Ala Tyr Val Ile Cys Arg Gln Val Glu
 130 135 140

Ala Ala Ala Lys Thr Leu Gly Ile Asp Ile Ala Gln Ala Thr Val Ala
 145 150 155 160

ES 2 600 354 T3

Val Val Gly Ala Thr Gly Asp Ile Gly Ser Ala Val Cys Arg Trp Leu
165 170 175

Asp Leu Lys Leu Gly Val Gly Asp Leu Ile Leu Thr Ala Arg Asn Gln
180 185 190

Glu Arg Leu Asp Asn Leu Gln Ala Glu Leu Gly Arg Gly Lys Ile Leu
195 200 205

Pro Leu Glu Ala Ala Leu Pro Glu Ala Asp Phe Ile Val Trp Val Ala
210 215 220

Ser Met Pro Gln Gly Val Val Ile Asp Pro Ala Thr Leu Lys Gln Pro
225 230 235 240

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

Gln Gly Glu Gly Ile Tyr Val Leu Asn Gly Gly Val Val Glu His Cys
260 265 270

Phe Asp Ile Asp Trp Gln Ile Met Ser Leu Ala Glu Met Ala Arg Pro
275 280 285

Glu Arg Gln Met Phe Ala Cys Phe Ala Glu Ala Met Leu Leu Glu Phe
290 295 300

Glu Gly Trp His Thr Asn Phe Ser Trp Gly Arg Asn Gln Ile Thr Ile
305 310 315 320

Glu Lys Met Glu Ala Ile Gly Glu Ala Ser Val Arg His Gly Phe Gln
325 330 335

Pro Leu Ala Leu Ala Ile
340

<210> 64

<211> 342

<212> PRT

5 <213> *Synechococcus elongatus*

<400> 64

Met Ala Phe Gly Leu Ile Gly His Leu Thr Ser Leu Glu Gln Ala Arg
1 5 10 15

Asp Val Trp Arg Arg Met Gly Tyr Glu Glu Tyr Ala Asp Gln Gly Leu
20 25 30

Glu Phe Trp Ser Ser Ala Pro Pro Gln Ile Val Asp Glu Ile Thr Val
35 40 45

Thr Ser Ala Thr Gly Lys Val Ile His Gly Arg Tyr Ile Glu Ser Gly

325 330 335

Pro Leu Ala Leu Ala Ile
340

<210> 65

<211> 347

<212> PRT

5 <213> *Prochlorococcus marinus*

<400> 65

ES 2 600 354 T3

Met Ala Phe Gly Leu Ile Gly His Ser Thr Ser Phe Glu Asp Ala Lys
1 5 10 15

Arg Lys Ala Ser Leu Leu Gly Phe Asp His Ile Ala Asp Gly Asp Leu
20 25 30

Asp Val Trp Cys Thr Ala Pro Pro Gln Leu Val Glu Asn Val Glu Val
35 40 45

Lys Ser Ala Ile Gly Ile Ser Ile Glu Gly Ser Tyr Ile Glu Ser Cys
50 55 60

Phe Val Pro Glu Met Leu Ser Arg Phe Lys Thr Ala Arg Arg Lys Val
65 70 75 80

Leu Asn Ala Met Glu Leu Ala Gln Lys Lys Gly Ile Asn Ile Thr Ala
85 90 95

Leu Gly Gly Phe Thr Ser Ile Ile Phe Glu Asn Phe Asn Leu Leu Gln
100 105 110

His Lys Gln Ile Arg Asn Thr Ser Leu Glu Trp Glu Arg Phe Thr Thr
115 120 125

Gly Asn Thr His Thr Ala Trp Val Ile Cys Arg Gln Leu Glu Met Asn
130 135 140

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

Gly Ala Thr Gly Asp Ile Gly Ser Ala Val Cys Arg Trp Leu Ile Asn
165 170 175

Lys Thr Gly Ile Gly Glu Leu Leu Leu Val Ala Arg Gln Lys Glu Pro
180 185 190

Leu Asp Ser Leu Gln Lys Glu Leu Asp Gly Gly Thr Ile Lys Asn Leu
195 200 205

Asp Glu Ala Leu Pro Glu Ala Asp Ile Val Val Trp Val Ala Ser Met
210 215 220

ES 2 600 354 T3

```

Pro Lys Thr Met Glu Ile Asp Ala Asn Asn Leu Lys Gln Pro Cys Leu
225      230      235

Met Ile Asp Gly Gly Tyr Pro Lys Asn Leu Asp Glu Lys Phe Gln Gly
      245      250

Asn Asn Ile His Val Val Lys Gly Gly Ile Val Arg Phe Phe Asn Asp
      260      265      270

Ile Gly Trp Asn Met Met Glu Leu Ala Glu Met Gln Asn Pro Gln Arg
      275      280      285

Glu Met Phe Ala Cys Phe Ala Glu Ala Met Ile Leu Glu Phe Glu Lys
      290      295      300

Cys His Thr Asn Phe Ser Trp Gly Arg Asn Asn Ile Ser Leu Glu Lys
305      310      315      320

Met Glu Phe Ile Gly Ala Ala Ser Val Lys His Gly Phe Ser Ala Ile
      325      330      335

Gly Leu Asp Lys His Pro Lys Val Leu Ala Val
      340      345

```

<210> 66

<211> 347

<212> PRT

5 <213> *Prochlorococcus marinus*

<400> 66

ES 2 600 354 T3

Met Ala Phe Gly Leu Ile Gly His Ser Thr Ser Phe Glu Asp Ala Lys
 1 5 10 15

Arg Lys Ala Ser Leu Leu Gly Phe Asp His Ile Ala Asp Gly Asp Leu
 20 25 30

Asp Val Trp Cys Thr Ala Pro Pro Gln Leu Val Glu Asn Val Glu Val
 35 40 45

Lys Ser Ala Ile Gly Ile Ser Ile Glu Gly Ser Tyr Ile Asp Ser Cys
 50 55 60

Phe Val Pro Glu Met Leu Ser Arg Phe Lys Thr Ala Arg Arg Lys Val
 65 70 75 80

Leu Asn Ala Met Glu Leu Ala Gln Lys Lys Gly Ile Asn Ile Thr Ala
 85 90 95

Leu Gly Gly Phe Thr Ser Ile Ile Phe Glu Asn Phe Asn Leu Leu Gln
 100 105 110

His Lys Gln Ile Arg Asn Thr Ser Leu Glu Trp Glu Arg Phe Thr Thr

ES 2 600 354 T3

Met Ala Phe Gly Leu Ile Gly His Ser Thr Ser Phe Glu Asp Ala Lys
 1 5 10 15
 Arg Lys Ala Ser Leu Leu Gly Phe Asp His Ile Ala Asp Gly Asp Leu
 20 25 30
 Asp Val Trp Cys Thr Ala Pro Pro Val Leu Val Glu Asn Val Glu Val
 35 40 45
 Lys Ser Ala Ile Gly Ile Ser Ile Glu Gly Ser Tyr Ile Asp Ser Cys
 50 55 60
 Phe Val Pro Glu Met Leu Ser Arg Phe Lys Thr Ala Arg Arg Lys Val
 65 70 75 80
 Leu Asn Ala Met Glu Leu Ala Gln Lys Lys Gly Ile Asn Ile Thr Ala
 85 90 95
 Leu Gly Gly Phe Thr Ser Ile Ile Phe Glu Asn Phe Asn Leu Leu Gln
 100 105 110
 His Lys Gln Ile Arg Asn Thr Ser Leu Glu Trp Glu Arg Phe Thr Thr
 115 120 125
 Gly Asn Thr His Thr Ala Trp Val Ile Cys Arg Gln Leu Glu Met Asn
 130 135 140
 Ala Pro Lys Ile Gly Ile Asp Leu Lys Ser Ala Thr Val Ala Val Val
 145 150 155 160
 Gly Ala Thr Gly Asp Ile Gly Ser Ala Val Cys Arg Trp Leu Ile Asn
 165 170 175
 Lys Thr Gly Ile Gly Glu Leu Leu Leu Val Ala Arg Gln Lys Glu Pro
 180 185 190
 Leu Asp Ser Leu Gln Lys Glu Leu Asp Gly Gly Thr Ile Lys Asn Leu
 195 200 205
 Asp Glu Ala Leu Pro Glu Ala Asp Ile Val Val Trp Val Ala Ser Met
 210 215 220
 Pro Lys Thr Met Glu Ile Asp Ala Asn Asn Leu Lys Gln Pro Cys Leu
 225 230 235 240
 Met Ile Asp Gly Gly Tyr Pro Lys Asn Leu Asp Glu Lys Phe Gln Gly
 245 250 255
 Asn Asn Ile His Val Val Lys Gly Gly Ile Val Arg Phe Phe Asn Asp
 260 265 270
 Ile Gly Trp Asn Met Met Glu Leu Ala Glu Met Gln Asn Pro Gln Arg
 275 280 285

