

19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 817 001**

51 Int. Cl.:

**C12N 9/06** (2006.01)

**C12N 15/63** (2006.01)

12

## TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **17.02.2016 PCT/NL2016/050108**

87 Fecha y número de publicación internacional: **25.08.2016 WO16133384**

96 Fecha de presentación y número de la solicitud europea: **17.02.2016 E 16715126 (5)**

97 Fecha y número de publicación de la concesión europea: **17.06.2020 EP 3259349**

54 Título: **Producción de FDCA catalizada por deshidrogenasa**

30 Prioridad:

**17.02.2015 EP 15155401**

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

**06.04.2021**

73 Titular/es:

**PURAC BIOCHEM B.V. (100.0%)**

**Arkelsedijk 46**

**4206 AC Gorinchem, NL**

72 Inventor/es:

**RUIJSSENAARS, HARALD JOHAN**

74 Agente/Representante:

**TOMAS GIL, Tesifonte Enrique**

ES 2 817 001 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

Producción de FDCA catalizada por deshidrogenasa

5 Campo de la invención

[0001] La invención se refiere a los campos de la enzimología, genética molecular, biotransformación y tecnología de fermentación. En particular, la invención se refiere a deshidrogenasas que oxidan el ácido 5-(hidroximetil)-2-furoico en ácido 5-formil-2-furoico, y a polinucleótidos que codifican dichas deshidrogenasas y su uso en la biotransformación de hidroximetilfurfural en ácido 2,5-furandicarboxílico.

10 Antecedentes de la invención

[0002] El ácido 2,5-furandicarboxílico (FDCA) es un compuesto monomérico que se puede aplicar en la producción de poliésteres que tienen un tremendo impacto económico. Un compuesto muy importante en el campo es el polietilentereftalato (PET) que se produce a partir de ácido tereftálico (PTA) y etilenglicol. El FDCA puede sustituir al PTA en el poliéster PET, en cuyo caso da lugar a polietilénfurandicarboxilato (PEF). El PEF tiene un buen potencial para reemplazar al PET en el gran mercado del poliéster. No solo porque tiene propiedades superiores en comparación con el PET, sino también porque puede obtenerse a partir de materias primas renovables. El FDCA se puede producir a partir de azúcares, ya sea químicamente (De Jong et al. 2012. En: Biobased Monomers, Polymers, and Materials; Smith, P. et al.; Serie de simposios ACS; ACS Symposium Series; American Chemical Society: Washington, DC) o en una ruta química-biológica combinada (Wierckx et al 2011. Appl Microbiol Biotechnol 92: 1095-1105). En el último caso, un azúcar monomérico como glucosa o fructosa se transforma químicamente en 5-(hidroximetil)-2-furaldehído (HMF) que posteriormente puede ser oxidado por enzimas en FDCA.

[0003] Se ha desarrollado una ruta biológica para producir FDCA a partir de HMF basándose en el aislamiento de la cepa de degradación del HMF de *Cupriavidus basilensis* HMF14 (Wierckx et al 2010. Microbial Technology 3: 336-343). Se identificó un grupo de genes que codifican enzimas involucradas en la ruta de degradación del HMF en *C. basilensis* HMF14 y los genes relevantes se expresaron de manera heteróloga en una cepa de *Pseudomonas putida* (Koopman et al 2010 PNAS 107: 4919-4924) que de ese modo adquirió la capacidad de metabolizar el HMF. El primer paso oxidativo en la ruta de degradación implicó la formación de ácido 5-(hidroximetil)-2-furoico (HMFCFA) que a su vez se oxidó en ácido 5-formil-2-furoico (FFA) y luego en FDCA. En trabajos posteriores (Koopman et al 2010. Bioresource Technology 101: 6291-6296; y WO 2011/026913), solo el gen *hmfH* de *C. basilensis* HMF14 que codifica la enzima HMF oxidorreductasa se introdujo en *P. putida*. La oxidorreductasa actúa como oxidasa principalmente en HMFCFA, pero también puede oxidar HMF o FFA. Solo la expresión heteróloga del gen *hmfH* permite a *P. putida* producir FDCA a partir de HMF. En un trabajo de optimización adicional (Wierckx et al 2011, *supra*; y WO 2012/064195), dos genes adicionales se expresaron en *P. putida* y codificaron un transportador HMFCFA y una aldehído deshidrogenasa con especificidad desconocida, respectivamente.

[0004] Sin embargo, la ruta catalizada por oxidasa para la producción de FDCA a partir de HMF tiene varias desventajas inherentes en comparación con una ruta catalizada por deshidrogenasa, que incluyen al menos la producción de H<sub>2</sub>O<sub>2</sub> tóxico, la falta de ganancia de energía del paso oxidativo y la escasa afinidad por el O<sub>2</sub> y la alta demanda de oxígeno asociada del sistema. Por lo tanto, un objetivo de la presente invención es abordar estas desventajas proporcionando medios y métodos para una nueva ruta catalizada por deshidrogenasa para la producción de FDCA a partir de precursores furánicos como el HMF, así como proporcionar medios y métodos para usar un nuevo transportador de HMFCFA en tales procesos.

50 Resumen de la invención

[0005] En un primer aspecto, la invención se refiere a una célula que comprende un constructo de expresión para la expresión de una secuencia de nucleótidos que codifica una HMFCFA deshidrogenasa que tiene una secuencia de aminoácidos con al menos un 81.65 % de identidad con la secuencia de aminoácidos de la SEC ID N.º: 1, en la que el constructo de expresión se puede expresar en la célula y la expresión de la deshidrogenasa confiere o aumenta en la célula la capacidad de oxidar el ácido 5-hidroximetil-2-furancarboxílico (HMFCFA) a ácido 5-formil-2-furoico (FFA), en comparación con una célula de tipo salvaje correspondiente que carece del constructo de expresión. Preferiblemente, la célula tiene además: a) una actividad de aldehído deshidrogenasa que oxida los aldehídos furánicos a los correspondientes ácidos carboxílicos furánicos, donde preferiblemente la célula comprende un segundo constructo de expresión para la expresión de una secuencia de nucleótidos que codifica una aldehído deshidrogenasa que comprende una secuencia de aminoácidos con al menos un 45 % de identidad con cualquiera de las secuencias de aminoácidos SEC ID N.º: 24, 25, 26, 27, 28, 29 y 30, donde el segundo constructo de expresión es expresable en la célula y la expresión de la aldehído deshidrogenasa confiere o aumenta en la célula al menos una de las capacidades de i) oxidar 5-hidroximetilfurfural (HMF) a HMFCFA, ii) oxidar DFF a FFA y iii) oxidar FFA a FDCA, en comparación con una célula de tipo salvaje correspondiente que carece del segundo constructo de expresión; y b) la capacidad de transportar compuestos furánicos hacia dentro

y/o hacia fuera de la célula, donde preferiblemente, la célula comprende un tercer constructo de expresión para la expresión de una secuencia de nucleótidos que codifica un polipéptido que tiene la capacidad de transportar al menos HMFCa al interior de la célula, cuyo polipéptido comprende una secuencia de aminoácidos con al menos un 45 % de identidad con cualquiera de las secuencias de aminoácidos SEC ID N.º: 17, 31, 32, 33 y 34, donde el tercer constructo de expresión es expresable en la célula y la expresión del polipéptido confiere o aumenta en la célula la capacidad de transportar al menos HMFCa al interior de la célula, en comparación con una célula de tipo salvaje correspondiente que carece del tercer constructo de expresión.

[0006] En otro aspecto, la divulgación se refiere a una célula que comprende un constructo de expresión para la expresión de una secuencia de nucleótidos que codifica un polipéptido que tiene la capacidad de transportar al menos HMFCa al interior de la célula, donde el polipéptido comprende una secuencia de aminoácidos que tiene al menos un 86.5 % de identidad con la secuencia de aminoácidos de la SEC ID N.º: 17, en la que el constructo de expresión se puede expresar en la célula y la expresión del polipéptido confiere o aumenta en la célula al menos la capacidad de transportar al menos HMFCa al interior de la célula, en comparación con la correspondiente célula de tipo salvaje que carece del constructo de expresión, y donde la célula comprende además enzimas para convertir HMF en FDCA, donde las enzimas para convertir HMF en FDCA incluyen preferiblemente al menos una de: a) alcohol deshidrogenasa que oxida HMFCa a FFA y una actividad de aldehído deshidrogenasa que oxida los aldehídos furánicos a los correspondientes ácidos carboxílicos furánicos; y b) una oxidorreductasa que oxida a FDCA uno o más compuestos entre HMF, 2,5-dihidroximetil furano, HMFCa, FFA y 2,5-diformil furano y opcionalmente una actividad de aldehído deshidrogenasa que oxida los aldehídos furánicos a los correspondientes ácidos carboxílicos furánicos.

[0007] Una célula según la invención preferiblemente es una célula microbiana, por ejemplo, una célula bacteriana, de levadura o fúngica filamentosa. Una levadura o una célula fúngica filamentosa de la invención se seleccionan preferiblemente de un género del grupo que consta de *Candida*, *Hansenula*, *Kluyveromyces*, *Pichia*, *Saccharomyces*, *Schizosaccharomyces*, *Yarrowia*, *Acremonium*, *Agaricus*, *Aspergillus*, *Aureobasidium*, *Myceliophthora*, *Chrysosporium*, *Coprinus*, *Cryptococcus*, *Filibasidium*, *Fusarium*, *Mupocoricola*, *Necalicola*, *Magdalena Piromyces*, *Panerochaete*, *Pleurotus*, *Schizophyllum*, *Talaromyces*, *Thermoascus*, *Thielavia*, *Tolypocladium* y *Trichoderma*, más preferiblemente una levadura o una célula fúngica filamentosa seleccionada de una especie del grupo que consta de *Kluyveromyces lactis*, *S. cerevisiae*, *Hansenula polymorpha*, *Yarrowia lipolytica*, *Pichia pastoris*, *Aspergillus niger*, *Aspergillus awamori*, *Aspergillus foetidus*, *Aspergillus sojae*, *Aspergillus fumigatus*, *Talaromyces emersonii*, *Aspergillus oryzae*, *Myceliophthora thermophila*, *Trichoderma reesei* y *Penicillium chrysogenum*. Una célula bacteriana de la invención se selecciona preferiblemente de un género del grupo que consta de *Escherichia*, *Anabaena*, *Aeribacillus*, *Aneurinibacillus*, *Burkholderia*, *Bradyrhizobium*, *Caulobacter*, *Cupriavidus*, *Desulfotomaculum*, *Desulfurispora*, *Gluconobacter*, *Rhodobacter*, *Pelotomaculum*, *Pseudomonas*, *Paracoccus*, *Bacillus*, *Geobacillus*, *Brevibacillus*, *Brevibacterium*, *Corynebacterium*, *Rhizobium* (*Sinorhizobium*), *Flavobacterium*, *Klebsiella*, *Enterobacter*, *Lactobacillus*, *Lactococcus*, *Methylobacterium*, *Ralstonia*, *Rhodopseudomonas*, *Staphylococcus* y *Streptomyces*, más preferiblemente una célula bacteriana seleccionada de una especie del grupo que consta de *A. pallidus*, *A. terranovensis*, *B. subtilis*, *B. amyloliquefaciens*, *B. coagulans*, *B. kribbensis*, *B. licheniformis*, *B. puntis*, *B. megaterium*, *B. halodurans*, *B. pumilus*, *B. thermoruber*, *B. panacihumi*, *C. basilensis*, *D. kuznetsovii*, *D. thermophila*, *G. kaustophilus*, *Gluconobacter oxydans*, *Caulobacter crescentus* CB 15, *Methylobacterium extorquens*, *Rhodobacter sphaeroides*, *Pelotomaculum thermopropionicum*, *Pseudomonas zeaxanthinificans*, *Pseudomonas putida*, *Paracoccus denitrificans*, *E. coli*, *C. glutamicum*, *Staphylococcus carnosus*, *Streptomyces lividans*, *Sinorhizobium melioli* y *Rhizobium radiobacter*.

[0008] En otros aspecto, la invención se refiere a un proceso para preparar un polipéptido que tiene una actividad de HMFCa deshidrogenasa tal como se define en los aspectos anteriores, y/o para preparar un polipéptido que tiene capacidades de transporte de compuestos furánicos tal como se define en los aspectos anteriores. El método comprende preferiblemente el paso de cultivar una célula tal como se define en los aspectos anteriores, en condiciones propicias para la expresión del polipéptido o polipéptidos y, opcionalmente, recuperar el polipéptido o polipéptidos.

[0009] En otro aspecto, la invención se refiere a un proceso para oxidar HMFCa a FFA, que comprende el paso de incubar una célula de acuerdo con cualquiera de los aspectos anteriores en presencia de HMFCa, preferiblemente en condiciones propicias para la oxidación del HMFCa por parte la célula.

[0010] En otro aspecto más, la invención se refiere a un proceso para producir FDCA, que comprende el paso de incubar una célula de acuerdo con cualquiera de los aspectos anteriores, en un medio que comprende uno o más precursores furánicos de FDCA, preferiblemente en condiciones propicias para la oxidación de los precursores furánicos de FDCA por la célula a FDCA y, opcionalmente, la recuperación de FDCA, donde preferiblemente, al menos un precursor furánico de FDCA se selecciona del grupo que consiste en HMF, 2,5-dihidroximetil furano (DHF o HMF-OH), HMFCa, FFA y 2,5-diformil furano (DFF), de los cuales HMF es el preferido, en el que los precursores furánicos de FDCA se obtienen a partir de uno o más azúcares hexosa, preferiblemente uno o más azúcares hexosa obtenidos de biomasa lignocelulósica, preferiblemente mediante deshidratación catalizada por

ácido, donde preferiblemente el FDCA se recupera del medio por un proceso que comprende precipitación ácida o salina seguida de cristalización por enfriamiento y/o extracción con disolvente.

5 [0011] En un aspecto adicional, la invención se refiere a un proceso para producir un polímero a partir de uno o más monómeros de FDCA. Dicho proceso comprende los pasos de: a) preparar un monómero de FDCA en un proceso según el aspecto anterior; y producir un polímero a partir del monómero de FDCA obtenido en a).

10 [0012] La invención también se refiere al uso de una célula según cualquiera de los aspectos anteriores, para la biotransformación de uno o más precursores furánicos de FDCA a FDCA, donde preferiblemente, al menos un precursor furánico de FDCA se selecciona del grupo que consiste en HMF, DHF, HMFCA, FFA y DFF, de los cuales el HMF es el preferido.

15 [0013] En otro aspecto, la invención se refiere a un polipéptido que tiene actividad de HMFCA deshidrogenasa, cuyo polipéptido comprende una secuencia de aminoácidos que tiene al menos un 81.85 % de identidad de secuencia con la secuencia de aminoácidos de la SEC ID N.º: 1. En este aspecto, esta invención también se refiere a una molécula de ácido nucleico que comprende al menos uno de los siguientes elementos: a) una secuencia de nucleótidos que codifica un polipéptido que tiene actividad de HMFCA deshidrogenasa, cuyo polipéptido comprende una secuencia de aminoácidos que tiene al menos un 81.85 % de identidad de secuencia con la secuencia de aminoácidos de la SEC ID N.º: 1; b) una secuencia de nucleótidos establecida en las SEC ID N.º: 12 o 13; c) una secuencia de nucleótidos cuya secuencia difiere de la secuencia de una secuencia de nucleótidos de b) debido a la degeneración del código genético; y d) una secuencia de nucleótidos que es el complemento inverso de una secuencia de nucleótidos tal como se define en a) a c), donde, preferiblemente, la molécula de ácido nucleico es un vector. En este aspecto, la invención se refiere además a una célula que comprende un polipéptido de este aspecto y/o una molécula de ácido nucleico de este aspecto, donde preferiblemente la célula es una célula cultivada.

25

[0014] En un aspecto final, la divulgación se refiere a un polipéptido que tiene la capacidad de transportar al menos HMFCA al interior de la célula, cuyo polipéptido comprende una secuencia de aminoácidos que tiene al menos un 86.5 % de identidad de secuencia con la secuencia de aminoácidos de la SEC ID N.º: 17. En este aspecto, la divulgación también se refiere a una molécula de ácido nucleico que comprende al menos uno de los siguientes elementos: a) una secuencia de nucleótidos que codifica un polipéptido que tiene la capacidad de transportar al menos HMFCA al interior de la célula, cuyo polipéptido comprende una secuencia de aminoácidos que tiene al menos un 86.5 % de identidad de secuencia con la secuencia de aminoácidos de la SEC ID N.º: 17; b) una secuencia de nucleótidos establecida en las SEC ID N.º: 18; c) un fragmento de una secuencia de nucleótidos tal como se define en a) o b) que tiene 10, 15, 20, 30, 50 o 100 nucleótidos de longitud; d) una secuencia de nucleótidos cuya secuencia difiere de la secuencia de nucleótidos de b) o c) debido a la degeneración del código genético; y e) una secuencia de nucleótidos que es el complemento inverso de una secuencia de nucleótidos tal como se define en a) a d), donde, preferiblemente, la molécula de ácido nucleico es un vector. En este aspecto, la divulgación se refiere además a una célula que comprende un polipéptido de este aspecto y/o una molécula de ácido nucleico de este aspecto, donde preferiblemente la célula es una célula cultivada.

30

35

40

#### Descripción de la invención

#### 45 Definiciones

[0015] Los términos "homología", "identidad de secuencia" y similares se usan indistintamente en el presente documento. La identidad de secuencia se define en el presente documento como una relación entre dos o más secuencias de aminoácidos (polipéptido o proteína) o dos o más secuencias de ácidos nucleicos (polinucleótidos), según se determine comparando las secuencias. En la técnica, "identidad" también significa el grado de relación de secuencia entre las secuencias de aminoácidos o ácidos nucleicos, según sea el caso, de acuerdo con lo determinado mediante la coincidencia entre cadenas de tales secuencias. La "similitud" entre dos secuencias de aminoácidos se determina comparando la secuencia de aminoácidos y sus sustitutos de aminoácidos conservados de un polipéptido con la secuencia de un segundo polipéptido. La "identidad" y la "similitud" se pueden calcular fácilmente mediante métodos conocidos.

50

55

[0016] La "identidad de secuencia" y la "similitud de secuencia" se pueden determinar mediante la alineación de dos secuencias de péptidos o de dos nucleótidos usando algoritmos de alineación global o local, dependiendo de la longitud de las dos secuencias. Las secuencias de longitudes similares se alinean preferiblemente usando un algoritmo de alineación global (por ejemplo, Needleman Wunsch) que alinea las secuencias de manera óptima en toda la longitud, mientras que las secuencias de longitudes sustancialmente diferentes se alinean preferiblemente usando un algoritmo de alineación local (por ejemplo, Smith Waterman). Las secuencias pueden entonces denominarse "sustancialmente idénticas" o "esencialmente similares" (cuando se alinean de manera óptima, por ejemplo, mediante los programas GAP o BESTFIT usando parámetros predeterminados) cuando comparten al menos un cierto porcentaje mínimo de identidad de secuencia (como se define a continuación). GAP utiliza el algoritmo de alineación global de Needleman y Wunsch para alinear dos secuencias en toda su

60

65

longitud (longitud completa), maximizando el número de coincidencias y minimizando el número de gaps. Una alineación global se usa adecuadamente para determinar la identidad de secuencia cuando las dos secuencias tienen longitudes similares. Generalmente, se utilizan los parámetros predeterminados de GAP, con una penalización por creación de gap = 50 (nucleótidos)/8 (proteínas) y una penalización por extensión de gap = 3 (nucleótidos)/2 (proteínas). Para los nucleótidos, la matriz de puntuación predeterminada utilizada es nwsgapdna y para las proteínas, la matriz de puntuación predeterminada es Blosum62 (Henikoff y Henikoff, 1992, PNAS 89, 915-919). Las alineaciones de secuencia y las puntuaciones para el porcentaje de identidad de secuencia se pueden determinar utilizando programas informáticos, como el paquete Wisconsin de GCG, versión 10.3, disponible en Accelrys Inc., 9685 Scranton Road, San Diego, CA 92121-3752 EE. UU., o utilizando software de código abierto, como el programa "needle" (usando el algoritmo global de Needleman Wunsch) o "water" (usando el algoritmo local de Smith Waterman) en EmbossWIN versión 2.10.0, usando los mismos parámetros que para GAP anteriormente, o usando la configuración predeterminada (ambos para "needle" y para "water" y tanto para proteínas como para alineaciones de ADN, la penalización por apertura de gap predeterminada es 10.0 y la penalización por extensión de gap predeterminada es 0.5; las matrices de puntuación predeterminadas son Blossum62 para proteínas y DNAFull para ADN). Cuando las secuencias tienen longitudes generales sustancialmente diferentes, se prefieren las alineaciones locales, como las que utiliza el algoritmo de Smith Waterman.

[0017] Alternativamente, el porcentaje de similitud o identidad se puede determinar buscando en bases de datos públicas, usando algoritmos como FASTA, BLAST, etc. Por lo tanto, las secuencias de ácidos nucleicos y proteínas de la presente invención pueden usarse además como una "secuencia de consulta" para realizar búsquedas contra bases de datos públicas con el fin de, por ejemplo, identificar a otros miembros de la familia o secuencias relacionadas. Estas búsquedas se pueden realizar utilizando los programas BLASTn y BLASTx (versión 2.0) de Altschul et al. (1990) J. Mol. Biol. 215: 403-10. Las búsquedas de nucleótidos con BLAST se pueden realizar con el programa NBLAST, puntuación = 100, longitud de palabra = 12 para obtener secuencias de nucleótidos homólogas a las moléculas de ácidos nucleicos de oxidoreductasa de la invención. Las búsquedas de proteínas con BLAST se pueden realizar con el programa BLASTx, puntuación = 50, longitud de palabra = 3 para obtener secuencias de aminoácidos homólogas a las moléculas de proteínas de la invención. Para obtener alineaciones con gaps para fines de comparación, se puede utilizar Gapped BLAST tal como se describe en Altschul et al., (1997) Nucleic Acids Res. 25 (17): 3389-3402. Al utilizar los programas BLAST y Gapped BLAST, se pueden utilizar los parámetros predeterminados de los respectivos programas (por ejemplo, BLASTx y BLASTn). Consulte la página de inicio del National Center for Biotechnology Information en <http://www.ncbi.nlm.nih.gov/>.

[0018] Opcionalmente, al determinar el grado de similitud de aminoácidos, el experto en la materia también puede tener en cuenta las denominadas sustituciones de aminoácidos "conservadoras", como será evidente para el experto. Las sustituciones conservadoras de aminoácidos se refieren a la intercambiabilidad de los residuos que tienen cadenas laterales similares. Por ejemplo, un grupo de aminoácidos con cadenas laterales alifáticas se compone de glicina, alanina, valina, leucina e isoleucina; un grupo de aminoácidos con cadenas laterales de hidroxilo alifático se compone de serina y treonina; un grupo de aminoácidos con cadenas laterales que contienen amidas se compone de asparaginas y glutamina; un grupo de aminoácidos con cadenas laterales aromáticas es fenilalanina, tirosina y triptófano; un grupo de aminoácidos que tienen cadenas laterales básicas se compone de lisina, arginina e histidina; y un grupo de aminoácidos con cadenas laterales que contienen azufre se compone de cisteína y metionina. Los grupos de sustitución de aminoácidos conservadores preferidos son: valina-leucina-isoleucina, fenilalanina-tirosina, lisina-arginina, alanina-valina y asparagina-glutamina. Las variantes de sustitución de la secuencia de aminoácidos descrita en el presente documento son aquellas en las que se ha eliminado al menos un residuo de las secuencias descritas y se ha insertado un residuo diferente en su lugar. Preferiblemente, el cambio de aminoácidos es conservador. Las sustituciones conservadoras preferidas para cada uno de los aminoácidos de origen natural son las siguientes: ala por ser; arg por lys; asn por gln o his; asp por glu; cys por ser o ala; gln a asn; glu a asp; gly por pro; his por asn o gln; ile por leu o val; leu por ile o val; lys por arg, gln o glu; met por leu o ile; phe por met, leu o tyr; ser por thr; thr por ser; trp por tyr; tyr por trp o phe; y val por ile o leu.

[0019] Tal como se usa en el presente documento, los términos "hibridación selectiva", "se hibrida selectivamente" y términos similares están destinados a describir condiciones de hibridación y lavado en las que las secuencias de nucleótidos que tienen una homología de al menos un 66 %, al menos un 70 %, al menos un 75 %, al menos un 80 %, más preferiblemente al menos un 85 %, incluso más preferiblemente al menos un 90 %, preferiblemente al menos un 95 %, más preferiblemente al menos un 98 % o más preferiblemente al menos un 99 % entre sí permanecen típicamente hibridadas entre sí. Es decir, dichas secuencias de hibridación pueden compartir al menos un 45 %, al menos un 50 %, al menos un 55 %, al menos un 60 %, al menos un 65 %, al menos un 70 %, al menos un 75 %, al menos un 80 %, más preferiblemente al menos un 85 %, incluso más preferiblemente al menos un 90 %, más preferiblemente al menos un 95 %, más preferiblemente al menos un 98 % o más preferiblemente al menos un 99 % de identidad de secuencia.

[0020] Un ejemplo preferido, no limitante de tales condiciones de hibridación es la hibridación en cloruro de sodio/citrato de sodio (SSC) 6X a aproximadamente 45 °C, seguida de uno o más lavados en SSC 1 X, SDS al

0.1 % a aproximadamente 50 °C, preferiblemente a aproximadamente 55 °C, preferiblemente a aproximadamente 60 °C e incluso más preferiblemente a aproximadamente 65 °C.

5 [0021] Las condiciones muy astringentes incluyen, por ejemplo, hibridación a aproximadamente 68 °C en SSC 5x/solución de Denhardt 5x/SDS al 1.0 % y lavado en SSC 0.2x/SDS al 0.1 % a temperatura ambiente. Alternativamente, el lavado se puede realizar a 42 °C.

10 [0022] El experto en la materia sabrá qué condiciones aplicar para condiciones de hibridación astringentes y muy astringentes. En la técnica se encuentra fácilmente disponible orientación adicional con respecto a dichas condiciones, por ejemplo, en Sambrook et al., 1989, Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, Nueva York.; y Ausubel et al. (eds.), Sambrook y Russell (2001) "Molecular Cloning: A Laboratory Manual (3ª edición), Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Nueva York 1995, Current Protocols in Molecular Biology, (John Wiley & Sons, N.Y.).

15 [0023] Por supuesto, un polinucleótido que se hibrida solo con una secuencia poli (A) (como el tracto poli (A) del extremo 3' de ARNm), o con un tramo complementario de residuos de T (o U) no se incluiría en un polinucleótido de la invención utilizado para hibridarse específicamente con una porción de un ácido nucleico de la invención, ya que dicho polinucleótido se hibridaría con cualquier molécula de ácido nucleico que contenga un tramo de poli (A) o el complemento del mismo (por ejemplo, prácticamente cualquier clon de ADNc de doble hebra).

20 [0024] Un "constructo de ácidos nucleicos" o "vector de ácidos nucleicos" se entiende en el presente documento como una molécula de ácido nucleico artificial resultante del uso de tecnología de ADN recombinante. El término "constructo de ácidos nucleicos", por lo tanto, no incluye moléculas de ácidos nucleicos de origen natural, aunque un constructo de ácidos nucleico puede comprender (partes de) moléculas de ácidos nucleicos de origen natural. Los términos "vector de expresión" o "constructo de expresión" se refieren a secuencias de nucleótidos que son capaces de efectuar la expresión de un gen en células huésped u organismos huésped compatibles con tales secuencias. Estos vectores de expresión típicamente incluyen al menos secuencias reguladoras de la transcripción adecuadas y, opcionalmente, señales de terminación de la transcripción 3'. También pueden estar presentes factores adicionales necesarios o útiles para efectuar la expresión, tales como elementos potenciadores de la expresión. El vector de expresión se introducirá en una célula huésped adecuada y podrá efectuar la expresión de la secuencia codificante en un cultivo celular in vitro de la célula huésped. El vector de expresión será adecuado para la replicación en la célula u organismo huésped de la invención.

35 [0025] Según se usa en el presente documento, los términos "promotor" o "secuencia reguladora de la transcripción" se refieren a un fragmento de ácidos nucleicos que funciona para controlar la transcripción de una o más secuencias codificantes, y está ubicado corriente arriba con respecto a la dirección de transcripción del sitio de inicio de la transcripción de la secuencia codificante, y se identifica estructuralmente por la presencia de un sitio de unión para la ARN polimerasa dependiente de ADN, sitios de inicio de la transcripción y cualquier otra secuencia de ADN, incluyendo, entre otros, los sitios de unión al factor de transcripción, los sitios de unión a las proteínas represora y activadora y cualquier otra secuencia de nucleótidos conocida por un experto en la técnica que actúe directa o indirectamente para regular la cantidad de transcripción del promotor. Un promotor "constitutivo" es un promotor que está activo en la mayoría de los tejidos en la mayoría de las condiciones fisiológicas y de desarrollo. Un promotor "inducible" es un promotor que está regulado fisiológicamente o en cuanto a desarrollo, p. ej., mediante la aplicación de un inductor químico.

45 [0026] El término "marcador seleccionable" es un término familiar para un experto en la técnica y se usa en el presente documento para describir cualquier entidad genética que, cuando se expresa, puede usarse para seleccionar una célula o células que contienen el marcador seleccionable. El término "reportero" puede usarse indistintamente con marcador, aunque se usa principalmente para referirse a marcadores visibles, como la proteína verde fluorescente (GFP). Los marcadores seleccionables pueden ser dominantes, recesivos o bidireccionales.

50 [0027] Tal como se usa en el presente documento, el término "operativamente enlazado" se refiere a un enlace de elementos polinucleotídicos en una relación funcional. Un ácido nucleico está "operativamente enlazado" cuando se coloca en una relación funcional con otra secuencia de ácido nucleico. Por ejemplo, una secuencia reguladora de la transcripción está operativamente enlazada a una secuencia codificante si afecta a la transcripción de la secuencia codificante. Operativamente enlazado significa que las secuencias de ADN que se enlazan son típicamente contiguas y, cuando es necesario para unir dos regiones codificantes de proteínas, son contiguas y están en el marco de lectura.

60 [0028] Los términos "proteína" o "polipéptido" se usan indistintamente y se refieren a moléculas que constan de una cadena de aminoácidos, sin referencia a un modo de acción, tamaño, estructura tridimensional u origen específicos.

65 [0029] El término "gen" significa un fragmento de ADN que comprende una región (región transcrita), que se transcribe en una molécula de ARN (por ejemplo, un ARNm) en una célula, operativamente enlazada a regiones

reguladoras adecuadas (por ejemplo, un promotor). Un gen comprenderá habitualmente varios fragmentos operativamente enlazados, tales como un promotor, una secuencia líder 5', una región codificante y una secuencia no traducida 3' (extremo 3') que comprende un sitio de poliadenilación. "Expresión de un gen" se refiere al proceso en el que una región de ADN que está operativamente enlazada a regiones reguladoras apropiadas, en particular un promotor, se transcribe en un ARN que es biológicamente activo, es decir, que puede traducirse en una proteína o péptido biológicamente activos. El término "homólogo", cuando se usa para indicar la relación entre una molécula de ácido nucleico o de polipéptido (recombinante) dada y un organismo o célula huésped dados, se entiende que significa que, en la naturaleza, la molécula de ácido nucleico o polipéptido es producida por una célula huésped u organismos de la misma especie, preferiblemente de la misma variedad o cepa. Si es homóloga a una célula huésped, una secuencia de ácidos nucleicos que codifica un polipéptido estará de manera habitual (pero no necesariamente) operativamente enlazada a otra secuencia promotora (heteróloga) y, si corresponde, a otra secuencia señal secretora (heteróloga) y/o secuencia terminadora que en su entorno natural. Se entiende que las secuencias reguladoras, secuencias señal, secuencias terminadoras, etc. también pueden ser homólogas a la célula huésped. En este contexto, usar solo elementos de secuencia "homólogos" permite la construcción de organismos genéticamente modificados (OGM) "autoclonados" (la autoclonación se define en el presente documento como en la Directiva Europea 98/81/EC Anexo II). Cuando se usa para indicar la relación de dos secuencias de ácidos nucleicos, el término "homólogo" significa que una secuencia de ácido nucleico monocatenario puede hibridarse con una secuencia complementaria de ácido nucleico monocatenario. El grado de hibridación puede depender de una serie de factores, incluido el grado de identidad entre las secuencias y las condiciones de hibridación, tales como la temperatura y la concentración de sal, según se comenta más adelante.

