

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 671 223**

51 Int. Cl.:

C12N 9/02	(2006.01)
C12N 9/10	(2006.01)
C12N 15/52	(2006.01)
C12P 7/64	(2006.01)
C12P 19/12	(2006.01)
C12P 19/44	(2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **10.06.2011 PCT/EP2011/059683**

87 Fecha y número de publicación internacional: **15.12.2011 WO11154523**

96 Fecha de presentación y número de la solicitud europea: **10.06.2011 E 11730241 (4)**

97 Fecha y número de publicación de la concesión europea: **11.04.2018 EP 2580321**

54 Título: **Cepas de levaduras modificadas en su producción de sofrolípidos y sus usos**

30 Prioridad:

11.06.2010 GB 201009882

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

05.06.2018

73 Titular/es:

**UNIVERSITEIT GENT (100.0%)
Sint-Pietersnieuwstraat 25
9000 Gent, BE**

72 Inventor/es:

**SOETAERT, WIM;
DE MAESENEIRE, SOFIE;
SAERENS, KAREN;
ROELANTS, SOPHIE y
VAN BOGAERT, INGE**

74 Agente/Representante:

LEHMANN NOVO, María Isabel

ES 2 671 223 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

Cepas de levaduras modificadas en su producción de soforolípidos y sus usos

Campo técnico de la invención

5 La presente invención se refiere a especies de levaduras que normalmente son capaces de producir soforolípidos, pero que están tan modificadas que se vuelven incapaces de producir estos últimos compuestos. Estas cepas negativas a soforolípidos exhiben sorprendentemente características de crecimiento y formación de biomasa iguales a las de sus homólogos de tipo salvaje y, por lo tanto, son útiles para la producción de numerosos compuestos útiles tales como proteínas recombinantes, glicolípidos, polihidroxicanoatos, azúcares especiales, ácidos grasos especiales, escualeno, ácidos orgánicos, compuestos hidrofóbicos y carotenoides. Además, la presente invención describe dos genes de glucosiltransferasa con funciones clave en la producción de soforolípidos y su uso.

Técnica de antecedentes

15 La levadura no patógena *Candida bombicola* y otras especies de levadura tales como *Candida apicola*, *Candida batistae*, *Rhodotorula bogoriensis* y *Wickerhamiella domericqiae* son conocidas por su producción de soforolípidos durante la fase estacionaria (Spencer *et al.*, 1970, Gorin *et al.*, 1961, Tulloch *et al.* 1968, documentos EP0837140A1, US6433152, US 4215213). *C. bombicola* y otras son especies de levadura oleaginosas, es decir, pueden utilizar sustratos oleaginosos tales como alcanos y aceites como fuente de carbono, y pueden manejar esos sustratos en concentraciones relativamente altas. Además, *C. bombicola* puede producir soforolípidos en altas cantidades (más de 400 g/L), que son excretadas en el medio de fermentación.

20 *Candida bombicola* ATCC 22214 ya se aplica comercialmente para la producción de soforolípidos. Estos biotensioactivos de glicolípidos están constituidos por un grupo de cabeza de soforosa (2-O-β-D-glucopiranosil-β-D-glucopiranososa) a partir del cual el átomo de C anomérico se une a un ácido graso C₁₆ o C₁₈ (ω) o (ω-1) hidroxilado. Ocurren como estructuras de anillo abierto (forma ácida) o como lactonas con una intra-esterificación entre el grupo carboxilo de ácido graso y el átomo de carbono 4", 6' o 6" del grupo de cabeza de soforosa. Además, los grupos acetilo se pueden unir en las posiciones 6' y/o 6" (Asmer *et al.*, 1988).

25 En una fermentación típica de *Candida bombicola* con, p. ej., aceite de colza como fuente de carbono hidrofóbica, los soforolípidos están presentes como una mezcla compleja de moléculas estructuralmente relacionadas, siendo los soforolípidos lactónicos mono- y di-acetilados los más importantes.

30 Si bien muchas investigaciones se han centrado en las condiciones de fermentación para optimizar la producción de soforolípidos por parte de *C. bombicola* (Daniel *et al.*, 1998a, Daniel *et al.*, 1998b, Casas *et al.*, 1999, Cavalero *et al.*, 2003, Kim *et al.*, 2009), y los soforolípidos han servido como sustratos para modificaciones (químico)-enzimáticas (Bisht *et al.*, 1999, Hu *et al.*, 2003, Rau *et al.*, 1999), la vía bioquímica de estos bioproductos económicamente importantes sigue sin estar clara y no hay información disponible sobre los genes implicados (Ochsner *et al.*, 1994a, Ochsner *et al.*, 1994b). Esta falta de información dificulta la implementación de técnicas modernas como la ingeniería metabólica para aumentar los rendimientos de soforolípidos. Los únicos datos sobre enzimas implicadas en la producción de soforolípidos por parte de *C. bombicola* sugieren la implicación de una monooxigenasa del citocromo P450 de la familia CYP52 (Van Bogaert *et al.*, 2009). Los datos sobre enzimas en otras levaduras se limitan a experimentos de proteínas con lisados de *C. bogoriensis* y datan de principios de los años 1970 y 1980 (Esders *et al.*, 1972, Bucholtz *et al.*, 1976, Breithaupt *et al.*, 1982). Además de una acetiltransferasa se supuso que dos glucosiltransferasas estaban implicadas en la producción escalonada de soforolípidos por parte de este organismo, aunque la separación de las dos actividades no tuvo éxito (Breithaupt *et al.*, 1982).

40 Además del hecho de que la vía bioquímica de la producción de soforolípidos es en gran parte desconocida, *C. bombicola* siempre producirá soforolípidos sin importar qué fuente de carbono se emplee. Por lo tanto, los soforolípidos siempre formarán una mezcla con otros compuestos de interés potencial y la vía biosintética competirá por el uso de sustratos. Consecuentemente, es casi imposible combinar la producción (recombinante) de un compuesto de interés con la síntesis de soforolípidos sin una pérdida sustancial de eficacia y sin costes de purificación adicionales. La opción de producir sólo el compuesto de interés durante la fase de crecimiento exponencial cuando no se producen o se producen cantidades bajas de soforolípidos, no sólo dará como resultado rendimientos más bajos debido a la menor cantidad de biomasa disponible y al menor tiempo de producción, sino, a menudo, una baja tolerancia de las células en crecimiento hacia los sustratos oleaginosos utilizados. Estos problemas se omiten cuando se utilizan cepas que carecen de esta producción de soforolípidos. Sin embargo, Ito e Inoue (Ito *et al.*, 1982; Inoue *et al.*, 1982) demostraron que los soforolípidos estimulan el crecimiento en sustratos oleaginosos, mientras que una cantidad de tensioactivos no iónicos sintéticos no tiene efecto alguno. Afirman, además, que los soforolípidos actúan como factores específicos estimulantes del crecimiento y se necesitan para emulsionar los sustratos oleaginosos insolubles. Por lo tanto, cepas incapaces de producir soforolípidos muestran un crecimiento inferior en comparación con el tipo salvaje y no pueden manejar sustratos oleaginosos.

Tomados en conjunto, está claro que las especies de levadura productoras de soforolípidos podrían ser muy útiles para producir, además de soforolípidos, otros numerosos compuestos útiles tales como proteínas recombinantes, glicolípidos, polihidroxicanoatos, soforosa, ramnosa, ácidos grasos especiales, escualeno, ácidos orgánicos, compuestos hidrofóbicos y compuestos oleaginosos. Sin embargo, la falta de comprensión de la vía bioquímica subyacente de la síntesis de soforolípidos y el problema de combinar la producción de un compuesto útil de interés con la síntesis de soforolípidos sin una pérdida sustancial de eficacia y sin costes adicionales de purificación dificulta seriamente el uso de estas especies de levadura.

Descripción de las tablas:

10	Tabla 1	Cebadores utilizados para inactivar el gen <i>CYP52M1</i> de <i>C. bombicola</i> . Todos los cebadores se obtuvieron de Sigma Genosys.
	Tabla 2	Cebadores utilizados para el aislamiento del gen <i>UGTA1</i> y construcción del casete inactivado. Todos los cebadores se obtuvieron de Sigma Genosys.
	Tabla 3	Diez mejores puntuaciones de homología para la secuencia <i>UGTA1</i> traducida
15	Tabla 4	Cebadores utilizados para el aislamiento del gen <i>UGTB1</i> y construcción del casete inactivado. Todos los cebadores se obtuvieron de Sigma Genosys.
	Tabla 5	Diez mejores puntuaciones de homología para la secuencia <i>UGTB1</i> traducida
	Tabla 6	Cebadores utilizados para la creación del casete de expresión de GFP. Todos los cebadores se obtuvieron de Sigma Genosys. Los caracteres en negrita representan extensiones sin unión.
20	Tabla 7	Cebadores utilizados para crear el casete de expresión de amilasa y el control de transformantes de amilasa.
	Tabla 8	Cebadores utilizados para crear el casete de expresión de PHA. Todos los cebadores se obtuvieron de Sigma Genosys.
	Tabla 9	Cebadores utilizados para la creación de los casetes de expresión de <i>UGT1</i> y <i>CepB</i> . Todos los cebadores se obtuvieron de Sigma Genosys.

25 Descripción de las Figuras:

	Figura 1	Análisis por HPLC-ELSD del medio de producción de soforolípidos para <i>Candida bombicola</i> de tipo salvaje (arriba) y una de las cepas <i>cyp52M1</i> -negativas (abajo). Los soforolípidos de clase principal se detectan entre 25 y 30 min. El ácido oleico y el ácido linoleico eluyen a 33,5 y 36,2 min, respectivamente.
30	Figura 2	Crecimiento de <i>C. bombicola</i> de tipo salvaje, A113 (mutante de delección <i>ugtA1</i>) y B11 (mutante de delección <i>ugtB1</i>) en medio Lang. Todos los cultivos se inocularon a partir de un precultivo de Lang de tal manera que todos los cultivos comenzaron con una DO de 0,2.
	Figura 3	Secuencia completa de ADN del gen <i>UGTA1</i> (número de acceso GenBank HM440973, SEQ ID N° 1). Los elementos putativos de promotor y terminador están indicados por recuadros, mientras que un posible motivo reglamentario tipo GATA está sombreado en gris. También se representa la secuencia de aminoácidos codificada (SEQ ID N° 2) de la proteína UgtA1 glucosiltransferasa.
35	Figura 4	Cromatogramas por HPLC-ELSD de extractos de cultivo, 10 días después de la incubación de <i>C. bombicola</i> ATCC22214 (arriba) y <i>C. bombicola</i> A113 (abajo) con aceite de colza. Los soforolípidos <i>de novo</i> eluyen típicamente entre 25 y 30 min.
40	Figura 5	Cromatograma por HPLC-ELSD de extractos de muestra de un ensayo de glucosiltransferasa I en fracciones de proteínas solubles de levadura de tipo salvaje (arriba) y mutante A113 (abajo). Los picos a los 27 min y 29 min proceden de la síntesis de soforolípidos <i>de novo</i> , mientras que los picos que eluyen después de 30 minutos proceden de constituyentes de células apolares co-extraídas. FA 17-OH C18:1 = ácido 17-hidroxiocetadecenoico, GL = glucolípido (ácido 17-O-(β-D-glucopiranosil)-octadecenoico), SL = ácido soforolípido diacetilado (ácido 17-O-(2-O-β-D-glucopiranosil)-glucopiranosil)-octadecenoico), lacSL 2Ac = lactona soforolípida diacetilada
45	Figura 6	Cromatograma por HPLC-ELSD de extractos de muestra de un ensayo de glucosiltransferasa II en fracciones de proteínas solubles de levadura de tipo salvaje (arriba) y mutante A113 (abajo). Los picos a los 27 min y 29 min proceden de la síntesis de soforolípidos <i>de novo</i> , mientras que los picos que eluyen después de 30 min proceden de constituyentes de células apolares co-extraídas. FA 17-OH C18:1 = ácido 17-hidroxiocetadecenoico, GL = glucolípido (ácido 17-O-(β-D-glucopiranosil)-octadecenoico), SL = ácido soforolípido diacetilado (ácido 17-O-(2-O-β-D-glucopiranosil)-glucopiranosil)-octadecenoico), lacSL 2Ac = lactona soforolípida diacetilada
50	Figura 7	Secuencia completa de ADN del gen <i>UGTB1</i> (número de acceso GenBank HM440974, SEQ ID N° 3). Los sitios de los cebadores están subrayados. Los elementos putativos de promotor y terminador están indicados por cuadros, mientras que los posibles motivos reguladores tipo GATA están sombreados en gris. También se representa la secuencia de aminoácidos codificada (SEQ ID N° 4) de la proteína UgtB1 glucosiltransferasa.
55	Figura 8	Alineamiento global por pares de las secuencias de proteínas de <i>C. bombicola</i> UgtA1 (superior) y UgtB1 (inferior). Los aminoácidos similares están sombreados en gris, mientras que los residuos
60		

- idénticos están sombreados en negro. Los 14 residuos conservados del dominio GT1_Gtf_like se indican con flechas hacia abajo para UgtA1 y flechas hacia arriba para UgtB1.
- Figura 9 Cromatogramas por HPLC-ELSD de extractos de cultivo, 7 días después de la incubación de *C. bombicola* ATCC22214 (arriba) y *C. bombicola* B11 (abajo) con aceite de colza. Los soforolípidos *de novo* eluyen generalmente entre 25 y 30 min. SL = soforolípidos, GL = glucolípidos.
- 5 Figura 10 Cromatograma por HPLC-ELSD de extractos de muestras de un ensayo de glucosiltransferasa I en fracciones de proteínas solubles de levadura de tipo salvaje (arriba) y mutante B11 (abajo). Los picos a los 27 min y 29 min provienen de la síntesis de soforolípidos *de novo*, mientras que los picos que eluyen después de 30 min provienen de constituyentes de células apolares co-extraídas. FA 17-OH C18:1 = ácido 17-hidroxiocetadecenoico, GL = glucolípidos (ácido 17-O-(β-D-glucopiranosil)-octadecenoico), SL = ácido soforolípidos diacetilado (ácido 17-O-(2-O-β-D-glucopiranosil-glicopiranosil)-octadecenoico), lacSL 2Ac = lactona soforolípidos diacetilada
- 10 Figura 11 Cromatograma por HPLC-ELSD de extractos de muestras de una prueba de actividad de glucosiltransferasa II en fracciones de proteínas solubles de levadura de tipo salvaje (arriba) y mutante B11 (abajo). Los picos a los 27 min y 29 min proceden de la síntesis de soforolípidos *de novo*, mientras que los picos que eluyen después de 30 min proceden de constituyentes de células apolares co-extraídas. FA 17-OH C18:1 = ácido 17-hidroxiocetadecenoico, GL = glucolípidos (ácido 17-O-(β-D-glucopiranosil)-octadecenoico), SL = ácido soforolípidos diacetilado (ácido 17-O-(2-O-β-D-glucopiranosil-glicopiranosil)-octadecenoico), lacSL 2Ac = lactona soforolípidos diacetilada
- 15 Figura 12 Señal fluorescente en función del tiempo para el tipo salvaje y el mutante E1 GFP
- Figura 13 Mapa molecular de la construcción sintética de amilasa, precedida por el promotor parcial de GAPD
- Figura 14 Secuencia de la construcción sintética amilasa, precedida por parte del promotor de GAPD
- Figura 15 Vector p_sAmyAO-pGapd_iUra
- 25 Figura 16 Producción de amilasa por el transformante de amilasa soforolípidos-negativo (SL-Amy+) y el tipo salvaje (WT) de *C. bombicola*
- Figura 17 Perfil metabólico del transformante de amilasa soforolípidos-negativo (SL-Amy+, izquierda) y el tipo salvaje (WT, derecha) de *C. bombicola*
- Figura 18 Casete de expresión de PHA y su integración en el locus *cyp52m1*
- 30 Figura 19 Análisis GC-MS de los FAMES derivados de muestras finales del mutante de *Candida bombicola* PHAC1 A8 cultivado en medio de Lang con adición de aceite de colza después de 48 h de cultivo
- Figura 20 Espectros de masas de los tres glucolípidos predominantes encontrados en un extracto de cultivo B11 14 días después del cultivo en aceite de colza (37,5 g/L).

35 Descripción detallada de la invención

La presente invención se define en las reivindicaciones adjuntas.

La memoria descriptiva demuestra que, sorprendentemente, las cepas soforolípidos-negativas mostraron iguales características de crecimiento y formación de biomasa en comparación con el tipo salvaje y no se inhibieron por la presencia de sustratos oleaginosos. La memoria descriptiva describe, además, que las cepas soforolípidos-negativas pueden utilizarse para producir compuestos útiles. Además, la presente invención también describe dos genes glucosiltransferasa con funciones clave en la producción de soforolípidos.

40

Por lo tanto, la memoria descriptiva se refiere a una cepa de levadura modificada perteneciente a una especie de levadura capaz de producir soforolípidos, caracterizada porque dicha cepa de levadura tiene, en comparación con una cepa de tipo salvaje, no modificada: a) al menos una mutación en su genoma, y b) una reducción en su capacidad para producir soforolípidos de al menos 75%, y en donde dichos soforolípidos están constituidos por el azúcar soforosa unido a ácido graso C₁₆, C₁₈, C₂₂ o C₂₄ hidroxilado. La reducción en dicha capacidad debe medirse por fermentación de dicho mutante y dicha cepa de tipo salvaje exactamente en las mismas condiciones de fermentación, y de forma similar, la producción de soforolípidos por dicho mutante y dicho tipo salvaje se debe medir exactamente por el mismo método. La cuantificación de dicha producción de soforolípidos puede realizarse - por ejemplo - después de la precipitación desde el medio de fermentación o después de la extracción con acetato de etilo tal como se describe por Lang et al. (2000)

45

50

La expresión "especie de levadura capaz de producir soforolípidos" se refiere a un grupo filogenéticamente diverso de levaduras (predominantemente Ascomycetes y unos pocos Basidiomycetes) que sintetizan espontáneamente soforolípidos constituidos por el azúcar soforosa unido a un ácido graso hidroxilado. Dicho grupo filogenéticamente diverso de levaduras comprende la especie *Candida apicola* (Gorin et al., 1961) que se identificó inicialmente como *C. magnolia*, *C. bombicola* (Spencer et al., 1970), *Wickerhamiella domericqiae* (Chen et al., 2006), *Rhodotorula bogoriensis* (Tulloch et al., 1968), *Pichia anómala* PY1, *Candida batistae* (Konishi et al., 2008), *Candida florícola* (Imura et al., 2010), *Candida riodocensis*, *Candida stellata* y *Candida* sp. NRRL Y-27208 (Kurtzman et al., 2010) y

55

otras especies del llamado clado Starmerella que incorpora a más de 40 especies. *C. bombicola* ha sido recientemente reasignada al género Starmerella (Rosa y Lachance, 1998).

Por lo tanto, la memoria descriptiva se refiere, además, a una cepa de levadura modificada tal como se indicó anteriormente, en donde dicha especie de levadura se selecciona del grupo de levaduras productoras de soforolípidos, que consiste en *Candida bombicola*, *Candida apicola*, *Candida batistae*, *Candida floricola*, *Candida riodecensis*, *Candida stellata*, *Candida sp.* NRRL Y-27208, *Rhodotorula bogoriensis*, *Pichia anómala* PY1, *Wickerhamiella domericqiae* y especies productoras de soforolípidos del clado Starmerella. Más específicamente, la memoria descriptiva se refiere a una cepa de levadura modificada tal como se indicó arriba, en donde dicha *Candida bombicola* es la cepa *Candida bombicola* ATCC 22214.

5 La expresión "cepa de levadura modificada" indica que el material genético de dicha cepa de levadura ha sido alterado de modo que la cepa de levadura es incapaz, o es casi incapaz de producir soforolípidos. La expresión "casi incapaz" indica que dicha cepa de levadura modificada es capaz únicamente de producir un máximo de 25% de la producción total por unidad de tiempo de soforolípidos mediante una denominada cepa de tipo salvaje no modificada. Las cepas de levadura modificadas pueden ser únicamente capaces de producir un máximo de 20%,
15, 10% o 5% de la producción total por unidad de tiempo de soforolípidos mediante una denominada cepa salvaje no modificada. Se prefieren las cepas de levadura modificadas que son completamente incapaces (0%) de producir soforolípidos. La expresión 'cepa salvaje no modificada' se refiere a una cepa de levadura tal como se produce en la naturaleza y que es totalmente capaz (100%) de producir soforolípidos.

20 Más específicamente, la expresión "cepa de levadura modificada" se refiere a una cepa de levadura que contiene al menos una mutación en su genoma. El término mutación se refiere a una mutación espontánea y/o a una mutación inducida en el genoma de dicha cepa de levadura. Dicha mutación puede ser una mutación puntual, delección, inserción o cualquier otro tipo de mutación. La expresión lo más específicamente se refiere a las inactivaciones (KO) a través de la inserción de un casete KO.

25 La inducción de una mutación en el genoma de una cepa de levadura se puede llevar a cabo mediante cualquier método en la técnica conocido por una persona experta, tal como la inserción de un casete KO en un gen de interés.

De forma similar, el rastreo o la detección de si existe una mutación en el genoma de un gen modificado, en comparación con una cepa de tipo salvaje, también se puede determinar mediante cualquier método conocido en la técnica.

30 El término "soforolípidos" se refiere a biotensioactivos anfífilos basados en hidratos de carbono que están constituidos por el azúcar soforosa unido a un ácido graso C₁₆, C₁₈, C₂₂ o C₂₄ hidroxilado, es decir, ácidos grasos hidroxilados en los que la cadena de ácidos grasos se compone de 16, 18, 22 o 24 átomos de C. Más específicamente, el término se refiere a biotensioactivos glicolípidos que están constituidos por un grupo de cabeza de soforosa (2-O-β-D-glucopiranosil-D-glucopiranososa) a partir de la cual el átomo de C anomérico se une a un ácido graso C₁₆, C₁₈ (ω) o (ω-1) hidroxilado o a un ácido graso C₂₂ o C₂₄ hidroxilado en la posición C₁₃. Con respecto a *C. bombicola*, dicha cadena de ácido graso está compuesta por 16 o 18 átomos de C. Se producen como estructuras de anillo abierto (forma ácida) o como lactonas con una intra-esterificación entre el grupo carboxilo del ácido graso y el átomo de carbono 4", 6' o 6" del grupo de cabeza de soforosa. Además, los grupos acetilo se pueden unir en las posiciones 6' y/o 6" (Asmer *et al.*, 1988). En una fermentación típica de *Candida bombicola* con, p. ej., aceite de colza como fuente de carbono hidrofóbica, los soforolípidos están presentes como una mezcla compleja de moléculas estructuralmente relacionadas, siendo los más importantes los soforolípidos lactónicos mono- y di-acetilados.

45 La memoria descriptiva describe una cepa de levadura modificada tal como se indicó anteriormente, en donde dicha mutación es una delección en un gen que codifica una proteína, más específicamente en donde dicha proteína es una enzima o una proteína reguladora, implicada en la vía biosintética de soforolípidos. Más específicamente, la memoria descriptiva describe mutaciones en los genes *CYP52M1* (Van Bogaert *et al.*, 2009) que codifican una monooxigenasa del citocromo P450, *UGTA1* (representada por SEQ ID N° 1) que codifica la glucosiltransferasa 1 y *UGTB1* (representada por SEQ ID N° 3) que codifica la glucosiltransferasa 2.

50 Por lo tanto, y más específicamente, la memoria descriptiva se refiere a una cepa de levadura modificada tal como se indicó arriba, en donde dicho gen codifica una monooxigenasa del citocromo P450 o una glucosiltransferasa, y más específicamente, en donde dicho gen que codifica una monooxigenasa del citocromo P450 es el gen *CYP52M1* que tiene número de acceso de Genbank EU552419 y en donde dicho gen que codifica una glucosiltransferasa es el gen *UGTA1* que tiene una secuencia tal como se representa en la SEQ ID N° 1 y que tiene el número de acceso Genbank HM440973 o es el gen *UGTB1* que tiene una secuencia tal como se representa por la SEQ ID N° 3 y que tiene el número de acceso Genbank HM440974.

La memoria descriptiva describe, además, una secuencia de ácido nucleico tal como se representa mediante la SEQ ID N° 1 que codifica la UDP-glucosiltransferasa UgtA1 responsable del primer paso de glucosilación en la vía biosintética de soforolípidos de *Candida bombicola*, o un fragmento del mismo que codifica una proteína que conserva dicha actividad de UDP-glucosiltransferasa, o una variante de la misma que codifica una proteína que tiene al menos un 50% de identidad de secuencia con SEQ ID N° 2 y que tiene dicha actividad de UDP-glucosiltransferasa. El término "fragmento" se refiere a una secuencia de ácidos nucleicos que contiene menos nucleótidos que la secuencia de ácidos nucleicos tal como se representa en la SEQ ID N°: 1 y que codifica una proteína que conserva dicha actividad de UDP-glucosiltransferasa. El término "variante" se refiere a un ácido nucleico que codifica una proteína que tiene al menos un 50% de identidad de secuencia, preferiblemente que tiene al menos un 51-70% de identidad de secuencia, más preferiblemente que tiene al menos un 71-90% de identidad de secuencia o más preferiblemente que tiene al menos un 91, 92, 93, 94, 95, 96, 97, 98 o 99% de identidad de secuencia con SEQ ID N°: 2 o con un fragmento de la misma, y que codifica una proteína que conserva dicha actividad de UDP-glucosiltransferasa.

La memoria descriptiva también describe una secuencia de aminoácidos tal como se representa por SEQ ID N° 2 y correspondiente a la UDP-glucosiltransferasa UgtA1 responsable del primer paso de glucosilación en la vía biosintética de soforolípidos de *Candida bombicola* o un fragmento del mismo que tiene dicha actividad de UDP-glucosiltransferasa o una variante de la misma que tiene al menos un 50% de identidad de secuencia con SEQ ID N° 2 y que tiene dicha actividad de UDP-glucosiltransferasa. El término 'fragmento' se refiere a una proteína que contiene menos aminoácidos que la secuencia de aminoácidos tal como se representa por SEQ ID N° 2 y que conserva dicha actividad de UDP-glucosiltransferasa. Un fragmento de este tipo puede, por ejemplo, omitir los extremos N y C de la proteína, dejando una proteína acortada de 366 aminoácidos que comienza en I10 y termina en A375, sin tocar con ello la actividad de UDP-glucosiltransferasa. El término 'variante' se refiere a una proteína que tiene al menos un 50% de identidad de secuencia, preferiblemente que tiene al menos un 51-70% de identidad de secuencia, más preferiblemente que tiene al menos un 71-90% de identidad de secuencia, o lo más preferiblemente que tiene al menos un 91, 92, 93, 94, 95, 96, 97, 98 o 99% de identidad de secuencia con SEQ ID N° 2 o con un fragmento de la misma, y que codifica una proteína que conserva dicha actividad de UDP-glucosiltransferasa. esto último puede diferir de la proteína representada por SEQ ID N° 2 o un fragmento de la misma sólo en sustituciones y/o modificaciones conservativas, de modo que se conserva la capacidad de la proteína de tener actividad de UDP-glucosiltransferasa. Una "sustitución conservativa" es aquella en la que un aminoácido es sustituido con otro aminoácido que tiene propiedades similares, de modo que un experto en la técnica de la química de proteínas esperaría que la naturaleza de la proteína se mantuviera sustancialmente sin cambios. En general, los siguientes grupos de aminoácidos representan cambios conservadores: (1) ala, pro, gly, glu, asp, gln, asn, ser, thr; (2) cys, ser, tyr, thr; (3) val, ile, leu, met, ala, phe; (4) lys, arg, his; y (5) phe, tyr, trp, his.

Las variantes también pueden (o alternativamente) ser proteínas tal como se describe en esta memoria, modificadas, por ejemplo, mediante la delección o adición de aminoácidos que tienen una influencia mínima sobre la actividad de UDP-glucosiltransferasa, la estructura secundaria y la naturaleza hidropática de la proteína. Regiones dentro de SEQ ID N° 2 que contribuyen en la actividad de las proteínas son regiones determinadas por los residuos I10 a V50, G131 a L155, M180 a H196, W213 a F233, S266 a Y282, V341 a H355, A375 a T386. Por lo tanto, las variantes, tal como se definen arriba, comprenden preferiblemente al menos una de las últimas regiones. Más preferiblemente, las últimas variantes comprenden al menos uno de los siguientes residuos o aminoácidos de SEQ ID N° 2: G19, H20, M22, N181, L184, E185, F189, S216, T277, N343, G345, G347, G348 o H351.