ES 2 600 354 T3

Glu Met Phe Ala Cys Phe Ala Glu Ala Met Ile Leu Glu Phe Glu Lys
 290 295 300

Cys His Thr Asn Phe Ser Trp Gly Arg Asn Asn Ile Ser Leu Glu Lys
 305 310 315 320

Met Glu Phe Ile Gly Ala Ala Ser Val Lys His Gly Phe Ser Ala Ile
 325 330 335

Gly Leu Asp Lys His Pro Lys Val Leu Ala Val
 340 345

<210> 68

<211> 347

<212> PRT

5 <213> *Prochlorococcus marinus*

<400> 68

Met Ala Phe Gly Leu Ile Gly His Ser Thr Ser Phe Glu Asp Ala Lys
 1 5 10 15

Arg Lys Ala Ser Leu Leu Gly Phe Asp His Ile Ala Asp Gly Asp Leu
 20 25 30

Asp Val Trp Cys Thr Ala Pro Pro Gln Leu Val Glu Asn Val Glu Val
 35 40 45

Lys Ser Ala Ile Gly Ile Ser Ile Glu Gly Ser Tyr Ile Asp Ser Cys
 50 55 60

Phe Val Pro Glu Met Leu Ser Arg Phe Lys Thr Ala Arg Arg Lys Val
 65 70 75 80

Leu Asn Ala Met Glu Leu Ala Gln Lys Lys Gly Ile Asn Ile Thr Ala
 85 90 95

Leu Gly Gly Phe Thr Ser Ile Ile Phe Glu Asn Phe Asn Leu Leu Gln
 100 105 110

His Lys Gln Ile Arg Asn Thr Ser Leu Glu Trp Glu Arg Phe Thr Thr
 115 120 125

Gly Asn Thr His Thr Ala Trp Val Ile Cys Arg Gln Leu Glu Met Asn
 130 135 140

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

Gly Ala Thr Gly Asp Ile Gly Ser Ala Val Cys Arg Trp Leu Ile Asn
 165 170 175

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

ES 2 600 354 T3

Leu Asn Ala Met Glu₈₅ Leu Ala Gln Lys Lys₉₀ Gly Ile Asn Ile Thr Ala
 Leu Gly Gly Phe₁₀₀ Thr Ser Ile Ile Phe₁₀₅ Glu Asn Phe Asn Leu₁₁₀ Leu Gln
 His Lys Gln₁₁₅ Ile Arg Asn Thr Ser₁₂₀ Leu Glu Trp Glu Arg₁₂₅ Phe Thr Thr
 Gly Asn Thr His Thr Ala Trp₁₃₅ Val Ile Cys Arg Gln₁₄₀ Leu Glu Met Asn
 Ala Pro Lys Ile Gly Ile₁₅₀ Asp Leu Lys Ser Ala Thr Val Ala Val Val₁₆₀
 Gly Ala Thr Gly Asp₁₆₅ Ile Gly Ser Ala Val₁₇₀ Cys Arg Trp Leu Ile Asn
 Lys Thr Gly Ile₁₈₀ Gly Glu Leu Leu Leu₁₈₅ Val Ala Arg Gln Lys₁₉₀ Glu Pro
 Leu Asp Ser₁₉₅ Leu Gln Lys Glu Leu₂₀₀ Asp Gly Gly Thr Ile₂₀₅ Lys Asn Leu
 Asp Glu Ala Leu Pro Glu Ala₂₁₅ Asp Ile Val Val Trp₂₂₀ Val Ala Ser Met
 Pro Lys Thr Met Glu Ile₂₃₀ Asp Ala Asn Asn Leu₂₃₅ Lys Gln Pro Cys Leu₂₄₀
 Met Ile Asp Gly Gly₂₄₅ Tyr Pro Lys Asn Leu₂₅₀ Asp Glu Lys Phe Gln Gly
 Asn Asn Ile His₂₆₀ Val Val Lys Gly Gly₂₆₅ Ile Val Arg Phe Phe₂₇₀ Asn Asp
 Ile Gly Trp₂₇₅ Asn Met Met Glu Leu Ala Glu Met Gln Asn₂₈₅ Pro Gln Arg
 Glu Met Phe Ala Cys Phe Ala₂₉₅ Glu Ala Met Ile Leu₃₀₀ Glu Phe Glu Lys
 Cys His Thr Asn Phe Ser₃₁₀ Trp Gly Arg Asn Asn₃₁₅ Ile Ser Leu Glu Lys₃₂₀
 Met Glu Phe Ile Gly₃₂₅ Ala Ala Ser Val Lys₃₃₀ His Gly Phe Ser Ala₃₃₅ Ile
 Gly Leu Asp Lys₃₄₀ His Pro Lys Val Ser Ala Val

