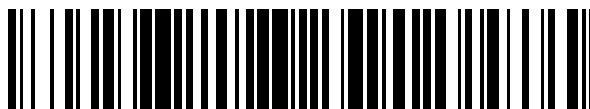


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 660 301**

51 Int. Cl.:

C12N 9/10 (2006.01)

C12P 7/42 (2006.01)

C12P 7/66 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **18.12.2014** **PCT/EP2014/078540**

87 Fecha y número de publicación internacional: **25.06.2015** **WO15091843**

96 Fecha de presentación y número de la solicitud europea: **18.12.2014** **E 14828027 (4)**

97 Fecha y número de publicación de la concesión europea: **29.11.2017** **EP 3083950**

54 Título: **Glicosiltransferasa que glicosila ácido flavokermésico y/o ácido kermésico**

30 Prioridad:

18.12.2013 EP 13198110

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

21.03.2018

73 Titular/es:

KØBENHAVNS UNIVERSITET (50.0%)
Nørregade 10

1017 København K, DK y
DANMARKS TEKNISKE UNIVERSITET (50.0%)

72 Inventor/es:

KANNANGARA, RUBINI MAYA;
BENNEDSEN, MADS;
MADSEN, BJØRN;
BINDERUP, KIM;
THRANE, ULF;
FRANDSEN, RASMUS JOHN NORMAND;
MORTENSEN, UFFE HASBRO;
MØLLER, BIRGER LINDBERG y
OKKELS, FINN THYGE

74 Agente/Representante:

CARPINTERO LÓPEZ, Mario

ES 2 660 301 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

Glicosiltransferasa que glicosila ácido flavokermésico y/o ácido kermésico

CAMPO DE LA INVENCION

- 5 La presente invención se refiere a un polipéptido de tipo glicosiltransferasa (GT) aislado capaz de: (I): conjugar glucosa con ácido flavokermésico (FK); y/o (II): conjugar glucosa con ácido kermésico (AK) y el uso de esta GT para preparar, p. ej., ácido carmínico.

ANTECEDENTES DE LA INVENCION

El pigmento natural ácido carmínico es uno de los colorantes utilizados más frecuentemente en alimentos, medicinas, productos cosméticos y textiles.

- 10 El ácido carmínico es un colorante, el cual se puede extraer de los cuerpos de insectos hembra de *Dactylopius coccus costa* (nombre alternativo *Coccus cacti* L.). Los insectos viven en *Nopalea cochinellifera*, *Opuntia fidus indica* y otras plantas de la familia *Cactaceae* que se cultivan, por ejemplo, en las áreas desérticas de México, América Central y del Sur y las Islas Canarias. Dependiendo del pH, el colorante puede ser un color comprendido en un espectro que varía desde naranja pasando por rojo hasta lila y se conoce generalmente como cochinilla o color de cochinilla. El colorante
15 de carmín se utiliza extensamente en alimentos y bebidas.

Según existe constancia en la técnica, *Porphyrophora polonica* también produce ácido carmínico y se ha cultivado para la producción de ácido carmínico, p. ej., en Polonia.

Respecto a la producción industrial relevante actual, el ácido carmínico se recolecta mediante la extracción de los cuerpos secos del insecto con agua o alcohol.

- 20 Los insectos (*Dactylopius coccus*) se cultivan en cactus y, por consiguiente, el suministro puede ser relativamente caro y está sometido a variaciones no deseables y fluctuaciones del precio.

Con el fin de intentar resolver el problema de las variaciones no deseables y las fluctuaciones del precio, el documento US5424421 (European Colour, publicado en 1995) describe la síntesis química del ácido carmínico mediante una ruta de síntesis que implica diferentes intermedios.

- 25 Según se discute, p. ej., en el documento WO2006/056585A1 (Chr. Hansen A/S), durante la extracción acuosa del ácido carmínico a partir del insecto, también se libera cierta cantidad de proteína del insecto a partir del insecto y esta quedará contenida en el extracto de colorante y se ha publicado que las proteínas de insectos de tipo cochinilla podrían crear ciertos problemas relacionados con alergias. En el documento WO2006/056585A1 se describe un proceso especial para reducir la cantidad de proteína del insecto a partir de la solución de extracto del insecto; sin embargo, el
30 producto/composición de color producido final del documento WO2006/056585A1 seguirá comprendiendo ciertas cantidades de proteínas del insecto *Dactylopius coccus costa*.

La estructura del ácido carmínico se muestra en la Figura 1; según se puede observar, es lo que se denomina un C-glucósido (es decir, donde la glucosa está unida/conjugada con el aglucón mediante un enlace carbono-carbono).

- 35 De acuerdo con la técnica, el término «aglicón» denota la parte que no es un carbohidrato de la forma glicosilada correspondiente del aglicón. Cuando el azúcar es glucosa, el aglicón se puede denominar aglucón.

De acuerdo con la técnica, el término «glicósido» denota un compuesto que por hidrólisis da como resultado un azúcar y un residuo que no es un azúcar (aglicón), p. ej., los glucósidos pueden dar glucosa, los galactósidos pueden dar galactosa. Según se muestra en la Figura 1, la hidrólisis del C-glucósido ácido carmínico da como resultado glucosa y el aglucón ácido kermésico (AK).

- 40 En la actualidad, no se ha descrito detalladamente la ruta biosintética del insecto *in vivo* (*Dactylopius coccus*) implicada en la producción de carmín; por consiguiente, basándose en la técnica anterior, el experto no sabe qué compuesto es el aglucón durante la producción biosintética de ácido carmínico por parte de *Dactylopius coccus in vivo*.

El análisis de *D. coccus* ha mostrado que en el extracto de *D. coccus* están presentes una amplia gama de compuestos relacionados con el ácido carmínico y en principio numerosos de estos compuestos se podrían glucosilar durante la
45 producción biosintética de ácido carmínico por parte de *Dactylopius coccus in vivo*.

Por ejemplo, el artículo de Stathopoulou *et al.* (*Analytica Chimica Acta* 804 (2013) 264-272) describe seis antraquinonas nuevas que están presentes en el extracto de *D. coccus* y cualquiera de estas seis antraquinonas nuevas (remítase, p. ej., a la Figura 1 del artículo) podría ser, en principio, la molécula que se glucosila durante la producción biosintética de ácido carmínico por parte de *Dactylopius coccus in vivo*.

- 50 Además, y según existe constancia en la técnica, el compuesto glucosilado principal formado durante la producción biosintética *in vivo* del producto final de tipo glucósido puede ser un compuesto intermedio inestable que no se

identificará en un extracto aislado de *D. coccus* como, p. ej., se analiza en el artículo discutido anteriormente de Stathopoulou *et al.*

Basándose en la técnica anterior, se podría especular que un compuesto glucosilado principal relevante durante la producción biosintética de ácido carmínico por parte de *Dactylopius coccus in vivo* podría ser, p. ej., un compuesto de tipo policétido intermedio inestable con aproximadamente el mismo número de átomos de carbono que, p. ej., el ácido flavokermésico.

De acuerdo con la técnica, el término «glicosiltransferasa» (GT) denota una enzima glicosiltransferasa capaz de transferir un azúcar de un azúcar de nucleótido activado a un aglicón para formar un glicósido.

En la técnica anterior no se ha descrito explícitamente una secuencia de aminoácidos o ADN relevante para la presente de una glicosiltransferasa implicada en la ruta biosintética del ácido carmínico por parte del insecto (*Dactylopius coccus in vivo*).

Según existe constancia en la técnica, para insectos que acumulan productos químicos de bajo peso molecular, en ocasiones los genes de las rutas biosintéticas relevantes no están presentes en el genoma del insecto. Por ejemplo, algunos insectos captan los glicósidos de las plantas de las que se alimentan; remítase, p. ej., al artículo de Zagrobelny *et al.* (*Cyanogenic glucosides and plant-insect interactions; Phytochemistry*. Feb de 2004;65(3):293-306) o el artículo de Geuder *et al.* (*Journal of Chemical Ecology*, Vol. 23, N.º 5, 1997). Además, los genes de las rutas biosintéticas relevantes se encuentran en ocasiones en los microorganismos que viven en los insectos, remítase, p. ej., al artículo de Genta *et al.*, (*Potential role for gut microbiota in cell wall digestion and glucoside detoxification in Tenebrio molitor larvae*), *Journal of Insect Physiology* 52 (2006) 593-601.

Los insectos de *Dactylopius coccus* se alimentan de plantas de cactus y podría darse el caso de que los insectos de *D. coccus* (al igual que otros insectos) captaran los glicósidos relevantes de los cactus de los que se alimentan.

Por consiguiente, basándose en la técnica anterior, el experto no podría saber si el genoma de *Dactylopius coccus* comprendería de hecho un gen que codificase una glicosiltransferasa implicada en la ruta biosintética *in vivo* que conduce al ácido carmínico.

El documento WO2004/111254A1 (Poalis A/S) describe la producción *in vivo* de una forma glucosilada de la vanilina en, p. ej., células eucariotas, células de levadura y/o células de *E. coli* procariotas utilizando una glicosiltransferasa para conjugar glucosa al aglicón de la vanilina *in vivo* dentro de una célula del microorganismo. La vanilina natural se obtiene de la semilla de la planta de vainilla. Por consiguiente, en la técnica anterior se ha descrito la producción *in vivo* con éxito en células de microorganismos de compuestos que son glicósidos de plantas (tales como, p. ej., un glicósido de la vanilina).

COMPENDIO DE LA INVENCIÓN

El problema que ha de resolver la presente invención se refiere a proporcionar una glicosiltransferasa (GT) implicada en una ruta biosintética que pueda conducir al ácido carmínico y al uso de esta glicosiltransferasa para preparar, p. ej., ácido carmínico.

Según se discute en los ejemplos prácticos de la presente, los inventores de esta han secuenciado todo el genoma y transcriptoma (es decir, el conjunto de moléculas de ARN, incluido el ARNm) de *Dactylopius coccus* y organismos simbióticos microbianos.

Las secuencias de oligonucleótidos identificadas obtenidas a partir del genoma y transcriptoma se analizaron para determinar su similitud respecto a secuencias de C-glicosiltransferasas de dominio público y el resultado fue negativo. Ninguna de las secuencias génicas identificadas del genoma/transcriptoma mostró una similitud significativa respecto a las secuencias de C-glicosiltransferasas de dominio público.

Según se ha discutido anteriormente, basándose en la técnica anterior, el experto no podría saber si el genoma de *Dactylopius coccus* comprendería de hecho un gen que codificase una glicosiltransferasa implicada en la ruta biosintética *in vivo* que conduce al ácido carmínico. Sin embargo, los inventores de la presente siguieron investigando la materia en cuestión.

Según se discute en los ejemplos prácticos de la presente, los inventores de esta identificaron un extracto de *Dactylopius coccus* (incluidos extractos de los organismos endosimbióticos presentes en *D. coccus*) con una actividad de GT relevante y, mediante una combinación de pasos de evaluación y purificación relevantes, los inventores fueron finalmente capaces de obtener una composición/fracción relativamente pura a partir de la cual fue posible obtener varias secuencias de aminoácidos parciales de posibles candidatos de enzimas GT.

Las secuencias de aminoácidos parciales de estos candidatos de enzimas se compararon con las secuencias génicas identificadas del transcriptoma y, después de un trabajo detallado adicional, se identificó una secuencia que codificaba una secuencia de enzima de tipo glicosiltransferasa; la secuencia del polinucleótido que codifica esta

glicosiltransferasa novedosa clonada/aislada se muestra en la SEQ ID NO: 1 y la secuencia de aminoácidos del polipéptido se muestra en la SEQ ID NO: 2.

La enzima de tipo glicosiltransferasa de la SEQ ID NO: 2 se puede denominar «DcUGT2».

5 Se cree que la glicosiltransferasa clonada/aislada descrita es la primera glicosiltransferasa derivada de un insecto descrita. Según se ha descrito, las secuencias génicas identificadas del transcriptoma de *Dactylopius coccus* se analizaron para determinar su similitud respecto a secuencias de glicosiltransferasas de dominio público relevantes y el resultado fue negativo.

10 Los inventores de la presente han descubierto que las secuencias de glicosiltransferasas de la técnica anterior de dominio público presentan una identidad inferior a un 45% respecto a la secuencia del polipéptido de glicosiltransferasa novedosa que se muestra en la SEQ ID NO: 2.

Según se discute en los ejemplos prácticos de la presente, los inventores de esta evaluaron la actividad de la glicosiltransferasa novedosa clonada/aislada y descubrieron que era capaz de conjugar glucosa con los aglicones ácido flavokermésico (FK) y ácido kermésico (AK) - remítase a la Figura 1.

15 El análisis del producto glucosilado mostró que, entre los muchos productos de tipo O- y C-glucósido posibles que se podrían formar potencialmente, se formó únicamente un único producto de tipo glucósido con cada uno de los dos sustratos. El análisis mostró que cada uno de los productos fue C-glucosilado en la posición 7 de la estructura de antraquinona, más exactamente el ácido 7- α -D-glucopiranosil-9,10-dihidro-3,6,8-trihidroxi-1-metil-9,10-dioxoantracenocarboxílico (DcII - remítase a la Figura 1) y el ácido 7- α -D-glucopiranosil-9,10-dihidro-3,5,6,8-tetrahidroxi-1-metil-9,10-dioxoantracenocarboxílico (ácido carmínico - remítase a la Figura 1), por parte de la GT
20 cuando se utilizó UDP-glucosa como el sustrato dador de azúcar.

El artículo de Gutmann *et al.* (*Pure Appl. Chem*, 2013-07-09) describe que, a pesar de que se han aislado una serie de C-glicósidos a partir de fuentes naturales, las enzimas responsables de su biosíntesis son solo conocidas en muy pocos casos y las estrategias biocatalíticas para la producción de C-glicósidos todavía no se han establecido.

25 El artículo de Baig *et al.* (*Angew Chem Int Ed Engl.* 27 de noviembre de 2006;45(46):7842-6) describe la glicosiltransferasa (GT) denominada UrdGT2 y explica que es capaz de conjugar un azúcar con una serie de aglicones que se pueden considerar en la presente relativamente similares al ácido flavokermésico(FK) y el ácido kermésico (AK).

30 Por consiguiente, se puede decir que en principio UrdGT2 sería una conjetura cualificada para una GT que podría ser capaz de conjugar un azúcar con el ácido flavokermésico (FK) y/o el ácido kermésico (AK). Según se discute en un Ejemplo práctico de la presente, los inventores de esta clonaron UrdGT2 y la evaluaron para determinar su actividad de GT para el ácido flavokermésico (FK) y/o el ácido kermésico (AK) y se observó que UrdGT2 era capaz de utilizar UDP-glucosa como dador de azúcar, pero UrdGT2 no glucosiló ninguno de los posibles aglicones evaluados, es decir, no se identificó ninguna actividad de GT en relación con estos aglicones.

35 La UrdGT2 tiene aproximadamente una identidad de los aminoácidos de un 15-20% con la SEQ ID NO:2 descrita en la presente.

Basándose tanto en secuencias de GT de dominio público como en secuencias de GT que no son de dominio público, los inventores de la presente realizaron diferentes investigaciones de alineaciones de secuencias detalladas.

Basándose en estas investigaciones de alineaciones de secuencias, se cree que los aminoácidos del 20 a aproximadamente el 468 de la SEQ ID NO:2 comprenden las partes esenciales del dominio catalítico.

40 Basándose en estas investigaciones de alineaciones de secuencias, se cree que los aminoácidos de aproximadamente el 291 a aproximadamente el 383 de la SEQ ID NO:2 comprenden el denominado sitio de unión al azúcar del nucleótido activado.

45 Basándose en estas investigaciones de alineaciones de secuencias, se cree que los aminoácidos de aproximadamente el 1 a aproximadamente el 20 de la SEQ ID NO:2 comprenden el denominado péptido señal y se cree que este péptido señal se puede eliminar sin que la actividad de GT de la enzima relevante de la presente se vea afectada de forma significativa.

50 Además, se cree que el sitio de unión al azúcar del nucleótido activado se puede sustituir por secuencias de sitio de unión al azúcar del nucleótido activado de GT similares (p. ej., conocidos en la técnica anterior), tales como, p. ej., el sitio de unión al azúcar del nucleótido activado que se describe en Radominska-Pandya A, Bratton SM, Redinbo MR, Miley MJ. *Drug Metab Rev.* Febrero de 2010;42(1): 133-44) y *Plant Physiology*, Noviembre de 2008, Vol. 148, págs. 1295-1308.

Por consiguiente, un primer aspecto de la presente invención se refiere a un polipéptido de tipo glicosiltransferasa aislado capaz de:

(I): conjugar glucosa con ácido flavokermésico (FK); y/o

(II): conjugar glucosa con ácido kermésico (AK);

y donde el polipéptido de tipo glicosiltransferasa es al menos un polipéptido seleccionado del grupo constituido por:

5 (a) un polipéptido que comprende una secuencia de aminoácidos que presenta una identidad de al menos un 70% con los aminoácidos 1-515 de la SEQ ID NO:2;

(b) un polipéptido que comprende una secuencia de aminoácidos que presenta una identidad de al menos un 70% con los aminoácidos 20-468 de la SEQ ID NO:2;

(c) un polipéptido que está codificado por un polinucleótido que se hibrida en condiciones de rigurosidad al menos media con (i) los nucleótidos 1-1548 de la SEQ ID NO:1 o (ii) una hebra complementaria de (i); y

10 (d) un fragmento de los aminoácidos 1-515 de la SEQ ID NO:2, que presenta la actividad de glicosiltransferasa que se ha especificado en (I) y/o (II).

Según sobreentiende el experto en el presente contexto, la expresión «un polipéptido de tipo glicosiltransferasa capaz de» del primer aspecto se refiere a que la glicosiltransferasa será capaz de llevar a cabo la actividad de glicosiltransferasa (I) y/o (II), pero también puede ser capaz o no de llevar a cabo otras actividades de glicosiltransferasa.

Según sobreentiende el experto en el presente contexto, la divulgación de la secuencia de glicosiltransferasa novedosa descrita en la presente es una herramienta importante para identificar glicosiltransferasas similares en, p. ej., otros insectos que no sean *Dactylopius coccus* y, sin pretender vincularse a ninguna teoría, se cree que una secuencia con una identidad de al menos un 70% con la SEQ ID NO:2 sería un buen candidato razonable para otra glicosiltransferasa relevante de la presente.

Un segundo aspecto de la presente invención se refiere a un polinucleótido aislado que comprende una secuencia de nucleótidos que codifica el polipéptido del primer aspecto y/o realizaciones relevantes en la presente de este.

Un tercer aspecto de la presente invención se refiere a un constructo de ácido nucleico que comprende el polinucleótido aislado del segundo aspecto y/o realizaciones relevantes en la presente de este, unido operablemente a una o más secuencias de control que dirigen la producción del polipéptido en un hospedador de expresión.