[0030] Los términos "heterólogo" y "exógeno" cuando se usan con respecto a un ácido nucleico (ADN o ARN) o proteína se refieren a un ácido nucleico o proteína que no se produce de modo natural como parte del organismo, célula, genoma o secuencia de ADN o ARN en el que está presente, o que se encuentra en una célula o ubicación o ubicaciones en el genoma o secuencia de ADN o ARN que difieren de aquella en la que se encuentra en la naturaleza. Los ácidos nucleicos o proteínas heterólogos y exógenos no son endógenos a la célula en la que se introducen, sino que se han obtenido de otra célula o se han producido de forma sintética o recombinante. Generalmente, aunque no necesariamente, dichos ácidos nucleicos codifican proteínas, es decir, proteínas exógenas, que normalmente no son producidas por la célula en la que se transcribe o expresa el ADN. De manera similar, el ARN exógeno codifica proteínas que normalmente no se expresan en la célula en la que está presente el ARN exógeno. Los ácidos nucleicos y proteínas heterólogos/exógenos también pueden denominarse ácidos nucleicos o proteínas extranjeros. Los términos ácido nucleico o proteína heterólogos o exógenos abarcan en el presente documento cualquier ácido nucleico o proteína que un experto en la técnica reconocería como extranjeros a la célula en la que se expresa. Los términos heterólogo y exógeno también se aplican a combinaciones no naturales de secuencias de ácidos nucleicos o aminoácidos, es decir, combinaciones en las que al menos dos de las secuencias combinadas son extranjeras entre sí.

[0031] La "actividad específica" de una enzima se entiende en el presente documento como la cantidad de actividad de una enzima particular por cantidad de proteína total de la célula huésped, normalmente expresada en unidades de actividad enzimática por mg de proteína total de la célula huésped. En el contexto de la presente invención, la actividad específica de una enzima particular puede aumentarse o disminuirse en comparación con la actividad específica de esa enzima en una célula huésped de tipo salvaje (por lo demás idéntica).

[0032] "Compuestos furánicos" se entiende en el presente documento como ácido 2,5-furano-dicarboxílico (FDCA) así como cualquier compuesto que tenga un grupo furano que pueda oxidarse a FDCA, este último se denomina en este documento un "precursor de FDCA" o un "precursor furánico de FDCA". Los precursores de FDCA incluyen al menos: 5-hidroximetilfurfural (HMF), 2,5-dihidroximetil furano (DHF o HMF-OH) o 2,5-bis(hidroximetil) furano (BHF), ácido 5-hidroximetil-2-furancarboxílico o ácido 5-hidroximetil-2-furoico (HMFCFA), ácido 5-formil-2-furoico (FFA) y 2,5-diformil furano (DFF). Se entiende además que en los "compuestos furánicos", el anillo de furano o cualquiera de sus grupos laterales sustituibles pueden estar sustituidos, p. ej., con fracciones de éter de OH, alquilo C1-C10, alquilo, alilo, arilo o RO-, incluyendo los grupos cíclicos, en el anillo de furano en cualquier posición disponible.

[0033] Cualquier referencia a secuencias de nucleótidos o aminoácidos accesibles en las bases de datos de secuencias públicas en este documento se hace a la versión de la entrada de la secuencia disponible en la fecha de presentación de este documento.

#### Descripción detallada de la invención

##### Células que expresan una HMFCFA deshidrogenasa

[0034] En un primer aspecto, la invención se refiere a una célula que tiene la capacidad de oxidar el ácido 5-hidroximetil-2-furancarboxílico (HMFCFA) a ácido 5-formilfuroico (FFA). La capacidad de oxidar HMFCFA a FFA preferiblemente se confiere a la célula o aumenta en la célula mediante la transformación de la célula con un constructo de ácidos nucleicos que comprende una secuencia de nucleótidos que codifica una deshidrogenasa

que tiene la capacidad de oxidar HMFCa a FFA. La deshidrogenasa es preferiblemente una alcohol deshidrogenasa (es decir, que tiene actividad EC 1.1). Por tanto, la célula es preferiblemente una célula que comprende un constructo de expresión para la expresión de una secuencia de nucleótidos que codifica una deshidrogenasa que tiene la capacidad de oxidar HMFCa a FFA. En una célula preferida de la invención, el constructo de expresión se puede expresar en la célula y la expresión de la deshidrogenasa preferiblemente confiere o aumenta en la célula la capacidad de oxidar HMFCa a FFA, en comparación con una célula correspondiente que carece del constructo de expresión, p. ej., una célula de tipo salvaje. La actividad específica de la enzima que oxida el HMFCa a FFA se incrementa preferiblemente en la célula en al menos un factor 1.05, 1.1, 1.2, 1.5, 2.0, 5.0, 10, 20, 50 o 100 en comparación con una célula correspondiente que carece de constructo de expresión.

[0035] Una deshidrogenasa que tiene la capacidad de oxidar HMFCa a FFA es, por tanto, una alcohol deshidrogenasa que tiene actividad de HMFCa deshidrogenasa. Si un polipéptido tiene actividad de HMFCa deshidrogenasa o no, se puede ensayar mediante la expresión del polipéptido en una célula huésped adecuada que sea incapaz de oxidar HMFCa a FFA y detectar si la expresión del polipéptido confiere a la célula la capacidad de oxidar HMFCa a FFA. Preferiblemente, la actividad de HMFCa deshidrogenasa se ensaya como se describe en el ejemplo IV en el presente documento, donde una secuencia de nucleótidos que codifica el polipéptido que se va a ensayar para determinar la actividad de HMFCa deshidrogenasa reemplaza el gen *hmfH* de *C. basilensis* en pBT'*hmfH*-adh (descrito en WO2012/064195), después de lo cual se introduce el plásmido que comprende la secuencia codificante del polipéptido que se va a ensayar para determinar la actividad de HMFCa deshidrogenasa en *P. putida* KT2440Δ*gcd* que contiene pJNN*hmfT*1(t) (descrito en WO2012064195). Los transformantes de *P. putida* que expresan el polipéptido que se va a ensayar para determinar la actividad de HMFCa deshidrogenasa se incuban con HMF y se extraen muestras a intervalos regulares para el análisis de FDCA. Un aumento de la producción de FDCA, en comparación con los transformantes de *P. putida* correspondientes que carecen del polipéptido que se va a ensayar para determinar la actividad de HMFCa deshidrogenasa (y el gen *hmfH*) se toma como una indicación de que el polipéptido tiene actividad de HMFCa deshidrogenasa.

[0036] La HMFCa deshidrogenasa expresada en la célula de la invención es preferiblemente una deshidrogenasa que depende de un cofactor seleccionado entre un dinucleótido de adenina, por ejemplo NADH o NADPH, un dinucleótido de flavina y adenina (FAD), un mononucleótido de flavina (FMN) y pirroloquinolina quinolona (PQQ).

[0037] La HMFCa deshidrogenasa expresada en la célula de la invención es además preferiblemente una alcohol deshidrogenasa que (también) tiene la capacidad de oxidar otros alcoholes furánicos, preferiblemente alcoholes furánicos con un grupo hidroxilo en la posición 2, a los aldehídos correspondientes. Por tanto, la HMFCa deshidrogenasa tiene preferiblemente la capacidad de oxidar 5-hidroximetilfurfural (HMF) a 2,5-diformil furano (DFF).

[0038] En una realización, la secuencia de nucleótidos que codifica la deshidrogenasa con la capacidad de oxidar HMFCa a FFA se selecciona del grupo que consta de:

(a) una secuencia de nucleótidos que codifica un polipéptido con actividad de HMFCa deshidrogenasa, cuyo polipéptido comprende una secuencia de aminoácidos que tiene al menos un 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 81.65, 81.7, 81.8, 81.85, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 95, 96, 97, 98, 99 o 100 % de identidad de secuencia con la secuencia de aminoácidos de cualquiera de las SEC ID N.º: 1 (*Aeribacillus pallidus*), SEC ID N.º: 2 (*Bacilo kribbensis*), SEC ID N.º: 3 (*Geobacillus kaustophilus*), SEC ID N.º: 4 (*Aneurinibacillus terranovensis*), SEC ID N.º: 5 (*Brevibacillus thermoruber*), SEC ID N.º: 6 (*Brevibacillus panacihumi*), SEC ID N.º: 7 (*Bacillus sp. FJAT-14578*), SEC ID N.º: 8 (*Desulfotomaculum kuznetsovii*), SEC ID N.º: 9 (*Desulfurispora thermophila*), SEC ID N.º: 10 (*Bacillus sp. L1 (2012)*) y SEC ID N.º: 11 (*Pelotomaculum thermopropionicum*);

(b) una secuencia de nucleótidos cuya hebra complementaria se hibrida con una secuencia de nucleótidos de (a); y

(c) una secuencia de nucleótidos cuya secuencia difiere de la secuencia de una secuencia de nucleótidos de (b) debido a la degeneración del código genético.

[0039] Por tanto, una secuencia de nucleótidos preferida de la invención codifica una HMFCa deshidrogenasa con una secuencia de aminoácidos que es idéntica a la de una HMFCa deshidrogenasa que se puede obtener de (o se encuentra de manera natural en) una bacteria de los órdenes de Bacillales o Clostridiales. En una realización preferida, la bacteria es de la familia Bacillaceae, más preferiblemente la bacteria es del género *Aeribacillus*, *Geobacillus* y *Bacillus*, de los cuales las especies *Aeribacillus pallidus*, *Bacillus kribbensis*, *Geobacillus kaustophilus*, *Aneurinibacillus terranovensis*, *Bacillus sp. FJAT-14578* y *Bacillus sp. L1 (2012)* son las más preferidas. En otra realización preferida, la bacteria es de la familia Paenibacillaceae, más preferiblemente una bacteria de los géneros *Aneurinibacillus* y *Brevibacillus*, de los cuales las especies *Aneurinibacillus terranovensis*, *Brevibacillus thermoruber* y *Brevibacillus panacihumi*, son las más preferidas. En

otra realización preferida, la bacteria es de la familia Peptococcaceae, más preferiblemente la bacteria es de los géneros *Desulfotomaculum*, *Desulfurispora* y *Pelotomaculum*, de los cuales las especies *Desulfotomaculum kuznetsovii*, *Desulfurispora thermophila* y *Pelotomaculum thermopropionicum* son las más preferidas.

5 [0040] En una realización, una secuencia de nucleótidos preferida de la invención codifica una HMFCA deshidrogenasa de una bacteria mesófila, es decir, una bacteria que crece mejor a temperatura moderada, típicamente entre 20 y 45 °C. Preferiblemente, la secuencia de nucleótidos de la invención codifica una HMFCA deshidrogenasa mesófila con una actividad y estabilidad óptimas en el intervalo entre 20 y 45 °C. Ejemplos de tales deshidrogenasas mesófilas son, p. ej., las deshidrogenasas de *Bacillus kribbensis* (30 °C), *Aneurinibacillus terranovensis* (40 °C), *Brevibacillus thermoruber* (45 °C), *Brevibacillus panacihumi* (30 °C), *Bacillus sp. FJAT-14578* (30 °C) y *Bacillus sp. L1 (2012)* (30 - 50 °C) y deshidrogenasa relacionada con los mismos.

15 [0041] En una realización, una secuencia de nucleótidos preferida de la invención codifica una HMFCA deshidrogenasa de una bacteria termófila, es decir, una bacteria que crece mejor a temperaturas relativamente altas, típicamente entre 45 y 122 °C. Preferiblemente, la secuencia de nucleótidos de la invención codifica así una HMFCA deshidrogenasa termófila con una actividad y estabilidad óptimas en el intervalo entre 45 y 122 °C. Ejemplos de tales deshidrogenasas termófilas son, p. ej., las deshidrogenasas de *Aeribacillus pallidus* (55 °C), *Geobacillus kaustophilus* (55 °C), *Desulfotomaculum kuznetsovii* (60 °C), *Desulfurispora thermophila* (50 °C), *Pelotomaculum thermopropionicum* (55 °C) y *Bacillus sp. L1 (2012)* (30 - 50 °C) y deshidrogenasa relacionada con los mismos.

25 [0042] En una realización, la secuencia de nucleótidos codifica un polipéptido con actividad de HMFCA deshidrogenasa tal como ocurre en la naturaleza, p. ej., ya que puede aislarse de un organismo fuente de tipo salvaje. Alternativamente, la secuencia de nucleótidos puede codificar formas diseñadas de cualquiera de las HMFCA deshidrogenasas definidas anteriormente y que comprenden una o más sustituciones, inserciones y/o deleciones de aminoácidos en comparación con la correspondiente HMFCA deshidrogenasa de origen natural pero que están dentro de los rangos de identidad o similitud tal como se define en este documento. Por tanto, en una realización, la secuencia de nucleótidos de la invención codifica una HMFCA deshidrogenasa cuya secuencia de aminoácidos comprende al menos en cada una de las posiciones invariables (que se indican en la tabla 2 con un "\*\*\*") el aminoácido presente en una posición invariable. Preferiblemente, la secuencia de aminoácidos también comprende en las posiciones fuertemente conservadas (que se indican en la tabla 2 con un ":",) uno de los aminoácidos presentes en una posición fuertemente conservada. Más preferiblemente, la secuencia de aminoácidos comprende además en las posiciones menos fuertemente conservadas (que se indican en la tabla 2 con un ".") uno de los aminoácidos presentes en una posición menos fuertemente conservada. Es menos improbable que las sustituciones de aminoácidos fuera de estas posiciones invariables y conservadas afecten a la actividad deshidrogenasa de HMFCA.

35 [0043] Las secuencias de nucleótidos de la invención, que codifican polipéptidos con actividad de HMFCA deshidrogenasa, se pueden obtener a partir de ADN genómico y/o de ADNc de un hongo, levadura o bacteria, p. ej., uno que pertenece al mismo filo, clase o género que los organismos fuente descritos anteriormente, usando métodos para el aislamiento de secuencias de nucleótidos que son bien conocidos en la técnica per se (ver, p. ej., Sambrook y Russell (2001) *Molecular Cloning: A Laboratory Manual* (tercera edición), Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Nueva York). Las secuencias de nucleótidos de la invención se pueden obtener, p. ej., en un proceso en el que a) se usan cebadores de PCR degenerados (diseñados sobre la base de secuencias de aminoácidos conservadas) en ADNc y/o ADN genómico de un organismo adecuado para generar un fragmento de PCR que comprende parte de las secuencias de nucleótidos que codifican los polipéptidos con actividad de HMFCA deshidrogenasa; b) el fragmento de PCR obtenido en a) se usa como sonda para cribar una biblioteca de ADNc y/o genómica del organismo; y c) producir un ADNc o ADN genómico que comprende la secuencia de nucleótidos que codifica un polipéptido con actividad de HMFCA deshidrogenasa.

40 [0044] Para aumentar la probabilidad de que una HMFCA deshidrogenasa de la invención se exprese en niveles suficientes y en forma activa en las células transformadas de la invención, la secuencia de nucleótidos que codifica a estas enzimas, así como otras enzimas de la invención (ver más adelante), se adaptan preferiblemente para optimizar su uso de codones al de la célula huésped en cuestión. La adaptabilidad de una secuencia de nucleótidos que codifica un polipéptido para el uso de codones de una célula huésped puede expresarse como índice de adaptación de codones (CAI). El índice de adaptación de codones se define en el presente documento como una medida de la adaptabilidad relativa del uso de codones de un gen al uso de codones de genes altamente expresados en una célula u organismo huésped particular. La adaptabilidad relativa (w) de cada codón es la relación entre el uso de cada codón y el del codón más abundante para el mismo aminoácido. El índice CAI se define como la media geométrica de estos valores de adaptabilidad relativa. Se excluyen los codones no sinónimos y los codones de terminación (que dependen del código genético). Los valores de CAI van de 0 a 1, y los valores más altos indican una mayor proporción de los codones más abundantes (ver Sharp y Li, 1987, *Nucleic Acids Research* 15: 1281-1295; ver también: Jansen et al., 2003, *Nucleic Acids Res.* 31 (8): 2242-51).

55 Una secuencia de nucleótidos adaptada tiene preferiblemente un CAI de al menos 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8

o 0.9. Las preferidas son las secuencias enumeradas en las SEC ID N.º: 13 o 14, que han sido optimizadas con respecto a codones para su expresión en células de *P. putida*.

[0045] La célula huésped que se va a transformar con un constructo de ácidos nucleicos para la expresión de la secuencia de nucleótidos que codifica una HMFCa deshidrogenasa de la invención puede ser, en principio, cualquier célula huésped en la que la invención de HMFCa deshidrogenasa pueda expresarse adecuadamente, preferiblemente en forma funcional, es decir, activa. La célula huésped de la invención es preferiblemente un huésped capaz de transportar compuestos furánicos de forma activa o pasiva tanto dentro como fuera de la célula. Una célula huésped preferida de la invención no tiene o carece de actividades detectables que descarboxilen compuestos furánicos carboxilados, tales como, en particular, HMFCa, FFA y FDCA. Preferiblemente, dicha célula huésped carece naturalmente de la capacidad de descarboxilar compuestos furánicos carboxilados.

[0046] Preferiblemente, la célula huésped es una célula cultivada, p. ej., una célula que puede cultivarse en un proceso de fermentación, preferiblemente en fermentación sumergida.

[0047] Según una realización, la célula huésped según la invención es una célula huésped eucariótica. Preferiblemente, la célula eucariótica es una célula de mamífero, insecto, planta, hongo o alga. Las células de mamífero preferidas incluyen, p. ej., células de ovario de hámster chino (CHO), células COS, células 293, células PerC6 e hibridomas. Las células de insecto preferidas incluyen, p. ej., células Sf9 y Sf21 y derivados de las mismas.

[0048] Sin embargo, preferiblemente, la célula huésped es una célula microbiana. La célula puede ser una célula microbiana eucariótica, preferiblemente una célula fúngica, como, p. ej., una levadura o una célula fúngica filamentosa. Las células huésped de levadura preferidas incluyen, p. ej., células de levaduras de géneros como *Candida*, *Hansenula*, *Kluyveromyces*, *Pichia*, *Saccharomyces*, *Schizosaccharomyces* y *Yarrowia*. Más preferiblemente levaduras de especies como *Kluyveromyces lactis*, *S. cerevisiae*, *Hansenula polymorpha*, *Yarrowia lipolytica* y *Pichia pastoris*. Las células fúngicas filamentosas preferidas incluyen, p. ej., células de hongos filamentosos de géneros como *Acremonium*, *Agaricus*, *Aspergillus*, *Aureobasidium*, *Myceliophthora*, *Chrysosporium*, *Coprinus*, *Cryptococcus*, *Filibasidium*, *Fusarium*, *Humicola*, *Magnaporthe*, *Mucor*, *Myceliophthora*, *Neocallimastix*, *Neurospora*, *Paecilomyces*, *Penicillium*, *Piromyces*, *Panerochaete*, *Pleurotus*, *Schizophyllum*, *Talaromyces*, *Thermoascus*, *Thielavia*, *Tolypocladium* y *Trichoderma*. Las células fúngicas filamentosas preferidas pertenecen a una especie de un género de *Aspergillus*, *Myceliophthora*, *Penicillium*, *Myceliophthora*, *Talaromyces* o *Trichoderma*, y más preferiblemente una especie seleccionada entre *Aspergillus niger*, *Aspergillus awamori*, *Aspergillus foetidus*, *Aspergillus sojae*, *Aspergillus fumigatus*, *Talaromyces emersonii*, *Aspergillus oryzae*, *Myceliophthora thermophila*, *Trichoderma reesei* y *Penicillium chrysogenum*.

[0049] La célula huésped microbiana también puede ser una célula procariótica, preferiblemente una célula bacteriana. El término "célula bacteriana" incluye microorganismos tanto gram negativos como gram positivos. Las bacterias adecuadas pueden seleccionarse del género *Escherichia*, *Anabaena*, *Aeribacillus*, *Aneurinibacillus*, *Burkholderia*, *Bradyrhizobium*, *Caulobacter*, *Cupriavidus*, *Desulfotomaculum*, *Desulfurispora*, *Gluconobacter*, *Rhodobacter*, *Pelotomaculum*, *Pseudomonas*, *Paracoccus*, *Bacillus*, *Geobacillus*, *Brevibacillus*, *Brevibacterium*, *Corynebacterium*, *Rhizobium* (*Sinorhizobium*), *Flavobacterium*, *Klebsiella*, *Enterobacter*, *Lactobacillus*, *Lactococcus*, *Methylobacterium*, *Ralstonia*, *Rhodopseudomonas*, *Staphylococcus* y *Streptomyces*. Preferiblemente, la célula bacteriana se selecciona de una especie del grupo que consta de *A. pallidus*, *A. terranovensis*, *B. subtilis*, *B. amyloliquefaciens*, *B. coagulans*, *B. kribbensis*, *B. licheniformis*, *B. punitis*, *B. megaterium*, *B. halodurans*, *B. pumilus*, *B. thermoruber*, *B. panacihumi*, *C. basiliensis*, *D. kuznetsovii*, *D. thermophila*, *G. kaustophilus*, *Gluconobacter oxydans*, *Caulobacter crescentus* CB 15, *Methylobacterium extorquens*, *Rhodobacter sphaeroides*, *Pelotomaculum thermopropionicum*, *Pseudomonas zeaxanthinifaciens*, *Pseudomonas putida*, *Paracoccus denitrificans*, *E. coli*, *C. glutamicum*, *Staphylococcus carnosus*, *Streptomyces lividans*, *Sinorhizobium meliloti* y *Rhizobium radiobacter*. Dentro de la especie *Pseudomonas putida*, se prefieren las cepas de *P. putida* S12 y *P. putida* KT2440.

[0050] Para usos específicos de un compuesto producido en una célula huésped según la invención, la selección de la célula huésped puede realizarse según dicho uso. Donde, p. ej., el compuesto producido en una célula huésped según la invención se va a utilizar en aplicaciones alimentarias, una célula huésped puede seleccionarse de un organismo de calidad alimentaria como *Saccharomyces cerevisiae*. Los usos específicos incluyen, entre otros, alimentos, piensos (animales), farmacéuticos, agrícolas tales como protección de cultivos y/o aplicaciones de cuidado personal.

[0051] El constructo de expresión para la expresión de una secuencia de nucleótidos que codifica una HMFCa deshidrogenasa de la invención es preferiblemente un constructo de expresión que es heterólogo o exógeno a la célula huésped transformada con el constructo. En el presente documento se entiende que un constructo es heterólogo o exógeno a la célula huésped que comprende el constructo cuando el constructo comprende al menos una secuencia o elemento de secuencia que no se produce de forma natural en la célula huésped y/o cuando el constructo comprende al menos dos elementos de secuencia en una combinación y/u orden que no

ocurre naturalmente en la célula huésped, incluso si los elementos mismos ocurren de manera natural en la célula huésped.

5 [0052] Los vectores y constructos de expresión para la expresión de una secuencia de nucleótidos que codifica una HMFCa deshidrogenasa de la invención en células huésped apropiadas se describen con más detalle en el presente documento más adelante.

10 [0053] Una célula transformada que expresa una HMFCa deshidrogenasa de la invención, además, preferiblemente tiene actividad de aldehído deshidrogenasa (es decir, tiene actividad de EC 1.2). Preferiblemente, la actividad de la aldehído deshidrogenasa es capaz de convertir aldehídos furánicos. Más preferiblemente, la actividad de aldehído deshidrogenasa es capaz de oxidar los aldehídos furánicos a los correspondientes ácidos carboxílicos furánicos. Más específicamente, la actividad de aldehído deshidrogenasa es preferiblemente capaz de al menos una de las siguientes opciones: i) oxidar HMF a HMFCa, ii) oxidar 2,5-diformil furano (DFF) a ácido 5-formil-2-furoico (FFA) y iii) FFA en FDCA. Dicha actividad de aldehído deshidrogenasa furánica puede ser una actividad endógena de la célula o puede ser una actividad exógena conferida a la célula. Preferiblemente, la actividad de aldehído deshidrogenasa furánica se confiere o aumenta en la célula mediante la transformación de la célula con un segundo constructo de expresión. En una célula preferida de la invención, el segundo constructo de expresión se puede expresar en la célula y la expresión de la aldehído deshidrogenasa furánica confiere o aumenta preferiblemente en la célula la capacidad de al menos una de las siguientes opciones: i) oxidar HMF a HMFCa, ii) oxidar DFF a FFA y iii) oxidar FFA a FDCA, en comparación con una célula correspondiente que carece del constructo de expresión, p. ej., una célula de tipo salvaje. La actividad específica de la aldehído deshidrogenasa furánica se incrementa preferiblemente en la célula en al menos un factor 1.05, 1.1, 1.2, 1.5, 2.0, 5.0, 10, 20, 50 o 100 en comparación con una célula correspondiente que carece de constructo de expresión. El segundo constructo de expresión comprende preferiblemente una secuencia de nucleótidos que codifica un polipéptido:

25 a) que tiene al menos una de las capacidades de i) oxidar HMF a HMFCa, ii) oxidar DFF a FFA y iii) oxidar FFA a FDCA; y  
 30 b) que comprende una secuencia de aminoácidos que tiene al menos un 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99 o 100 % de identidad de secuencia con la secuencia de aminoácidos de cualquiera de las SEC ID N.º: 24, 25, 26, 27, 28, 29 y 30.

35 [0054] La capacidad de un polipéptido para oxidar al menos una de las siguientes: i) HMF a HMFCa, ii) DFF a FFA y iii) FFA a FDCA, puede ensayarse mediante la coexpresión de una secuencia de nucleótidos que codifica el polipéptido en una célula huésped de *P. putida*, preferiblemente una célula huésped de *P. putida* KT2440, junto con los genes HmfH y HmfT1 de *C. basilensis* HMF 14, incubando células de *P. putida* en HMF 10 mM y detectando un aumento en la acumulación de FDCA en comparación con las células correspondientes de *P. putida* que no expresan el polipéptido, p. ej. como se describe en el ejemplo IV de WO2012/064195. La capacidad de un polipéptido para oxidar HMF a HMFCa también puede ensayarse como describen Koopman et al. 2010, PNAS *supra*). Cepas que expresan el gen HmfT1 de *C. basilensis* HMF14 en el presente documento se entiende que expresa un producto génico que tiene la secuencia de aminoácidos de la SEC ID N.º: 31.

45 [0055] Una célula transformada que expresa una HMFCa deshidrogenasa de la invención, además, preferiblemente tiene la capacidad de transportar compuestos furánicos hacia y/o desde la célula. Preferiblemente, la célula tiene la capacidad de transportar compuestos furánicos que son precursores de FDCA al interior de la célula y preferiblemente la capacidad de transportar FDCA fuera de la célula. Dichas capacidades de transporte de compuestos furánicos pueden ser capacidades endógenas de la célula y/o pueden ser capacidades exógenas conferidas a la célula. Por tanto, una célula preferida de la invención expresa un polipéptido que tiene capacidades de transporte de compuestos furánicos. Más preferiblemente, la célula expresa un polipéptido que tiene capacidades de transporte de HMFCa. Se entiende que las capacidades de transporte de HMFCa incluyen al menos la capacidad de transportar HMFCa a la célula. La expresión de un polipéptido que tiene capacidades de transporte de HMFCa aumentará el transporte de HMFCa al interior de la célula, lo que aumenta su disponibilidad para la conversión intracelular a FDCA. Por tanto, se puede mejorar la bioconversión de HMFCa.

60 [0056] Preferiblemente, la capacidad para transportar compuestos furánicos hacia y/o desde de la célula se confiere o aumenta en la célula mediante la transformación de la célula con un tercer constructo de expresión. En una célula preferida de la invención, el tercer constructo de expresión se puede expresar en la célula y la expresión del transportador del compuesto furánico preferiblemente confiere o aumenta en la célula la capacidad de transportar al menos HMFCa al interior de la célula, en comparación con una célula correspondiente que carece del constructo de expresión, p. ej., una célula de tipo salvaje. El tercer constructo de expresión comprende preferiblemente una secuencia de nucleótidos que codifica un polipéptido:

65 a) que tiene al menos capacidad de transporte de HMFCa; y

b) que comprende una secuencia de aminoácidos que tiene al menos un 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99 o 100 % de identidad de secuencia con la secuencia de aminoácidos de cualquiera de las SEC ID N.º: 17, 31, 32, 33 y 34.

5 [0057] La capacidad de un polipéptido para transportar compuestos furánicos, en particular HMFCa, al interior de la célula puede ensayarse mediante la coexpresión de una secuencia de nucleótidos que codifica el polipéptido transportador en una célula huésped de *P. putida*, preferiblemente una célula huésped de *P. putida* KT2440, junto con el gen HmfH de *C. basilensis* HMF 14 y un gen que codifica una aldehído deshidrogenasa furánica asociada con el operón de degradación de HMF de *C. basilensis* HMF 14 (que tiene la secuencia de aminoácidos de la SEC ID N.º: 19 de WO2012/064195), incubando células de *P. putida* en HMF 10 mM y detectando un aumento en la acumulación de FDCA en comparación con las células correspondientes de *P. putida* que no expresan el polipéptido transportador, p. ej. como se describe en el ejemplo IV de WO2012/064195.

15 [0058] En una realización, la secuencia de nucleótidos codifica un polipéptido que tiene capacidad de transporte de HMFCa tal como ocurre en la naturaleza, p. ej., ya que puede aislarse de un organismo fuente de tipo salvaje. Alternativamente, la secuencia de nucleótidos puede codificar formas diseñadas de cualquiera de los polipéptidos que tienen capacidad de transporte de HMFCa tal como se definió anteriormente y que comprenden una o más sustituciones, inserciones y/o deleciones de aminoácidos en comparación con el correspondiente polipéptido natural con capacidad de transporte de HMFCa pero que están dentro de los rangos de identidad o similitud tal como se define en el presente documento. Por tanto, en una realización, la secuencia de nucleótidos de la invención codifica un polipéptido que tiene capacidad de transporte de HMFCa, cuya secuencia de aminoácidos comprende al menos en cada una de las posiciones invariables (que se indican en la tabla 3 con un "\*\*") el aminoácido presente en una posición invariable. Preferiblemente, la secuencia de aminoácidos también comprende en las posiciones fuertemente conservadas (que se indican en la tabla 3 con un ":") uno de los aminoácidos presentes en una posición fuertemente conservada. Más preferiblemente, la secuencia de aminoácidos comprende además en las posiciones menos fuertemente conservadas (que se indican en la tabla 3 con un ".") uno de los aminoácidos presentes en una posición menos fuertemente conservada. Es menos probable que las sustituciones de aminoácidos fuera de estas posiciones invariables y conservadas afecten a la capacidad de transporte de HMFCa.

25 [0059] Las secuencias de nucleótidos de la invención, que codifican polipéptidos que tienen capacidad de transporte de HMFCa, se pueden obtener a partir de ADNc y/o ADN genómico de un hongo, levadura o bacteria, p. ej., uno que pertenece al mismo filo, clase o género que los organismos fuente descritos anteriormente, utilizando métodos para el aislamiento de secuencias de nucleótidos que son bien conocidos en la técnica per se, de una manera similar a la descrita anteriormente para las secuencias de nucleótidos que codifican las HMFCa deshidrogenasas de la invención.

#### 40 Células que expresan un transportador de compuestos furánicos.

45 [0060] En un segundo aspecto, la divulgación se refiere a una célula que expresa una secuencia de nucleótidos que codifica un polipéptido que tiene capacidades de transporte de compuestos furánicos. Preferiblemente, la célula se transforma con un constructo de expresión para la expresión de una secuencia de nucleótidos que codifica un polipéptido que tiene capacidades de transporte de compuestos furánicos. El polipéptido que tiene capacidades de transporte de compuestos furánicos es preferiblemente un polipéptido que tiene capacidades de transporte de HMFCa, que al menos incluye la capacidad de transportar HMFCa al interior de la célula. Preferiblemente, la célula comprende un constructo de expresión para la expresión de una secuencia de nucleótidos que codifica un polipéptido que tiene la capacidad de transportar al menos HMFCa al interior de la célula, donde el polipéptido comprende una secuencia de aminoácidos con al menos un 86.5, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99 o 100 % de identidad con la secuencia de aminoácidos de la SEC ID N.º: 17, donde, el constructo de expresión se puede expresar en la célula y la expresión del polipéptido confiere a o aumenta en la célula la capacidad de transportar al menos HMFCa al interior de la célula, en comparación con una célula de tipo salvaje correspondiente que carece del constructo de expresión. La capacidad de un polipéptido para transportar compuestos furánicos, en particular HMFCa, al interior de la célula puede ensayarse como se ha descrito anteriormente.

55 [0061] Preferiblemente, una célula transformada que expresa un transportador de compuestos furánicos de este aspecto de la divulgación comprende además actividades enzimáticas para convertir HMF en FDCA, donde las actividades para convertir HMF en FDCA incluyen preferiblemente al menos una de las siguientes opciones:

- 60 a) una alcohol deshidrogenasa que oxida el HMFCa a FFA y una actividad de aldehído deshidrogenasa que oxida los aldehídos furánicos a los correspondientes ácidos carboxílicos furánicos; y  
 b) una oxidorreductasa, preferiblemente una oxidasa, que oxida uno o más de los siguientes: HMF, 2,5-dihidroximetil furano, HMFCa, FFA y 2,5-diformil furano a FDCA y opcionalmente una actividad aldehído deshidrogenasa que oxida los aldehídos a los correspondientes ácidos carboxílicos furánicos.