<210> 70

<211> 347

ES 2 600 354 T3

<212> PRT

<213> *Prochlorococcus marinus*

<400> 70

Met Ala Phe Gly Leu Ile Gly His Ser Thr Ser Phe Glu Asp Ala Lys
1 5 10 15

Arg Lys Ala Ser Leu Leu Gly Phe Asp His Ile Ala Asp Gly Asp Leu
20 25 30

Asp Val Trp Cys Thr Ala Pro Pro Gln Leu Val Glu Asn Val Glu Val
35 40 45

Lys Ser Ala Ile Gly Ile Ser Ile Glu Gly Ser Tyr Ile Asp Ser Cys
50 55 60

Phe Val Pro Glu Met Leu Ser Arg Phe Lys Thr Ala Arg Arg Lys Val
65 70 75 80

Leu Asn Ala Met Glu Leu Ala Gln Lys Lys Gly Ile Asn Ile Thr Ala
85 90 95

Leu Gly Gly Phe Thr Ser Ile Ile Phe Glu Asn Phe Asn Leu Leu Gln
100 105 110

His Lys Gln Ile Arg Asn Thr Ser Leu Glu Trp Glu Arg Phe Thr Thr
115 120 125

Gly Asn Thr His Thr Ala Trp Val Ile Cys Arg Gln Leu Glu Met Asn
130 135 140

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

Gly Ala Thr Gly Asp Ile Gly Ser Ala Val Cys Arg Trp Leu Ile Asn
165 170 175

Lys Thr Gly Ile Gly Glu Leu Leu Leu Val Ala Arg Gln Lys Glu Pro
180 185 190

Leu Asp Ser Leu Gln Lys Glu Leu Asp Gly Gly Thr Ile Lys Asn Leu
195 200 205

Asp Glu Ala Leu Pro Glu Ala Asp Ile Val Val Trp Val Ala Ser Met
210 215 220

Pro Lys Thr Met Glu Ile Asp Ala Asn Asn Leu Lys Gln Pro Cys Leu
225 230 235 240

Met Ile Asp Gly Gly Tyr Pro Lys Asn Leu Asp Glu Lys Phe Gln Gly

ES 2 600 354 T3

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

Gly Ala Thr Gly Asp Ile Gly Ser Ala Val Cys Arg Trp Leu Ile Asn
 165 170 175

Lys Thr Gly Ile Gly Glu Leu Leu Leu Val Ala Arg Gln Lys Glu Pro
 180 185 190

Leu Asp Ser Leu Gln Lys Glu Leu Asp Gly Gly Thr Ile Lys Asn Leu
 195 200 205

Asp Glu Ala Leu Pro Glu Ala Asp Ile Val Val Trp Val Ala Ser Met
 210 215 220

Pro Lys Thr Met Glu Ile Asp Ala Asn Asn Leu Lys Gln Pro Cys Leu
 225 230 235 240

Met Ile Asp Gly Gly Tyr Pro Lys Asn Leu Asp Glu Lys Phe Gln Gly
 245 250 255

Asn Asn Ile His Val Val Lys Gly Gly Ile Val Arg Phe Phe Asn Asp
 260 265 270

Ile Gly Trp Asn Met Met Glu Leu Ala Glu Met Gln Asn Pro Gln Arg
 275 280 285

Glu Met Phe Ala Cys Phe Ala Glu Ala Met Ile Leu Glu Phe Glu Lys
 290 295 300

Cys His Thr Asn Phe Ser Trp Gly Arg Asn Asn Ile Ser Leu Glu Lys
 305 310 315 320

Met Glu Phe Ile Gly Ala Ala Ser Val Lys His Gly Phe Ser Ala Ile
 325 330 335

Gly Leu Asp Lys His Pro Lys Val Asp Ala Val
 340 345

<210> 72

<211> 347

<212> PRT

5 <213> *Prochlorococcus marinus*

<400> 72

ES 2 600 354 T3

Met Ala Phe Gly Leu Ile Gly His Ser Thr Ser Phe Glu Asp Ala Lys
1 5 10 15

Arg Lys Ala Ser Leu Leu Gly Phe Asp His Ile Ala Asp Gly Asp Leu
20 25 30

Asp Val Trp Cys Thr Ala Pro Pro Gln Leu Val Glu Asn Val Glu Val

ES 2 600 354 T3

Asp Glu Ala Leu Pro Glu Ala Asp Ile Val Val Trp Val Ala Ser Met
210 215 220

Pro Lys Thr Met Glu Ile Asp Ala Asn Asn Leu Lys Gln Pro Cys Leu
225 230 235 240

Met Ile Asp Gly Gly Tyr Pro Lys Asn Leu Asp Glu Lys Phe Gln Gly
245 250 255

Asn Asn Ile His Val Val Lys Gly Gly Ile Val Arg Phe Phe Asn Asp
260 265 270

Ile Gly Trp Asn Met Met Glu Leu Ala Glu Met Gln Asn Pro Gln Arg
275 280 285

Glu Met Phe Ala Cys Phe Ala Glu Ala Met Ile Leu Glu Phe Glu Lys
290 295 300

Cys His Thr Asn Phe Ser Trp Gly Arg Asn Asn Ile Ser Leu Glu Lys
305 310 315 320

Met Glu Phe Ile Gly Ala Ala Ser Val Lys His Gly Phe Ser Ala Ile
325 330 335

Gly Leu Asp Lys His Pro Lys Val Thr Ala Val
340 345

<210> 74

<211> 347

<212> PRT

5 <213> *Prochlorococcus marinus*

<400> 74

ES 2 600 354 T3

Met Ala Phe Gly Leu Ile Gly His Ser Thr Ser Phe Glu Asp Ala Lys
1 5 10 15

Arg Lys Ala Ser Leu Leu Gly Phe Asp His Ile Ala Asp Gly Asp Leu
20 25 30

Asp Val Trp Cys Thr Ala Pro Pro Gln Leu Val Glu Asn Val Glu Val
35 40 45

Lys Ser Ala Ile Gly Ile Ser Ile Glu Gly Ser Tyr Ile Asp Ser Cys
50 55 60

Phe Val Pro Glu Met Leu Ser Arg Phe Lys Thr Ala Arg Arg Lys Val
65 70 75 80

Leu Asn Ala Met Glu Leu Ala Gln Lys Lys Gly Ile Asn Ile Thr Ala
85 90 95

Leu Gly Gly Phe Thr Ser Ile Ile Phe Glu Asn Phe Asn Leu Leu Gln

ES 2 600 354 T3

<400> 75

Met Ala Phe Gly Leu Ile Gly His Ser Thr Ser Phe Glu Asp Ala Lys
 1 5 10 15

Arg Lys Ala Ser Leu Leu Gly Phe Asp His Ile Ala Asp Gly Asp Leu
 20 25 30

Asp Val Trp Cys Thr Ala Pro Pro Gln Leu Val Glu Asn Val Glu Val
 35 40 45

Lys Ser Ala Ile Gly Ile Ser Ile Glu Gly Ser Tyr Ile Asp Ser Cys
 50 55 60

Phe Val Pro Glu Met Leu Ser Arg Phe Lys Thr Ala Arg Arg Lys Val
 65 70 75 80

Leu Asn Ala Met Glu Leu Ala Gln Lys Lys Gly Ile Asn Ile Thr Ala
 85 90 95

Leu Gly Gly Phe Thr Ser Ile Ile Phe Glu Asn Phe Asn Leu Leu Gln
 100 105 110

His Lys Gln Ile Arg Asn Thr Ser Leu Glu Trp Glu Arg Phe Thr Thr
 115 120 125

Gly Asn Thr His Thr Ala Trp Val Ile Cys Arg Gln Leu Glu Met Asn
 130 135 140

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

Gly Ala Thr Gly Asp Ile Gly Ser Ala Val Cys Arg Trp Leu Ile Asn
 165 170 175

Lys Thr Gly Ile Gly Glu Leu Leu Leu Val Ala Arg Gln Lys Glu Pro
 180 185 190

Leu Asp Ser Leu Gln Lys Glu Leu Asp Gly Gly Thr Ile Lys Asn Leu
 195 200 205

Asp Glu Ala Leu Pro Glu Ala Asp Ile Val Val Trp Val Ala Ser Met
 210 215 220

Pro Lys Thr Met Glu Ile Asp Ala Asn Asn Leu Lys Gln Pro Cys Leu
 225 230 235 240

Met Ile Asp Gly Gly Tyr Pro Lys Asn Leu Asp Glu Lys Phe Gln Gly
 245 250 255

Asn Asn Ile His Val Val Lys Gly Gly Ile Val Arg Phe Phe Asn Asp
 260 265 270

ES 2 600 354 T3

Ile Gly Trp Asn Met Met Glu Leu Ala Glu Met Gln Asn Pro Gln Arg
275 280 285

Glu Met Phe Ala Cys Phe Ala Glu Ala Met Ile Leu Glu Phe Glu Lys
290 295 300

Cys His Thr Asn Phe Ser Trp Gly Arg Asn Asn Ile Ser Leu Glu Lys
305 310 315 320

Met Glu Phe Ile Gly Ala Ala Ser Val Lys His Gly Phe Ser Ala Ile
325 330 335

Gly Leu Asp Lys His Pro Lys Val Ala Ala Val
340 345

<210> 76

<211> 347

<212> PRT

5 <213> *Prochlorococcus marinus*

<400> 76

ES 2 600 354 T3

Met Ala Phe Gly Leu Ile Gly His Ser Thr Ser Phe Glu Asp Ala Lys
 1 5 10 15

Arg Lys Ala Ser Leu Leu Gly Phe Asp His Ile Ala Asp Gly Asp Leu
 20 25 30

Asp Val Trp Cys Thr Ala Pro Pro Gln Leu Val Glu Asn Val Glu Val
 35 40 45

Lys Ser Ala Ile Gly Ile Ser Ile Glu Gly Ser Tyr Ile Asp Ser Cys
 50 55 60

Phe Val Pro Glu Met Leu Ser Arg Phe Lys Thr Ala Arg Arg Lys Val
 65 70 75 80

Leu Asn Ala Met Glu Leu Ala Gln Lys Lys Gly Ile Asn Ile Thr Ala
 85 90 95

Leu Gly Gly Phe Thr Ser Ile Ile Phe Glu Asn Phe Asn Leu Leu Gln
 100 105 110

His Lys Gln Ile Arg Asn Thr Ser Leu Glu Trp Glu Arg Phe Thr Thr
 115 120 125

Gly Asn Thr His Thr Ala Trp Val Ile Cys Arg Gln Leu Glu Met Asn
 130 135 140

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

Gly Ala Thr Gly Asp Ile Gly Ser Ala Val Cys Arg Trp Leu Ile Asn

ES 2 600 354 T3

phe Val Pro Glu Met Leu Ser Arg Phe Lys Thr Ala Arg Arg Lys Val
 65 70 75 80
 Leu Asn Ala Met Glu Leu Ala Gln Lys Lys Gly Ile Asn Ile Thr Ala
 85 90
 Leu Gly Gly Phe Thr Ser Ile Ile Phe Glu Asn Phe Asn Leu Leu Gln
 100 105 110
 His Lys Gln Ile Arg Asn Thr Ser Leu Glu Trp Glu Arg Phe Thr Thr
 115 120 125
 Gly Asn Thr His Thr Ala Trp Val Ile Cys Arg Gln Leu Glu Met Asn
 130 135
 Ala Pro Lys Ile Gly Ile Asp Leu Lys Ser Ala Thr Val Ala Val Val
 145 150 155 160
 Gly Ala Thr Gly Asp Ile Gly Ser Ala Val Cys Arg Trp Leu Ile Asn
 165 170
 Lys Thr Gly Ile Gly Glu Leu Leu Leu Val Ala Arg Gln Lys Glu Pro
 180 185 190
 Leu Asp Ser Leu Gln Lys Glu Leu Asp Gly Gly Thr Ile Lys Asn Leu
 195 200 205
 Asp Glu Ala Leu Pro Glu Ala Asp Ile Val Val Trp Val Ala Ser Met
 210 215 220
 Pro Lys Thr Met Glu Ile Asp Ala Asn Asn Leu Lys Gln Pro Cys Leu
 225 230 235 240
 Met Ile Asp Gly Gly Tyr Pro Lys Asn Leu Asp Glu Lys Phe Gln Gly
 245 250 255
 Asn Asn Ile His Val Val Lys Gly Gly Ile Val Arg Phe Phe Asn Asp
 260 265 270
 Ile Glu Trp Asn Met Met Glu Leu Ala Glu Met Gln Asn Pro Gln Arg
 275 280 285
 Glu Met Phe Ala Cys Phe Ala Glu Ala Met Ile Leu Glu Phe Glu Lys
 290 295 300
 Cys His Thr Asn Phe Ser Trp Gly Arg Asn Asn Ile Ser Leu Glu Lys
 305 310 315 320
 Met Glu Phe Ile Gly Ala Ala Ser Val Lys His Gly Phe Ser Ala Ile
 325 330 335
 Gly Leu Asp Lys His Pro Lys Val Leu Ala Val
 340 345