Un cuarto aspecto de la presente invención se refiere a un vector de expresión recombinante que comprende el constructo de ácido nucleico del tercer aspecto y/o realizaciones relevantes en la presente de este.

Un quinto aspecto de la presente invención se refiere a una célula hospedadora recombinante que comprende el constructo de ácido nucleico del tercer aspecto y/o realizaciones relevantes en la presente de esta.

30 Según se ha discutido anteriormente, basándose en la técnica anterior, el experto no sabe qué compuesto es el compuesto glicosilado principal durante la producción biosintética de ácido carmínico *in vivo* en *Dactylopius coccus*.

Se ha demostrado que *D. coccus* contiene una GT capaz de C-glicosilar el ácido flavokermésico (FK) y/o el ácido kermésico (AK).

35 Es obvio que este conocimiento importante es suficiente para producir, p. ej., ácido carmínico sin la necesidad de realizar una extracción a partir de los insectos y, de este modo, poder preparar un producto/composición colorante de ácido carmínico esencialmente exento de, p. ej., proteínas del insecto *Dactylopius coccus* costa no deseadas.

Aunque el experto no sabía qué compuesto se glicosilaba durante la producción biosintética de ácido carmínico por parte de *Dactylopius coccus in vivo*, de hecho el experto sí que sabía si existía realmente en la naturaleza una glicosiltransferasa capaz de C-glicosilar el aglicón ácido flavokermésico y/o el aglicón ácido kermésico.

40 Se cree que la glicosiltransferasa novedosa descrita en la presente representa la primera glicosiltransferasa aislada capaz de glicosilar el aglicón ácido flavokermésico y/o ácido kermésico.

Por consiguiente, basándose en la divulgación técnica de la presente, se cree que el experto sería capaz de identificar otras glicosiltransferasas adecuadas capaces de glicosilar el ácido flavokermésico (FK) y/o ácido kermésico (AK).

45 El experto apreciaría que una manera de intentar identificar si un organismo/planta comprendería una glicosiltransferasa relevante sería poner en contacto aglicones relevantes (es decir, FK y/o AK) con el organismo/planta (*in vivo* y/o *in vitro*) y a continuación cuantificar si el organismo/planta produce glicósidos de FK y/o AK relevantes.

Según se sobreentiende en la presente, si el organismo/planta produce glicósidos de FK y/o AK relevantes, entonces el organismo/planta comprenderá una glicosiltransferasa relevante, es decir, una glicosiltransferasa que es capaz de glicosilar el ácido flavokermésico con el fin de producir un glicósido del ácido flavokermésico; y/o capaz de glicosilar el ácido kermésico con el fin de producir un glicósido del ácido kermésico.

Según se discute más adelante, basándose en la estrategia anterior, los inventores de la presente han descubierto que se pueden identificar glicosiltransferasas relevantes en plantas de *Aloe*, plantas de *Haworthia*, Sorgo o plantas de arroz.

- 5 Por consiguiente, un sexto aspecto de la presente invención se refiere a un método para producir un glicósido del ácido flavokermésico (FK) y/o un glicósido del ácido kermésico (AK), donde el método comprende los siguientes pasos:

(A): poner en contacto *in vivo* o *in vitro* en una célula hospedadora recombinante que comprende un gen de glicosiltransferasa que codifica una glicosiltransferasa:

(a1): ácido flavokermésico (FK) con una glicosiltransferasa capaz de glicosilar el ácido flavokermésico en condiciones adecuadas, donde se produce el glicósido del ácido flavokermésico; y/o

- 10 (a2): ácido kermésico (AK) con una glicosiltransferasa capaz de glicosilar el ácido kermésico en condiciones adecuadas, donde se produce el glicósido del ácido kermésico.

La expresión «célula hospedadora recombinante» se debe interpretar en la presente de acuerdo con la técnica. Según existe constancia en la técnica, las moléculas de polinucleótidos recombinantes (p. ej., ADN) son moléculas de polinucleótidos (p. ej., ADN) formadas mediante métodos de laboratorio de recombinación genética (tales como clonación molecular) para agrupar material genético de múltiples fuentes, con lo que se crean secuencias que no se encontrarían de otro modo en organismos biológicos. Según sobreentiende el experto, una célula hospedadora recombinante comprende moléculas de polinucleótidos recombinantes (p. ej., ADN) y, por consiguiente, no se sobreentenderá que una célula hospedadora recombinante cubra una célula de origen natural en sí tal como, p. ej., una célula de *Dactylopius coccus* de origen natural.

- 20 Dicho de otro modo y según sobreentiende el experto, por ejemplo, una célula de *Dactylopius coccus* de origen natural en sí no contiene ningún gen de glicosiltransferasa recombinante que codifique una glicosiltransferasa.

Se puede preferir que la célula hospedadora recombinante del paso (A) sea una célula hospedadora recombinante que comprenda un gen de glicosiltransferasa recombinante que codifique una glicosiltransferasa.

- 25 Según se discute en la presente, en los Ejemplos prácticos se puso en contacto *in vitro* ácido flavokermésico (FK) y/o ácido kermésico (AK) con la glicosiltransferasa de la SEQ ID NO:2. Para el experto, se puede ver como un trabajo rutinario el hecho de llevar a cabo un paso de puesta en contacto *in vitro* de este tipo.

La glicosiltransferasa de la SEQ ID NO:2 se expresó recombinantemente en una célula de levadura (remítase al Ejemplo práctico), por consiguiente, se preparó una célula hospedadora de levadura recombinante que comprendía un gen de glicosiltransferasa recombinante que codificaba una glicosiltransferasa de SEQ ID NO:2.

- 30 Se cree que si se añadiera ácido flavokermésico (FK) y/o ácido kermésico (AK) en condiciones adecuadas a un medio de fermentación, el o los compuestos FK y/o AK entrarían en, p. ej., células de levadura fermentadas en el medio; por consiguiente, si, p. ej., las células de levadura son células hospedadoras de levadura recombinantes que comprenden un gen de glicosiltransferasa recombinante que codifica una glicosiltransferasa, entonces se produciría una puesta en contacto *in vivo* en una célula hospedadora recombinante de FK y/o AK con una glicosiltransferasa.

- 35 En, p. ej., los documentos WO2004/111254A1 (Poalis A/S) discutidos anteriormente, se llevaron a cabo tales puestas en contacto *in vivo* de diferentes compuestos aglicones en diferentes células hospedadoras recombinantes y el experto sabría cómo realizar tales puestas en contacto *in vivo* en una célula hospedadora recombinante de un aglicón relevante (en la presente ácido flavokermésico (FK) y/o ácido kermésico (AK)) y una glicosiltransferasa expresada recombinantemente.

- 40 Según se ha discutido anteriormente, se cree que la glicosiltransferasa novedosa descrita en la presente representa la primera vez que se ha descrito una glicosiltransferasa aislada capaz de glicosilar el aglicón ácido flavokermésico y/o el aglicón ácido kermésico.

Se cree que secuencias parciales relevantes de la glicosiltransferasa novedosa descrita en la presente de SEQ ID NO:2 se pueden introducir recombinantemente en otra secuencia de glicosiltransferasa con el fin de construir una nueva secuencia de glicosiltransferasa híbrida capaz de glicosilar ácido flavokermésico y/o ácido kermésico. Tales GT con una reducción de K_M o un incremento de V_{max} podrían resultar importantes a la hora de garantizar una glucosilación rápida de los sustratos que podrían presentar efectos tóxicos que inhibirían el crecimiento de levaduras si se acumulan en niveles elevados (Esben Halkjaer Hansen *et al. Phytochemistry* 70(4): 473-482). Del mismo modo, si así se desea, se prevé posible modificar la especificidad por el sustrato hacia la glucosilación de intermedios previos de la ruta.

- 50 Por consiguiente, un aspecto adicional de la presente invención se refiere a un método para construir un polipéptido de tipo glicosiltransferasa híbrido aislado novedoso capaz de:

(I): conjugar glucosa con ácido flavokermésico (FK); y/o

(II): conjugar glucosa con ácido kermésico (AK),

donde el método comprende los siguientes pasos:

(i): insertar una secuencia de polinucleótido que codifica un fragmento de una secuencia de aminoácidos que presenta una identidad de al menos un 70% con los aminoácidos 1-515 de la SEQ ID NO:2 (preferentemente un fragmento de una secuencia de aminoácidos que presenta una identidad de al menos un 90% con los aminoácidos 1-515 de la SEQ ID NO:2, más preferentemente un fragmento de una secuencia de aminoácidos que presenta una identidad de al menos un 99% con los aminoácidos 1-515 de la SEQ ID NO:2) donde el fragmento comprende al menos 75 aminoácidos (preferentemente al menos 100 aminoácidos, más preferentemente al menos 150 aminoácidos y aún más preferentemente al menos 468 aminoácidos), en otra secuencia de polinucleótido derivada de una glicosiltransferasa con el fin de construir de este modo una secuencia de polinucleótido híbrido recombinante novedoso;

(ii): expresar el polipéptido híbrido novedoso que está codificado por la secuencia de polinucleótido híbrido recombinante novedoso del paso (i);

(iii): aislar el polipéptido híbrido novedoso expresado del paso (ii);

(iv): evaluar si el polipéptido híbrido novedoso aislado del paso (iii) es capaz de:

(I): conjugar glucosa con ácido flavokermésico (FK); y/o

(II): conjugar glucosa con ácido kermésico (AK); y

(v) si la prueba del paso (iv) da positivo, entonces se ha construido el polipéptido de tipo glicosiltransferasa híbrido aislado novedoso capaz de:

(I): conjugar glucosa con ácido flavokermésico (FK); y/o

(II): conjugar glucosa con ácido kermésico (AK).

DEFINICIONES

Todas las definiciones de términos relevantes están de acuerdo con lo que sobreentendería el experto respecto al contexto técnico relevante.

El término «aglicón» denota la parte que no es un carbohidrato de la forma glicosilada correspondiente del aglicón. Cuando el azúcar es glucosa, el aglicón se puede denominar aglucón. Además, cuando el azúcar es glucosa, se puede utilizar el término glucosilado en lugar de glicosilado. Cuando el aglicón está glicosilado en un grupo hidroxilo, generalmente se crea lo que se denomina un enlace O-glicosídico, es decir, lo que se denomina un O-glicósido (u O-glucósido si el azúcar es glucosa). Cuando el aglicón está glicosilado mediante un enlace carbono-carbono, se puede denominar un enlace C-glicosídico, es decir, lo que se denomina un C-glicósido (o C-glucósido si el azúcar es glucosa).

El término «glicósido» denota un compuesto que cuando se somete a hidrólisis puede dar un azúcar y un residuo que no es un azúcar (aglicón), p. ej., los glucósidos pueden dar glucosa y los galactósidos pueden dar galactosa.

El término «glicosiltransferasa» denota una enzima capaz de conjugar un azúcar activado con un nucleótido a un compuesto (p. ej., un compuesto de tipo aglicón). El azúcar puede ser, p. ej., los isómeros D y L de la galactosa, glucosamina, *N*-acetilglucosamina, xilosa, ácido glucurónico, ramnosa, arabinosa, manosa o glucosa. Como alternativa, el azúcar puede ser un derivado de tipo carbohidrato tal como, p. ej., inositol, olivosa, rodinosa, etc. disponible como difosfatos de nucleótidos. Además, el azúcar puede ser, p. ej., un monosacárido, un disacárido o un trisacárido. En el caso de oligo- y polisacáridos, los azúcares se enlazan uno a uno, p. ej., implicando el uso de una o varias glicosiltransferasas. Si el azúcar es glucosa, la glicosiltransferasa se puede denominar glucosiltransferasa.

Cuando la glicosiltransferasa conjuga un azúcar activado con un nucleótido a un compuesto mediante un enlace C-glicosídico, se puede denominar una C-glicosiltransferasa.

Cuando la glicosiltransferasa conjuga un azúcar a un aglicón mediante un enlace O-glicosídico, se puede denominar una O-glicosiltransferasa.

La expresión «se hibrida», en relación con un polinucleótido que se hibrida en condiciones de rigurosidad al menos media con (i) los nucleótidos 1-1548 de la SEQ ID NO:1 o (ii) una hebra complementaria de (i), se refiere a la secuencia de nucleótidos que se hibrida con una sonda de ácido nucleico marcada correspondiente a la secuencia de nucleótidos mostrada en la SEQ ID NO:1 o su hebra complementaria en condiciones de rigurosidad de media a muy alta. Las moléculas con las cuales se hibrida la sonda de ácido nucleico en estas condiciones se pueden detectar utilizando, p. ej., una película de rayos X.

En la presente, las condiciones de rigurosidad relevantes para la hibridación son las condiciones de rigurosidad que sobreentendería normalmente el experto que son relevantes, es decir, para las condiciones de rigurosidad media, las condiciones que el experto sobreentendería que son condiciones de rigurosidad media. El experto estará familiarizado

con las condiciones de rigurosidad relevantes para la hibridación; remítase, p. ej., a (J. Sambrook, E.F. Fritsch y T. Maniatus, 1989, *Molecular Cloning, A Laboratory Manual*, 2.^a edición, Cold Spring Harbor, Nueva York).

De acuerdo con la técnica, para sondas largas de al menos 100 nucleótidos de longitud, unas condiciones de rigurosidad de muy baja a muy alta se definen como una prehibridación e hibridación a 42 °C en 5X SSPE, un 0.3% de SDS, 200 µg/ml de ADN de esperma de salmón desnaturalizado y troceado, y o bien un 25% de formamida para rigurosidades muy bajas y bajas, un 35% de formamida para rigurosidades medias y medias-altas, o un 50% de formamida para rigurosidades altas y muy altas, y a continuación procedimientos de inmunotransferencia de Southern estándar durante 12-24 horas de forma óptima.

Para sondas largas de al menos 100 nucleótidos de longitud, el material portador se lava finalmente tres veces, cada una de ellas durante 15 minutos utilizando 2X SSC y un 0.2% de SDS, preferentemente al menos a 45 °C (rigurosidad muy baja), más preferentemente al menos a 50 °C (rigurosidad baja), más preferentemente al menos a 55 °C (rigurosidad media), más preferentemente al menos a 60 °C (rigurosidad media-alta), aún más preferentemente al menos a 65°C (rigurosidad alta) y de la forma más preferida al menos a 70 °C (rigurosidad muy alta).

La expresión «*in vitro*» (latín: en vidrio) se refiere a estudios que se llevan a cabo utilizando componentes de un organismo que han sido aislados de su entorno biológico habitual con el fin de permitir un análisis más detallado o más adecuado del que se puede realizar con organismos enteros. Estos experimentos se denominan habitualmente «experimentos de tubo de ensayo». Por el contrario, los estudios *in vivo* son aquellos que se llevan a cabo con organismos vivos en su estado intacto normal.

La expresión «*in vivo*» (latín para «dentro del que vive») se refiere a experimentación que utiliza un organismo vivo entero en lugar de un organismo muerto o parcial, o un entorno controlado *in vitro* («dentro del vidrio», p. ej., en un tubo de ensayo o una cápsula de Petri).

La expresión «polinucleótido aislado» se refiere esencialmente en la presente a que el polinucleótido está aislado de su entorno natural; dicho de otro modo, que el preparado del polinucleótido está esencialmente exento de otro material polinucleotídico con el que se asocia de forma nativa. La secuencia de polinucleótido que codifica la glicosiltransferasa novedosa clonada/aislada descrita en la presente se muestra en la SEQ ID NO: 1 y se aisló del insecto *Dactylopius coccus*. Por consiguiente, según sobrentiende el experto, la expresión polinucleótido aislado no cubre el polinucleótido de la SEQ ID NO: 1 cuando se encuentra presente de forma natural en el genoma de *Dactylopius coccus*. La expresión «polinucleótido aislado» se refiere esencialmente a que el polinucleótido aislado se encuentra en una forma adecuada para su uso en sistemas de producción de proteínas modificadas genéticamente. De este modo, un polinucleótido aislado contiene como máximo un 10%, preferentemente como máximo un 8%, más preferentemente como máximo un 6%, más preferentemente como máximo un 5%, más preferentemente como máximo un 4%, más preferentemente como máximo un 3%, aún más preferentemente como máximo un 2%, de la forma más preferida como máximo un 1% y de la forma aún más preferida como máximo un 0.5% en peso de otro material polinucleotídico con el que se asocia de forma nativa. La expresión «polinucleótido aislado» se puede denominar de forma alternativa en la presente «polinucleótido clonado».

La expresión «polipéptido aislado» se refiere esencialmente en la presente a que el polipéptido está aislado de su entorno natural. El polipéptido de tipo glicosiltransferasa novedoso que se muestra en la SEQ ID NO: 2 se aisló del insecto *Dactylopius coccus*. Por consiguiente, según sobrentiende el experto, la expresión «polipéptido aislado» no cubre el polipéptido de tipo glicosiltransferasa de la SEQ ID NO: 2 cuando se encuentra presente de forma natural en el genoma de *Dactylopius coccus*. La expresión «polipéptido aislado» denota un preparado del polipéptido que contiene como máximo un 10%, preferentemente como máximo un 8%, más preferentemente como máximo un 6%, más preferentemente como máximo un 5%, más preferentemente como máximo un 4%, como máximo un 3%, aún más preferentemente como máximo un 2%, de la forma más preferida como máximo un 1% y de la forma aún más preferida como máximo un 0.5% en peso de otro material polipeptídico con el que se asocia de forma nativa. La expresión «otro material polipeptídico con el que se asocia de forma nativa», en relación con el polipéptido de tipo glicosiltransferasa novedoso que se muestra en la SEQ ID NO: 2, se puede interpretar como que se relaciona con otro material polipeptídico con el que se asocia de forma nativa en *Dactylopius coccus*. En algún caso, se puede preferir que el «polipéptido aislado» se refiera a un polipéptido con una pureza de al menos un 20%, preferentemente una pureza de al menos un 40%, más preferentemente una pureza de al menos un 60%, aún más preferentemente una pureza de al menos un 80%, de la forma más preferida una pureza de al menos un 90% y de la forma aún más preferida una pureza de al menos un 95%, que se determina mediante SDS-PAGE.