[0062] La alcohol deshidrogenasa que oxida el HMFCa a FFA y la actividad de aldehído deshidrogenasa que oxida los aldehídos furánicos son preferiblemente como se han definido en el presente documento anteriormente. La oxidorreductasa que oxida uno o más de los compuestos HMF, 2,5-dihidroximetil furano, HMFCa, FFA y 2,5-diformal furano a FDCA es preferiblemente una oxidorreductasa que tiene actividades de EC 1.1 y EC 1.2, como se describe en WO2011/026913.

[0063] A menos que se especifique lo contrario, una célula transformada que expresa un transportador de compuestos furánicos de este aspecto de la divulgación puede tener además las características de una célula que expresa una HMFCa deshidrogenasa del primer aspecto de la invención tal como se ha definido anteriormente.

#### Vectores, constructos y método para la expresión de polipéptidos de la invención.

[0064] Otro aspecto de la invención se refiere a constructos de ácidos nucleicos, tales como vectores, que incluyen vectores de clonación y expresión, que comprenden un polinucleótido de la invención, p. ej., una secuencia de nucleótidos que codifica una HMFCa deshidrogenasa o un transportador de la invención o un equivalente funcional de los mismos y métodos para hacer crecer, transformar o transfectar dichos vectores en una célula huésped adecuada, por ejemplo en condiciones en las que se produce la expresión de un polipéptido de la invención. Según se usa en el presente documento, los términos "vector" y "constructo" se usan de manera intercambiable y se refieren a una molécula de ácidos nucleicos construida que comprende y es preferiblemente capaz de transportar un polinucleótido de la invención.

[0065] Los polinucleótidos de la invención pueden incorporarse a un vector replicable recombinante, por ejemplo, un vector de clonación o expresión. El vector puede usarse para replicar el ácido nucleico en una célula huésped compatible. Por tanto, en una realización adicional, la invención proporciona un método para fabricar polinucleótidos de la invención mediante la introducción de un polinucleótido de la invención en un vector replicable, la introducción del vector en una célula huésped compatible y el crecimiento de la célula huésped en condiciones que provocan la replicación del vector. El vector puede recuperarse de la célula huésped. Las células huésped adecuadas se han descrito anteriormente.

[0066] El vector en el que se inserta el casete de expresión o polinucleótido de la invención puede ser cualquier vector que se pueda someter convenientemente a procedimientos de ADN recombinante, y la elección del vector dependerá a menudo de la célula huésped en la que se va a introducir.

[0067] Un vector según la invención puede ser un vector de replicación autónoma, es decir, un vector que existe como una entidad extracromosómica, cuya replicación es independiente de la replicación cromosómica, p. ej., un plásmido. Alternativamente, el vector puede ser tal que, cuando se introduce en una célula huésped, está integrado en el genoma de la célula huésped y se replica junto con el cromosoma o cromosomas en los que se ha integrado.

[0068] Un tipo de vector es un "plásmido", que se refiere a un bucle circular de ADN bicatenario en el que se pueden ligar segmentos de ADN adicionales. Otro tipo de vector es un vector vírico, en el que se pueden ligar segmentos de ADN adicionales en el genoma vírico. Ciertos vectores son capaces de replicación autónoma en una célula huésped en la que se introducen (por ejemplo, vectores bacterianos que tienen un origen de replicación bacteriano y vectores de mamíferos episomales). Otros vectores (p. ej., vectores de mamíferos no episomales) se integran en el genoma de una célula huésped tras la introducción en la célula huésped y, por tanto, se replican junto con el genoma huésped. Además, ciertos vectores son capaces de dirigir la expresión de genes a los que están operativamente enlazados. Dichos vectores se denominan en el presente documento "vectores de expresión". En general, los vectores de expresión útiles para las técnicas de ADN recombinante se encuentran a menudo en forma de plásmidos. Los términos "plásmido" y "vector" pueden usarse indistintamente en este documento, ya que el plásmido es la forma de vector más comúnmente utilizada. Sin embargo, la invención pretende incluir otras formas de vectores de expresión, tales como cósmidos, vectores víricos (por ejemplo, retrovirus de replicación defectuosa, adenovirus y virus adenoasociados) y vectores de fagos que cumplen funciones equivalentes.

[0069] Se pueden utilizar *in vitro* vectores según la invención, por ejemplo, para la producción de ARN o para transfectar o transformar una célula huésped.

[0070] Un vector de la invención puede comprender dos o más, por ejemplo tres, cuatro o cinco, polinucleótidos de la invención, por ejemplo, para sobreexpresión.

[0071] Los vectores de expresión recombinantes de la invención comprenden un ácido nucleico de la invención en una forma adecuada para la expresión del ácido nucleico en una célula huésped, lo que significa que el vector de expresión recombinante incluye una o más secuencias reguladoras, seleccionadas sobre la base de las células huésped que se van a utilizar para la expresión, que están operativamente enlazadas a la secuencia de ácidos nucleicos que se va a expresar. Una secuencia reguladora, por ejemplo, un promotor, un potenciador u

otra señal de regulación de la expresión, "operativamente enlazada" a una secuencia codificante se coloca de tal manera que la expresión de la secuencia codificante se logra bajo condiciones compatibles con las secuencias de control o las secuencias se disponen de manera que funcionan conjuntamente para su propósito previsto, por ejemplo, la transcripción se inicia en un promotor y avanza a través de la secuencia de ADN que codifica el polipéptido. El término "secuencia reguladora" o "secuencia de control" incluye promotores, potenciadores y otros elementos de control de la expresión (por ejemplo, señales de poliadenilación). Dichas secuencias reguladoras se describen, por ejemplo, en Goeddel; Gene Expression Technology: Methods in Enzymology 185, Academic Press, San Diego, CA (1990). El término secuencias reguladoras o de control incluye aquellas secuencias que dirigen la expresión constitutiva de una secuencia de nucleótidos en muchos tipos de células huésped y aquellas que dirigen la expresión de la secuencia de nucleótidos solo en una determinada célula huésped (por ejemplo, secuencias reguladoras específicas de tejido).

[0072] Un vector o constructo de expresión para una célula huésped determinada puede comprender, por tanto, los siguientes elementos operativamente enlazados entre sí en un orden consecutivo desde el extremo 5' hasta el extremo 3' en relación con la cadena codificante de la secuencia que codifica el polipéptido de la primera invención: (1) una secuencia promotora capaz de dirigir la transcripción de la secuencia de nucleótidos que codifica el polipéptido en la célula huésped dada; (2) secuencias de iniciación de la traducción, como la secuencia de consenso de Kozak eucariótica o la secuencia del sitio de unión al ribosoma procariótica/Shine-Dalgarno, (3) opcionalmente, una secuencia señal capaz de dirigir la secreción del polipéptido de la célula huésped dada a un medio de cultivo; (4) una secuencia de ADN de la invención que codifica una forma madura y preferiblemente activa de un polipéptido de la invención; y preferiblemente también (5) una región de terminación de la transcripción (terminador) capaz de terminar la transcripción corriente abajo de la secuencia de nucleótidos que codifica el polipéptido.

[0073] Corriente abajo de la secuencia de nucleótidos según la invención puede haber una región 3' sin traducir que contiene uno o más sitios de terminación de la transcripción (por ejemplo, un terminador). El origen del terminador es menos crítico. El terminador puede, por ejemplo, ser nativo de la secuencia de ADN que codifica el polipéptido. Sin embargo, preferiblemente se usa un terminador de levadura en células huésped de levadura y un terminador fúngico filamentososo se usa en células huésped fúngicas filamentosas. Más preferiblemente, el terminador es endógeno a la célula huésped (en la que se va a expresar la secuencia de nucleótidos que codifica el polipéptido). En la región transcrita, puede estar presente un sitio de unión al ribosoma para traducción. La parte codificante de los transcritos maduros expresados por los constructos incluirá un AUG que iniciará la traducción al principio y un codón de terminación posicionado apropiadamente al final del polipéptido que se va a traducir.

[0074] La expresión mejorada del polinucleótido de la invención también se puede lograr mediante la selección de regiones reguladoras heterólogas, p. ej., regiones promotoras, líderes de secreción y/o terminadoras, que pueden servir para aumentar la expresión y, si se desea, los niveles de secreción de la proteína de interés desde el huésped de expresión y/o para proporcionar el control inducible de la expresión de un polipéptido de la invención.

[0075] Los expertos en la técnica apreciarán que el diseño del vector de expresión puede depender de factores tales como la elección de la célula huésped que se va a transformar, el nivel de expresión de la proteína deseada, etc. Los vectores de la invención, tales como los vectores de expresión, pueden introducirse en células huésped para producir proteínas o péptidos, codificados por ácidos nucleicos como se describe en el presente documento (por ejemplo, HMFCA deshidrogenasa o un transportador de la invención, formas mutantes de la misma, fragmentos, variantes o equivalentes funcionales de la misma, proteínas de fusión, etc.).

[0076] Según se ha expuesto anteriormente, el término "secuencias de control" o "secuencias reguladoras" se define en el presente documento para incluir al menos cualquier componente que pueda ser necesario y/o beneficioso para la expresión de un polipéptido. Cualquier secuencia de control puede ser nativa o extranjera a la secuencia de ácidos nucleicos de la invención que codifica un polipéptido. Dichas secuencias de control pueden incluir, entre otras, un promotor, un líder, secuencias óptimas de iniciación de traducción (tal como se describe en Kozak, 1991, J. Biol. Chem. 266: 19867-19870) o las secuencias procarióticas Shine-Dalgarno, una secuencia señal de secreción, una secuencia propéptido, una secuencia de poliadenilación y un terminador de la transcripción. Como mínimo, las secuencias de control incluyen típicamente un promotor y señales de inicio y parada traduccional.

[0077] Un microorganismo transformado de forma estable es aquel al que se le han introducido uno o más fragmentos de ADN de manera que las moléculas introducidas se mantienen, replican y segregan en un cultivo en crecimiento. La transformación estable puede deberse a integraciones cromosómicas múltiples o únicas o a un elemento o elementos extracromosómicos como un vector plasmídico. Un vector plasmídico es capaz de dirigir la expresión de polipéptidos codificados por fragmentos de ADN particulares.

[0078] La expresión puede ser constitutiva o regulada por promotores inducibles (o reprimibles) que permiten

altos niveles de transcripción de fragmentos de ADN funcionalmente asociados que codifican polipéptidos específicos.

5 [0079] Independientemente del mecanismo exacto utilizado para la expresión de polipéptidos de la invención, se contempla que dicha expresión sea transferible mediante la introducción de genes que codifican estos polipéptidos en otra célula huésped mediante métodos conocidos en la técnica. Los elementos genéticos tal como se definen en el presente documento incluyen ácidos nucleicos (generalmente ADN o ARN) que tienen secuencias codificantes expresables para productos tales como proteínas, específicamente enzimas, apoproteínas o ARN antisentido, que expresan o regulan la expresión de polipéptidos relevantes. Las proteínas expresadas pueden funcionar como enzimas, reprimir o desreprimir la actividad enzimática o controlar la expresión de enzimas o funcionar como transportadores de compuestos, p. ej., metabolitos. El ADN recombinante que codifica estas secuencias expresables puede ser cromosómico (integrado en el cromosoma de la célula huésped mediante, por ejemplo, recombinación homóloga) o extracromosómico (por ejemplo, transportado por uno o más plásmidos, cósmidos y otros vectores capaces de autorreplicarse). Se entiende que el ADN recombinante utilizado para transformar la célula huésped según esta invención puede incluir, además de genes estructurales y factores de transcripción, secuencias de control de expresión, incluyendo promotores, represores y potenciadores, que actúan para controlar la expresión o desrepresión de secuencias codificantes de proteínas, apoproteínas o ARN antisentido. Por ejemplo, dichas secuencias de control pueden insertarse en células huésped de tipo salvaje para promover la sobreexpresión de polipéptidos seleccionados ya codificados en el genoma de la célula huésped, o alternativamente pueden usarse para controlar la síntesis de polipéptidos codificados extracromosómicamente.

25 [0080] El ADN recombinante puede introducirse en la célula huésped por cualquier medio, incluyendo, entre otros, plásmidos, cósmidos, fagos, cromosomas artificiales de levadura u otros vectores que median la transferencia de elementos genéticos a una célula huésped. Estos vectores pueden incluir un origen de replicación, junto con elementos de control que actúan en cis que controlan la replicación del vector y los elementos genéticos transportados por el vector. Pueden estar presentes marcadores seleccionables en el vector para ayudar en la identificación de células huésped en las que se han introducido elementos genéticos.

30 [0081] Los medios para introducir elementos genéticos en una célula huésped (por ejemplo, clonación) son bien conocidos por los expertos en la materia. Se puede utilizar un vector plasmídico extracromosómico de múltiples copias para insertar los elementos genéticos según la presente invención. La introducción del elemento genético a través del plásmido en las células huésped implica una escisión inicial de un vector plasmídico con una enzima de restricción, seguida de la ligación del plásmido y los elementos genéticos que codifican la especie enzimática diana de acuerdo con la invención. Tras la recircularización del plásmido recombinante ligado, se utiliza una infección (por ejemplo, empaquetamiento en fago lambda) u otro mecanismo de transferencia de plásmido (por ejemplo, electroporación, microinyección, etc.) para transferir el plásmido a la célula huésped. Los plásmidos adecuados para la inserción de elementos genéticos en la célula huésped son bien conocidos por los expertos en la materia.

40 [0082] Otros métodos de clonación de genes incluyen, entre otros, la integración directa del material genético en el cromosoma. Esto puede ocurrir por una variedad de medios, incluida la clonación de los elementos genéticos descritos en el presente documento en plásmidos no replicantes flanqueados por secuencias de ADN homólogas del cromosoma huésped; tras transformar dicho plásmido recombinante en un huésped, los elementos genéticos pueden introducirse en el cromosoma mediante recombinación de ADN. Estas cepas recombinantes se pueden recuperar si los fragmentos de ADN que se integran contienen un marcador seleccionable, como la resistencia a los antibióticos. Alternativamente, los elementos genéticos se pueden introducir directamente en el cromosoma de una célula huésped sin el uso de un plásmido no replicante. Esto puede realizarse produciendo de manera sintética fragmentos de ADN de los elementos genéticos de acuerdo con la presente invención que también contienen secuencias de ADN homólogas del cromosoma huésped. De nuevo, si estos fragmentos de ADN sintéticos también contienen un marcador seleccionable, los elementos genéticos pueden insertarse en el cromosoma del huésped.

55 [0083] La invención se refiere además a un método para la preparación de un polipéptido que tiene actividad de HMFCa deshidrogenasa de la invención y/o un polipéptido que tiene capacidades de transporte de compuestos furánicos de la invención, el cual comprende cultivar una célula de acuerdo con la invención en condiciones propicias para la expresión del polipéptido y, opcionalmente, recuperar el polipéptido expresado, así como a un polipéptido obtenible mediante dicho método.

#### 60 Procesos de oxidación de compuestos furánicos

65 [0084] En un aspecto adicional, la invención se refiere a procesos para oxidar compuestos furánicos. En particular, la invención se refiere a un proceso en el que se oxidan precursores furánicos de FDCA. Un proceso de la invención puede comprender una única etapa de reacción de oxidación que da como resultado un producto (por ejemplo, la oxidación de HMFCa a FFA). Alternativamente, un proceso de la invención puede comprender más de una etapa de reacción de oxidación, donde cada etapa da como resultado un intermedio y el último

intermedio es el producto final. Los ejemplos de dicha serie de pasos, en los que el HMF se oxida a FDCA a través de pasos de oxidación secuenciales incluyen, por ejemplo: 1) El HMF se oxida primero a HMFCa, que en un segundo paso se oxida a FFA, que luego finalmente se oxida a FDCA, o alternativamente, como lo describen Dijkman et al. (2014, *Angew. Chem.* 53 (2014) 6515-8.) 2) El HMF se oxida primero a DFF, que en un segundo paso se oxida a FFA, que finalmente se oxida a FDCA. Por tanto, en un proceso preferido de la invención, uno o más precursores furánicos de FDCA se oxidan en una serie de pasos hasta llegar finalmente a FDCA.

[0085] En una realización, la invención se refiere a procesos que comprenden al menos la oxidación de HMFCa a FFA. Preferiblemente, el proceso es un proceso para oxidar HMFCa a FFA, donde el proceso comprende el paso de incubar una célula en presencia de HMFCa, donde la célula es una célula que expresa una HMFCa deshidrogenasa tal como se ha definido en el presente documento anteriormente, o una célula que expresa un polipéptido que tiene capacidades de transporte de compuestos furánicos y que además comprende actividades de HMFCa deshidrogenasa u oxidasa tal como se ha definido en el presente documento anteriormente. Preferiblemente, la célula se incuba en presencia de HMFCa en condiciones que conducen a la oxidación del HMFCa por la célula, como, p. ej., se especifica a continuación.

[0086] En otra realización, la invención se refiere a procesos para producir FDCA. Un proceso para producir FDCA comprende preferiblemente el paso de incubar una célula en un medio que comprende uno o más precursores furánicos de FDCA, en el que la célula es una célula que expresa una HMFCa deshidrogenasa tal como se ha definido en el presente documento anteriormente, o una célula que expresa un polipéptido que tiene capacidades de transporte de compuestos furánicos y que comprende además una actividad de HMFCa deshidrogenasa u oxidasa tal como se ha definido en el presente documento anteriormente. Preferiblemente, la célula se incuba en presencia de HMFCa en condiciones propicias para la oxidación a FDCA de los precursores furánicos de FDCA por parte de la célula, como, p. ej., se especifica a continuación.

[0087] Preferiblemente, en el proceso, al menos un precursor furánico de FDCA se selecciona del grupo que consta de HMF, DHF, HMFCa, FFA y DFF, de los cuales HMF es el preferido. Los precursores furánicos del FDCA se obtienen preferiblemente a partir de uno o más azúcares de hexosa, preferiblemente mediante deshidratación catalizada por ácido, p. ej., mediante calentamiento en presencia de ácido, de manera convencional. La tecnología para generar HMF a partir de fructosa está bien establecida y es robusta (ver p. ej., van Putten et al., 2013, *Chem. Rev.* 113, 1499-1597). También se puede utilizar materia prima rica en glucosa, pero la formación termoquímica de HMF procede de forma más eficaz a partir de la fructosa. Por lo tanto, se puede incluir un paso enzimático adicional para convertir la glucosa en fructosa, utilizando glucosa isomerasa. Este último proceso está bien establecido en la industria alimentaria, p. ej., para producir jarabe de maíz con alto contenido de fructosa (JMAF) a partir de almidón hidrolizado. La glucosa también se puede isomerizar químicamente a fructosa usando combinaciones de catalizadores y disolventes como se describe, p. ej., en van Putten et al. (2013, *supra*).

[0088] Los azúcares hexosa normalmente se obtendrán a partir de biomasa. El término "biomasa" se entiende como la fracción biodegradable de productos, desechos y residuos de origen biológico de la agricultura (incluyendo sustancias vegetales, tales como residuos de cultivos, y animales), la silvicultura (por ejemplo, recursos madereros) e industrias relacionadas, incluidas la pesca y la acuicultura, así como la fracción biodegradable de residuos industriales y municipales, tales como residuos sólidos urbanos o papel de desecho. En una realización preferida, la biomasa es biomasa vegetal, más preferiblemente una biomasa (fermentable) rica en hexosa/glucosa/azúcar, como p. ej., caña de azúcar, una biomasa que contiene almidón, por ejemplo, grano de trigo o paja de maíz, o incluso granos de cereales, como maíz, trigo, cebada o mezclas de los mismos. Se prefieren los cultivos agrícolas naturalmente ricos en fructanos (por ejemplo, raíces de topinambur o achicoria).

[0089] Los azúcares de hexosa se pueden obtener por hidrólisis de dicha biomasa. Los métodos para la hidrólisis de biomasa son conocidos en la técnica per se e incluyen el uso de, p. ej., vapor y/o carbohidrasas tales como glucoamilasas.

[0090] Otro tipo preferido de biomasa para usar en el proceso de la invención es la denominada materia prima lignocelulósica de "segunda generación", que se prefiere si se van a producir grandes volúmenes de FDCA de una manera más sostenible. Las materias primas lignocelulósicas pueden obtenerse de cultivos energéticos dedicados, p. ej., cultivados en terrenos marginales, por lo que no compite directamente con los cultivos alimentarios. O pueden obtenerse materias primas lignocelulósicas como subproductos, p. ej., se pueden considerar los desechos sólidos municipales, el papel de desecho, los residuos de madera (incluidos los desechos de aserraderos y fábricas de papel) y los residuos de cultivos. Ejemplos de residuos de cultivos incluyen el bagazo de caña de azúcar y también varios desechos de maíz y trigo. En el caso de los subproductos del maíz, tres residuos son fibra, mazorcas y rastrojo. Además, la biomasa forestal se puede utilizar como materia prima. Para convertir las materias primas de segunda generación en productos de fermentación de la invención, es necesario liberar la celulosa y la hemicelulosa como monosacáridos. Para ello, se aplican enfoques termoquímicos (normalmente denominados pretratamientos), enfoques enzimáticos o una combinación de las dos metodologías. Un pretratamiento puede servir para liberar completamente los azúcares o para hacer que los

compuestos poliméricos sean más accesibles al ataque enzimático posterior. Los diferentes tipos de pretratamiento incluyen agua caliente líquida, explosión de vapor, pretratamiento con ácido, pretratamiento con bases y pretratamientos con líquidos iónicos. Las cantidades relativas de los diversos compuestos dependerán tanto de la materia prima utilizada como del pretratamiento empleado. Para la liberación de azúcares monosacáridos de dicha materia prima lignocelulósica, se emplean carbohidrasas apropiadas, incluyendo p. ej., arabinasas, xilanasas, glucanasas, amilasas, celulasas, glucanasas y similares.

[0091] El proceso de la invención comprende además preferiblemente la etapa de recuperación del producto o productos de oxidación producidos en el proceso, tales como FDCA o HMFCa. Preferiblemente, el producto de oxidación se recupera del medio en el que se incubó la célula que lleva a cabo las etapas de oxidación. Los productos de oxidación tales como FDCA, HMFCa, etc. pueden recuperarse de la mezcla o medio de reacción, p. ej., precipitación (ácida o salina), posterior cristalización por enfriamiento y separación del producto de oxidación cristalizado, por ejemplo, FDCA cristalizado. Sin embargo, son adecuados otros métodos de recuperación, como p. ej., precipitación ácida o salina y extracción con disolvente, como se conoce en la técnica. La precipitación salina para la recuperación del FDCA puede, por ejemplo, realizarse usando cationes divalentes (metálicos), como p. ej.,  $Mg^{2+}$ .

[0092] Las reacciones de oxidación se llevan a cabo preferiblemente a la temperatura óptima para la célula y las enzimas oxidorreductasa contenidas en la célula. Por tanto, en el caso de células y enzimas termofílicas, la temperatura preferiblemente es de 45 °C o superior, p. ej., en el rango entre 45 y 122 °C, p. ej., superior a 50, 55, 60 o 65 °C. Sin embargo, en el caso de una célula mesófila que contiene enzimas de mesófilos, las reacciones de oxidación se llevan a cabo preferiblemente a una temperatura relativamente suave, p. ej., 10 - 80 °C, más preferiblemente 20 - 45 °C, lo más preferible aproximadamente de 25 - 40 °C.

[0093] Las reacciones de oxidación se llevan a cabo preferiblemente a un pH en el que el FDCA está en una forma neutra o en una forma completamente disociada, de modo que se pueda controlar la formación de sal. En vista de la presencia de dos fracciones ácidas en FDCA, hay dos intervalos de pH preferidos separados. El pH durante la reacción puede ser de 1 a 6, preferiblemente de 1 a 4, más preferiblemente de 1 a 3. Alternativamente, el pH durante la reacción puede ser de 5 a 9, preferiblemente de 5 a 8, más preferiblemente de 5 a 7. La persona experta comprenderá que los requisitos de la célula huésped también influirán en la selección de un valor de pH adecuado para el proceso. La selección de los valores de pH que son apropiados para una célula huésped particular está dentro del ámbito de la persona experta y puede obtenerse de libros de texto estándar. Para *Pseudomonas putida*, incluyendo, p. ej., las cepas de *Pseudomonas putida* S12 o KT2440, el rango de pH preferido es de 5 a 7.

[0094] El tiempo de la reacción puede ser de 6 a 150 horas, más preferiblemente de 6 a 18 horas. Preferiblemente, se suministra oxígeno a las células en el medio de reacción desde una fuente de oxígeno, tal como oxígeno molecular, p. ej., como oxígeno puro o en el aire, o agua, o una fuente diferente de oxígeno dependiendo de los requisitos de la enzima oxidante furánica. El aire puede usarse convenientemente como fuente de oxígeno molecular.

[0095] El reactor puede ser cualquier biorreactor adecuado (aireado). Puede operarse en lotes, en continuo o preferiblemente en lotes alimentados.

[0096] Los procesos de la invención para oxidar compuestos furánicos pueden aplicarse de manera ventajosa para la eliminación de compuestos furánicos de materias primas en las que los compuestos furánicos se consideran perjudiciales, tales como materias primas para fermentaciones para la producción de biocombustibles y productos bioquímicos. Más preferiblemente, los procesos para oxidar compuestos furánicos se aplican en la bioproducción de FDCA como un precursor monomérico para la producción de poliésteres (plásticos), donde el FDCA puede sustituir al PTA en el poliéster PET en cuyo caso se produce el polietilénfurandicarboxilato (PEF) de base biológica. El FDCA también puede usarse como sustrato para una gran variedad de compuestos valiosos, que incluyen p. ej., como sustrato para la producción de ácido succínico, 2,5-bis (aminometil)-tetrahydrofurano, 2,5-dihidroximetil-tetrahydrofurano, 2,5-dihidroximetilfurano y 2,5-furandicarbaldehído. El FDCA puede usarse en la producción de revestimientos, p. ej., en resinas alquídicas y revestimientos termoplásticos. También se puede utilizar como equivalente de xileno en biocombustibles y como disolvente. El FDCA puede esterificarse y los ésteres pueden usarse como plastificantes. El FDCA puede convertirse en su diol, que puede usarse en poliésteres y poliuretanos similares al PET. Además, el FDCA se puede convertir en su diamina, la diamina se puede usar como extensor de cadena y se puede convertir en diisocianato, que se puede usar en la producción de poliuretanos.

[0097] Por tanto, en un aspecto adicional, la invención se refiere a un proceso para producir un polímero a partir de uno o más monómeros de FDCA, donde el proceso comprende las etapas de: a) preparar un monómero FDCA en un proceso de oxidación de la invención tal como se ha descrito anteriormente; y b) producir un polímero a partir del monómero de FDCA obtenido en a). Preferiblemente, el polímero es polietilénfurandicarboxilato (PEF).

[0098] En otro aspecto, la invención se refiere al uso de una célula de la invención, para la biotransformación de uno o más precursores furánicos a FDCA, donde la célula es una célula que expresa una HMFCa deshidrogenasa tal como se ha definido en el presente documento anteriormente o una célula que expresa el polipéptido que tiene capacidades de transporte de compuestos furánicos y que comprende además una actividad de HMFCa deshidrogenasa u oxidasa tal como se ha definido en el presente documento anteriormente. Preferiblemente, al menos un precursor furánico del FDCA que se biotransforma en FDCA se selecciona del grupo que consta de HMF, DHF, HMFCa, FFA y DFF, de los cuales HMF es el preferido.

#### 10 Polipéptidos de HMFCa deshidrogenasa y ácidos nucleicos que codifican HMFCa deshidrogenasas

[0099] En un aspecto adicional, la invención se refiere a un polipéptido que tiene actividad de HMFCa deshidrogenasa. El polipéptido con actividad de HMFCa deshidrogenasa comprende o consta de una secuencia de aminoácidos que tiene al menos un 81.65, 81.7, 81.8, 81.85, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 95, 96, 97, 98, 99 o 100 % de identidad de secuencia con la secuencia de aminoácidos de la SEC ID N.º: 1 (*Aeribacillus pallidus*) pero por lo demás es tal como se ha definido en el presente documento anteriormente. Preferiblemente, el polipéptido es un polipéptido aislado.

[0100] La invención se refiere además a una molécula de ácidos nucleicos que comprende al menos una de las siguientes:

- a) una secuencia de nucleótidos que codifica un polipéptido que tiene actividad de HMFCa deshidrogenasa, cuyo polipéptido comprende o consta de una secuencia de aminoácidos que tiene al menos un 81.65, 81.7, 81.8, 81.85, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 95, 96, 97, 98, 99 o 100 % de identidad de secuencia con la secuencia de aminoácidos de la SEC ID N.º: 1;
- b) una secuencia de nucleótidos establecida en las SEC ID N.º: 12 o 13;
- c) una secuencia de nucleótidos cuya secuencia difiere de la secuencia de una secuencia de nucleótidos de b) debido a la degeneración del código genético; y
- d) una secuencia de nucleótidos que es el complemento inverso de una secuencia de nucleótidos tal como se ha definido en los puntos a) a c).

[0101] Otro aspecto de la invención se refiere a vectores, incluyendo los vectores de clonación y expresión, que comprenden una secuencia de nucleótidos tal como se ha definido en los puntos a) a d) anteriormente en esta sección, donde dichos vectores son por lo demás tal como se han descrito en el presente documento anteriormente.

[0102] En otro aspecto, la invención se refiere a una célula que comprende al menos uno de los siguientes: i) un polipéptido que tiene actividad de HMFCa deshidrogenasa tal como se ha definido anteriormente en esta sección, y ii) una molécula de ácidos nucleicos tal como se ha definido anteriormente en esta sección. Preferiblemente, la célula es una célula que comprende o se transforma con una secuencia de nucleótidos tal como se ha definido en los puntos a) a d) anteriormente en esta sección, o un vector que comprende dicha secuencia de nucleótidos. La célula es preferiblemente una célula aislada o una célula cultivada, la célula preferiblemente es por lo demás como se ha descrito en el presente documento anteriormente y preferiblemente la célula comprende una o más de las modificaciones genéticas descritas en el presente documento anteriormente. La célula se puede aplicar en cualquiera de los métodos, procesos y usos descritos anteriormente.

#### 50 Polipéptidos transportadores de compuestos furánicos y ácidos nucleicos que codifican dichos polipéptidos transportadores

[0103] En otro aspecto adicional, la divulgación se refiere a un polipéptido que tiene capacidades de transporte de compuestos furánicos. El polipéptido preferiblemente tiene al menos la capacidad de transportar HMFCa al interior de la célula. Preferiblemente, el polipéptido comprende o consta de una secuencia de aminoácidos que tiene al menos un 86.5, 87, 88, 89, 90, 91, 92, 93, 94, 95, 95, 96, 97, 98, 99 o 100 % de identidad de secuencia con la secuencia de aminoácidos de la SEC ID N.º: 17 (*Aeribacillus pallidus*) pero por lo demás es tal como se ha definido en el presente documento anteriormente. Preferiblemente, el polipéptido es un polipéptido aislado.

[0104] La divulgación se refiere además a una molécula de ácidos nucleicos que comprende al menos una de las siguientes:

- a) una secuencia de nucleótidos que codifica un polipéptido que tiene la capacidad de transportar al menos HMFCa al interior de la célula, donde el polipéptido comprende una secuencia de aminoácidos con al menos un 86.5, 87, 88, 89, 90, 91, 92, 93, 94, 95, 95, 96, 97, 98, 99 o 100 % de identidad con la secuencia de aminoácidos de la SEC ID N.º: 17;
- b) una secuencia de nucleótidos establecida en las SEC ID N.º: 18;

c) un fragmento de una secuencia de nucleótidos tal como se define en (a) o (b) que tiene 10, 15, 20, 30, 50 o 100 nucleótidos de longitud;

d) una secuencia de nucleótidos cuya secuencia difiere de la secuencia de una secuencia de nucleótidos de b) o c) debido a la degeneración del código genético; y

e) una secuencia de nucleótidos que es el complemento inverso de una secuencia de nucleótidos tal como se ha definido en los puntos a) a d).

[0105] Otro aspecto de la divulgación se refiere a vectores, incluyendo los vectores de clonación y expresión, que comprenden una secuencia de nucleótidos tal como se ha definido en los puntos a) a e) anteriormente en esta sección, donde dichos vectores son por lo demás tal como se han descrito en el presente documento anteriormente.

[0106] En otro aspecto, la divulgación se refiere a una célula que comprende al menos uno de los siguientes: i) un polipéptido que tiene capacidades de transporte de compuestos furánicos tal como se ha definido anteriormente en esta sección, y ii) una molécula de ácidos nucleicos tal como se ha definido anteriormente en esta sección. Preferiblemente, la célula es una célula que comprende o se transforma con una secuencia de nucleótidos tal como se ha definido en los puntos a) a e) anteriormente en esta sección, o un vector que comprende dicha secuencia de nucleótidos. La célula es preferiblemente una célula aislada o una célula cultivada, la célula preferiblemente es por lo demás como se ha descrito en el presente documento anteriormente y preferiblemente la célula comprende una o más de las modificaciones genéticas descritas en el presente documento anteriormente. La célula se puede aplicar en cualquiera de los métodos, procesos y usos descritos anteriormente.