ES 2 600 354 T3

<210> 78

<211> 345

<212> PRT

<213> *Prochlorococcus marinus*

5 <400> 78

Met Ala Phe Gly Leu Ile Gly His Ser Thr Ser Phe Glu Asp Ala Lys
 1 5 10 15

Arg Lys Ala Ser Leu Leu Gly Phe Asp His Ile Ala Asp Gly Asp Leu
 20 25 30

Asp Val Trp Cys Thr Ala Pro Pro Gln Leu Val Glu Asn Val Glu Val
 35 40 45

Lys Ser Ala Ile Gly Ile Ser Ile Glu Gly Ser Tyr Ile Asp Ser Cys
 50 55 60

Phe Val Pro Glu Met Leu Ser Arg Phe Lys Thr Ala Arg Arg Lys Val
 65 70 75 80

Leu Asn Ala Met Glu Leu Ala Gln Lys Lys Gly Ile Asn Ile Thr Ala
 85 90 95

Leu Gly Gly Phe Thr Ser Ile Ile Phe Glu Asn Phe Asn Leu Leu Gln
 100 105 110

His Lys Gln Ile Arg Asn Thr Ser Leu Glu Trp Glu Arg Phe Thr Thr
 115 120 125

Gly Asn Thr His Thr Ala Trp Val Ile Cys Arg Gln Leu Glu Met Asn
 130 135 140

Ala Pro Lys Ile Gly Ile Asp Leu Lys Ser Ala Thr Val Ala Val Val
 145 150 155 160 165

Gly Ala Thr Gly Asp Ile Gly Ser Ala Val Cys Arg Trp Leu Ile Asn
 165 170 175

Lys Thr Gly Ile Gly Glu Leu Leu Leu Val Ala Arg Gln Lys Glu Pro
 180 185 190

Leu Asp Ser Leu Gln Lys Glu Leu Asp Gly Gly Thr Ile Lys Asn Leu
 195 200 205

Asp Glu Ala Leu Pro Glu Ala Asp Ile Val Val Trp Val Ala Ser Met
 210 215 220

Pro Lys Thr Met Glu Ile Asp Ala Asn Asn Leu Lys Gln Pro Cys Leu

REIVINDICACIONES

1. Polipéptido de acil-ACP reductasa (AAR) variante que comprende una identidad de secuencia de al menos el 90% con la secuencia de aminoácidos de SEQ ID NO: 57, en el que dicho polipéptido de AAR variante comprende una mutación en la posición de aminoácido 18, en el que la mutación es S18W, y en el que dicho polipéptido de AAR cataliza la conversión de una acil-ACP en un aldehído graso.
2. Polipéptido de AAR variante según la reivindicación 1, en el que la expresión de dicho polipéptido de AAR variante en una célula huésped recombinante da como resultado un mayor título de una composición de alcohol graso en comparación con el título producido mediante la expresión de un polipéptido de AAR silvestre en una célula huésped silvestre correspondiente.
3. Polipéptido de AAR variante según la reivindicación 2, en el que dicha composición de alcohol graso es una composición de alcohol graso C12 o C14.
4. Polipéptido de AAR variante según cualquiera de las reivindicaciones 1-4, en el que dicho polipéptido de AAR variante comprende además una mutación en una posición de aminoácido seleccionada del grupo que consiste en los aminoácidos 8, 16, 21, 24, 31, 34, 35, 43, 50, 63, 86, 112, 113, 116, 118, 120, 135, 148, 153, 154, 155, 157, 159, 168, 172, 187, 188, 191, 209, 210, 211, 236, 277, 281, 283, 285, 291, 324, 328, 335, 337 y 338.
5. Polipéptido de AAR variante según la reivindicación 4, en el que dicha mutación se selecciona del grupo que consiste en L8A, D16L, M21L, D24E, D24Y, D24V, D24P, L31V, L31M, W34F, W35F, D43E, A50Q, C63A, C63G, C63Y, S86G, A112R, S113K, Q116G, R118Q, T120S, A135S, T148C, T148E, T148V, I153P, T154A, Q155C, Q155L, T157V, A159V, I168V, C172L, T187V, T188H, T188V, Q191A, L209R, E210Y, A211W, T236C, Q277V, A281L, E283G, E283S, A285V, M291V, A324T, A328S, Q335N, L337V y L338W.
6. Polipéptido de AAR variante según la reivindicación 5, en el que dicho polipéptido de AAR variante comprende una mutación M21L, una mutación C63GA, una mutación S113K, una mutación T154A, una mutación A281L o una mutación L8A.
7. Polipéptido de AAR variante según cualquiera de las reivindicaciones 1-3, en el que dicho polipéptido de AAR variante comprende además
 - (a) una mutación M21L, una mutación C63G, una mutación S113K, T154A y una mutación A281L, preferiblemente en el que el polipéptido de AAR variante comprende la secuencia de aminoácidos de SEQ ID NO: 58;
 - (b) una mutación L8A, una mutación M21L, una mutación C63G, una mutación S113K, una mutación T154A y una mutación A281L, preferiblemente en el que el polipéptido de AAR variante comprende la secuencia de aminoácidos de SEQ ID NO: 59;
 - (c) una mutación D16L, una mutación M21L, una mutación C63G, una mutación S113K, T154A y una mutación A281L, preferiblemente en el que el polipéptido de AAR variante comprende la secuencia de aminoácidos de SEQ ID NO: 60;
 - (d) una mutación L8A, una mutación D24V, una mutación C63G, una mutación S113K, una mutación Q155L y una mutación A281L, preferiblemente en el que el polipéptido de AAR variante comprende la secuencia de aminoácidos de SEQ ID NO: 61;
 - (e) una mutación D24P, una mutación L31M, una mutación C63G, una mutación S113K, una mutación T154A y una mutación A281L, preferiblemente en el que el polipéptido de AAR variante comprende la secuencia de aminoácidos de SEQ ID NO: 62;
 - (f) una mutación L8A, una mutación D16L, una mutación D24V, una mutación C63G, una mutación S113K, una mutación T154A y una mutación A281L, preferiblemente en el que el polipéptido de AAR variante comprende la secuencia de aminoácidos de SEQ ID NO: 63; o
 - (g) una mutación D24E, una mutación C63G, una mutación S113K, una mutación T154A y una mutación A281L, preferiblemente en el que el polipéptido de AAR variante comprende la secuencia de aminoácidos de SEQ ID NO: 64.
8. Célula huésped recombinante que expresa el polipéptido de AAR variante según cualquiera de las reivindicaciones 1 a 7.
9. Célula huésped recombinante según la reivindicación 8, en la que la célula huésped recombinante produce una composición de alcohol graso con un título que es al menos el 10% mayor, al menos el 15% mayor, al menos el 20% mayor, al menos el 25% mayor, o al menos el 30% mayor que el título de una composición de alcohol graso producida por una célula huésped que expresa un polipéptido de AAR silvestre

correspondiente, cuando se cultiva en medio que contiene una fuente de carbono en condiciones eficaces para expresar el polipéptido de AAR variante, preferiblemente en la que la composición de alcohol graso se produce de manera extracelular.

- 5 10. Célula huésped recombinante según la reivindicación 9, en la que la composición de alcohol graso se produce a un título de aproximadamente 30 g/l a aproximadamente 250 g/l.
11. Cultivo celular que comprende la célula huésped recombinante según cualquiera de las reivindicaciones 8-10.
12. Cultivo celular según la reivindicación 11, en el que la composición de alcohol graso comprende uno o más de un alcohol graso C6, C8, C10, C12, C13, C14, C15, C16, C17, y C18.
- 10 13. Cultivo celular según la reivindicación 12, en el que la composición de alcohol graso comprende
- (a) uno o más de un alcohol graso insaturado C10:1, C12:1, C14:1, C16:1, y C18:1;
 - (b) un alcohol graso insaturado;
 - (c) un alcohol graso que tiene un doble enlace en la posición 7 en la cadena de carbonos entre C₇ y C₈ desde el extremo reducido del alcohol graso; o
 - 15 (d) un alcohol graso saturado.
14. Método de producción de una composición de alcohol graso que tiene un aumento del título, que comprende:
- i. cultivar la célula huésped según la reivindicación 8 con una fuente de carbono; y
 - ii. recoger una composición de alcohol graso.
- 20 15. Método según la reivindicación 14, en el que el título del alcohol graso es de al menos del 20% al 30% mayor que el título de una composición de alcohol graso producida por una célula huésped que expresa AAR silvestre.