La expresión «constructo de ácido nucleico», tal como se utiliza en la presente, se refiere a una molécula de ácido nucleico, ya sea mono- o bicatenaria, que se aísla de un gen de origen natural o que se modifica para que contenga segmentos de ácidos nucleicos de una manera que no existiría de otro modo en la naturaleza. La expresión constructo de ácido nucleico es sinónima de la expresión «casete de expresión» cuando el constructo de ácido nucleico contiene las secuencias de control requeridas para la expresión de una secuencia codificante de la presente invención. Según existe constancia en la técnica, las secuencias de control incluyen todos los componentes que son necesarios o favorables para la expresión de un polinucleótido que codifica un polipéptido de la presente invención. Cada secuencia de control puede ser nativa o exógena respecto a la secuencia de nucleótidos que codifica el polipéptido. Tales secuencias de control incluyen, sin carácter limitante, un líder, una secuencia de poliadenilación, una secuencia de

propéptido, un promotor, una secuencia de péptido señal y un terminador de la transcripción. Como mínimo, las secuencias de control incluyen un promotor y señales de detención de la traducción y la transcripción. Las secuencias de control se pueden proporcionar con conectores con el fin de introducir sitios de restricción específicos que faciliten el ligamento de las secuencias de control con la región codificante de la secuencia de nucleótidos que codifica un polipéptido.

La expresión «vector de expresión recombinante» se refiere a vectores de expresión recombinantes que comprenden un polinucleótido de la presente invención, un promotor y señales de detención de la traducción y la transcripción. Los diferentes ácidos nucleicos y secuencias de control que se han descrito anteriormente se pueden unir entre sí para producir un vector de expresión recombinante, el cual puede incluir uno o más sitios de restricción adecuados para permitir la inserción o sustitución de la secuencia de nucleótidos que codifica el polipéptido en tales sitios.

La expresión «célula hospedadora recombinante» se debe interpretar en la presente de acuerdo con la técnica. Según existe constancia en la técnica, las moléculas de polinucleótidos recombinantes (p. ej., ADN) son moléculas de polinucleótidos (p. ej., ADN) formadas mediante métodos de laboratorio de recombinación genética (tales como clonación molecular) para agrupar material genético de múltiples fuentes, con lo que se crean secuencias que no se encontrarían de otro modo en organismos biológicos. Según sobreentiende el experto, una célula hospedadora recombinante comprende moléculas de polinucleótidos recombinantes (p. ej., ADN) y, por consiguiente, no se sobreentenderá que una célula hospedadora recombinante cubra una célula de origen natural tal como, p. ej., una célula de *Dactylopius coccus* de origen natural.

La expresión «Identidad secuencial» se refiere al grado de relación existente entre dos secuencias de aminoácidos o entre dos secuencias de nucleótidos.

A los efectos de la presente invención, el grado de identidad secuencial entre dos secuencias de aminoácidos se determina utilizando el algoritmo de Needleman-Wunsch (Needleman y Wunsch, 1970, *J. Mol. Biol.* 48: 443-453) según se implementa en el programa Needle del paquete EMBOSS (*EMBOSS: The European Molecular Biology Open Software Suite*, Rice *et al.*, 2000, *Trends Genet.* 16: 276-277), preferentemente la versión 3.0.0 o una versión posterior. Los parámetros opcionales utilizados son una penalización por la abertura de un hueco de 10, una penalización por la extensión de un hueco de 0.5 y la matriz de sustitución EBLOSUM62 (versión EMBOSS de BLOSUM62). El resultado de la «identidad más larga» etiquetada por Needle (que se obtiene utilizando la opción no resumida) se utiliza como el porcentaje de identidad y se calcula como se indica continuación:

$$(\text{Residuos idénticos} \times 100) / (\text{Longitud de la alineación} - \text{Número total de huecos en la alineación}).$$

A los efectos de la presente invención, el grado de identidad secuencial entre dos secuencias de nucleótidos se determina utilizando el algoritmo de Needleman-Wunsch (Needleman y Wunsch, 1970, *supra*) según se implementa en el programa Needle del paquete EMBOSS (*EMBOSS: The European Molecular Biology Open Software Suite*, Rice *et al.*, 2000, *supra*), preferentemente la versión 3.0.0 o una versión posterior. Los parámetros opcionales utilizados son una penalización por la abertura de un hueco de 10, una penalización por la extensión de un hueco de 0.5 y la matriz de sustitución EDNAFULL (versión EMBOSS de NCBI NUC4.4). El resultado de la «identidad más larga» etiquetada por Needle (que se obtiene utilizando la opción no resumida) se utiliza como el porcentaje de identidad y se calcula como se indica a continuación:

$$(\text{Desoxirribonucleótidos idénticos} \times 100) / (\text{Longitud de la alineación} - \text{Número total de huecos en la alineación}).$$

A continuación se describen las realizaciones de la presente invención únicamente a modo de ejemplo.

DIBUJOS

Figura 1: Presentación esquemática de la actividad de glicosiltransferasa relevante de la glicosiltransferasa novedosa clonada/aislada descrita en la presente de SEQ ID NO:2; según se ilustra en la figura, se determinó que era capaz de conjugar glucosa con los aglicones ácido flavokermésico (FK) y ácido kermésico (AK).

Figura 2: Producción de glucósidos del ácido flavokermésico y el ácido kermésico utilizando OsCGT y SbUGT85B1. Análisis de LC-MS de productos glucosilados formados en ensayos que contienen UDP-glucosa y ácido flavokermésico (FK) o ácido kermésico (AK). Lisado crudo procedente de la cepa Xjb de *E.coli* (control negativo) incubada con FK (A) o AK (C). Lisado crudo procedente de células Xjb que expresan OsCGT incubadas con FK (B) o AK (D). Lisado crudo procedente de células Xjb que expresan SbUGT85B1 incubadas con FK (E) o AK (F). FK (G) y AK (H) como sustratos solos. Se indican los cromatogramas de iones totales (TIC, por sus siglas en inglés) y cromatogramas de iones extraídos para m/z 313[M-H]⁻, m/z 329[M-H]⁻, m/z 475 [M-H]⁻, m/z 491 [M-H]⁻, correspondientes a FK, AK, FK-monoglucósido y AK-monoglucósido. Los tiempos de retención de los picos se indican en minutos.

DESCRIPCIÓN DETALLADA DE LA INVENCION

Un polipéptido de tipo glicosiltransferasa aislado novedoso según se describe en la presente

Cuando en la presente se hace referencia a un polipéptido de tipo glicosiltransferasa aislado según se describe en la presente, se hace referencia a un polipéptido de tipo glicosiltransferasa aislado del primer aspecto de la invención y/o realizaciones relevantes en la presente de este.

- 5 Según se ha discutido anteriormente, la expresión «polipéptido aislado» se refiere esencialmente a que el polipéptido está aislado de su entorno natural. El polipéptido de tipo glicosiltransferasa novedoso que se describe en la presente según se muestra en la SEQ ID NO: 2 se aisló del insecto *Dactylopius coccus*. Por consiguiente, según sobrentiende el experto en el presente contexto, la expresión «polipéptido aislado» no cubre el polipéptido de tipo glicosiltransferasa de la SEQ ID NO: 2 cuando se encuentra presente de forma natural en el genoma de *Dactylopius coccus*.
- 10 Preferentemente, el polipéptido de tipo glicosiltransferasa aislado según se describe en la presente denota un preparado del polipéptido que contiene como máximo un 10%, preferentemente como máximo un 8%, más preferentemente como máximo un 6%, más preferentemente como máximo un 5%, más preferentemente como máximo un 4%, como máximo un 3%, aún más preferentemente como máximo un 2%, de la forma más preferida como máximo un 1% y de la forma aún más preferida como máximo un 0.5% en peso de otro material polipeptídico con el
- 15 que se asocia de forma nativa.

Según sobreentiende el experto, la expresión «otro material polipeptídico con el que se asocia de forma nativa», en relación con el polipéptido de tipo glicosiltransferasa novedoso que se muestra en la SEQ ID NO: 2, se puede interpretar como que se relaciona con otro material polipeptídico con el que se asocia de forma nativa en *Dactylopius coccus*.

- 20 En algún caso, se puede preferir que el polipéptido de tipo glicosiltransferasa aislado según se describe en la presente se refiera a un polipéptido con una pureza de al menos un 20%, preferentemente una pureza de al menos un 40%, más preferentemente una pureza de al menos un 60%, aún más preferentemente una pureza de al menos un 80%, de la forma más preferida una pureza de al menos un 90% y de la forma aún más preferida una pureza de al menos un 95%, que se determina mediante SDS-PAGE.

- 25 Basándose, p. ej., en la información de la secuencia descrita en la presente, el hecho de obtener un polipéptido de tipo glicosiltransferasa aislado según se describe en la presente es un trabajo rutinario para el experto.

Esto se puede llevar a cabo, p. ej., mediante expresión recombinante en una célula hospedadora recombinante adecuada de acuerdo con procedimientos conocidos en la técnica.

- 30 Por consiguiente, no se cree necesario describir tales procedimientos de expresión recombinante conocidos estándar con gran detalle en la presente.

Preferentemente, el polipéptido de tipo glicosiltransferasa aislado según se describe en la presente es capaz de:

(I): conjugar glucosa con ácido flavokermésico (FK); y

(II): conjugar glucosa con ácido kermésico (AK).

Una realización preferida se refiere a donde el polipéptido de tipo glicosiltransferasa del primer aspecto es:

- 35 (a) un polipéptido que comprende una secuencia de aminoácidos que presenta una identidad de al menos un 80% con los aminoácidos 1-515 de la SEQ ID NO:2; más preferentemente

- (a) un polipéptido que comprende una secuencia de aminoácidos que presenta una identidad de al menos un 90% con los aminoácidos 1-515 de la SEQ ID NO:2; aún más preferentemente (a) un polipéptido que comprende una secuencia de aminoácidos que presenta una identidad de al menos un 95% con los aminoácidos 1-515 de la SEQ ID NO:2; y de la forma más preferida
- 40

(a) un polipéptido que comprende una secuencia de aminoácidos que presenta una identidad de al menos un 98% con los aminoácidos 1-515 de la SEQ ID NO:2.

Puede preferirse que el polipéptido de tipo glicosiltransferasa del primer aspecto sea un polipéptido que comprenda una secuencia de aminoácidos con los aminoácidos 1-515 de la SEQ ID NO:2.

- 45 Una realización preferida se refiere a donde el polipéptido de tipo glicosiltransferasa del primer aspecto es:

(b) un polipéptido que comprende una secuencia de aminoácidos que presenta una identidad de al menos un 80% con los aminoácidos 20-468 de la SEQ ID NO:2; más preferentemente

(b) un polipéptido que comprende una secuencia de aminoácidos que presenta una identidad de al menos un 90% con los aminoácidos 20-468 de la SEQ ID NO:2; aún más preferentemente

(b) un polipéptido que comprende una secuencia de aminoácidos que presenta una identidad de al menos un 95% con los aminoácidos 20-468 de la SEQ ID NO:2; y de la forma más preferida

(b) un polipéptido que comprende una secuencia de aminoácidos que presenta una identidad de al menos un 98% con los aminoácidos 20-468 de la SEQ ID NO:2.

- 5 Puede preferirse que el polipéptido de tipo glicosiltransferasa del primer aspecto sea un polipéptido que comprenda una secuencia de aminoácidos con los aminoácidos 20-468 de la SEQ ID NO:2.

Una realización preferida se refiere a donde el polipéptido de tipo glicosiltransferasa del primer aspecto es:

- 10 (c) un polipéptido que está codificado por un polinucleótido que se hibrida en condiciones de rigurosidad al menos media-alta con (i) los nucleótidos 1-1548 de la SEQ ID NO:1 o (ii) una hebra complementaria de (i); más preferentemente

(c) un polipéptido que está codificado por un polinucleótido que se hibrida en condiciones de rigurosidad al menos alta con (i) los nucleótidos 1-1548 de la SEQ ID NO:1 o (ii) una hebra complementaria de (i); y aún más preferentemente

- 15 (c) un polipéptido que está codificado por un polinucleótido que se hibrida en condiciones de al menos mucha rigurosidad con (i) los nucleótidos 1-1548 de la SEQ ID NO:1 o (ii) una hebra complementaria de (i).

El hecho de preparar una variante de un polipéptido de tipo glicosiltransferasa aislado según se describe en la presente, es decir, una variante donde, p. ej., uno o más aminoácidos de, p. ej., la SEQ ID NO:2 haya sido modificado/alterado, es un trabajo rutinario para el experto.

- 20 Además, según tiene constancia el experto, si tales cambios de la variante no son demasiado drásticos, será razonable asumir que la enzima mantendría su actividad de GT relevante.

Una realización preferida se refiere a donde el polipéptido de tipo glicosiltransferasa del primer aspecto es:

- 25 (a) un polipéptido que comprende una secuencia de aminoácidos con los aminoácidos 1-515 de la SEQ ID NO:2 o una variante de este, donde la variante comprende una alteración en una o más (varias) posiciones de la SEQ ID NO:2 y donde la variante comprende menos de 50 alteraciones, más preferentemente menos de 40 alteraciones, aún más preferentemente menos de 20 alteraciones y de la forma más preferida menos de 10 alteraciones.

En una realización preferida, la expresión «una alteración en una o más (varias) posiciones de la SEQ ID NO:2» se refiere a de 1 a 10 alteraciones en la SEQ ID NO:2.

- 30 De acuerdo con la técnica, el término «variante» significa en la presente un péptido que tiene una actividad de GT relevante que comprende una alteración, es decir, una sustitución, inserción y/o delección, en una o más (varias) posiciones. Una sustitución significa un reemplazo de un aminoácido que ocupa una posición por un aminoácido diferente; una delección significa la eliminación de un aminoácido que ocupa una posición; y una inserción significa añadir 1-3 aminoácidos adyacentes a un aminoácido que ocupa una posición.

El aminoácido puede ser un aminoácido natural o no natural, por ejemplo, la sustitución con, p. ej., en particular los isómeros D (o formas D) de, p. ej., D-alanina podría ser posible teóricamente.

- 35 En una realización preferida, el polipéptido de tipo glicosiltransferasa del primer aspecto es una GT que está unida a la membrana o que es insoluble en agua.

Polinucleótido aislado que comprende una secuencia de nucleótidos que codifica el polipéptido de tipo glicosiltransferasa según se describe en la presente

- 40 Según se ha discutido anteriormente, un segundo aspecto de la presente invención se refiere a un polinucleótido aislado que comprende una secuencia de nucleótidos que codifica el polipéptido del primer aspecto y/o realizaciones relevantes en la presente de este.

La expresión «polinucleótido aislado» se puede denominar de forma alternativa en la presente «polinucleótido clonado».

- 45 Según se ha discutido anteriormente, la expresión «polinucleótido aislado» se refiere esencialmente a que el polinucleótido está aislado de su entorno natural; dicho de otro modo, que el preparado del polinucleótido está esencialmente exento de otro material polinucleotídico con el que se asocia de forma nativa. La secuencia de polinucleótido que codifica la glicosiltransferasa novedosa clonada/aislada descrita se muestra en la SEQ ID NO: 1 y se aisló del insecto *Dactylopius coccus*. Por consiguiente, según sobreentiende el experto, la expresión polinucleótido aislado no cubre el polinucleótido de la SEQ ID NO: 1 cuando se encuentra presente de forma natural en el genoma de *Dactylopius coccus*.
- 50

La expresión «polinucleótido aislado» se refiere esencialmente a que el polinucleótido aislado se encuentra en una forma adecuada para su uso en sistemas de producción de proteínas modificadas genéticamente.

De este modo, un polinucleótido aislado contiene como máximo un 10%, preferentemente como máximo un 8%, más preferentemente como máximo un 6%, más preferentemente como máximo un 5%, más preferentemente como máximo un 4%, más preferentemente como máximo un 3%, aún más preferentemente como máximo un 2%, de la forma más preferida como máximo un 1% y de la forma aún más preferida como máximo un 0.5% en peso de otro material polinucleotídico con el que se asocia de forma nativa.

Basándose, p. ej., en la información de la secuencia descrita en la presente, el hecho de obtener un polinucleótido aislado según se describe en la presente es un trabajo rutinario para el experto.

Esto se puede llevar a cabo, p. ej., mediante expresión recombinante en una célula hospedadora recombinante adecuada de acuerdo con procedimientos conocidos en la técnica.

Por consiguiente, no se cree necesario describir tales procedimientos de expresión recombinante conocidos estándar con gran detalle en la presente.

Un constructo de ácido nucleico que comprende el polinucleótido aislado según se describe en la presente

Según se ha discutido anteriormente, un tercer aspecto de la presente invención se refiere a un constructo de ácido nucleico que comprende el polinucleótido aislado del segundo aspecto y/o realizaciones relevantes en la presente de este, unido operablemente a una o más secuencias de control que dirigen la producción del polipéptido en un hospedador de expresión.

De acuerdo con la técnica, la expresión «constructo de ácido nucleico», tal como se utiliza en la presente, se refiere a una molécula de ácido nucleico, ya sea mono- o bicatenaria, que se aísla de un gen de origen natural o que se modifica para que contenga segmentos de ácidos nucleicos de una manera que no existiría de otro modo en la naturaleza.

La expresión constructo de ácido nucleico es sinónima de la expresión «casete de expresión» cuando el constructo de ácido nucleico contiene las secuencias de control requeridas para la expresión de una secuencia codificante de la presente invención. Según existe constancia en la técnica, las secuencias de control incluyen todos los componentes que son necesarios o favorables para la expresión de un polinucleótido que codifica un polipéptido de la presente invención.

Basándose, p. ej., en la información de la secuencia que se describe en la presente, el hecho de preparar un constructo de ácido nucleico relevante es un trabajo rutinario para el experto, por ejemplo, basándose en la técnica anterior el experto conoce numerosas secuencias de control adecuadas diferentes para la expresión de un polinucleótido que codifica un polipéptido de la presente invención.

Por consiguiente, no se cree necesario describir tales elementos técnicos conocidos estándar con gran detalle en la presente.

Un vector de expresión recombinante que comprende el constructo de ácido nucleico según se describe en la presente

Según se ha discutido anteriormente, un cuarto aspecto de la presente invención se refiere a un vector de expresión recombinante que comprende el constructo de ácido nucleico del tercer aspecto y/o realizaciones relevantes en la presente de este.

De acuerdo con la técnica, la expresión «vector de expresión recombinante» se refiere a vectores de expresión recombinantes que comprenden un polinucleótido de la presente invención, un promotor y señales de detención de la traducción y la transcripción. Los diferentes ácidos nucleicos y secuencias de control que se han descrito anteriormente se pueden unir entre sí para producir un vector de expresión recombinante, el cual puede incluir uno o más sitios de restricción adecuados para permitir la inserción o sustitución de la secuencia de nucleótidos que codifica el polipéptido en tales sitios.

Basándose, p. ej., en la información de la secuencia que se describe en la presente, el hecho de preparar un vector de expresión recombinante relevante es un trabajo rutinario para el experto, por ejemplo, basándose en la técnica anterior el experto conoce numerosos promotores y señales de detención de la traducción y la transcripción adecuados diferentes.

Por consiguiente, no se cree necesario describir tales elementos técnicos conocidos estándar con gran detalle en la presente.