[0107] En este documento y en sus reivindicaciones, el verbo "comprender" y sus conjugaciones se utilizan en su sentido no limitante para indicar que se incluyen los elementos que siguen a la palabra, pero no se excluyen los elementos que no se mencionan específicamente. Además, la referencia a un elemento por los artículos indefinidos "un" o "una" no excluye la posibilidad de que esté presente más de uno de los elementos, a menos que el contexto claramente requiera que haya uno y sólo uno de los elementos. Por tanto, el artículo indefinido "un" o "una" normalmente significa "al menos uno".

[0108] Los siguientes ejemplos se ofrecen solo con fines ilustrativos y no pretenden limitar el alcance de la presente invención de ninguna manera.

#### Descripción de las figuras

[0109]

Figura 1A: Biotransformación de HMF por *P. putida* CA2046 (*P. putida*; círculo abierto: HMF (5-hidroxiacetilfurfural); cuadrado abierto: HMFCFA (ácido 5-hidroxiacetilfuroico); rombo relleno: FDCA (ácido 2,5-furano dicarboxílico); círculo gris relleno: OD600.

Figura 1B: Biotransformación de HMF por *P. putida* CA2101; círculo abierto: HMF (5-hidroxiacetilfurfural); cuadrado abierto: HMFCFA (ácido 5-hidroxiacetilfuroico); rombo relleno: FDCA (ácido 2,5-furano dicarboxílico); círculo gris relleno: OD600.

Figura 2: Biotransformación de HMF por *P. putida* CA2111, coexpresando *YiaY* con *Aldh* y *HmfT1* de *C. basilensis* HMF14; círculo abierto: HMF (5-hidroxiacetilfurfural); cuadrado abierto: HMFCFA (ácido 5-hidroxiacetilfuroico); rombo relleno: FDCA (ácido 2,5-furano dicarboxílico); círculo gris relleno: OD600. Se muestran los promedios de cultivos duplicados.

Figura 3: Biotransformación de HMF por *P. putida* CA2112, coexpresando *YiaY* con *Aldh* y *HmfT1* de *C. basilensis* HMF14; círculo abierto: HMF (5-hidroxiacetilfurfural); cuadrado abierto: HMFCFA (ácido 5-hidroxiacetilfuroico); rombo relleno: FDCA (ácido 2,5-furano dicarboxílico); círculo gris relleno: OD600. Se muestran los promedios de cultivos duplicados.

Figura 4: Biotransformación de HMF por *P. putida* CA21780, coexpresando *YiaY* de *Bacillus kribbensis* DSM17871 y *Aldh* y *HmfT1* de *C. basilensis*. HMF-OH es dihidroxiacetilfurfural, también denominado "DHF" en el presente documento.

Figura 5: Biotransformación de HMF por *P. putida* CA21781, coexpresando *YiaY* de *Aneurinibacillus terranovensis* DSM18919 y *Aldh* y *HmfT1* de *C. basilensis*. HMF-OH es dihidroxiacetilfurfural, también denominado "DHF" en el presente documento.

Figura 6: Biotransformación de HMF por *P. putida* CA21783, coexpresando *YiaY* de *Brevibacillus panacihumi* W25 y *Aldh* y *HmfT1* de *C. basilensis*. HMF-OH es dihidroxiacetilfurfural, también denominado "DHF" en el presente documento.

#### **Ejemplos**

Metodología general

*Cepas y plásmidos*

[0110] *Pseudomonas putida* S12Δgcd o *P. putida* KT2440Δgcd (mutantes deficientes en glucosa deshidrogenasa de *P. putida* S12 (ATCC 700801), resp., *P. putida* KT2440 (DSM6125)) o *P. putida* S12 de tipo salvaje, se utilizaron como huésped para la expresión del gen *yiaY* de *Aeribacillus pallidus* cepa CA1828 (ver más adelante). La cepa TG90 de *Escherichia coli* se utilizó para fines generales de clonación.

[0111] Para la expresión episomal del gen de *A. pallidus* se usó el pBT<sup>mcs</sup> derivado de pBBR1MCS (Koopman et al., 2010a, *Biores Technol* 101: 6291-6196). En el pBT<sup>mcs</sup>, la expresión del gen diana es impulsada por el promotor constitutivo *tac*.

#### Medios y condiciones de cultivo

[0112] El medio de sales minerales mesófilas (MMM) contenía lo siguiente (por litro de agua desmineralizada): 15.52 g de K<sub>2</sub>HPO<sub>4</sub>, 6.52 g de NaH<sub>2</sub>PO<sub>4</sub>, 2.0 g de (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 0.1 g de MgCl<sub>2</sub>·6H<sub>2</sub>O, 10 mg de EDTA, 2 mg de ZnSO<sub>4</sub>·7H<sub>2</sub>O, 1 mg de CaCl<sub>2</sub>·2H<sub>2</sub>O, 5 mg de FeSO<sub>4</sub>·7H<sub>2</sub>O, 0.2 mg de Na<sub>2</sub>MoO<sub>4</sub>·2H<sub>2</sub>O, 0.2 mg de CuSO<sub>4</sub>·5H<sub>2</sub>O, 0.4 mg de CoCl<sub>2</sub>·6H<sub>2</sub>O y 1 mg de MnCl<sub>2</sub>·2H<sub>2</sub>O, suplementado con una fuente de carbono según se especifica.

[0113] El medio de sales minerales termófilas (TMM) contenía lo siguiente (por litro de agua desmineralizada): 10 g de Bis-Tris, FeSO<sub>4</sub>·7H<sub>2</sub>O 10 μM, tricina 4 mM, K<sub>2</sub>HPO<sub>4</sub> 1.32 mM, NH<sub>4</sub>Cl 9.53 mM, 0.2 g de extracto de levadura, 5 g de NaCl, 1.47 g de Na<sub>2</sub>SO<sub>4</sub>, 0.08 g de NaHCO<sub>3</sub>, 0.25 g de KCl, 1.87 g de MgCl<sub>2</sub>·6H<sub>2</sub>O, 0.41 g de CaCl<sub>2</sub>·2H<sub>2</sub>O, 0.008 g de SrCl<sub>2</sub>·6H<sub>2</sub>O, 0.008 g de H<sub>3</sub>BO<sub>3</sub>, 0.90 g de NaNO<sub>3</sub> y 1 ml de solución vitamínica (tiamina, 0.1 g/l; riboflavina, 0.1 g/l; ácido nicotínico, 0.5 g/l, ácido pantoténico, 0.1 g/l; piridoxamina- HCl, 0.5 g/l; piridoxal-HCl, 0.5 g/l; D-biotina, 0.1 g/l; ácido fólico, 0.1 g/l; ácido p-aminobenzoico, 0.1 g/l; cobalamina, 0.1 g/l). Las fuentes de carbono se suplementaron según lo especificado.

[0114] Como medio completo para la propagación de mesófilos se utilizó caldo Luria-Bertani (LB): 10 g/l de Bacto triptona (Difco), 5 g/l de extracto de levadura (Difco), 10 g/l de NaCl. Para el cultivo en placas, LB se solidificó con un 1.5 % (p/v) de agar (Difco). Para la selección de los transformantes de *E. coli*, *P. putida* S12 o *P. putida* KT2440 que portaban los plásmidos derivados de pBT<sup>mcs</sup>, se añadieron al medio 50 μg/ml de kanamicina (Km). Los antibióticos se adquirieron de Sigma-Aldrich. *P. putida* se cultivó a 30 °C; *E. coli* se cultivó a 37 °C.

[0115] Como medio completo para la propagación de termófilos se utilizó caldo TGP: 17 g/l de triptona, 3 g/l de peptona de soja, 5 g/l de NaCl, 2.5 g/l de K<sub>2</sub>HPO<sub>4</sub>, 4 g/l de glicerol y 4 g/l de Na-piruvato (pH 7). Para el cultivo en placas, se solidificó TGP con 1.5 % (p/v) de agar (Difco). *Aeribacillus pallidus* se cultivó a 60 °C.

#### Ensayos y métodos analíticos

##### Medición del peso seco de la célula (CDW):

[0116] El contenido de CDW de cultivos bacterianos se determinó midiendo la densidad óptica a 600 nm (OD<sub>600</sub>) utilizando un medidor de densidad celular Biowave (WPA Ltd) o un espectrofotómetro de microplacas universal μQuant MQX200 (Biotek), utilizando microplacas de fondo plano de 96 pocillos (Greiner). Una OD<sub>600</sub> de 1.0 corresponde a 0.56 g CDW/l (Biowave) o 1.4 g CDW/l (μQuant) para *P. putida*.

##### Análisis HPLC:

[0117] Los compuestos de furano (FDCA, HMF, HMF-alcohol, HMFA y FFA) se analizaron mediante RP-HPLC como describen Koopman et. (2010a, *Biores Technol* 101: 6291-6196).

#### Sustancias químicas

[0118] El 5-Hidroximetilfurfural (HMF) se adquirió en Eurolabs Ltd (Poynton, Reino Unido). Los estándares analíticos del FDCA y el ácido 5-hidroximetilfuroico (HMFA) se adquirieron de Immunosource B.V. (Halle-Zoersel, Bélgica), respectivamente, Matrix Scientific (Columbia SC, EE. UU.). Todos los demás productos químicos se adquirieron de Sigma-Aldrich Chemie B.V. (Zwijndrecht, Países Bajos).

#### Técnicas moleculares y genéticas:

[0119] Se aisló ADN genómico de *A. pallidus* CA1828 utilizando el kit de purificación de ADN gram positivo MasterPure™ (Epicenter). El ADN plasmídico se aisló con el kit de purificación de plásmidos JETSTAR Maxi (GENOMED, ITK diagnostics). Los fragmentos de ADN atrapados en agarosa se aislaron con el

[0120] DNA Clean & Concentrator™ (Zymo research). Las reacciones de PCR se realizaron con Phusion Flash PCR Master Mix (Thermo Scientific) de acuerdo con las instrucciones del fabricante. Los cebadores de oligonucleótidos (especificados en los ejemplos) fueron sintetizados por Sigma-Aldrich. El ADN plasmídico se

introdujo en células electrocompetentes utilizando un dispositivo de electroporación Gene Pulser (BioRad). Se llevaron a cabo otras técnicas estándar de biología molecular de acuerdo con Sambrook y Russell (2001, *supra*).

#### Ejemplo I: Aislamiento de cepas de *Aeribacillus pallidus* que metabolizan HMF

[0121] Se mezcló abono (15 g) con 15 ml de solución de NaCl al 0.9 % (p/v) y se incubó durante 40 min a 750 rpm y 80 °C. La suspensión de abono resultante se incubó en TMM suplementado con 0.65 g/l de HMF en matraces de agitación a 60 °C y 180 rpm durante 3 días. El cultivo se transfirió en intervalos regulares a TMM-HMF reciente y se colocó en placas sobre TMM-HMF sólido. Se volvieron a sembrar colonias individuales en placas de TMM-HMF y TGP, y se volvió a evaluar su capacidad para metabolizar HMF y también FDCA. Dos cepas que metabolizaron tanto HMF como FDCA (cepas CA1809 y CA1828) se identificaron como *Aeribacillus pallidus* mediante secuenciación de ADNr 16S y se seleccionaron para estudios adicionales.

#### Ejemplo II: Identificación de una nueva ruta catabólica de HMF catalizada por deshidrogenasa en HMF degradando aislados de *A. pallidus*

[0122] Los genomas de las cepas CA1809 y CA1828 de *A. pallidus* se secuenciaron mediante secuenciación PacBio, y se realizaron llamadas y anotaciones ORF automatizadas. En los genomas anotados, se identificaron homólogos de los genes *hmfABCDE* de *Cupriavidus basilensis* HMF14 que constituyen el clúster de degradación del ácido furoico (Koopman et al., 2010, Proc Nat Acad Sci USA 107: 4919-4924).

[0123] Considerando la capacidad de las cepas CA1809 y CA1828 para metabolizar FDCA además de HMF se sugirió firmemente que HMF se metaboliza a través de FDCA como en *C. basilensis* HMF14. Sin embargo, inesperadamente no se encontró ningún homólogo del clúster *hmfFGH* de *C. basilensis* HMF14 que constituya la ruta de degradación de HMF a ácido furoico a través de FDCA. Este resultado sugirió que existía una vía alternativa para la oxidación de HMF a FDCA, y posiblemente la descarboxilación posterior a ácido furoico, en los aislados de *A. pallidus*. La extracción de genomas en busca de clústers de genes que comprenden genes que codifican actividades oxidantes y descarboxilantes dio como resultado la identificación de un clúster de degradación putativo de HMF, que comprende genes que codifican una alcohol deshidrogenasa, una aldehído deshidrogenasa y dos descarboxilasas (Tablas 1 A y B). Juntos, estos genes codifican una vía putativa para la oxidación de HMF a FDCA, a través del ácido hidroximetilfuroico (HMFCA), como en *C. basilensis* HMF14 pero que implica una actividad de alcohol deshidrogenasa para la oxidación de HMFCA a ácido formilfuroico (FFA) en lugar de una actividad de oxidasa.

Tabla 1 A: Clúster de degradación putativo de HMF de *A. pallidus* CA1809

ID de locus	mejor BLAST hit	% identidad (x AA/y AA)	% de similitud (x AA/y AA)	función putativa	Locus correspondiente en <i>C. basilensis</i> HMF14	% identidad (x AA/y AA)	% de similitud (x AA/y AA)
03430	Transportador MFS [Geobacillus kaustophilus] ID de secuencia: ref WP_011229501.1	86 (384/446)	93 (418/446)	Transportador MFS	<i>hmfT1</i>	60 (270/450)	75 (341/450)
03431	4-hidroxibenzoato descarboxilasa [Geobacillus kaustophilus] ID de secuencia: ref WP_011229502.1	87 (405/466)	94 (442/466)	Subunidad FDCA descarboxilasa	<i>hmfF</i>	52 (240/462)	68 (315/462)
03432	alcohol deshidrogenasa [Geobacillus kaustophilus] ID de secuencia: ref WP_011229504.1	82 (320/392)	92 (363/392)	HMFCA deshidrogenasa	-		
03433	aldehído deshidrogenasa [Geobacillus kaustophilus] ID de secuencia: ref	88 (429/488)	95 (466/488)	HMF/FFA deshidrogenasa	<i>aldh</i>	37 (174/470)	55 (261/470)

ID de locus	mejor BLAST hit	% identidad (x AA/y AA)	% de similitud (x AA/y AA)	función putativa	Locus correspondiente en <i>C. basilensis</i> HMF14	% identidad (x AA/y AA)	% de similitud (x AA/y AA)
	WP_011229505.1						
03434	subunidad B de descarboxilasa de ácido fenólico [Geobacillus kaustophilus] ID de secuencia: ref   WP_011229508.1	91 (162/179)	96 (172/179)	Subunidad FDCA descarboxilasa	<i>hmfG</i>	54 (99/183)	74 (136/183)

Tabla 1 B: Clúster de degradación putativo de HMF de *A. pallidus* CA1828

ID de locus	mejor BLAST hit	% identidad (x AA/y AA)	% de similitud (x AA/y AA)	función putativa	Locus correspondiente en <i>C. basilensis</i> HMF14	% identidad (x AA/y AA)	% de similitud (x AA/y AA)
03227	Transportador MFS [Geobacillus kaustophilus] ID de secuencia: ref   WP_011229501.1	86 (384/446)	93 (418/446)	Transportador MFS	<i>hmfT1</i>	61 (232/383)	76 (293/383)
03228	4-hidrobenzoato descarboxilasa [Geobacillus kaustophilus] ID de secuencia: ref   WP_011229502.1	87 (405/466)	94 (442/466)	Subunidad FDCA descarboxilasa	<i>hmfF</i>	52 (240/462)	68 (315/462)
03229	alcohol deshidrogenasa [Geobacillus kaustophilus] ID de secuencia: ref   WP_011229504.1	82 (320/392)	92 (363/392)	HMFCa deshidrogenasa	-		
03230	aldehído deshidrogenasa [Geobacillus kaustophilus] ID de secuencia: ref   WP_011229505.1	88 (429/488)	95 (466/488)	HMF/FFA deshidrogenasa	<i>aldh</i>	37 (174/470)	55 (261/470)
03231	subunidad B de descarboxilasa de ácido fenólico [Geobacillus kaustophilus] ID de secuencia: ref   WP_011229508.1	91 (162/179)	96 (172/179)	Subunidad FDCA descarboxilasa	<i>hmfG</i>	54 (99/183)	74 (136/183)

**Ejemplo III: La expresión de *YiaY* de *A. pallidus* en *P. putida* S12 confiere la capacidad de oxidar HMF a FDCA**

5

[0124] El gen *yiaY* se clonó como un fragmento XbaI-SalI sintético de 1988 pb (SEC ID N.º: 15), incluyendo la región promotora PldhL1 de *B. coagulans* DSM1, en pBT<sup>mcs</sup> produciendo el plásmido pKW007. El plásmido pKW007 se introdujo en *P. putida* KT2440Δgcd (CA1877), produciendo *P. putida* CA2101. Se ensayó *P. putida* KT2440Δgcd que portaba pBT<sup>mcs</sup> (cepa CA2046) como control de vector vacío.

10

[0125] Las cepas CA2101 y CA2046 de *P. putida* se cultivaron en matraces de agitación de 100 ml que contenían 10 ml de MM + glicerol 80 mM y glucosa 2 mM suplementada con 50 mg/l de kanamicina. Las células

se recolectaron al final de la fase logarítmica ( $DO_{600} \approx 4$ ), se lavaron y se volvieron a suspender en MM suplementado con 19.4 g/l de  $K_2HPO_4$ , 8.15 g/l de  $NaH_2PO_4$ , glicerol 80 mM y 50 mg/l de kanamicina. Se incubaron alícuotas (10 ml) de suspensiones de células lavadas (a una  $DO_{600}$  de 1-2) con HMF en matraces Erlenmeyer de 100 ml y se extrajeron muestras a intervalos regulares para el análisis de FDCA. La figura 1A muestra que el HMF se oxida rápidamente a ácido hidroximetilfuroico (HMFCa) en el control de vector vacío, mientras que la formación del FDCA estuvo totalmente ausente. Cuando se expresó YiaY (Fig. 1B), el HMFCa acumulado se oxidó lentamente a FDCA, lo que demostró la funcionalidad de YiaY como deshidrogenasa oxidante de HMFCa.

#### 10 **Ejemplo IV: Oxidación optimizada de HMF a FDCA mediante la coexpresión de YiaY de *A. pallidus* y Aldh y HmfT1 de *C. basilensis* HMF14**

[0126] Se sintetizó el gen *yiaY* de *A. pallidus* CA1828 incluyendo el sitio de unión al ribosoma TAGGAAAGGAAGATTAACCC (SEC ID N.º: 21). El fragmento de *yiaY* (SEC ID N.º: 16) se digirió con KpnI y XbaI para reemplazar el gen *hmfH* en pBT'*hmfH*-adh (WO2012064195) lo que produjo el plásmido pKW010. El plásmido pKW010 se introdujo en *P. putida* S12Δgcd que albergaba pJNN*hmfT1*(t) (WO2012064195) lo que produjo *P. putida* CA2111, y en *P. putida* KT2440Δgcd (que también albergaba pJNN*hmfT1*(t)) lo que produjo *P. putida* CA2112. Por tanto, la alcohol deshidrogenasa oxidante HMFCa codificada por *yiaY* podría coexpresarse con la HMF deshidrogenasa y el transportador de HMFCa de *C. basilensis* HMF14 para eliminar los cuellos de botella de la oxidación de HMF a HMFCa y la absorción de HMFCa.

[0127] Se cultivaron CA2111 y CA2112 de *P. putida* en matraces de agitación de 100 ml que contenían 10 ml de MM + glicerol 80 mM y glucosa 2 mM suplementada con 50 mg/l de kanamicina, 30 mg/l de gentamicina y 100 μM de ácido salicílico. Las células se recolectaron al final de la fase logarítmica ( $DO_{600} \approx 4$ ), se lavaron y se resuspendieron en MM 50 mg/l de kanamicina, 30 mg/l de gentamicina y 10 μM de ácido salicílico. Se incubaron alícuotas (10 ml) de suspensiones de células lavadas (a una  $DO_{600}$  de 1-2) con HMF en matraces Erlenmeyer de 100 ml y se extrajeron muestras a intervalos regulares para el análisis de FDCA. Las figuras 2 y 3 muestran que el HMF se oxida rápidamente a HMFCa, que se oxida además a FDCA. Está claro que la coexpresión de YiaY con Aldh y HmfT1 acelera considerablemente la oxidación de HMF a FDCA.

#### 30 **Ejemplo V: Construcción de cepas optimizadas para la oxidación de HMF a FDCA a través de la coexpresión de alcohol deshidrogenasas de HMFCa mesófilas y Aldh y HmfT1 de *C. basilensis* HMF14**

[0128] Los homólogos de *yiaY* de *Bacillus kribbensis* DSM17871, *Brevibacillus thermoruber* 423, *Bacillus* sp. FJAT-14578 y *Bacillus* sp. L1 (2012) se sintetizaron incluyendo un sitio de unión a ribosomas que contenía el espaciador TAGGAAAGGAAGATTAACCC (SEC ID N.º: 21) así como sitios de reconocimiento para enzimas de restricción (KpnI, resp., NheI; compatible con XbaI) para la clonación (SEC ID N.º: 19, 36, 38 y 39).

[0129] Los homólogos de *yiaY* de *Aneurinibacillus terranovensis* DSM18919 y *Brevibacillus panacihumi* se sintetizaron W25 incluyendo un sitio de unión a ribosomas que contenía el espaciador GAATTCCACATGACAAGGGGAGACCGC (SEC ID N.º: 40) así como sitios de reconocimiento para enzimas de restricción (KpnI, resp., XbaI) para la clonación (SEC ID N.º: 35 y 37). La secuencia de nucleótidos codificante de la enzima de *B. kribbensis* (SEC ID N.º: 19), la enzima de *B. thermoruber* (SEC ID N.º: 36) y las dos enzimas de *Bacillus* sp. (SEC ID N.º: 38 y 39) se obtuvieron mediante la traducción inversa de las secuencias de aminoácidos ([http://www.bioinformatics.org/sms2/rev\\_trans.html](http://www.bioinformatics.org/sms2/rev_trans.html)) utilizando la tabla de uso de codones de *P. putida* de <http://www.kazusa.or.jp/codon/>. La secuencia de nucleótidos codificante de la enzima de *A. terranova* y *B. panacihumi* se obtuvo mediante la traducción inversa de las secuencias de aminoácidos utilizando la herramienta de optimización de secuencia de *E. coli* de GeneArt (<https://www.thermofisher.com/nl/en/home/life-science/cloning/gene-synthesis/geneart-gene-synthesis/geneoptimizer.html>).

[0130] Los fragmentos de homólogos de *yiaY* de *B. kribbensis*, *B. thermoruber*, *Bacillus* sp. FJAT-14578 y *Bacillus* sp. L1 (2012) fueron digeridos con KpnI y NheI (compatible con XbaI en pBT'*hmfH*-adh) para reemplazar el gen *hmfH* en pBT'*hmfH*-adh (WO2012064195) lo que produjo los plásmidos pKW2210, pKW2212, pKW2214 y pKW2215. Los fragmentos de homólogos de *yiaY* de *A. terranovensis* y *B. panacihumi* fueron digeridos con KpnI y XbaI para reemplazar el gen *hmfH* en pBT'*hmfH*-adh (WO2012064195) lo que produjo los plásmidos pKW2211 y pKW2213.

[0131] Los plásmidos pKW2210, pKW2211, pKW2212, pKW2213, pKW2214 y pKW2215 se introdujeron en *P. putida* KT2440Δgcd\_pJNN*hmfT1* (CA1965), lo que produjo respectivamente, *P. putida* CA21780, CA21781, CA21782, CA21783, CA21784 y CA21785 para la expresión de los homólogos de YiaY en un entorno huésped optimizado que incluía *aldh* y *hmfT1*. Para la evaluación del rendimiento, las cepas de *P. putida* CA21780, CA21781, CA21782, CA21783, CA21784 y CA21785 se cultivaron en matraces de agitación de 100 ml que contenían 10 ml de MM + glicerol 80 mM y glucosa 2 mM suplementada con 50 mg/l de kanamicina, 30 mg/l de gentamicina y 100 μM de ácido salicílico. Las células se recolectaron al final de la fase logarítmica ( $DO_{600} \approx 4$ ), se lavaron y se resuspendieron en MM 50 mg/l de kanamicina, 30 mg/l de gentamicina y 10 μM de ácido

salicílico. Se incubaron alícuotas (10 ml) de suspensiones de células lavadas (DO600 de 1-2) con HMF en matraces Erlenmeyer de 100 ml y se extrajeron muestras a intervalos regulares para el análisis de FDCA. Los resultados para *P. putida* CA21780, CA21781 y CA21783 se muestran en las figuras 4, 5 y 6, respectivamente. Las tres cepas transformadas produjeron FDCA a partir de HMF. Sin embargo, las diferentes cepas mostraron marcadas diferencias en la acumulación transitoria de HMFCa y la reducción parcial de HMF a dihidroximetil furano (HMF-OH o DHF). También se encontró que las cepas de *P. putida* CA21782, CA21784 y CA21785 producen FDCA a partir de HMF, lo que demuestra la funcionalidad de las seis alcohol deshidrogenasas como enzimas oxidantes de HMFCa.

10 **Ejemplo VI: Construcción de una cepa de *P. putida* que exprese el transportador HMFCa codificado por *proP* de *Aeribacillus pallidus***

[0132] El gen *proP* (SEC ID N.º: 18) se amplificó a partir de ADN genómico de *Aeribacillus pallidus* CA1828 por PCR usando los cebadores *proP*(f) (gccgaattcATGAAGAATATCGCTAATACG; SEC ID N.º: 22) y *proP*(r) (gccgctagcTTATTTGAGGTTTCCTTTTGTTC; SEC ID N.º: 23). El producto de PCR se introdujo como un fragmento de *EcoRI-NheI* de 1350 bp (SEC ID N.º: 20) en pJNNmcs(t) lo que produjo pJNNproP(t). Los plásmidos pBT<sup>+</sup>hmfH<sub>aldh</sub> y pJNNproP(t) se introdujeron sucesivamente en *P. putida* KT2440Δgcd (CA1877), lo que produjo *P. putida* CA21783. Se cultivó *P. putida* CA2111 y CA2112 en matraces de agitación de 100 ml que contenían 10 ml de MM + glicerol 80 mM y glucosa 2 mM suplementada con 50 mg/l de kanamicina, 30 mg/l de gentamicina y 100 μM de ácido salicílico. Las células se recolectaron al final de la fase logarítmica (DO600 ≈ 4), se lavaron y se resuspendieron en MM 50 mg/l de kanamicina, 30 mg/l de gentamicina y 10 μM de ácido salicílico. Se incubaron alícuotas (10 ml) de suspensiones de células lavadas (DO600 de 1-2) con HMF en matraces Erlenmeyer de 100 ml y se extrajeron muestras a intervalos regulares para el análisis de FDCA. Es evidente que la expresión del transportador de HMFCa codificado por *proP* acelera considerablemente la oxidación de HMF a FDCA en comparación con una cepa de control correspondiente que no expresa *proP*.

Tabla 2: Alineación de secuencias de aminoácidos YiaY

```

Adh_Bp      -----MESPFSFHLPTNVQFGVGSASRLGEMLLSMGVRRVFLVTDQGVRAAGLLDE
Adh_Bk      -----MDVEFSFHLPTLIEFGFGKASLLGERLLKLGVGNVFLVSDKGVASAGLLQK
Adh_Bt      --MSQTVQGTDFAFS FHLPTLIEFGYGRASRLGERLQHLGVTNVFVVDKGVAAAGLLNG
Adh_At      --MSPAVKAINFEFSFNLPTLIEFGYGKMEKFGQQLISIGVKRIFMVTDKGVESAGLLAA
YiaY        MIGNYAKKAIDFEFTFYLP LTLIEFGYGKASRMGEMLEQMGIKNVFLVTDKGVAAAGLLAG
Adh_Gk      MVGHYIQKEVEFEFSFHLPTS IQFGYGKASQLGNQLVDMGIKSAFLVTDKGVVATGLLAG
Adh_Bsp     -----MYPSEFHLPTKIHFYNTIKQLDH--LPFEIKRAFI VTDQGVLSGLVEN
Adh_BspL1   -----
Adh_Pt      -----
Adh_Dk      -----
Adh_Dt      -----MKTTVCFGANIVSSIDDRCDYNARHVLIVTDQGVKAGILEK
    
```

```

Adh_Bp      VIHSLEEKGLHFQIYADVEPDPSLETIQAGAAMFQQSFDCMVAIGGGSPIDTAKGIRVL
Adh_Bk      LEQSLQTSDIHFKTYLEVEPDPSLETIDLGAEFNSGKYDCIVAVGGGSAIDTAKGIRVV
Adh_Bt      LVGSLQSAGIAFDLYTEVEPDPLETIDRGA AVFRAKPYDCLVAVGGGSPIDAAKMRVV
Adh_At      LTDSLQAAAIQFDIYTDVESDPSLETIDRGVEVFQKPYDCIVAVGGGSPIDTAKGIRVV
YiaY        IVQSLESSNIRYVIYSDVEPDPSLETIDRGASVFKEQSFDCILAVGGGSPIDTAKGIRVV
Adh_Gk      IIQSLESSNIQYCVYADVEPDPSLETIDQGA AFKEQPFDCIVAIGGGSPIDTAKGIRVV
Adh_Bsp     VTNILKDHQISYVIYSEVEPDPSVETVDKAAQMFQREEADALIAIGGGSPIDTAKGVRVI
Adh_BspL1   -----PSVETVDKAAKAF AEACDLLIAVGGGSPIDTAKGVRVV
Adh_Pt      -----VEPDPLETVHKAA AFLGRTRPDCLVALGGGSSIDVAKGARVI
Adh_Dk      -----DPGLETIHRCASC FRENKCDLILAVGGGSPIDTAKGARVI
Adh_Dt      VEKVLSDAGIENVVFDVDPGLETIHRCASC FRENKCDLFLAIGGGSPIDTAKGARI I
                *.:**:. . : * :*:**** *.*** *.:
    
```

```

Adh_Bp      AANGGGIGQYAGVNRVPAASAIPLIAIPTTSGTGSEVTIFGVYSDWENHVKITVTS PHMA
Adh_Bk      AGNGGSIGDFAGVDKIGKAPQIPLIAVPTTSGTGSEVTIFGVYSDWVKNVKT VTSQYMA
Adh_Bt      TSCGGS IADYAGVNRVPMAPAVPLVAVPTTSGTGSEVTMFGVYSDWVHNHVKT VTS PHMA
Adh_At      AANGGNIGHYAGVNQIPVAPTIPLLA IPTTSGTGSEVTNFGVYSDWQNNVKT VTSQYMA
YiaY        VTNGGNIGDYAGVNRVAKKSEIPLVAVPTTSGTGSEVTIFGVYSDWENQVKVTVTSPYMA
Adh_Gk      ATNGGSIGDYAGVNR I KKKSEIPLIALPTTSGTGSEVTIFGVYSDWKNVKT VTS PYMA
Adh_Bsp     AGNGGSIRDYAGVNLIKQKSNIPLIAIPTTSGTGSEVTIFAVFSDWEENRKT VTS PFLA
Adh_BspL1   ASNGGSIRNYSGVNLVKEAPSVPLVA IPTTAGTGSEVTIFAVFSDDKENRKT VTS SHLS
Adh_Pt      YDNGGKISDYAGVNKVKVPSLPLMAVPTTAGTGSEVTVFAVLSDWENQIKITV TSEYLA
Adh_Dk      VENGHIRDYAGVNKVPRAPVTPLIAIPTTSGTGSEVTTFAVLSDWENRMKITISS PFLA
Adh_Dt      VDNGGHIRDYAGVNKVPRAPRTPLLA IPTTSGTGSEVTTFAVLSDWENRMKITISS PFLA
                ** * .:***: : ***:***:***** *. * ** :. *:*:* *.:
    
```