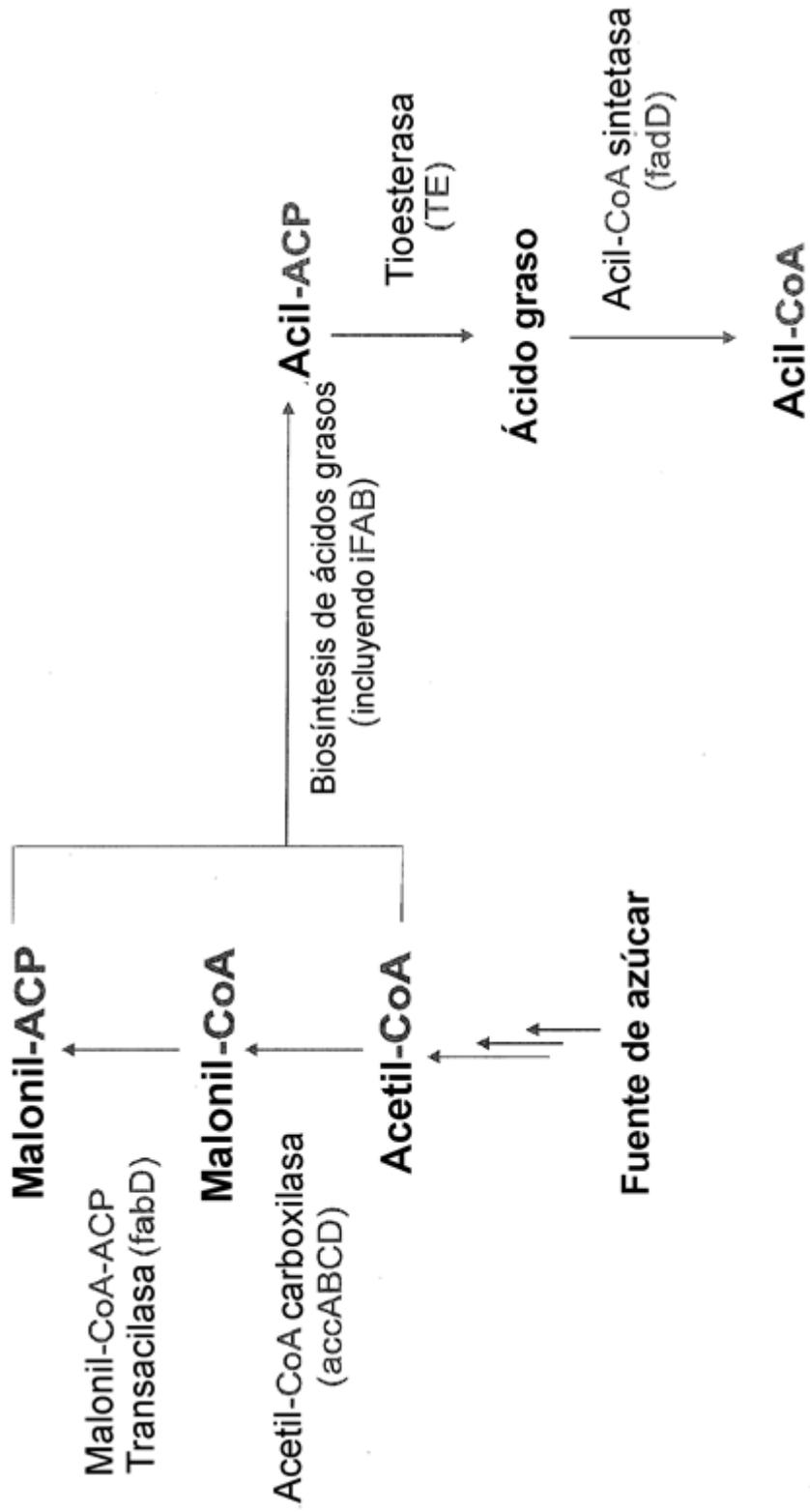


Figura 1

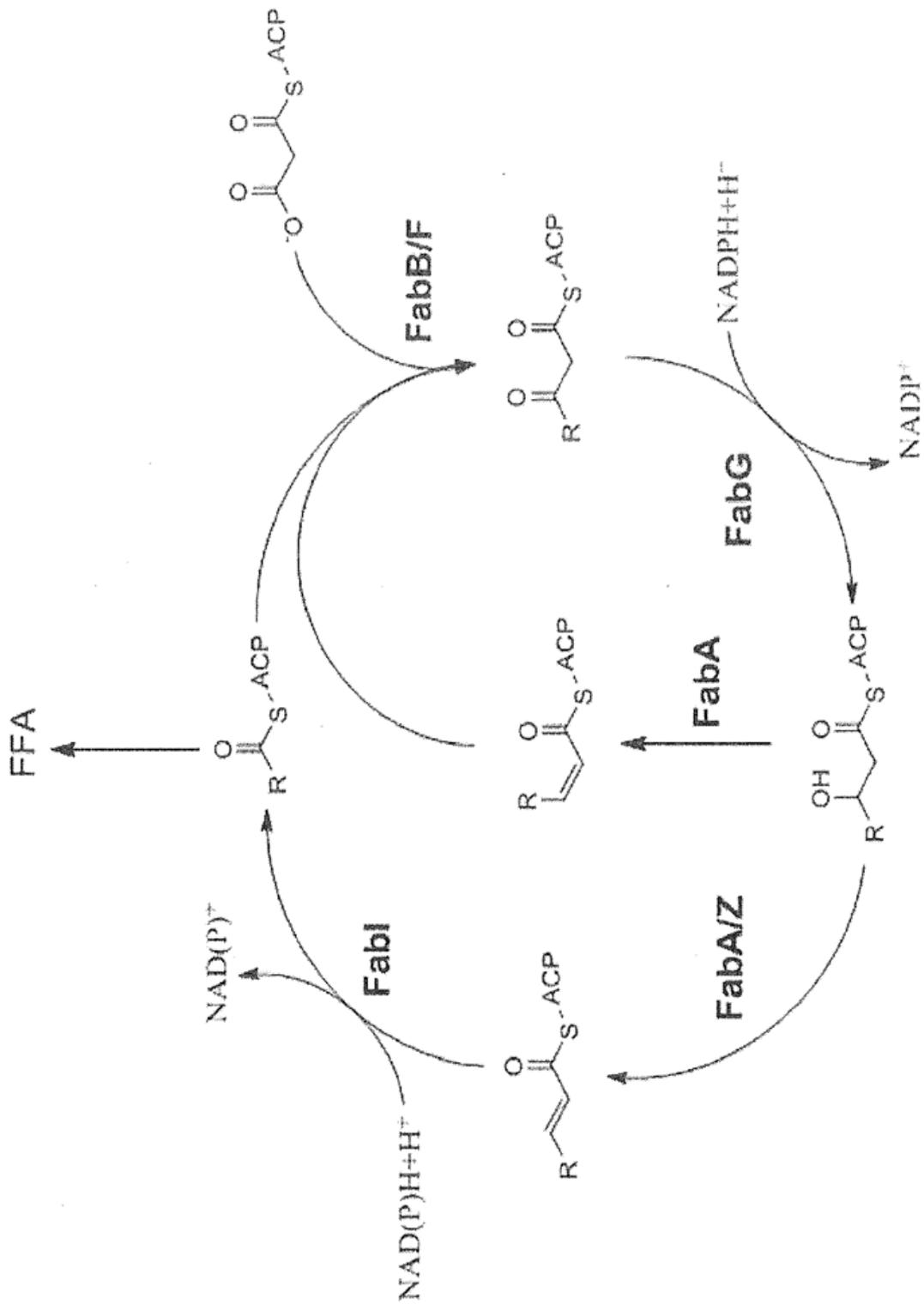


Figura 2

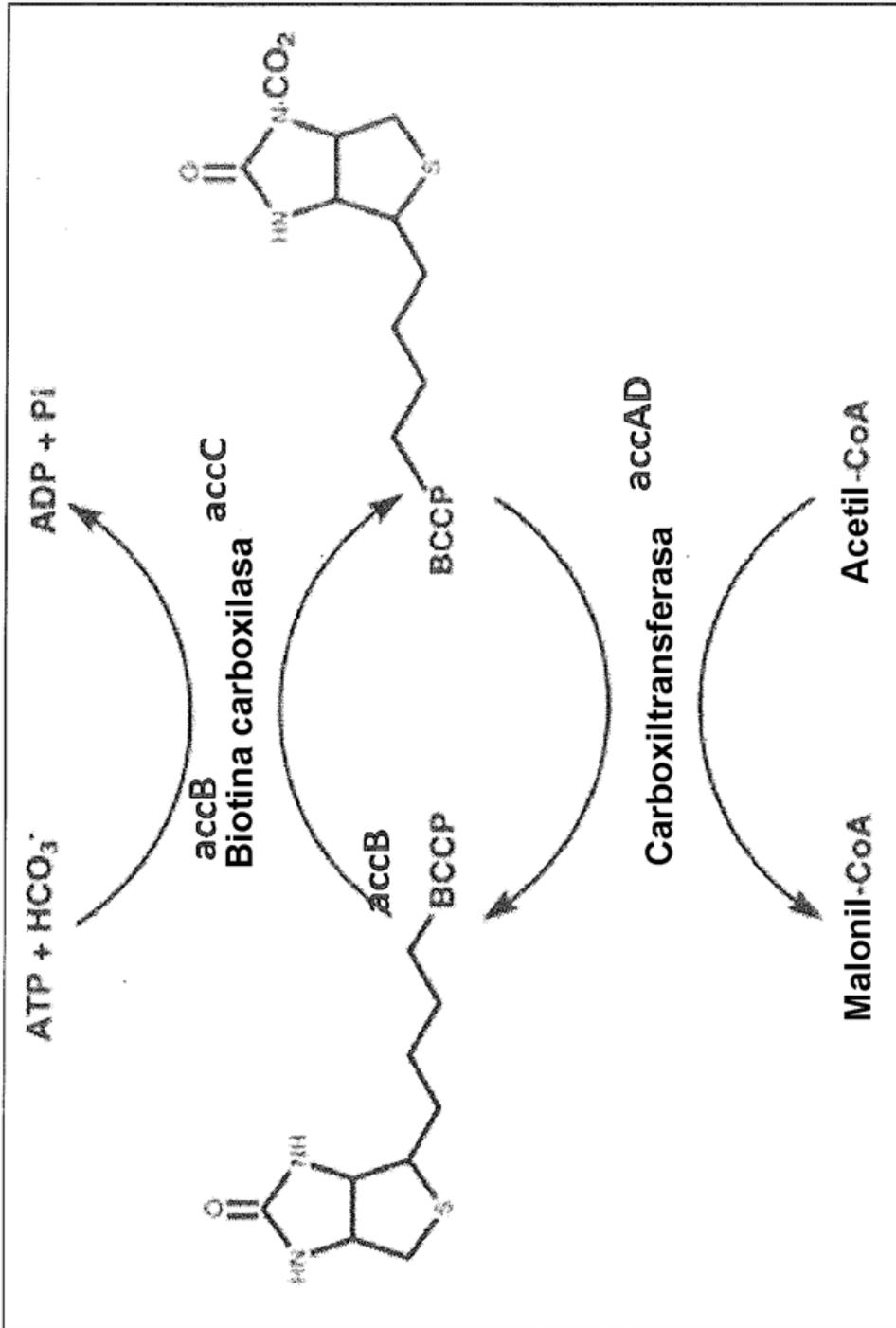