Una célula hospedadora recombinante que comprende el constructo de ácido nucleico según se describe en la presente

Según se ha discutido anteriormente, un quinto aspecto de la presente invención se refiere a una célula hospedadora recombinante que comprende el constructo de ácido nucleico del tercer aspecto y/o realizaciones relevantes en la presente de esta.

La expresión «célula hospedadora recombinante» se debe interpretar en la presente de acuerdo con la técnica. Según existe constancia en la técnica, las moléculas de polinucleótidos recombinantes (p. ej., ADN) son moléculas de polinucleótidos (p. ej., ADN) formadas mediante métodos de laboratorio de recombinación genética (tales como clonación molecular) para agrupar material genético de múltiples fuentes, con lo que se crean secuencias que no se encontrarían de otro modo en organismos biológicos. Según sobreentiende el experto, una célula hospedadora recombinante comprende moléculas de polinucleótidos recombinantes (p. ej., ADN) y, por consiguiente, no se sobreentenderá que una célula hospedadora recombinante cubra una célula de origen natural tal como, p. ej., una célula de *Dactylopius coccus* de origen natural.

Basándose, p. ej., en la información de la secuencia que se describe en la presente, el hecho de preparar una célula hospedadora recombinante relevante es un trabajo rutinario para el experto, por ejemplo, basándose en la técnica anterior el experto conoce numerosas células hospedadoras recombinantes adecuadas diferentes que se han utilizado durante años como células hospedadoras recombinantes para, p. ej., la expresión de diferentes polipéptidos de interés.

La célula hospedadora recombinante puede ser cualquier célula adecuada tal como cualquier célula eucariota [p. ej., células de mamífero (tales como, p. ej., células de ovario de hámster chino (CHO)) o una célula vegetal] o cualquier célula procariota.

Se prefiere en particular cuando la célula hospedadora recombinante es una célula vegetal que produce ácido flavokermésico/ácido kermésico u otro compuesto relacionado tal como, p. ej., una célula vegetal de ruibarbo.

Preferentemente, la célula hospedadora recombinante es una célula seleccionada del grupo constituido por una célula de hongo filamentoso y una célula de un microorganismo.

Los hongos filamentosos incluyen todas las formas filamentosas de la subdivisión *Eumycota* y *Oomycota* (según definen Hawksworth *et al.*, 1995, *supra*). Los hongos filamentosos se caracterizan por un micelio vegetativo compuesto por quitina, celulosa, glucano, quitosano, manano y otros polisacáridos complejos. Su crecimiento vegetativo se produce mediante la elongación de las hifas y el catabolismo de carbono es obligatoriamente aeróbico. Por el contrario, el crecimiento vegetativo por parte de levaduras tales como *Saccharomyces cerevisiae* se produce formando brotes a partir de un tallo unicelular y el catabolismo de carbono puede ser fermentativo.

Puede preferirse que la célula de hongo filamentoso sea una célula de una especie de, sin carácter limitante, *Acremonium*, *Aspergillus*, *Fusarium*, *Humicola*, *Mucor*, *Myceliophthora*, *Neurospora*, *Penicillium*, *Thielavia*, *Tolypocladium* y *Trichoderma* o un teleomorfo o sinónimo de estas.

Una célula de *Aspergillus* preferida es *Aspergillus niger* o *Aspergillus oryzae*.

Una célula de un microorganismo preferida en la presente es una célula de un microorganismo seleccionada del grupo constituido por una célula de levadura y una célula procariota.

Una célula de levadura preferida es una célula de levadura seleccionada del grupo constituido por ascomicetos, basidiomicetos y hongos imperfectos. Preferentemente, una célula de levadura seleccionada del grupo constituido por ascomicetos.

Una célula de levadura de ascomicetos preferida seleccionada del grupo constituido por *Ashbya*, *Botryosascus*, *Debaryomyces*, *Hansenula*, *Kluveromyces*, *Lipomyces*, *Saccharomyces* spp, p. ej., *Saccharomyces cerevisiae*, *Pichia* spp., *Schizosaccharomyces* spp.

Una célula de levadura preferida es una célula de levadura seleccionada del grupo constituido por *Saccharomyces* spp, p. ej., *Saccharomyces cerevisiae*, y *Pichia* spp.

Una célula procariota preferida se selecciona del grupo constituido por *Bacillus*, *Streptomyces*, *Corynebacterium*, *Pseudomonas*, bacterias del ácido láctico y una célula de *E. coli*.

Una célula de *Bacillus* preferida es *B. subtilis*, *B. amyloliquefaciens* o *B. licheniformis*.

Una célula de *Streptomyces* preferida es *S. setonii* o *S. coelicolor*.

Una célula de *Corynebacterium* preferida es *C. glutamicum*.

Una célula de *Pseudomonas* preferida es *P. putida* o *P. fluorescens*.

Un método para producir un glicósido del ácido flavokermésico (FK) y/o un glicósido del ácido kermésico (AK)

Según se ha discutido anteriormente, un sexto aspecto de la presente invención se refiere a un método para producir un glicósido del ácido flavokermésico (FK) y/o un glicósido del ácido kermésico (AK), donde el método comprende los siguientes pasos:

5 (A): poner en contacto *in vivo* o *in vitro* en una célula hospedadora recombinante que comprende un gen de glicosiltransferasa que codifica una glicosiltransferasa:

(a1): ácido flavokermésico (FK) con una glicosiltransferasa capaz de glicosilar el ácido flavokermésico en condiciones adecuadas, donde se produce el glicósido del ácido flavokermésico; y/o

(a2): ácido kermésico (AK) con una glicosiltransferasa capaz de glicosilar el ácido kermésico en condiciones adecuadas, donde se produce el glicósido del ácido kermésico.

10 Se puede preferir que la célula hospedadora recombinante del paso (A) sea una célula hospedadora recombinante que comprenda un gen de glicosiltransferasa recombinante que codifique una glicosiltransferasa.

Preferentemente, la glicosiltransferasa del paso (a2) es una glucosiltransferasa y, por consiguiente, en el paso (a2) se produce un glucósido del ácido kermésico, preferentemente donde el glucósido del ácido kermésico producido es ácido carmínico (la Figura 1 de la presente muestra la estructura del ácido carmínico).

15 Se puede preferir que la glicosiltransferasa del paso (a1) sea una glucosiltransferasa y, por consiguiente, en el paso (a1) se produce un glucósido del ácido flavokermésico, preferentemente donde el glucósido del ácido flavokermésico producido es el compuesto DcII (la Figura 1 de la presente muestra la estructura del compuesto DcII).

Cuando el compuesto producido en el paso (a1) es DcII, se puede preferir utilizar este DcII como un intermedio para preparar ácido carmínico.

20 Esto se puede llevar a cabo mediante síntesis química y el experto sabe cómo hacerlo basándose en su conocimiento común general.

Como alternativa, se puede llevar a cabo enzimáticamente, p. ej., utilizando una oxigenasa adecuada. Un ejemplo de una oxigenasa adecuada es una enzima de la superfamilia de monooxigenasas del citocromo P450 (que se abrevia oficialmente como CYP). Otros ejemplos son flavina-monooxigenasas o diferentes tipos de dioxigenasas; no se debe considerar que esta lista excluya la participación de otras clases de enzimas.

Según existe constancia en la técnica, la reacción más habitual catalizada por los citocromos P450 es una reacción de tipo monooxigenasa, p. ej., la inserción de un átomo de oxígeno en un sustrato.

30 Según sobreentiende el experto en el presente contexto, se debe sobreentender que las expresiones aglicones ácido flavokermésico (FK) y/o ácido kermésico (AK) del paso (a) del método del sexto aspecto según se discute en la presente son los compuestos específicos de FK y/o AK que se muestran en la Figura 1 y análogos equivalentes de estos compuestos específicos con sustituyentes menores (p. ej., un éster metílico del FK).

35 Según sobreentiende el experto, si se utiliza el éster metílico del FK como aglicón en el paso (a) del método del sexto aspecto, entonces se generará, mediante el paso de glicosilación, un glicósido del éster metílico del FK, el cual mediante la eliminación rutinaria del grupo metilo generará DcII; por consiguiente, el aglicón éster metílico del FK se puede ver como un equivalente del aglicón FK en relación con el método del sexto aspecto según se discute en la presente.

En el paso (a) del método del sexto aspecto se especifica que se utiliza una glicosiltransferasa capaz de glicosilar FK y/o AK; por consiguiente, se sobreentiende que la GT debe ser capaz de realizar esta transformación.

Se puede preferir purificar el glicósido producido en el paso (A), es decir, en el paso (a1) y/o en el paso (a2).

40 Por consiguiente, se puede preferir que el método del sexto aspecto comprenda un paso adicional (B) con los siguientes pasos:

(B): purificar el glicósido producido en el paso (a1) y/o en el paso (a2) de modo que se obtenga una composición, donde al menos un 5% p/p (preferentemente al menos un 10% p/p, más preferentemente al menos un 50% p/p y de la forma más preferida al menos un 80% p/p) de los compuestos en la composición sea el glicósido del ácido flavokermésico y/o el glicósido del ácido kermésico producido.

El experto sabe cómo purificar tales compuestos glicósidos y esto se puede llevar a cabo de acuerdo con la técnica.

El paso de purificación (B) puede ser particularmente preferido cuando:

- el glicósido producido en el paso (a2) es ácido carmínico;
- el glicósido producido en el paso (a1) es el compuesto DcII; y/o

- el glicósido producido en el paso (a1) es el compuesto DcII y se utiliza como intermedio para preparar ácido carmínico.

Según se discute en la presente, en los Ejemplos prácticos se puso en contacto *in vitro* ácido flavokermésico (FK) y/o ácido kermésico (AK) con la glicosiltransferasa de la SEQ ID NO:2. Para el experto, se puede ver como un trabajo rutinario el hecho de llevar a cabo un paso de puesta en contacto *in vitro* de este tipo.

La glicosiltransferasa de la SEQ ID NO:2 se expresó recombinantemente en una célula de levadura (remítase al Ejemplo práctico de la presente), por consiguiente, en un Ejemplo práctico de la presente se preparó una célula hospedadora de levadura recombinante que comprendía un gen de glicosiltransferasa recombinante que codificaba una glicosiltransferasa de SEQ ID NO:2.

Se cree que si se añadiera ácido flavokermésico (FK) y/o ácido kermésico (AK) en condiciones adecuadas a un medio de fermentación, el o los compuestos FK y/o AK entrarían en, p. ej., células de levadura fermentadas en el medio; por consiguiente, si, p. ej., las células de levadura son células hospedadoras de levadura recombinantes que comprenden un gen de glicosiltransferasa recombinante que codifica una glicosiltransferasa, entonces se produciría una puesta en contacto *in vivo* en una célula hospedadora recombinante de FK y/o AK con una glicosiltransferasa.

En una realización preferida, la puesta en contacto del paso (A) se realiza *in vivo* y la célula hospedadora recombinante es una célula de levadura, preferentemente donde la célula de levadura se selecciona del grupo constituido por *Saccharomyces spp* (p. ej., *Saccharomyces cerevisiae*) y *Pichia spp*.

Anteriormente se han descrito células hospedadoras recombinantes preferidas; estas células hospedadoras recombinantes preferidas también pueden ser células hospedadoras recombinantes preferidas en relación con el método del sexto aspecto de la presente invención.

En el presente contexto, se puede decir que se encuentra dentro del conocimiento común del experto el hecho de identificar una célula hospedadora recombinante adecuada para llevar a cabo el paso (A) de puesta en contacto *in vivo* del método del sexto aspecto y no se cree que sea necesario describir esto con gran detalle en la presente.

Anteriormente se ha discutido que las células hospedadoras recombinantes preferidas pueden ser, p. ej., una célula de un microorganismo o una célula de hongo filamentoso; estas células pueden ser células hospedadoras recombinantes preferidas en relación con el método del sexto aspecto.

Quizá sea posible obtener una célula hospedadora recombinante (p. ej., una célula hospedadora recombinante de un microorganismo) que comprenda un gen que codifique un producto implicado en la ruta biosintética que conduce al ácido flavokermésico (FK) y/o al ácido kermésico (AK) y una célula hospedadora recombinante de este tipo podría preferirse en la presente.

Por consiguiente, se puede preferir que la puesta en contacto en el paso (A) sea una puesta en contacto *in vivo* en una célula hospedadora recombinante que comprenda un gen de glicosiltransferasa recombinante que codifique una glicosiltransferasa y un gen que codifique un producto implicado en la ruta biosintética que conduce al ácido flavokermésico(FK) y/o al ácido kermésico (AK).

Según se discute en un Ejemplo práctico de la presente, la GT de SEQ ID NO:2 está unida a la membrana o es hidrófoba/insoluble *in vivo* y en agua. Cuando las células de producción o fracciones de células que contienen la GT unida a la membrana se separan del producto (p. ej., ácido carmínico), la GT puede no estar esencialmente presente en la fracción en la que esté presente el producto más soluble/producto hidrófilo. Esto supone una ventaja a la hora de obtener un producto final (p. ej., composición/producto de ácido carmínico) que esté de forma esencial totalmente exento de la GT recombinante.

Debido a que los sustratos glicosilados por la GT pueden ser aglicones hidrófobos, cabría esperar que los aglicones se acumularan parcialmente en las membranas y otras partes hidrófobas de las células de producción. Mediante el uso de una GT unida a la membrana, se obtiene una glicosilación más eficaz de los compuestos hidrófobos presentes en, p. ej., las membranas.

Por consiguiente, en una realización preferida, la glicosiltransferasa utilizada en el método del sexto aspecto es una GT que está unida a la membrana o que es insoluble en agua.

En una realización preferida, la glicosiltransferasa en el paso (A) del método del sexto aspecto es una glicosiltransferasa del primer aspecto y/o realizaciones relevantes en la presente de esta.

Según se discute en la presente, los resultados/datos identificados de los Ejemplos prácticos 4 muestran que las enzimas de tipo GT relevantes de la presente se pueden identificar en, p. ej., plantas de arroz y sorgo.

La secuencia del polipéptido de sorgo (número de ID de Genbank: AAF17077.1) se muestra como la SEQ ID NO: 4 de la presente.

La secuencia del polipéptido de arroz (número de ID de Genbank: CAQ77160.1) se muestra como la SEQ ID NO: 5 de la presente.

5 Puede ser relevante que la glicosiltransferasa en el paso (A) del método del sexto aspecto sea una glicosiltransferasa que comprenda una secuencia de aminoácidos que presente una identidad de al menos un 70% (preferentemente al menos un 80%, más preferentemente al menos un 90% y aún más preferentemente al menos un 98%) con los aminoácidos 1-492 de la SEQ ID NO:4.

10 Puede ser relevante que la glicosiltransferasa en el paso (A) del método del sexto aspecto sea una glicosiltransferasa que comprenda una secuencia de aminoácidos que presente una identidad de al menos un 70% (preferentemente al menos un 80%, más preferentemente al menos un 90% y aún más preferentemente al menos un 98%) con los aminoácidos 1-471 de la SEQ ID NO:5.

Ejemplos

Ejemplo 1 - Clonación de GT de *D. coccus* y evaluación de su actividad FK y AK

Materiales y métodos

Purificación de ADN y ARNm

15 *Dactylopius coccus* congelados frescos (se obtuvieron de Lanzarote). *Porphyrophora polonica* congelada fresca se obtuvo de Polonia. Los insectos congelados se trituraron hasta obtener un polvo en nitrógeno líquido y se purificó el ADN/ARN: el ADN se purificó utilizando el kit *DNeasy Blood & Tissue* (Qiagen), de acuerdo con el protocolo del proveedor. El ARN se purificó utilizando el kit *RNeasy mini* (Qiagen) de acuerdo con el protocolo del proveedor.

20 El ARNm eucariota se convirtió en ADNc utilizando el kit *RT² Easy First Strand* (Qiagen) de acuerdo con el protocolo del proveedor utilizando cebadores de poli-dT en la reacción de la transcriptasa inversa.

Secuenciación de ADN y ARN:

El ADN y ADNc se enviaron para la secuenciación a BGI (Shenzhen, China) para una secuenciación utilizando tecnología de Illumina de extremos apareados de 100 pb de acuerdo con el protocolo de Illumina con una cobertura de aproximadamente 60-100X y el resultado en formato de datos fastq.

Análisis de las secuencias de ADN y ARN/ADNc:

Las secuencias fastq obtenidas de ADN y ARN/ADNc se importaron en Genomic Workbench versión 5.4 (CLC-bio, Århus, Dinamarca) y se acoplaron en contigs utilizando el algoritmo de acoplamiento *de novo*. Los archivos de los resultados se exportaron en formato FASTA.

30 A continuación, los archivos FASTA de ARN/ADNc se importaron en IOGMA v. 10 (Genostar, Grenoble, Francia) y los posibles genes se identificaron utilizando el buscador de genes procariotas basado en la matriz oculta de Markov.

Los posibles genes se anotaron utilizando BLAST (siglas en inglés referentes a la herramienta de búsqueda por alineación local básica) frente a genbank (NCBI) utilizando además la secuencia de nucleótidos como la secuencia de la proteína traducida. Los posibles genes también se anotaron mediante una comparación por similitud respecto a bases de datos PFAM de familias de proteínas.

Preparación de fracciones proteicas de *D. coccus*

35 Se homogeneizaron tres gramos de insectos *D. coccus* frescos en 120 mL de tampón de aislamiento [sacarosa 350 mM, Tricina 20 mM (pH 7.9), NaCl 10 mM, DTT 5 mM, PMSF 1 mM] que contenía 0.3 g de polivinilpolipirrolidona. El homogenato se filtró a través de un trapo de nailon (malla de 22 µm) y se centrifugó durante (10 min, 10 000 x g a 4 °C). El sobrenadante se centrifugó (1 h, 105 000 x g, a 4 °C), para proporcionar una fracción proteica soluble y una fracción proteica unida a la membrana. La fracción proteica soluble se concentró hasta 1 mL y se intercambió el tampón con Tricina 20 mM (pH 7.9), DTT 5 mM utilizando dispositivos de ultracentrifugación Amicon 3K (Millipore). El pellet de proteína unida a la membrana se lavó 3 veces resuspendiendo el pellet en 60 mL de Tricina 20 mM (pH 7.9), DTT 5 mM utilizando un pincel *marten* y a continuación volviendo a centrifugar. El pellet de proteína unida a la membrana se resuspendió finalmente en 1 mL de Tricina 20 mM (pH 7.9), DTT 5 mM. La fracción proteica soluble y la fracción proteica unida a la membrana se analizaron para determinar su actividad de glicosilación.