ES 2 817 001 T3

Adh\_Bp PSTALIDPALTLSPAKMTAATGIDALAHGIETFFSLRSSPASDALAIHAMKMIAPHLRR  
 Adh\_Bk PTIALVDPELTMRLPRKMTAASGIDALAHGIESYFSLRSTSASRALSLEAINIVGNHLRQ  
 Adh\_Bt PTIALVDPALTVSLPAKMTAASGIDALAHGIETFFSVRSRPASDALAMEAIAAVNAHLRR  
 Adh\_At PTIAWVDPALTMSLPAKMTAASGIDALAHGIETFFSLGSSPASDALAIEAHTVNRYSR  
 YiaY PEIALVDPELTMSLPQKMTAASGIDALAHGIETFFSLRSRPASDALAVEAMATVSAYLRR  
 Adh\_Gk PEIALVDPKLTMSLPKKITAASGIDALAHGIETFFSLRSQPISDVLAIEAMTTVNRYLRR  
 Adh\_Bsp PDISIVDPKMTMTAPPAITAASGFDAFAHGAETFVSRASQPASDVLAFSAMSTVSKYLRR  
 Adh\_BspL1 PDVSIIDPKLTLTAPPSITAAAGFDAFAHAAEAFVSRISQPPSDALALSAMKTVHTYLRR  
 Adh\_Pt PEAAFVDPLAMVSAPPGITAASGIDALSHAVEAYVSRAASPVSDNLALGAVELIGGHLRQ  
 Adh\_Dk PEVAVVDPLLMTAPPVTAASGIDALSHAIETYVSLKAQPPAEALALKAIELIGESLRT  
 Adh\_Dt PEVAVVDPILTLTAPPSVTAASGIDALSHAIETYVSLKAQPPAEALALKAIELIGESLRA  
 \* : : \*\* : \* : \*\*\* : \* : \* : \* : \* : \* : \* : \* : \* : \* : \*

Adh\_Bp AVR DGADMEARIGMSQGSVLAGMAFNNGFLGLAHAIGSALSNGHCHVPHGVAIGLLLPHVV  
 Adh\_Bk SVANGEDKEARCGMSHGSLLAGMAFNNGFLGLAHAIGSALSNGHCHVPHGVAIGLLLPHVV  
 Adh\_Bt AVHDGSDVEARIGMSHGSLLAGMAFTNGFLGLAHAIGSALSNGHCHVPHGVAIGLLLPHVV  
 Adh\_At AVHNGSDMEARIGMSHGSLLAGMAFNNGFLGLAHAIGSALSNGHCHVPHGVAIGLLLPHVV  
 YiaY AVEDGTDKEARIGMSQGSVLAGMAFNNGFLGLAHAIGSALSNGHCHVSHGVAIGLLLPHVV  
 Adh\_Gk AVEDGTNKEARIGMSYGSLLAGMAFNNGFLGLAHAIGSALSNGHCHVSHGVAIGLLLPHVV  
 Adh\_Bsp AVYNGEDVEARIKMAEASLLAGMAFNQSYLGLTHAIGSALSNGHCHVSHGVAIGLLLPHVV  
 Adh\_BspL1 AVYNGDDIEARMKMAEASLLAGMAFNQSYLGLAHAIGSAISVHAHVSHGVVIGLLLPHVV  
 Adh\_Pt AVANGGDLAARTGAALGSLLAGMAFNNAFLGLTHSIGAALSNGHVHVSHGVAVGLLLPYVM  
 Adh\_Dk AVADGSDKEARTRMSLGSLLAGMAFNNSLLGLTHSIGAALSNGHCHVSHGMAIGLLLPHVV  
 Adh\_Dt AVADGSDKEARTKMSLGSLLAGMAFNNSLLGLTHSIGAALSNGHCHVSHGMAVGLLLPYVM  
 : \* : \* : \*\* : . \* : \* \* \* \* \* : . \* \* \* \* \* : \* \* \* \* \* : \* \* \* \* \* : \*

Adh\_Bp AFNTPVRPEKAELIADVLSV--QKET----GTAAELVQQLVQDIGLPQRLQEVGVPEAK  
 Adh\_Bk EFNSSECPDQAAEIAKILGVK--AEDERQLAEQASHAVGDLVKDIGLPTRLRDMNVPEEK  
 Adh\_Bt AFNAPARPDKAAQLARLLGVE--ANPREERGEETSAAVARMVADIGLPTRLRDVGVPEEK  
 Adh\_At EFNATVRPDKAAKIAGLMGMK--GEHSEELALQASPAVARLVEDIGLPTRLREVDVTEKK  
 YiaY EFNARVRPEKAAKIAELLGVK--GDREEVLAEQAAPAVASLVKEIGLPTRLRDVDVSEEK  
 Adh\_Gk EFNSVQPEKAAKIAELLGRK--GNQNT-IVQQAALAVASLVKEIGLPTRLRDVDVPEEK  
 Adh\_Bsp RYNSISRMDKHIEMAGAFREIDRSLSDWEIIDQLIEDVSRLRDDIGLPQRLQVGVKEDQ  
 Adh\_BspL1 EYNLVAKIDKYAEAGKYIEQSSHGLSNYEAALFSETVTQLRNDIGLPKQLREVNVKEAQ  
 Adh\_Pt EYNLMKPKDFARLARAMGEVTEGKSLYRAASLAPRAVKAMVKSIGLPVRLKEIGVPEGA  
 Adh\_Dk EFNAMARMEKFSKIAVALGEDVKGLSLREAALRSVKAVRELVEDISLPRRLGDVGVGTGDM  
 Adh\_Dt EFNAMARLEKYGKIAIALGEDVKGLSLREAALRSVKAVRELVEDISLPRRLGEVGVGTGDM  
 : \* : : . : \* : . \* . \* \* : \* : : \*

```

Adh_Bp      LVDI AKDSFKSGMMKWNPRLPTEQEVL ELLQKAF
Adh_Bk      LADI ARDSFQSGMMKFNPRRASESEVLELLHRVY
Adh_Bt      LPAIAKDAFKSGMMTCNPRQPTEQEVR ELLRRAF
Adh_At      LFEIAKDSFKSGMMKFNPRQPSESEVLQ LLLKEIF
YiaY       LPDI ARDAFKSGMMKFNPRQPSLSEVLT LLQQIY
Adh_Gk      LPDI AKDSFKSGMMRFNPRQPSEAEVMT LLQQIY
Adh_Bsp     LKMIAADSVKSGMWKFNPRQASEEE ILELLKELY
Adh_BspL1   LEAISKDSIKSGMWQFNPRRASEQDVYQ MLREML
Adh_Pt      LAAIAETALKHGMKFNPRVPSREDILDI VKKAY
Adh_Dk      IEGMAKDAMGHGMLKFNPRAVTEKDI IAILRKAL
Adh_Dt      IEGMAKDAMGHGMLKFNPRVVTEKDI MAILQKAL
           :  ::  :.  **  ***  :  ::  :::
    
```

5 [0133] Adh\_Bp = SEC ID N.º: 6 (*Brevibacillus panacihumi*), Adh\_Bk = SEC ID N.º: 2 (*Bacilo kribbensis*); Adh\_Bt = SEC ID N.º: 5 (*Brevibacillus thermoruber*); Adh\_At = SEC ID N.º: 4 (*Aneurinibacillus terranovensis*); YiaY = SEC ID N.º: 1 (*Aeribacillus pallidus*); Adh\_Gk = SEC ID N.º: 3 (*Geobacillus kaustophilus*); Adh\_Bsp = SEC ID N.º: 7 (*Bacillus sp. FJAT-14578*); Adh\_BspL1 = SEC ID N.º: 10 (*Bacillus sp. L1 (2012)*); Adh\_Pt = SEC ID N.º: 11 (*Pelotomaculum thermopropionicum*); Adh\_Dk = SEC ID N.º: 8 (*Desulfotomaculum kuznetsovii*); y Adh\_Dt = SEC ID N.º: 9 (*Desulfurispora thermophila*). Los símbolos debajo de la alineación indican: \* = posiciones invariantes; : = posiciones fuertemente conservadas; . = posiciones menos conservadas; ningún símbolo indicaba posiciones no conservadas.

10 Tabla 3: Alineación de secuencia de aminoácidos (Clustal Omega) del transportador MFS de *A. Pallidus* (transportador HMFA) con los 10 mejores resultados de BLAST

15 Transportador de Aeribacillus

```

MKNIANTSTERPVNDASVKNRQMRATIASLIGWSLDLYDLFLLLVATTIGNLFPPASN
gi|499548718|ref|WP_011229501.1|:1-445      MDNITKTNIERPVE-
VSIKNSQMVRATIASLIGWALDLYDLFLLLVATTIGNLFPPASN
gi|651977233|ref|WP_026691821.1|:7-435      -----
NNRQLVSATMASLLGWSFDLYDLFILLVYVPTIGSLFFPSSN
gi|654945126|ref|WP_028395291.1|:1-445      MSNVVAT--
HSKQESVTVSKREVR SAMVASLLGWSFDLYDLFLLLVVAPTISVLFPPPTN
gi|737333963|ref|WP_035316274.1|:3-438      -----
SSNQPPKVEISRRQMVNASIASLLGWALDLDLFDLVLLYVAPVIGKLFPPTEL
gi|558617199|gb|EST53422.1|:11-446         -----
SSNQPPKVEISRRQMVNASIASLLGWALDLDLFDLVLLYVAPVIGKLFPPTEL
gi|737314460|ref|WP_035297308.1|:18-449     -----
RPPAAVGRKQMITAVLASLLGWSLDLYDLFILLVYVTPVLGKLFPPADN
gi|656061131|ref|WP_029098927.1|:18-449     -----
RPPAAVGRKQMITAVLASLLGWSLDLYDLFILLVYVTPVLGKLFPPADN
gi|503166469|ref|WP_013401130.1|:2-445
SANMETPVQQASALAAAISRKQMI IAVMASLLGWSLDLYDLFILLYVAPELGKLFPPPTDK
gi|505187461|ref|WP_015374563.1|:8-445      -----
NVQQTSSLTVSISKQMITAVTASLLGWSLDLYDLFILLYVAPELGKLFPPADK
gi|612120256|gb|EzP78263.1|:2-445
SVNTETTQQASPLTVSISRKQMI IAVMSSLLGWSLDLYDLFILLYVAPELGKLFPPPTDK
           .  ::  *
***:***:***:***:***:  ..  ****:
    
```

# ES 2 817 001 T3

## Transportador de Aeribacillus

QTLSLAAVYASFAVTLLMRPLGSAIFGIYADKNGRKKAMTVAIIGAGLCTAAFGLLPTIH  
gi|499548718|ref|WP\_011229501.1|:1-445  
QTLSLAAVYASFAVTLLMRPLGSAIFGVYADKNGRKKAMTVAIIGAGLSTTAFGLLPTIH  
gi|651977233|ref|WP\_026691821.1|:7-435  
PTLSLAAVYASFAVTLLMRPLGSAIFGSYADKNGRKKAMTVAVIGVGVSTAVFGLLPTVP  
gi|654945126|ref|WP\_028395291.1|:1-445  
PTLSLAAVYASFAVTLLMRPLGSAIFGSYADKNGRKKAMIVSVVGVGVSTAAFGLLPTVP  
gi|737333963|ref|WP\_035316274.1|:3-438  
PTLSLAAVYASFAVTLLMRPIGSALFGSYADRKGRKKAMIVAVIGVGVATALFGALPTVH  
gi|558617199|gb|EST53422.1|:11-446  
PTLSLAAVYASFAVTLLMRPIGSALFGSYADRKGRKKAMIVAVIGVGVATALFGALPTVH  
gi|737314460|ref|WP\_035297308.1|:18-449  
PTLSLAAVYASFAVTLLRPFSGALFGSYADRNGRKKRAMVVAVSGVGISTALFGVLPTVA  
gi|656061131|ref|WP\_029098927.1|:18-449  
PTLSLAAVYASFAVTLLRPFSGALFGSYADRNGRKKRAMVVAVSGVGISTALFGVLPTVA  
gi|503166469|ref|WP\_013401130.1|:2-445  
PTLSLAAVYASFAVTLMRPLGSAIFGAYADRNGRKKRAMVVAVSGVGISTALFGALPTVA  
gi|505187461|ref|WP\_015374563.1|:8-445  
PTLSLAAVYASFAVTLMRPLGSAIFGSYADRNGRKKRAMVVAVSGVGISTALFGALPTVE  
gi|612120256|gb|EYP78263.1|:2-445  
PTLSLAAVYASFAVTLMRPLGSAIFGTYADRNGRKKRAMVVAVSGVGISTALFGALPTVA  
\*\*\*\*\*:\*\*\*:\*\*

\*\*\*:\*\*:\*\* \*:\* \*:\*:\* \*\* \*\*:

## Transportador de Aeribacillus

QVGVVAAIAFLILRLVQGVFVGGVVASTHTIGTESASPKYRFGMSGLIGGGGAGLGALFA  
gi|499548718|ref|WP\_011229501.1|:1-445  
QVGVVAAIAFLILRLVQGVFVGGVVASTHTIGTESASPKYRGLMSGLIGGGGAGLGALFA  
gi|651977233|ref|WP\_026691821.1|:7-435  
QIGVFATIIFLVRLCQGIQGVVASSHTIGTESAPPKLRGLMSGLIGGGGAGLGALFA  
gi|654945126|ref|WP\_028395291.1|:1-445  
QIGFMASIIFLVRLVQGVFVGGVVASTHTIGTESAPPKWRGLMSGLIGGGGAGLGALFA  
gi|737333963|ref|WP\_035316274.1|:3-438  
QIGVGASIIFLILRLVQGVFVGGVVASTHTIGTESVPPKWRGFMMSGVGGGAGLGALLA  
gi|558617199|gb|EST53422.1|:11-446  
QIGVGASIIFLILRLVQGVFVGGVVASTHTIGTESVPPKWRGFMMSGVGGGAGLGALLA  
gi|737314460|ref|WP\_035297308.1|:18-449  
HIGAAATILFIILRLIQGVFVGGVVASTHTIGTESVPEKWRGLMSGLVGGGAGLGALLA  
gi|656061131|ref|WP\_029098927.1|:18-449  
HIGAAATILFIILRLIQGVFVGGVVASTHTIGTESVPEKWRGLMSGLVGGGAGLGALLA  
gi|503166469|ref|WP\_013401130.1|:2-445  
QIGAAAAIIFILRLVQGVFVGGVVASTHTIGTESVPEKWRGLMSGLVGGGAALGALLA  
gi|505187461|ref|WP\_015374563.1|:8-445  
QIGAAAAIIFILRLIQGVFVGGVVASTHTIGTESVPEKWRGLMSGLVGGGAALGALLA  
gi|612120256|gb|EYP78263.1|:2-445  
QIGAAAAIIFIVRLIQGVFVGGVVASTHTIGTESVPEKWRGLMSGLVGGGAALGALLA  
:\* \*:\* \*:\*:\*\*

\*\*\*:\*\*:\*\* \*:\* \*:\*:\* \*\* \*\*:

Transportador de Aeribacillus

SISYSVVTAIFPGEAFDVGWRVWFPTGIIGSLFGLFIFRSLLEESPLWKQLKEENSKGEV  
gi|499548718|ref|WP\_011229501.1|:1-445  
SIAYSIVSAIFPGDAFDTLGWRIMFFPTGIIGALFGLFIFRSLDESPLWKQLKEKQSKDKM  
gi|651977233|ref|WP\_026691821.1|:7-435  
SIAFTVVSSFFPGEAFSEWGWVWFPTGILGAIAGLVFVRTLDESPLWKGLQEEKKGKAV  
gi|654945126|ref|WP\_028395291.1|:1-445  
SIAFAIISALFPGEAFNEWGWVWFPTGILGAGAGLIVFRSLNESPLWAQLHEEKKTNE  
gi|737333963|ref|WP\_035316274.1|:3-438  
SIVYFIVSEAFPGEAFDAWGWVWFPTGILSAVLGVFVFKSLEESPLWLQAQQKKE---A  
gi|558617199|gb|EST53422.1|:11-446  
SIVYFIVSEAFPGEAFDAWGWVWFPTGILSAVLGVFVFKSLEESPLWLQAQQKKE---A  
gi|737314460|ref|WP\_035297308.1|:18-449  
SIVYFVLSLFPGEAFSEWGWVWFPTGILCSVLGLFVFRMLEESPLWVQHNEQA---A  
gi|656061131|ref|WP\_029098927.1|:18-449  
SIVYFVLSLFPGEAFSEWGWVWFPTGILCSVLGLFVFRMLEESPLWVQHNEQA---A  
gi|503166469|ref|WP\_013401130.1|:2-445  
SIVYFVLSVFSGPEFSEWGWVWFPTGILSSVLGLFVFKKLEESPLWVQHKKKQE---T  
gi|505187461|ref|WP\_015374563.1|:8-445  
SIVYFVLSVFPGEAFSEWGWVWFPTGILSSVLGLFVFKKLEESPLWVQHKKVQE---T  
  
gi|612120256|gb|EZP78263.1|:2-445  
SIVYFVLSNIFSGSEFSEWGWVWFPTGILCSVLGLFVFKKLEESPLWVQHKKKQE---M  
  
\*\* : ::: \* \* \* . \*\*\*.:\*\*:::

Transportador de Aeribacillus

SEFQKAPLKTFFTKRYKVLVNLMIIVIGGGSGYLLTSGFIPTFLKVVNKVASVSSGVL  
gi|499548718|ref|WP\_011229501.1|:1-445 -  
VEQQKSPFKMFLTKRYKVLVNLMIIVIGGGSGYLLTAGFIPTFLKVVNKVPAAVSSGVL  
gi|651977233|ref|WP\_026691821.1|:7-435  
SHTIEQKPVKTLFTTYSKVLVNLMIIVIGGGTGYLLTAGFIPTFLTIINDVSPGTSKSGIL  
gi|654945126|ref|WP\_028395291.1|:1-445  
EDAVPQSPKMLFKQYPGVLLVNVMIIVMGGGSAYLLTSGFVPTFLKVVNEAPPNVISGVL  
gi|737333963|ref|WP\_035316274.1|:3-438  
AKKPEGSPVKMIFPTQYRNVLLVNLMLVTTGGTAYLLTSGYLPFTFLNVINKVSSGTASLIL  
gi|558617199|gb|EST53422.1|:11-446  
AKKPEGSPVKMIFPTQYRNVLLVNLMLVTTGGTAYLLTSGYLPFTFLNVINKVSSGTASLIL  
gi|737314460|ref|WP\_035297308.1|:18-449  
KPAGQQSPVKMVFVKYLPVLLVNLMIIVIGGGSAAYLLTSGYLPFTFLNVINHVPQTSMIL  
gi|656061131|ref|WP\_029098927.1|:18-449  
KPAGQQSPVKMVFVKYLPVLLVNLMIIVIGGGSAAYLLTSGYLPFTFLNVINHVPQTSMIL  
gi|503166469|ref|WP\_013401130.1|:2-445  
KPEYQQSPVKMVFVKYLVLLVNLMIIVIGGGSAAYLLTCGYLPFTFLKVINNIPTVSSIIL  
gi|505187461|ref|WP\_015374563.1|:8-445  
KPEYQQSPVKIVFTKYLINLMIIVIGGGSAAYLLTCGYLPFTFLKVINNIPTVSSMIL  
gi|612120256|gb|EZP78263.1|:2-445  
KPEYQQSPVKMVFVKYLVLLINLMIIVIGGGSAAYLLTCGYLPFTFLKVINNIPTVSSMIL  
  
\* . \* \*\*.:\*\*:::

Transportador de Aeribacillus

IATSIMTIVA AVL VGH LSEV IGRKKT FLL I GILC I VGLP YF YLS LANSTTTTGIYLNALG  
 gi|499548718|ref|WP\_011229501.1|:1-445  
 IATSITLILAAI VVGH LSEL IGRKKT FMI I GILC V FGLP YF YLS LAHSTTTT SIYLN AIG  
 gi|651977233|ref|WP\_026691821.1|:7-435  
 IASSVVTI ISALL VGH LSEI IGRKKT FLA IGVVNI IGLPFFYLS LADAATTPSIYFYTMC  
 gi|654945126|ref|WP\_028395291.1|:1-445  
 IASSIVTI ISALL VGH LSEL IGRKKT V FLL VGLV LNI IGLP YF YLALGDSVTTLSIYLN TMG  
 gi|737333963|ref|WP\_035316274.1|:3-438  
 MGASVSAI ISAVL FGYLSDV IGRKKT FLL I GFINL ILLV LFIQLG SATSIPMITFYALA  
 gi|558617199|gb|EST53422.1|:11-446  
 MGASVSAI ISAVL FGYLSDV IGRKKT FLL I GFINL ILLV LFIQLG SATSIPMITFYALA  
 gi|737314460|ref|WP\_035297308.1|:18-449  
 AASSIAAI ISVALGHLSTV IGRKKT FVLLG I LNL MALP YLYTE LAAAQDLSRIALYAMG  
 gi|656061131|ref|WP\_029098927.1|:18-449  
 AASSISAI ISAVL VGH LSTI IGRKKT FVLLG I LNL MALP YLYTE LAAAQDLSRIALYAMG  
 gi|503166469|ref|WP\_013401130.1|:2-445  
 MVSSISAMVA AVL VGH LSTI IGRKKT FILLG IVNL FALP YLYTE LADAQDLTMITLYAMG  
 gi|505187461|ref|WP\_015374563.1|:8-445  
 IVSSISAMIAA IALGHLSTI IGRKKT FILLG IVNL IALP YLYTE LADAQDMTSITLYAMG

gi|612120256|gb|E2P78263.1|:2-445  
 MVSSISAMIASIVLGH LSTI IGRKKT FILLG TVNL IALP YLYTE LAAAQDLTLIILYAMG  
 \*: ::::: .\*:\*\* :\*\*\*\*\*.\*:  
 :\* : .. \*\* :: \* . : \* : ::

Transportador de Aeribacillus

LIFLGNAA YAPVLI FLNERFP TSIRSTGTGLSWNMGF AIGGMMPTFVN LASGTVEHIPYT  
 gi|499548718|ref|WP\_011229501.1|:1-445  
 LVFLGNASYAPVLI FLNERFPTEVRSTGTGLSWNVGF AIGGMMPTFVN LASGTVEHIPYT  
 gi|651977233|ref|WP\_026691821.1|:7-435  
 VVFLGNAA YAPVLI FLNERFP TSIRSTGTGISWNMGFAVGGMMPTFVTLASGSVKNI PHT  
 gi|654945126|ref|WP\_028395291.1|:1-445  
 LAFLGNAA YAPVLI FLNERFP TVIRSTGTGLSWNMGF AIGGMMPTFVTLASGKVENIPTT  
 gi|737333963|ref|WP\_035316274.1|:3-438  
 LAFLGNAA YAPILIFLNERFP TSIRSSGTGLSWNMGF AIGGMMPTFVTLASGT TENIPYS  
 gi|558617199|gb|EST53422.1|:11-446  
 LAFLGNAA YAPILIFLNERFP TSIRSSGTGLSWNMGF AIGGMMPTFVTLASGT TENIPYS  
 gi|737314460|ref|WP\_035297308.1|:18-449  
 LAFLGNASYAPVLI FLNERFP TAIRSTGTGLSWNMGF AIGGMMPTFVTMASGQTSEIPFF  
 gi|656061131|ref|WP\_029098927.1|:18-449  
 LAFLGNASYAPVLI FLNERFP TAIRSTGTGLSWNMGF AIGGMMPTFVTMASGQTSEIPFY  
 gi|503166469|ref|WP\_013401130.1|:2-445  
 LAFLGNASYAPVLI FLNERFP TSIRSTGTGLSWNMGF AIGGMMPTFVTMASRQTS DIPSS  
 gi|505187461|ref|WP\_015374563.1|:8-445  
 LAFLGNASYAPVLI FLNERFP TIRSTGTGLSWNMGF AIGGMMPTFVTMASSQTS DIPLS  
 gi|612120256|gb|E2P78263.1|:2-445  
 LAFLGNASYAPVLI FLNERFP TAIRSTGTGLSWNMGF AIGGMMPTFVTMTSSQTS DIPLS

: \*\*\*\*.\*\*\*:\*\*\*\*\*  
 \*\*:\*\*\*:\*\*\*:\*\*\*:\*\*\*\*\*.:\* ...\*\*

# ES 2 817 001 T3

Transportador de Aeribacillus gi 499548718 ref WP_011229501.1 :1-445 gi 651977233 ref WP_026691821.1 :7-435 gi 654945126 ref WP_028395291.1 :1-445 gi 737333963 ref WP_035316274.1 :3-438 gi 558617199 gb EST53422.1 :11-446 gi 737314460 ref WP_035297308.1 :18-449 gi 656061131 ref WP_029098927.1 :18-449 gi 503166469 ref WP_013401130.1 :2-445 gi 505187461 ref WP_015374563.1 :8-445 gi 612120256 gb EZP78263.1 :2-445	LMYFTIGIYLVYILGSLIIPETKGNLK LMYFTIVYIYLVYILGSLIIPETKGNLK LMYFFIGIFLLYLIGSAVIKETKGNLN LMYFAIGIFLVYIIGSIIVPETKGNLK LMGFSIAVFVVYVIGSLVIPETKGNFE LMGFSIAVFVVYVIGSLVIPETKGNFE LAYFSIGLFLLYLVGSLIIPETKGNFQ LAYFSIGLFLLYLVGSLIIPETKGNFQ LAYFFIALFLLYLLGSLIIPETKGNFK LTYFSIALFLLYLLGSLIIPETKGNFK LAYFSIALFLLYLLGSLIIPETKGNFK * * * ::::*:*:* :: *****:
--	---

5 Los símbolos por debajo de la alineación indican: \* posiciones invariables; : = posiciones fuertemente conservadas; . = posiciones menos conservadas; la ausencia de símbolo indica posiciones no conservadas.

## LISTADO DE SECUENCIAS

10 [0134]

<110> Purac Biochem B.V.

<120> Producción de FDCA catalizada por deshidrogenasa

<130> P6054479PCT

<150> EP15155401

15 <151> 2015-02-17

<160> 40

<170> versión de PatentIn 3.3

<210> 1

<211> 392

20 <212> PRT

<213> Aeribacillus pallidus

<400> 1

ES 2 817 001 T3

Met Ile Gly Asn Tyr Ala Lys Lys Ala Ile Asp Phe Glu Phe Thr Phe  
1 5 10 15

Tyr Leu Pro Thr Leu Ile Glu Phe Gly Tyr Gly Lys Ala Ser Arg Met  
20 25 30

Gly Glu Met Leu Glu Gln Met Gly Ile Lys Asn Val Phe Leu Val Thr  
35 40 45

Asp Lys Gly Val Glu Ala Ala Gly Leu Leu Ala Gly Ile Val Gln Ser  
50 55 60

Leu Glu Ser Ser Asn Ile Arg Tyr Val Ile Tyr Ser Asp Val Glu Pro  
65 70 75 80

Asp Pro Ser Leu Glu Thr Ile Asp Arg Gly Ala Ser Val Phe Lys Glu  
85 90 95

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

Thr Ala Lys Gly Ile Arg Val Val Val Thr Asn Gly Gly Asn Ile Gly  
115 120 125

Asp Tyr Ala Gly Val Asn Arg Val Ala Lys Lys Ser Glu Ile Pro Leu  
130 135 140

Val Ala Val Pro Thr Thr Ser Gly Thr Gly Ser Glu Val Thr Ile Phe  
145 150 155 160

Gly Val Tyr Ser Asp Trp Glu Asn Gln Val Lys Val Thr Val Thr Ser



# ES 2 817 001 T3

<210> 2  
<211> 383  
<212> PRT  
<213> *Bacillus kribbensis*

5 <400> 2

ES 2 817 001 T3

Met Asp Val Glu Phe Ser Phe His Leu Pro Thr Leu Ile Glu Phe Gly  
 1 5 10 15

Phe Gly Lys Ala Ser Leu Leu Gly Glu Arg Leu Leu Lys Leu Gly Val  
 20 25 30

Gly Asn Val Phe Leu Val Ser Asp Lys Gly Val Ala Ser Ala Gly Leu  
 35 40 45

Leu Gln Lys Leu Glu Gln Ser Leu Gln Thr Ser Asp Ile His Phe Lys  
 50 55 60

Thr Tyr Leu Glu Val Glu Pro Asp Pro Ser Leu Glu Thr Ile Asp Leu  
 65 70 75 80

Gly Ala Glu Ala Phe Asn Ser Gly Lys Tyr Asp Cys Ile Val Ala Val  
 85 90 95

Gly Gly Gly Ser Ala Ile Asp Thr Ala Lys Gly Ile Arg Val Val Ala  
 100 105 110

Gly Asn Gly Gly Ser Ile Gly Asp Phe Ala Gly Val Asp Lys Ile Gly  
 115 120 125

Lys Ala Pro Gln Ile Pro Leu Ile Ala Val Pro Thr Thr Ser Gly Thr  
 130 135 140

Gly Ser Glu Val Thr Ile Phe Gly Val Tyr Ser Asp Trp Val Lys Asn  
 145 150 155 160

Val Lys Val Thr Val Thr Ser Gln Tyr Met Ala Pro Thr Ile Ala Leu  
 165 170 175

Val Asp Pro Glu Leu Thr Met Arg Leu Pro Arg Lys Met Thr Ala Ala  
 180 185 190

Ser Gly Ile Asp Ala Leu Ala His Gly Ile Glu Ser Tyr Phe Ser Leu  
 195 200 205

Arg Ser Thr Ser Ala Ser Arg Ala Leu Ser Leu Glu Ala Ile Asn Ile  
 210 215 220

Val Gly Asn His Leu Arg Gln Ser Val Ala Asn Gly Glu Asp Lys Glu  
 225 230 235 240

ES 2 817 001 T3

Ala Arg Cys Gly Met Ser His Gly Ser Leu Leu Ala Gly Met Ala Phe  
 245 250 255

Asn Asn Gly Phe Leu Gly Leu Ala His Ala Ile Gly Ser Ala Leu Ser  
 260 265 270

Gly His Cys His Val Pro His Gly Val Ala Ile Gly Leu Leu Leu Pro  
 275 280 285

His Val Val Glu Phe Asn Ser Ser Glu Cys Pro Asp Gln Ala Ala Glu  
 290 295 300

Ile Ala Lys Ile Leu Gly Val Lys Ala Glu Asp Glu Arg Gln Leu Ala  
 305 310 315 320

Glu Gln Ala Ser His Ala Val Gly Asp Leu Val Lys Asp Ile Gly Leu  
 325 330 335

Pro Thr Arg Leu Arg Asp Met Asn Val Pro Glu Glu Lys Leu Ala Asp  
 340 345 350

Ile Ala Arg Asp Ser Phe Gln Ser Gly Met Met Lys Phe Asn Pro Arg  
 355 360 365

Arg Ala Ser Glu Ser Glu Val Leu Glu Leu Leu His Arg Val Tyr  
 370 375 380

<210> 3

<211> 391

<212> PRT

<213> Geobacillus kaustophilus

5

<400> 3

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

His Leu Pro Thr Ser Ile Gln Phe Gly Tyr Gly Lys Ala Ser Gln Leu  
 20 25 30

Gly Asn Gln Leu Val Asp Met Gly Ile Lys Ser Ala Phe Leu Val Thr  
 35 40 45

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

Leu Glu Ser Ser Asn Ile Gln Tyr Cys Val Tyr Ala Asp Val Glu Pro  
 65 70 75 80

ES 2 817 001 T3

Asp Pro Ser Leu Glu Thr Ile Asp Gln Gly Ala Ala Ala Phe Lys Glu  
85 90 95

Gln Pro Phe Asp Cys Ile Val Ala Ile Gly Gly Gly Ser Pro Ile Asp  
100 105 110

Thr Ala Lys Gly Ile Arg Val Val Ala Thr Asn Gly Gly Ser Ile Gly  
115 120 125

Asp Tyr Ala Gly Val Asn Arg Ile Lys Lys Lys Ser Glu Ile Pro Leu  
130 135 140

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

Gly Val Tyr Ser Asp Trp Lys Asn Asn Val Lys Val Thr Val Thr Ser  
165 170 175

Pro Tyr Met Ala Pro Glu Ile Ala Leu Val Asp Pro Lys Leu Thr Met  
180 185 190

Ser Leu Pro Lys Lys Ile Thr Ala Ala Ser Gly Ile Asp Ala Leu Ala  
195 200 205

His Gly Ile Glu Thr Phe Phe Ser Leu Arg Ser Gln Pro Ile Ser Asp  
210 215 220

Val Leu Ala Ile Glu Ala Met Thr Thr Val Asn Arg Tyr Leu Arg Arg  
225 230 235 240

Ala Val Glu Asp Gly Thr Asn Lys Glu Ala Arg Ile Gly Met Ser Tyr  
245 250 255

Gly Ser Leu Leu Ala Gly Met Ala Phe Asn Asn Gly Phe Leu Gly Leu  
260 265 270

Ala His Ala Ile Gly Ser Ala Leu Ser Gly His Cys His Val Ser His  
275 280 285

Gly Val Ala Ile Gly Leu Leu Leu Pro Lys Val Val Glu Phe Asn Ser  
290 295 300

Val Val Gln Pro Glu Lys Ala Ala Lys Ile Ala Glu Leu Leu Gly Arg  
305 310 315 320