Figura 3

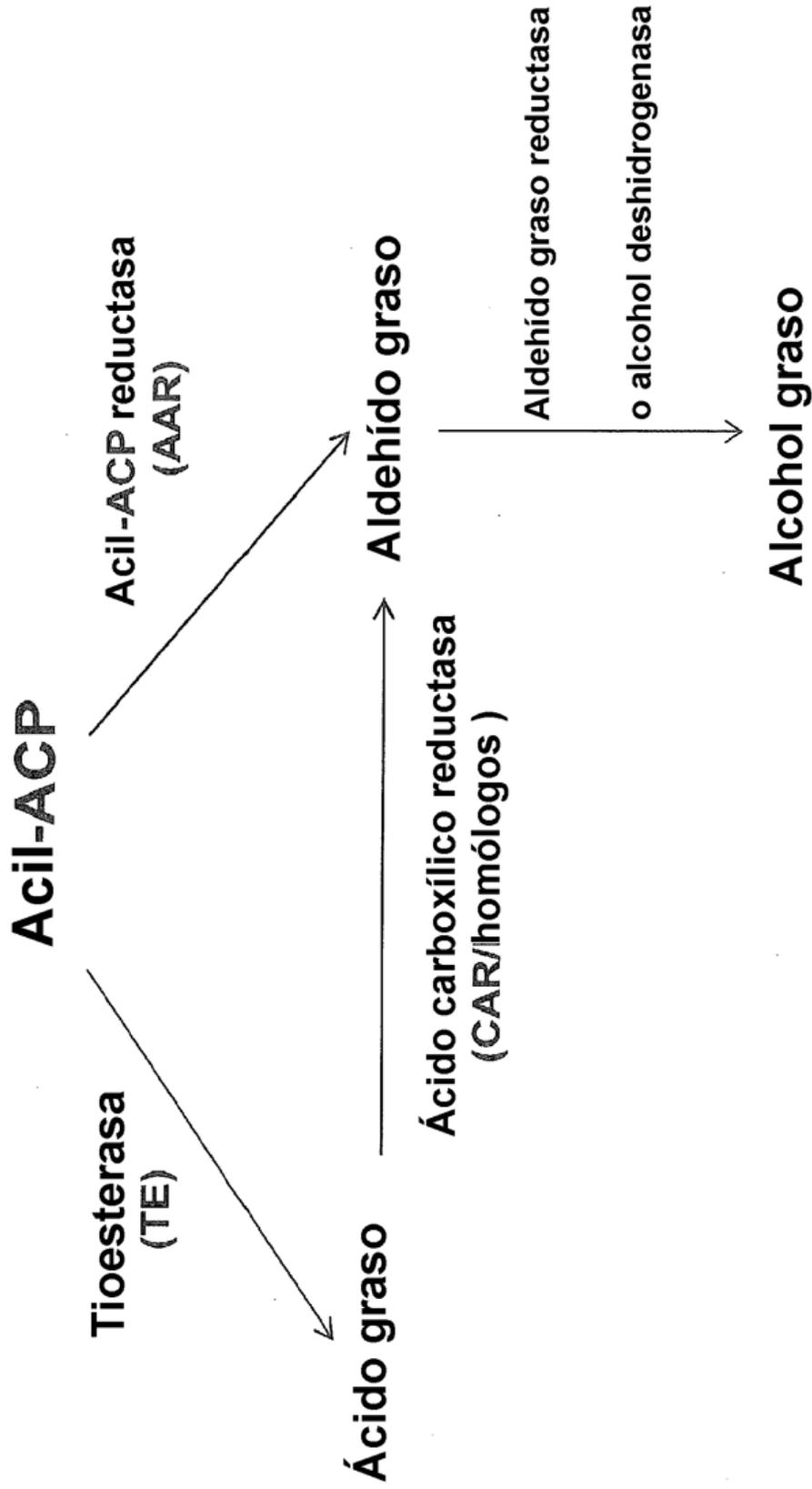


Figura 4

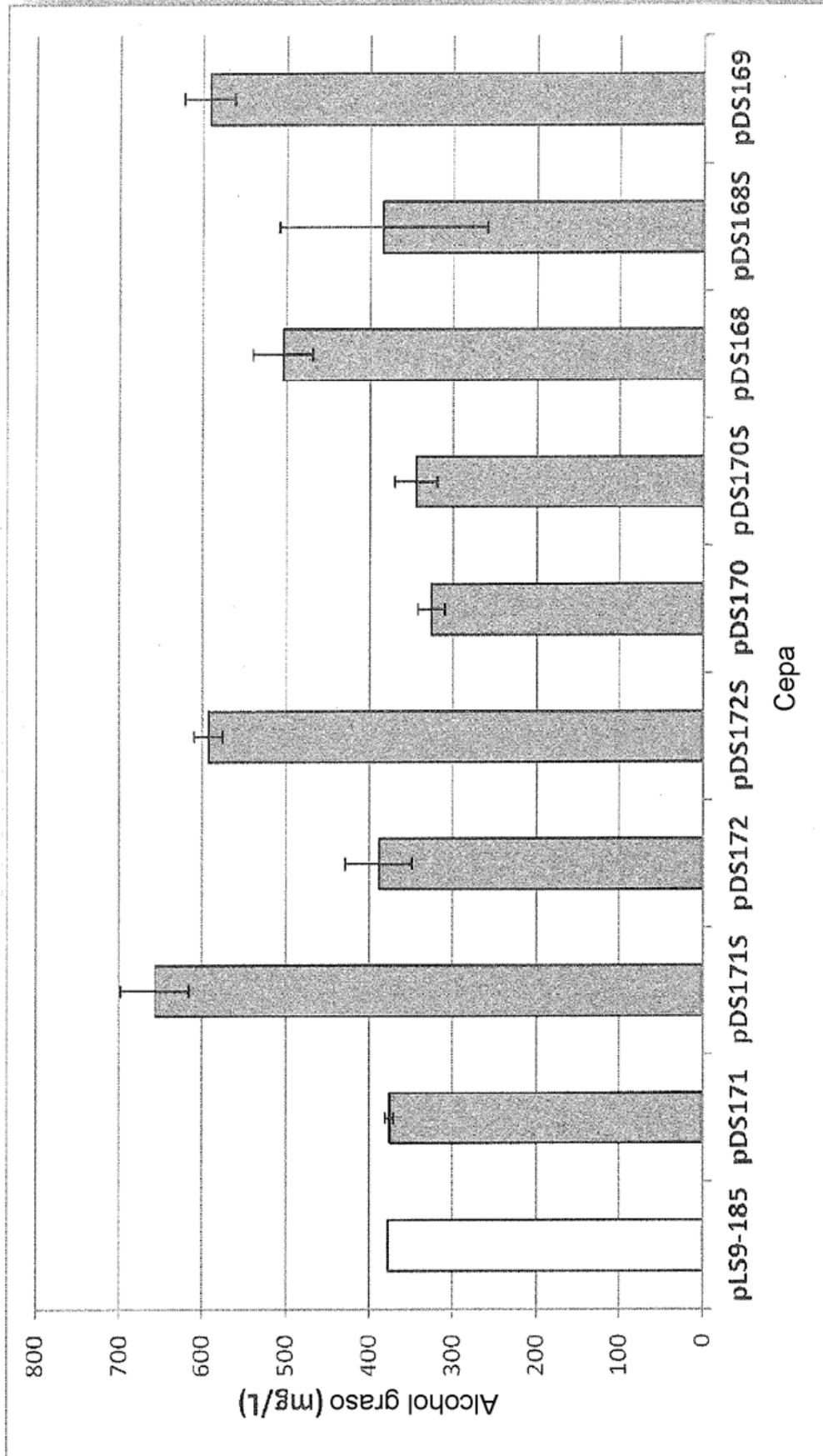


Figura 5