La memoria descriptiva describe, además, una secuencia de ácidos nucleicos tal como se representa por SEQ ID N° 3 que codifica la UDP-glucosiltransferasa UgtB1 responsable de la segunda etapa de glucosilación en la vía biosintética de soforolípidos de *Candida bombicola* o un fragmento de la misma que codifica una proteína que conserva dicha actividad de UDP-glucosiltransferasa, o una variante de la misma que codifica una proteína que tiene al menos un 50% de identidad de secuencia con SEQ ID N° 4 y que tiene dicha actividad de UDP-glucosiltransferasa. El término 'fragmento' se refiere a una secuencia de ácidos nucleicos que contiene menos nucleótidos que la secuencia de ácidos nucleicos tal como se representa por SEQ ID N° 3 y que codifica una proteína que conserva dicha actividad de UDP-glucosiltransferasa. El término "variante" se refiere a un ácido nucleico que codifica una proteína que tiene al menos un 50% de identidad de secuencia, preferiblemente que tiene al menos un 51-70% de identidad de secuencia, más preferiblemente que tiene al menos un 71-90% de identidad de secuencia, o lo más preferiblemente que tiene al menos un 91, 92, 93, 94, 95, 96, 97, 98 o 99% de identidad de secuencia con SEQ ID N° 4 o con un fragmento de la misma, y que codifica una proteína que conserva dicha actividad de UDP-glucosiltransferasa.

La memoria descriptiva también describe una secuencia de aminoácidos tal como se representa por SEQ ID N° 4 y correspondiente a la UDP-glucosiltransferasa UgtB1 responsable de la segunda etapa de glucosilación en la vía biosintética de soforolípidos de *Candida bombicola*, o un fragmento de la misma que tiene dicha actividad de UDP-glucosiltransferasa o una variante de la misma que tiene al menos un 50% de identidad de secuencia con SEQ ID N° 4 y que tiene dicha actividad de UDP-glucosiltransferasa. El término "fragmento" se refiere a una proteína que contiene menos aminoácidos que la secuencia de aminoácidos tal como se representa por SEQ ID N° 4 y que

5 conserva dicha actividad de UDP-glucosiltransferasa. Un fragmento de este tipo puede, por ejemplo, omitir los extremos N y C de la proteína, dejando una proteína acortada de 351 aminoácidos que comienza en I8 y termina en G358 sin tocar con ello la actividad de UDP-glucosiltransferasa. El término "variante" se refiere a una proteína que tiene al menos un 50% de identidad de secuencia, preferiblemente que tiene al menos un 51-70% de identidad de secuencia, más preferiblemente que tiene al menos un 71-90% de identidad de secuencia, o más preferiblemente que tiene al menos un 91, 92, 93, 94, 95, 96, 97, 98 o 99% de identidad de secuencia con SEQ ID N° 4 o con un fragmento de la misma, y que codifica una proteína que conserva dicha actividad de UDP-glucosiltransferasa. Esta última puede diferir de la proteína tal como se representa por SEQ ID N° 4 o un fragmento de la misma solo en sustituciones y/o modificaciones conservativas, de manera que se conserva la capacidad de la proteína de tener actividad de UDP-glucosiltransferasa. Una "sustitución conservativa" es aquella en la que un aminoácido está sustituido con otro aminoácido que tiene propiedades similares, de modo que un experto en la técnica de la química de proteínas esperaría que la naturaleza de la proteína se mantuviera sustancialmente sin cambios. En general, los siguientes grupos de aminoácidos representan cambios conservadores: (1) ala, pro, gly, glu, asp, gin, asn, ser, thr; (2) cys, ser, tyr, thr; (3) val, ile, leu, met, ala, phe; (4) lys, arg, his; y (5) phe, tyr, trp, his.

15 Las variantes también pueden (o alternativamente) ser proteínas tal como se describe en esta memoria, modificadas, por ejemplo, por la delección o la adición de aminoácidos que tienen una influencia mínima sobre la actividad de UDP-glucosiltransferasa, la estructura secundaria y la naturaleza hidropática de la proteína. Regiones dentro de SEQ ID N° 4 que contribuyen en la actividad de las proteínas son los residuos I8 a V48, G129 a C153, I170 a E186, L203 a F223, S261 a Y277, V334 a H350 y A371 a T382. Por lo tanto, las variantes tal como se definen arriba comprenden preferiblemente al menos una de las últimas regiones. Más preferiblemente, las últimas variantes comprenden al menos uno de los siguientes residuos o aminoácidos de SEQ ID N° 4: G17, H18, G20, R175, V178, F179, G182, R212, T272, N338, G340, G342, G343 o H346.

25 La presente memoria descriptiva describe, además, la expresión de SEQ ID N° 1, SEQ ID N° 2, SEQ ID N° 3 y SEQ ID N° 4 o variantes de las mismas en otros organismos distintos de las diversas especies de levadura capaces de producir soforolípidos tal como se especifica arriba. Estos "otros" organismos pueden ser microorganismos tales como bacterias, levaduras y hongos, plantas, animales y algas. La expresión de los genes en estos organismos conduce a la producción de compuestos útiles tales como compuestos glicosilados, glicolípidos o soforolípidos.

30 Tal como se especificó arriba, la presente memoria descriptiva describe específicamente el uso de SEQ ID N° 1, SEQ ID N° 2, SEQ ID N° 3 y SEQ ID N° 4 y sus variantes en especies de levaduras capaces de producir soforolípidos. Estas secuencias pueden manipularse en la cepa endógena (p. ej., en *C. bombicola*) inactivando el gen o pueden someterse a otras alteraciones tales como mutación o sobre-expresión. Estas manipulaciones se pueden obtener por diversos métodos conocidos por los expertos en la técnica. La alteración de la expresión dará como resultado la modificación de una vía biosintética de glucolípidos, más en particular la vía biosintética de soforolípidos.

35 La memoria descriptiva se refiere, además, al uso de una cepa de levadura modificada tal como se indica arriba, para la producción de compuestos tales como proteínas recombinantes (p. ej., GFP y amilasa), ácidos beta-hidroxi grasos y polihidroxialcanoatos (PHA), ácidos dicarboxílicos, ácidos grasos poliinsaturados, ácidos grasos hidroxilados: hidroxilados en posición terminal o subterminal o en cualquier otra posición, glicolípidos tales como celobiosalípidos, glucolípidos, trehalosalípidos, ramnolípidos, soforolípidos con una cola ácido graso especial, soforolípidos con una cola ácido graso que oscila entre 10 y 15 átomos de carbono, soforolípidos con una cola ácido graso de 17 átomos de carbono, soforolípidos con una cola ácido graso que oscila entre 19 y 25 átomos de carbono, soforolípidos con una cola ácido graso ramificada, soforolípidos con múltiples colas ácido graso hidroxiladas, soforolípidos completamente lactonizados y soforolípidos completamente ácidos; ramnosa, soforosa, antibióticos policetídicos, estructuras lactónicas (basadas en ácidos grasos), ácidos orgánicos tales como succinato, adipato y citrato; compuestos oleaginosos, compuestos hidrofóbicos, soforosa, ramnosa, escualeno, vitamina D, resveratrol, esteroides y carotenoides.

La memoria descriptiva se refiere además, específicamente, al uso tal como se indicó arriba, en donde dichos compuestos son glicolípidos (tales como celobiosalípidos y glucolípidos), polihidroxialcanoatos o cualesquiera otros compuesto oleaginosos y proteínas recombinantes tales como proteína fluorescente verde y amilasa.

50 Los usos arriba indicados pueden realizarse utilizando técnicas bien conocidas de la ingeniería genética.

Los usos arriba indicados se ilustrarán mediante los siguientes ejemplos no limitantes.

Ejemplos

Ejemplo 1: Inactivación de la monooxigenasa CYP52M1 del citocromo P450

Introducción

Esta cepa está inactivada en el gen *CYP52M1* (número de acceso de GenBank EU552419), que codifica la enzima responsable de la hidroxilación de ácidos grasos con una longitud preferida de 16 o 18 átomos de carbono. El ácido graso se convierte en un ácido graso ω o ω -1 hidroxilado.

- 5 En esta cepa, apenas se detecta producción de soforolípidos, mientras que el crecimiento y la viabilidad de las células es equiparable al tipo salvaje.

Materiales y métodos

Cepas, plásmidos y condiciones de cultivo

- 10 *Candida bombicola* ATCC 22214 se utilizó en todos los experimentos. Cuando se pretendía la producción de soforolípidos, se utilizó el medio descrito por Lang *et al.* (Lang *et al.*, 2000). Los cultivos de levadura se incubaron a 30°C y 200 rpm.

- 15 Todos los productos de PCR destinados para el análisis de secuencias se clonaron en el vector pGEM-T[®] (Promega). *Escherichia coli* DH5 α se utilizó en todos los experimentos de clonación y se transformó tal como se describe por Sambrook y Russell (2001). Células de *E. coli* se cultivaron en medio Luria-Bertani (LB) (triptona 1%, extracto de levadura 0,5% y cloruro de sodio 0,5%) complementado con 100 mg/L de ampicilina y 40 mg/L de X-gal si fuese necesario. Cultivos líquidos de *E. coli* se incubaron a 37°C y 200 rpm.

Aislamiento y secuenciación de ADN

El ADN genómico de levadura se aisló con el kit de ADN genómico bacteriano GenElute[™] (Sigma). La formación previa de protoplastos se realizó mediante incubación a 30°C durante 90 minutos con zimoliasa (Sigma).

- 20 El ADN del plásmido bacteriano se aisló con el kit QIAprep Spin Miniprep (Qiagen). Todas las secuencias de ADN se determinaron en VIB Genetic Service Facility (Bélgica).

Transformación

- 25 Células de *C. bombicola* se transformaron con el método de acetato de litio (Gietz *et al.*, 1995), pero se utilizó LiAc 50 mM en lugar de 100. Los transformantes se seleccionaron en placas de peptona dextrosa (YPD) de levadura (1% de extracto de levadura, 2% de peptona, 2% de glucosa y 2% de agar), 2% de glucosa y 2% de agar) que contiene 500 μ g/mL de higromicina o en placas sintéticas de dextrosa (SD) [0,67% de base nitrogenada de levadura sin aminoácidos (DIFCO) y 2% de glucosa]. Células de *E. coli* se transformaron tal como se describe por Sambrook (Sambrook *et al.*, 2001).

Creación del casete de inactivación

- 30 El fragmento codificante de 1617 pb y 218 y 1060 pb aguas arriba y aguas abajo del gen *CYP52M1* se amplificaron con los cebadores A21 TotFor y A21 TotRev (Tabla 1), proporcionando un fragmento de 2869 pb que se clonó en el vector pGEM-T[®] (Promega). El vector creado se digirió con *Ava*I, cortando dos veces la secuencia codificante de *CYP52M1*, eliminando de este modo 308 pb de la secuencia *CYP52M1*. El gen de resistencia a higromicina de *E. coli* controlado por el promotor GPD de *Candida bombicola* (Van Bogaert *et al.*, 2008a) se insertó por medio del kit de clonación por PCR Dry-Down In-Fusion[™] 2.0 (Clontech). Los cebadores GHInfA21For e HygroInfA21Rev se diseñaron de acuerdo con las directrices del manual y se utilizaron para la integración del casete de resistencia a higromicina (1968 pb) en *CYP52M1*. El par de cebadores A21KnockHygroCasFor y A21KnockHygroCasRev se utilizaron para la amplificación de un fragmento de 4003 pb que contiene el casete de resistencia a higromicina con aproximadamente 1000 pb de la secuencia de *CYP52M1* en cada sitio. Los fragmentos se utilizaron para transformar la cepa de tipo salvaje de *Candida bombicola*.

Muestreo

- 40 Muestras analíticas de soforolípidos se prepararon de la siguiente manera: se añadieron 440 μ L de acetato de etilo y 11 μ L de ácido acético a 1 mL de caldo de cultivo y se agitaron vigorosamente durante 5 min. Después de la centrifugación a 9 000 g durante 5 min, la capa de disolvente superior se retiró y se colocó en un tubo eppendorf nuevo con 600 μ L de etanol. Al final del período de incubación, se añadieron 3 volúmenes de etanol al caldo de cultivo para la extracción total de soforolípidos. Los desechos celulares se separaron por centrifugación a 1500 g durante 10 min. Las muestras se analizaron mediante HPLC y Detección de Dispersión de la Luz Evaporativa.

Se midió el peso seco celular (CDW) por centrifugación de 2 mL de caldo de cultivo durante 5 min a 9 000 g. Los sedimentos se lavaron dos veces con etanol para separar los soforolípidos y el sustrato hidrofóbico y finalmente se disolvieron en agua destilada. La suspensión se transfirió a un filtro de nitrato de celulosa con un diámetro de poro de 0,45 μ m (Sartorius) y el peso seco se determinó en el horno automático XM60 de Precisa Instruments Ltd.

Las unidades formadoras de colonias (UFC) se determinaron sembrando diluciones decimales en placas de agar con 10% de glucosa, 1% de extracto de levadura y 0,1% de urea que se incubaron a 30 °C durante tres días.

Análisis por HPLC de soforolípidos

5 Muestras de soforolípidos se analizaron por HPLC en un sistema HPLC Varian Prostar utilizando una columna Chromolith® Performance RP-18e de 100-4,6 mm de Merck KGaA a 30 °C y Detección de Dispersión de la Luz Evaporativa (Alltech). Se tuvo que utilizar un gradiente de dos eluyentes, una solución acuosa de ácido acético al 0,5% y acetonitrilo, para separar los componentes. El gradiente comenzó con 5% de acetonitrilo y aumentó linealmente hasta 95% en 40 min. La mezcla se mantuvo de esta manera durante 10 min y luego se devolvió a 5% de acetonitrilo en 5 min. Se empleó un caudal de 1 mL/min.

10 Resultados y discusión

15 El casete de inactivación *CYP52M1* se construyó tal como se describe en la sección Materiales y Métodos. Este fragmento lineal se utilizó para transformar células de tipo salvaje de *Candida bombicola*. El genotipo de 10 transformantes se controló mediante PCR de colonia de levaduras con el cebador HygroInserCheckFor, uniéndose al casete inactivado y al cebador A21totRev, uniéndose al gen genómico *CYP52M1* (Tabla 1). Todos los 10 transformantes mostraron el genotipo correcto.

El efecto del gen *CYP52M1* alterado se sometió a ensayo mediante la evaluación de tres transformantes seleccionados al azar para su producción de soforolípidos en medio líquido. CDW y la unidad de formación de colonias eran similares en comparación con la cepa de tipo salvaje, lo que indica que la alteración del gen no afectó el crecimiento celular.

20 Sin embargo, se observaron diferencias claras con respecto al consumo de glucosa. Mientras que durante la fase de crecimiento exponencial no se observaron diferencias, el tipo salvaje muestra un consumo mucho más rápido en la fase de crecimiento estacionario, cuando tiene lugar la síntesis de soforolípidos. Dado que la producción de soforolípidos exige un gran aporte de glucosa, la baja utilización de glucosa de las inactivaciones sugiere la ausencia de síntesis de soforolípidos. Además, durante todo el período de incubación, el aceite de colza flotaba sobre la superficie del medio de cultivo de las inactivaciones, lo que indica que tampoco se consume la fuente de carbono hidrofóbica necesaria para la producción de soforolípidos.

30 Finalmente, se verificó la producción de biotensioactivos mediante análisis de HPLC de muestras tomadas durante y al final del período de incubación. Mientras que había una clara producción para el tipo salvaje, no se pudieron detectar soforolípidos en el medio o las células de todas las tres inactivaciones (Figura 1). Los picos observados para el transformante son productos de degradación del aceite de colza no consumido: ácido oleico, el principal constituyente de los ácidos grasos de aceite de colza (60%) se detecta a los 36,2 minutos y también se identifica ácido linoleico (23%) (33,5 min).

35 Los resultados mencionados anteriormente demuestran que *CYP52M1* es la monooxigenasa del citocromo P450 responsable de la síntesis de ácidos grasos hidroxilados, que son esenciales para la producción de soforolípidos. La desactivación del gen inhibe la síntesis de soforolípidos, pero sorprendentemente no tiene efecto alguno sobre el crecimiento o la viabilidad de las células. Las cepas creadas son incapaces de producir soforolípidos y, por consiguiente, pueden utilizarse como huésped de producción. Dado que *C. bombicola* no está intoxicada por sustratos oleosos o hidrofóbicos y tiene vías metabólicas para metabolizar o convertir a los mismos, la cepa de inactivación se puede utilizar como una cepa de plataforma para la síntesis de otros productos oleaginosos tales como polihidroxialcanoatos. Además, las cepas se pueden utilizar para expresar enzimas P450 heterólogas con el fin de crear soforolípidos con una cola de ácido graso hecha a medida.

Ejemplo 2: Identificación del gen *UGTA1* de UDP-glucosiltransferasa, responsable del primer paso de glucosilación en la vía biosintética de soforolípidos de *Candida bombicola* ATCC 22214

Introducción

45 Esta cepa está inactivada en el gen *UGTA1* (número de acceso GenBank HM440973), que codifica la enzima responsable del primer paso de glucosilación en la vía biosintética de soforolípidos. Esta enzima transfiere glucosa de UDP-glucosa a un ácido graso hidroxilado, dando como resultado la producción de un glucolípidos.

En esta cepa, no se detecta producción de soforolípidos alguna, mientras que el crecimiento y la viabilidad celular son equiparables al tipo salvaje (Figura 2).

50 Materiales y Métodos

Cepas, plásmidos y medios de cultivo

Candida bombicola ATCC 22214 se utilizó para el aislamiento del gen *UGTA1* y *C. bombicola* G9 (Van Bogaert *et al.*, 2008b) para la creación del mutante por delección *ugtA1*. *Escherichia coli* DH5 α F' se utilizó para el mantenimiento del plásmido. Se cultivaron células de levadura en medio YPD que contenía 1% de extracto de levadura, 2% de peptona y 2% de dextrosa, medio SD que contenía 0,67% de base nitrogenada de levadura sin aminoácidos (Difco) y 2% de glucosa o medio 3C que contenía 10% de glucosa, 1% de extracto de levadura y 0,1% de urea. Los medios líquidos se incubaron a 30°C y 200 rpm. *E. coli* se cultivó en medio Luria Bertani (extracto de levadura Bacto 0,5%, Triptona Bacto 1%, NaCl 0,5%) que contenía 0,01% de ampicilina y se incubó a 37°C y 200 rpm. Los plásmidos se aislaron de *E. coli* DH5 α F' por medio del kit de aislamiento de plásmidos MiniPrep de Qiagen.

*Clonación del gen UGTA1 de C. bombicola**Diseño de los cebadores y secuencia de análisis*

El diseño de los cebadores, el análisis de la secuencia y el diseño de la estrategia se realizaron con el software Clone Manager Professional Suite (Versión 8.0). Los cebadores se solicitaron de Sigma y todos los plásmidos creados se enviaron para su secuenciación a VIB (Bélgica) o AGOWA (Alemania). Las búsquedas de homología se realizaron con el programa BLAST (Altschul *et al.*, 1997) contra las bases de datos disponibles en el sitio web del NCBI (<http://ncbi.nlm.nih.gov>).

Aislamiento de ADN genómico

C. bombicola ATCC22214 se hizo crecer durante la noche en medio 3C. Se separaron cantidades menores de soforolípidos extrayendo 500 μ l de muestras de cultivo con un volumen de acetato de etilo. La pared celular de la levadura se separó enzimáticamente incubando con 200 unidades de Enzima Lítica de Levadura (Sigma) en tampón SCE (sorbitol 1 M, acetato de sodio 0,1 M y EDTA 60 mM, pH 7,0) durante 90 min a 37°C en presencia de 3,75 μ l de mercaptoetanol. El ADN genómico se aisló de los protoplastos restantes por medio del kit de ADN genómico bacteriano GenElute™ (Sigma).

Caminar del Genoma

Para el aislamiento de la secuencia del gen *UGTA1* del ADN genómico de *C. bombicola* ATCC 22214, se utilizó el kit universal BD GenomeWalker™ (BD Biosciences). Cebadores específicos para el gen para PCR primaria y anidada se diseñaron basándose en la secuencia parcial de *UGTA1*, obtenida a partir de datos preliminares de secuenciación del genoma y se proporcionan en la Tabla 2. Las reacciones de amplificación se realizaron con el sistema de PCR Expand Long Template (Roche) siguiendo el protocolo tal como se describe en (De Maeseneire *et al.*, 2006). Los productos de amplificación por PCR se purifican a partir de la mezcla de PCR o mediante extracción en gel utilizando el kit de purificación Qiaquick PCR (Qiagen) o el kit de extracción de gel Qiaexll (Qiagen), respectivamente. Los fragmentos purificados se clonaron en pGEM-T® utilizando el sistema de vectores pGEM-T® (Promega) y los plásmidos resultantes se utilizaron para la transformación de *E. coli* DH5 α F' según Sambrook (Sambrook *et al.*, 2001). Las colonias correctas se cultivaron en LB líquido para el posterior aislamiento y secuenciación del plásmido.

Clonación del gen UGTA1 completo

La secuencia completa del gen *UGTA1* se amplificó a partir del ADN genómico de *C. bombicola* ATCC 22214 por medio del kit High Fidelity PCR Master (Roche). El fragmento de 2566 pb obtenido se purificó y clonó en pGEM-T® tal como se describió anteriormente. El plásmido resultante se denominó pGugtA1 Tot.

Creación del casete de inactivación UGTA1

El procedimiento que conduce al casete de inactivación para el gen *UGTA1* de *C. bombicola* se basa en una inserción mediada por enzima de restricción del marcador seleccionable URA3 (Van Bogaert *et al.*, 2007) entre regiones de homología con los extremos 5' y 3' del gen *UGTA1*. Todas las reacciones PCR se realizaron con el sistema Roche High Fidelity. En un primer paso, las regiones 5' (A1P) y 3' (A1T) del gen *UGTA1* se amplificaron a partir del plásmido pGugtA1 Tot mediante los cebadores de restricción A1P RevNheI y A1T ForNheI en combinación con UDPGTA1 TotF y A1T Rev para amplificar las regiones *UGTA1* 5' (A1P) y *UGTA1* 3' (A1T), respectivamente (Tabla 2). Estos fragmentos de 853 pb y 1121 pb respectivamente se purificaron y se digirieron con la enzima de restricción NheI (New England Biosystems) de acuerdo con Sambrook (Sambrook *et al.*, 2001). Los fragmentos digeridos se purificaron a partir de las mezclas de restricción y se ligaron por medio de la T4 ADN ligasa (Fermentas). El producto de ligamiento se amplificó de nuevo y se purificó mediante extracción en gel. Posteriormente, este fragmento (A1PT) se clonó en pGEM-T®, el plásmido pGA1PT se obtuvo de *E. coli* y se verificó

mediante secuenciación tal como se describió anteriormente. Dado que pGEM-T® carece de un sitio de restricción *NheI*, el plásmido pGA1PT obtenido se puede linearizar únicamente en el sitio *NheI* introducido entre los fragmentos A1P y A1T, dando lugar a extremos adhesivos en ambos sitios. Por consiguiente, el marcador URA3 se amplifica a partir del plásmido pCbura3 (Van Bogaert *et al.*, 2007) haciendo uso de los cebadores de restricción Ura3 FbisNheI y Ura3 RbisNheI que contienen ambos extensiones del sitio de restricción *NheI* en sus extremos 5'. Después de la purificación, el fragmento obtenido se digirió con *NheI* y se ligó en el vector pGA1PT linearizado por medio de la T4 ADN ligasa (Fermentas). La mezcla de ligamiento se utilizó para la transformación de *E. coli* DH5α y los transformantes correctos se cultivaron en LB para el posterior aislamiento del plásmido. La integración del marcador URA3 se verificó mediante digestión de control con la enzima de restricción *AccI* (New England Biolabs). La secuencia completa del gen *URA3* en el plásmido pGKO_A1 obtenido se confirmó mediante secuenciación.

Creación de un mutante por delección de *ugtA1*

Se produjo un casete de inactivación lineal a partir del plásmido pGKO_A1 mediante PCR de alta fidelidad haciendo uso de los cebadores externos UDPGTA1 TotF y A1T Rev, ya que se ha demostrado que la linearización aumenta en gran medida la frecuencia de recombinación (Van Bogaert *et al.*, 2008b). El fragmento obtenido se purificó y se utilizaron 2,5 µg para la transformación del *ura3⁻* *C. bombicola* G9. Se utilizó el protocolo descrito para *Saccharomyces* (Gietz *et al.*, 1995) con ligeras modificaciones. Se utilizó una solución de LiAc 50 mM en lugar de 100 mM, la incubación de células con el casete antes de que se produjera el choque térmico durante 90 minutos en lugar de 30 y no se añadió DMSO. Después de la transformación, las células se sembraron en medio de agar SD y se incubaron a 30°C hasta que aparecieron las colonias transformantes.

Caracterización del mutante por delección de *ugtA1*

La correcta integración del casete en el genoma se verificó mediante PCR de colonias de levaduras utilizando cebadores que se reasocian fuera de los sitios de recombinación. Transformantes con el genotipo correcto fueron luego transferidos a placas de agar 3C y testados en cuanto a la producción de soforolípidos. Para eso, la colonia de levadura mutante se inoculó en medio líquido tal como se describe por Lang (Lang *et al.*, 2000) y se cultivó durante 48 horas antes de la adición de aceite de colza (30 g/L). *Candida bombicola* de tipo salvaje ATCC 22214 se cultivó de la misma manera y sirvió como referencia. Diez días después de la adición de la fuente de carbono hidrofóbica, la producción de soforolípidos se verificó extrayendo 1 ml de medio de cultivo con 400 µl de acetato de etilo técnico en presencia de 10 µl de ácido acético. Después de agitar en vórtice durante 5 minutos, se diluyeron 300 µl de la fase de disolvente en 1,7 ml de etanol absoluto y se analizaron en HPLC-ELSD utilizando una HPLC Varian ProStar (Varian) equipada con una columna Chromolith® Performance RP-18e [100 mm (l) x 4.6 mm (D.I.)] (Merck) y se conectó a un Detector de Dispersión de la Luz Evaporativa (Alltech). Los compuestos se eluyeron por medio de un gradiente de acetonitrilo/ácido acético (0,5% en agua) (5/95 a 95/5 en 40 min) bajo un caudal constante de 1 ml/min. La temperatura de la columna se ajustó a 30°C.

Función biocatalítica de la glucosiltransferasa *UgtA1*

Preparación de lisados celulares

La levadura de tipo salvaje y el mutante por delección de *ugtA1* A113 se inocularon del medio de agar 3C en 5 ml de medio de Lang y se cultivaron durante la noche a 30°C y 200 rpm. Con este precultivo se inocularon 50 ml de medio de Lang fresco con una DO inicial de 0,2 y se incubaron de la misma manera hasta que se lisaron las células. Las células se recogieron por centrifugación a 3000 rpm y 4°C con una centrifuga de cubeta oscilante y se lavó con 10 ml de agua destilada. Después, el sedimento se resuspendió en tampón de lisis de pH 7,7 que contenía KH₂PO₄ 50 mM, 5% de glicerol, MgCl₂ 0,5 mM, DTT 0,5 mM y PMSF 1 mM a DO₁₀₀. Se añadió un volumen igual de perlas de vidrio lavadas con ácido (150-212 µm de diámetro, Sigma) y las células se rompieron agitando en vórtice durante 15 minutos con intervalos de 30 segundos en hielo. Las fracciones de proteína solubles se utilizaron para ensayos enzimáticos después de la centrifugación del lisado crudo a 3000 rpm a 4°C. La concentración de proteína en el lisado se determinó por medio del kit de ensayo de proteínas BCA™ (Pierce). Las soluciones de proteína se almacenaron a 4°C.

Ensayos enzimáticos

La UDP-glucosa se obtuvo de Sigma, el ácido 17-hidroxil-octadecenoico y el glucolípido se obtuvieron a partir de soforolípidos tal como se describió anteriormente (Saerens *et al.*, 2009). Todas las soluciones de sustrato se prepararon recientemente en KH₂PO₄ 50 mM, pH 7.7. Los ensayos enzimáticos contenían UDP-glucosa 2 mM, aceptor 2 mM y 200 µl de solución de proteína fresca en un volumen total de 250 µl. Para las reacciones en blanco, tampón fosfato reemplazó al donador, al aceptor o a la solución de proteínas. Para cada uno de los ensayos también se realizó una reacción en blanco con 200 µl de tampón. Todas las reacciones enzimáticas se incubaron a 30°C durante 3 horas. Las reacciones se detuvieron mediante la adición de 200 µl de HCl (2N) y los glicolípidos se extrajeron con 800 µl de dietiléter/acetato de etilo (1/1). A partir de la fase de disolvente se recuperaron 700 µl, se

evaporaron a sequedad y se redisolieron en 300 µl de etanol absoluto antes del análisis por HPLC tal como se describió anteriormente.

Resultados y Discusión

Aislamiento de la secuencia del gen UGTA1

5 Los datos preliminares de secuenciación del genoma proporcionaron parte de una glucosiltransferasa putativa con homología a glicosiltransferasas de la familia MGT y otras UDP-glucuronosilo/UDP-glucosiltransferasas disponibles en las bases de datos del NCBI. Los alineamientos ilustraron que este ORF parcial contenía un codón de inicio apropiado, pero carecía de un codón de parada. Con el fin de aislar el ORF completo, se diseñaron cebadores para el posterior caminar del genoma aguas abajo de este gen putativo. De esta forma, se podrían obtener otros 1600 pb que después del ensamblaje con la secuencia primaria dieron como resultado un fragmento de secuencia bruta total de 2566 pb. Este fragmento contenía un marco de lectura abierto completo de 1392 pb que codifica un gen putativo de UDP-glucuronosilo/UDP-glucosiltransferasa de 463 aminoácidos que los autores de la invención denominaron UgtA1.