Lys Gly Asn Gln Asn Thr Leu Val Gln Gln Ala Ala Leu Ala Val Ala  
325 330 335

ES 2 817 001 T3

Ser Leu Val Lys Glu Ile Gly Leu Pro Thr Arg Leu Arg Asp Val Asp  
 340 345 350

Val Pro Lys Glu Lys Leu Pro Asp Ile Ala Lys Asp Ser Phe Lys Ser  
 355 360 365

Gly Met Met Arg Phe Asn Pro Arg Gln Pro Ser Glu Ala Glu Val Met  
 370 375 380

Thr Leu Leu Gln Gln Ile Tyr  
 385 390

<210> 4

<211> 390

<212> PRT

5 <213> Aneurinibacillus terranovensis

<400> 4

Met Ser Pro Ala Val Lys Ala Ile Asn Phe Glu Phe Ser Phe Asn Leu  
 1 5 10 15

Pro Thr Leu Ile Glu Phe Gly Tyr Gly Lys Met Glu Lys Phe Gly Gln  
 20 25 30

Gln Leu Ile Ser Ile Gly Val Lys Arg Ile Phe Met Val Thr Asp Lys  
 35 40 45

Gly Val Glu Ser Ala Gly Leu Leu Ala Ala Leu Thr Asp Ser Leu Gln  
 50 55 60

Ala Ala Ala Ile Gln Phe Asp Ile Tyr Thr Asp Val Glu Ser Asp Pro  
 65 70 75 80

Ser Leu Glu Thr Ile Asp Arg Gly Val Glu Val Phe Gln Gln Lys Pro  
 85 90 95

Tyr Asp Cys Ile Val Ala Val Gly Gly Gly Ser Pro Ile Asp Thr Ala  
 100 105 110

Lys Gly Ile Arg Val Val Ala Ala Asn Gly Gly Asn Ile Gly His Tyr  
 115 120 125

Ala Gly Val Asn Gln Ile Pro Val Ala Pro Thr Ile Pro Leu Leu Ala  
 130 135 140

Ile Pro Thr Thr Ser Gly Thr Gly Ser Glu Val Thr Asn Phe Gly Val  
 145 150 155 160

ES 2 817 001 T3

Tyr Ser Asp Trp Gln Asn Asn Val Lys Val Thr Val Thr Ser Gln Tyr  
 165 170 175  
 Met Ala Pro Thr Ile Ala Trp Val Asp Pro Ala Leu Thr Met Ser Leu  
 180 185 190  
 Pro Ala Lys Met Thr Ala Ala Ser Gly Ile Asp Ala Leu Ala His Gly  
 195 200 205  
 Ile Glu Thr Phe Phe Ser Leu Gly Ser Ser Pro Ala Ser Asp Ala Leu  
 210 215 220  
 Ala Ile Glu Ala Ile His Thr Val Asn Arg Tyr Leu Ser Arg Ala Val  
 225 230 235 240  
 His Asn Gly Ser Asp Met Glu Ala Arg Ile Gly Met Ser His Gly Ser  
 245 250 255  
 Leu Leu Ala Gly Met Ala Phe Asn Asn Gly Phe Leu Gly Leu Ala His  
 260 265 270  
 Ala Ile Gly Ser Ala Leu Ser Gly His Cys His Val Pro His Gly Val  
 275 280 285  
 Ala Ile Gly Leu Leu Leu Pro Lys Val Val Glu Phe Asn Ala Thr Val  
 290 295 300  
 Arg Pro Asp Lys Ala Ala Lys Ile Ala Gly Leu Met Gly Met Lys Gly  
 305 310 315 320  
 Glu His Ser Glu Glu Leu Ala Leu Gln Ala Ser Pro Ala Val Ala Arg  
 325 330 335  
 Leu Val Glu Asp Ile Gly Leu Pro Thr Arg Leu Arg Glu Val Asp Val  
 340 345 350  
 Thr Glu Lys Lys Leu Phe Glu Ile Ala Lys Asp Ser Phe Lys Ser Gly  
 355 360 365  
 Met Met Lys Phe Asn Pro Arg Gln Pro Ser Glu Ser Glu Val Leu Gln  
 370 375 380  
 Leu Leu Lys Glu Ile Phe  
 385 390

<210> 5

<211> 390

<212> PRT

5 <213> Brevibacillus thermoruber

<400> 5

ES 2 817 001 T3

Met Ser Gln Thr Val Gln Gly Thr Asp Phe Ala Phe Ser Phe His Leu  
1 5 10 15

Pro Thr Leu Ile Glu Phe Gly Tyr Gly Arg Ala Ser Arg Leu Gly Glu  
20 25 30

Arg Leu Gln His Leu Gly Val Thr Asn Val Phe Val Val Thr Asp Lys  
35 40 45

Gly Val Glu Ala Ala Gly Leu Leu Asn Gly Leu Val Gly Ser Leu Gln  
50 55 60

Ser Ala Gly Ile Ala Phe Asp Leu Tyr Thr Glu Val Glu Pro Asp Pro  
65 70 75 80

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

Tyr Asp Cys Leu Val Ala Val Gly Gly Gly Ser Pro Ile Asp Ala Ala  
100 105 110

Lys Gly Met Arg Val Val Thr Ser Cys Gly Gly Ser Ile Ala Asp Tyr  
115 120 125

Ala Gly Val Asn Arg Val Pro Met Ala Pro Ala Val Pro Leu Val Ala  
130 135 140

Val Pro Thr Thr Ser Gly Thr Gly Ser Glu Val Thr Met Phe Gly Val  
145 150 155 160

Tyr Ser Asp Trp His Asn His Val Lys Val Thr Val Thr Ser Pro His  
165 170 175

Met Ala Pro Thr Ile Ala Leu Val Asp Pro Ala Leu Thr Val Ser Leu  
180 185 190

Pro Ala Lys Met Thr Ala Ala Ser Gly Ile Asp Ala Leu Ala His Gly  
195 200 205

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

Ala Met Glu Ala Ile Ala Ala Val Asn Ala His Leu Arg Arg Ala Val  
225 230 235 240

ES 2 817 001 T3

His Asp Gly Ser Asp Val Glu Ala Arg Ile Gly Met Ser His Gly Ser  
 245 250 255

Leu Leu Ala Gly Met Ala Phe Thr Asn Gly Phe Leu Gly Leu Ala His  
 260 265 270

Ala Ile Gly Ser Ala Leu Ser Gly His Cys His Val Pro His Gly Ile  
 275 280 285

Ala Ile Gly Leu Leu Leu Pro His Val Val Ala Phe Asn Ala Pro Ala  
 290 295 300

Arg Pro Asp Lys Ala Ala Gln Leu Ala Arg Leu Leu Gly Val Glu Ala  
 305 310 315 320

Asn Pro Arg Glu Glu Arg Gly Glu Glu Thr Ser Ala Ala Val Ala Arg  
 325 330 335

Met Val Ala Asp Ile Gly Leu Pro Thr Arg Leu Arg Asp Val Gly Val  
 340 345 350

Pro Glu Glu Lys Leu Pro Ala Ile Ala Lys Asp Ala Phe Lys Ser Gly  
 355 360 365

Met Met Thr Cys Asn Pro Arg Gln Pro Thr Glu Gln Glu Val Arg Glu  
 370 375 380

Leu Leu Arg Arg Ala Phe  
 385 390

<210> 6

<211> 379

<212> PRT

5 <213> Brevibacillus panacihumi

<400> 6

Met Glu Ser Pro Phe Ser Phe His Leu Pro Thr Asn Val Gln Phe Gly  
 1 5 10 15

Val Gly Ser Ala Ser Arg Leu Gly Glu Met Leu Leu Ser Met Gly Val  
 20 25 30

Arg Arg Val Phe Leu Val Thr Asp Gln Gly Val Arg Gln Ala Gly Leu  
 35 40 45

Leu Asp Glu Val Ile His Ser Leu Glu Glu Lys Gly Leu His Phe Gln  
 50 55 60

ES 2 817 001 T3

Ile Tyr Ala Asp Val Glu Pro Asp Pro Ser Leu Glu Thr Ile Gln Ala  
65 70 75 80

Gly Ala Ala Met Phe Gln Gln Gln Ser Phe Asp Cys Met Val Ala Ile  
85 90 95

Gly Gly Gly Ser Pro Ile Asp Thr Ala Lys Gly Ile Arg Val Leu Ala  
100 105 110

Ala Asn Gly Gly Gly Ile Gly Gln Tyr Ala Gly Val Asn Arg Val Pro  
115 120 125

Ala Ala Ser Ala Ile Pro Leu Ile Ala Ile Pro Thr Thr Ser Gly Thr  
130 135 140

Gly Ser Glu Val Thr Ile Phe Gly Val Tyr Ser Asp Trp Glu Asn His  
145 150 155 160

Val Lys Ile Thr Val Thr Ser Pro His Met Ala Pro Ser Thr Ala Leu  
165 170 175

Ile Asp Pro Ala Leu Thr Leu Ser Leu Pro Ala Lys Met Thr Ala Ala  
180 185 190

Thr Gly Ile Asp Ala Leu Ala His Gly Ile Glu Thr Phe Phe Ser Leu  
195 200 205

Arg Ser Ser Pro Ala Ser Asp Ala Leu Ala Ile His Ala Met Lys Met  
210 215 220

Ile Ala Pro His Leu Arg Arg Ala Val Arg Asp Gly Ala Asp Met Glu  
225 230 235 240

Ala Arg Ile Gly Met Ser Gln Gly Ser Val Leu Ala Gly Met Ala Phe  
245 250 255

Asn Asn Gly Phe Leu Gly Leu Ala His Ala Ile Gly Ser Ala Leu Ser  
260 265 270

Gly His Cys His Val Pro His Gly Val Ala Ile Gly Leu Leu Leu Pro  
275 280 285

His Val Val Ala Phe Asn Thr Pro Val Arg Pro Glu Lys Ala Glu Leu  
290 295 300

Ile Ala Asp Val Leu Gly Ser Val Gln Lys Glu Thr Gly Thr Ala Ala



ES 2 817 001 T3

Met Tyr Pro Ser Phe Glu Phe His Leu Pro Thr Lys Ile His Phe Gly  
 1 5 10 15

Tyr Asn Thr Ile Lys Gln Leu Asp His Leu Pro Phe Glu Ile Lys Arg  
 20 25 30

Ala Phe Ile Val Thr Asp Gln Gly Val Leu Asn Ser Gly Leu Val Glu  
 35 40 45

Asn Val Thr Asn Ile Leu Lys Asp His Gln Ile Ser Tyr Val Ile Tyr  
 50 55 60

Ser Glu Val Glu Pro Asp Pro Ser Val Glu Thr Val Asp Lys Ala Ala  
 65 70 75 80

Gln Met Phe Gln Arg Glu Glu Ala Asp Ala Leu Ile Ala Ile Gly Gly  
 85 90 95

Gly Ser Pro Ile Asp Thr Ala Lys Gly Val Arg Val Ile Ala Gly Asn  
 100 105 110

Gly Gly Ser Ile Arg Asp Tyr Ala Gly Val Asn Leu Ile Lys Gln Lys  
 115 120 125

Ser Asn Ile Pro Leu Ile Ala Ile Pro Thr Thr Ser Gly Thr Gly Ser  
 130 135 140

Glu Val Thr Ile Phe Ala Val Phe Ser Asp Trp Glu Glu Asn Arg Lys



ES 2 817 001 T3

<400> 8

Met Glu Ala Phe Thr Phe Gln Leu Lys Thr Thr Val Cys Phe Gly Ala  
 1 5 10 15

Asn Val Val Ser Gly Ile Val Asp Trp Cys Arg Asn Tyr Asn Ala Lys  
 20 25 30

Arg Val Leu Ile Val Thr Asp Gln Gly Val Arg Lys Ala Gly Ile Leu  
 35 40 45

Glu Lys Val Glu Lys Ile Leu Ser Asp Ala Gly Ile Glu Asn Val Val  
 50 55 60

Phe Asp Asp Val Glu Pro Asp Pro Gly Leu Glu Thr Ile His Arg Cys  
 65 70 75 80

Ala Ser Cys Phe Arg Glu Asn Lys Cys Asp Leu Ile Leu Ala Val Gly  
 85 90 95

Gly Gly Ser Pro Ile Asp Thr Ala Lys Gly Ala Arg Val Ile Val Glu  
 100 105 110

Asn Gly Gly His Ile Arg Asp Tyr Ala Gly Val Asn Lys Val Pro Arg  
 115 120 125

Ala Pro Val Thr Pro Leu Ile Ala Ile Pro Thr Thr Ser Gly Thr Gly  
 130 135 140

Ser Glu Val Thr Thr Phe Ala Val Leu Ser Asp Trp Glu Asn Arg Met  
 145 150 155 160

Lys Ile Thr Ile Ser Ser Pro Phe Leu Ala Pro Glu Val Ala Val Val  
 165 170 175

Asp Pro Leu Leu Thr Met Thr Ala Pro Pro Ser Val Thr Ala Ala Ser  
 180 185 190

Gly Ile Asp Ala Leu Ser His Ala Ile Glu Thr Tyr Val Ser Leu Lys  
 195 200 205

Ala Gln Pro Pro Ala Glu Ala Leu Ala Leu Lys Ala Ile Glu Leu Ile  
 210 215 220

Gly Glu Ser Leu Arg Thr Ala Val Ala Asp Gly Ser Asp Lys Glu Ala  
 225 230 235 240

ES 2 817 001 T3

Arg Thr Arg Met Ser Leu Gly Ser Leu Leu Ala Gly Met Ala Phe Asn  
 245 250 255

Asn Ser Leu Leu Gly Leu Thr His Ser Ile Gly Ala Ala Leu Ser Gly  
 260 265 270

His Ala His Val Ser His Gly Met Ala Ile Gly Leu Leu Leu Pro Tyr  
 275 280 285

Val Met Glu Phe Asn Ala Met Ala Arg Met Glu Lys Phe Ser Lys Ile  
 290 295 300

Ala Val Ala Leu Gly Glu Asp Val Lys Gly Leu Ser Leu Arg Glu Ala  
 305 310 315 320

Ala Leu Arg Ser Val Lys Ala Val Arg Glu Leu Val Glu Asp Ile Ser  
 325 330 335

Leu Pro Arg Arg Leu Gly Asp Val Gly Val Thr Gly Asp Met Ile Glu  
 340 345 350

Gly Met Ala Lys Asp Ala Met Gly His Gly Met Leu Lys Phe Asn Pro  
 355 360 365

Arg Ala Val Thr Glu Lys Asp Ile Ile Ala Ile Leu Arg Lys Ala Leu  
 370 375 380

<210> 9

5 <211> 377

<212> PRT

<213> Desulfurispora thermophila

<400> 9

Met Lys Thr Thr Val Cys Phe Gly Ala Asn Ile Val Ser Ser Ile Asp  
 1 5 10 15

Asp Arg Cys Arg Asp Tyr Asn Ala Arg His Val Leu Ile Val Thr Asp  
 20 25 30

Gln Gly Val Glu Lys Ala Gly Ile Leu Glu Lys Val Glu Lys Val Leu  
 35 40 45

Ser Asp Ala Gly Ile Glu Asn Val Val Phe Asp Asp Val Glu Pro Asp  
 50 55 60

Pro Gly Leu Glu Thr Ile His Arg Cys Ala Ser Cys Phe Arg Glu Asn  
 65 70 75 80

ES 2 817 001 T3

Lys Cys Asp Leu Phe Leu Ala Ile Gly Gly Gly Ser Pro Ile Asp Thr  
85 90 95

Ala Lys Gly Ala Arg Ile Ile Val Asp Asn Gly Gly His Ile Arg Asp  
100 105 110

Tyr Ala Gly Val Asn Lys Val Pro Arg Ala Pro Arg Thr Pro Leu Leu  
115 120 125

Ala Ile Pro Thr Thr Ser Gly Thr Gly Ser Glu Val Thr Thr Phe Ala  
130 135 140

Val Leu Ser Asp Trp Glu Asn Arg Met Lys Ile Thr Ile Ser Ser Pro  
145 150 155 160

Phe Leu Ala Pro Glu Val Ala Val Val Asp Pro Ile Leu Thr Leu Thr  
165 170 175

Ala Pro Pro Ser Val Thr Ala Ala Ser Gly Ile Asp Ala Leu Ser His  
180 185 190

Ala Ile Glu Thr Tyr Val Ser Leu Lys Ala Gln Pro Pro Ala Glu Ala  
195 200 205

Leu Ala Leu Lys Ala Ile Glu Leu Ile Gly Glu Ser Leu Arg Ala Ala  
210 215 220

Val Ala Asp Gly Ser Asn Lys Glu Ala Arg Thr Lys Met Ser Leu Gly  
225 230 235 240

Ser Leu Leu Ala Gly Met Ala Phe Asn Asn Ser Leu Leu Gly Leu Thr  
245 250 255

His Ser Ile Gly Ala Ala Leu Ser Gly His Ala His Val Ser His Gly  
260 265 270

Met Ala Val Gly Leu Leu Leu Pro Tyr Val Met Glu Phe Asn Ala Met  
275 280 285

Ala Arg Leu Glu Lys Tyr Gly Lys Ile Ala Ile Ala Leu Gly Glu Asp  
290 295 300

Val Lys Gly Leu Ser Leu Arg Glu Ala Ala Leu Arg Ser Val Lys Ala  
305 310 315 320

Val Arg Glu Leu Val Glu Asp Ile Ser Leu Pro Arg Arg Leu Gly Glu  
325 330 335

ES 2 817 001 T3

Val Gly Val Thr Gly Asp Met Ile Glu Gly Met Ala Lys Asp Ala Met  
 340 345 350

Gly His Gly Met Leu Lys Phe Asn Pro Arg Val Val Thr Glu Lys Asp  
 355 360 365

Ile Met Ala Ile Leu Gln Lys Ala Leu  
 370 375

<210> 10

<211> 383

<212> PRT

5 <213> Bacillus sp. L1(2012)

<400> 10

Met Tyr Thr Ser Phe Asn Phe His Leu Pro Thr Arg Ile Gln Phe Gly  
 1 5 10 15

Tyr Glu Lys Val Lys Glu Leu Lys Asn Leu Pro Phe Gln Ala Asn Arg  
 20 25 30

Ala Phe Ile Val Thr Asp Lys Gly Val Glu Lys Ala Gly Leu Leu Asn  
 35 40 45

Asp Val Ile Asp Ala Ile Lys Gln Ala Asn Met Thr Tyr Lys Ile Tyr  
 50 55 60

Arg Asp Val Glu Pro Asp Pro Ser Val Glu Thr Val Asp Lys Ala Ala  
 65 70 75 80

Lys Ala Phe Ala Glu Ala Glu Cys Asp Leu Leu Ile Ala Val Gly Gly  
 85 90 95

Gly Ser Pro Ile Asp Thr Ala Lys Gly Val Arg Val Val Ala Ser Asn  
 100 105 110

Gly Gly Ser Ile Arg Asn Tyr Ser Gly Val Asn Leu Val Lys Glu Ala  
 115 120 125

Pro Ser Val Pro Leu Val Ala Ile Pro Thr Thr Ala Gly Thr Gly Ser  
 130 135 140

Glu Val Thr Ile Phe Ala Val Phe Ser Asp Asp Lys Glu Asn Arg Lys  
 145 150 155 160

Val Thr Val Thr Ser Ser His Leu Ser Pro Asp Val Ser Ile Ile Asp  
 165 170 175

ES 2 817 001 T3

Pro Lys Leu Thr Leu Thr Ala Pro Pro Ser Ile Thr Ala Ala Ala Gly  
 180 185 190

Phe Asp Ala Phe Ala His Ala Ala Glu Ala Phe Val Ser Arg Ile Ser  
 195 200 205

Gln Pro Pro Ser Asp Ala Leu Ala Leu Ser Ala Met Lys Thr Val His  
 210 215 220

Thr Tyr Leu Arg Arg Ala Val Tyr Asn Gly Asp Asp Ile Glu Ala Arg  
 225 230 235 240

Met Lys Met Ala Glu Ala Ser Leu Leu Ala Gly Met Ala Phe Asn Gln  
 245 250 255

Ser Tyr Leu Gly Leu Ala His Ala Ile Gly Ser Ala Ile Ser Val His  
 260 265 270

Ala His Val Ser His Gly Val Val Ile Gly Leu Leu Leu Pro Lys Val  
 275 280 285

Ile Glu Tyr Asn Leu Val Ala Lys Ile Asp Lys Tyr Ala Glu Ala Gly  
 290 295 300

Lys Tyr Ile Glu Gln Ser Ser His Gly Leu Ser Asn Tyr Glu Ala Ala  
 305 310 315 320

Ala Leu Phe Ser Glu Thr Val Thr Gln Leu Arg Asn Asp Ile Gly Leu  
 325 330 335

Pro Lys Gln Leu Arg Glu Val Asn Val Lys Glu Ala Gln Leu Glu Ala  
 340 345 350

Ile Ser Lys Asp Ser Ile Lys Ser Gly Met Trp Gln Phe Asn Pro Arg  
 355 360 365

Arg Ala Ser Glu Gln Asp Val Tyr Gln Met Leu Arg Glu Met Leu  
 370 375 380

<210> 11

<211> 387

<212> PRT

<213> Pelotomaculum thermopropionicum

<400> 11

Met Ala Asp Tyr Asn Phe Ser Phe Ala Val Arg Thr Lys Val Phe Phe  
 1 5 10 15

5

ES 2 817 001 T3

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

Gly Cys Lys Lys Ala Val Leu Val Ser Asp Pro Gly Ile Val Gly Thr  
 35 40 45

Gly Leu Ala Asp Arg Val Lys Asp Leu Leu Ala Gly Gly Gly Val Ala  
 50 55 60

Val Glu Val Phe Ser Glu Val Glu Pro Asp Pro Gly Leu Glu Thr Val  
 65 70 75 80

His Lys Ala Ala Ala Phe Leu Gly Arg Thr Arg Pro Asp Cys Leu Val  
 85 90 95

Ala Leu Gly Gly Gly Ser Ser Ile Asp Val Ala Lys Gly Ala Arg Val  
 100 105 110

Ile Tyr Asp Asn Gly Gly Lys Ile Ser Asp Tyr Ala Gly Val Asn Lys  
 115 120 125

Val Lys Val Lys Pro Ser Leu Pro Leu Met Ala Val Pro Thr Thr Ala  
 130 135 140

Gly Thr Gly Ser Glu Val Thr Val Phe Ala Val Leu Ser Asp Trp Glu  
 145 150 155 160

Gln Asn Ile Lys Ile Thr Val Thr Ser Glu Tyr Leu Ala Pro Glu Ala  
 165 170 175

Ala Phe Val Asp Pro Leu Ala Met Val Ser Ala Pro Pro Gly Ile Thr  
 180 185 190

Ala Ala Ser Gly Ile Asp Ala Leu Ser His Ala Val Glu Ala Tyr Val  
 195 200 205

Ser Arg Ala Ala Ser Pro Val Ser Asp Asn Leu Ala Leu Gly Ala Val  
 210 215 220

Glu Leu Ile Gly Gly His Leu Arg Gln Ala Val Ala Asn Gly Gly Asp  
 225 230 235 240

Leu Ala Ala Arg Thr Gly Ala Ala Leu Gly Ser Leu Leu Ala Gly Met  
 245 250 255

Ala Phe Asn Asn Ala Phe Leu Gly Leu Thr His Ser Ile Gly Ala Ala  
 260 265 270

ES 2 817 001 T3

Leu Ser Gly His Val His Val Ser His Gly Val Ala Val Gly Leu Leu  
275 280 285

Leu Pro Tyr Val Met Glu Tyr Asn Leu Met Ala Lys Pro Asp Lys Phe  
290 295 300

Ala Arg Leu Ala Arg Ala Met Gly Glu Val Thr Glu Gly Lys Ser Leu  
305 310 315 320

Tyr Arg Ala Ala Ser Leu Ala Pro Arg Ala Val Lys Ala Met Val Lys  
325 330 335

Ser Ile Gly Leu Pro Val Arg Leu Lys Glu Ile Gly Val Pro Glu Gly  
340 345 350

Ala Leu Ala Ala Ile Ala Glu Thr Ala Leu Lys His Gly Met Ile Lys  
355 360 365

Phe Asn Pro Arg Val Pro Ser Arg Glu Asp Ile Leu Asp Ile Val Lys  
370 375 380

Lys Ala Tyr  
385

<210> 12

<211> 1176

5 <212> ADN

<213> Aeribacillus pallidus

<400> 12

ES 2 817 001 T3

atgatcggaa attacgcaaa aaagggcatt gatttcogagt tcacttttta tottcctaca 60  
 ttgatcgaat tcgatacgg caaggcttcc cgaatgggag agatgcttga acagatgggt 120  
 ataaagaacg tttttttggt taccgacaaa ggagtggaag ctgcgggtct gttggcagga 180  
 atcgttcagt ctctggaatc atccaatcgc cgatatgtta tttattcaga cgtagaacct 240  
 gacccgagct tagagacgat tgatcgtggt gcgtccggtt ttaaggagca gtcttttgac 300  
 tgtatcttag ctgtgggtgg aggaagtccg attgatacag ctaaggggat ccgtgtcgta 360  
 gtgacgaacg gaggaacat cggtgactat gccggtgtta accgtgttgc gaaaaaatct 420  
 gaaattcctt tgggtggctgt gccgactaca tccggcacgg gcagtgaagt aaccattttc 480  
 ggagtctact ccgattggga aatcaagta aaggtgacgg taacaagccc atatatggcg 540  
 ccggagatcg ctttggtaga cccggaactt accatgagtc taccgcaaaa aatgacagca 600  
 gcatcgggaa ttgatgctct agctcatggg attgaaactt tcttctcctt gcgttctcga 660  
 cctgcatccg atgccctagc ggtcgaagcg atggcgacgg tgagtgttga tttgcgccgt 720  
 gcggtggaag atggtacgga taaagaagcg aggatcggca tgtcccaggg cagtttggtg 780  
 gcagggatgg cattcaacaa tggcttctta ggtttggccc atgcgatcgg tagtgctttg 840  
 tctggccatt gtcatgtgtc ccatgggtgtc gcaatcgggt tgttgctacc gaaagtgggtg 900  
 gaatttaatg ctagggtgcg cccggaaaaa gctgcaaaaa tcgcagaatt gttgggagta 960  
 aaaggggatc gagaggaggt tcttgccggag caggcagctc ctgcagtcgc ctcgtagtc 1020  
 aaagagattg gtcttccac tcgtttgcgt gatgttgatg tttctgaaga aaagctccca 1080  
 gatatcgsaa gagatgcatt taaaagcgg atgatgaagt ttaaccacg ccaaccaagt 1140  
 ttgtcagaag tgcttacct tttgcagcag atttat 1176

- 5 <210> 13
- <211> 1179
- <212> ADN
- <213> Artificial

- <220>
- 10 <223> Codón sintético optimizado para la expresión en P. putida

<400> 13

ES 2 817 001 T3

atgatcggca actacgcca gaaggccatc gacttcgagt tcaccttcta cctgccgacc 60  
ctgatcgagt tcggctacgg caaggccagc cgcatggcg agatgctgga acaaatgggt 120  
atcaagaacg tgttcctggg gaccgacaag ggcgtggaag ccgccggtct gctggccggc 180  
atcgtgcaga gcctggaaag cagcaacatc cgctacgtga tctacagcga cgtggaaccg 240  
gacccgagcc tggaaacat cgaccgaggc gccagcgtgt tcaaagaaca gagcttcgac 300  
tgcacacctg ccgtggggcg cggcagcccg atcgacaccg ccaagggcat ccgctgggtg 360  
gtgaccaacg gcggcaacat cggcgactac gccggcgtga accgctggc caagaagtcg 420  
gagatcccgc tggctgccgt gcccaaccacc tcgggcaccg gcagcgaagt gaccatcttc 480  
ggcgtgtaca gcgactggga gaaccaggtg aaggtgaccg tgaccagccc gtacatggcc 540  
ccggaaatcg ccctggtgga cccggaactg accatgagcc tgccgcagaa gatgaccgcc 600  
gccagcggca tcgacgccct ggcccacggc atcgaaacct tcttcagcct gcgcagccgc 660  
ccagcctcgg atgccctggc ggtggaagcc atggccaccg tgagcgccta cctgcgccgc 720  
gccgtcgagg acggcaccga caaagaagcc cgcacggca tgagccaggg cagcctgctg 780  
gggggcatgg ccttcaacaa cggcttcctg ggcctggccc atgccatcgg cagcgcctg 840  
agcggccatt gccatgtgag ccacggcgtg gccatcggcc tgctgctgcc gaaggtggtg 900  
gaattcaacg cccgcgtgcg cccggaaaag gccgccaaga tcgccgaact gctgggcgtg 960  
aagggcgacc gcgaagaggt gctggccgaa caggccgccc cagccgtggc cagcctggtg 1020  
aaagaaatcg gcctgccgac ccgctgcgc gacgtggacg tgagcgaaga gaagctgccg 1080  
gacatcggcc gcgacgcctt caagagcggc atgatgaagt tcaaccgcg ccagccgagc 1140  
ctgagcgagg tgctgaccct gctgcagcag atctactga 1179

- 5 <210> 14
- <211> 1152
- <212> ADN
- <213> Artificial

- <220>
- 10 <223> Codón sintético optimizado para la expresión en P.putida

<400> 14

ES 2 817 001 T3

atggacgtgg aattcagctt ccatctgccg accctgatcg agttcggctt cggcaaggcc 60  
 agcctgctgg gcgagcgcct gctgaagctg ggcgtgggca acgtgttcct ggtgagcgac 120  
 aagggcgtgg ccagcgcagg cctgctgcag aagctggaac agagcctgca gaccagcgac 180  
 atccacttca agacctacct ggaagtggaa ccggacccga gcctggaaac catcgacctg 240  
 ggtgccgagg ccttcaacag cggcaagtac gactgcatcg tggccgtggg tggtagcgac 300  
 gccatcgaca ccgccaaggg catccgcgtg gtggcaggca acggtggcag catcggcgac 360  
 ttgcgaggcg tggacaagat cggcaaggca ccgcagatcc cgctgatcgc cgtgccgacc 420  
 acctcgggca ccggcagcga agtgaccatc ttcggcgtgt acagcgactg ggtgaagaac 480  
 gtgaagggtga ccgtgaccag ccagtacatg gcaccgacca ttgccctggt ggacccggaa 540  
 ctgaccatgc gcctgccacg caagatgacc gcagccagcg gcatcgacgc cctggcccac 600  
 ggcacgaga gctacttcag cctgcgcagc accagcgcca gccgtgccct gtcgctggaa 660  
 gccatcaaca tcgtgggcaa ccatctgcgc cagagcgtgg cgaacggcga ggacaaggaa 720  
 gcacgctgcg gcatgagcca cggcagcctg ctggcaggca tggcgttcaa caacggcttc 780  
 ctgggcctgg cccatgccat cggcagcgcga ctgagcggtc actgccacgt gccgcacggc 840  
 gtggccatcg gcctgctgct gccgcacgtg gtggaattca acagcagcga gtgccagac 900  
 caggcagccg agatcgccaa gatcctgggc gtgaaggccg aggacgaacg ccagctggcc 960  
 gaacaggcca gccacgccgt gggcgacctg gtgaaggaca tcggcctgcc gaccctctg 1020  
 cgcgacatga acgtgccgga agagaagctg gccgacattg cacgcgacag cttccagagc 1080  
 ggcatgatga agttcaacc acgtcgtgcc agcgagagcg aggtgctgga actgctgcac 1140  
 cgctgtact ga 1152

<210> 15  
 <211> 1994  
 <212> ADN  
 <213> Artificial

5

<220>  
 <223> Fragmento de XbaI-Sall sintético

<400> 15

ES 2 817 001 T3

tctagatcct ctttgataat aatgaaagc agccggtatg gagagaaaa agtgcactta 60  
tatgaagttg attttatggt cggctttatt ttgcccgtcg tactggctgt ccacacgatg 120  
ttcatttttg atgcacaatt gaatggctgt acagttgcgt ttttgtcgat gtctggcggg 180  
cacgcctcca tgcatgtgaa gcagattcct ttaagcgggc agcaccgct ttttggagg 240  
gcaggcattc aggagcaaaa atggcagaga tcagttgggc gggatcagcc atttattcct 300  
ccatccgggg cactttgtga aatcagcac aagaatgaat aacgctttca tatctggctt 360  
tttcaaataa aaccatttgt gaaaaatgta aacggatgat tttgaaaaac cgtcattttc 420  
ctttaaaacc gggcatttgg gcagataaat tttcaaattt tcgccataaa atatgtgaat 480  
ctaatacaca aatagtggt atacttacc atgtggaatg aaggaaaatg aacggaacga 540  
tccatttcag ccataaaagg gcatgccgtc catctatttc acaaaccgca cggcagcatt 600  
tgctgcaaaa gttaatcgt cctgctttaa aggaaaagca gtatggaatc cattaggagt 660  
tggcacaata tccatagact ggataggggc ccgccatgcc gggcttgcaa aactgctttc 720  
atacagtga aatattttt acttttgatg gggaggaaga ttatatatga tcggaatta 780  
cgcaaaaaag gcgattgatt tcgagttcac tttttatcct cctacattga tcgaattcgg 840  
atacggcaag gcttcccgaa tgggagagat gcttgaacag atgggtataa agaacgtttt 900  
tttggttacc gacaaaggag tggagctgc gggctctgtg gcaggaatcg ttcagtctct 960  
ggaatcatcc aatatccgat atgttattta ttcagacgta gaacctgacc cgagcttaga 1020  
gacgattgat cgtggtgcgt ccgtttttaa ggagcagtct tttgactgta tcttagctgt 1080  
gggtggagga agtccgattg atacagctaa ggggatccgt gtcgtagtga cgaacggagg 1140  
aaacatcggg gactatgccg gtgttaaccg tgttgcgaaa aaatctgaaa ttcctttggt 1200  
ggctgtgccg actacatccg gcacgggcag tgaagtaacc attttcggag tctactccga 1260  
ttgggaaaat caagtaaagg tgacggtaac aagccatat atggcgccgg agatcgcttt 1320  
ggtagacccc gaacttacca tgagtctacc gcaaaaaatg acagcagcat cgggaattga 1380  
tgctctagct catgggattg aaactttctt ctccttgcgt tctcgacctg catccgatgc 1440  
cctagcggtc gaagcgatgg cgacggtgag tgcttatttg cgccgtgcgg tggagatgg 1500  
tacggataaa gaagcgagga tcggcatgtc ccagggcagt ttgttggcag ggatggcatt 1560  
caacaatggc ttcttaggtt tggcccatgc gatcggtagt gctttgtctg gccattgtca 1620  
tgtgtcccat ggtgtcgcaa tcggtttgtt gctaccgaaa gtggtggaat ttaatgctag 1680  
ggtgcgcccg gaaaaagctg caaaaatcgc agaattgttg ggagtaaaag gggatcgaga 1740  
ggaggttctt gcggagcagg cagctcctgc agtcgcctcg ttagtcaaag agattggtct 1800  
tcccactcgt ttgctgatg ttgatgttc tgaagaaaag ctcccagata tcgcaagaga 1860