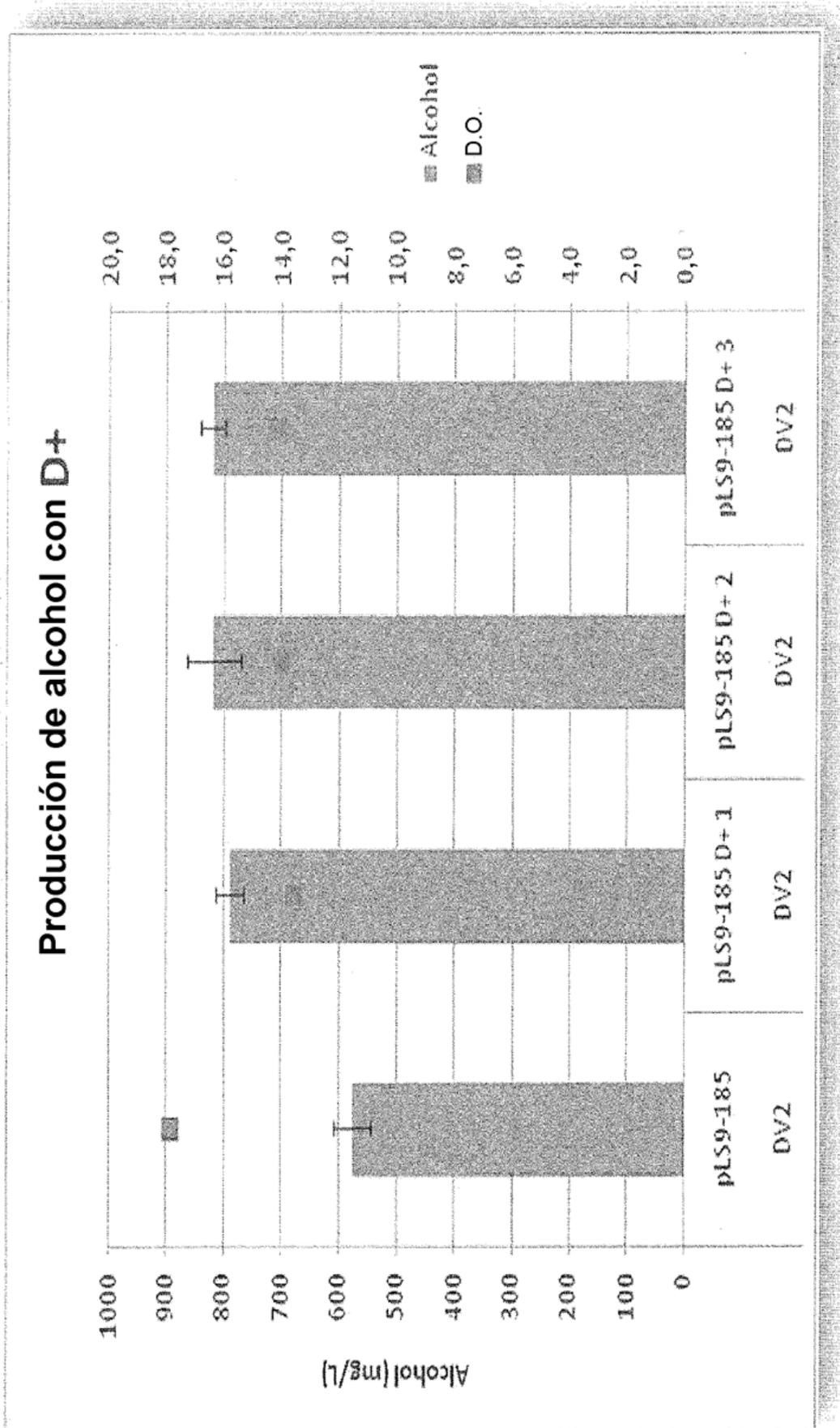


Figura 6

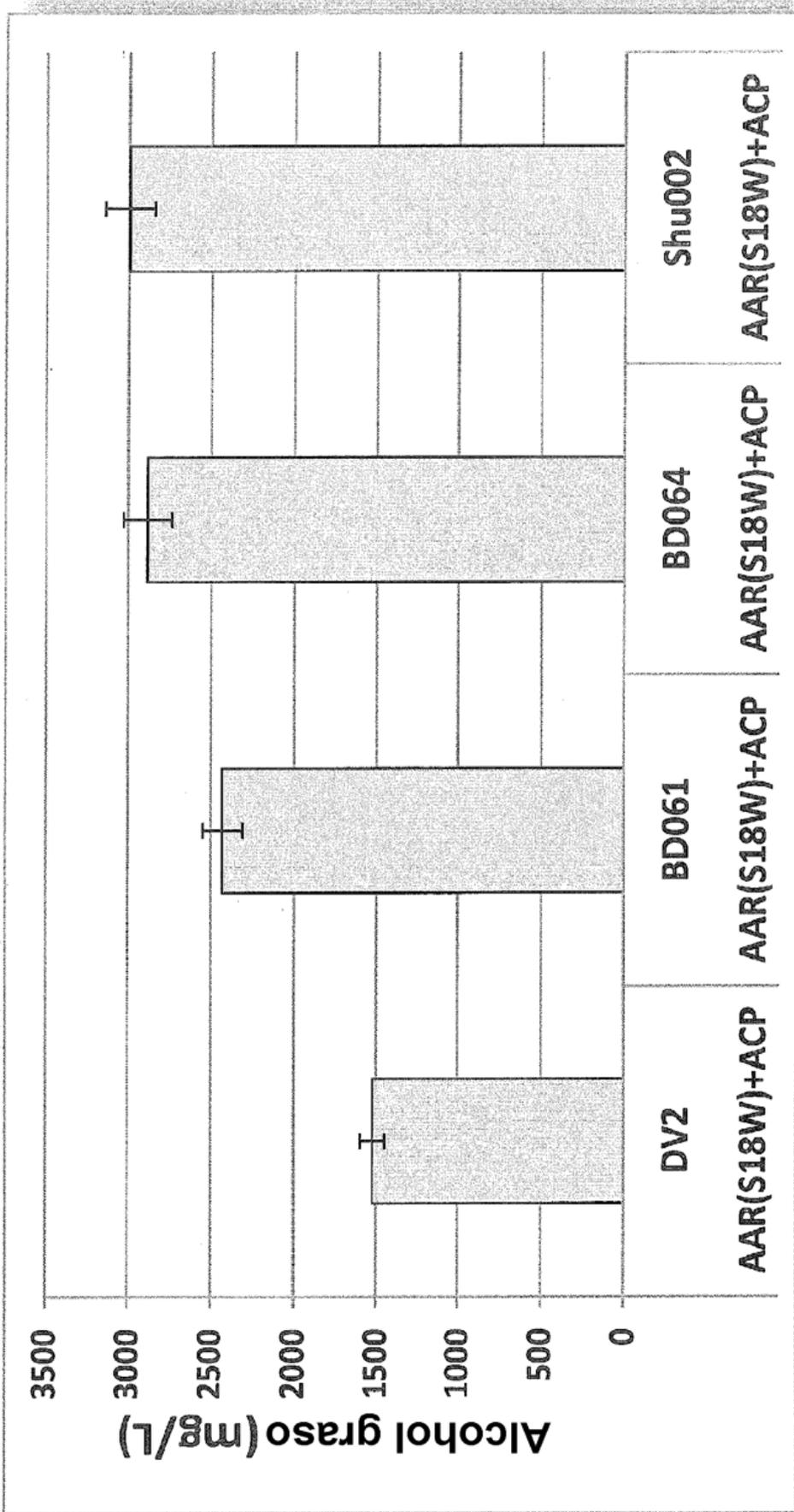


Figura 7

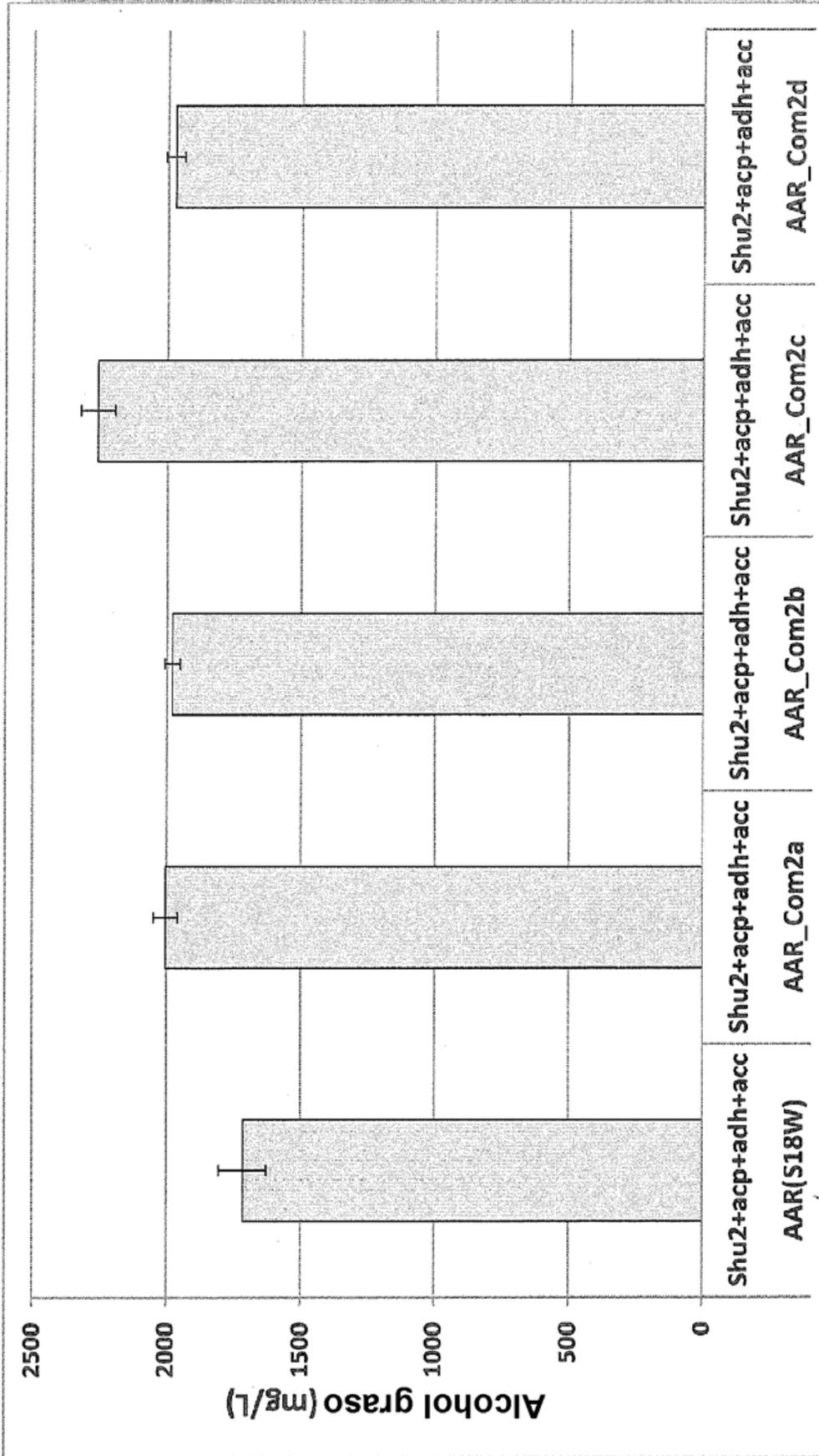