Clonación y análisis de la secuencia del gen UGTA1 completo

15 Para aislar el gen *UGTA1* completo del genoma de *Candida bombicola*, el fragmento de 2566 pb se amplificó mediante PCR de alta fidelidad en ADN genómico recién aislado. La región codificante con 233 pb y 153 pb de las regiones aguas arriba y aguas abajo, respectivamente, se da en la Figura 3. El marco de lectura abierto codifica una glicosiltransferasa putativa de 463 aminoácidos con una masa molecular estimada de 50,5 kDa y un pI estimado de 5,45.

20 Aunque la mayoría de los promotores de levadura carecen de una caja TATA clara, una secuencia consenso similar a TATA, TATA(A/T)A(A/T)(A/G), rica en A/T, podría identificarse 50-200 pb aguas arriba de genes de *Saccharomyces* altamente regulados o inducidos por estrés (Basehoar *et al.*, 2004). Si existieran mecanismos reguladores equiparables para la activación transcripcional de genes de control interno e inducidos por estrés en *C. bombicola*, y dado que la producción de soforolípidos está ligada a las condiciones limitantes del nitrógeno (Davila *et al.*, 1992) se podría esperar un consenso similar a TATA en genes implicados en la producción de soforolípidos por esta levadura. Un posible elemento promotor con alta homología con el consenso de *Saccharomyces* podría identificarse 415 pb aguas arriba del codón de inicio *UGTA1* (resultados no mostrados), mientras que una segunda región rica en A/T se encuentra más cerca del codón de inicio (Figura 3). También se encuentran varias secuencias de poliadenilación posibles, pero no se pueden distinguir elementos de señalización específicos implicados en la formación del extremo 3' de ARNm tal como se describe para la levadura (Guo *et al.*, 1996).

Otra característica típica de los genes regulados por el metabolismo del nitrógeno es la presencia de secuencias reguladoras similares a GATA en sus regiones aguas arriba. Dichos motivos GATA son reconocidos por factores de transcripción de tipo GATA que son activados fuertemente por agotamiento de nitrógeno (Magasanik *et al.*, 2002). Un ejemplo típico de un factor de transcripción de tipo GATA de este tipo es el gen *AreA* de *Aspergillus nidulans* (Marzluf *et al.*, 1997) y varios elementos similares a GATA se reconocen en genes de sintetización de glicolípidos de *Ustilago maydis* [Hewald *et al.*, 2005]. En la región 5' aguas arriba de la región codificante de *UGTA1* se encuentran varios elementos posibles similares a GATA.

Secuencia de homología con otras UDP-glucosiltransferasas

40 El análisis de la secuencia de la proteína UgtA1 a través de la Base de Datos de Dominio Conservado disponible en el sitio web NCBI (Marchler-Bauer *et al.*, 2009) muestra que esta proteína pertenece a la familia GT1 de glicosiltransferasas (E.C. 2.4.1), una familia poliespecífica que alberga hasta la fecha más de 3200 glicosiltransferasas diferentes, de las cuales la mayoría son enzimas inversoras utilizando azúcares activados por nucleótidos o fosfatos de azúcar como moléculas donantes (Campbell *et al.*, 1997, Coutinho *et al.*, 2003). Las estructuras de las proteínas GT1 muestran la típica topología de tipo GT-B, que se refleja en un dominio conservado (cl 10013) que también se encuentra en la secuencia UgtA1 y, por lo tanto, esta familia está recogida en la superfamilia de glicosiltransferasas tipo GT-B (Teichmann *et al.*, 2007). Otras glicosiltransferasas tales como la RhIB de *P. aeruginosa* (Ochsner *et al.*, 1994a) y las proteínas tanto Emt1 como Hgt1/Ugt1 de *U. maydis* (Hewald *et al.*, 2005, Teichmann *et al.*, 2007), que catalizan reacciones de glicosilación análogas en las vías biosintéticas de glicolípidos de estos organismos pertenecen a la misma familia GT1. Sorprendentemente, la secuencia total de proteínas de UgtA1 muestra sólo una baja similitud con estas proteínas (34% de similitud con *Ustilago* Hgt1/Ugt1 y 38% de similitud con *P. aeruginosa* RhIB) tal como se deriva de un alineamiento global por pares utilizando el Software BioEdit (Hall, 1999). Además, una búsqueda de homología Blastp ilustra que la proteína UgtA1 muestra una pobre similitud con cualquier otra secuencia de proteína. La homología más alta se encuentra con las UDP-glucosa/UDP-glucuronosiltransferasas bacterianas, de las cuales la mayoría pertenece a la subfamilia MGT (Tabla 3), aunque la identidad de la secuencia es aún baja. Las proteínas de la subfamilia MGT están implicadas en la

biosíntesis o inactivación de antibióticos macrólidos y muestran un dominio conservado (TIGR01426/cd 03784) que también se encuentra en la secuencia de UgtA1, indicando la transferencia de un resto de azúcar desde un residuo de azúcar activado de nucleótido a un aceptor complejo tal como el núcleo heptapeptídico de un antibiótico. Estos hallazgos indican que la proteína UgtA1 está muy probablemente implicada en la síntesis de un metabolito complejo.

5 Creación y caracterización de un mutante por delección *ugtA1*

El casete de inactivación utilizado es un fragmento lineal basado en una inserción mediada por enzima de restricción del marcador seleccionable URA3 entre 853 pb y 1024 pb de homología con las regiones 5' y 3' del gen *UGTA1*, respectivamente. Este casete se utiliza para la transformación de la cepa G9 deficiente en *ura3*. Sólo cuando se ha producido una recombinación homóloga y el casete está integrado en el genoma, se recuperará la funcionalidad de URA3 y los transformantes podrán crecer en medio SD, sin uracilo. Cinco días después de la transformación de la cepa G9, se obtuvieron 40 colonias en las placas selectivas, de las cuales 17 se sometieron a una PCR de colonia de levadura para verificar su genotipo. Sólo una colonia mostró el genotipo correcto y se la conoce como *C. bombicola* A113. Para los otros 16 transformantes seleccionados, la complementación de *ura3* se habrá producido por recombinación sólo entre el marcador y el alelo mutante *ura3* de la cepa G9 o por recombinación unilateral o ilegítima.

Para controlar la implicación de la glucosiltransferasa UgtA1 en la producción de soforolípidos, el mutante de *C. bombicola* A113 obtenido se cultivó en medio de producción tal como se describió anteriormente y se comparó la producción de soforolípidos con la levadura de tipo salvaje. La Figura 4 muestra los cromatogramas de HPLC de extractos de cultivo tanto de *C. bombicola* A113 como de levadura salvaje cuando se utilizó aceite de colza como fuente de carbono hidrofóbica. Sólo la cepa de levadura de tipo salvaje produce soforolípidos, mientras que el mutante por delección *ugtA1* creció bien, pero perdió por completo su capacidad de producir cualquier glucolípido. Esto indica claramente que la glucosiltransferasa UgtA1 tiene una función clave en la producción de soforolípidos por parte de la levadura. Dado que el mutante muestra un crecimiento equiparable al de la levadura de tipo salvaje, es poco probable que la glucosiltransferasa UgtA1 tenga otra función en el metabolismo primario.

25 Función biocatalítica de la glucosiltransferasa *UgtA1*

Con el fin de vincular la glucosiltransferasa UgtA1 a un paso específico en la vía biosintética de soforolípidos, se midieron las actividades de glucosiltransferasa en lisados celulares del mutante A113 y se compararon con las de la levadura de tipo salvaje. Dado que la producción de soforolípidos está vinculada a las condiciones de inanición de nitrógeno, es muy probable que la expresión de genes implicados en la vía biosintética sólo comience en la fase estacionaria temprana. Para determinar el momento mejor para recoger células para la extracción de proteínas y las pruebas de actividad posteriores, *C. bombicola* de tipo salvaje se cultivó en medio de Lang y se determinaron las actividades de glucosiltransferasa en muestras de cultivo de 5 ml tomadas en diferentes momentos. Para la detección de la primera actividad de glucosiltransferasa (GTI), se utilizaron UDP-glucosa y ácido 17-hidroxiocetadecenoico como donante y aceptor de glucosilo, respectivamente. Para la detección de la segunda actividad (GTII), el ácido graso fue reemplazado por el glucolípido. En la levadura de tipo salvaje ambas actividades de la glucosiltransferasa se detectaron a partir de la fase exponencial tardía y aumentaron aún más hasta un máximo en la fase estacionaria temprana. La adición de aceite de colza al medio de cultivo después de 48 horas no dio como resultado un aumento de las actividades, mientras que no se pudo detectar actividad significativa alguna cuando las células se cultivaron en medios de levadura estándar tales como YPD (resultados no mostrados). Por lo tanto, la mejor configuración para los ensayos enzimáticos es recolectar células al cabo de 42 a 48 horas de incubación en medio de Lang sin la adición de fuente de carbono hidrofóbica.

Bajo estas condiciones, se prepararon lisados de levadura tanto de tipo salvaje y como mutante A113 y se utilizaron fracciones de proteínas solubles para los mismos ensayos de glucosiltransferasa. Los resultados se presentan en las Figuras 5 y 6. De la Figura 5 resulta claro que las actividades tanto de GTI como de GTII están presentes en el lisado celular de la levadura de tipo salvaje, ya que tanto los glucolípidos (actividad de GTI) como los soforolípidos (actividad de GTII) se detectan como productos. Ninguna de estas actividades se observaron cuando el lisado celular, la UDP-glucosa o el ácido graso se omitieron de la reacción. En contraste con el lisado de tipo salvaje, el lisado celular del mutante A113 no muestra producto de glucosilación del ácido graso alguno, lo cual indica que la proteína UgtA1 es al menos responsable de la primera reacción de glucosilación en la vía biosintética de soforolípidos. Las señales a los 27 y 29 minutos en los cromatogramas de los ensayos de tipo salvaje provienen la síntesis *de novo* de soforolípidos durante el crecimiento en medio de Lang. Estos picos están ausentes en los cromatogramas del mutante A113, lo que confirma que la síntesis *de novo* en este mutante está completamente bloqueada. Con el fin de verificar si la segunda etapa de glucosilación (GTII) se ve afectada por la alteración de la proteína UgtA1, el ensayo se repitió con el glucolípido como el aceptor. De la Figura 6 se puede ver claramente que la segunda actividad de glucosilación todavía está presente en el lisado A113 y que esta actividad es equiparable a la del lisado de tipo salvaje. De este modo, los autores de la invención proporcionan evidencia de que la producción de soforolípidos en *C. bombicola* implica dos glucosiltransferasas diferentes que transfieren glucosa desde UDP-glucosa a sus respectivos sustratos aceptores de una manera escalonada pero independiente. Demostraron que la

glucosiltransferasa UgtA1 aislada aquí es responsable de la primera reacción de glucosilación y la alteración del gen no tiene influencia alguna en la segunda etapa de glucosilación. Esta segunda glucosiltransferasa parece ser altamente específica a su propio sustrato aceptor, ya que no se observa formación de glucolípidos cuando los ácidos grasos están presentes como aceptor en los ensayos A113 (Figura 5). Basados en estos hallazgos, los autores de la invención creen que la producción de soforolípidos en *C. bogoriensis* se basa en una vía escalonada equiparable y que es dudoso que exista una proteína multifuncional que acepte tanto ácidos grasos como glucolípidos en su centro catalítico.

Conclusión

Aquí, los autores de la invención identificaron el gen *UGTA1* con una función clara en la producción de soforolípidos por *Candida bombicola*. El gen *UGTA1* codifica una glucosiltransferasa de 463 aminoácidos y un peso molecular estimado de 50,5 kDa. La proteína se puede clasificar dentro de la familia GT1 de glucosiltransferasas (EC 2.4.1.x). Mediante la creación de un mutante por delección *ugtA1* pudieron identificar que esta proteína tiene una función clara en la producción de soforolípidos y los ensayos enzimáticos en lisados celulares proporcionaron evidencia adicional de que la glucosiltransferasa UgtA1 está catalizando la primera etapa de glucosilación en la vía biosintética de soforolípidos de *C. bombicola*. Además, demostraron que la segunda reacción de glucosilación es catalizada por otra glucosiltransferasa independiente y que esta transferasa es altamente específica para los glucolípidos como sustrato, ya que la alteración de la actividad de UgtA1 da como resultado una pérdida completa de la producción de soforolípidos tanto *in vivo* como *in vitro*. La cepa A113 creada aquí ya no es capaz de convertir los ácidos grasos ω o ω -1 hidroxilados en glucolípidos y, por lo tanto, es útil en la producción de compuestos basados en hidroxiácidos grasos. La falta de actividad de GTI conducirá a una agrupación intracelular de ácidos grasos hidroxilados que puede utilizarse como bloques de construcción para otros compuestos oleaginosos.

Ejemplo 3: Identificación del gen UDP-glucosiltransferasa *UGTB1* responsable de la segunda etapa de glucosilación en la vía biosintética de soforolípidos de *Candida bombicola* ATCC 22214

Introducción

Esta cepa se inactivó en el gen *UGTB1* (número de acceso de GenBank HM440974), que codifica la enzima responsable de la segunda etapa de glucosilación en la vía biosintética de soforolípidos. La enzima transfiere glucosa de UDP-glucosa a un glucolípido, lo que resulta en la producción de un soforolípido.

En esta cepa, no se detecta producción de soforolípidos, mientras que el crecimiento y la viabilidad celular son equiparables a los del tipo salvaje (Figura 2).

30 Materiales y Métodos

Cepas, plásmidos y medios de cultivo

Candida bombicola ATCC 22214 se utilizó para el aislamiento del gen *UGTB1* y *C. bombicola* G9 (derivado de ATCC 22214, véase Van Bogaert *et al.*, 2008b) para la creación del mutante por delección Δ *ugtB1*. *Escherichia coli* DH5 α F' se utilizó para el mantenimiento del plásmido. Se cultivaron células de levadura en medio YPD que contenía 1% de extracto de levadura, 2% de peptona y 2% de dextrosa, medio SD que contenía 0,67% de base nitrogenada de levadura sin aminoácidos (Difco) y 2% de glucosa, o medio 3C que contenía 10% de glucosa, 1% de extracto de levadura y 0,1% de urea. Los medios líquidos se incubaron a 30°C y 200 rpm. *E. coli* se cultivó en medio Luria Bertani (extracto de levadura Bacto 0,5%, Triptona Bacto 1%, NaCl 0,5%) que contenía ampicilina 0,01% y se incubó a 37°C y 200 rpm.

Los plásmidos se aislaron de *E. coli* DH5 α F' por medio del kit de Aislamiento de Plásmidos MiniPrep de Qiagen y se secuenciaron en AGOWA (Alemania).

Diseño del cebador y análisis de la secuencia

El diseño del cebador, el análisis de la secuencia y el diseño de la estrategia se realizaron con el software Clone Manager Professional Suite (Versión 8.0). Los cebadores se solicitaron en Sigma.

45 Clonación del gen *UGTB1* completo

Para el aislamiento de ADN genómico, *C. bombicola* ATCC22214 se cultivó durante la noche en medio 3C. Se separaron cantidades menores de soforolípidos extrayendo 500 μ l de muestras de cultivo con un volumen de acetato de etilo. La pared celular se separó enzimáticamente incubando con 200 unidades de Enzima Lítica de Levadura (Sigma) en tampón SCE (sorbitol 1 M, acetato de sodio 0,1 M y EDTA 60 mM, pH 7,0) durante 90 min a 37°C en

presencia de β -mercaptoetanol al 0,75%. El ADN genómico se aisló de los protoplastos restantes por medio del kit de ADN genómico bacteriano GenElute™ (Sigma).

5 La secuencia completa del gen *UGTB1* se amplificó a partir del ADN genómico de *C. bombicola* ATCC 22214 por medio del kit High Fidelity PCR Master (Roche) y los cebadores GTII-472For y GTII + 239Rev (Tabla 4). El fragmento obtenido se purificó por medio del kit de purificación PCR Qiaquick (Qiagen) y se clonó en pGEM-T® haciendo uso del sistema de vector pGEM-T® (Promega). Esto condujo al plásmido pGugtB1 Tot que luego se utilizó para la transformación de *E. coli* DH5 α F' de acuerdo con (Sambrook y Russell, 2001). Los transformantes correctos se aislaron después de la PCR de colonias y se cultivaron en LB líquido para el posterior aislamiento del plásmido.

Creación del casete de inactivación *UGTB1*

10 El casete de inactivación para el gen *UGTB1* de *C. bombicola* se basa en la integración del marcador seleccionable URA3 entre regiones de homología con los extremos 5' y 3' del gen *UGTB1*. En una primera etapa, el plásmido pGugtB1Tot se lineariza mediante un doble digestión con enzimas de corte único *AvaI* y *KasI* (New England Biolabs) de acuerdo con Sambrook (Sambrook y Russell, 2001). De esta forma, se separa un fragmento de 100 pb (desde pb 950-1049 de la inserción) de la secuencia codificante de *UgtB1*, dejando fragmentos pegajosos en ambos extremos
15 del vector linearizado. El marcador de selección de URA3 se obtuvo después de PCR de alta fidelidad con el sistema de PCR de alta fidelidad PfuUltra (Stratagene) en el plásmido pCbura3 (Van Bogaert *et al.*, 2008). Los cebadores utilizados fueron *ura3infugtB1F* y *ura3infugtB1R* (Tabla 4). Estos cebadores contienen 15 pb de homología con, respectivamente, el extremo 3' y 5' del plásmido linearizado pGugtB1 Tot. El amplicón obtenido se utilizó directamente para la clonación en el plásmido linearizado haciendo uso del kit de clonación In-Fusion Dry
20 Down PCR (Clontech). Esto condujo al plásmido pGKO_ugtB1 que ahora contiene el gen *URA3* completo (2043 pb) de *C. bombicola* ATCC22214, flanqueado por 949 pb y 961 pb de regiones homólogas a las secuencias 5' y 3' de *UGTB1*, respectivamente. La clonación y la posterior transformación de células de *E. coli* competentes de Fusion Blue, para el mantenimiento del plásmido, se realizaron tal como se describe en el manual del kit. Transformantes de *E. coli* se analizaron mediante PCR de colonias para la inserción del marcador URA3 en el plásmido y
25 posteriormente se cultivaron en LB para el aislamiento del plásmido. La secuencia del gen *URA3* se confirmó por secuenciación.

Creación de un mutante por delección Δ ugtB1

30 Se produjo un casete de inactivación lineal a partir del plásmido pGKO_B1 mediante PfuUltra High Fidelity PCR (Stratagene) haciendo uso de los cebadores GTII -472F y GTII +239R (Tabla 4) ya que la frecuencia de recombinación se incrementa fuertemente utilizando fragmentos lineales (Van Bogaert *et al.*, 2008). El fragmento obtenido se purificó en columna y se utilizó 1 μ g para la transformación del *ura3* de *C. bombicola* G9.

35 Para la transformación de *C. bombicola* G9 se utilizó el protocolo tal como se describe para *Saccharomyces* (Gietz y Schiestl, 1995) con algunas ligeras modificaciones. Se utilizó una solución de LiAc 50 mM en lugar de 100 mM, la incubación de células con el casete antes de producirse el choque de calor durante 90 minutos en lugar de 30 y no se añadió DMSO. Después de la transformación, las células se sembraron en medio de agar SD y se incubaron a 30°C hasta que aparecieron colonias transformantes.

Caracterización del mutante por delección Δ ugtB1

40 El genotipo de los mutantes obtenidos se controló mediante PCR de colonias de levaduras utilizando los cebadores KOugtB1 CtrlF y KOugtB1 CtrlR (Tabla 4) que se reasocian aguas arriba del sitio de recombinación del lado izquierdo y el marcador URA3, respectivamente. Para estudiar el fenotipo, se cultivaron colonias de levaduras mutantes en medio líquido tal como se describe por Lang *et al.* (2000) durante 48 horas antes de la adición de aceite de colza (37,5 g/L). *Candida bombicola* de tipo salvaje ATCC 22214 sirvió de referencia. Siete días después de la adición de la fuente de carbono hidrofóbica, se verificó la producción de soforolípidos extrayendo 1 ml de medio de cultivo con 400 μ l de acetato de etilo técnico en presencia de 10 μ l de ácido acético.

45 Después de agitar en vórtice durante 5 minutos, se diluyeron 300 μ l de la fase de disolvente en 1,7 ml de etanol absoluto y se analizaron en HPLC-ELSD utilizando un Varian ProStar HPLC (Varian) equipado con una columna Chromolith® Performance RP-18e [100 mm (l) x 4,6 mm (D.I.)] (Merck) y conectado a un Detector de Dispersión de la Luz Evaporativa (Alltech). Los compuestos se eluyeron por medio de un gradiente de acetonitrilo/ácido acético (0,5% en agua) (5/95 a 95/5 en 40 min) bajo caudal constante de 1 ml/min. La temperatura de la columna se ajustó a
50 30°C. Para verificar las masas moleculares de los glucolípidos producidos, se analizaron las mismas muestras bajo las mismas condiciones en un sistema de HPLC Shimadzu LC-10-AD conectado a un espectrómetro de masas cuadrupolar (Waters). Las moléculas se identificaron por sus masas moleculares nativas después de ESI (ionización por atomización de electrones) sin colisión.

Función biocatalítica de la glucosiltransferasa *UgtB1*

C. *bombicola* ATCC 22214 y el mutante por delección *ΔugtB1* B11 se inocularon del medio de agar 3C en 5 ml de medio de Lang y se cultivaron durante la noche a 30°C y 200 rpm. Con este precultivo, se inocularon 50 ml de medio de Lang fresco con una DO inicial de 0,2 y se incubaron de la misma manera durante 60 horas. Las células se recogieron por centrifugación a 5600 g y 4°C con una centrifuga de cubeta oscilante y se lavó con 10 ml de agua destilada. Después, el sedimento se resuspendió en tampón de lisis de pH 7,7 que contenía KH₂PO₄ 50 mM, 5% de glicerol, MgCl₂ 0,5 mM, DTT 0,5 mM y PMSF 1 mM a DO₁₀₀. Se añadió un volumen igual de perlas de vidrio lavadas con ácido (150-212 μm de diámetro, Sigma) y las células se rompieron agitando en vórtice durante 15 minutos con intervalos de 30 segundos en hielo. Las fracciones de proteína soluble se utilizaron para ensayos enzimáticos después de la centrifugación del lisado crudo a 5600 g a 4°C. La concentración de proteínas en el lisado se determinó por medio del Kit de Ensayo de Proteínas BCA™ (Pierce).

La UDP-glucosa se obtuvo de Sigma, el ácido 17-hidroxi-octadecenoico y el glucolípido se obtuvieron a partir de soforolípidos tal como se describió anteriormente (Saerens *et al.*, 2009). Todas las soluciones de sustrato se prepararon recientemente en KH₂PO₄ 50 mM, pH 7.7. Los ensayos enzimáticos contenían UDP-glucosa 2 mM, aceptor 2 mM y 200 μl de solución de proteína reciente en un volumen total de 250 μl. Para las reacciones en blanco, el tampón reemplazó ya sea UDP-glucosa, aceptor o solución de proteínas. Todas las reacciones enzimáticas se incubaron a 30°C durante 3 horas. Las reacciones se detuvieron mediante la adición de 200 μl de HCl (2 N) y los glicolípidos se extrajeron con 800 μl de dietiléter/acetato de etilo (1/1) de acuerdo con Breithaupt y Light (1982). A partir de la fase de disolvente se recuperaron 700 μl, se evaporaron a sequedad y se redisolvieron en 300 μl de etanol absoluto antes del análisis por HPLC tal como se describió anteriormente. Las áreas pico se convirtieron en concentraciones de producto y la actividad enzimática se expresó en μM/mg min.

Resultados y discusión

Clonación y análisis de secuencia del gen *UGTB1* completo

Los datos preliminares de secuenciación del genoma de *Candida bombicola* ATCC 22214 revelaron un marco de lectura abierto (ORF) putativo de 1299 pb con homología con un gran número de UDP-glucosiltransferasas/glicuroniltransferasas de origen microbiano, en su mayoría hipotéticas, tal como se muestra mediante una búsqueda de homología BLASTx (Altschul *et al.*, 1997). Los cebadores se diseñaron a 472 pb aguas arriba y 239 pb aguas abajo, respectivamente, del codón de inicio y parada del gen putativo y la secuencia de gen completa se aisló y se clonó en el plásmido pGugtB1 Tot. El gen, al que se alude como *UGTB1* (Figura 7), codifica una proteína putativa de 432 aminoácidos con una masa molecular estimada de 46,2 kDa y un pI estimado de 4,98.

Aunque la mayoría de los promotores de *Saccharomyces* carecen de una caja TATA clara, se podía identificar una secuencia consenso TATA(A/T)A(A/T)(A/G) similar a TATA, rica en A/T, 50-200 pb aguas arriba de genes altamente regulados o inducidos por estrés (Basehoar *et al.*, 2004). Si existieran mecanismos reguladores diferentes equiparables para la activación transcripcional de genes de control interno e inducidos (por estrés) en *C. bombicola*, se podría esperar un consenso similar a TATA equiparable en los genes implicados en la producción de soforolípidos por esta levadura también, dado que la síntesis de soforolípidos se observa en condiciones limitantes de nitrógeno (Davila *et al.*, 1992). Se puede identificar un elemento similar a TATA claro correspondiente a la secuencia de consenso de *Saccharomyces* 50 pb aguas arriba del codón de inicio de *UGTB1* (Figura 7). Por el contrario, no se puede encontrar una señal de poliadenilación clara o cualquier otro elemento de señal que pueda estar implicado en la formación del extremo 3' del ARNm de levadura (Guo y Sherman, 1996). Dado que la síntesis de soforolípidos está estrechamente vinculada a la limitación de nitrógeno, cabría esperar secuencias reguladoras similares a GATA en las regiones aguas arriba de los genes implicados en esta vía. Dichos motivos GATA son reconocidos por factores de transcripción de tipo GATA que se activan fuertemente por agotamiento de nitrógeno (Magasanik y Kaiser, 2002). Un ejemplo típico de dicho factor de transcripción de tipo GATA es el gen *AreA* de *Aspergillus nidulans* (Marzluf *et al.*, 1997) y se reconocen varios elementos similares a GATA en genes que sintetizan glicolípidos de *Ustilago maydis* (Hewald *et al.*, 2005). En la región 5' aguas arriba del gen *UGTB1* se pueden encontrar dos elementos posibles similares a GATA (Figura 7). Tomados en conjunto, es muy probable que la expresión del gen *UGTB1* aislado esté bajo el control de un sistema de regulación relacionado con el metabolismo del nitrógeno.

Homología de la secuencia con otras UDP-glucosiltransferasas

El análisis de la secuencia de la proteína UgtB1 frente a la Base de Datos de Dominio Conservado (CDD) (Marchler-Bauer *et al.*, 2009) ilustra que la proteína pertenece a la familia de glicosiltransferasas GT1 poliespecífica (Campbell *et al.*, 1997; Coutinho *et al.*, 2003) que se caracteriza por un dominio conservado GT1_Gtf_like (CDD/cd03784) y que incluye, entre otros, un grupo de glicosiltransferasas homólogas implicadas en las fases finales de la biosíntesis de vancomicina y cloroeremomicina. Estas proteínas transfieren restos de azúcar desde un donante de NDP-azúcar activado al núcleo heptapeptídico del antibiótico. Los 14 aminoácidos catalíticos conservados se encuentran en la secuencia UgtB1 (Figura 8). A partir del alineamiento con otras proteínas GT1_Gtf_like, se supone que el residuo H18 corresponde al único aminoácido conservado del bolsillo de unión del sustrato aceptor, mientras que se supone

que los residuos G17, T272, N338, G340, G342 y G343 constituyen los 6 aminoácidos conservados desde el sitio de enlace UDP del dominio.

5 Dado que todos los miembros de la familia GT1 muestran la topología de tipo GT-B, esta familia se recoge en la amplia superfamilia de glicosiltransferasas tipo GTB (Breton *et al.*, 2006), caracterizada por otro dominio conservado (CDD/cl10013) presente en la secuencia de UgtB1. La homología de secuencia con glicosiltransferasas relacionadas con antibióticos se confirma mediante el reconocimiento de un multidominio conservado MGT (macrólido glicosiltransferasa) (CDD/TIGR01426) y los resultados obtenidos de una búsqueda de homología BLASTp frente a todas las secuencias de proteínas no redundantes disponibles en las bases de datos del NCBI. Numerosos éxitos con homología moderada parecen ser glicosiltransferasas hipotéticas de origen tanto bacteriano como fúngico que surgen del gran número de proyectos de secuenciación del genoma. La homología de secuencia más alta del 57% (38% de identidad de secuencia) se encuentra para una proteína hipotética del hongo patógeno de plantas *Sclerotinia sclerotiorum* 1980 (gb/EDN94128). Si bien no se puede atribuir ninguna función bioquímica a la mayoría de las proteínas coincidentes, un número limitado se asocia con la producción secundaria de metabolitos y más específicamente con la síntesis de antibióticos (Tabla 5). Estos hallazgos sugieren que la proteína UgtB1 está muy probablemente implicada en la biosíntesis de un metabolito secundario complejo.