Purificación de una actividad de GT específica del ácido flavokermésico/ácido kermésico procedente de proteínas de membrana de *D. coccus*

50 Una fracción proteica unida a la membrana aislada a partir de 3 g de insectos *D. coccus* frescos se solubilizó añadiendo un 1% (v/v) de Triton x-100 (forma reducida) y agitando suavemente durante 1.5 h en frío. La solución tratada con Triton x-100 se centrifugó (1 h, 105 000 x g, a 4 °C) y el sobrenadante se aisló y se aplicó a una columna empaquetada con 2 mL de Q-sepharose de flujo rápido (Pharmacia). La columna se lavó en 4 mL de tampón A [Tricina 20 mM (pH

7.9), un 0.1 % (v/v) de Triton x-100 (forma reducida), NaCl 50 mM] y las proteínas se eluyeron con Tricina 20 mM (pH 7.9), un 0.1 % (v/v) de Triton x-100 (forma reducida)] utilizando un gradiente discontinuo de NaCl desde 100 mM hasta 500 mM (con incrementos de 50 mM). Se recolectaron fracciones de 0.5 ml, se desalaron, se analizaron mediante SDS-PAGE y se monitorizaron para determinar su actividad de glucosilación utilizando el ensayo descrito para enzimas de glucosilación radiomarcadas. Una fracción que contenía actividad de GT específica del ácido flavokermésico/ácido kermésico enriquecida se sometió a un análisis de la huella de la masa peptídica.

Ensayos enzimáticos y detección del producto glucosídico

Se llevó a cabo la glucosilación del ácido flavokermésico y del ácido kermésico en mezclas de ensayo de 60 µL que contenían Tricina 20 mM (pH 7.9), UDP[14C]glucosa 3.3 µM y 20 µL de extracto proteico (proteína soluble o unida a la membrana). Las reacciones se incubaron durante 0.5 h a 30 °C y se terminaron añadiendo 180 µL de metanol. Las muestras se centrifugaron a 16 000 x g durante 5 min a 4 °C y el sobrenadante se aplicó sobre placas de TLC (placas de gel de sílice 60 F254; Merck). Los productos del ensayo se resolvieron en diclorometano:metanol:ácido fórmico (7:2:2, en volumen). Los productos radiomarcados se visualizaron utilizando un instrumento STORM 840 PhosphorImager (Molecular Dynamics, <http://www.moleculardynamics.com>).

Expresión de DcUGT2 con codones optimizados, DcUGT4 y DcUGT5 en *S. cerevisiae*

Se adquirió de GenScript una versión sintética con codones optimizados de DcUGT2 y dos posibles secuencias de GT diferentes procedentes del transcriptoma de *D. coccus* denominadas DcUGT4 y DcUGT5 para la expresión en levadura con sitios attL de recombinación Gateway flanqueantes. Los fragmentos sintéticos se utilizaron como modelos de PCR con cebadores específicos para generar las versiones marcadas con StrepII C-terminales correspondientes. Los seis constructos génicos (fragmentos marcados y no marcados) se clonaron en el plásmido de destino Gateway pYES-DEST52 (Invitrogen) utilizando la mezcla de enzimas LR clonaseII. Los seis constructos plasmídicos pYES-DEST52 se transformaron por separado en la cepa de levadura Invsc1 (Invitrogen) y los transformantes positivos se verificaron mediante PCR. La producción de proteínas heterólogas se llevó a cabo de acuerdo con las instrucciones del vector Gateway pYES-DEST52 (Invitrogen). La producción de proteína heteróloga marcada con StrepII se verificó mediante inmunotransferencia de Western utilizando un anticuerpo anti-Strep. Se preparó una fracción proteica unida a la membrana a partir de transformantes de levadura verificados según se describe en (D. Pompon, B. Louerat, A. Bronine, P. Urban, *Yeast expression of animal and plant P450s in optimized redox environments, Methods Enzymol.* 272 (1996) 51-64) y se cribó para determinar su actividad de glucosilación frente al ácido flavokermésico/ácido kermésico. La secuencia optimizada de levadura se muestra en la SEQ ID NO: 3 de la presente.

LC-MS-MS

La LC/MS se llevó a cabo en un instrumento Agilent Q-TOF con el siguiente sistema de HPLC:

Columna Kinetix 2.6 µ XB-C18 100A (100 x 4.60 mm, Phenomenex). Disolvente A (900 ml de agua desionizada, 100 ml de metanol y 50 ml de ácido fórmico). Disolvente B (700 ml de metanol, 300 ml de agua desionizada y 50 ml de ácido fórmico).

Flujo 0.8 ml/min. 35 °C.

Elución en gradiente. 0-1 min 100% de A, gradiente lineal hasta un 83% de A 3 min, gradiente lineal hasta un 63% de A 6 min, gradiente lineal hasta un 45% de A 9 min, gradiente lineal hasta un 27% de A 12 min, gradiente lineal hasta un 10% de A 15 min, gradiente lineal hasta un 3% de A 17 min, gradiente lineal hasta un 2% de A 19 min, gradiente lineal hasta un 0% de A 20 min, 0% de A 22 min, gradiente lineal hasta un 100% de A 25 min. Los tiempos de retención fueron de 7.6 min para el ácido carmínico, 7.8 min para DC II, 13.7 min para el ácido flavokermésico y 13.9 min Para el ácido kermésico.

Resultados

La capacidad para glicosilar ácido flavokermésico/ácido kermésico utilizando C14-UDP-glucosa como sustrato se detectó en insectos de *D. coccus* homogeneizados. Se demostró que la actividad estaba unida a la membrana y la actividad se purificó y las proteínas purificadas se sometieron a un análisis proteómico. Se demostró que la actividad enzimática procedía de un polipéptido con una secuencia correspondiente a nuestro gen candidato DcUGT2.

Según se ha discutido anteriormente, la enzima de tipo glicosiltransferasa relevante de la presente de SEQ ID NO: 2 se puede denominar en la presente «DcUGT2».

La secuencia de aminoácidos de DcUGT2 presenta menos de un 45% de homología respecto a cualquier glicosiltransferasa conocida.

Sabiendo que la clonación de la secuencia de origen natural en levadura no proporcionó ninguna actividad enzimática relevante, rediseñamos la secuencia de nucleótidos de DcUGT2 para obtener una secuencia que codificara el mismo polipéptido pero que empleara codones de nucleótidos optimizados para *S. cerevisiae*, un proceso denominado optimización de codones (la secuencia optimizada de *S. cerevisiae* se muestra como SEQ ID No. 3 en la presente).

Posteriormente, la secuencia con codones optimizados de DcUGT2 se clonó y se expresó en levadura. La cepa de levadura heteróloga contiene una actividad enzimática unida a la membrana capaz de glucosilar ácido kermésico y ácido flavokermésico.

- 5 Después de obtener los datos de la huella de la masa peptídica a partir de una fracción proteica de *D. coccus* enriquecida con actividad de GT frente al ácido flavokermésico/ácido kermésico, hicimos coincidir las masas peptídicas con el conjunto de datos transcriptómicos e identificamos tres posibles UGT (DcUGT2, DcUGT4 y DcUGT5).

La expresión heteróloga de los tres candidatos en levadura reveló que únicamente una de estas UGT, denominada DcUGT2, era la responsable de la actividad de glucosilación observada frente al ácido flavokermésico/ácido kermésico en la fracción proteica de *D. coccus*.

- 10 Un tratamiento con Viscozyme del glucósido radiomarcado con C-14 generado mostró que era resistente a la hidrólisis, lo cual sugiere además que DcUGT2 es una C-GT, responsable de la producción de DCII y ácido carmínico.

Una LC-MS-MS mostró la formación de productos con el mismo tiempo de retención, espectro, masa molecular y patrón de degradación molecular que DcII y el ácido carmínico, respectivamente.

Conclusión

- 15 El resultado de este ejemplo 1 demostró que no era una tarea fácil aislar/clonar la enzima de tipo glicosiltransferasa relevante de la presente de SEQ ID NO: 2, la cual se puede denominar en la presente «DcUGT2».

- 20 Por ejemplo, las secuencias génicas identificadas del genoma y transcriptoma de insectos de *D. coccus* se analizaron para determinar su similitud respecto a secuencias de C-glicosiltransferasa de dominio público relevantes en la presente y el resultado fue negativo en el sentido de que ninguna de las secuencias génicas identificadas del genoma/transcriptoma mostró en la presente una similitud significativa respecto a secuencias de C-glicosiltransferasa relevantes en la presente de dominio público.

- 25 Sin embargo, aunque se podría decir que el análisis bioinformático de similitud de secuencias indica que el genoma de *Dactylopius coccus* no comprendería ningún gen que codifique una glicosiltransferasa relevante en la presente, los inventores de la presente siguieron investigando la cuestión y los inventores de la presente identificaron un extracto de *Dactylopius coccus* (incluidos extractos de organismos endosimbióticos presentes en *D. coccus*) con actividad de GT relevante en la presente y, mediante una combinación de pasos de prueba y purificación relevantes en la presente, los inventores fueron finalmente capaces de obtener una composición/fracción relativamente pura a partir de la cual fue posible obtener varias secuencias de aminoácidos parciales de posibles candidatos de enzimas GT.

- 30 Los inventores de la presente evaluaron la actividad de la glicosiltransferasa novedosa clonada/aislada descrita en la presente de SEQ ID NO: 2 (DcUGT2) y descubrieron que era capaz de conjugar glucosa con los aglicones ácido flavokermésico (FK) y ácido kermésico (AK); remítase a la Figura 1 de la presente.

Ejemplo 2 Evaluación de la actividad de GT frente a AK de UrdGT2 conocida en la técnica anterior

Según se ha discutido anteriormente, UrdGT2 se describe en el artículo de Baig *et al.* (*Angew Chem Int Ed Engl.* 27 de noviembre de 2006;45(46):7842-6).

- 35 Según se ha discutido anteriormente, este artículo describe que UrdGT2 es capaz de glucosilar diferentes moléculas de aglicones que se pueden considerar estructuralmente similares a los aglicones ácido kermésico (AK) y ácido flavokermésico (FK) relevantes en la presente.

- 40 Se clonó una versión sintética con codones optimizados de UrdGT2 para la expresión en *E. coli* y se expresó recombinantemente en *E. coli*. Se obtuvo un extracto proteico soluble crudo que contenía UrdGT2 recombinante, es decir, un extracto que comprendía UrdGT2.

La actividad de UrdGT2 GT se analizó *in vitro* utilizando o bien UDP-glucosa o TDP-glucosa como dador de azúcar y FK/AK como sustratos de tipo aglicón. No se detectó ninguna actividad frente a estos aglicones, es decir, no se identificó ninguna actividad de GT relevante en la presente respecto a estos aglicones.

- 45 Sin embargo, se confirmó que UrdGT2 recombinante era activa, lo cual se demostró mediante la formación *in vitro* de un glucósido radiomarcado con C14 derivado de la glucosilación de un compuesto no identificado en el extracto crudo de *E. coli*.

Ejemplo 3 Actividad de GT en una planta de Aloe y una planta de Haworthia

Aislamiento y prueba de actividad de GT a partir de Aloe

- 50 1) La planta se lavó para eliminar partículas de tierra y se separó en: A) Raíz, B) Tejido foliar verde y C) el material de tipo gel procedente de la hoja.

2) Se congelaron inmediatamente 5 g de tejido en nitrógeno líquido y se trituraron en un mortero frío con una mano de mortero hasta obtener un polvo fino.

3) Se añadieron al polvo 20 mL de tampón de extracción frío [Tricina-HCl 20 mM, NaCl 10 mM, DTT 5 mM, PMSF 1 mM, pH 7.9] que contenía un inhibidor de proteasas completo sin EDTA (Roche), un 0.1% (p/v) de sulfato de proteamina y 0.5 g de PVPP, y se mezclaron con un vórtex.

4) El homogenato se agitó suavemente a 4 °C durante 10 min y a continuación se centrifugó a 12 000 x g a 4 °C durante 5 min.

5) Se aisló el sobrenadante y se añadió gota a gota 1 mL de sulfato de proteamina al 2% (p/v) en Tricina-HCl 20 mM, pH 7.9 durante 2 min a 4 °C con agitación constante.

6) El sobrenadante se filtró a través de 2 piezas de malla de nailon. A continuación, el sobrenadante filtrado se centrifugó a 12 000 x g a 4 °C durante 5 min.

7) El sobrenadante se aisló y se ultracentrifugó a 110 000 x g a 4 °C durante 1 h.

8) Se aisló la fracción proteínica soluble (sobrenadante) y se intercambió el tampón 5 veces con Tricina-HCl 20 mM, pH 7.9 que contenía DTT 5 mM utilizando un dispositivo de filtro de ultracentrifugación Amicon 3K (Millipore).

9) Se incubaron 20 µL de extracto proteico soluble en un volumen de reacción total de 60 µL que contenía UDP-glucosa (conc. final 1.25 mM) y o bien FK (conc. final 50 µM), AK (conc. final 50 µM) o MeO-FK/EtO-FK (conc. final 50 µM/50 µM) durante 2 h a 30 °C, agitando a 650 rpm.

10) Las reacciones enzimáticas se terminaron con 180 µL de metanol frío, se filtraron a través de un filtro de 0.45 micras y se sometieron a un análisis de HPLC-MS.

Tabla 1. Glucósidos formados en los ensayos de glucosilación *in vitro* utilizando extractos enzimáticos de *Aloe*.

Valores de m/z [M-H] ⁻	475 m/z [M-H] ⁻ FK-monogluc	491 m/z [M-H] ⁻ AK-monogluc	489 m/z [M-H] ⁻ MeOFK-monogluc	503 m/z [M-H] ⁻ EtOFK-monogluc
Proteína soluble de <i>Aloe</i>				
Hoja	3.73	3.71	5.81	6.63
Gel				
Raíz		3.71		

Los extractos enzimáticos solubles crudos de tres tejidos de *Aloe*, material foliar verde (Hoja), material de tipo gel procedente de la hoja (Gel) y Raíz, se evaluaron para determinar su actividad de glucosilación frente al ácido flavokermésico (FK), el ácido kermésico (AK), el éster metílico del ácido flavokermésico (MeOFK) y el éster etílico del ácido flavokermésico (EtOFK). Los números corresponden a los tiempos de retención (min) tras la separación mediante HPLC-MS de los glucósidos novedosos formados *in vitro* (Tabla 1).

Los valores m/z de 475 y 491 son los mismos valores m/z que se obtienen para Dcll y AC, respectivamente, solubilizados en soluciones similares. Ambos valores m/z son 162 (valor m/z de la glucosa en un glucósido) unidades más elevados que los valores m/z del FK y AK, lo cual indica que el resto de glucosa procedente de UDP-glucosa en el tampón de reacción ha sido transferido al aglicón mediante una GT en el extracto. Los valores m/z [M-H]⁻ de 489 y 503 también son 162 unidades más elevados que los valores m/z obtenidos con MeOFK y EtOFK, respectivamente, lo cual indica que ha sido añadida una unidad de glucosa tanto a MeOFK como a EtOFK mediante una GT presente en el extracto.

Aislamiento y prueba de actividad de GT a partir de *Haworthia limifolia*

El procedimiento fue como el descrito para *Aloe*, pero los tejidos vegetales analizados fueron los siguientes: A) Tejido foliar verde, B) Material de tipo gel procedente de la hoja, C) Tejido basal (parte rosada entre la raíz y el tallo) y D) Tejido de la raíz.

Los extractos enzimáticos solubles crudos de cuatro tejidos de *Haworthia limifolia*, material foliar verde (Hoja), material de tipo gel procedente de la hoja (Gel), tejido rosado entre la raíz y el tallo (Base) y Raíz, se evaluaron para determinar su actividad de glucosilación frente al ácido flavokermésico (FK), el ácido kermésico (AK), el éster metílico del ácido

flavokermésico (MeOFK) y el éster etílico del ácido flavokermésico (EtOFK). Los números corresponden a los tiempos de retención (min) tras la separación mediante HPLC-MS de los glucósidos novedosos formados *in vitro* (Tabla 2).

Valores de m/z [M-H] ⁻ ----- Proteína soluble de Haworthia	475 m/z [M-H] ⁻ FK-monoglc	491 m/z [M-H] ⁻ KA-monoglc	489 m/z [M-H] ⁻ MeOFK-monoglc	503 m/z [M-H] ⁻ EtOFK-monoglc
Hoja	3.73	3.71	5.81	6.63
Gel				
Base	3.73	3.71	5.81	6.63
Raíz	3.73	3.71	5.81	6.63

Tabla 2. Glucósidos formados en ensayos de glucosilación *in vitro* utilizando extractos enzimáticos de *Haworthia limifolia*.

Los valores m/z de 475 y 491 son los mismos valores m/z que se obtienen para DcII y AC, respectivamente, solubilizados en soluciones similares. Ambos valores m/z son 162 (valor m/z de la glucosa en un glucósido) unidades más elevados que los valores m/z del FK y AK, lo cual indica que el resto de glucosa procedente de UDP-glucosa en el tampón de reacción ha sido transferido al aglicón mediante una GT en el extracto. Los valores m/z [M-H] de 489 y 503 también son 162 unidades más elevados que los valores m/z obtenidos con MeOFK y EtOFK, respectivamente, lo cual indica que ha sido añadida una unidad de glucosa tanto a MeOFK como a EtOFK mediante una GT presente en el extracto.

Conclusión

Los resultados de este ejemplo demuestran que las enzimas de tipo glicosiltransferasa (GT) relevantes de la presente pueden ser identificadas en plantas de *Aloe* y plantas de *Haworthia*.

Dicho de otro modo, las plantas de *Aloe* y las plantas de *Haworthia* comprenden una glicosiltransferasa que es capaz de glicosilar el ácido flavokermésico con el fin de producir un glicósido del ácido flavokermésico; y/o capaz de glicosilar el ácido kermésico con el fin de producir un glicósido del ácido kermésico.

Ejemplo 4 Actividad de GT en una planta de sorgo y de arroz

Según existe constancia en la técnica, las plantas de sorgo y de arroz comprenden glicosiltransferasas.

Según existe constancia en la técnica, algunas de las glicosiltransferasas de sorgo y arroz pueden glicosilar compuestos de tipo aglicón de bajo peso molecular.

Las glicosiltransferasas descritas en la técnica procedentes de plantas de sorgo y arroz presentan una identidad significativamente inferior a un 70% con los aminoácidos 1-515 de la SEQ ID NO:2 que se describe en la presente.

En la técnica existe constancia de si las glicosiltransferasas de plantas de sorgo y/o arroz serían una glicosiltransferasa relevante en la presente, es decir, una glicosiltransferasa capaz de glicosilar el ácido flavokermésico con el fin de producir glicósidos del ácido flavokermésico; y/o capaz de glicosilar el ácido kermésico con el fin de producir glicósidos del ácido kermésico.