Figura 8

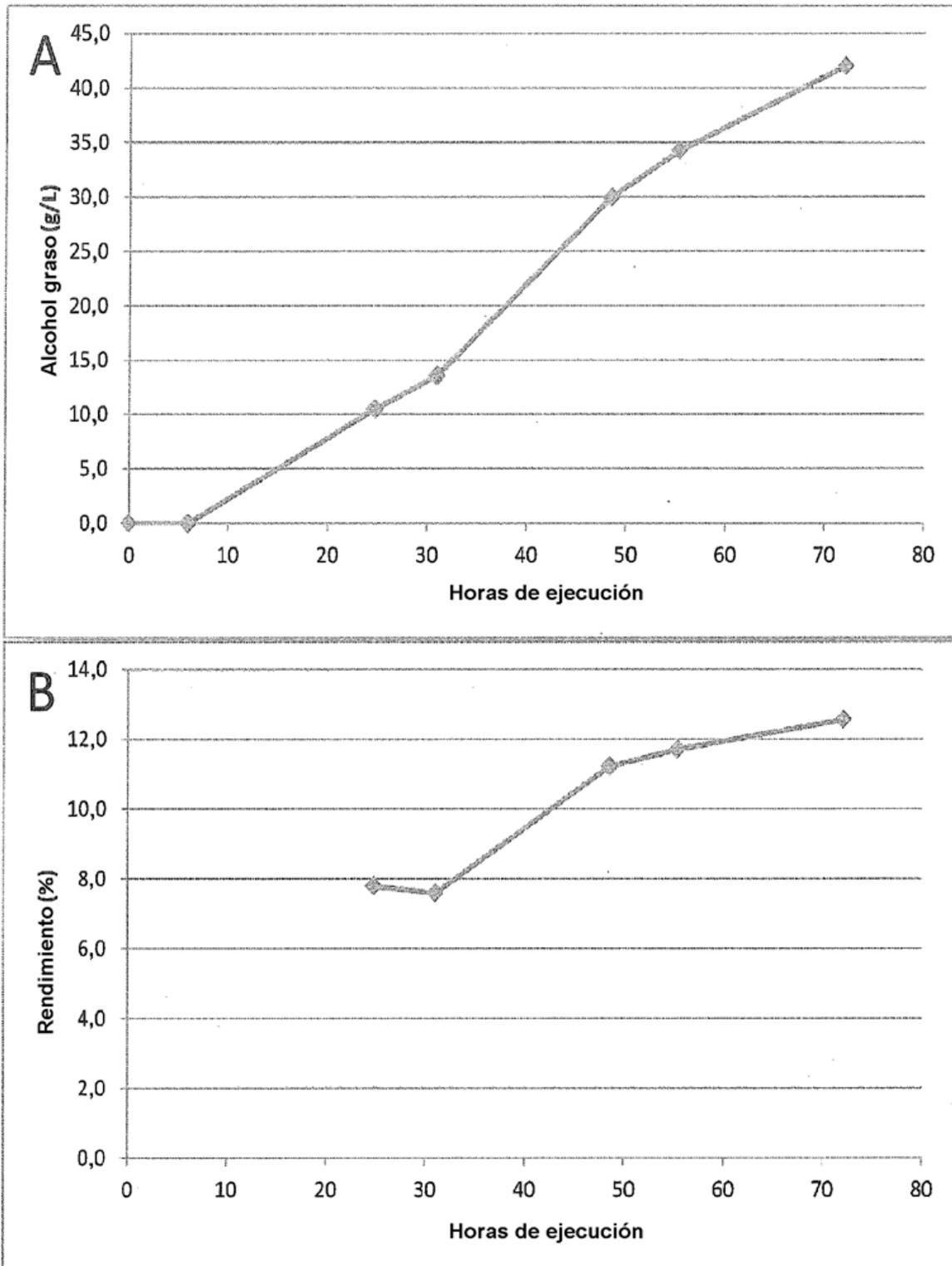


Figura 9

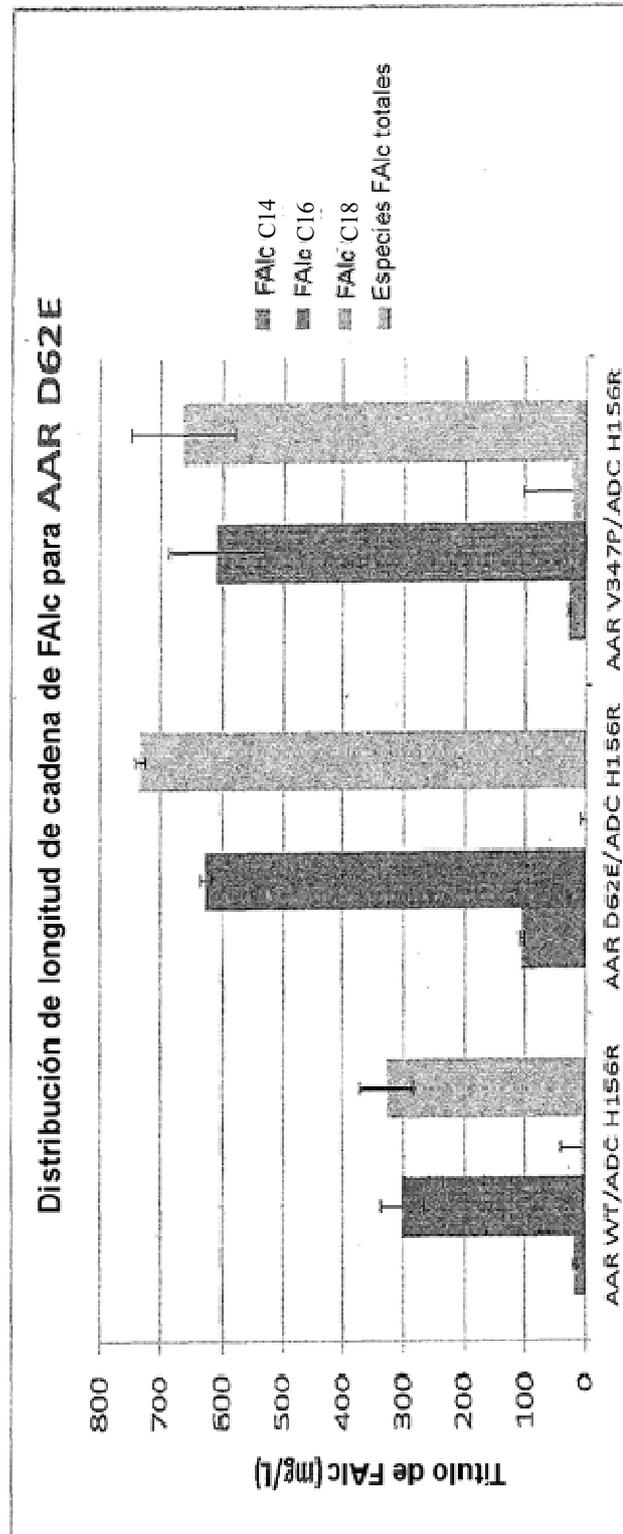


Figura 10