20 Dado que ramnosiltransferasa RhIB (gb/L28170) de *Pseudomonas aeruginosa* (Ochsner *et al.*, 1994) y tanto eritritol β -manosiltransferasa Emt1 (gb/XP_400732) como hidroxipalmitato glucosiltransferasa Hgt1 (gb/EAK87174) de *Ustilago maydis* (Hewald *et al.*, 2005; Teichmann *et al.*, 2007) catalizan etapas bioquímicas equiparables durante la síntesis de otros glicolípidos, y debido a que estas proteínas (todas EC 2.4.1.-) se clasifican dentro de la misma familia GT1 de glicosiltransferasas, se esperaba que UgtB1 mostrara una homología significativa de la secuencia con aquellos. Sorprendentemente, un alineamiento de pares global óptimo utilizando el software BioEdit (matrix blosum 62) (Hall *et al.*, 1999) muestra un 36% de similitud de secuencia (22% de identidad) con RhIB, un 33,5% (18%) con Emt1 y un 33% (19%) con Hgt1, valores que son incluso más bajos en comparación con aquellos para antibióticos que sintetizan proteínas de origen bacteriano (Tabla 5).

25 Con estas características, UgtB1 parece ser muy comparable a otra UDP-glucosiltransferasa a la que se alude como UgtA1 (gb/HM440973) que los autores de la invención aislaron de *C. bombicola* y a la que podría atribuirse la primera etapa de glucosilación en la biosíntesis de soforolípidos (resultados no mostrados). Por lo tanto, la homología de secuencia entre estas dos proteínas se verificó mediante otro alineamiento óptimo por pares global (matrix blosum 62). Ambas proteínas mostraron un 45,2% de identidad de secuencia y un 61% de similitud de secuencia (Figura 8). Además de los residuos conservados GT1_Gtf_like, que también están presentes en la secuencia UgtA1, varios otros aminoácidos son constitutivos. Los modelos de homología de la estructura que los autores de la invención crearon para UgtA1 y UgtB1 sugieren que estos residuos probablemente estén situados en la vecindad del centro catalítico o en bucles externos, por lo que sugieren su implicación en el reconocimiento y la orientación del sustrato.

35 *Creación y caracterización de un mutante por delección Δ ugtB1*

Se utilizó un casete de inactivación lineal para la transformación del *ura3⁻ Candida bombicola* G9, derivado de la levadura de tipo salvaje ATCC22214 (Van Bogaert *et al.*, 2008), y los transformantes se seleccionaron por complementación como resultado de la alteración del gen *UGTB1* mediante la inserción del marcador URA3. Un par de días después de la transformación, se obtuvieron varios transformantes en las placas selectivas y 28 se sometieron a una PCR de colonias para verificar su genotipo. Como consecuencia de los sitios de reasociación de cebadores, los eventos de doble cruzamiento o los cruzamientos simples del lado izquierdo conducirán a un amplicón. De los 28 transformantes, 13 proporcionaron el amplicón esperado. Los otros mutantes probablemente surgieron de eventos cruzados simples del lado derecho, recombinación ilegítima o de la recombinación del marcador URA3 con el alelo *ura3* de G9.

45 Tres transformantes correctos (B11, B14 y B19) se cultivaron posteriormente en medio de Lang líquido para investigar la influencia de la inactivación en la producción de soforolípidos. Se utilizó *C. bombicola* de tipo salvaje ATCC 22214 como referencia. El crecimiento de los transformantes pareció ser equiparable al de la levadura de tipo salvaje, lo que indica que el gen *UGTB1* no tiene contribución a ninguna vía en el metabolismo primario. Siete días después de la adición de aceite de colza, se verificó la producción de soforolípidos en los medios de cultivo. La Figura 9 muestra los resultados para la levadura de tipo salvaje y el mutante B11 (B14 y B19 dieron resultados idénticos).

55 En contraposición con la levadura de tipo salvaje, no se produjeron soforolípidos por los mutantes por delección Δ ugtB1, lo que indica que UgtB1 tiene una función clave en la producción de soforolípidos. En cambio, los mutantes produjeron glucolípidos, lo que indica que la primera glucosilación todavía se está realizando. El que UgtB1 está catalizando la segunda glucosilación se confirmó mediante ensayos enzimáticos en lisados celulares de la levadura de tipo salvaje ATCC 22214 y el mutante por delección Δ ugtB1 seleccionado al azar B11. Para la detección de la actividad de glucosiltransferasa I (GTI), se utilizaron UDP-glucosa y ácido 17-hidroxi-octadecenoico como donante y

5 aceptor, respectivamente. Para la detección de la actividad de glucosiltransferasa II (GTII), el ácido graso fue reemplazado por glucolípido como aceptor. La Figura 10 muestra los análisis de extractos de muestras después de un ensayo de GTI en los lisados de tipo salvaje y B11, respectivamente. En el lisado celular del tipo salvaje, la conversión del ácido graso en glucolípidos por la acción de la primera glucosiltransferasa es seguida por una
 10 segunda glucosilación que conduce a soforolípidos. Los picos a los 27 y 29 minutos corresponden a soforolípidos lactonizados diacetilados a partir de la síntesis *de novo* durante la incubación en medio de Lang. Estos soforolípidos *de novo* están ausentes en el cromatograma del mutante B11, confirmando la alteración de la vía. Dado que no se observa formación de soforolípidos de carácter ácido con el lisado de B11, mostrando una formación menor de glucolípidos solamente, se puede concluir que la segunda glucosilación está bloqueada en el mutante por delección.
 15 Esto se confirma mediante un ensayo GTII en los mismos lisados (Figura 11): cuando los glucolípidos se convierten en soforolípidos de carácter ácido con el lisado del tipo salvaje, no se observa conversión de glucolípidos con el lisado del mutante B11. El pequeño pico de ácidos grasos observado en estos cromatogramas tiene su origen en la preparación de glucolípidos (Saerens *et al.*, 2009).

Conclusión

15 El conocimiento de los mecanismos genéticos detrás de la producción de biotensioactivos por parte de diferentes microorganismos está aumentando, abriendo una nueva forma de aumentar los rendimientos mediante la creación de mutantes súper-productores y de esa manera hacer que los biotensioactivos sean económicamente más competitivos. Sin embargo, la genética detrás de la producción de soforolípidos, uno de los grupos más prometedores de biotensioactivos de glicolípidos, sigue sin estar clara. Aquí, los autores de la invención aislaron un gen de *C. bombicola* empleado industrialmente con una función clave en esta vía importante. El gen *UGTB1* codifica la UDP-glucosiltransferasa responsable de la segunda etapa de glucosilación tal como se demostró mediante la creación de un mutante por delección Δ *ugtB1* y posteriores ensayos enzimáticos *in vitro* con lisados celulares. Aquí, demostraron que dos UDP-glucosiltransferasas independientes, denominadas UgtA1 y UgtB1, actúan de forma escalonada. La presencia de dominios conservados en la secuencia de la proteína UgtB1 indica que la proteína pertenece a la familia GT1 de glicosiltransferasas. Sorprendentemente, la homología de secuencia con las glicosiltransferasas bacterianas implicadas en la síntesis de antibióticos es mayor que la homología con otras glicosiltransferasas conocidas hasta ahora por estar implicadas en la biosíntesis de glicolípidos microbianos. Esto podría explicarse por la distancia filogenética entre los diferentes organismos. *Ustilago maydis*, por ejemplo, es un basidiomiceto dimórfico y la agrupación de genes que sintetizan glucolípidos en *Ustilago* sugiere que estos han surgido de una transferencia horizontal de genes (Hewald *et al.*, 2006). En general, sin embargo, la homología de la secuencia entre el UgtB1 y otras glicosiltransferasas de la familia GT1 es baja, lo que indica que el UgtB1 podría considerarse como una nueva enzima dentro de esta familia.

Ejemplo 4: ejemplos de componentes producidos

Ejemplo 4.1 Expresión de proteína heteróloga

35 Una cepa derivada de *Candida bombicola* ATCC 22214, es decir, una cepa inactivada en el gen *cyp52M1* (número de acceso de GenBank EU552419), se utilizó para la producción de proteínas. En esta cepa, apenas se detecta producción de soforolípidos, mientras que el crecimiento y la viabilidad celular son equiparables a los del tipo salvaje (véase el ejemplo 1).

40 Con el fin de permitir la integración del gen de la proteína heteróloga en el genoma y la selección para este evento, se utilizó la cepa PT36 *ura3*-negativa. Estas cepas se derivan de la cepa de *Candida bombicola* de tipo salvaje ATCC 22214, pero sólo alberga el promotor y el terminador del gen *ura3*, mientras que se separa la secuencia codificante de *ura3*. Esto se hizo por recombinación homóloga con un casete que contenía la región no codificante *ura3* 5' aguas arriba fusionada a su región 3' no codificante aguas abajo. El mutante PT36 es auxotrófico para uracilo o uridina (*ura3*⁻) y se puede transformar nuevamente en prototrofia con un gen *ura3* funcional. Los transformantes se pueden seleccionar en medio SD. El gen *cyp52M1* de la cepa PT36 se inactivó tal como se describe en el Ejemplo 1.

Ejemplo 4.1.1 Proteína verde fluorescente

Introducción

50 Esta cepa contiene el gen *yEGFP* que fue optimizado en codones para *Candida albicans* (Cormack *et al.*, 1997). El fuerte promotor GAPD constitutivo de *Candida bombicola* (Van Bogaert *et al.*, 2008a) se utilizó para impulsar la expresión del gen.

Materiales y Métodos

Cepas, plásmidos y condiciones de cultivo

Se utilizaron células ultracompetentes XL10GOLD de *Escherichia coli* para el mantenimiento del plásmido y para todos los experimentos de clonación. *C. bombicola* se cultivó en medio de peptona dextrosa de levadura (YPD) (1% de extracto de levadura, 2% de peptona, 2% de glucosa) o en medio sintético de dextrosa (SD) (0,67% de base nitrogenada de levadura sin aminoácidos (DIFCO) y 2% de glucosa). Se incubaron cultivos de matraces agitados de levadura líquida a 30°C y 200 rpm. *E. coli* se cultivó en medio Luria-Bertani (LB) (triptona 1%, extracto de levadura 0,5% y cloruro de sodio 0,5%) complementado con 100 mg/L de ampicilina, de ser necesario. Cultivos líquidos de *E. coli* se incubaron a 37°C y 200 rpm. El plásmido pGALyEGFPTU que alberga la variante yEGFP GFP, se obtuvo del Laboratorio de Biología Molecular de la Universidad de Ghent (LMBP). Esta variante de GFP contiene dos mutaciones (S65G y S72A) en relación con la GFP de tipo salvaje y adicionalmente está optimizada para codones para la levadura *Candida albicans* (Cormack *et al.*, 1997).

Aislamiento y secuenciación de ADN

El vector pGEM®-T (Promega) se utilizó para todos los experimentos de clonación. Se utilizaron T4 ADN ligasa (Fermentas) y el kit de clonación In Fusion Dry Down PCR (Clontech) para la clonación y el ligamiento. Todas las nucleasas de restricción se obtuvieron de New England Biolabs (NEB) y los digestiones de restricción se realizaron según lo especificado por el proveedor. Se aisló ADN genómico de cultivos de levaduras durante la noche cultivados en YPD. La pared celular de las levaduras se separó enzimáticamente mediante incubación del sedimento celular derivado de 1 ml de cultivo de levadura con 0,80 g de Enzima Lítica de Levadura (Sigma)/g de peso en húmedo de la célula en tampón SCE (sorbitol 1 M, acetato de sodio 0,1 M y EDTA 60 mM, pH 7,5) durante 90 min a 37°C en presencia de 3,75 µl de mercaptoetanol. El ADN genómico se aisló de los protoplastos restantes mediante el Kit DNeasy® Plant Maxi (Qiagen). El ADN del plásmido se aisló utilizando el Kit QIAprep Spin Miniprep (Qiagen). Las mezclas de la reacción PCR se purificaron utilizando el Kit de purificación QiAquick PCR (Qiagen) o el Kit Sure Clean Plus (Bioline) libre de columna para fragmentos de PCR mayores que 4 kb. Este último también se utilizó para la purificación de mezclas de restricción. Todas las secuencias de ADN se determinaron en LGC Genomics, Alemania.

Diseño de los cebadores y reacciones PCR

El diseño de los cebadores, el análisis de la secuencia y el diseño de la estrategia se realizaron con el software Clone Manager Professional Suite (Versión 8.0). Los cebadores se solicitaron de Sigma. Todas las reacciones PCR de alta fidelidad se realizaron utilizando la polimerasa de alta fidelidad *Pfu*, a menos que se establezca de otro modo. Las PCRs de colonias tanto en *E. coli* como en *C. bombicola* se realizaron utilizando *Taq* polimerasa.

Creación del casete de expresión de GFP

Se creó un vector 'central' (pGEM-T_cassette_yEGFP) que contenía el gen yEGFP flanqueado en ambos lados por una secuencia de aproximadamente 1 kb de longitud para la recombinación en el locus P/T *ura3* genómico en mutantes PT de *C. bombicola* ATCC 22214 (arriba descrito) y adicionalmente contiene el marcador *ura3* para la selección. El vector se construyó de tal manera que se puede cortar con dos enzimas de restricción únicas, de las cuales una (*SapI*) corta justo antes del codón de 'inicio' ATG del gen *gfp*. Esta característica del vector permite clonar fragmentos de promotor exactamente delante de la variante de GFP. Por lo tanto, se construyó un sitio de clonación múltiple (MCS3) del cual se añadió la secuencia como una extensión no vinculante al cebador P5_FOR_yEGFP extMCS3 (caracteres en negrita en la Tabla 6). Este MCS contiene 11 sitios de restricción únicos, de los cuales dos (*SapI* y *AvaI*) se utilizaron para linearizar el vector. El vector linearizado se utilizó entonces para clonar el promotor GAPD delante de la variante de GFP utilizando el kit de clonación In Fusion Dry Down PCR.

El vector pGEM-T_cassette_yEGFP se construyó en ocho pasos. Primero se amplificaron el 5'UTR del gen *ura3* (incluido el promotor *ura3*), la secuencia codificante de *ura3* y 3'UTR de *ura3* (incluido el terminador *ura3*) como un fragmento de PCR (1970 pb) a partir de ADN genómico de *C. bombicola* ATCC de tipo salvaje 22214, utilizando los cebadores P1_FOR_URA3v y P2_REV_URA3v. Dado que el vector pGEM-T se utilizó para la clonación, se utilizó el Kit High Fidelity PCR Master (Roche) para la amplificación. Este primer fragmento de PCR (PRODUCTO 1) proporciona el casete final con 1000 pares de bases necesarios para la recombinación homóloga en mutantes PT de *C. bombicola*, mientras que al mismo tiempo proporciona el casete con un marcador de selección. Este vector pGEM-T_ura3* resultante todavía contenía un sitio de reconocimiento *SapI*, que estaba originalmente presente en pGEM-T. Este sitio de reconocimiento debe separarse para hacer de *SapI* una enzima de restricción única. Esto último se realizó utilizando el kit Quick Change Site Directed Mutagenesis (Stratagene) con el par de cebadores de mutagénesis P7_FOR_QCSapI pGEM-T y P8_REV_QCSapI pGEM-T. El sitio de reconocimiento de *SapI* '5-GCTCTTC-3' se transformó así en la siguiente secuencia: '5-GCTCCTC-3' que ya no será reconocida por la endonucleasa de restricción *SapI*, haciendo que el sitio de reconocimiento *SapI* de MCS3 sea único. El vector resultante se denominó pGEM-T_ura3.

Un segundo paso implicó la amplificación del 3'UTR del gen *ura3* del ADN genómico utilizando los cebadores P3_FOR_URA3t_extyEGFP y P4_REV_URA3t_extNotI (PRODUCTO 2) y la amplificación de la variante yEGFP y el

terminador MAL2 de *Candida albicans* de pGALyEGFPTU utilizando los cebadores P5_FOR_yEGFP_extMCS3 y P6_REV_yEGFP_extURA3t. El sitio de clonación múltiple MCS3 y un sitio de restricción *NotI* se añadieron a PRODUCTO3 y PRODUCTO2, respectivamente, mediante extensiones no vinculantes en los cebadores P4_REV_URA3t_ext*NotI* y P5_FOR_yEGFP_extMCS3. PRODUCTO3 y PRODUCTO2 se fusionaron posteriormente por PCR solapada. Los productos de PCR del molde ya contenían los quince pares de bases complementarios necesarios para realizar la PCR solapada que se añadieron como extensiones no vinculantes en los cebadores P3_FOR_URA3t_extyEGFP y P6_REV_yEGFP_extURA3t. Los productos de PCR del molde se limpiaron utilizando el Kit de Purificación de PCR Qiaquick, se midieron las concentraciones y se establecieron tres reacciones de PCR con diferentes concentraciones del molde; 0,5 ng, 5 ng y 50 ng de PRODUCTO2 y PRODUCTO3 se añadieron a tres tubos de PCR separados. Posteriormente, se realizaron quince ciclos de PCR sin cebador siguiendo el siguiente programa de temperatura; una desnaturalización inicial a 95°C durante 2 min; una repetición de 15 veces de los siguientes tres pasos: 95°C durante 30 s, un reasociación específica a 60°C durante 30 s y un paso de elongación a 72°C durante 3 min. Estos quince ciclos fueron seguidos por un paso de elongación final a 72°C durante 7 min. Después de esta primera PCR sin cebador, 3 µl de cada uno de los dos cebadores para la amplificación del producto de fusión, P5_FOR_yEGFP_extMCS3 y P4_REV_URA3t_ext*NotI*, se añadió a cada uno de los tubos de PCR. A continuación, se realizaron 30 ciclos de PCR regulares. El producto de PCR de fusión, P3_P2 (= 2275 pb), se cortó posteriormente con *SpeI* y *NotI* al igual que el vector pGEM-T_ura3. El vector y la inserción se ligaron posteriormente utilizando T4 ADN ligasa y la mezcla de ligamiento se transformó en células de *E. coli* competentes XL10Gold. El vector resultante se denominó pGEM-T_cassette_yEGFP.

El vector construido se utilizó posteriormente para clonar el promotor de GAPD delante del codón de inicio yEGFP. El vector se linearizó utilizando las nucleasas de restricción *SapI* y *AvaI* y se utilizó el Kit de Clonación PCR In Fusion Dry Down para clonar el promotor en el vector linearizado. Para este fin, se proporcionaron al promotor los quince pares de bases necesarios para clonar fragmentos de PCR lineales en un vector linearizado. Estos se añadieron como fragmentos no vinculantes en los cebadores P18_REV_GAPDprom y P19_FOR_pGAPD1560 utilizados para la amplificación del promotor a partir de ADN genómico de *C. bombicola* y se representan en negrita en la Tabla 6. El vector construido se denominó pGEM-T_pGAPD1555_yEGFP.

Transformación

Células de levaduras se transformaron mediante electroporación estándar. El ADN lineal para la transformación se obtuvo por PCR estándar con los cebadores P1_FOR_URA3v y P31_REV_cassette (Tabla 6). Los transformantes se seleccionaron en placas SD. Las células de *E. coli* se transformaron tal como se describe por Sambrook (Sambrook *et al.*, 2001) y la selección se produjo en placas de LB complementadas con ampicilina.

Fluorimetría

La fluorescencia se midió utilizando un dispositivo Spectramax Gemini XS (Molecular Devices, St. Grégoire, Francia) con placas negras de 96 pocillos. La emisión de fluorescencia se midió a 511 nm después de la excitación a 488 nm y se cuantificó en unidades de fluorescencia relativa (RFUs). Los cultivos se cultivaron en SD y antes de medir la fluorescencia. *C. bombicola* de tipo salvaje ATCC 22214 se utilizó como referencia en blanco.

Resultados

C. bombicola se transformó con el casete de expresión de GFP tal como se describe en la sección de materiales y métodos. Después de cinco días de incubación aparecieron colonias en las placas SD selectivas. La PCR de colonias se realizó con el par de cebadores P37_FOR_checkGFP y P35_REV_checkcasIN y se obtuvo una colonia positiva. Se aisló el ADN genómico de E1 mutante (*Candida bombicola*_pGAPD1555_yEGFP) y se realizaron varias PCRs control para verificar el genotipo correcto. Estas PCRs control confirmaron que los mutantes contenían los casetes de integración completos.

Tres réplicas biológicas del mutante y WT, respectivamente, se cultivaron en 3 mL de medio SD en placas profundas de 24 pocillos a 30°C, 200 rpm. Los pocillos se inocularon en la misma DO (0,2) a partir de precultivos en SD a partir de colonias individuales. Se tomaron varias muestras a lo largo de la curva de crecimiento. Se dispusieron 100 microlitros (x2) de cada uno de los pocillos en una placa negra de 96 pocillos y se midió la fluorescencia después de la excitación a 488 nm. Los valores en blanco obtenidos para el tipo salvaje se sustrajeron de los valores de fluorescencia del mutante y WT, respectivamente (fondo) y los valores obtenidos se representaron frente al tiempo de incubación.

Como se puede ver en la Figura 12 se detectó una señal fluorescente significativa entre 40 y 63 horas de incubación.

Ejemplo 4.1.2 Amilasa

Introducción

Con el fin de explorar sus capacidades para producir enzimas (heterólogas), *Candida bombicola* ATCC 22214 soforolípido-negativa *cyp52M1* fue testada en cuanto a su producción de una -amilasa. Las -amilasas constituyen una clase importante de enzimas que encuentran muchas aplicaciones biotecnológicas en procesos que implican, por ejemplo, la degradación de almidón y la determinación de fibra dietética soluble e insoluble en salvado de arroz y trigo. Tales aplicaciones se encuentran en industrias de panificación, elaboración de cerveza, detergentes y textiles (Roy et al., 2000). En la panificación, por ejemplo, las -amilasas se utilizan porque aumentan el volumen del pan, porque mejoran el grano de la miga, la corteza y el color de la miga, y por su desarrollo del sabor fomentado en el producto final (Rosell et al., 2001). La -amilasa es una endoenzima que hidroliza aleatoriamente los enlaces -1,4 glucosídicos en polisacáridos.

Para la expresión heteróloga en *Candida bombicola*, se ha seleccionado la α -amilasa de *Aspergillus oryzae* (*amy3*). El fuerte promotor de GAPD constitutivo de *Candida bombicola* (número de acceso de GenBank EU315245) se utilizó para impulsar la expresión del gen. Con el fin de obtener la enzima secretada, se ha proporcionado una señal de secreción N-terminal (la señal de secreción del factor de apareamiento α de *S. cerevisiae*). Dado que el uso de codones de *A. oryzae* difiere significativamente del de *C. bombicola*, y *C. bombicola* no tiene múltiples genes exón-intrón, se ha diseñado una secuencia de ADN optimizada para codones.

Materiales y Métodos

Cepas, plásmidos y condiciones de cultivo

C. bombicola inactivada en el gen *cyp52M1* y en el gen *ura3* (cepa PT36) se utilizó para la expresión de amilasa. Más específicamente, el casete de expresión de amilasa se utilizó para transformar un 'mutante PT' *ura3* *C. bombicola* auxotrófico, soforolípido-negativo (descrito en la introducción del ejemplo 4).

Escherichia coli DH5 α se utilizó en experimentos de clonación y para el mantenimiento de plásmidos.

El plásmido pGEM-T_pGAPD1555_yEGFP (según se describe en el Ejemplo 4.1.1: producción de la proteína GFP en *Candida bombicola*) se utilizó como estructura principal del vector. Este vector contiene un marcador de selección bacteriana (Amp^R) y origen de replicación, una copia funcional del gen *ura3* de *C. bombicola*, y el gen yEGFP. La transcripción de yEGFP es controlada por 1560 pb de la región de promotor de GAPD de *C. bombicola* (número de acceso GenBank EU315245) y el terminador MAL2 de *Candida albicans* (número de acceso GenBank M94674). La construcción promotor-yEGFP-terminador está flanqueada en ambos lados por una secuencia de aproximadamente 1 kb de longitud para la recombinación en el locus *ura3* de PT genómico en mutantes PT de *C. bombicola* ATCC 22214, estando la copia funcional del gen *ura3* de *C. bombicola* en el lado 5' aguas arriba.

Para experimentos de rutina, las células de levadura se cultivaron en medio YPD (1% de extracto de levadura, 2% de peptona y 2% de dextrosa), mientras que la selección después de la transformación se realizó en medio de dextrosa sintética (SD) (0,67% de base nitrogenada de levadura sin aminoácidos (DIFCO) y 2% de glucosa) y la producción de amilasa se realizó en medio 3C con 3% de sacarosa (10% de glucosa, 1% de extracto de levadura, 0,1% de urea, 3% de sacarosa). Los cultivos de levadura se incubaron a 30 °C y 200 rpm.

Células de *E. coli* se cultivaron en medio Luria-Bertani (LB) (triptona 1%, extracto de levadura 0,5%, cloruro de sodio 0,5% (y 15% de agar para las placas)) complementado con 100 mg/L de ampicilina. Los cultivos de *E. coli* se incubaron a 37 °C y 200 rpm.

Manipulación estándar de ADN

La metodología de ADN recombinante de rutina se realizó de acuerdo con Sambrook y Russell (2001) (Sambrook y Russell, 2001). Las concentraciones de ADN se midieron con el espectrofotómetro NanoDrop® ND-1000 UV-Vis (NanoDrop Technologies). La escala doble logarítmica de Westburg BV se utilizó para controlar la longitud de los productos de ADN. Las reacciones PCR se realizaron con *Taq*-ADN polimerasa estándar (Westburg BV) o con PfuUltra™ High-Fidelity DNA Polymerase AD (Stratagene) de acuerdo con el procedimiento descrito por el fabricante. Fragmentos de gel se purificaron con el kit de Purificación de Gel Qiaexll (Qiagen). El ligamiento se realizó con la T₄ ADN ligasa de Fermentas, de acuerdo con el protocolo del fabricante. El ADN del plásmido se aisló con el Kit QIAprep Spin Miniprep (Qiagen). La secuenciación se realizó en AGOWA (LGC genomics). El diseño del cebador, el análisis de la secuencia y el diseño de la estrategia se realizaron con el software Clone Manager Professional Suite (Versión 8.0). Todos los cebadores se obtuvieron de Sigma-Aldrich Co. Las enzimas de restricción se obtuvieron de New England Biolabs (Westburg BV) y se utilizaron según lo indicado por el fabricante.

Creación del casete de expresión de amilasa

Dado que el uso de codón de *A. oryzae* difiere significativamente del de *C. bombicola*, y *C. bombicola* no tiene múltiples genes exón-intrón, se diseñó una secuencia de ADNc optimizada por codones (Figura 13). La secuencia proteica de la α -amilasa madura de *Aspergillus oryzae* (TAKA-amilasa A, EC 3.2.1.1., codificada por *amy3*, número de acceso de Genbank CAA31220.1) se tradujo empleando el uso del codón promedio de 10 genes de *C. bombicola* conocido por ser bien expresado. En la construcción sintética, la secuencia codificante de amilasa (sAmyAO) está precedida por la secuencia que codifica 85 aminoácidos de la señal de secreción del factor de apareamiento α de *S. cerevisiae* (número de acceso de GenBank NP_015137), que también se tradujo para la optimización del codón (sMFaScss). Parte del promotor GAPD de *C. bombicola* (número de acceso de GenBank EU315245) se dispuso 5' aguas arriba de la construcción sintética (pGAPD). Finalmente, se realizaron algunas mutaciones para introducir un sitio de clonación múltiple delante del codón de inicio ATG y en el extremo 3' de la secuencia codificante. La secuencia final de la construcción ordenada se da en la Figura 14. La construcción (2153 pb) se ordenó como tal en GenScript (Piscataway, EE.UU.) y se obtuvo clonada en el sitio *EcoRI* en el vector de clonación de plásmido pUC57.

El casete de expresión de amilasa sintética se aisló de pUC57 por restricción con *Bss*HI y *Bs*WI. La estructura principal del vector pGEM-T_pGAPD1555_yEGFP se cortó con *Bss*HI y *Bsr*GI, creando extremos compatibles. Después de la extracción con gel, el casete de expresión de amilasa se ligó en la estructura principal del vector pGEM-T_pGAPD1555_yEGFP, dando como resultado el vector p_sAmyAO_pGapd_iUra (Figura 15). La mezcla de ligamiento se transformó en *E. coli* Dh5, y las colonias positivas se seleccionaron mediante PCR de colonias con los cebadores sAmyAOfw y sAmyAOrv (Tabla 7). La secuencia del plásmido construido se confirmó mediante secuenciación en Agowa (LCG Genomics). El plásmido p_sAmyAO_pGapd_iUra es el mismo que pGEM-T_pGAPD1555_yEGFP, pero la región codificante yEGFP ha sido reemplazada por la región codificante de la amilasa (y la señal de secreción).