Las glicosiltransferasas conocidas de sorgo (*Sorghum bicolor*), SbUGT85B1, con número de ID de Genbank AF199453.1 (sec. de nucleótidos) / AAF17077.1 (sec. de polipéptido), y de arroz (*Oryza sativa*), OsCGT, con número de ID de Genbank FM179712.1 (sec. de nucleótidos) / CAQ77160.1 (sec. de polipéptido), se expresaron en la cepa Xjb de *E.coli* y se prepararon extractos proteicos crudos de *E.coli* y se evaluaron para determinar su actividad de glucosilación sobre los sustratos ácido kermésico y ácido flavokermésico según describen Kannangara *et al.* (2011) y Augustin *et al.* (2012).

La Figura 2 muestra análisis de LC-MS de los productos glucosilados formados en ensayos que contenían un lisado crudo de una cepa Xjb de *E.coli* que expresaba o bien SbUGT85B1 u OsCGT, UDP-glucosa y ácido flavokermésico (FK) o ácido kermésico (AK). En los ensayos se utilizó como control negativo un extracto crudo procedente de la cepa Xjb de *E.coli*.

Se identificaron glicósidos de AK (491 m/z [M-H], el valor de m/z [M-H] del AC) para ambas glicosiltransferasas y glicósidos de FK (475 m/z [M-H], el valor m/z[M-H] de DcII) para OsCGT.

Conclusión

El resultado de este ejemplo demuestra que las enzimas de tipo glicosiltransferasa (GT) relevantes de la presente pueden ser identificadas en plantas de sorgo y/o arroz.

- 5 Dicho de otro modo, las plantas de sorgo y/o arroz comprenden una glicosiltransferasa que es capaz de glicosilar el ácido flavokermésico con el fin de producir un glicósido del ácido flavokermésico; y/o capaz de glicosilar el ácido kermésico con el fin de producir un glicósido del ácido kermésico.

Ejemplo 5 Uso de actividad de GT o gen de GT endógenos

- 10 Según existe constancia en la técnica, las glicosiltransferasas capaces de glicosilar un peso molecular bajo están presentes en muchos organismos diferentes. Un método para poner en contacto la glicosiltransferasa de las células de un organismo con un compuesto de peso molecular bajo consiste en introducir uno o más genes que dirijan la biosíntesis del compuesto de peso molecular bajo y de este modo se posibilita que las células glicosilen el compuesto de peso molecular bajo. El compuesto de peso molecular bajo puede ser, p. ej., ácido flavokermésico o ácido kermésico o versiones modificadas de estas moléculas.

- 15 Se introducen uno o más genes que dirigen la biosíntesis del ácido flavokermésico o ácido kermésico o una versión modificada de estas moléculas en un organismo que contiene una glicosiltransferasa, p. ej., la planta de tabaco, *Nicotiana benthamiana*.

- 20 Cuando el gen o genes se expresan transitoriamente de acuerdo con los métodos descritos en D'Aoust *et al.* (2008) en, p. ej., tejido vegetal, se produce el compuesto o compuestos de peso molecular bajo. Se producen células que expresan de forma estable el gen o genes y se seleccionan de acuerdo con los métodos descritos en Gelvin (2003).

- 25 En las células que contienen el gen o genes ya sean expresados de forma transitoria y/o expresados de forma estable, los compuestos de peso molecular bajo entran en contacto con las glicosiltransferasas endógenas, lo cual da como resultado la formación de uno o más glicósidos del ácido flavokermésico, ácido kermésico o versiones modificadas de estas moléculas.

- 30 La presencia de los glicósidos se demuestra mediante la extracción y los métodos analíticos descritos en el ejemplo 3.

Las muestras se preparan para LC/MS mediante el método de extracción descrito por Rauwald y Sigler (1994).

Conclusión

- 35 Los resultados de este ejemplo demuestran que las glicosiltransferasas endógenas presentes en las células de un organismo recombinante se pueden utilizar para convertir ácido flavokermésico, ácido kermésico o versiones modificadas de estas moléculas en glicósidos cuando un gen o genes que dirigen la biosíntesis de los aglicones se introducen en el organismo.

- 40 Dicho de otro modo, la introducción de un gen o genes que dirigen la biosíntesis de ácido flavokermésico, ácido kermésico, versiones modificadas de estas moléculas o compuestos relacionados de peso molecular bajo es un método para poner el compuesto de peso molecular bajo en contacto con glicosiltransferasas y, por lo tanto, un método para producir glicósidos del ácido flavokermésico, ácido kermésico o una versión modificada de estos compuestos.

REFERENCIAS

- 1: US5424421 (European Colour, publicada en 1995)
- 2: WO2006/056585A1 (Chr. Hansen A/S)
- 3: Stathopoulou *et al.* (*Analytica Chimica Acta* 804 (2013) 264- 272)
- 40 4: Zagrobelny *et al.* (*Cyanogenic glucosides and plant-insect interactions; Phytochemistry*. Febrero de 2004;65(3):293-306)
- 5: Geuder *et al.* (*Journal of Chemical Ecology*, Vol. 23, N.º 5, 1997)
- 6: Genta *et al.*, (*Potential role for gut microbiota in cell wall digestion and glucoside detoxification in Tenebrio molitor larvae*), *Journal of Insect Physiology* 52 (2006) 593-601
- 45 7: WO2004/111254A1 (Poalis A/S)
- 8: Gutmann *et al.* (*Pure Appl. Chem*, 2013-07-09)
- 9: Pompon *et al.* (*Methods Enzymol.* 272 (1996):51-64)

- 10: Baig *et al.* (*Angew Chem Int Ed Engl.* 27 de noviembre de 2006;45(46):7842-6)
- 11: Radomska-Pandya A, Bratton SM, Redinbo MR, Miley MJ. *Drug Metab Rev.* Febrero de 2010;42(1):133-44)
- 12: *Plant Physiology*, Noviembre de 2008, Vol. 148, págs. 1295-1308
- 13: Esben Halkjaer Hansen *et al.* *Phytochemistry* 70(4): 473-482
- 14: Kannangara *et al.* (*Plant Journal.* 68 (2011): 287-301)
- 15: Augustin *et al.* (*Plant Physiology.* 160 (2012): 1881-1895)
- 16: D'Aoust *et al.* (*Methods Mol Biol* 483 (2009): 41-50)
- 17: Gelvin (*Microbiol Mol Biol Rev* 67(1) (2003): 16-37)
- 18: Rauwald y Sigler (*Phytochemical Analysis* 5 (1994):266-270)

LISTA DE SECUENCIAS

<110> Chr. Hansen A/S

<120> Glicosiltransferasa que glicosila ácido flavokermésico y/o ácido kermésico

<130> P4547PC00

5 <150> EP13198110.2

< 151> 18-12-2013

<160> 5

<170> PatentIn versión 3.5

<210> 1

10 < 211> 1548

< 212> ADN

< 213> Dactylopius coccus costa

<220>

< 221> CDS

15 < 222> (1)..(1548)

<400> 1

atg gaa ttt cgt tta cta atc ctg gct ctt ttt tct gta ctt atg agt	48
Met Glu Phe Arg Leu Leu Ile Leu Ala Leu Phe Ser Val Leu Met Ser	
1 5 10 15	
act tca aac gga gca gaa att tta gct ctt ttc cct att cac ggt atc	96
Thr Ser Asn Gly Ala Glu Ile Leu Ala Leu Phe Pro Ile His Gly Ile	
20 25 30	
agt aat tat aat gtt gct gaa gca ctg ctg aag acc tta gct aac cgg	144
Ser Asn Tyr Asn Val Ala Glu Ala Leu Leu Lys Thr Leu Ala Asn Arg	
35 40 45	
ggg cat aat gtt aca gtt gtc aca tct ttt cct caa aaa aaa cct gta	192
Gly His Asn Val Thr Val Thr Ser Phe Pro Gln Lys Lys Pro Val	
50 55 60	
cct aat ttg tac gaa att gac gta tct gga gct aaa ggc ttg gct act	240
Pro Asn Leu Tyr Glu Ile Asp Val Ser Gly Ala Lys Gly Leu Ala Thr	
65 70 75 80	
aat tca ata cat ttt gaa aga tta caa acg att att caa gat gta aaa	288
Asn Ser Ile His Phe Glu Arg Leu Gln Thr Ile Ile Gln Asp Val Lys	
85 90 95	
tcg aac ttt aag aac atg gta cga ctt agc aga aca tac tgt gag att	336
Ser Asn Phe Lys Asn Met Val Arg Leu Ser Arg Thr Tyr Cys Glu Ile	
100 105 110	
atg ttt tct gat ccg agg gtt ttg aac att cga gac aag aaa ttc gat	384
Met Phe Ser Asp Pro Arg Val Leu Asn Ile Arg Asp Lys Lys Phe Asp	
115 120 125	
ctc gta ata aac gcc gta ttt ggc agt gac tgc gat gcc gga ttc gca	432
Leu Val Ile Asn Ala Val Phe Gly Ser Asp Cys Asp Ala Gly Phe Ala	
130 135 140	

ES 2 660 301 T3

tgg aaa agt caa gct cca ttg att tca att ctc aat gct aga cat act	480
Trp Lys Ser Gln Ala Pro Leu Ile Ser Ile Leu Asn Ala Arg His Thr	
145 150 155 160	
cct tgg gcc cta cac aga atg gga aat cca tca aat cca gcg tat atg	528
Pro Trp Ala Leu His Arg Met Gly Asn Pro Ser Asn Pro Ala Tyr Met	
165 170 175	
cct gtc att cat tct aga ttt cct gta aaa atg aat ttc ttc caa aga	576
Pro Val Ile His Ser Arg Phe Pro Val Lys Met Asn Phe Phe Gln Arg	
180 185 190	
atg ata aat acg ggt tgg cat ttg tat ttt ctg tac atg tac ttt tat	624
Met Ile Asn Thr Gly Trp His Leu Tyr Phe Leu Tyr Met Tyr Phe Tyr	
195 200 205	
tat ggt aat gga gaa gat gcc aac aaa atg gcg aga aaa ttt ttt ggc	672
Tyr Gly Asn Gly Glu Asp Ala Asn Lys Met Ala Arg Lys Phe Phe Gly	
210 215 220	
aac gac atg ccc gac ata aat gaa atg gtt ttt aat aca tct tta tta	720
Asn Asp Met Pro Asp Ile Asn Glu Met Val Phe Asn Thr Ser Leu Leu	
225 230 235 240	
ttc gta aat act cac ttt tcg gtt gat atg cca tat cct ttg gtt cca	768
Phe Val Asn Thr His Phe Ser Val Asp Met Pro Tyr Pro Leu Val Pro	
245 250 255	
aac tgc att gaa ata gga gga ata cat gta aaa gag cca caa cca ctg	816
Asn Cys Ile Glu Ile Gly Gly Ile His Val Lys Glu Pro Gln Pro Leu	
260 265 270	
cct ttg gaa ata caa aaa ttc atg gac gaa gca gaa cat ggg gtc att	864
Pro Leu Glu Ile Gln Lys Phe Met Asp Glu Ala Glu His Gly Val Ile	
275 280 285	
ttc ttc acg cta gga tca atg gtg cgt act tcc acg ttt cca aat caa	912
Phe Phe Thr Leu Gly Ser Met Val Arg Thr Ser Thr Phe Pro Asn Gln	
290 295 300	
act att caa gca ttt aag gaa gct ttt gcc gaa tta cct caa aga gtc	960
Thr Ile Gln Ala Phe Lys Glu Ala Phe Ala Glu Leu Pro Gln Arg Val	
305 310 315 320	
tta tgg aag ttt gag aat gaa aat gag gat atg cca tca aat gta ctc	1008
Leu Trp Lys Phe Glu Asn Glu Asn Glu Asp Met Pro Ser Asn Val Leu	
325 330 335	
ata agg aaa tgg ttt cca caa aat gat ata ttc ggt cat aag aat atc	1056
Ile Arg Lys Trp Phe Pro Gln Asn Asp Ile Phe Gly His Lys Asn Ile	
340 345 350	
aaa gca ttc att agt cac ggt gga aat tct gga gct ctg gag gct gtt	1104
Lys Ala Phe Ile Ser His Gly Gly Asn Ser Gly Ala Leu Glu Ala Val	
355 360 365	
cat ttc gga gta ccg ata att gga att cct tta ttc tac gat cag tac	1152
His Phe Gly Val Pro Ile Ile Gly Ile Pro Leu Phe Tyr Asp Gln Tyr	
370 375 380	
agg aat att ttg agt ttc gtt aaa gaa ggt gtt gcc gtt ctt ttg gat	1200
Arg Asn Ile Leu Ser Phe Val Lys Glu Gly Val Ala Val Leu Leu Asp	
385 390 395 400	

ES 2 660 301 T3

gtg aat gat ctg acg aaa gat aat att tta tct tct gtc agg act gtt	1248
Val Asn Asp Leu Thr Lys Asp Asn Ile Leu Ser Ser Val Arg Thr Val	
405 410 415	
ggt aat gat aag agt tac tca gaa cgt atg aaa gca ttg tca caa cta	1296
Val Asn Asp Lys Ser Tyr Ser Glu Arg Met Lys Ala Leu Ser Gln Leu	
420 425 430	
ttc cga gat cga cca atg agt cct ctt gac aca gct gtt tac tgg aca	1344
Phe Arg Asp Arg Pro Met Ser Pro Leu Asp Thr Ala Val Tyr Trp Thr	
435 440 445	
gaa tat gtc atc cgc cat aga gga gcc cat cac ctc aag acc gct ggc	1392
Glu Tyr Val Ile Arg His Arg Gly Ala His His Leu Lys Thr Ala Gly	
450 455 460	
gca ttt ttg cat tgg tat cag tat tta ctt ttg gac gtt att acc ttc	1440
Ala Phe Leu His Trp Tyr Gln Tyr Leu Leu Leu Asp Val Ile Thr Phe	
465 470 475 480	
tta tta gtc aca ttc tgc gct ttt tgt ttt att gtg aaa tat ata tgt	1488
Leu Leu Val Thr Phe Cys Ala Phe Cys Phe Ile Val Lys Tyr Ile Cys	
485 490 495	
aaa gct ctc att cat cat tat tgg agc agt tcg aaa tct gaa aag ttg	1536
Lys Ala Leu Ile His His Tyr Trp Ser Ser Ser Lys Ser Glu Lys Leu	
500 505 510	
aaa aaa aat taa	1548
Lys Lys Asn	
515	

<210> 2

< 211> 515

< 212> PRT

5 < 213> Dactylopius coccus costa

<400> 2

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

ES 2 660 301 T3

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

ES 2 660 301 T3

340

345

350

Lys Ala Phe Ile Ser His Gly Gly Asn Ser Gly Ala Leu Glu Ala Val
355 360 365

His Phe Gly Val Pro Ile Ile Gly Ile Pro Leu Phe Tyr Asp Gln Tyr
370 375 380

Arg Asn Ile Leu Ser Phe Val Lys Glu Gly Val Ala Val Leu Leu Asp
385 390 395 400

Val Asn Asp Leu Thr Lys Asp Asn Ile Leu Ser Ser Val Arg Thr Val
405 410 415

Val Asn Asp Lys Ser Tyr Ser Glu Arg Met Lys Ala Leu Ser Gln Leu
420 425 430

Phe Arg Asp Arg Pro Met Ser Pro Leu Asp Thr Ala Val Tyr Trp Thr
435 440 445

Glu Tyr Val Ile Arg His Arg Gly Ala His His Leu Lys Thr Ala Gly
450 455 460

Ala Phe Leu His Trp Tyr Gln Tyr Leu Leu Leu Asp Val Ile Thr Phe
465 470 475 480

Leu Leu Val Thr Phe Cys Ala Phe Cys Phe Ile Val Lys Tyr Ile Cys
485 490 495

Lys Ala Leu Ile His His Tyr Trp Ser Ser Ser Lys Ser Glu Lys Leu
500 505 510

Lys Lys Asn
515

<210> 3

< 211> 1548

< 212> ADN

5 < 213> Secuencia artificial

<220>

< 223> Secuencia optimizada en levadura

<400> 3

atggaattca gattgttgat attggccttg ttctccgtat tgatgtctac ctctaattggt	60
gccgaaatct tggctttatt ccctattcat ggtatatcta actacaacgt agctgaagca	120
ttgttgaaga ctttggctaa cagaggtcac aacgttaccg ttgtaacttc atttccacaa	180
aagaaaccag ttcctaattt gtacgaaatt gatgtatcag gtgcaaaggg tttagccaca	240

ES 2 660 301 T3

```

aactccatcc atttcgaaag attgcaaacc atcatccaag atgtcaagag taacttcaag      300
aacatgggta gattgtctag aacatactgt gaaatcatgt tctcagaccc aagagttttg      360
aacatcagag ataaaaagtt tgacttggtt ataaacgccg tattcggttc agattgcgac      420
gctggttttg catggaaaag tcaagctcct ttaatatcta tcttgaatgc cagacataca      480
ccatgggctt tgcacagaat gggtaatcct tccaaccag catatatgcc tgtaatccat      540
agtagattcc cagtcaagat gaatttcttt caaagaatga taaacaccgg ttggcactta      600
tactttttgt acatgtactt ctactacggt aatggtgaag atgctaacaa aatggcaaga      660
aagtttttcg gtaatgatat gcctgacata aacgaaatgg tttttaacac ctccttggtg      720
ttcgtaaaca ctcatcttcag tgtcgatatg ccataccctt tagtcccaa ctgtatcgaa      780
atcgggtggt tccatgttaa ggaaccacaa cctttgccat tggaaatcca aaagtttatg      840
gatgaagcag aacatgggtg aatctttttc accttgggta gtatggtcag aacttctaca      900
ttccctaata aaactattca agccttttaa gaagccttcg ctgaattacc acaaagagtt      960
ttgtggaagt tcgaaaacga aaacgaagat atgccttcca acgttttgat cagaaagtgg     1020
ttcccacaaa acgacatctt cggtcataag aacatcaagg ctttcatttc acacggtggt     1080
aattccgggtg ccttggaagc tgtccatttc ggtgttcta tcataggtat occattgttt     1140
tatgatcaat acagaaacat cttgtctttc gttaaagaag gtgtagctgt cttgttggat     1200
gtaaacgact taactaagga taacatcttg tcttcagtta gaacagtcgt taacgacaag     1260
tcatactccg aaagaatgaa ggcattgtct caattgttta gagatagacc tatgtcacca     1320
ttagacacag ctgtttattg gaccgaatac gtaattagac atagaggtgc acatcactta     1380
aaaactgcag gtgccttttt gcaactggtat caatacttgt tgttggatgt catcacattt     1440
ttgttggtta cattctgtgc attctgcttc atcgttaagt acatctgcaa ggccttaata     1500
catcactact ggtccagttc taaatctgaa aagttgaaaa agaattaa     1548