ES 2 817 001 T3

tgcatttaaa agcggatga tgaagtttaa cccacgcaa ccaagtttgt cagaagtgct 1920  
 tacacttttg cagcagattt attaattggt cgggtttcag tgttccattt tcaaatttc 1980  
 cgtaagggt cgac 1994

5 <210> 16  
 <211> 1258  
 <212> ADN  
 <213> Artificial

<220>  
 <223> Fragmento de KpnI-XbaI sintético

<400> 16  
 ggtaccttca cacaggaaac aggaggtaca atgatcggaa attacgcaa aaaggcgatt 60  
 gatttcgagt tcaacttttta tcttcttaca ttgatcgaat tcggatacgg caaggcttcc 120  
 cgaatgggag agatgcttga acagatgggt ataaagaacg tttttttggt taccgacaaa 180  
 ggagtggaag ctgccccgtct gttggcagga atcgttcagt ctctggaatc atccaatc 240  
 cgatatgtta tttattcaga cgtagaacct gacccgagct tagagacgat tgatcgtggt 300  
 gcgtccggtt ttaaggagca gtcttttgac tgtatcttag ctgtgggtgg aggaagtccg 360  
 attgatacag ctaaggggat ccgtgctgta gtgacgaacg gaggaacat cggtgactat 420  
 gccggtgtta accgtgttgc gaaaaaatct gaaattcctt tgggtggctgt gccgactaca 480  
 tccggcacgg gcagtgaagt aaccattttc ggagtctact ccgattggga aaatcaagta 540  
 aaggtgacgg taacaagccc atatatggcg ccggagatcg ctttggtaga ccccgaactt 600  
 accatgagtc taccgacaaa aatgacagca gcacgaggaa ttgatgctct agctcatggg 660  
 attgaaactt tcttctcctt gcgttctcga cctgcatccg atgccctagc ggtcgaagcg 720  
 atggcgacgg tgagtgctta tttgcgccgt gcggtggaag atggtagcga taaagaagcg 780  
 aggatcggca tgtcccaggg cagtttggtg gcagggatgg cattcaacaa tggcttctta 840  
 ggtttggccc atgcgatcgg tagtgctttg tctggccatt gtcatgtgtc ccatggtgtc 900  
 gcaatcgggt tgttgctacc gaaagtgggt gaatttaatg ctagggtgcg cccggaaaaa 960  
 gctgcaaaaa tcgcagaatt gttgggagta aaaggggatc gagaggaggt tcttgcgag 1020  
 caggcagctc ctgcagtcgc ctcgttagtc aaagagattg gtcttcccac tcgtttgct 1080  
 gatgttgatg tttctgaaga aaagctocca gatatcgaag gagatgcatt taaaagcgg 1140  
 atgatgaagt ttaaccacg ccaaccaagt ttgtcagaag tgcttacct tttgcagcag 1200  
 10 atttattaat tgttcgggtt tcagtgttcc attttcaaat attccgtaa ggtctaga 1258

<210> 17  
 <211> 446  
 <212> PRT  
 <213> Aeribacillus pallidus

15 <400> 17

ES 2 817 001 T3

Met Lys Asn Ile Ala Asn Thr Ser Thr Glu Arg Pro Val Asn Asp Ala  
 1 5 10 15

Ser Val Lys Asn Arg Gln Met Val Arg Ala Thr Ile Ala Ser Leu Ile  
 20 25 30

Gly Trp Ser Leu Asp Leu Tyr Asp Leu Phe Leu Leu Leu Phe Val Ala  
 35 40 45

Thr Thr Ile Gly Asn Leu Phe Phe Pro Ala Ser Asn Gln Thr Leu Ser  
 50 55 60

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

Leu Gly Ser Ala Ile Phe Gly Ile Tyr Ala Asp Lys Asn Gly Arg Lys  
 85 90 95

Lys Ala Met Thr Val Ala Ile Ile Gly Ala Gly Leu Cys Thr Ala Ala  
 100 105 110

Phe Gly Leu Leu Pro Thr Ile His Gln Val Gly Val Val Ala Ala Ile  
 115 120 125

Ala Phe Leu Ile Leu Arg Leu Val Gln Gly Val Phe Val Gly Gly Val  
 130 135 140

Val Ala Ser Thr His Thr Ile Gly Thr Glu Ser Ala Ser Pro Lys Tyr  
 145 150 155 160

Arg Gly Phe Met Ser Gly Leu Ile Gly Gly Gly Gly Ala Gly Leu Gly  
 165 170 175

Ala Leu Phe Ala Ser Ile Ser Tyr Ser Val Val Thr Ala Ile Phe Pro  
 180 185 190

Gly Glu Ala Phe Asp Val Trp Gly Trp Arg Val Met Phe Phe Thr Gly  
 195 200 205

Ile Ile Gly Ser Leu Phe Gly Leu Phe Ile Phe Arg Ser Leu Glu Glu  
 210 215 220

Ser Pro Leu Trp Lys Gln Leu Lys Glu Glu Asn Ser Lys Gly Glu Val  
 225 230 235 240

ES 2 817 001 T3

Ser Glu Phe Gln Lys Ala Pro Leu Lys Thr Phe Phe Thr Lys Tyr Tyr  
 245 250 255

Lys Val Leu Leu Val Asn Leu Met Ile Val Ile Gly Gly Gly Ser Gly  
 260 265 270

Tyr Tyr Leu Thr Ser Gly Phe Ile Pro Thr Phe Leu Lys Val Val Asn  
 275 280 285

Lys Val Ser Ala Ser Val Ser Ser Gly Val Leu Ile Ala Thr Ser Ile  
 290 295 300

Met Thr Ile Val Ala Ala Val Leu Val Gly His Leu Ser Glu Val Ile  
 305 310 315 320

Gly Arg Lys Lys Thr Phe Leu Leu Ile Gly Ile Leu Cys Leu Val Gly  
 325 330 335

Leu Pro Tyr Phe Tyr Leu Ser Leu Ala Asn Ser Thr Thr Thr Thr Gly  
 340 345 350

Ile Tyr Leu Asn Ala Leu Gly Leu Ile Phe Leu Gly Asn Ala Ala Tyr  
 355 360 365

Ala Pro Val Leu Ile Phe Leu Asn Glu Arg Phe Pro Thr Ser Ile Arg  
 370 375 380

Ser Thr Gly Thr Gly Leu Ser Trp Asn Met Gly Phe Ala Ile Gly Gly  
 385 390 395 400

Met Met Pro Thr Phe Val Asn Leu Ala Ser Gly Thr Val Glu His Ile  
 405 410 415

Pro Tyr Thr Leu Met Tyr Phe Thr Ile Gly Ile Tyr Leu Val Tyr Ile  
 420 425 430

Leu Gly Ser Leu Ile Ile Pro Glu Thr Lys Gly Asn Leu Lys  
 435 440 445

- <210> 18
- <211> 1338
- <212> ADN
- <213> Aeribacillus pallidus

<400> 18

ES 2 817 001 T3

gtgaagaata tgcctaatac gagtaccgaa cgacctgtaa atgatgcttc agttaagaat 60  
cgtcaaattg tgcgagctac gattgcctcg ctcatagggg ggtcactcga tctttacgat 120  
ttattttctgc tgctttttgt tgcgacgacc ataggaatt tgttttttcc cgccagcaat 180  
caaacacttt ctttggctgc cgtgtatgct tcctttgccg ttacgctttt gatgcggcct 240  
ttgggttccg ccattttcgg catttatgcg gataaaaacg ggagaaagaa agcgatgact 300  
gtggcaatca ttggagcagg cttgtgcacg gcggctttcg gtctgttacc tacgatccac 360  
caagttggag tggtcgctgc gatcgccttc ttgattttac gtttagttca aggagtgtt 420  
gtcggcggag tggttgcttc caccatacag ataggaacgg aatccgcac cccaaaatat 480  
cgggggttta tgtcgggatt gatcgggtgg ggaggagcag gattgggagc actgtttgct 540  
tctattttctt attcggttgt gacggcaatt tttccgggag aggcttttga tgtttgggga 600  
tggcgtgtca tgtttttcac aggcattatc ggttccctct tcggcctttt catattccgg 660  
tcccttgagg aatctcctct ctggaacaa ttgaaagaag aaaatagtaa aggcgaagtg 720  
tccgagtttc agaaagcacc gctgaagacg tttttacta aatattacaa ggtattgctc 780  
gtcaacctta tgatcgtcat cgggtggggc tccggttatt atctgactag tggatttatt 840  
cctacatttt taaaggtagt taacaaagta tcagcctctg tttcgtcggg ggtactcatt 900  
gcgacaagta ttatgaccat tgtagccgcc gttctcgtgg gacacctgag cgaggctac 960  
ggcagaaaga aaacatttct gttaatcggg attctttgtc ttgtcggact tccgtatttt 1020  
tatctgtcat tggcaaacct aactacgaca acgggcatct acttaaatgc tcttggactc 1080  
atattcttgg ggaatgctgc atatgcaccg gtactcatct tcttgaacga acgttttccc 1140  
acatcgatcc gttcaacagg taccggatta tcatggaaca tgggtttcgc cattggcggg 1200  
atgatgccga cgtttgtgaa cttagccagt ggtacgggtg aacatattcc ttacacgctg 1260  
atgtatttta ctatcggaat ttacttggtt tatatccttg gcagcctgat tattccggaa 1320  
acaaaaggaa acctcaaa 1338

5 <210> 19  
<211> 1184  
<212> ADN  
<213> Artificial

10 <220>  
<223> Fragmento de KpnI - NheI sintético YiaY de B.kribbensis optimizado para codón

<400> 19

ES 2 817 001 T3

ggtacctagg aaaggaagat taacctatgg acgtggaatt cagcttccat ctgccgaccc 60  
 tgatcgagtt cggcttcggc aaggccagcc tgctgggcga gcgcctgctg aagctgggcg 120  
 tgggcaacgt gttcctggtg agcgacaagg gcgtggccag cgcaggcctg ctgcagaagc 180  
 tggaacagag cctgcagacc agcgacatcc acttcaagac ctacctgga gtggaaccgg 240  
 acccgagcct ggaaaccatc gacctgggtg ccgaggcctt caacagcggc aagtacgact 300  
 gcatcgtggc cgtgggtggt ggcagcgcca tcgacaccgc caagggcatc cgcgtggtgg 360  
 caggcaacgg tggcagcatc ggcgacttcg caggcgtgga caagatcggc aaggcaccgc 420  
 agatcccgt gatcgccgtg ccgaccacct cgggcaccgg cagcgaagtg accatcttcg 480  
 gcgtgtacag cgactgggtg aagaacgtga aggtgaccgt gaccagccag tacatggcac 540  
 cgaccattgc cctggtggac ccggaactga ccatgcgcct gccacgcaag atgaccgcag 600  
 ccagcggcat cgacgcctg gccacggca tcgagagcta cttcagcctg cgcagcacca 660  
 gcgccagccg tgcctgtcg ctggaagcca tcaacatcgt gggcaaccat ctgcgccaga 720  
 gcgtggcgaa cggcgaggac aaggaagcac gctgcggcat gagccacggc agcctgctgg 780  
 caggcatggc gttcaacaac ggcttcctgg gcctggccca tgccatcggc agcgactga 840  
 gcggtcactg ccacgtgccg cacggcgtgg ccatcggcct gctgctgccg cacgtggtgg 900  
 aattcaacag cagcgagtgc ccagaccagg cagccgagat cgccaagatc ctgggcgtga 960  
 aggccgagga cgaacgccag ctggccgaac aggccagcca cgccgtgggc gacctggtga 1020  
 aggacatcgg cctgccgacc cgtctgcgcg acatgaacgt gccggaagag aagctggccg 1080  
 acattgcacg cgacagcttc cagagcggca tgatgaagtt caaccacgt cgtgccagcg 1140  
 agagcgaggt gctggaactg ctgcaccgcg tgtactgagc tagc 1184

5 <210> 20  
 <211> 1359  
 <212> ADN  
 <213> Artificial

<220>  
 10 <223> fragmento de PCR de EcoRI-NheI con secuencia codificante de proP de A.pallidus

<400> 20

ES 2 817 001 T3

gccgaattca tgaagaatat cgctaatacg agtaccgaac gacctgtaa tgatgcttca 60  
 gttaagaatc gtcaaatggt gcgagctacg attgcctcgc tcataggggtg gtcactcgat 120  
 ctttacgatt tatttctgct gctttttggt gcgacgacca tagggaattt gttttttccc 180  
 gccagcaatc aaacactttc tttggctgcc gtgtatgctt cctttgccgt tacgcttttg 240  
 atgcggcctt tgggttccgc cattttcggc atttatgcgg ataaaaacgg gagaaagaaa 300  
 gcgatgactg tggcaatcat tggagcaggc ttgtgcacgg cggctttcgg tctgttacct 360  
 acgatccacc aagttggagt ggtcgctgcg atcgccttct tgattttacg tttagttaa 420  
 ggagtgtttg tcggcggagt ggttgcttcc acccatacga taggaacgga atccgcatcg 480  
 ccaaaatatc gggggtttat gtcgggattg atcgggtggtg gcggagcagg attgggagca 540  
 ctgtttgctt ctatttctta ttcggttgtg acggcaattt ttccgggaga ggcttttgat 600  
 gtttggggat ggcgtgtcat gtttttcaca ggcattatcg gttccctctt cggccttttc 660  
 atattccggt cccttgagga atctcctctc tggaaacaat tgaaagaaga aaatagtaaa 720  
 ggcgaagtgt ccgagtttca gaaagcaccg ctgaagacgt ttttactaa atattacaag 780  
 gtattgctcg tcaaccttat gatcgtcacg ggtggtggct ccggttatta tctgactagt 840  
 ggatttattc ctacattttt aaaggtagtt aacaaagtat cagcctctgt ttcgtcgggg 900  
 gtactcattg cgacaagtat tatgaccatt gtagccgccg ttctcgtggg acacctgagc 960  
 gaggtcatcg gcagaaagaa aacatttctg ttaatcggta ttctttgtct tgtcggactt 1020  
 ccgtattttt atctgtcatt ggcaaactca actacgacaa cgggcatcta cttaaagtct 1080  
 cttggactca tattcttggg gaatgctgca tatgcaccgg tactcatctt cttgaacgaa 1140  
 cgttttccca catcgatccg ttcaacagggt accggattat catggaacat gggtttcgcc 1200  
 attggcggga tgatgccgac gtttgtgaac ttagccagtg gtacgggtgga acatattcct 1260  
 tacacgctga tgtattttac tatcgggaatt tacttggttt atataccttg cagcctgatt 1320  
 attccggaaa caaaaggaaa cctcaaataa gctagcggc 1359

5 <210> 21  
 <211> 20  
 <212> ADN  
 <213> Artificial

10 <220>  
 <223> sitio de unión al ribosoma

<400> 21  
 taggaaagga agattaacct 20

<210> 22  
 <211> 30

<212> ADN  
<213> Artificial

<220>  
<223> Cebador PCRproP(f)

5 <400> 22  
gccgaattca tgaagaatat cgctaatacg 30

<210> 23  
<211> 34  
<212> ADN  
10 <213> Artificial

<220>  
<223> Cebador PCRproP(r)

<400> 23  
gccgctagct tatttgaggt ttcctttgt ttcc 34

15 <210> 24  
<211> 500  
<212> PRT  
<213> Cupriavidus basilensis

<400> 24

ES 2 817 001 T3

Met Asn Ala Gln His Trp Ile Ala Gly Ala Trp Thr Gly Glu Pro Ser  
 1 5 10 15

Ala Asp Ser Val Asn Pro Ala Asp Gly Thr Leu Ile Gly Gln Phe Ala  
 20 25 30

Asp Gly Gly Thr Trp Gln Ala Glu Ala Ala Ile Ala Ala Ala Arg His  
 35 40 45

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

Val Leu Leu Ala Trp Ala Gly Ala Leu Glu Ala Glu Arg Glu Arg Leu  
 65 70 75 80

Ala Ser Leu Leu Thr Ala Glu Asn Gly Lys Pro Val Ala Gln Ala Arg  
 85 90 95

Gly Glu Val Gly Ala Ala Ile Ser Glu Val Arg Tyr Tyr Ala Gly Leu  
 100 105 110

Ala Arg His Ile Pro Gly His Val Leu Glu Pro Glu Pro Gly Thr Ile  
 115 120 125

Ser Thr Ile Leu Arg Glu Pro Ala Gly Val Ala Ala Ile Ile Val Pro  
 130 135 140

Trp Asn Ala Pro Ala Val Leu Leu Val Arg Ser Leu Ala Pro Ala Leu  
 145 150 155 160

Ala Ala Gly Cys Thr Ala Val Val Lys Ser Ala Ala Gln Thr Thr Leu  
 165 170 175

Phe Thr Ala Ala Met Leu Arg Leu Phe Glu Arg Thr Ala Leu Pro Ala  
 180 185 190

Gly Ala Val Asn Leu Val Cys Glu Thr Gly Tyr Ala Ala Ala Asp His  
 195 200 205

Leu Val Arg Ser Arg Asp Val Asp Val Val Ser Phe Thr Gly Ser Thr  
 210 215 220

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

ES 2 817 001 T3

Leu Ser Leu Glu Leu Gly Gly Lys Ser Cys Cys Leu Val Phe Asp Asp  
 245 255

Val Asp Ala Gln Ala Val Ala Lys Arg Leu Ala Leu Ala Ala Thr Val  
 260 265 270

Ile Ser Gly Gln Gln Cys Thr Ala Ala Arg Arg Val Leu Val His Glu  
 275 280 285

Ala Ile Ala Pro Gln Met Arg Arg His Leu Thr Glu Ala Leu Ala Ala  
 290 295 300

Leu Arg Leu Gly Pro Gly Ile Glu Pro Asp Thr Gln Ile Gly Pro Leu  
 305 310 315 320

Ile Asp His Pro Thr Arg Ala Met Val Ser Ala Gln Val Glu Arg Ala  
 325 330 335

Cys Asp Glu Ala Asp Thr Val Leu Leu Arg Gly Thr Met Pro Gly Gly  
 340 345 350

Ala Leu Ala Arg Gly Ala Phe Leu Ser Pro Thr Leu Val Glu His Ser  
 355 360 365

Asp Pro Gly Ala Phe Phe Cys Gln Glu Glu Ile Phe Gly Pro Phe Val  
 370 375 380

Thr Phe Glu Thr Phe Ala Thr Glu Asp Glu Ala Leu Ala Lys Ala Asn  
 385 390 395 400

Asn Thr Val Phe Gly Leu Ser Ala Ser Val Trp Thr His His Gly Glu  
 405 410 415

Arg Ala Ile Arg Leu Ala Arg Ala Leu Arg Asn Gly Thr Val Trp Val  
 420 425 430

Asn Asp His Asn Arg Leu Phe Ala Glu Ala Glu Thr Gly Gly Tyr Arg  
 435 440 445

Gln Ser Gly Leu Gly Arg Leu His Gly Tyr Asp Ala Leu Ala Asp Phe  
 450 455 460

Thr Glu Leu Lys His Ile Cys Ile Gln Ala Gly Leu Pro Lys Gly Met  
 465 470 475 480

Ser Gln Ala Gly Cys Arg Leu Ser Gly Val Ala Ala Arg Glu Arg Met  
 485 490 495

Gly Val Ser Val  
500

<210> 25

<211> 479

<212> PRT

5 <213> Burkholderia sp. CCGE1002

<400> 25

ES 2 817 001 T3

Met Asn Ala Arg His Trp Ile Ala Gly Glu Trp Thr Gly Thr Pro Asn  
1 5 10 15

Ile Asp Ser Ile Asp Pro Ala Thr Gly Asp Ala Ile Gly Arg Phe Ala  
20 25 30

Asp Gly Gly Ser Ser Glu Ala Asp Ala Ala Ile Ala Ala Ala Arg His  
35 40 45

Ala Phe Asp Arg Thr Thr Trp Ala Gln Asp Ala Arg Leu Arg Gln Asp  
50 55 60

Val Leu Leu Gly Trp Ala Ser Ala Leu Glu Ala Glu Arg Asp Met Leu  
65 70 75 80

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

Asp Glu Ile Ala Gly Ala Ile Ser Glu Val Arg Tyr Tyr Ala Gly Leu  
100 105 110

Ala Arg His Ile Ala Gly His Val Leu Glu Pro Glu Pro Gly Thr Ile  
115 120 125

Ser Thr Met Leu Arg Glu Ala Ala Gly Val Ala Ala Ile Ile Val Pro  
130 135 140

Trp Asn Ala Pro Ala Val Leu Leu Val Arg Ser Leu Ala Pro Ala Leu  
145 150 155 160

Ala Ala Gly Cys Thr Val Ile Val Lys Pro Ala Ala Gln Thr Ser Leu  
165 170 175

Leu Thr Ala Ala Met Leu Arg Cys Phe Glu His Thr Ala Leu Pro Glu  
180 185 190

Gly Ala Val Asn Leu Val Asn Glu Arg Gly Tyr Ala Ala Ser Gln Arg  
195 200 205

ES 2 817 001 T3

Leu Val Asp Ser His Gly Val Asp Val Val Ser Phe Thr Gly Ser Thr  
 210 215 220

Ala Thr Gly Lys Lys Ile Met Ala Ala Ala Ala Asp Ser Met Lys Lys  
 225 230 235 240

Leu Ser Leu Glu Leu Gly Gly Lys Ser Cys Cys Val Val Phe Asp Asp  
 245 250 255

Ala Asp Val Ala Ala Ile Ala Pro Arg Leu Ala Arg Ala Ala Thr Ile  
 260 265 270

Ile Ser Gly Gln Gln Cys Thr Ala Ala Arg Arg Val Leu Val His Ala  
 275 280 285

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

Leu Arg Val Gly Pro Gly Ile Asp Pro Ala Thr Asp Ile Gly Ala Leu  
 305 310 315 320

Ile Asp Gly Thr Thr Arg Asp Ala Val Ala Arg Thr Ile Glu Arg Ala  
 325 330 335

Cys Gly Thr Ala Glu Arg Val Leu Leu Arg Gly Thr Cys Ser Gly His  
 340 345 350

Ala Phe Leu Ser Pro Thr Leu Val Glu His Asp Asp Pro Lys Ala Phe  
 355 360 365

Phe Cys Gln Asp Glu Ile Phe Gly Pro Phe Val Thr Leu Glu Val Phe  
 370 375 380

Glu Asn Glu Met Glu Ala Ile Glu Lys Ala Asn Asp Thr Val Phe Gly  
 385 390 395 400

Leu Ser Ala Ser Val Trp Thr His Asp Gly Ala Arg Ala Leu Arg Val  
 405 410 415

Ala Arg Ala Leu Arg Asn Gly Thr Val Trp Ile Asn Asp His Asn Lys  
 420 425 430

Leu Phe Ala Glu Ala Glu Thr Gly Gly Tyr Arg Gln Ser Gly Leu Gly  
 435 440 445

Arg Leu His Gly Tyr Asp Ala Leu Ala Asp Phe Thr Glu Leu Lys His  
 450 455 460

Ile Cys Met Pro Ala Gly Val Ala Glu Gly Ile Ala Pro Leu Arg  
465 470 475

<210> 26

<211> 483

5 <212> PRT

<213> Burkholderia graminis C4D1M

<400> 26

ES 2 817 001 T3

Met Glu Arg Asp Ala Met Asn Trp Ile Ala Gly Glu Trp Ala Gly Lys  
1 5 10 15

Pro Val Leu Ala Ser Ser Asp Pro Ser Asn Gly Glu Thr Leu Gly Arg  
20 25 30

Phe Val Ser Ser Asn Thr Gln Asp Ala Asp Ala Ala Val Ser Ala Ala  
35 40 45

Arg His Thr Phe Asp His Thr Thr Trp Ala Gln Asp Ala Arg Arg Arg  
50 55 60

Gln Asp Val Leu Leu Arg Trp Ala Gln Ala Leu Glu Leu Ser Val Glu  
65 70 75 80

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

Ala Arg Gly Glu Met Arg Ala Ala Ile Ser Glu Val Arg Tyr Tyr Ala  
100 105 110

Gly Leu Ala Arg His Ile Ala Gly His Val Ile Glu Pro Glu Pro Gly  
115 120 125

Thr Ile Ser Thr Met Leu Arg Glu Ala Ala Gly Val Ala Ala Ile Ile  
130 135 140

Val Pro Trp Asn Ala Pro Ala Val Leu Leu Val Arg Ser Leu Ala Pro  
145 150 155 160

Ala Leu Ala Ala Gly Cys Thr Ala Ile Val Lys Pro Ala Ala Gln Thr  
165 170 175

Ser Leu Ile Thr Ala Ala Met Ile Arg Cys Leu Asp Gln Pro Ala Leu  
180 185 190

Pro Ala Gly Ala Val Asn Leu Leu Leu Glu Asn Gly Ala Glu Ala Ala  
195 200 205

ES 2 817 001 T3

Ala Arg Leu Val Glu Ser Ala Asp Val Asp Val Ile Ser Phe Thr Gly  
 210 215 220

Ser Thr Glu Val Gly Lys Arg Ile Met Arg Ala Ala Ala Asp Ser Met  
 225 230 235 240

Lys Arg Leu Ser Leu Glu Leu Gly Gly Lys Ser Cys Cys Leu Val Phe  
 245 250 255

Glu Asp Ser Asp Val Lys Ala Ile Ala Pro Arg Leu Ala Arg Ala Ala  
 260 265 270

Thr Ile Ile Ser Gly Gln Gln Cys Thr Ala Ala Arg Arg Ile Leu Val  
 275 280 285

His Val Ser Lys Ala Asp Gln Met Arg Asp Glu Leu Val Lys Ala Leu  
 290 295 300

Ala Ser Leu Lys Val Gly Pro Gly Ile Asp Pro Ala Ser Asp Ile Gly  
 305 310 315 320

Ala Leu Ile Asp Ala Ala Ser Arg Asp Ala Val Gln Thr Thr Val Glu  
 325 330 335

Arg Ala Cys Asp Leu Ala Asp Arg Val Leu Leu Arg Gly Thr Ser Ser  
 340 345 350

Gly Pro Gly Ala Phe Leu Ser Pro Thr Leu Val Glu His Gly Glu Pro  
 355 360 365

His Ala Phe Phe Cys Gln Asp Glu Ile Phe Gly Pro Phe Val Thr Leu  
 370 375 380

Glu Thr Phe Val Thr Glu Lys Glu Ala Val Glu Lys Ala Asn Asn Thr  
 385 390 395 400

Val Phe Gly Leu Ser Ala Ser Val Trp Thr His Asp Ser Ala Arg Ala  
 405 410 415

Phe Arg Ile Ala Arg Ala Leu Arg Asp Gly Thr Val Trp Ile Asn Asp  
 420 425 430

His Asn Arg Leu Phe Ala Glu Ala Glu Thr Gly Gly Tyr Arg Gln Ser  
 435 440 445

Gly Leu Gly Arg Leu His Gly Tyr Asp Ala Leu Ala Asp Phe Thr Glu



ES 2 817 001 T3

Met Thr Asn Leu Asp Ser Arg His Trp Ile Asp Gly Ala Trp Val Pro  
 1 5 10 15

Gly Thr Asp Arg Phe Ala Ser Ile Asn Pro Ala Asp Gly Ser Val Leu  
 20 25 30

Gly His Ala Ala Asp Gly Gly Arg Ala Glu Ala Glu Ala Ala Ile Ala  
 35 40 45

Ala Ala His Ala Ala Phe Asn Arg Pro Asp Trp Ala Gln Asn Pro Arg  
 50 55 60

Leu Arg Gln Ser Ile Leu Leu Gly Trp Ala Asp Arg Leu Asp Thr Gln  
 65 70 75 80

Ala Glu Asp Leu Ala Arg Leu Leu Thr Leu Glu Asn Gly Lys Ala Ile  
 85 90 95

Ala Gln Ser Arg Gly Glu Ile Ala Gly Ala Ile Ser Glu Ile Arg Tyr  
 100 105 110

Tyr Gly Gly Leu Ala Arg His Val Pro Gly His Val Leu Glu Val Glu  
 115 120 125

Pro Gly Val Leu Ser Thr Met Leu Arg Glu Pro Ala Gly Val Ala Ala  
 130 135 140

Leu Ile Ile Pro Trp Asn Ala Pro Ala Val Leu Leu Ala Arg Ala Ile  
 145 150 155 160

Gly Pro Ala Leu Ala Cys Gly Cys Thr Val Val Val Lys Pro Ala Ala  
 165 170 175

Gln Thr Thr Leu Leu Thr Ala Ala Phe Leu Arg Ala Leu Ser Glu Val

ES 2 817 001 T3

			180					185						190			
Pro	Ser	Leu	Pro	Arg	Gly	Val	Cys	Asn	Met	Ile	Ser	Glu	Thr	Gly	His		
		195					200					205					
Ala	Ala	Ala	Ala	Arg	Leu	Val	Asp	Ser	Pro	Leu	Val	Asp	Val	Val	Ser		
	210					215					220						
Phe	Thr	Gly	Ser	Thr	Ala	Thr	Gly	Lys	Arg	Ile	Met	Val	Ala	Ala	Ala		
225					230					235						240	
Asp	Thr	Met	Lys	Lys	Leu	Ser	Leu	Glu	Leu	Gly	Gly	Lys	Ser	Cys	Cys		
				245					250					255			
Leu	Val	Phe	Pro	Asp	Ala	Asp	Pro	Ala	Glu	Thr	Ala	Ala	Arg	Ile	Ala		
			260					265					270				
Thr	Ala	Ala	Thr	Ile	Ile	Ser	Gly	Gln	Gln	Cys	Thr	Ala	Ala	Arg	Arg		
	275						280					285					
Val	Leu	Val	His	Ala	Ser	Ala	Phe	Asp	Ala	Met	Lys	Thr	His	Leu	Arg		
	290					295					300						
Ala	Ala	Leu	Ala	Ala	Met	Thr	Val	Gly	Asn	Gly	Leu	Asp	Pro	Ala	Ile		
305					310					315					320		
Arg	Met	Gly	Pro	Leu	Ile	Asp	Arg	Pro	Ala	Arg	Asp	Gln	Val	Gln	Thr		
				325					330					335			
Gln	Val	Glu	Arg	Ala	Phe	Asp	Ala	Cys	Asp	Glu	Val	Leu	Leu	Arg	Gly		
			340					345					350				
Gly	Val	Pro	Thr	Asp	Ser	Pro	Ala	Ala	Ala	Ser	Phe	Leu	Thr	Pro	Ser		
		355					360					365					
Leu	Val	Ala	His	Asp	Asp	Pro	Ser	Ala	Phe	Phe	Cys	Gln	Asp	Glu	Ile		
	370					375					380						
Phe	Gly	Pro	Phe	Val	Val	Leu	Glu	Arg	Phe	Glu	Thr	Glu	Ala	Glu	Ala		
385					390					395					400		
Val	Ala	Lys	Ala	Asn	Asn	Thr	Val	Phe	Gly	Leu	Ser	Ala	Ser	Val	Trp		
				405					410					415			
Thr	Arg	Asp	Gly	Ala	Arg	Ala	Leu	Arg	Met	Ala	Arg	Ala	Leu	Arg	Asn		
			420					425					430				