Antes de la transformación, el casete de expresión de amilasa junto con el marcador *ura3* (6426 pb) se linearizó a partir del plásmido p_sAmyAO_pGapd_iUra por restricción con *Bgl*I y *Eag*I. El casete linearizado contenía regiones para la recombinación homóloga en el locus *ura3*, de aproximadamente 1 kb de longitud (Figura 15).

Transformación de levadura y verificación de los transformantes

Células de *C. bombicola* se transformaron con el método de acetato de litio (Gietz *et al.*, 1995), pero en lugar de 100 se utilizó LiAc 50 mM. Los transformantes se seleccionaron en placas sintéticas de dextrosa (SD) (prototofía de uracilo restaurada con *C. bombicola ura3*).

La PCR de colonias de levadura se realizó con *Taq* ADN polimerasa (Westburg BV), de acuerdo con el protocolo del fabricante, con la excepción de que la desnaturalización inicial se realizó durante 7 minutos.

Ensayo de enzima amilasa

Se tomaron muestras de 200 μ L de cultivos de 3C + 3% de sacarosa. Después de la medición de la DO, las muestras se sedimentaron por centrifugación a 12000 rpm durante 3 minutos. El sobrenadante se utilizó para ensayos de enzimas con el kit de ensayo de amilasa AMYL® (Roche Diagnostics).

La mezcla de ensayo de enzima contenía 1 μ L de muestra, 50 μ L de R₁, 10 μ L de R₂ y 139 μ L de tampón citrato-fosfato, pH 7,0. La mezcla se incubó en el espectrofotómetro a 37°C. La absorbancia se mide cada 30 s durante 30 minutos.

Una unidad de amilasa se define como la cantidad de enzima capaz de elevar la DO_{405 nm} en 1 durante 15 minutos en una mezcla de reacción que consiste en 1 μ L de muestra + 50 μ L de R₁ + 10 μ L de R₂ + 139 μ L de tampón incubado a 37°C.

Resultados

El casete de expresión de amilasa, junto con el marcador *ura3*, se construyó tal como se describe en la sección "Materiales y Métodos". Este fragmento lineal (6426 pb) se utilizó para transformar células PT de *Candida bombicola* soforolípido- y uracilo/uridina-negativas. Una colonia apareció en las placas SD selectivas después de 4-10 días de incubación. El genotipo de este transformante se verificó mediante PCR de colonias de levaduras con 2 pares de cebadores. En primer lugar, se utilizaron cebadores sAmyAOfw y sAmyAOrv (Tabla 7), ambos se reasocian en la región codificante de amilasa del casete. De esta manera, se aseguró la integridad de la parte que expresa amilasa. El segundo par de cebadores utilizado fue sAmyAOfw2 y P35 (Tabla 7). El cebador sAmyAOfw2 se une a la región codificante de amilasa, mientras que el último se une fuera de las regiones homólogas del casete, en el ADN genómico. Por lo tanto, se verifica la correcta integración del casete en el locus *ura3*. La colonia mostró el genotipo correcto y se la conoce como 'SL Amy⁺'.

El transformante 'SL⁻Amy⁺' y el tipo salvaje se cultivaron posteriormente en medio 3C con 3% de sacarosa. La producción de amilasa, el crecimiento celular (DO 600 nm) y el pH de los cultivos se siguieron durante 48 h, tomando muestras aproximadamente cada 2 horas. Los ensayos de la enzima amilasa se realizaron tal como se explica en la sección "Materiales y Métodos". Los resultados de este ensayo de crecimiento y producción se muestran en las Figuras 16 y 17. De la Figura 16 resulta claro que el transformante 'SL⁻Amy⁺' produce amilasa, hasta 106 unidades/ μ L de sobrenadante de cultivo, mientras que la cepa de tipo salvaje no produce amilasa alguna (9 unidades/ μ L máximo, fondo). Las curvas de crecimiento y pH (Figura 17) son similares para ambas cepas, aunque el pH cae más en el cultivo del transformante 'SL⁻Amy⁺' al final del ensayo. La producción de amilasa no tiene un efecto negativo sobre el crecimiento de *C. bombicola*. En conclusión, se creó un transformante de expresión de amilasa que se inactiva en producción de soforolípidos, pero en su lugar produce -amilasa, hasta 106 unidades/ μ L de sobrenadante de cultivo cuando se cultiva en medio 3C con 3% de sacarosa. La cepa de tipo salvaje no produjo amilasa alguna. Esta es la primera vez que se describe la producción de enzimas (extracelulares) para una cepa soforolípidos-negativa de *C. bombicola*.

Ejemplo 4.2 Creación de una cepa que sintetiza polihidroxialcanoatos.

15 Introducción

Esta cepa se inactiva en el gen *cyp52M1* (número de acceso de GenBank EU552419) mencionado anteriormente y en su lugar porta un gen de PHA sintasa (*phaC1*), cuya expresión depende de las secuencias reguladoras aguas arriba y aguas abajo del gen *cyp52M1*. La secuencia proteica de PHAC1_{pr} (número de acceso ENA AAD26365.2) de *Pseudomonas resinovorans* se volvió a traducir empleando el uso del codón promediado de los genes de la vía de soforolípidos. La cepa resultante ya no produce soforolípidos, sino que, en su lugar, produce MCL-polihidroxialcanoatos (PHA) cuando se cultiva en glucosa con la adición de aceite de colza. El crecimiento celular y la viabilidad de dicha cepa son equiparables a los de tipo salvaje.

Materiales y Métodos

Cepas, plásmidos y condiciones de cultivo

25 *Candida bombicola* ATCC 22214 se utilizó para el aislamiento del gen *ura3* con secuencias reguladoras aguas arriba y aguas abajo y para el aislamiento de la región aguas abajo del gen *cyp52M1* (downCYP). La cepa PT36 *ura3⁻* *Candida bombicola* auxotrófica (que se describe en el Ejemplo 4.1.1) se utilizó para la inserción del casete de expresión de PHA. Células de levadura se cultivaron en medio YPD (1% de extracto de levadura, 2% de peptona y 2% de dextrosa), medio SD (0,67% de base nitrogenada de levadura sin aminoácidos (Difco), 2% de glucosa), medio 3C (10% de glucosa, 1% de extracto de levadura, 0,1% de urea y 15% de agar) o el medio descrito por Lang (Lang *et al.*, 2000) al que se añadió aceite de colza (Sigma) después de 48 h cuando se mencionó. Los cultivos de levadura se incubaron a 30 °C y 200 rpm.

35 *Escherichia coli* DH5 α o XL10GOLD se utilizaron en los experimentos de clonación y la transformación se produjo tal como se describe por Sambrook y Russell (2001). Células de *E. coli* se cultivaron en medio Luria-Bertani (LB) (triptona 1%, extracto de levadura 0,5%, cloruro de sodio 0,5% (y 15% de agar para las placas)) complementado con 100 mg/L de ampicilina. Los cultivos líquidos de *E. coli* se incubaron a 37°C y 200 rpm.

Aislamiento y secuenciación de ADN

40 El ADN genómico de levadura se aisló con el kit de ADN genómico bacteriano GenElute™ (Sigma) de cultivos de levadura nocturnos cultivados en YPD. La pared celular de la levadura se separó enzimáticamente mediante incubación del sedimento celular derivado de 1 ml de cultivo de levadura con 0,80 g de Enzima Lítica de Levadura (Sigma)/ g de peso en húmedo de la célula en tampón SCE (sorbitol 1 M, acetato de sodio 0,1 M y EDTA 60 mM, pH 7,5) durante 90 min a 37°C en presencia de 3,75 μ l de mercaptoetanol. El ADN genómico se aisló de los protoplastos restantes por medio del kit de ADN genómico bacteriano GenElute™ (Sigma).

45 El ADN del plásmido bacteriano se aisló con el Kit QIAprep Spin Miniprep (Qiagen). Todos los productos de PCR se clonaron en el vector pGEM-T® (Promega) o sus derivados y se enviaron como tales a AGOWA (LGC genomics) para el análisis de la secuencia. El vector pGEM-T® era la estructura principal de todos los vectores construidos. Cuando fue necesario, el ADN se aisló del gel utilizando el kit de extracción de gel QIAquick (Qiagen).

Diseño de cebadores y reacciones PCR

50 El diseño del cebador, el análisis de la secuencia y el diseño de la estrategia se realizaron con el software Clone Manager Professional Suite (Versión 8.0). Los cebadores se solicitaron de Sigma. Todas las reacciones de PCR de alta fidelidad se realizaron utilizando la polimerasa de alta fidelidad *Pfu*. Las PCR de colonias tanto en *E. coli* como en *C. bombicola* se realizaron utilizando *Taq* polimerasa.

Transformación

Las células de *C. bombicola* se transformaron utilizando un protocolo de electroporación estándar. Los transformantes se seleccionaron en placas SD. Células de *E. coli* se transformaron tal como se describe por Sambrook (Sambrook *et al.*, 2001) y la selección se produjo en placas de LB complementadas con ampicilina.

5 Construcción sintética

La secuencia de la proteína PHAC1 de *Pseudomonas resinovorans* (número de acceso ENA AAD26365.2) se volvió a traducir empleando el uso de codones promedio de los genes de la vía SL, que se determinó utilizando una herramienta en línea (Stothard, 2000). Se añadió una secuencia diana peroxisomal (PTS) SKL (TCTAAGCTG) en el extremo 3' del gen, así como en las regiones reguladoras aguas arriba y aguas abajo del gen *cyp52M1*, respectivamente, en el lado 5' (488 pb) y 3' (190 pb) de la secuencia *phac1* optimizada por codones. La región 5' UTR se extendió a 1098 pb para obtener suficiente homología para la recombinación homóloga en el locus *cyp52M1*.

La construcción se solicitó como tal en GenScript (Piscataway, EE.UU.) y se obtuvo clonada en un vector. La construcción se amplificó con los cebadores P55_FOR_upCYP_ext*NheI* y P58_REV_PHAC1+tCYP_ext*EcoRI* (Tabla 8), produciendo un fragmento de 2986 pb. Los cebadores contenían respectivamente extensiones *NheI* y *EcoRI*, de modo que el fragmento se pudo digerir posteriormente con dichas enzimas de restricción para la subclonación ulterior de la construcción sintética.

Creación del casete de expresión de PHA

El casete de expresión contenía el gen *phaC1* optimizado por codones y el marcador *ura3* y se construyó de modo que contenía las regiones aguas arriba y aguas abajo del gen *cyp52M1* para la recombinación homóloga en el locus *cyp52M1* en el ADN genómico de una cepa PT36 auxotrófica de *C. bombicola*. La construcción del casete se produjo en tres pasos. Primero, la región para la recombinación homóloga en el extremo 3' del gen *cyp52M1* (*downcyp*) se amplificó a partir del ADN genómico de *C. bombicola* utilizando los cebadores P53_FOR_downCYP_ext*SpeI* y P54_REV_downCYP_ext*NotI* (Tabla 8). El amplicón resultante y el vector pGEM-T®*ura3* (véase el Ejemplo 4.1.1) se digirieron con los cortadores únicos *SpeI* y *NotI* y posteriormente se ligaron utilizando T4 ligasa (NEB). En segundo lugar, el vector resultante se digirió utilizando los cortadores únicos *NheI* y *EcoRI*. Esta doble restricción proporcionó dos fragmentos (5644 pb y 358 pb) de los cuales el más grande se purificó en gel y posteriormente se ligó con la construcción sintética amplificada que se sometió primero a restricción con las mismas enzimas de restricción. En un tercer y último paso, el casete de expresión se amplificó utilizando los cebadores P63_FOR_cassPHAC1 y P64_REV_cassPHAC1 y el fragmento de PCR lineal se purificó y se utilizó para la transformación de la cepa de *C. bombicola* PT36 (Figura 18). Los transformantes se seleccionaron en placas SD.

Muestreo

Las muestras analíticas de soforolípidos se prepararon de la siguiente manera: se añadieron 440 µL de acetato de etilo y 11 µL de ácido acético a 1 mL de caldo de cultivo y se agitaron vigorosamente durante 5 min. Después de la centrifugación a 12000 rpm durante 5 min, se separó la capa de disolvente superior y se dispuso en un tubo eppendorf nuevo con 700 µL de etanol. Las muestras se analizaron mediante HPLC y Detección de Dispersión de la Luz Evaporativa.

El peso seco de la célula (CDW) se midió transfiriendo 2 mL de caldo de cultivo a un filtro de nitrato de celulosa con un diámetro de los poros de 0,45 µm (Sartorius) y el peso seco se determinó en el horno automático XM60 de Precisa Instruments Ltd.

Se realizó la hidrólisis de PHA y la formación de éster metílico de ácidos grasos (FAME) realizando metanolisis ácida de 30 mg de material celular liofilizado en 4 ml de mezcla de cloroformo/metanol 1:1 + H₂SO₄ al 3% a 95°C durante 4 h. Antes de la metanolisis las células se recogieron (4000 rpm, 4°C) a partir de cultivos de matraces, después de lo cual los sedimentos celulares se congelaron a -80°C y se liofilizaron durante 24 h. 30 mg del material liofilizado resultante se lavaron varias veces con 25 ml de metanol caliente (65°C) para separar el aceite y los ácidos grasos libres. Después de la metanolisis, se añadieron 4 ml de NaCl al 0,9% (p/v) a los tubos y la fase orgánica se recogió para el análisis por GC-MS. Se añadió 1 mg de patrón interno (ácido 2-hidroxihexanoico) antes de la metanolisis y se añadió 1 mg de patrón externo (ácido 12-hidroxidodecanoico) justo antes de la inyección de las muestras.

Análisis HPLC de soforolípidos

Muestras de soforolípidos se analizaron por HPLC en un sistema Varian Prostar HPLC utilizando una columna Chromolith® Performance RP-18e de 100-4,6 mm de Merck KGaA a 30°C y Detección de Dispersión de la Luz Evaporativa (Alltech). Se tuvo que utilizar un gradiente de dos eluyentes, una solución acuosa de ácido acético al

0,5% y acetonitrilo para separar los componentes. El gradiente comenzó con 5% de acetonitrilo y aumentó linealmente hasta 95% en 40 min. La mezcla se mantuvo así durante 10 min y luego se devolvió a 5% de acetonitrilo en 5 min. Se aplicó un caudal de 1 mL/min.

Análisis GC-MS de FAMES

- 5 La GC (TraceGC ultra, Interscience) contiene una columna Rxi® de 0.25 mm-1 ml (Restek) que está recubierta con dimetil polisiloxano. El gas portador utilizado era helio. Se utilizó el siguiente perfil de temperaturas: 2 minutos a 64°C, seguido de un aumento lineal de 30°C/min a 200°C. Cuando la columna alcanzó 200°C, se produjo un segundo aumento lineal de 50°C/min a 310°C. Los compuestos de elución se inyectaron posteriormente en la MS (DSQ, Interscience) en donde se ionizaron y se detectaron para una identificación adicional. Esto último se realizó
10 utilizando el software Xcalibur, que se acopló a la biblioteca NIST MS Search 2.0. Sólo se detectaron compuestos entre 40 en 650 g/mol.

Resultados

- 15 El casete de expresión de PHA se construyó tal como se describe en la sección "Materiales y Métodos". Este fragmento lineal se utilizó para transformar células de *Candida bombicola* PT36 (Figura 18). 16 colonias aparecieron en las placas SD selectivas después de 4-11 días de incubación. El genotipo de estos 16 transformantes se verificó mediante PCR de colonias de levadura con el par de cebadores P9_FOR_seqQCSap/_URA3down que se unen en el casete de expresión y A21totRev, uniéndose al ADN genómico aguas abajo del sitio de recombinación correcto (Tabla 8). Diez de las colonias mostraron el genotipo correcto. Se realizaron otras dos PCRs de colonias en las colonias positivas para controlar la inserción correcta del casete. El ADN genómico de un mutante (A8) se aisló
20 posteriormente y se realizó una reacción PCR con el par de cebadores UDPGTA1R y A21TotRev que se unen al ADN genómico de *C. bombicola* justo aguas arriba y aguas abajo de los sitios de recombinación izquierda y derecha, respectivamente. Este fragmento de PCR se envió para la secuenciación y el análisis reveló que el casete de expresión de PHA se había integrado correcta y completamente en el locus *cyp52M1* de *C. bombicola* y no contenía errores.

- 25 El mutante A8 se cultivó posteriormente en el medio descrito por Lang. Tres réplicas biológicas se complementaron con aceite de colza después de 48 h de incubación, y otras tres no. Cada uno de los matraces de agitación se inoculó de un precultivo diferente de cultivos desarrollados durante la noche en medio de Lang (5 mL) inoculados a partir de una colonia de una placa 3C. El tipo salvaje se cultivó en paralelo como control. Se tomaron muestras para la extracción de SL a lo largo de la curva de crecimiento al igual que las muestras para el consumo de glucosa y la
30 determinación de CDW. Trece días después de la adición del aceite, las células se recogieron y la hidrólisis de PHA y la conversión en FAMES se realizaron tal como se describe en la sección "Material y Métodos".

- 35 El efecto de la alteración del gen *cyp52M1* fue similar al descrito en el Ejemplo 1. El CDW y el crecimiento fueron similares para el mutante de expresión de PHAC1 y el consumo de tipo salvaje y de glucosa en fase estacionaria fue más lento para el mutante de expresión de PHAC1. Mientras que claramente había producción de soforolípidos para el tipo salvaje, no se pudieron detectar soforolípidos en el medio para el mutante de expresión de PHA, con o sin adición de aceite.

- 40 El análisis GC-MS de los FAMES derivados de muestras finales del experimento de crecimiento del mutante PHAC1 A8 (véase materiales y métodos) al que se añadió aceite de colza condujo a la identificación de compuestos derivados de MCL-PHA producidos en el mutante A8 de PHAC1 de *Candida bombicola* (Figura 19). Se detectaron tres compuestos derivados de PHA en dos de las tres réplicas biológicas: 3-metilhidroxioctanoato (0,50% p/dp), 3-metilhidroxidecanoato (0,54% p/dp) y 3-metilhidroxidodecanoato (0,32% p/dp) con un total de 1,36 p/dp de PHA. En la tercera réplica biológica se detectó adicionalmente 3-metilhidroxitetradecanoato (0,50% p/dp) y las cantidades de los otros monómeros de PHA fueron ligeramente más altas para este matraz: 3-metilhidroxioctanoato (0,64% p/dp),
45 3-metilhidroxidecanoato (0,73 % p/dp) y 3-metilhidroxidodecanoato (0,29% p/dp) con una cantidad total de 2,16% p/dp de PHA. Estos picos no se detectaron para las muestras derivadas de los cultivos de tipo salvaje (con y sin adición de aceite de colza) ni para las muestras del mutante A8 de PHAC1 a las que no se añadió aceite de colza. Por lo tanto, los picos no se derivaron de compuestos intermedios de la beta-oxidación que se convirtieron en FAMES durante la metanolisis. El aceite de colza fue necesario como una fuente lipogénica para producir el PHA (en cantidad sustancial). La cantidad de PHA producido se cuantificó utilizando un patrón interno (ácido 2-hidroxihexanoico) que se añadió a las muestras antes de la metanolisis.
50

- 55 Las secuencias reguladoras de *cyp52M1* impulsaron la expresión del gen PHAC1 en este experimento. Se descubrió además que se necesita una concentración sustancial de glucosa, en combinación con la inanición de N o P, para activar este promotor. La glucosa, por otro lado, reprime la expresión de los genes de la beta-oxidación. Para obtener una mayor producción de PHA, los promotores de la catalasa (pCTA) y la isocitrato liasa (pICL) se aíslan del genoma y la expresión de PHAC1 se deriva de estos promotores que están reprimidos por altas concentraciones de

glucosa. La interrupción de la beta-oxidación y la alimentación con sustratos con longitud de cadena larga (C16 y C18) conduce a la producción de PHA compuesto solo por monómeros C18 y/o C16.

En conclusión, se creó un mutante de expresión de PHAC1 que se inactiva en la producción de soforolípidos, pero en su lugar produce MCL-PHA hasta 0,99 p/dp cuando se cultiva en glucosa con la adición de aceite de colza.

5 Ejemplo 4.3 Producción de Glicolípidos

Ejemplo 4.3.1 Glucolípidos

Debido al grupo carboxilo libre, una cadena de carbono insaturada y la presencia de un grupo de cabeza de hidrato de carbono, los glucolípidos son productos intermedios interesantes para varios tipos de reacciones de conversión biocatalíticas o químicas. Dado que pequeñas variaciones estructurales pueden tener una influencia significativa sobre la actividad biológica o propiedades físico-químicas de un glicolípido, la síntesis enzimática o quimio-enzimática de derivados de soforolípidos ha sido objeto de algunos trabajos de investigación (Bisht *et al.*, 1999; Carr y Bisht, 2003 ; Rau *et al.*, 2001) y solicitudes de patente (WO2004/044216 y US05/0164955). Hasta la fecha, los glucolípidos con un extremo carboxílico libre sólo se producen por conversión enzimática de soforolípidos de carácter ácido (anillo abierto) que a su vez se obtienen después de la hidrólisis alcalina del bioproducto bruto de *C. bombicola* (Rau *et al.*, 1999; Saerens *et al.*, 2009). Por otro lado, los alquilglucósidos podrían obtenerse por conversión microbiana de alcoholes secundarios (Brakemeier *et al.*, 1998) o alcoholes grasos ramificados (Palme *et al.*, 2010). Los glucolípidos de carácter ácido (y soforolípidos) atraen especialmente la atención debido a que son bola-anfífilos asimétricos que, además de las estructuras supramoleculares que típicamente forman, también tienen una mayor versatilidad química en comparación con los simétricos químicamente sintetizados (Zhou *et al.*, 2004). El mutante de delección *UgtB1* creado en el Ejemplo 3 es una cepa interesante que ahora ofrece un ahorro de tiempo en la producción *in vivo* de estas biomoléculas a partir de sustratos renovables económicos. Cuando los autores de la invención repitieron la fermentación en matraz con el mutante B11, pero prolongaron la incubación en aceite de colza a 14 días, los glucolípidos producidos se identificaron como una mezcla de moléculas estructuralmente relacionadas reveladas por análisis de espectrometría de masas de extractos de cultivo (véase la sección de material y método, el Ejemplo 3 y la Figura 20). Además de los glucolípidos de carácter ácido monoacetilado más predominantes ($m/z = 502$), se pudieron identificar cantidades menores de glucolípidos no acetilados (sub-) hidroxilados en posición terminal ($m/z = 460$) (Figura 20). Hasta ahora, no se observó lactonización de glucolípidos. La aparición de glucolípidos acetilados ilustra que la acetiltransferasa, que normalmente decora los soforolípidos *de novo* en sus posiciones 6' y/o 6" con grupos acetilo, también muestra actividad hacia los glucolípidos. A este respecto, es posible que la acetilación de glucolípidos se produzca antes de la adición de la segunda unidad de glucosilo durante la síntesis *de novo* de soforolípidos, sin ser necesaria para esta segunda reacción de glucosilación.

Después de 14 días de incubación en aceite de colza, el aceite residual todavía flota en la parte superior del cultivo B11, mientras que una fermentación natural de tipo salvaje después de este tiempo utilizó completamente aceite de colza para la producción de soforolípidos con un rendimiento común en torno a 50 g/L, indicando una conversión de sustrato menos eficiente en glucolípidos en comparación con soforolípidos.

La producción de glucolípidos por parte del mutante de delección *UgtB1* crea ahora un proceso de producción *in vivo* eficiente en el tiempo de estos interesantes compuestos intermedios de glicolípidos y esto mediante la fermentación convencional en sustratos económicos.

Ejemplo 4.3.2 Celobiosalípidos

40 Introducción

Los celobiosalípidos son producidos en la naturaleza por varias levaduras tales como *Cryptococcus* (Puchkov *et al.*, 2002), *Pseudozyma* (Kulakovskaya *et al.*, 2005) y *Sympodiomyces* (Kulakovskaya *et al.*, 2004), y el hongo dimórfico *Ustilago maydis* (Spoeckner *et al.*, 1999, Teichmann *et al.*, 2007). Su estructura general es equiparable a los soforolípidos, sin embargo las dos unidades de glucosa están unidas por un enlace β -1,4, la cola de ácido graso puede mostrar hidroxilaciones múltiples (α -, ω - y ω -1) y la molécula de celobiosa está acetilada y/o acilada con ácidos β -hidroxilados grasos de cadena corta (C6/C8). Aunque los celobiosalípidos son agentes antimicrobianos prometedores, hasta la fecha no hay producción industrial debido a los muy bajos rendimientos generales. Los autores de la invención creamos aquí un mutante de *C. bombicola* que produce celobiosalípidos en lugar de soforolípidos cambiando el gen *UGTB1* de tipo salvaje (véase el ejemplo 3) por el gen *UGT1* de *U. maydis* o el gen *CepB* de *Clostridium stercoararium* (Reichenbecher *et al.*, 1997). La glucosiltransferasa Ugt1 de *U. maydis* actúa de forma muy similar a la glucosiltransferasa UgtB1 de *C. bombicola*, utilizando la misma UDP-glucosa que un donante de glucosilo y un glucolípido equiparable como aceptor, uniendo ambos por un enlace β -1,4 en lugar del β -1,2 formado por la glucosiltransferasa *UgtB1* (Teichmann *et al.*, 2007). La celodextrina fosforilasa bidireccional (CepB) de la bacteria termófila celulolítica *Clostridium stercoararium* está implicada en la degradación de la celulosa, pero la dirección sintética de la enzima puede utilizar ácido (17-O- β -D-glucopiranosil)-octadecenoico obtenido de

soforolípidos de *C. bombicola* (Saerens *et al.*, 2009) como un aceptor para la formación de celobiosalípidos, utilizando glucosa-1-Pi como donante de glucosilo.

Materiales y Métodos

Cepas y plásmidos

- 5 *C. bombicola* G9 (derivada de ATCC 22214, véase Van Bogaert *et al.*, 2008) se utilizó para la creación del mutante productor de celobiosalípidos. Se cultivaron células de levadura en medio YPD que contenía 1% de extracto de levadura, 2% de peptona y 2% de dextrosa, medio SD que contenía 0,67% de base nitrogenada de levadura sin aminoácidos (Difco) y 2% de glucosa o medio 3C que contenía 10% de glucosa, 1% de extracto de levadura y 0,1% de urea. Los medios líquidos se incubaron a 30°C y 200 rpm. Se cultivó *E. coli* en medio Luria Bertani (extracto de levadura Bacto 0,5%, Triptona Bacto 1%, NaCl 0,5%) que contenía ampicilina al 0,01% y se incubó a 37°C y 200 rpm. Los plásmidos se aislaron de *E. coli* mediante el kit de aislamiento de plásmidos MiniPrep de Qiagen y se secuenciaron en LGC genomics (Alemania).

Creación de los casetes de expresión *UGT1* y *CepB*

- 15 Tanto los casetes de expresión *UGT1* como *CepB* se crearon mediante acoplamiento mediado por enzimas de restricción del promotor *UGTB1* de *C. bombicola* de tipo salvaje y un terminador adecuado para los genes *UGT1* y *CepB*, respectivamente, seguido del acoplamiento de la secuencia de promotor-terminador del gen al marcador seleccionable *URA3*. Como terminador, se utilizaron tanto el terminador de *C. bombicola* *UGTB1* de tipo salvaje como el terminador de tirosina quinasa (TK). Todas las reacciones de PCR se realizaron con el sistema de PCR PfuUltra High Fidelity (Stratagene) a menos que se indique lo contrario. Como primer paso, el promotor *UGTB1* (P) delante del gen *UGT1* o *CepB* se amplificó a partir de los plásmidos pG_PUgt1T y PG_PCepBT, respectivamente, plásmidos derivados de pGEM_T® que albergan el gen de interés entre las secuencias de promotor y terminador de *UGTB1* nativas. Los pares de cebadores fueron MDR505Rev y Ugt1 1737Rev_Sall para el fragmento PUgt1 de 2736 pb y MDR505Rev y CepB 2355Rev_Sall para el fragmento PCepB de 3354 pb (Tabla 9). El terminador de TK de 134 pb se amplificó a partir de un plásmido que contenía el marcador de resistencia a higromicina entre el terminador de tirosina quinasa (TK) y el promotor GAPD de *C. bombicola* (Van Bogaert *et al.*, 2008) haciendo uso de los cebadores TK_F_Sall y TK_R_Mlul (Tabla 9). Todos los fragmentos obtenidos se purificaron por medio del kit de purificación de PCR Qiaquick (Qiagen) y se sometieron a una digestión durante la noche con *Sall* (New England Biolabs) de acuerdo con Sambrook (Sambrook y Russell, 2001), pero con una adición secundaria de 20U de enzima de restricción después de 2 horas de incubación. Después de la purificación de los fragmentos digeridos por medio del kit de limpieza de reacción MinElute de Qiagen, los dos fragmentos PCepB y PUgt1 se ligaron al terminador de TK digerido. Para ello, se añadieron 2 U de T4 ADN ligasa de fermentas a 100 ng de los fragmentos PCepB y PUgt1 y se añadió una cantidad adecuada de fragmento de TK purificado de manera que se obtuvo una relación de gen:terminador de 3:1. La mezcla se incubó durante la noche a 22°C. Posteriormente, los productos de ligamiento se amplificaron a partir de las mezclas de reacción por medio de los cebadores MDR505Rev y TKR_Mlul, proporcionando el fragmento de 2856 pb PUgt1 TK y de 3474 pb PCepBTK, respectivamente. Después de la purificación, los fragmentos se clonaron en el Vector de Clonación pJET1.2/blunt utilizando el kit de Clonación por PCR ClonJet™ de Fermentas y resultando los plásmidos pJ_PUgt1TK y pJ_PCepBTK para la transformación de células competentes de *E. coli* Fusión Blue y XL10Gold, respectivamente, de acuerdo con Sambrook (Sambrook y Russell, 2001). Los transformantes correctos se identificaron mediante PCR de colonias y los plásmidos derivados se enviaron para la secuenciación. Para acoplar el marcador seleccionable *URA3*, las inserciones de los plásmidos pJ_PUgt1TK y pJ_PCepBTK se amplificaron nuevamente mediante los cebadores MDR505Rev y TKR_Mlul, se purificaron y se sometieron a una digestión durante la noche con *Mlul* (New England Biolabs) de acuerdo con Sambrook (Sambrook y Russell, 2001), pero con una adición secundaria de 20 U de enzima de restricción después de 2 horas de incubación. Por consiguiente, el marcador seleccionable *URA3*, además del extremo 3' *UGTB1* de *C. bombicola* se amplificó a partir del plásmido pG_KOugtB1 (véase el Ejemplo 3) por medio de los cebadores URA3 677F_Mlul (Tabla 9) y GTII + 239Rev (Tabla 4), se purificó y se sometió a la misma digestión de *Mlul* durante la noche. Después de la purificación de los fragmentos PUgt1TK, PCepBTK y URA3GT2T digeridos, los genes de interés se acoplaron al marcador de selección. Para ello, se añadieron 100 ng del fragmento PUgt1 TK y PCepBTK, respectivamente a una cantidad apropiada de fragmento URA3GT2T, de modo que se obtuvo una relación gen:marcador de 3:1 y se realizó el ligamiento durante la noche a 22°C con 2 U de T4 ADN ligasa (Fermentas). Los productos de ligamiento finales PUgt1TK_URA3GT2T (5856 pb) y PCepBTK_URA3GT2T (6474 pb) se amplificaron a partir de las mezclas de ligamiento por medio del sistema de PCR Expand Long Template de Roche y los cebadores MDR505Rev y GTII + 239Rev.