```

<210> 4

< 211> 492

< 212> PRT

5 < 213> Sorghum bicolor

<400> 4

```

Met Gly Ser Asn Ala Pro Pro Pro Pro Thr Pro His Val Val Leu Val
1          5          10          15

Pro Phe Pro Gly Gln Gly His Val Ala Pro Leu Met Gln Leu Ala Arg
20          25          30

Leu Leu His Ala Arg Gly Ala Arg Val Thr Phe Val Tyr Thr Gln Tyr
35          40          45

```

ES 2 660 301 T3

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

ES 2 660 301 T3

290

295

300

Val Tyr Val Asn Phe Gly Ser Met Ala Val Met Thr Ala Ala Gln Ala
305 310 315 320

Arg Glu Phe Ala Leu Gly Leu Ala Ser Cys Gly Ser Pro Phe Leu Trp
325 330 335

Val Lys Arg Pro Asp Val Val Glu Gly Glu Glu Val Leu Leu Pro Glu
340 345 350

Ala Leu Leu Asp Glu Val Ala Arg Gly Arg Gly Leu Val Val Pro Trp
355 360 365

Cys Pro Gln Ala Ala Val Leu Lys His Ala Ala Val Gly Leu Phe Val
370 375 380

Ser His Cys Gly Trp Asn Ser Leu Leu Glu Ala Thr Ala Ala Gly Gln
385 390 395 400

Pro Val Leu Ala Trp Pro Cys His Gly Glu Gln Thr Thr Asn Cys Arg
405 410 415

Gln Leu Cys Glu Val Trp Gly Asn Gly Ala Gln Leu Pro Arg Glu Val
420 425 430

Glu Ser Gly Ala Val Ala Arg Leu Val Arg Glu Met Met Val Gly Asp
435 440 445

Leu Gly Lys Glu Lys Arg Ala Lys Ala Ala Glu Trp Lys Ala Ala Ala
450 455 460

Glu Ala Ala Ala Arg Lys Gly Gly Ala Ser Trp Arg Asn Val Glu Arg
465 470 475 480

Val Val Asn Asp Leu Leu Leu Val Gly Gly Lys Gln
485 490

<210> 5

< 211> 471

< 212> PRT

5 < 213> Oryza sativa

<400> 5

Met Pro Ser Ser Gly Asp Ala Ala Gly Arg Arg Pro His Val Val Leu
1 5 10 15

Ile Pro Ser Ala Gly Met Gly His Leu Val Pro Phe Gly Arg Leu Ala

ES 2 660 301 T3

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

ES 2 660 301 T3

Val	Val	Tyr	Val	Ser	Phe	Gly	Ser	Arg	Lys	Ala	Ile	Ser	Arg	Glu	Gln	275	280	285
Leu	Arg	Glu	Leu	Ala	Ala	Gly	Leu	Glu	Gly	Ser	Gly	His	Arg	Phe	Leu	290	295	300
Trp	Val	Val	Lys	Ser	Thr	Val	Val	Asp	Arg	Asp	Asp	Ala	Ala	Glu	Leu	305	310	315
Gly	Glu	Leu	Leu	Asp	Glu	Gly	Phe	Leu	Glu	Arg	Val	Glu	Lys	Arg	Gly	325	330	335
Leu	Val	Thr	Lys	Ala	Trp	Val	Asp	Gln	Glu	Glu	Val	Leu	Lys	His	Glu	340	345	350
Ser	Val	Ala	Leu	Phe	Val	Ser	His	Cys	Gly	Trp	Asn	Ser	Val	Thr	Glu	355	360	365
Ala	Ala	Ala	Ser	Gly	Val	Pro	Val	Leu	Ala	Leu	Pro	Arg	Phe	Gly	Asp	370	375	380
Gln	Arg	Val	Asn	Ser	Gly	Val	Val	Ala	Arg	Ala	Gly	Leu	Gly	Val	Trp	385	390	395
Ala	Asp	Thr	Trp	Ser	Trp	Glu	Gly	Glu	Ala	Gly	Val	Ile	Gly	Ala	Glu	405	410	415
Glu	Ile	Ser	Glu	Lys	Val	Lys	Ala	Ala	Met	Ala	Asp	Glu	Ala	Leu	Arg	420	425	430
Met	Lys	Ala	Ala	Ser	Leu	Ala	Glu	Ala	Ala	Ala	Lys	Ala	Val	Ala	Gly	435	440	445
Gly	Gly	Ser	Ser	His	Arg	Cys	Leu	Ala	Glu	Phe	Ala	Arg	Leu	Cys	Gln	450	455	460
Gly	Gly	Thr	Cys	Arg	Thr	Asn										465	470	

REIVINDICACIONES

1. Un polipéptido de tipo glicosiltransferasa aislado capaz de:

(I): conjugar glucosa activada con un nucleótido a ácido flavokermésico (FK); y/o

(II): conjugar glucosa activada con un nucleótido a ácido kermésico (AK).

5 y donde el polipéptido de tipo glicosiltransferasa es al menos un polipéptido seleccionado del grupo constituido por:

(a) un polipéptido que comprende una secuencia de aminoácidos que presenta una identidad de al menos un 70% con los aminoácidos 1-515 de la SEQ ID NO:2;

(b) un polipéptido que comprende una secuencia de aminoácidos que presenta una identidad de al menos un 70% con los aminoácidos 20-468 de la SEQ ID NO: 2;

10 (c) un polipéptido que está codificado por un polinucleótido que se hibrida en condiciones de rigurosidad al menos media con (i) los nucleótidos 1-1548 de la SEQ ID NO:1 o (ii) una hebra complementaria de (i); y

(d) un fragmento de los aminoácidos 1-515 de la SEQ ID NO:2, que presenta la actividad de glicosiltransferasa que se ha especificado en (I) y/o (II).

15 2. El polipéptido de tipo glicosiltransferasa aislado de la reivindicación 1, donde el polipéptido de tipo glicosiltransferasa aislado de la reivindicación 1 es capaz de:

(I): conjugar glucosa con ácido flavokermésico (FK); y

(II): conjugar glucosa con ácido kermésico (AK).

3. El polipéptido de tipo glicosiltransferasa aislado de cualquiera de las reivindicaciones precedentes, donde el polipéptido de tipo glicosiltransferasa de la reivindicación 1 es:

20 (a) un polipéptido que comprende una secuencia de aminoácidos que presenta una identidad de al menos un 80% con los aminoácidos 1-515 de la SEQ ID NO:2.

4. El polipéptido de tipo glicosiltransferasa aislado de cualquiera de las reivindicaciones precedentes, donde el polipéptido de tipo glicosiltransferasa de la reivindicación 1 es:

25 (a) un polipéptido que comprende una secuencia de aminoácidos que presenta una identidad de al menos un 95% con los aminoácidos 1-515 de la SEQ ID NO:2.

5. Un polinucleótido aislado que comprende una secuencia de nucleótidos que codifica el polipéptido de cualquiera de las reivindicaciones precedentes.

6. Un constructo de ácido nucleico que comprende el polinucleótido aislado de la reivindicación 5 unido operablemente a una o más secuencias de control que dirigen la producción del polipéptido en un hospedador de expresión.

30 7. Un vector de expresión recombinante que comprende el constructo de ácido nucleico de la reivindicación 6.

8. Una célula hospedadora recombinante que comprende el constructo de ácido nucleico de la reivindicación 7.

35 9. La célula hospedadora recombinante de la reivindicación 8, donde la célula hospedadora recombinante es una célula seleccionada del grupo constituido por una célula de hongo filamentoso y una célula de un microorganismo, preferentemente donde la célula del microorganismo es una célula de levadura, preferentemente donde la célula de levadura se selecciona del grupo constituido por *Saccharomyces spp* (p. ej., *Saccharomyces cerevisiae*) y *Pichia spp*.

10. Un método para producir un glicósido del ácido flavokermésico (FK) y/o un glicósido del ácido kermésico (AK), donde el método comprende los siguientes pasos:

(A): poner en contacto *in vivo* o *in vitro* en una célula hospedadora recombinante que comprende un gen de glicosiltransferasa que codifica una glicosiltransferasa:

40 (a1): ácido flavokermésico (FK) con una glicosiltransferasa capaz de glicosilar el ácido flavokermésico en condiciones adecuadas, donde se produce el glicósido del ácido flavokermésico; y/o

(a2): ácido kermésico (AK) con una glicosiltransferasa capaz de glicosilar el ácido kermésico en condiciones adecuadas, donde se produce el glicósido del ácido kermésico, y

donde la glicosiltransferasa del paso (A) es una glicosiltransferasa de cualquiera de las reivindicaciones 1 a 4.

45 11. El método de la reivindicación 10, donde

- la glicosiltransferasa del paso (a1) es una glucosiltransferasa y, por lo tanto, en el paso (a1) se produce un glucósido del ácido flavokermésico y donde el glucósido del ácido flavokermésico producido es el compuesto DcII; y/o

- la glicosiltransferasa del paso (a2) es una glucosiltransferasa y, por lo tanto, en el paso (a2) se produce un glucósido del ácido kermésico y el glucósido del ácido kermésico producido es ácido carmínico.

- 5 **12.** El método de cualquiera de las reivindicaciones 10 a 11, donde la célula hospedadora recombinante del paso (A) es una célula hospedadora recombinante que comprende un gen de glicosiltransferasa recombinante que codifica una glicosiltransferasa y donde el método comprende un paso adicional (B) con los siguientes pasos:

10 (B): purificar el glicósido producido en el paso (a1) y/o en el paso (a2) para obtener de este modo una composición, donde al menos un 50% p/p de los compuestos en la composición es el glicósido del ácido flavokermésico y/o el glicósido del ácido kermésico producido.

13. El método de cualquiera de las reivindicaciones 10 a 12, donde la puesta en contacto en el paso (A) de la reivindicación 1 se realiza *in vivo* y la célula hospedadora recombinante es una célula de levadura, preferentemente donde la célula de levadura se selecciona del grupo constituido por *Saccharomyces spp* (p. ej., *Saccharomyces cerevisiae*) y *Pichia spp*.

- 15 **14.** El método de cualquiera de las reivindicaciones 10 a 13, donde la glicosiltransferasa está unida a la membrana.

Figura 1

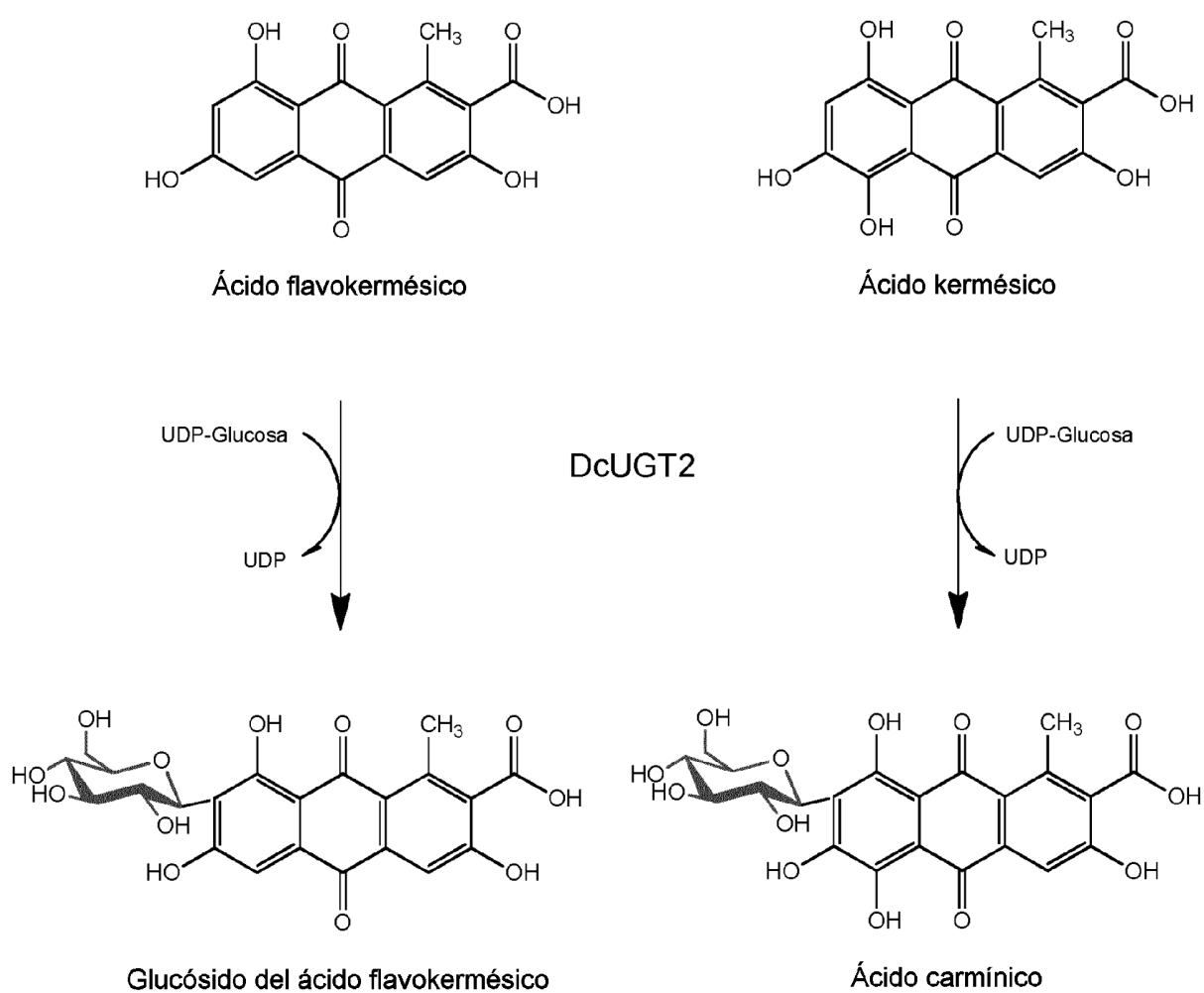


Figura 2 (A)

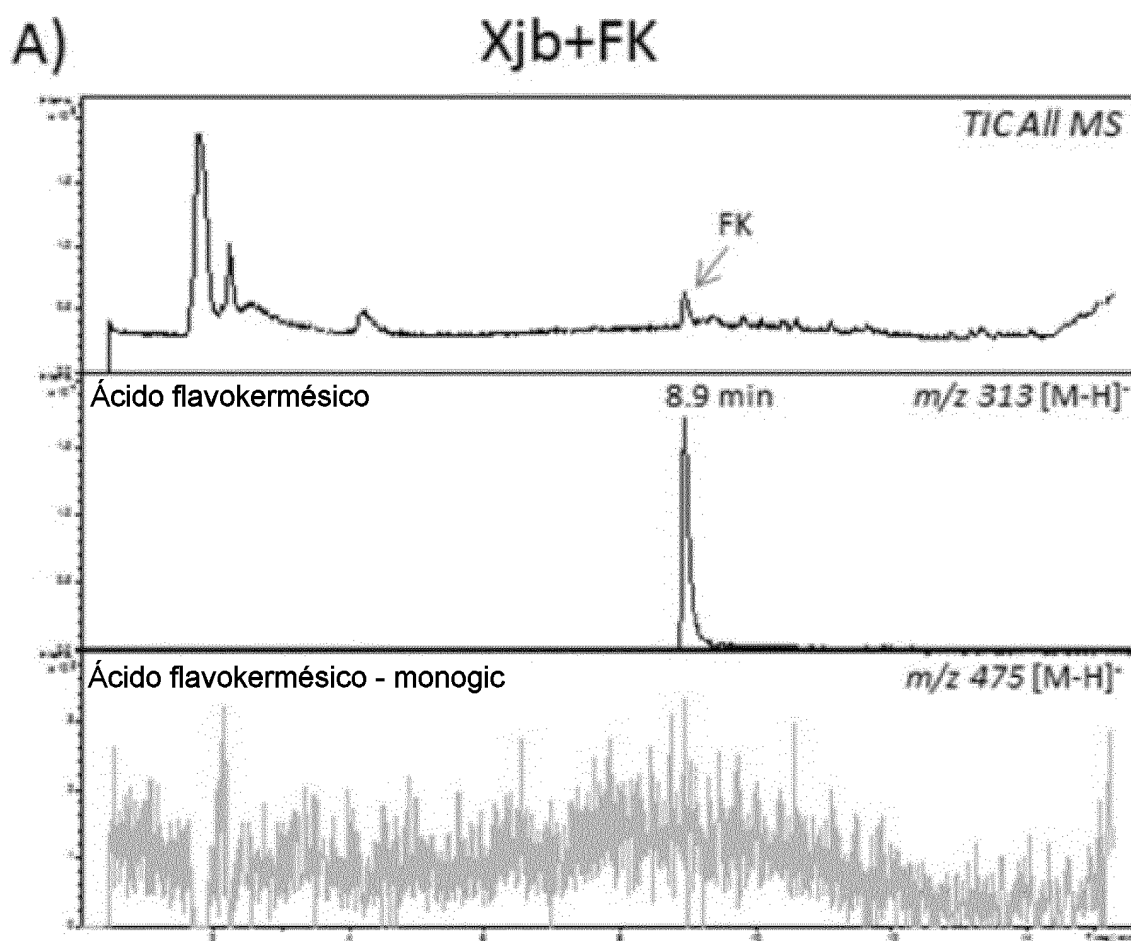


Figura 2 (B)

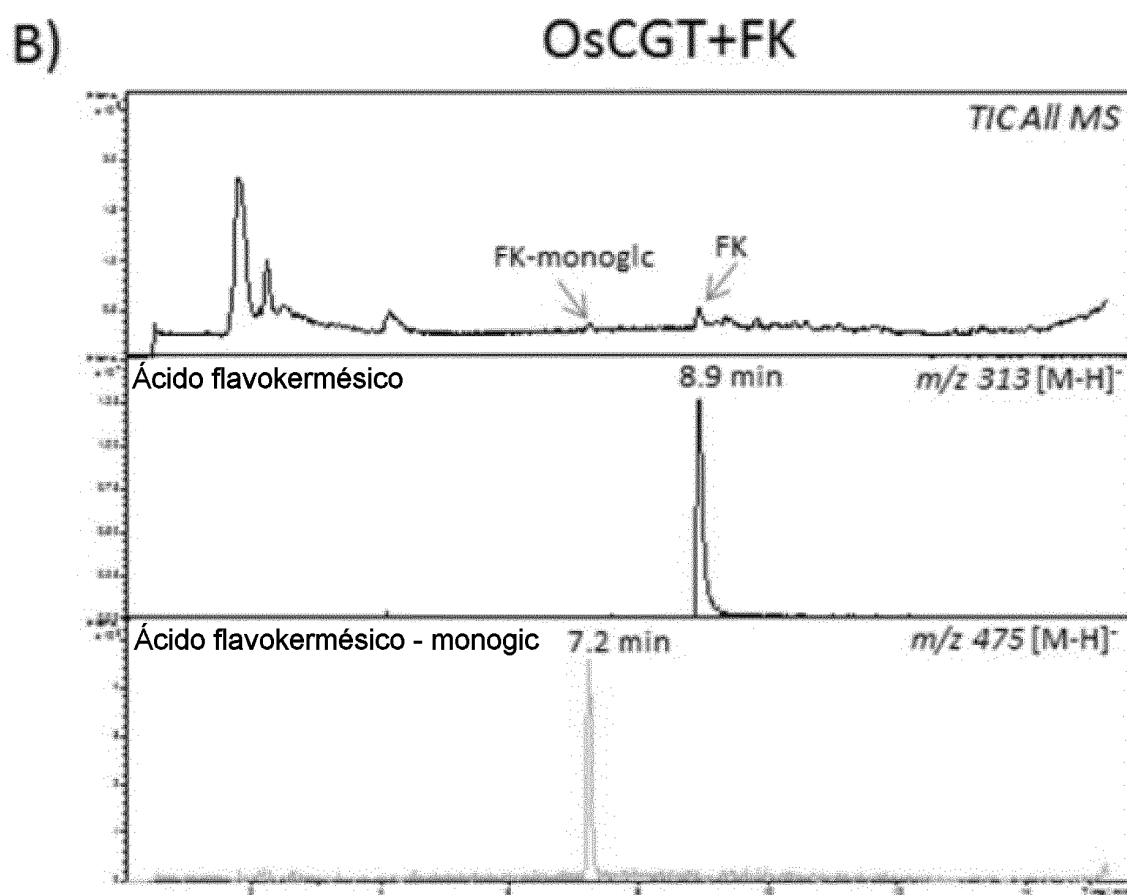


Figura 2 (C)

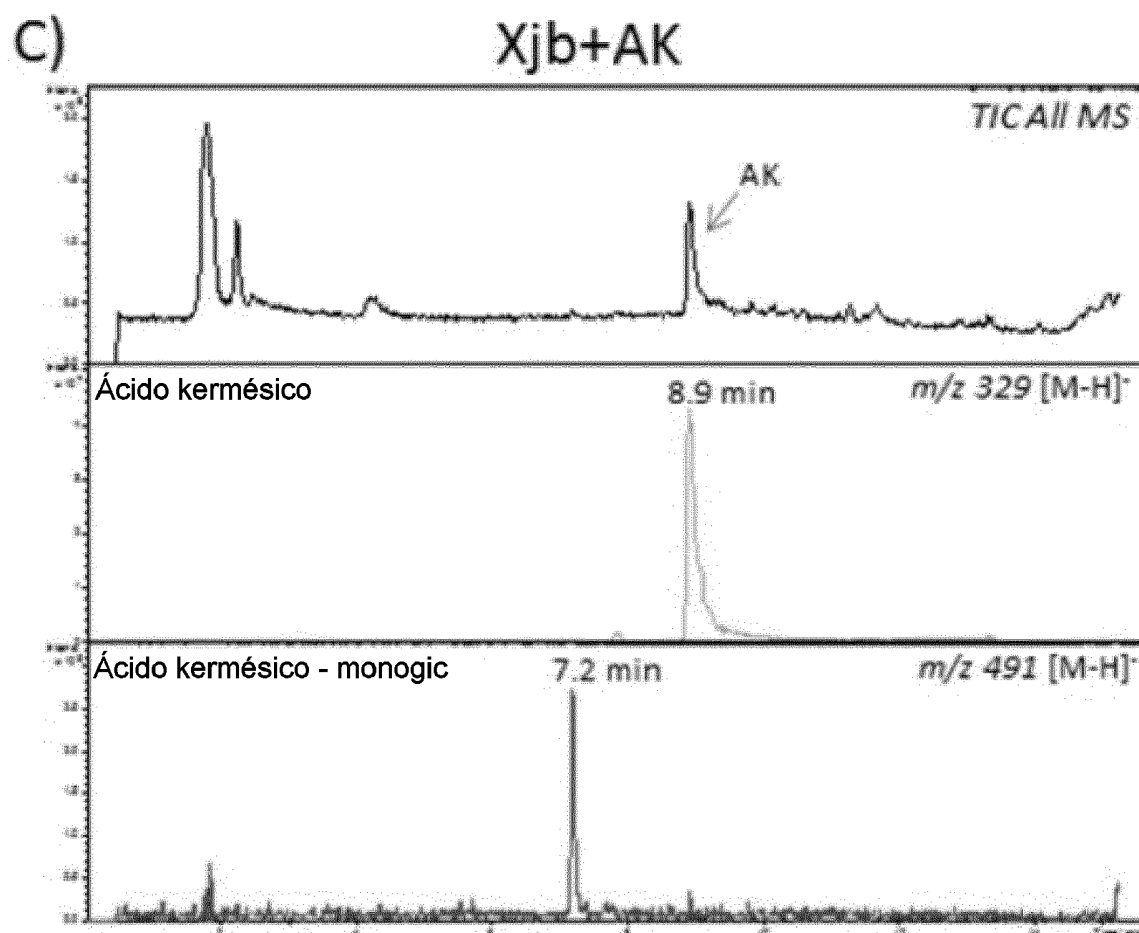


Figura 2 (D)

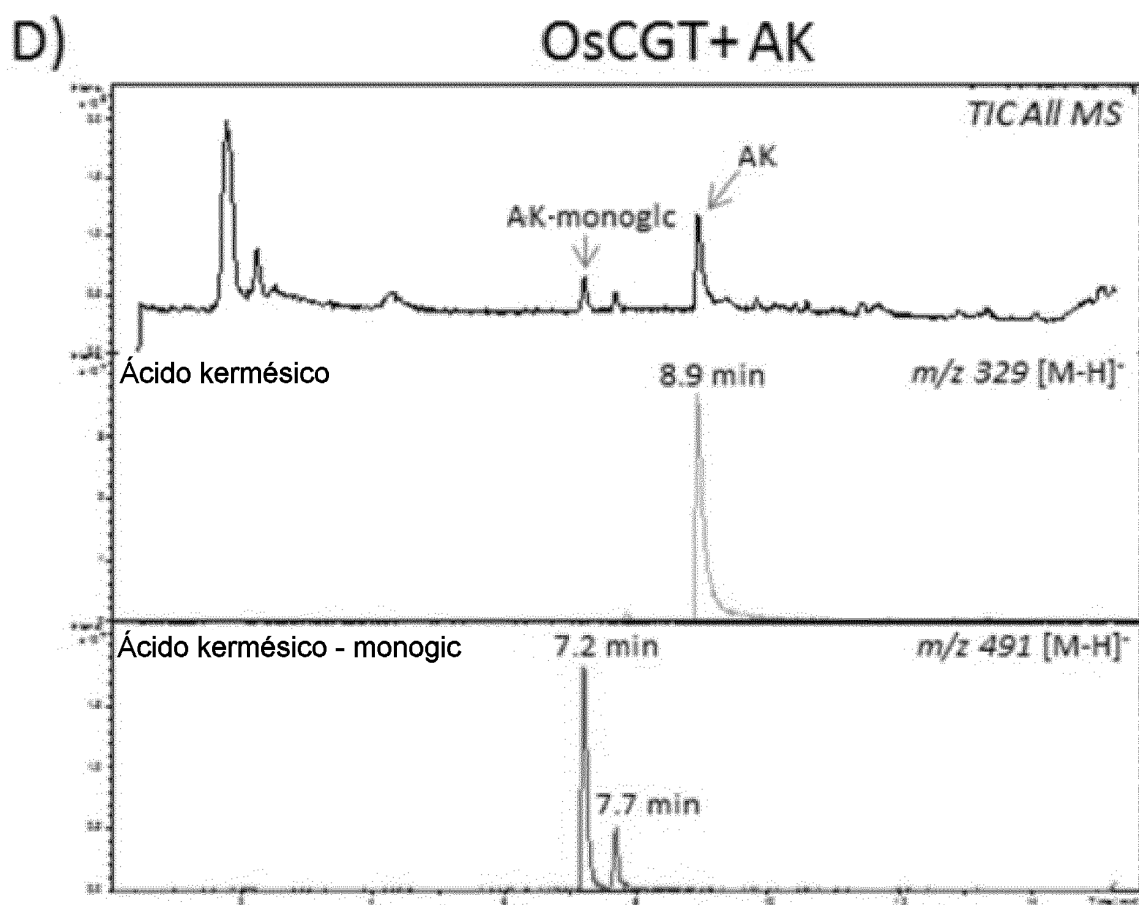


Figura 2 (E)

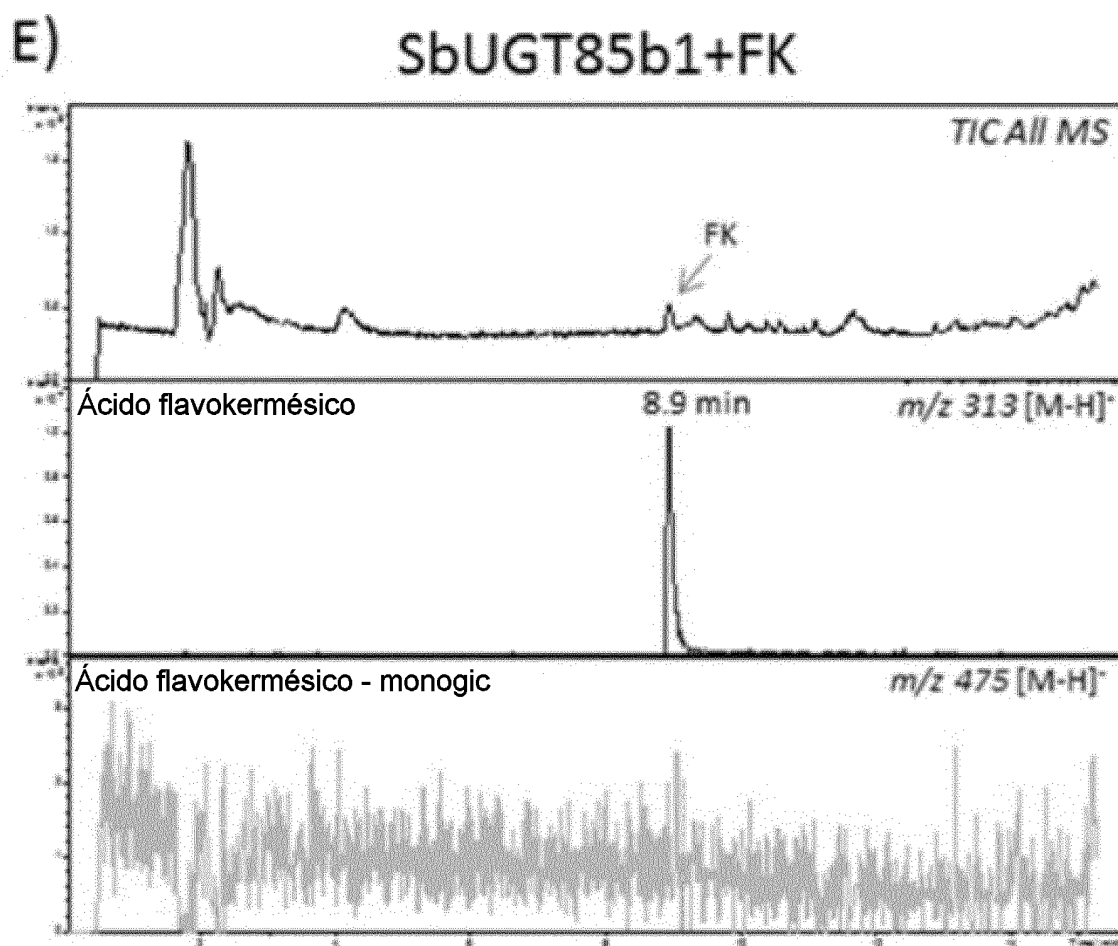


Figura 2 (F)

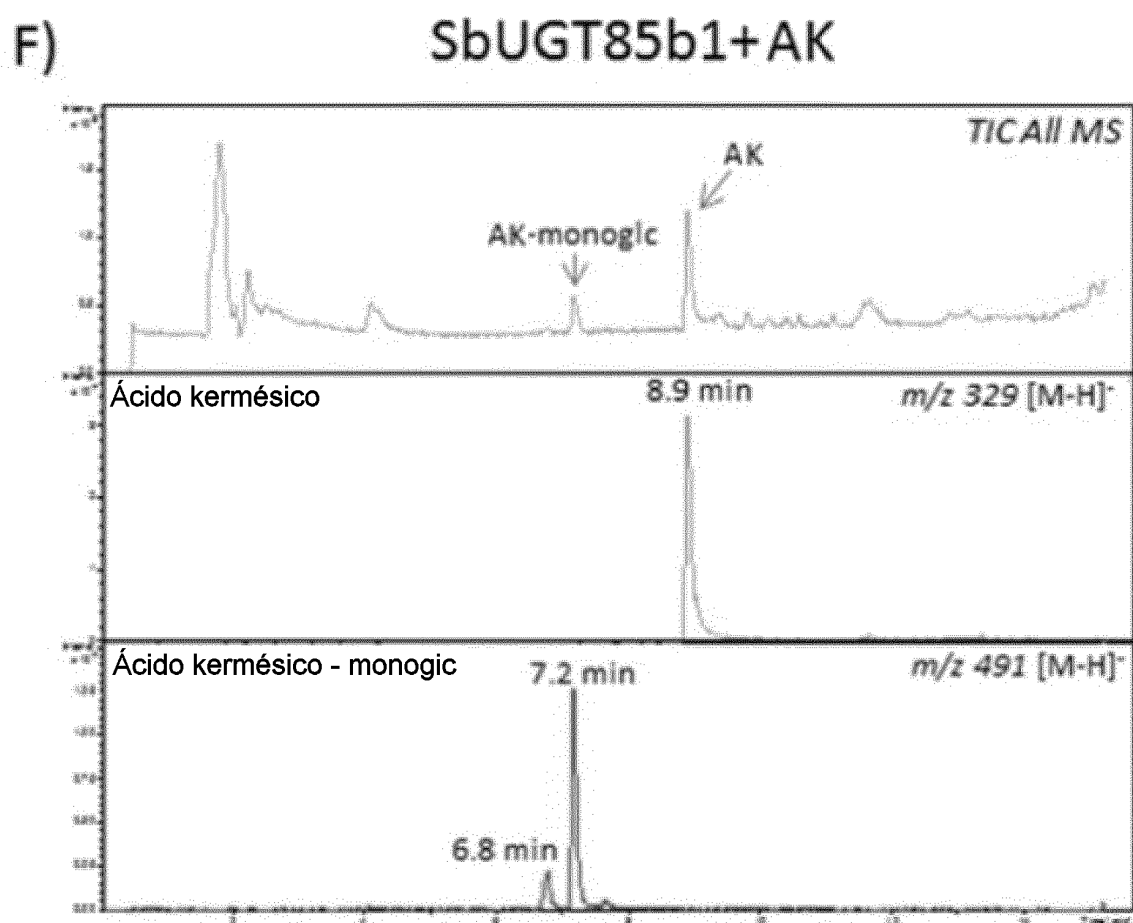


Figura 2 (G)

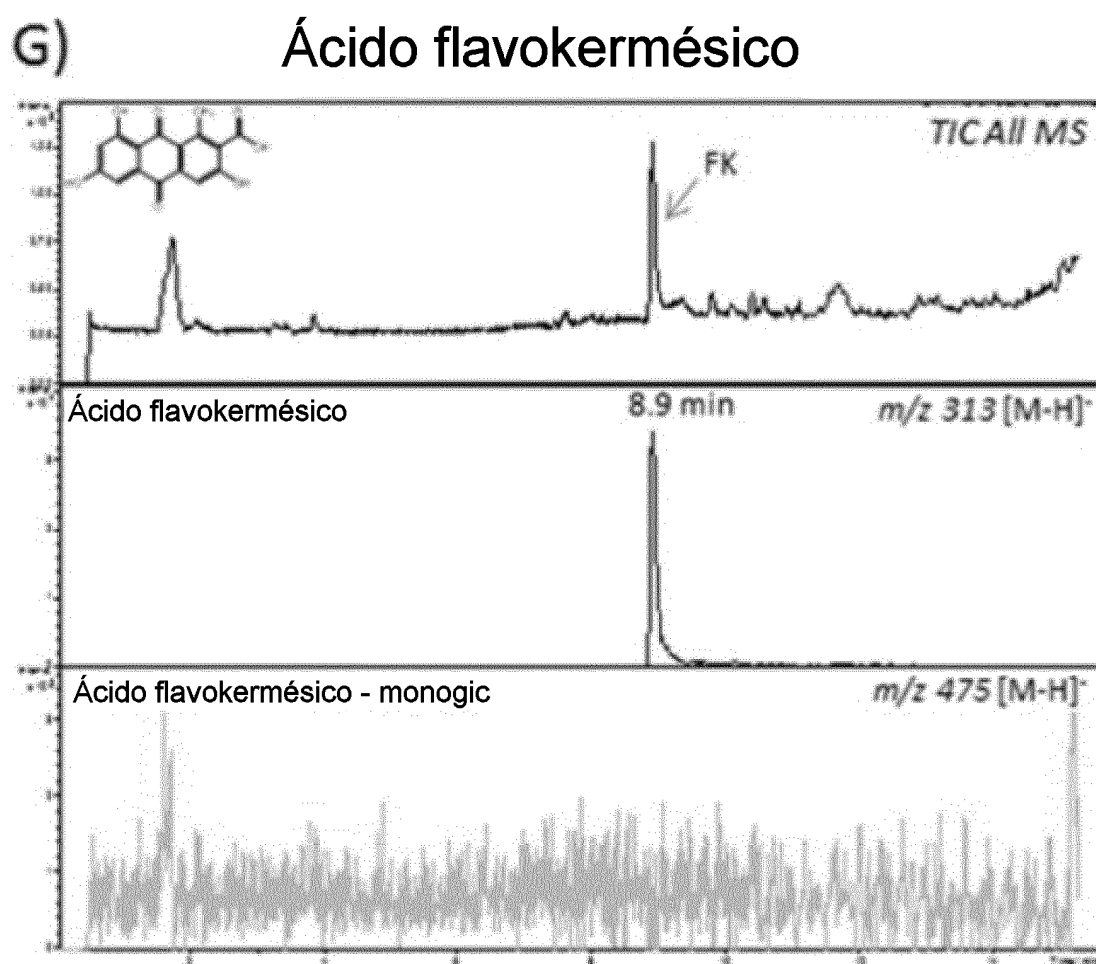


Figura 2 (H)

H) Ácido kermésico