- 55 Alternativamente, para crear casetes de expresión, en donde el gen de interés es seguido por el terminador nativo de *C. bombicola* *UGTB1* en lugar del terminador TK, las inserciones completas de los plásmidos pG_PUgt1T y PG_PCepBT se amplificaron por medio del par de cebadores MDR505Rev/PCepBT214_R_Mlul, proporcionando los fragmentos de 2967 pb PUgt1T y 3595 pb de PCepBT, respectivamente. El marcador seleccionable de *URA3*, seguido por el extremo 3' de *UGTB1*, pero sin secuencia de terminador, se amplificó a partir del plásmido

pG_KOugtB1 con los cebadores URA3677 F_Mlul y GTII + 1296Rev (Tabla 9). Los productos de PCR se purificaron, se digirieron durante la noche con *Mlul* y se ligaron tal como se describió anteriormente. Los casetes de expresión PUgt1T_URA3GT2 y PCepBT_URA3GT2 se amplificaron a partir de las mezclas de ligamiento por medio del sistema de PCR Expand Long Template de Roche utilizando los cebadores MDR505Rev y GTII + 1296Rev. Los cuatro casetes de expresión obtenidos se purificaron en gel utilizando el kit Qiaquick Gelextraction de Qiagen y se clonaron en pGEM-T® haciendo uso del sistema pGEM-T® Vector (Promega). Los plásmidos se utilizan para la transformación de células ultracompetentes de *E. coli* y los transformantes correctos se identifican por medio de PCR de colonias. Los plásmidos derivados se envían para la secuenciación.

Creación del mutante productor de celobiosalípidos

Los casetes de expresión lineal PUgt1TK_URA3GT2T, PCepBTK_URA3GT2T, PUgt1T_URA3GT2 y PCepBT_URA3GT2 se amplifican a partir de los plásmidos derivados de pGEM-T® haciendo uso del cebador MDR505Rev en combinación con GTII +239Rev o GTII +1296Rev, dependiendo de la presencia o ausencia del terminador *UGTB1* en el extremo 3' del casete. Los casetes se purifican y se utilizan para la transformación por electroporación de la cepa G9 de *C. bombicola* deficiente en *ura3*. Para ello, G9 de *C. bombicola* se cultiva durante la noche en 100 mL de YPD y cuando la DO alcanza 1, las células se recogen a partir de 50 mL por centrifugación durante 5 min a 4°C y 4300 g. Después de lavar dos veces con agua mQ fría y estéril, las células se resuspenden en 2 mL de solución de sorbitol estéril (1 M). Después de la centrifugación a 4°C y 4300 g, las células se resuspenden en 2 ml de acetato de litio estéril (0,1 M) en presencia de DTT 2,5 mM y se dejan reposar a temperatura ambiente durante 10 a 15 min. Las células se recogen entonces de nuevo y se lavan con 2 mL de sorbitol (1 M) antes de resuspender en 250 µl de sorbitol (1 M). A partir de esta suspensión, 50 µl se transfieren a un tubo eppendorf estéril, se añaden 500 ng-1 µg del casete de expresión lineal y la mezcla se incuba en hielo durante 2 min antes de transferirla a una cubeta de electroporación de 2 mm. Un pulso de 1,5 kV (200 Ohm) se administra durante 5 milisegundos y se añade 1 mL de YPD fría y estéril inmediatamente. Las células se incuban durante 1 h a 30°C y 200 rpm y se recogen por centrifugación a temperatura ambiente durante 5 min a 1500 g. Las células se resuspenden luego en 1 mL de sorbitol (1 M) y se extienden partes alícuotas de 200 µl en placas SD selectivas. Las placas se incuban a 30°C hasta que aparecen las colonias transformantes.

Caracterización de mutantes productores de celobiosalípidos

Colonias mutantes se verifican primero para la integración correcta del casete de expresión en el locus *UGTB1* del genoma por medio de PCR de colonias de levaduras. Los transformantes correctos, que aparecen a partir de eventos de doble cruzamiento, se transfieren a medio 3C que contiene 10% de glucosa, 1% de extracto de levadura, 0,1% de urea y 2% de agarosa antes de la inoculación al medio de producción líquido descrito por Lang *et al.* (2000) con el fin de verificar la producción de glicolípidos. Los medios líquidos se incuban a 30°C y 200 rpm durante dos días antes de la adición de aceite de colza (37,5 g/L). *C. bombicola* de tipo salvaje ATCC 22214 sirve como referencia. Una semana después de la adición de aceite de semilla de colza, se extraen glicolípidos de muestras de 1 ml de cultivo por medio de 400 µl de acetato de etilo en presencia de 10 µl de ácido acético. Las fracciones de acetato de etilo se analizan en HPLC-ELSD y LC-MS tal como se describe en el ejemplo 3. Para verificar la estructura molecular de los celobiosalípidos producidos, los extractos de los mutantes se rastrean en cuanto a compuestos con masa molecular en el intervalo de 600-800 (m/z).

Tabla 1. Cebadores utilizados para eliminar el gen *CYP52M1* de *C. bombicola*. Todos los cebadores se obtuvieron de Sigma Genosys.

Nombre	Característica	Secuencia
A21TotFor	clonación <i>CYP52M1</i>	CTGAGTGATAGGTTGAGCATTAG
A21TotRev	clonación <i>CYP52M1</i>	GCTCTTGTTCCGGTACTCTTATTG
GHInfA21For	Marcador de selección ligante en <i>CYP52M1</i>	GCTAAAGTTACCCGA-
HygroInfA21Rev	marcador de selección ligante en <i>CYP52M1</i>	CCAATGGCAGTGGCTTACCACTC GATCCTTCTGCTCGG-
A21KnockHygroCasFor	fragmento de inactivación por amplificación	CCGCGTTTATGAACAACGAGCC
A21KnockHygroCasRev	fragmento de inactivación por amplificación	GAGTCGGGCGTTATTTCTCC
HygroInsertCheckFor	comprobar genotipo de inactivación	AATCCATAAACGACTACTC
Ura3outEndfor	comprobar genotipo Cm2	TTCGACAGCGTCTCCGACCT
ATqRev	comprobar genotipo Cm2	TAAAGAAACGAAGGGCCCAGCAGTC CACCACAGTACGAGGAGGAGGAACA

Tabla 2. Cebadores utilizados para el aislamiento del gen *UGTA1* y construcción del casete de inactivación. Todos los cebadores se obtuvieron de Sigma Genosys.

Nombre	Característica	Secuencia
UDPGTA1 DS GSP1	Cebador GSP 3' primario	5' CAGCAGAGACCATCTGCCTACAACCTTC 3'
UDPGTA1 DS GSP2	Cebador GSP 3' anidado	5' CAACGCCCAAGCACCGAACTCAATTCAC 3'
UDPGTA1 TotF	Cebador directo Alta Fidelidad	5' GAAGATACGTCCGTGCTTTG 3'
A1P RevNheI	Cebador inverso Alta Fidelidad	5' CATGGCTAGCCGGGCATTATATGGCCTG 3'
A1T ForNheI	Cebador directo Alta Fidelidad	5' CATGGCTAGCCGCTATGAACACGCTCTTG 3'
A1T Rev	Cebador inverso Alta Fidelidad	5' CATGACAGCCTTTTCTTCTT 3'
Ura3 FbisNheI	Cebador directo Alta Fidelidad	5' CATGGCTAGCCTGACGGGCGGATAGTACAG 3'
Ura3 RbisNheI	Cebador inverso Alta Fidelidad	5' CATGGCTAGCGTCATCAACTCCATGGCGTGAGG 3'

Tabla 3. Diez mejores puntuaciones de homología para la secuencia *UGTA1* traducida

Gen	Organismo	Nº Acc. NCBI	% Id	Puntuación E
Proteína de la familia glicosiltransferasa	<i>Mycobacterium vanbaalenii</i>	YP_95481 7	35	8 e-55
UDP-glucuronosil/ glicosiltransferasa	<i>Mycobacterium radiotolerans</i>	YP_00175 4776	35	2 e-54
Proteína de la familia glicosiltransferasa	<i>Mycobacterium gilvum</i>	YP_00113 3857	33	2 e-53
Familia MGT de glicosiltransferasa	<i>Mycobacterium sp.</i>	YP_64073 6	32	2 e-52
Familia MGT de glicosiltransferasa	<i>Mycobacterium sp.</i>	YP_00107 1846	32	3 e-52
UDP-glucuronosil/glicosiltransferasa	<i>Mycobacterium vanbaalenii</i>	YP_95566 5	34	4 e-52
UDP-glucuronosil/glicosiltransferasa	<i>Acidovorax avenae</i>	YP_97296 8	31	3 e-50
UDP-glucuronosil/glicosiltransferasa	<i>Burkholderia ambifaria</i>	YP_77811 9	31	3 e-47
UDP-glucuronosil/glicosiltransferasa	<i>Methylocella silverstris</i>	YP_00236 4149	31	1 e-46
UDP-glucuronosil/glicosiltransferasa	<i>Mycobacterium gilvum</i>	YP_00113 3117	32	2

Tabla 4. Cebadores utilizados para el aislamiento del gen *UGTB1* y la construcción del casete de inactivación. Todos los cebadores se obtuvieron de Sigma Genosys.

Nombre	Característica	Secuencia
GTII -472For	Cebador directo Alta Fidelidad	5' GAGAGTGGGACCTGATTC 3'
GTII +239Rev	Cebador inverso Alta Fidelidad	5' CTGCTCTCAACACCGAGTGTAG 3'
Ura3inf ugtB1 F	Cebador directo Alta Fidelidad	5' AAGCAGAGAAGGCGCGATAGTACAGGCTTT GCC 3'
Ura3inf ugtB1 R	Cebador inverso Alta Fidelidad	5' CCTTCGTGGCCCCGATCATCGTCACTATAC ACATCG 3'
KOugtB1 Ctrl F	Cebador directo de control	5' AAGCCAAAATCAGAGAGTG 3'
KOugtB1 Ctrl R	Cebador inverso de control	5' GGTTCGTGCGAACTGGTATG 3'

Tabla 5. Diez mejores puntuaciones de homología para la secuencia *UGTB1* traducida

Gen	Organismo	Nº Acc. NCBI	% Id	Puntuación E
UDP-glucuronosil/glicosiltransferasa	<i>Mycobacterium radiotolerans</i>	YP_001754 776	32	1 e-15
UDP-glucuronosil/glicosiltransferasa	<i>Burkholderia ambifaria</i>	YP_778119	31	5 e-15
Proteína de la familia glicosiltransferasa	<i>Mycobacterium vanbaalenii</i>	YP_954817	32	2 e-14
UDP-glucuronosil/glicosiltransferasa	<i>Acidovorax avenae</i>	YP_972968	28	1 e-13
UDP-glucuronosil/glicosiltransferasa	<i>Methylocella silverstris</i>	YP_002364 149	28	2 e-13
Familia MGT de glicosiltransferasa	<i>Pectobacterium wasabiae</i>	YP_003258 26	29	1 e-12
Proteína de la familia glicosiltransferasa	<i>Mycobacterium gilvum</i>	YP_001133 857	31	2e-11
Familia 28 de glicosiltransferasa	<i>Methylobacterium nodulans</i>	YP_002500 506	28	1 e-10
Familia MGT de glicosiltransferasa	<i>Pectobacterium carotovorum</i>	YP_003018 782	28	2 e-10
Familia MGT de glicosiltransferasa	<i>Mycobacterium sp.</i>	YP_001071 846	26	7 e-10

Tabla 6. Cebadores utilizados para crear pGEM-T_cassette_yEGFP. Los caracteres en **negrita** representan extensiones no vinculantes.

Nombre	Característica	Secuencia
P1_FOR_URA3v	Recoge gen ura3 +5' UTR	AGAACAAGGCCGAGTATGTC
P2_REV_URA3v	Recoge gen ura3 +3' UTR	TGCCAGCAGATCATCATCAC
	Cebador de solapamiento	GGATCCCCGCAGGGC ATGCAACTTGCACATGAA TACC
P3_FOR_URA3t_exty EGFP	ura3-yEGFP	
P4_REV_URA3t_extN OTI	Amplificación ura3-yEGFP	TAGCGGCCGCGTC CAGATTAGCCTCCGACATAG
P5_FOR_yEGFP_extM CS3	Amplificación ura3-yEGFP	GCACTAGTATACCCGGGCGCCT- CAGCTCTTCGATGTCTAAAGGTGAAGAAT
	Cebador de solapamiento	TATTCATGTGCAAGTTGCATGCCCTGGGGGAT CCATACG
P6_REV_yEGFP_extU RA3t	ura3-yEGFP	
P7_FOR_QCSaplpGE M-t	Cebador de mutagénesis	CGTATTGGGCGCT CT CCGCTT CT CGCTCACTG ACTC
P8_REV_QCSaplpGE M-t	Cebador de mutagénesis	GAGTCAGTGAGCGAGGAAGCGGAG G AGCGCCC AATACG
P18_REV_GAPDprom	Recoge pGAPD	TTCACCTTTAGACATTTGTGTAGAGTTGTTTTTG
P19_FOR_pGAPD156 0	Recoge pGAPD	CACTAGTATACCCGGGACATCCGATGTGTAGTTA
P37_FOR_checkGFP	Cebador de PCR de colonias	GGTTGAATTAGATGGTGATGTTAATG
P35_REV_checkcasIN	Cebador de PCR de colonias	GAGCTCAAGACGCGTTTACTCAATGC
P31_REV_cassette	Casete de amplificación	GCGTCAGATTAGCCTCCGACATAG

Tabla 7. Cebadores utilizados para crear el casete de expresión de amilasa y el control de transformantes de amilasa.

Nombre	Característica	Secuencia
sAmyAOfw	Construcción de casete de control	GGTAGCAGCGTTGATTACTC
sAmyAOrv	Construcción de casete de control	ATCTGTGCCCTTACGCATAG
sAmyAOfw	Comprobar la integridad de la región de codificación amy	GGTAGCAGCGTTGATTACTC
sAmyAOrv	Comprobar la integridad de la región de codificación amy	ATCTGTGCCCTTACGCATAG
sAmyAOfw2	Comprobar el casete de integración	CCGACAGCGAGCTGTACAAG
P35	Comprobar el casete de integración	GAGCTCAAGACGCGTTTACTCAATGC

Tabla 8. Cebadores utilizados para crear el casete de expresión de PHA. Todos los cebadores se obtuvieron de Sigma Genosys.

Nombre	Característica	Secuencia
P53_FOR_downCYP_extSpeI2	clonar <i>región cyp52M1 aguas abajo</i>	T T A C T A G T G T T T C T T A G C C T C C C A T G
P54_REV_downCYP_extNotI2	clonar <i>región cyp52M1 aguas abajo</i>	G A A G A A A C G A A T T G G C C T T G C G G C C G C G G T G T C
P55_FOR_upCYP_extNheI	construcción sintética de amplificación	G A C T C G C C A A A T T C C A T C G T T G C T A G C T C T C G G C A G A T T T C C T
P58_REV_PHAC1+tCYP_extEcoRI	construcción sintética de amplificación	T G A G A A T T C G T C G G T T A A A C G C A C T C C
P63_FOR_cassPHAC1	amplificación del casete de expresión de PHA	T T C A C T C T C G G C A G A T T T C C T T G T G
P64_REV_cassPHAC1	amplificación del casete de expresión de PHA	G G T G T C G A C T C G C C A A A T T C
P9_FOR_seqQCSapI_URA3downA21TotRev	verificar casete de integración	G C A C A C T T C A A C C T T C C T A C
UDPGTA1R	verificar casete de integración cebador de secuenciación	G C T C T T G T T C G G T A C T C T T A T T G C C T A C C T C T C T T C C C T G A T C T

Tabla 9. Cebadores utilizados para la creación de los casetes de expresión *UGT1* y *CepB*. Todos los cebadores se obtuvieron de Sigma Genosys

Nombre	Característica	Secuencia
MDR 505Rev	Cebador directo Alta Fidelidad	5' C C T C G C C A C C A C C T A G T T T G 3'
Ugt1 1737Rev_Sall	Cebador inverso Alta Fidelidad	5' G A T C G T C G A C T C A A A A G A G G C G G A C T T C T G C C 3' 5' G A T C G T C G A C T C A T C C C A T T A T A A C A A C A C
CepB 2355Rev_Sall	Cebador inverso Alta Fidelidad	3' 5' G A T C G T C G A C T C A T C C C A T T A T A A C A A C A C
TK F_Sall	Cebador directo Alta Fidelidad	3' 5' A A T T G T C G A C G G G A G A T G G G G G A G G C T A A C
TK R_Sall	Cebador inverso Alta Fidelidad	3' 5' G A G T A C G C G T T G A A C A A A C G A C C C A A C A C C
URA3 677F_MluI	Cebador directo Alta Fidelidad	3' 5' G A G A A C G C G T G A T A G T A C A G G C T T T G C 3'
PCepBT214 R_MluI	Cebador inverso Alta Fidelidad	5' C A T A A C G C G T T T C T G C T C T C A A C A C C G A G 3'
GTII +1296Rev	Cebador inverso Alta Fidelidad	5' A G A A G C T A A T T C A C T A A T T G C C G A C 3'

5 Referencias:

Asmer H.J., Lang S., Wagner F., Wray V. (1988). Microbial-Production, Structure Elucidation and Bioconversion of Sophorose Lipids. *J. Am. Oil Chem. Soc.*, 65: p. 1460-1466.

10 Basehoar A.D., Zanton S.J., Pugh F. (2004). Identification and Distinct Regulation of Yeast TATA Box-Containing Genes. *Cell*, 116: p. 699-709.

Bisht K.S., Gross R.A., Kaplan D.L. (1999). Enzyme-mediated regioselective acylations of sophorolipids. *J. Org. Chem.*, 64: p. 780-789.

15 Brakemeier A., Wullbrandt D., Lang S. (1998). *Candida bombicola*: production of novel alkyl glycosides based on glucose/2-dodecanol. *Appl. Microbiol. Biotechnol.*, 50: p. 161-166.

Breithaupt T.B., Light R.J. (1982). Affinity chromatography and further characterization of the glucosyltransferases involved in hydroxydocosanoic acid sophoroside production in *Candida bogoriensis*. *J. Biol. Chem.*, 257: p. 9622-9628.

Bucholtz, M.L., Light R.J. (1976). Acetylation of 13-sophorosyloxydocosanoic Acid by an Acetyltransferase Purified from *Candida bogoriensis*. *J. Biol. Chem.*, 251: p. 424-430.

20 Campbell J.A., Davies G.J., Vulone V., Henrissat B. (1997). A classification of nucleotide-diphospho-sugar glucosyltransferases based on amino acid sequence similarities. *Biochem. J.*, 326: p. 929-942.

- Carr J.A., Bisht K.S. (2003). Enzyme-catalyzed regioselective transesterification of peracylated sophorolipids. *Tetrahedron*, 59: p. 7713-7724.
- Casas J.A., Garcia-Ochoa F. (1999). Sophorolipid production by *Candida bombicola*: Medium composition and culture methods. *J. Biosci. Bioeng.*, 88: p. 488-494.
- 5 Cavalero D.A., Cooper D.G. (2003). The effect of medium composition on the structure and physical state of sophorolipids produced by *Candida bombicola* ATCC 22214. *J. Biotechnol.*, 103: p. 31-41.
- Chen J., Song X., Zhang H., Qu Y.B., Miao J.Y. (2006). Production, structure elucidation and anticancer properties of sophorolipid from *Wickerhamiella domercqiae*. *Enzyme Microb. Technol.*, 39: p. 501-506.
- 10 Cormack B.P., Bertram G., Egerton M., Gow N.A., Falkow S., Brown A.J. (1997a). Yeast-enhanced green fluorescent protein (yEGFP): A reporter of gene expression in *Candida albicans*. *Microbiol-UK*, 143: p. 303-311.
- Coutinho P.M., Deleury E., Davies G.J., Henrissat B. (2003). An evolving hierarchical family classification for glycosyltransferases. *J. Mol. Biol.*, 328: p. 307-317.
- Daniel H.J., Otto R.T., Reuss M., Sylдатk C. (1998a). Sophorolipid production with high yields on whey concentrate and rapeseed oil without consumption of lactose. *Biotechnol. Lett.*, 20: p. 805-807.
- 15 Daniel H.J., Reuss M., Sylдатk C. (1998b). Production of sophorolipids in high concentration from deproteinized whey and rapeseed oil in a two stage fed batch process using *Candida bombicola* ATCC 22214 and *Cryptococcus curvatus* ATCC 20509. *Biotechnol. Lett.*, 20: p. 1153-1156.
- Davila A.M., Marchal R., Vandecasteele J.P. (1992). Kinetics and balance of a fermentation free from product inhibition-sophorose lipid production by *Candida bombicola*. *Appl. Microbiol. Biotechnol.*, 38: p. 6-11.
- 20 De Maeseneire S. L., De Groeve M. R. M., Dauvrin T., De Mey M., Soetaert W.K., Vandamme E.J. (2006). Cloning, sequence analysis and heterologous expression of the *Myrothecium gramineum* orotidine-5'-phosphate decarboxylase gene. *FEMS Microbiol. Lett.* 261: p. 262-271.
- Esders T.W., Light R.J. (1972). Glucosyl- and acetyltransferases involved in the biosynthesis of glycolipids from *Candida bogoriensis*. *J. Biol. Chem.*, 247: p. 1375-1386.
- 25 Gietz R.D., Schiestl R.H. (1995). Transforming yeast with DNA. *Methods Mol. Cell Biol.*, 5: p. 255-269.
- Gorin P.A.J., Spencer J.F.T., Tulloch A.P. (1961). Hydroxy fatty acid glycosides of sophorose from *Torulopsis magnoliae*. *Can. J. Chem.*, 39: p. 846-855.
- Gross R., Shah V. (2004). Antimicrobial properties of various forms of sophorolipids. *Patente Internacional WO 2004/044216 A1*.
- 30 Gross R., Shah V. (2005). Antifungal properties of various forms of sophorolipids *Patente de EE.UU. US 2005/0164955 A1*.
- Guo Z., Sherman F. (1996). Signals sufficient for 3'-End Formation of Yeast mRNA. *Mol. Cell. Biol.*, 16: p. 2772-2776.
- Hall T.A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for windows 95/98/NT. *Nucl. Acids. Symp. Ser.* 41: p. 95-98.
- 35 Heward S., Josephs K., Bölker M. (2005). Genetic Analysis of Biosurfactant Production in *Ustilago maydis*. *Appl. Environ. Microbiol.*, 71: p. 3033-3040.
- Heward S., Linne U., Scherer M., Marahiel M.A., Kämper J., Bölker M. (2006). Identification of a Gene Cluster for Biosynthesis of Mannosylerythritol Lipids in the Basidiomycetous Fungus *Ustilago maydis*. *Appl. Environ. Microbiol.*, 72: p. 5469-5477.
- 40 Hu Y., Ju L.K. (2003). Lipase-mediated deacetylation and oligomerization of lactonic sophorolipids. *Biotechnol. Progr.*, 19: p. 303-311.
- Imura T., Masuda Y., Minamikawa H., Fukuoka T., Konishi M., Morita T., Sakai H., Abe M., Kitamoto D. (2010). Enzymatic Conversion of Diacetylated Sophorolipid into Acetylated Glucoselipid: Surface-Active Properties of Novel Bolaform Biosurfactants. *J Oleo Sci.* 59: p. 495-501.
- 45 Inoue S., Ito S. (1982). Sophorolipids from *Torulopsis bombicola* as microbial surfactants in alkane fermentations. *Biotechnol. Lett.*, 4: p. 3-8.
- Ito S., Inoue S. (1982). Sophorolipids from *Torulopsis bombicola* - Possible relation to alkane uptake. *Appl. Environ. Microbiol.*, 43: p. 1278-1283.
- 50 Kim Y-B, Yun H.S., Kim E-K (2009). Enhanced sophorolipid production by feeding-rate-controlled fed-batch culture. *Bioresour. Technol.*, 100: p. 6028-6032.
- Kulakovskaya T.V., Shashkov A.S., Kulakovskaya E.V., Golubev W.I. (2004). Characterisation of an antifungal glycolipid secreted by the yeast *Syngliomyces paphiopedili*. *FEMS Yeast Res.*, 5: p. 247-252.
- Kulakovskaya T.V., Shashkov A.S., Kulakovskaya E.V., Golubev, W.I. (2005). Ustilagoc acid secretion by *Pseudozyma fusiformata* strains. *FEMS Yeast Res.*, 5: p. 919-923.
- 55 Kurtzman C.P., Price N.P.J., Ray K.J., Kuo T.M. (2010) Production of sophorolipid biosurfactants by multiple species of the *Starmerella (Candida) bombicola* yeast clade. *FEMS Microbiol. Lett.*, 311: p. 140-146.
- Lang S., Brakemeier A., Heckmann R., Spockner S., Rau U. (2000). Production of native and modified sophorose lipids. *Chim Oggi-Chem Today*, 18: p. 76-79.
- 60 Marchler-Bauer A. et al. (2009). CDD: Specific functional annotation with the Conserved Domain Database. *Nucleic Acids Res.*, 37 (D): p. 205-210.
- Marzluf G.A. (1997). Genetic Regulation of Nitrogen Metabolism in the Fungi. *Microbiol. Mol. Biol. Rev.*, 61: p. 17-32.
- Magasanik B., Kaiser C.A. (2002). Nitrogen regulation in *Saccharomyces cerevisiae*. *Gene*, 290: p. 1-18.

- Ochsner U.A., Fiechter A., Reiser J. (1994a). Isolation, characterization, and expression in *Escherichia coli* of the *Pseudomonas aeruginosa* rhlAB genes encoding a rhamnosyltransferase involved in rhamnolipid biosurfactant synthesis. *J. Biol. Chem.*, 269(31): p. 19787-19795.
- 5 Ochsner U.A., Koch A.K., Fiechter A., Reiser J. (1994b). Isolation and characterisation of a regulatory gene affecting rhamnolipid biosurfactant synthesis in *Pseudomonas aeruginosa*. *J. Bacteriol.*, 176: p. 2044-2054.
- Palme O., Comanescu G., Stoineva I., Radel S., Benes E., Develter D., Wray V., Lang S. (2010). Sophorolipids from *Candida bombicola*: Cell separation by ultrasonic particle manipulation. *Eur. J. Lipid Sci. Technol.*, 112: p. 663-673.
- 10 Puchkov E.O., Zähringer U., Lindner B., Kulakovskaya T.V., Seydel U., Wiese A. (2002). The mycocidal, membrane-active complex of *Cryptococcus humicola* is a new type of cellobiose lipid with detergent features. *Biochim. Biophys. Acta*, 1558: p. 161-170.
- Rau U., Heckmann R., Wray V., Lang, S. (1999). Enzymatic conversion of a sophorolipid into a glucose lipid. *Biotechnol. Lett.*, 21: p. 973-977.
- Rau U., Hammen S., Heckmann R., Wray V., Lang S. (2001). Sophorolipids: a source for novel compounds. *Ind. Crop Prod.*, 13: p. 85-92.
- 15 Reichenbecher M., Lottspeich F., Bronnenmeier K. (1997). Purification and properties of a cellobiose phosphorylase (CepA) and a cellodextrin phosphorylase (CepB) from the cellulolytic thermophile *Clostridium stercorarium*. *Eur. J. Biochem.*, 247: p. 262-267.
- Rosa C., Lachance M. (1998). The yeast genus *Starmerella* gen. nov. and *Starmerella bombicola* sp. nov., the teleomorph of *Candida bombicola* (Spencer, Gorin & Tulloch) Meyer & Yarrow. *Int. J. Syst. Bacteriol.*, 48: p. 1413-1417.
- 20 Rosell C.M., Haros M., Escriva C., de Barber C.B. (2001) Experimental approach to optimize the use of α -amylases in breadmaking. *J. Agric. Food Chem.*, 49: p. 2973-2977.
- Roy I., Sastry M.S.R., Johri B.N., Gupta, M.N. (2000) Purification of α -amylase isoenzymes from *Scytalidium thermophilum* on a fluidized bed of alginate beads followed by concanavalin A-agarose column chromatography. *Protein Expr. Purif.* 20: p. 162-168.
- 25 Saerens K., Van Bogaert I., Soetaert W. (2009). Production of glucolipids and specialty fatty acids from sophorolipids by *Penicillium decumbens* naringinase : Optimization and kinetics. *Biotechnol. J.*, 4: p. 517-524.
- Sambrook J., Russell D.W. (2001). *Molecular Cloning: A Laboratory Manual*. 3rd ed. Cold Spring Harbour Laboratory Press, Cold Spring Harbour, New York.
- 30 Spencer J.F.T., Gorin P.A.J., Tulloch A.P. (1970). *Torulopsis bombicola* sp.n. *Antonie Van Leeuwenhoek* 36: p. 129-133.
- Spoeckner S., Wray V., Nimitz M., Lang S. (1999). Glycolipids of the smut fungus *Ustilago maydis* from cultivation on renewable resources. *Appl. Microbiol. Biotechnol.*, 51: p. 33-39. Stothard P. (2000). The Sequence Manipulation Suite: JavaScript programs for analyzing and formatting protein and DNA sequences. *Biotechniques*, 28: p. 1102-1104.
- 35 Teichmann B., Linne U., Hewald S., Marahiel M.A., Bölker M. (2007). A biosynthetic gene cluster for a secreted cellobiose lipid with antifungal activity from *Ustilago maydis*. *Mol. Microbiol.*, 66: p. 525-533.
- Thaniyavarn J., Chianguthai T., Sangvanich P., Roongsawang N., Washio K., Morikawa M., Thaniyavarn S. (2008). Production of sophorolipid biosurfactant by *Pichia anomala*. *Biosci. Biotechnol. Biochem.*, 72: p.2061-2068.
- 40 Tulloch A.P., Spencer J.F.T., Deinema M.H. (1968). A new hydroxy fatty acid sophoroside from *Candida bogoriensis*. *Can. J. Chem.*, 46: p. 345-348.
- Van Bogaert I.N.A., De Maeseneire S.L., De Schampelaire W., Develter D., Soetaert W., Vandamme E.J. (2007). Cloning, characterization and functionality of the orotidine-5'-phosphate decarboxylase gene (URA3) of the glycolipid-producing yeast *Candida bombicola*. *Yeast*, 24: p. 201-208.
- 45 Van Bogaert I.N.A., De Maeseneire S.L., Develter D., Soetaert W., Vandamme E.J. (2008a). Cloning and characterisation of the glyceraldehyde 3-phosphate dehydrogenase gene of *Candida bombicola* and use of its promoter. *J. Ind. Microbiol. Biotechnol.*, 35: p. 1085-1092.
- Van Bogaert I.N.A., De Maeseneire S.L., Develter D., Soetaert W., Vandamme E.J. (2008b). Development of a transformation and selection system for the glycolipid producing yeast *Candida bombicola*. *Yeast*, 25: p. 272-278.
- 50 Van Bogaert I.N.A., De Mey M., Develter D., Soetaert W., Vandamme E.J. (2009). Importance of the cytochrome P450 monooxygenase CYP52 family for the sophorolipid-producing yeast *Candida bombicola*. *FEMS Yeast Res.*, 9: p. 87-94.
- Zhou S.Q., Xu C., Wang J., Gao W., Akhverdiyeva R., Shah V., Gross R. (2004). Supramolecular assemblies of a naturally derived sophorolipid. *Langmuir*, 20: p. 7926-7932.
- 55

REIVINDICACIONES

1. Uso de una cepa de levadura modificada para la producción de compuestos, en donde dicha cepa de levadura modificada pertenece a una especie de levadura capaz de producir soforolípidos, caracterizado por que dicha cepa de levadura modificada tiene, comparado con la cepa de tipo salvaje no modificada que es totalmente capaz de producir dichos soforolípidos, al menos un gen que codifica una enzima seleccionada del grupo que consiste en monooxigenasa del citocromo P450 o una glucosiltransferasa inactivada a través de la inserción de un casete de inactivación, en donde dicha inactivación provoca una reducción en su capacidad de producir dichos soforolípidos de al menos 95%, comparado con la producción total de soforolípidos por unidad de tiempo en dicha cepa de tipo salvaje no modificada, y en donde dichos soforolípidos están constituidos por el azúcar soforosa unida a un ácido graso C₁₆, C₁₈, C₂₂ o C₂₄ hidroxilado.
2. Uso de acuerdo con la reivindicación 1, en donde dicho gen que codifica una monooxigenasa del citocromo P450 es el gen *CYP52M21* con el número de acceso de Genbank EU552419, y en donde dicho gen que codifica una glucosiltransferasa es el gen *UGTA1* que tiene una secuencia tal como se representa por SEQ ID N° 1 y que tiene un número de acceso de Genbank HM440973 o es el gen *UGTB1* que tiene una secuencia tal como se representa por SEQ ID N° 3 y que tiene un número de acceso de Genbank HM440974.
3. Uso de acuerdo con cualquiera de las reivindicaciones 1 o 2, en donde dicha reducción en su capacidad de producir soforolípidos es 100%.
4. Uso de acuerdo con cualquiera de las reivindicaciones 1 a 3, en donde dicha especie de levadura se selecciona del grupo que consiste en *Candida bombicola*, *Candida apicola*, *Candida batistae*, *Candida floricola*, *Candida riodecensis*, *Candida stellata*, *Candida* sp. NRRL Y-27208, *Rhodotorula bogoriensis*, *Pichia anomala* PY1, *Wickerhamiella domericqiae* y especies productoras de soforolípidos del clado Starmerella.
5. Uso de acuerdo con cualquiera de las reivindicaciones 1 a 4, en donde dicha cepa de tipo salvaje no modificada es la cepa *Candida bombicola* ATCC 22214.
6. Uso de acuerdo con cualquiera de las reivindicaciones 1 a 5, en donde dichos compuestos son proteínas recombinantes, ácidos beta-hidroxi grasos y polihidroxialcanoatos, ácidos dicarboxílicos, ácidos grasos poliinsaturados, ácidos grasos hidroxilados, glicolípidos, glucolípidos, trehalosalípidos, ramnolípidos, soforolípidos con una cola ácido graso especial, soforolípidos con una cola ácido graso que oscila entre 10 y 15 átomos de carbono, soforolípidos con una cola ácido graso de 17 átomos de carbono, soforolípidos con una cola ácido graso que oscila entre 19 y 25 átomos de carbono, soforolípidos con una cola ácido graso ramificada, soforolípidos con múltiples colas ácido graso hidroxiladas, soforolípidos completamente lactonizados y soforolípidos completamente ácidos; ramnosa, soforosa, antibióticos policétidicos, estructuras lactónicas basadas en ácidos grasos, ácidos orgánicos, compuestos oleaginosos, compuestos hidrofóbicos, escualeno, vitamina D, resveratrol, esteroides y carotenoides.

Figura 2

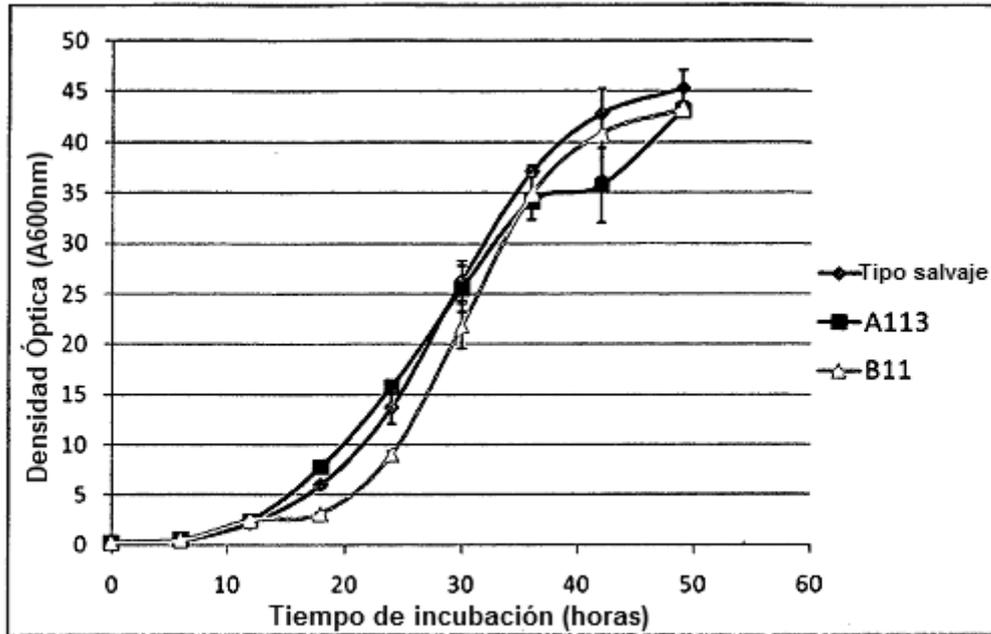


Figura 3

```

1  cagatatgca tcagggggcac agactaaaag ctgctctcag cgagtaccct tacctcttga
61  gaacctcaa aatttaccca gctgcagca tatcatgcac catgggtaaa ttcgaaatg
121 aattaccgg tggccttgaa ccacgttctc ccaattattt aaggcaataa cctgccactc
181 tcttgattg attaagaaag accttcaatt tagcttctcc ctacgaatat tcaatgagcc
                                     M S
241 cttcatcaca caaacccctg attctcgctt ggggettgcg tctttcaggc catataatgc
    P S S H K P L I L A C G L P L S G H I M
301 ccgttttgag tctggtacac ggccttacgg acgacggata egaagctact gttgtgacag
    P V L S L V H G L T D D G Y E A T V V T
361 gcagagcgtt tgaacaaaaa gttcgagatg tgggtgcaga ctttgttctc ttagaaggga
    G R A F E Q K V R D V G A D F V P L E G
421 acgcagattt tgatgaccac accttagacg atctggctcc gggccgtaaa gacatggccc
    N A D F D D H T L D D L V P G R K D M A
481 caagcttoga tegtacagtt caagatgtgg agcacatgat ggtagctact cttcctgagc
    P S F D R T V Q D V E H M M V A T L P E
541 agtttgccgc tattcagagg gctttcaaaa agctcagcgc aagcggecgc cctgtcgttc
    Q F A A I Q R A F K K L S A S G R P V V
601 ttgtcagtga agtgetgttt ttcggtgcac acctatcag cctcggtgct cctggtttca
    L V S E V L F F G A H P I S L G A P G F
    
```

Figura 3 cont.

661 aaccgctgg ctggatttgt ttaggggttt tgcctctttt gatccgcagt gatcatacct
 K P A G W I C L G V L P L L I R S D H T

721 taggacttga caacgacagg agccccgaag cacatgcaaa gaaactcgct atgaaccacg
 L G L D N D R S P E A H A K K L A M N H

781 ctcttgagca ccaaattttc gttaaagcca ctgctaagca caaggaaatc tgccgagagt
 A L E H Q I F V K A T A K H K E I C R E

841 taggttgacac tgaagatccc aaatttatct gggagcacag ttacattgct gcagacaagt
 L G C T E D P K F I W E H S Y I A A D K

901 tcctgcagct gtgcccgcct tctcttgagt tcagcagaga ccatctgcct agcaacttca
 F L Q L C P P S L E F S R D H L P S N F

961 aattcgccgg ctcaacgcc aagcaccgaa ctcaattcac cctcctctcc tgggtggggg
 K F A G S T P K H R T Q F T P P S W W G

1021 atgttctgag tgccaagcga gtcacatggt tcaactcaagg aacttttgct gtcagttaca
 D V L S A K R V I M V T Q G T F A V S Y

1081 agcatcttat tgtgcctact cttgaggcct tgaaggacga gcctgacact ttaacagtag
 K H L I V P T L E A L K D E P D T L T V

1141 ccatattggg ccgcccgggt gccaaagctac cggatgatgt tgtggttccct gagaatgctc
 A I L G R R G A K L P D D V V V P E N A

1201 gogtgatcga ctacttcaac tacgatgctc tacttctca cgttgatgct cttgtctaca
 R V I D Y F N Y D A L L P H V D A L V Y

1261 atggtggata tggcggactt cagcacagct taagccactc tgttccagtt gttattgctg
 N G G Y G G L Q H S L S H S V P V V I A

1321 gtgactctga agacaagcca atggtggcat cgagagctga ggccgctggc gtggcaattg
 G D S E D K P M V A S R A E A A G V A I

1381 atttgaaaac tggettgcct acagtggagc aaatcaaaga agctgttgat tcgataattg
 D L K T G L P T V E Q I K E A V D S I I

1441 gaaatccgaa attccacgaa gcctcgaaga aggttcaaat ggagttggaa agccacaact
 G N P K F H E A S K K V Q M E L E S H N

1501 ccttgaaaat tcttgaggaa agcatcgagg aaatcgccag ccatgacttt ggtcttttga
 S L K I L E E S I E E I A S H D F G L L

1561 ccaagagtga cgaggaaact gaagatatac ctgtcaaagg gccggcctta gcggtgagtt
 T K S D E E T E D I P V K G P A L A V S

1621 cttagaatcg tacga[ccaaa] tcagatcagg gaagagaggt agggtttttt ttatttatgt
 S -

1681 ctttgttttt attgattgaa atttacaata caacaaccat[caaa]ttaatt tgaacaaaca

1741 acaacacaca cacactgcaa ctt[caaaaa aataagta]

Figura 4

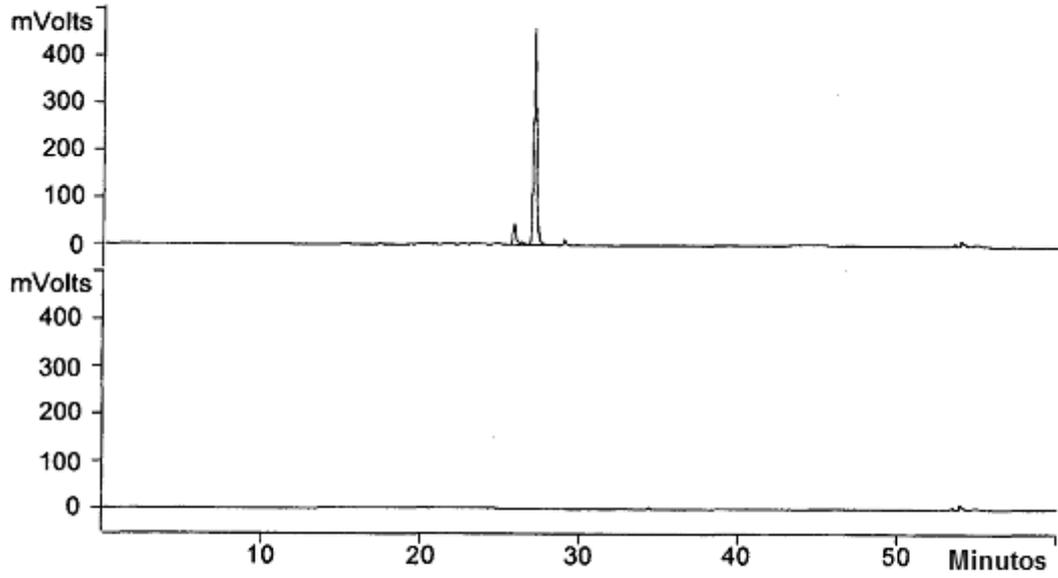


Figura 5

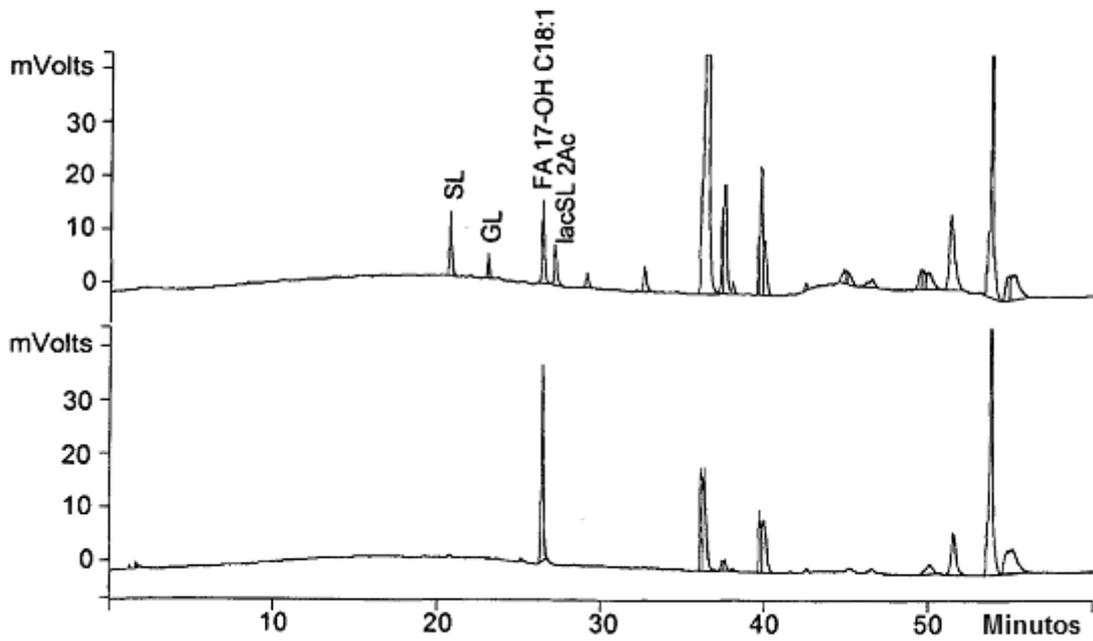


Figura 6

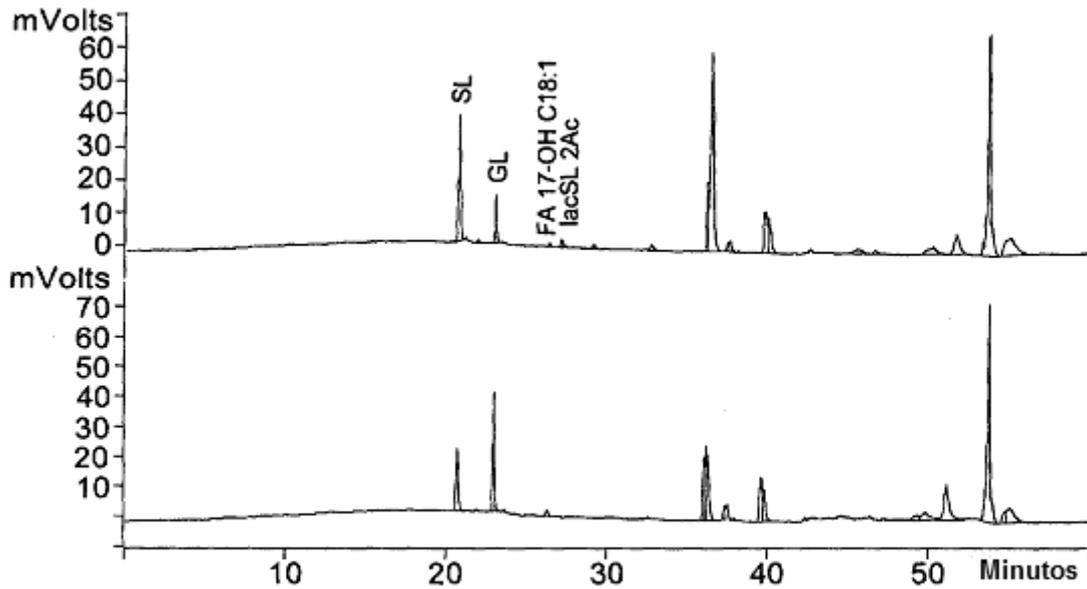


Figura 7

GTII -472For

```

1 gagagtggga cctgattcag aatcacacgg acccgtatat ataacaatca cttccaaca atatagcgag
71 tattaatata tttccgggta agggttgttc cggacttatg cattedaatca caggttgcat cagcetaata
141 tgtcagggcc gacggcgtaa attagaagg ttaggtcaag atccatcggc caggccaatg gagctctact
211 atgataggca gctgaagcga gacaagatat acttcagttg cgctctctga aaaaattatt ttgtgattct
281 cactcagttg agtgggcgac acacggaacc aataatctcg ccggaaggc ggctgaacat cagtcttgca
351 taagtgtgca agtgggcctg gcacagcgtg cattaccctt accatacatt cggggcaagt taaatccagc
421 attatataaa cttgattgac acaaatgggc ataaaaaat aaagtctcct atatggccat cgagaaacca
                                     M A I E K P
491 gtgatagttg cttgtgcctg cccactagcg gggcagctgg gccccagtgt cagcctggtc cgcggctctac
   V I V A C A C P L A G H V G P V L S L V R G L
561 tcaatagagg atatgaggtg actttcgtaa cagggaacgc attcaaggag aaagttattg aggcaggatg
   L N R G Y E V T F V T G N A F K E K V I E A G
631 cactttctgc cctctccaag gacgagctga ctaccatgaa tacaatctcc ctgaaatcgc tccaggattg
   C T F V P L Q G R A D Y H E Y N L P E I A P G L
701 ctcaagattc ctccaggcct tgagcagacc ggttactcaa tgaatgagat ttttgtgaag gcgattcctg
   L T I P P G L E Q T G Y S M N E I F V K A I P
771 agcagtaega tgcacttcaa actgctctaa aacaggttga ggctgaaaat aaatcagctg tggtgattgg
   E Q Y D A L Q T A L K Q V E A E N K S A V V I
841 cgagaccatg tttctagggg tgcateccgat atcactgggt gccccaggtc tcaagcccca aggcgtaatc
   G E T N F L G V H P I S L G A P G L K P Q G V I
911 acgttaggaa ctattccgtg catgctgaaa gcagagaagg cgccctggagt tcctagtctt gagccaatga
   T L G T I P C M L K A E K A P G V P S L E P M
    
```

Figura 7 cont.

981 ttgatacttt agtgcggcaa caagtatttc aaccaggaac tgactctgag aaggagatca tgaagacgct
 I D T L V R Q Q V F Q P G T D S E K E I M K T
 1051 cggggccacg aaggagcccg aatttctcct ggagaatata tacagcagcc ctgacagatt tttgcaactg
 L G A T K E P E F L L E N I Y S S P D R F L Q L
 1121 tgcctccat ctcttgaatt tcaactgact tgcctcctc ctggcttctc gttcgtggt agtgaccgcg
 C P P S L E P H L T S P P P G P S F A G S A P
 1191 atgtaaagtc tgctggatta gcaactccac ctcacctgcc gtcttgggtgg cctgatgtgc tgagtgcgaa
 H V K S A G L A T P P H L P S W W P D V L S A
 1261 gcgtctgatt gttgttacac aaggaacagc agccatcaac tatgaagatc tgctcattcc agcattgcag
 X R L I V V T Q G T A A I N Y E D L L I P A L Q
 1331 gcctttgctg acgaagaaga cactctcgta gttggatat tgggctcaa aggggctca ctctcctgata
 A P A D E E D T L V V G I L G V K G A S L P D
 1401 gcgttaaagt tcctgcaaac gctcgaattg ttgattattt tccttacgat gagctactac cgcctgctc
 S V K V P A N A R I V D Y P P Y D E L L P H A
 1471 tgttttcata tacaacggtg gatacggagg tctgcagcac agtttgagcc atggcgttcc cgtcatcatc
 S V P I Y N G G Y G G L Q H S L S H G V P V I I
 1541 ggaggaggaa tgttggtaga caagccagct gttgcttcac gagctgtatg ggctgggtgtt gtttatgatc
 G G G M L V D K P A V A S R A V W A G V G Y D
 1611 ttcaaacctt gcaggcaact tctgagctag tctccacggc cgttaaggag gtgttggtta ctccctcgta
 L Q T L Q A T S E L V S T A V K E V L A T P S
 1681 tcacgagaaa gccatggcag tcaagaaga gcttgaaaaa tacaagtctc ttgatattct agagtgcgca
 Y H E K A M A V K K E L E K Y K S L D I L E S A
 1751 attagtgaat tagcttctta acctggctct ttttctagat atgtctgctc cctgctcact gcttactggc
 I S E L A S -
 1821 ctaagctggt attacggacc ttaatcaagt atcaccccaa ggcaatcgag agtcttatcg agtctctagg
 1891 tagatagata cacgttttga tttttcggcc cactttgtag aaaaatctca gtgatttcat ggaattcagt
 1961 tacaaatact aatctgataa accaagaact acactcggtg ttgagagcag

GTII +239Rev

Figura 8

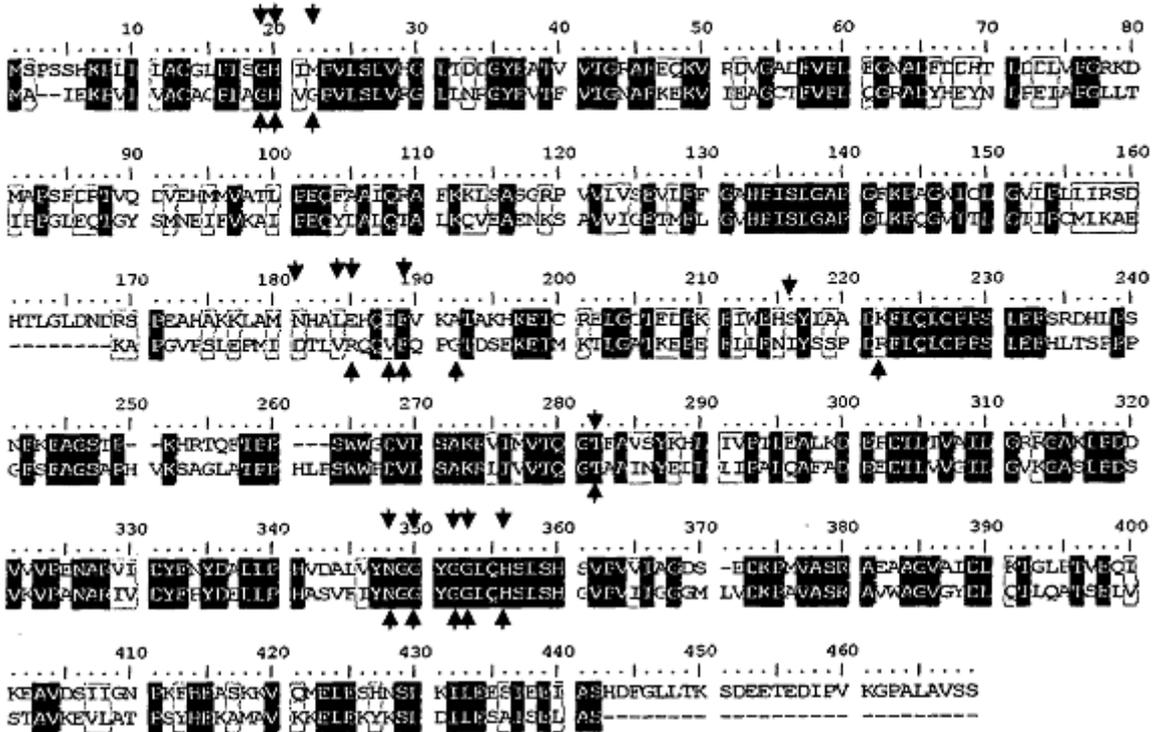


Figura 9

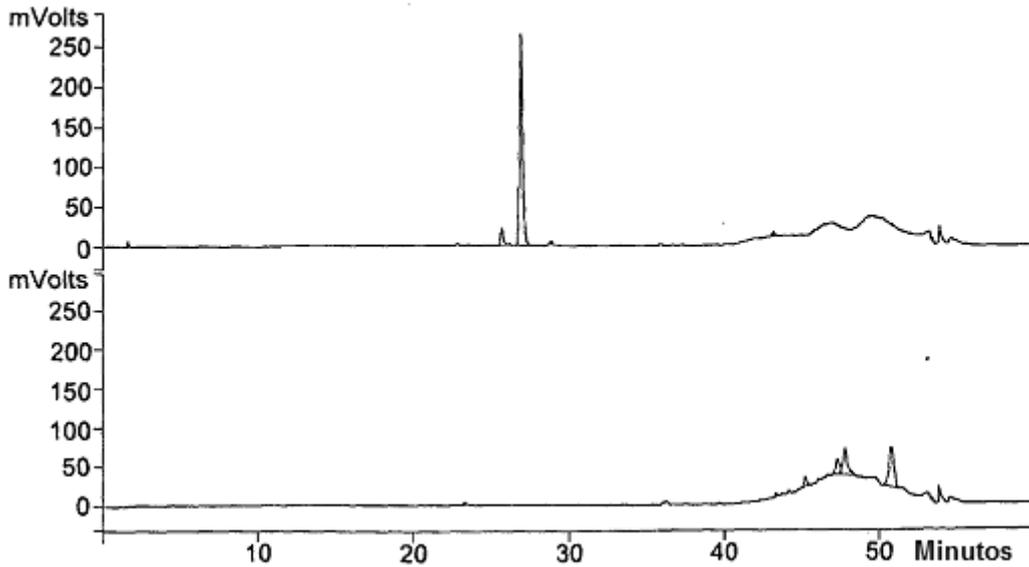


Figura 10

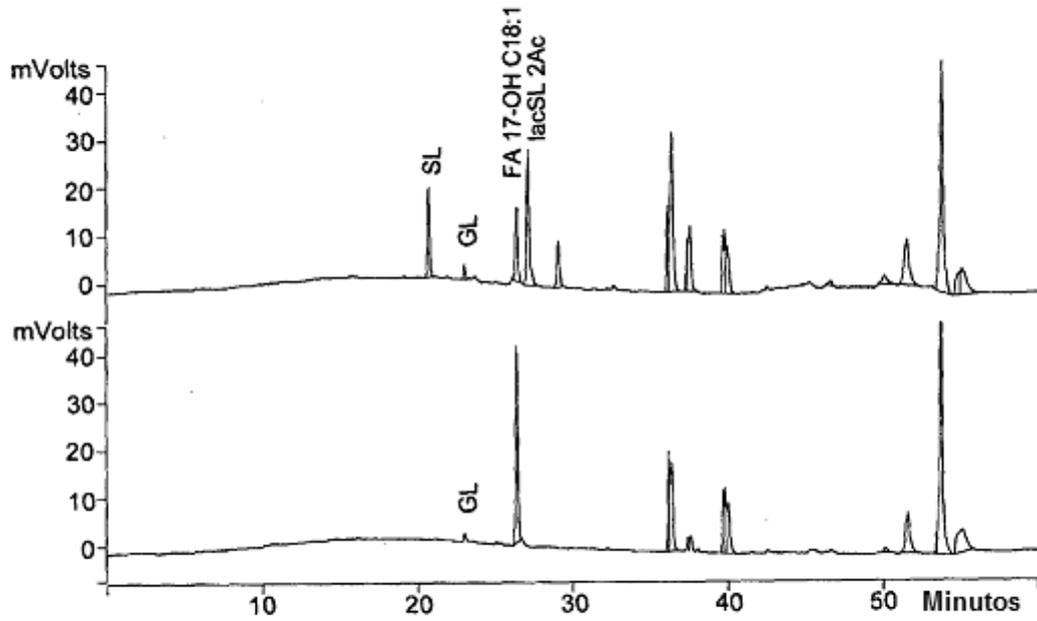


Figura 11

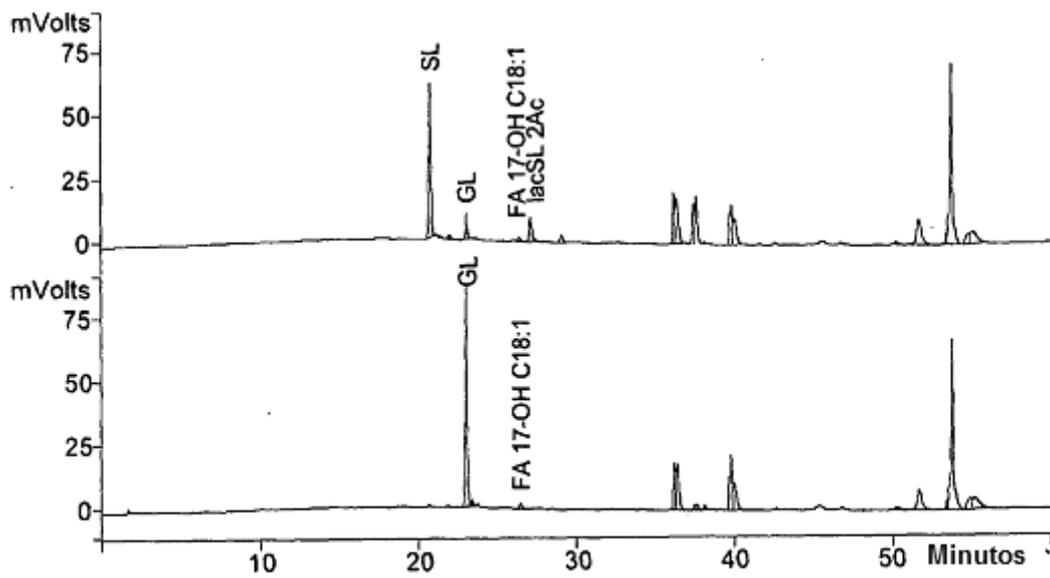


Figura 12

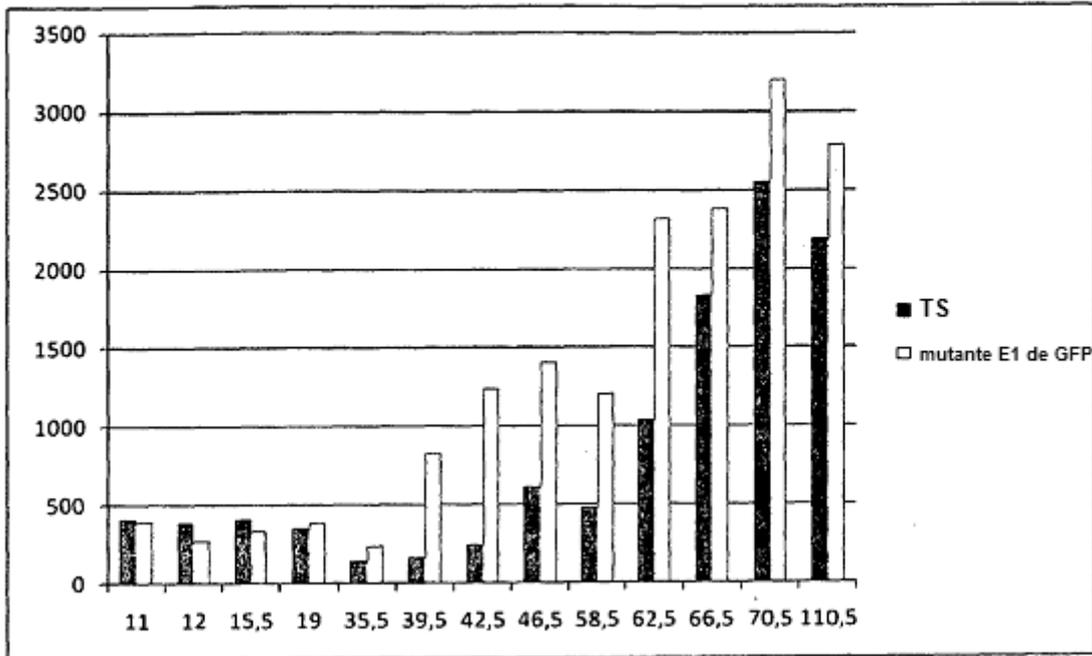


Figura 13

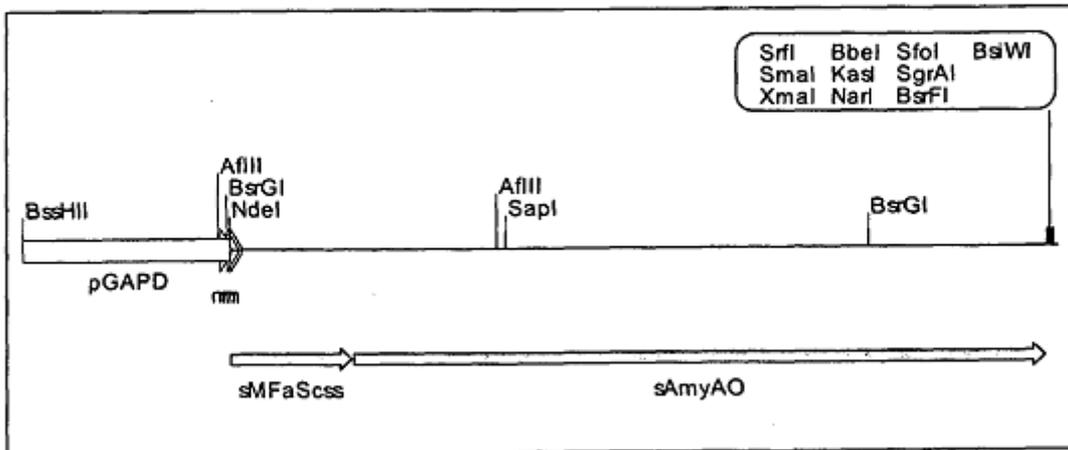


Figura 14

```

1  gcgcgcgcgcg aaacgggttag tataagcagt acgacgttgt tgctggagtc tcactctgcaa
>>.....pGAPD.....>
61  ggttgagtac caatccctgc cccaatacga gcaatcgaag ccttggggaa agatgcggcg
>.....pGAPD.....>
121  ggctagcttc agcaataaat agcaggcgac acacaaaaat taggcggcaa gcgcacgctc
>.....pGAPD.....>
181  agcatgccat ctaccagggc aaaaagcaag gcaacctett ttctgcatcc cgatttagag
>.....pGAPD.....>
241  cctaccgctc attgcagggt gtgcgtctac gatataacac gatcgacatc gcgctgggta
>.....pGAPD.....>
301  tgcttctggg taaggggtcg caacgtgtga gttgtcagca ctggccgata cccaaagtat
>.....pGAPD.....>
361  ataatgcgcc gttgaacggt tatagtcggt caagctctta aagaaagact taagaacaaa
>.....pGAPD.....>
421  aacaactgta cacatatgcg ttttctagc atttttactg ctgttctttt cgctgctagc
      m r f p s i f t a v l f a a s
>....pGAPD.....<.....sMFaScss.....>
481  agcgtctttg ctgctccagt taactactaca acagaggatg agactgctca gattccggct
      s a l a a p v n t t t e d e t a q i p a
>.....sMFaScss.....>
541  gaggtctgta ttggttacag cgatcttgag ggcgatttcg atgttgctgt tcttccattt
      e a v i g y s d l e g d f d v a v l p f
>.....sMFaScss.....>
601  agcaacagca caaataacgg ctttctttt ataaatacta ctattgctag cattgtgctc
      s n s t n n g l l f i n t t i a s i a a
>.....sMFaScss.....>
661  aaggaggagg gcgtttctct tgataagcgt gctactcctg ctgactggcg ttcgcagagc
      k e e g v s l d k r a t p a d w r s q s
>.....sMFaScss.....<.....sAmyAO.....>
721  atttatttcc ttctcactga tegttttgc cgtactgatg gctcgactac tgctacttgc
      i y f l l t d r f a r t d g s t t a t c
>.....sAmyAO.....>
781  aatctgctg atcgtaagta ctgcggtgga acatggcagg gcattattga caagcttgac
      n t a d r k y c g g t w q g i i d k l d
>.....sAmyAO.....>
841  tatattcagg gaatgggctt cacagctatt tggattacc cgttacagc tcagctgccc
      y i q g m g f t a i w i t p v t a q l p
>.....sAmyAO.....>
901  cagaccaccg cttatggaga tgcttaccac ggctactggc agcaggatat ttactctctg
      q t t a y g d a y h g y w q q d i y s l
>.....sAmyAO.....>
961  aacgagaact acggcactgc tgatgacctt aaggctctct cttcggtctc tcacgagcgt
      n e n y g t a d d l k a l s s a l h e r
>.....sAmyAO.....>
1021  ggcattgctc ttatggttga tgttgttgc aaccacatgg gctatgatgg agctggtagc
      g m y l m v d v v a n h m g y d g a g s
>.....sAmyAO.....>
1081  agcgttgatt actctgtttt taagccgttc tctagccagg actacttcca cccgttctgc
      s v d y s v f k p f s s q d y f h p f c
>.....sAmyAO.....>
1141  ctcattcaga actatgagga tcagactcag gttgaggatt gctggctcgg agataaact
      l i q n y e d q t q v e d c w l g d n t
>.....sAmyAO.....>

```

Figura 14 cont.

```

1201 gttagccttc ctgatctoga taccaccaag gatgttggtta agaatgagtg gtacgactgg
    v s l p d l d t t k d v v k n e w y d w
    >.....sAmyAO.....>
1261 gttggaagcc ttgtttcgaa ctacagcatt gacggcctcc gtattgacac agttaagcac
    v g s l v s n y s i d g l r i d t v k h
    >.....sAmyAO.....>
1321 gttcagaagg acttctggcc cggctacaac aaggetgctg gcgtttactg cattggcgag
    v q k d f w p g y n k a a g v y c i g e
    >.....sAmyAO.....>
1381 gttctcgacg gtgatccggc ttacacttgc cctaccaga acggttatgga cggcgttctg
    v l d g d p a y t c p y q n v m d g v l
    >.....sAmyAO.....>
1441 aactatccca ttactatcc actcctcaac gctttcaaga gcaccagcgg cagcatggac
    n y p i y y p l l n a f k s t s g s m d
    >.....sAmyAO.....>
1501 gacctctaca acatgattaa caccgttaag agcgactgcc cagacagcac actcctgggc
    d l y n m i n t v k s d c p d s t l l g
    >.....sAmyAO.....>
1561 acattcgttg agaaccacga caaccacgt ttcgcttctt acaccaacga cattgctctc
    t f v e n h d n p r f a s y t n d i a l
    >.....sAmyAO.....>
1621 gctaagaacg ttgctgcttt cattattctc aacgacggaa ttcccattat ttacgctggc
    a k n v a a f i i l n d g i p i i y a g
    >.....sAmyAO.....>
1681 caggagcagc actacgctgg cggaaacgac cccgctaacc gtgaggctac ctggctctcg
    q e q h y a g g n d p a n r e a t w l s
    >.....sAmyAO.....>
1741 ggctaccoga ccgacagcga gctgtacaag cttattgcta gcgctaacgc tattcgtaac
    g y p t d s e l y k l i a s a n a i r n
    >.....sAmyAO.....>
1801 tatgctatta gcaaggatac aggattcgtt acctacaaga actggcccat ttacaaggac
    y a i s k d t g f v t y k n w p i y k d
    >.....sAmyAO.....>
1861 gacacaacta ttgctatgcg taagggcaca gatggctcgc agattgttac tattcttagc
    d t t i a m r k g t d g s q i v t i l s
    >.....sAmyAO.....>
1921 aacaaggggtg cttcgggtga ttcgataacc ctcagccttt ctgggtgctgg ttacacagct
    n k g a s g d s y t l s l s g a g y t a
    >.....sAmyAO.....>
1981 ggccagcagc ttactgaggt tattggctgc actaccgta ctggttggtc ggatggaaat
    g q q l t e v i g c t t v t v g s d g n
    >.....sAmyAO.....>
2041 gttcctgttc ctatggctgg tggcctccct cgtgttcttt atccgactga gaagcttgct
    v p v p m a g g l p r v l y p t e k l a
    >.....sAmyAO.....>
2101 ggtagcaaga tttgctctag ctcgtgataa gccggggcgc cggcgtacga tta(SBQ ID NO 64)
    g s k i c s s s - -(SEQ ID NO 65)
    >.....sAmyAO.....>>

```

Figura 15

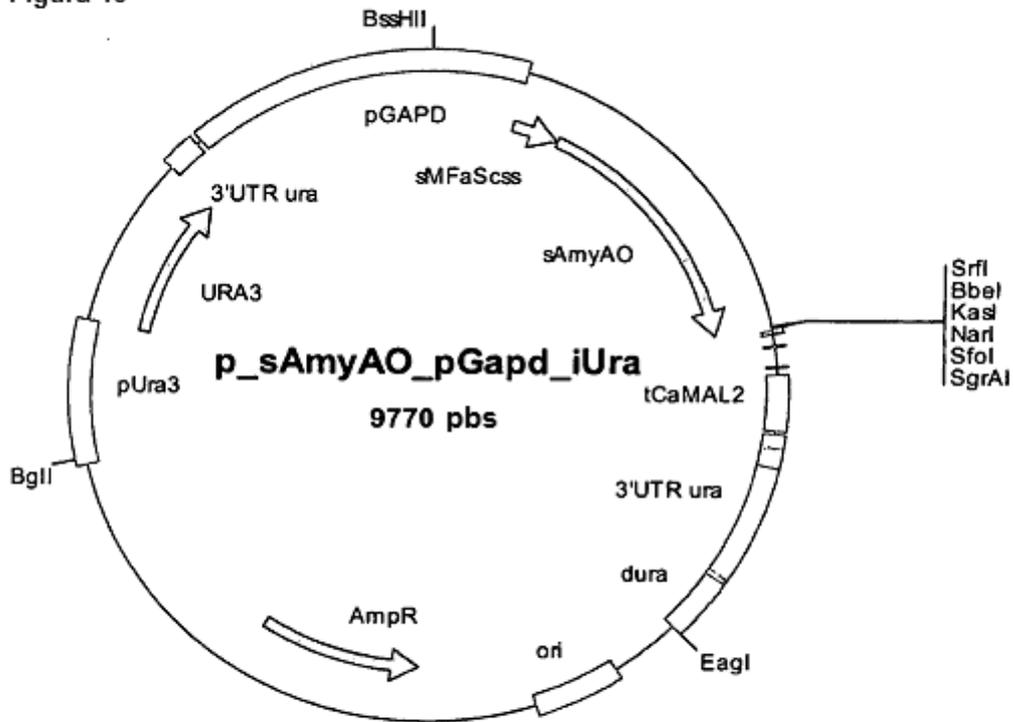


Figura 16

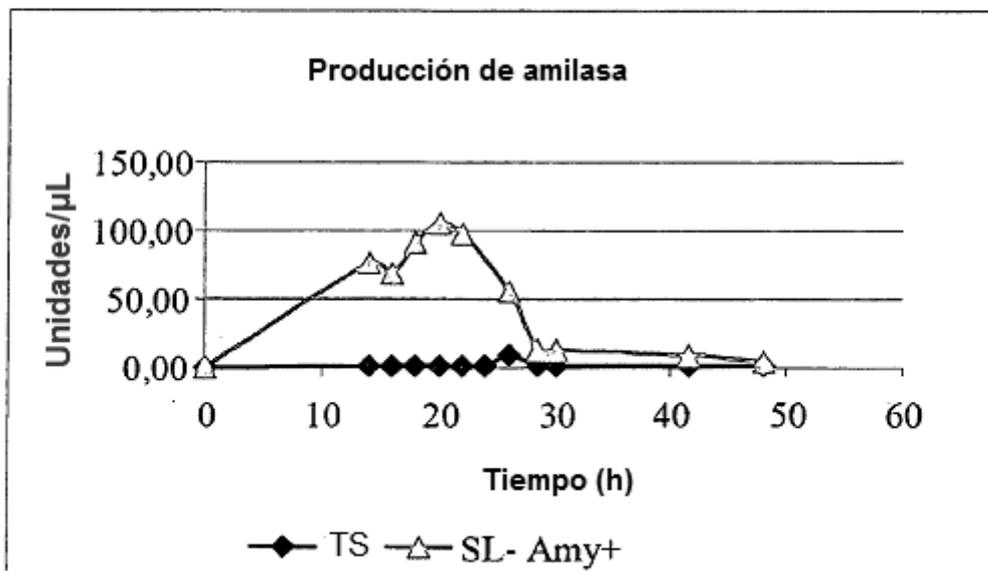


Figura 17

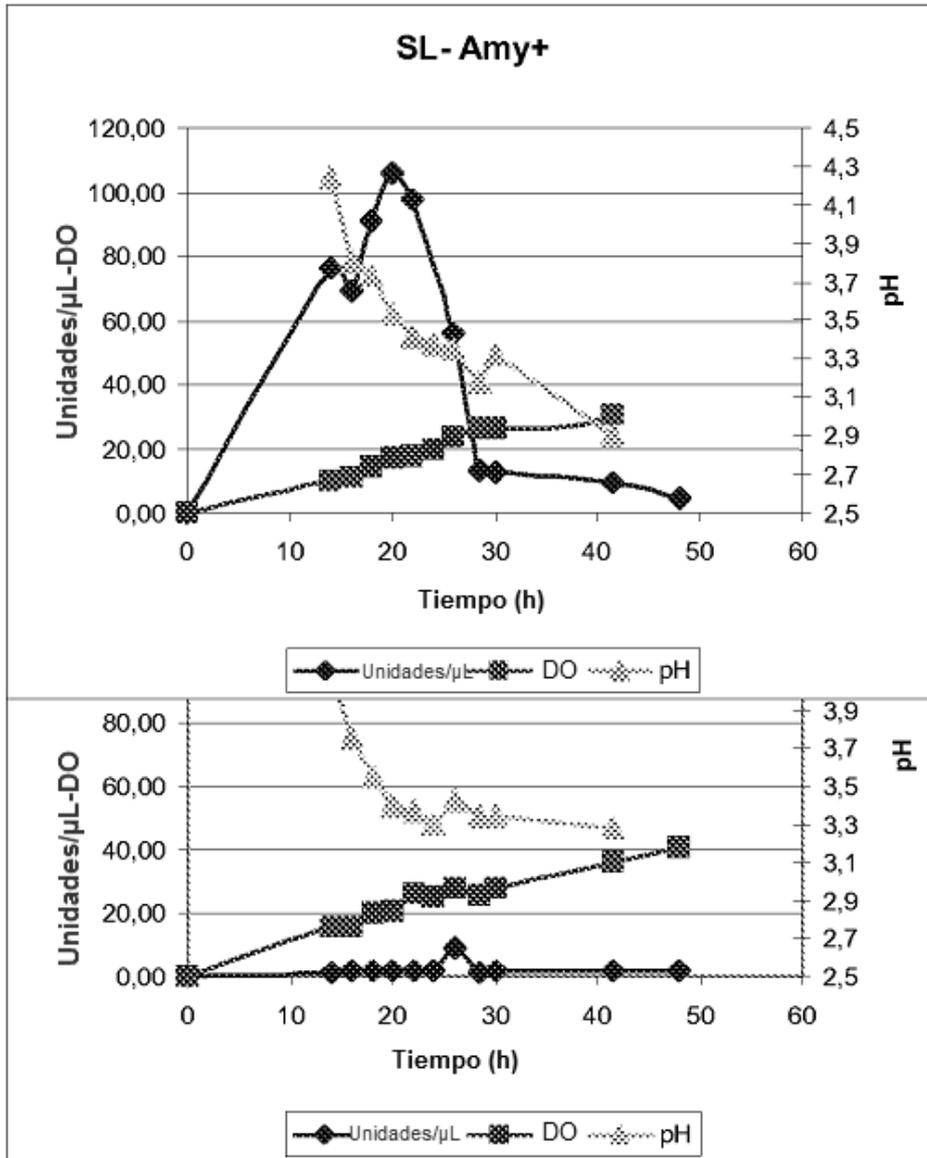


Figura 18

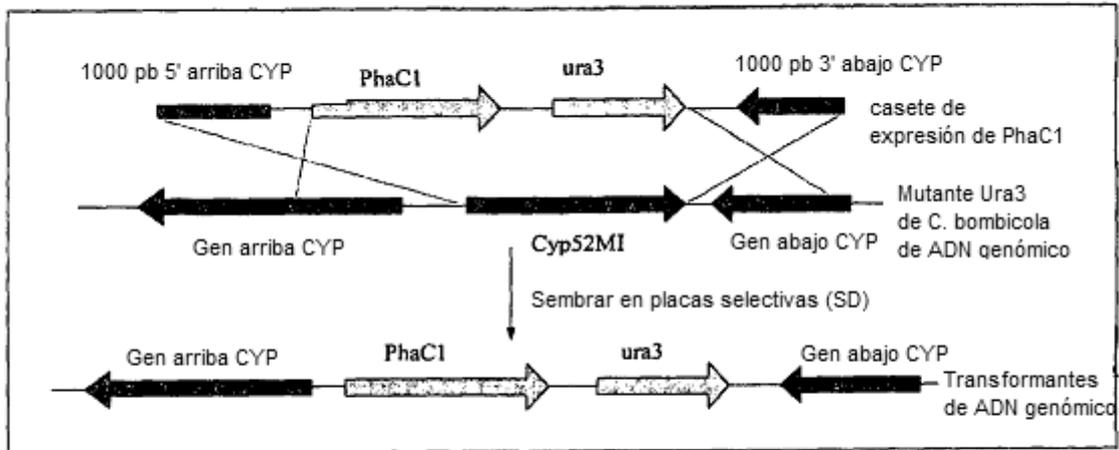


Figura 19 cont.

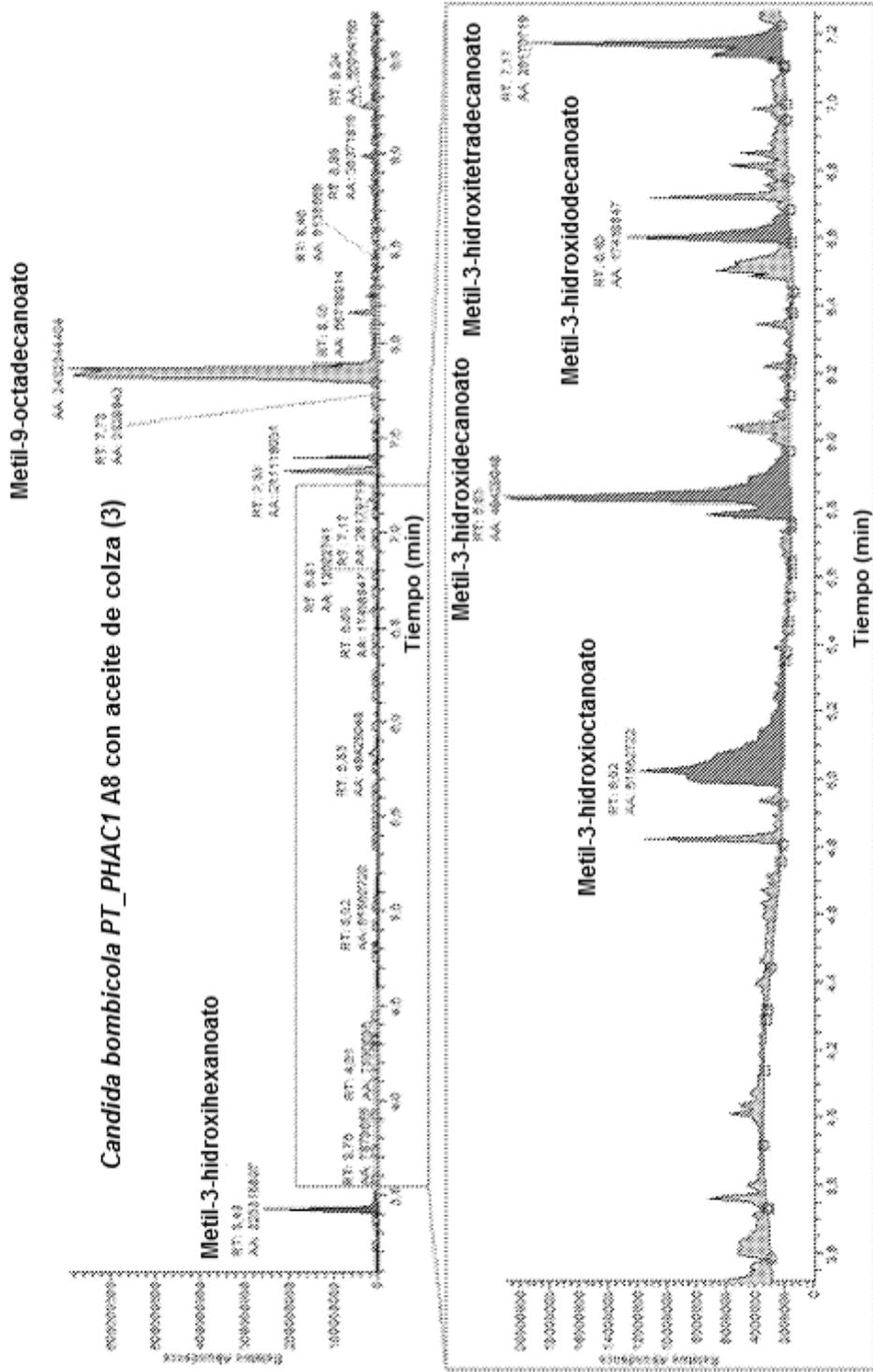


Figura 20