ES 2 817 001 T3

Gly Thr Val Trp Ile Asn Asp His Asn Arg Leu Phe Ala Glu Ala Glu  
435 440 445

Thr Gly Gly Tyr Arg Gln Ser Gly Leu Gly Arg Leu His Gly Tyr Asp  
450 455 460

Ala Phe Ala Asp Phe Thr Glu Leu Lys His Val Cys Gln Thr Val Gly  
465 470 475 480

Thr Ile Gly

<210> 28

<211> 480

5 <212> PRT

<213> Pseudomonas putida

<400> 28

ES 2 817 001 T3

Met Gln Ser Gln His Tyr Ile Asp Gly Gln Trp Thr Ser Thr Asp Arg  
 1 5 10 15

Trp Thr Asp Ser Leu Asp Pro Ala Ser Gly Glu Leu Ile Gly Cys Phe  
 20 25 30

Ala Asp Gly Gly Glu Ala Glu Ala Glu Ala Ala Val Ala Ala Ala Ala  
 35 40 45

Arg Ala Phe Asn Asp Pro Gln Trp Ala Gln Asn Pro Arg Leu Arg Gln  
 50 55 60

Gln Leu Leu Leu Glu Trp Ala Ala Gly Leu Lys Ala Arg Gln Glu Gln  
 65 70 75 80

Leu Ala Gln Leu Leu Thr Arg Glu Asn Gly Lys Ala Leu Ala Gln Ser  
 85 90 95

Arg Gly Glu Ile Gly Gly Ala Ile Ser Glu Ile Leu Tyr Tyr Ala Gly  
 100 105 110

Leu Ala Arg His Asn Pro Gly His Met Leu Glu Val Ala Pro Gly Glu  
 115 120 125

Phe Ser Ser Met Leu Arg Glu Pro Ala Gly Val Ala Gly Leu Ile Ile  
 130 135 140

Pro Trp Asn Ala Pro Ala Val Leu Leu Val Arg Ala Leu Ala Pro Ala  
 145 150 155 160

ES 2 817 001 T3

Ile Ala Ala Gly Cys Thr Val Val Ile Lys Pro Ala Pro Gln Thr Ala  
165 170 175

Leu Phe Asn Ala Ala Met Leu Glu Pro Leu Phe Ala Leu Pro Gly Leu  
180 185 190

Pro Ala Gly Ala Val Asn Leu Phe Ala Glu Ser Gly His Ala Gly Ala  
195 200 205

Ala His Leu Val Ala Ser Pro Arg Val Asp Val Leu Ser Phe Thr Gly  
210 215 220

Ser Thr Ala Thr Gly Gln Arg Ile Met Arg Asp Cys Ala Ala Thr Met  
225 230 235 240

Lys Lys Leu Ser Leu Glu Leu Gly Gly Lys Ser Cys Cys Leu Val Phe  
245 250 255

Glu Asp Ala Asp Ile Ala Ala Ile Ala Pro Lys Leu Ala Ala Ala Ala  
260 265 270

Thr Ile Ile Ser Gly Gln Gln Cys Thr Ala Ala Arg Arg Val Leu Val  
275 280 285

His Ala Ser Arg Phe Ala Glu Met Lys Thr Ala Leu Ser Ala Ala Leu  
290 295 300

Gly Gln Ile Arg Leu Gly Asn Gly Leu Asp Pro Ala Asn Asn Met Gly  
305 310 315 320

Pro Leu Ile Asp Trp His Ser Arg Asp Ser Val Glu Arg Arg Ile Gly  
325 330 335

Glu Ala Leu Asp Ser Cys Asp Glu Val Leu Leu Ala Gly Gly Arg Pro  
340 345 350

Gln Gly Glu Leu Ser Lys Gly Ala Phe Leu Ala Pro Ser Leu Ile Ala  
355 360 365

His Arg Asp Ser Ser Ala Phe Phe Cys Gln Glu Glu Ile Phe Gly Pro  
370 375 380

Leu Leu Val Leu Glu Ser Phe Glu Asp Glu Thr Glu Ala Val Ala Arg  
385 390 395 400

Ala Asn His Thr Glu Phe Gly Leu Ser Ala Ser Val Trp Thr Asp Gln  
405 410 415

ES 2 817 001 T3

Gly Ala Arg Ala Trp Arg Val Ala Arg Ala Leu Arg Asn Gly Thr Val  
420 425 430

Trp Leu Asn Asp His Asn Arg Leu Phe Ala Glu Ala Glu Thr Gly Gly  
435 440 445

Tyr Arg Lys Ser Gly Leu Gly Arg Leu His Gly Val Asp Ala Leu Leu  
450 455 460

Asp Phe Ser Glu Leu Lys His Ile Tyr Gln Asn Val Gly Thr Leu Gly  
465 470 475 480

<210> 29

<211> 486

5 <212> PRT

<213> Rhodopseudomonas palustris

<400> 29

ES 2 817 001 T3

Met Gly Met Thr Ala Leu His Ala Asp Asn Leu Ile Asp Gly Ala Trp  
 1 5 10 15

Gln Pro Ala Gln Ser Gly Ala Thr Ala Pro Ser Leu Asp Pro Ser Ser  
 20 25 30

Gly Gly Thr Ile Gly Gly Phe Ala Ala Gly Gly Ala Ala Asp Ala Gln  
 35 40 45

Ala Ala Val Ala Ala Ala Arg Arg Ala Phe Glu Arg Pro Glu Trp Ser  
 50 55 60

Gln Asn Pro Arg Ala Arg Gln Met Val Met Leu Arg Trp Ala Asp Arg  
 65 70 75 80

Met Glu Ala Gln Ala Asp Gln Leu Ala Arg Leu Leu Thr Leu Glu Asn  
 85 90 95

Gly Lys Pro Leu Pro Gln Ser Arg Gly Glu Ile Ala Gly Ser Val Ser  
 100 105 110

Glu Ile Arg Tyr Tyr Ala Gly Leu Thr Arg Tyr Ile Pro Gly His Val  
 115 120 125

Phe Glu Val Glu Pro Gly Ser Phe Ser Thr Leu Leu Lys Glu Pro Ala  
 130 135 140

Gly Val Ala Gly Leu Ile Ile Pro Trp Asn Ala Pro Ala Val Leu Leu  
 145 150 155 160

ES 2 817 001 T3

Ile Arg Ala Leu Thr Pro Ala Leu Ala Ala Gly Cys Thr Val Val Ile  
 165 170 175

Lys Pro Ala Pro Gln Thr Ala Gln Ile Thr Ala Ala Ile Ile Lys Cys  
 180 185 190

Leu His Glu Val Asp Gly Leu Pro Arg Gly Val Val Asn Leu Val Ser  
 195 200 205

Glu Gln Gly His Gln Val Ala Glu His Leu Val Thr Ser Asn Asp Val  
 210 215 220

Asp Val Ile Ser Phe Thr Gly Ser Asn Ala Thr Gly Ala Arg Ile Met  
 225 230 235 240

Ala Ala Ala Ala Pro Thr Met Lys Lys Leu Ser Leu Glu Leu Gly Gly  
 245 250 255

Lys Ser Ala Cys Leu Val Phe Asp Asp Ala Asp Ile Ala Asp Val Ala  
 260 265 270

Pro Lys Leu Ala Ala Ala Ala Thr Ile Ile Ala Gly Gln Gln Cys Thr  
 275 280 285

Ala Ala Arg Arg Val Leu Val His Ala Ser Arg Tyr Asp Glu Met Lys  
 290 295 300

Ala Ala Leu Lys Ala Ala Leu Ala Asn Ile Arg Ile Ala Pro Gly Ser  
 305 310 315 320

Ala Ala Gly Ala Glu Met Gly Pro Leu Ile Asp Ala Ala Ser Leu Ala  
 325 330 335

Ala Val Ala Lys Arg Ala Asp Glu Ala Met Gln Ala Ala Asp Glu Val  
 340 345 350

Val Leu Arg Gly Gly Arg Pro Ala Gly Asp Leu Ala Asn Gly Tyr Phe  
 355 360 365

Leu Ser Pro Thr Leu Val Ala His Arg Asp Thr Ser Ala Phe Phe Val  
 370 375 380

Gln Glu Glu Ile Phe Gly Pro Leu Val Val Leu Glu Lys Phe Glu Asp  
 385 390 395 400

Glu Lys Glu Ala Val Ala Arg Ala Asn His Ser Asp Tyr Gly Leu Ser  
 405 410 415

ES 2 817 001 T3

Ala Ser Val Trp Thr His Asp Gly Ala Arg Ala Met Arg Val Ala Arg  
420 425 430

Ala Leu Arg Asn Gly Thr Val Trp Ile Asn Asp His Asn Lys Leu Phe  
435 440 445

Ala Glu Ala Glu Thr Gly Gly Tyr Arg Arg Ser Gly Leu Gly Arg Leu  
450 455 460

His Gly Tyr Asp Ala Leu Ile Asp Phe Leu Glu Ile Lys His Val Tyr  
465 470 475 480

Gln Ser Cys Gly Val Val  
485

<210> 30

<211> 485

5 <212> PRT

<213> Dinoroseobacter shibae DFL 12

<400> 30

ES 2 817 001 T3

Met Thr Thr Thr Asp Leu Ile Ala Arg His Met Ile Gly Gly Ser Tyr  
 1 5 10 15

Ser Asp Ala Gly Asp Lys Ile Ala Ser Ile Asn Pro Ala Thr Gly Ala  
 20 25 30

Val Val Gly His Val Arg Ala Asp Gly Ala Ala Gln Ala Thr Ala Ala  
 35 40 45

Ile Ala Ala Ala Arg Ala Ala Phe Asp Thr Thr Leu Trp Pro Gln Ser  
 50 55 60

Pro Arg Asp Arg Gln Met Ala Leu Leu Arg Trp Ala Asp Ala Leu Glu  
 65 70 75 80

Ala Asp Leu Ala Arg Leu Ala Glu Leu Leu Thr Leu Thr Asn Gly Lys  
 85 90 95

Pro Leu Gly Ala Ser Lys Gly Glu Leu Gly Ala Ala Ile Ser Glu Ile  
 100 105 110

Arg Tyr Tyr Ala Gly Leu Thr Arg His Asn Pro Gly His Ala Met Glu  
 115 120 125

Val Ala Pro Gly Glu Leu Ser Val Met Leu Arg Glu Pro Ala Gly Val  
 130 135 140

ES 2 817 001 T3

Ala Gly Ile Ile Val Pro Trp Asn Ala Pro Ala Val Leu Leu Ile Arg  
145 150 155 160

Ser Leu Ala Pro Ala Leu Ala Val Gly Cys Thr Thr Val Thr Lys Pro  
165 170 175

Ala Pro Gln Thr Ala Leu Phe Thr Ala Ala Cys Met Ala Pro Leu Phe  
180 185 190

Glu Asp Ala Ala Ile Pro Ala Gly Val Val Asn Val Val Phe Glu Val  
195 200 205

Gly His Asp Ala Ala Gln Thr Leu Val Thr Ser Pro Asp Val Asp Val  
210 215 220

Ile Ser Phe Thr Gly Ser Asn Ala Val Gly Gln Arg Ile Met Ala Asp  
225 230 235 240

Ala Ala Pro Thr Met Lys Lys Leu Ser Leu Glu Leu Gly Gly Lys Ser  
245 250 255

Cys Cys Ile Val Leu Asp Asp Ala Asp Ile Gly Val Val Ala Pro Lys  
260 265 270

Leu Ala Ala Ala Ala Thr Ile Ile Ser Gly Gln Gln Cys Thr Ala Ala  
275 280 285

Arg Arg Val Leu Val His Glu Ser Arg Leu Asp Glu Ala Lys Ser Ala  
290 295 300

Leu Ser Ala Ala Leu Gln Ala Val Ser Ile Gly Asp Gly Met Ser Asp  
305 310 315 320

Gly Thr Ala Met Gly Pro Leu Ile Asp Ile Gln Ser Arg Asp Arg Val  
325 330 335

Met Arg Asp Cys Gly Thr Val Tyr Asp Thr Ala Asp Glu Val Val Leu  
340 345 350

Arg Gly Gly Pro Leu Asp Gly Pro Lys Gly Ser Ala Phe Met Ser Pro  
355 360 365

Ala Leu Val Val His Ser Asp Pro Asn Ala Ser Phe Val Gln Asp Glu  
370 375 380

Ile Phe Gly Pro Leu Val Val Leu Glu Thr Phe Arg Asp Glu Ala Asp

ES 2 817 001 T3

385					390						395					400
Ala	Val	Ala	Lys	Ala	Asn	Asn	Thr	Val	Tyr	Gly	Leu	Ser	Ala	Ser	Ile	
				405					410					415		
Trp	Thr	His	Arg	Gly	Asp	Ala	Ser	Trp	Arg	Leu	Ala	Arg	Ala	Leu	Arg	
			420					425					430			
Asn	Gly	Thr	Val	Trp	Ile	Asn	Asp	His	Asn	Arg	Leu	Phe	Ala	Glu	Ala	
		435					440					445				
Glu	Thr	Gly	Gly	Tyr	Arg	Arg	Ser	Gly	Leu	Gly	Arg	Leu	His	Gly	Phe	
	450					455					460					
Asp	Gly	Leu	Leu	Asp	Phe	Cys	Glu	Leu	Lys	His	Val	Tyr	Gln	Asn	Val	
465					470					475					480	
Gly	Val	Val	Gly	His												
				485												

<210> 31

<211> 447

<212> PRT

5 <213> Cupriavidus basilensis

<400> 31

ES 2 817 001 T3

Met Glu Ala Val Ala Lys Lys Arg Thr Glu Thr Ile Ser Glu Ala Leu  
 1 5 10 15

Pro Ala Ala Thr Asn Arg Gln Val Phe Gly Ala Val Thr Ala Ser Cys  
 20 25 30

Met Gly Trp Ala Leu Asp Leu Phe Asp Leu Phe Ile Leu Leu Phe Val  
 35 40 45

Ala Pro Val Ile Gly Arg Leu Phe Phe Pro Ser Glu His Ala Met Leu  
 50 55 60

Ser Leu Ala Ala Val Tyr Ala Ser Phe Ala Val Thr Leu Leu Met Arg  
 65 70 75 80

Pro Leu Gly Ser Ala Ile Phe Gly Thr Tyr Ala Asp Arg His Gly Arg  
 85 90 95

Lys Gly Ala Met Val Val Ala Val Thr Gly Val Gly Leu Ser Thr Ala  
 100 105 110

Ala Phe Gly Leu Leu Pro Thr Val Gly Gln Val Gly Leu Leu Ala Pro

ES 2 817 001 T3

115	120	125																
Ala	Leu	Phe	Ile	Leu	Leu	Arg	Leu	Val	Gln	Gly	Ile	Phe	Val	Gly	Gly			
130						135						140						
Val	Val	Ala	Ser	Thr	His	Thr	Ile	Gly	Thr	Glu	Ser	Val	Pro	Pro	Ser			
145					150						155				160			
Trp	Arg	Gly	Ala	Val	Ser	Gly	Leu	Val	Gly	Gly	Gly	Gly	Ala	Gly	Ile			
			165						170				175					
Gly	Ala	Leu	Leu	Ala	Ser	Ile	Thr	Tyr	Met	Ala	Met	Thr	Ala	Leu	Phe			
			180						185				190					
Pro	Gly	Glu	Ala	Phe	Asp	Ala	Trp	Gly	Trp	Arg	Cys	Met	Phe	Phe	Ser			
		195						200				205						
Gly	Ile	Ile	Ser	Ser	Val	Leu	Gly	Leu	Phe	Ile	Phe	Asn	Ser	Leu	Glu			
		210						215				220						
Glu	Ser	Pro	Leu	Trp	Lys	Gln	Leu	Gln	Ala	Ala	Lys	Gly	His	Ala	Ala			
225					230						235				240			
Pro	Val	Glu	Asn	Pro	Leu	Arg	Val	Ile	Phe	Ser	Arg	Gln	Tyr	Arg	Gly			
			245						250				255					
Val	Leu	Phe	Val	Asn	Ile	Leu	Leu	Thr	Val	Gly	Gly	Gly	Ser	Ala	Tyr			
			260						265				270					
Tyr	Leu	Thr	Ser	Gly	Tyr	Leu	Pro	Thr	Phe	Leu	Lys	Val	Val	Val	Lys			
		275						280				285						
Ala	Pro	Ala	Gly	Ala	Ser	Ala	Ala	Ile	Leu	Met	Ala	Ser	Ser	Val	Gly			
		290						295				300						
Val	Ile	Val	Ala	Ser	Ile	Ile	Ala	Gly	His	Leu	Ser	Thr	Leu	Ile	Gly			
305					310						315				320			
Arg	Lys	Arg	Ala	Phe	Leu	Leu	Ile	Gly	Ala	Leu	Asn	Val	Val	Leu	Leu			
			325						330				335					
Pro	Leu	Ile	Tyr	Gln	Arg	Met	Pro	Ala	Ala	Pro	Asp	Val	Thr	Thr	Leu			
			340						345				350					
Gly	Ile	Tyr	Ala	Val	Ala	Leu	Ala	Met	Leu	Gly	Ser	Thr	Gly	Phe	Ala			
		355						360				365						

ES 2 817 001 T3

Pro Ile Leu Ile Phe Leu Asn Glu Arg Phe Pro Thr Ser Ile Arg Ala  
370 375 380

Thr Gly Thr Gly Leu Ser Trp Asn Ile Gly Phe Ala Ile Gly Gly Met  
385 390 395 400

Met Pro Thr Phe Ala Ser Leu Cys Ala Ser Thr Pro Ala Asp Leu Pro  
405 410 415

Lys Val Leu Gly Ile Phe Val Ala Val Val Thr Ala Ile Tyr Leu Ala  
420 425 430

Gly Ala Ala Ile Val Pro Glu Thr Ala Gly Arg Leu Gly Glu Lys  
435 440 445

<210> 32

<211> 449

<212> **PRT**

5 <213> Cupriavidus basilensis

<400> 32

ES 2 817 001 T3

Met Glu Ala Val Ala Lys Lys Ser Ala Ala Thr Ile Ser Glu Ala Leu  
 1 5 10 15

Pro Ala Ala Ser Asn Arg Gln Val Phe Gly Ala Val Ala Ala Ser Cys  
 20 25 30

Met Gly Trp Ala Leu Asp Leu Phe Asp Leu Phe Ile Leu Leu Phe Val  
 35 40 45

Ala Pro Val Ile Gly Arg Leu Phe Phe Pro Ser Glu His Ala Met Leu  
 50 55 60

Ser Leu Ala Ala Val Tyr Ala Ser Phe Ala Val Thr Leu Leu Met Arg  
 65 70 75 80

Pro Leu Gly Ser Ala Ile Phe Gly Ser Tyr Ala Asp Arg His Gly Arg  
 85 90 95

Lys Gly Ala Met Val Val Ala Val Thr Gly Val Gly Leu Ser Thr Ala  
 100 105 110

Ala Phe Gly Leu Leu Pro Thr Val Gly Gln Val Gly Leu Leu Ala Pro  
 115 120 125

Ala Leu Phe Ile Leu Leu Arg Leu Val Gln Gly Ile Phe Val Gly Gly  
 130 135 140

ES 2 817 001 T3

Val Val Ala Ser Thr His Thr Ile Gly Thr Glu Ser Val Pro Pro Ser  
145 150 155 160

Trp Arg Gly Ala Val Ser Gly Leu Val Gly Gly Gly Gly Ala Gly Leu  
165 170 175

Gly Ala Leu Leu Ala Ser Ile Thr Tyr Met Ala Met Thr Ala Leu Phe  
180 185 190

Pro Gly Glu Ala Phe Asp Ala Trp Gly Trp Arg Cys Met Phe Phe Ser  
195 200 205

Gly Ile Ile Ser Ser Val Leu Gly Leu Phe Ile Phe Asn Ser Leu Glu  
210 215 220

Glu Ser Pro Leu Trp Lys Gln Leu Gln Ala Ala Lys Gly His Ala Ala  
225 230 235 240

Pro Val Glu Asn Pro Leu Arg Val Ile Phe Ser Arg Gln Tyr Arg Gly  
245 250 255

Val Leu Phe Val Asn Ile Leu Leu Thr Val Gly Gly Gly Ser Ala Tyr  
260 265 270

Tyr Leu Thr Ser Gly Tyr Leu Pro Thr Phe Leu Lys Val Val Val Lys  
275 280 285

Ala Ser Ala Gly Glu Ser Ala Ala Ile Leu Met Ala Ser Ser Leu Gly  
290 295 300

Val Ile Val Ala Ser Ile Leu Ala Gly His Leu Ser Thr Met Ile Gly  
305 310 315 320

Arg Lys Arg Ala Phe Leu Leu Ile Gly Ala Leu Asn Val Val Val Leu  
325 330 335

Pro Leu Leu Tyr Gln Trp Met Pro Ala Ala Pro Asp Thr Thr Thr Leu  
340 345 350

Gly Leu Tyr Ala Val Val Leu Ser Met Leu Gly Cys Ser Gly Phe Ala  
355 360 365

Pro Ile Leu Ile Phe Leu Asn Glu Arg Phe Pro Thr Ser Ile Arg Ala  
370 375 380

Thr Gly Thr Gly Leu Ser Trp Asn Ile Gly Phe Ala Val Gly Gly Met  
385 390 395 400

ES 2 817 001 T3

Met Pro Thr Phe Ala Ser Leu Cys Ala Ser Thr Pro Ala Glu Leu Pro  
405 410 415

Met Val Leu Gly Ile Phe Leu Ala Val Val Thr Ile Ile Tyr Leu Val  
420 425 430

Gly Ala Phe Ile Val Pro Glu Thr Val Gly Arg Leu Gly Asp Asn Gly  
435 440 445

Ala

<210> 33

<211> 449

5 <212> PRT

<213> Methylobacterium radiotolerans

<400> 33

ES 2 817 001 T3

Met Gln Thr Ala Ala Thr Phe Ala Ser Asp Pro Pro Ala Leu Ala Lys  
 1 5 10 15

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

Trp Gly Leu Asp Leu Phe Asp Leu Phe Ile Leu Leu Tyr Val Ala Pro  
 35 40 45

Val Val Gly Thr Leu Phe Phe Pro Ala Asp Lys Pro Met Leu Ser Leu  
 50 55 60

Ala Gly Ala Tyr Ala Ser Phe Ala Val Thr Leu Leu Ile Arg Pro Leu  
 65 70 75 80

Gly Ser Ala Leu Phe Gly Ser Tyr Ala Asp Arg Phe Gly Arg Arg Arg  
 85 90 95

Ala Leu Met Val Ala Val Val Gly Val Gly Ile Ser Thr Ala Val Phe  
 100 105 110

Gly Leu Leu Pro Thr Val Gly Gln Ile Gly Trp Leu Ala Thr Ala Val  
 115 120 125

Phe Leu Phe Phe Arg Leu Val Gln Gly Ile Phe Val Gly Gly Val Val  
 130 135 140

Ala Ala Ser His Thr Ile Gly Thr Glu Ser Val Pro Glu Arg Trp Arg  
 145 150 155 160

ES 2 817 001 T3

Gly Leu Met Ser Gly Ala Val Gly Gly Gly Gly Ser Ala Ile Gly Gly  
165 170 175

Leu Leu Ala Ser Leu Val Phe Tyr Val Val Ser Leu Met Ala Pro Gly  
180 185 190

Glu Ala Phe Ala Glu Trp Gly Trp Arg Leu Met Phe Phe Ser Gly Leu  
195 200 205

Leu Thr Ser Val Ile Gly Leu Ile Leu Phe Arg Asn Leu Glu Glu Ser  
210 215 220

Pro Ile Phe Lys Glu Leu Gln Ala Arg Lys Ala Ala Leu Arg Ala Gly  
225 230 235 240

Ala Pro Ala Glu Ala Ser Pro Ile Arg Ser Leu Phe Ser Pro Ser Asn  
245 250 255

Arg Gly Ser Phe Ala Val Ala Thr Leu Ile Ser Phe Gly Gly Gly Ala  
260 265 270

Ala Tyr Tyr Leu Thr Ser Gly Tyr Leu Pro Thr Leu Leu Lys Leu Val  
275 280 285

Asn Gly Val Pro Asn Ala Thr Ala Ser Met Ile Leu Ile Gly Ala Asn  
290 295 300

Val Ala Ala Ala Ile Gly Ala Cys Gly Met Gly Glu Leu Ser Gln His  
305 310 315 320

Ile Gly Arg Lys Arg Ser Phe Leu Leu Met Gly Val Ile Arg Leu Leu  
325 330 335

Ala Phe Pro Ala Leu Phe Leu Thr Met Ala Asn Thr Thr Ser Leu Val  
340 345 350

Gly Val Ala Ala Cys Ala Phe Leu Leu Ala Leu Ile Ala Asn Gly Ser  
355 360 365

Tyr Gly Pro Leu Leu Ile Phe Leu Asn Glu Lys Phe Pro Thr Ala Val  
370 375 380

Arg Ala Thr Gly Thr Gly Leu Thr Trp Asn Ile Gly Phe Ala Leu Gly  
385 390 395 400

Gly Met Leu Pro Thr Leu Val Ser Leu Val Ala Asp Gly Pro Thr Gln  
405 410 415

ES 2 817 001 T3

Ile Pro Met Val Leu Ala Val Ile Thr Thr Gly Val Thr Leu Val Tyr  
420 425 430

Leu Val Gly Ala Phe Leu Thr Asp Glu Thr Gln Gly Asn Leu Asp Arg  
435 440 445

Ala

<210> 34

<211> 443

5 <212> PRT

<213> Sulfolobus acidocaldarius

<400> 34

ES 2 817 001 T3

Met Lys Lys Glu Glu Lys Phe Thr Ser Asn His Phe Lys Trp Thr Leu  
1 5 10 15

Ala Thr Phe Phe Thr Trp Thr Phe Asp Leu Tyr Asp Leu Phe Thr Ile  
20 25 30

Leu Leu Val Ala Pro Tyr Ile Ser Ser Leu Phe Phe Pro Ser Ser Ile  
35 40 45

Thr Phe Leu Ser Ile Ala Ala Thr Tyr Ala Gly Phe Ala Thr Ser Leu  
50 55 60

Ile Met Arg Pro Val Gly Ala Thr Val Phe Gly Ser Arg Val Ser Asp  
65 70 75 80

Lys Val Gly Arg Lys Arg Ala Ile Phe Tyr Gly Leu Ile Gly Leu Val  
85 90 95

Ile Thr Ser Thr Leu Gln Gly Ala Leu Pro Thr Tyr Gln Val Val Gly  
100 105 110

Val Ile Ala Pro Ile Leu Leu Leu Ala Val Arg Leu Ile Gln Gly Val  
115 120 125

Phe Ile Gly Gly Ile Thr Ala Gly Ser His Val Ile Gly Pro Glu Ser  
130 135 140

Val Pro Glu Arg Tyr Arg Gly Ile Val Gly Gly Leu Gly Phe Ser Ala  
145 150 155 160

Ala Gly Val Ala Tyr Leu Ile Ala Ala Gly Trp Phe Phe Leu Thr Thr  
165 170 175

ES 2 817 001 T3

Ile Leu Tyr Pro Gly Ser Ser Tyr Leu Val Trp Gly Trp Arg Val Met  
180 185 190

Phe Phe Gly Gly Leu Leu Ser Leu Ala Val Leu Gly Phe Val Asn Tyr  
195 200 205

Leu Val Pro Glu Ser Glu Val Trp Thr Lys Ile Lys Lys Arg Gly Ser  
210 215 220

Val Val Lys Ser Pro Leu Lys Glu Ile Phe Ser Lys Tyr Arg Tyr Gln  
225 230 235 240

Leu Gly Val Ala Leu Leu Leu Ser Ile Gly Trp Gly Ala Ser Phe Tyr  
245 250 255

Val Thr Asp Gly Ile Leu Pro Thr Phe Leu Ser Ser Val Asn Lys Leu  
260 265 270

Ala Lys Thr Glu Ile Ala Ile Val Met Ile Ile Gly Ser Ile Gly Met  
275 280 285

Ser Ile Gly Pro Leu Ile Gly Gly Glu Ile Ser Gln Ile Ile Gly Arg  
290 295 300

Lys Ile Thr Ser Leu Ile Gly Ala Ile Ile Val Leu Ala Val Val Gly  
305 310 315 320

Pro Leu Phe Leu Ser Leu Gly Ser Leu Lys Ser Gly Asp Leu Asn Gln  
325 330 335

Ile Ile Leu His Ser Phe Ala Ile Leu Phe Leu Val Asp Ile Gly Gly  
340 345 350

Gly Met Leu Met Thr Tyr Leu Asn Glu Ile Tyr Pro Ala Ser Val Arg  
355 360 365

Gly Thr Gly Val Gly Phe Thr Trp Asn Thr Gly Phe Ala Ile Gly Gly  
370 375 380

Thr Ile Pro Thr Ile Ile Ser Leu Ala Val Ala Ser Ala Gly Leu Ser  
385 390 395 400

Ala Phe Pro Ser Ile Met Phe Tyr Thr Leu Ile Val Val Ser Val Ile  
405 410 415

Ile Leu Val Gly Thr Val Leu Thr Lys Glu Thr Lys Gly Thr Ile Ser

ES 2 817 001 T3

420

425

430

Lys Glu Glu Tyr Glu Ile Gln Lys Glu Thr Leu  
 435 440

<210> 35  
 <211> 1240  
 <212> ADN  
 <213> Artificial

5

<220>  
 <223> Fragmento de KpnI-NheI sintético para la expresión de alcohol deshidrogenasa de Aneurinibacillus terranovensis

<400> 35

10

```

ggtagcgaat tccacatgac aaggggagac cgcacgacca ttagtcgggc agttaaagcc      60
atcaactttg aatthtcatt taacctgccg accctgatcg aatttggtta tggtaaaatg      120
gaaaaattcg gccagcagct gattagcatt ggtgttaaac gcaththtat ggtgaccgat      180
aaaggtgttg aaagcgcagg tctgctggca gcaactgaccg attcactgca ggcagcagca      240
attcagtttg atatctatac cgatgtggaa agcgcaccca gcctggaaac cattgatcgt      300
ggtgttgaag ththtcagca gaaaccgat gattgcattg ttgcagttgg tgggtgtagc      360
ccgattgata ccgcaaaagg tattcgtgtt gttgcagcaa atggtggtaa tattggtcat      420
tatgccggtg ttaatcagat tccggttga cccgaccattc cgctgctggc aattccgacc      480
accagtggca ccggtagcga agttaccaat ththgtgttt atagcgattg gcagaacaac      540
gttaaagtta ccgttaccag ccagtatatg gcaccgacaa ttgcatgggt tgatccggca      600
ctgaccatga gcctgcctgc aaaaatgacc gcagcaagcg gtattgatgc actggcacat      660
ggtattgaaa cthththtag cthgggtagc agtccggcaa gtgatgccct ggcaattgaa      720
gcaattcata ccgttaatcg ttatctgagc cgtgcagttc ataatggtag cgatatggaa      780
gcacgtattg gtatgagcca tggtagcctg ctggctggca tggcatttaa caatggtttt      840
ctgggtctgg cccatgccat tggtagcga ctgagcggtc attgtcatgt tccgcatggt      900
gttgcaattg gtctgctgct gccgaaagt gttgaattta atgcaaccgt tcgtccggat      960
aaagcagcaa aaattgcagg tctgatgggt atgaaaggtg aacatagcga agaactggcc     1020
ctgcaggcat caccggcagt tgcacgtctg gttgaagata ttggcctgcc gacacgtctg     1080
cgtgaagttg atgttaccga aaaaaactg thcgagatcg ccaaagatag cththaaagc     1140
ggcatgatga aattcaatcc gcgtcagccg agcgaagcg aagttctgca gctgctgaaa     1200
gaaatcttht gaagaccgaa gcgaattcct cgagtctaga                                1240
    
```

<210> 36  
 <211> 1205

ES 2 817 001 T3

<212> ADN  
<213> Artificial

<220>  
<223> Fragmento de KpnI-NheI sintético para la expresión de alcohol deshidrogenasa de Brevibacillus  
thermoruber

5

<400> 36

```

ggtacctagg aaaggaagat taacccatga gccagaccgt gcagggcacc gacttcgcct      60
tcagcttcca cctgccgacc ctgatcgagt tcggctacgg ccgcgccagc cgcctgggcg      120
agcgcctgca gcacctgggc gtgaccaacg tgttcgtggt gaccgacaag ggcgtggagg      180
ccgccggcct gctgaacggc ctggtgggca gcctgcagag cgcgggcatc gccttcgacc      240
tgtacaccga ggtggagccg gacccgggcc tggagaccat cgaccgcggc gccgccgtgt      300
tccgcgccaa gccgtacgac tgcctggtgg ccgtgggchg cggcagcccg atcgacgccg      360
ccaagggcat gcgcgtggtg accagctgcg gcggcagcat cgcggactac gccggcgtga      420
accgcgtgcc gatggccccg gccgtgccgc tgggtggccgt gccgaccacc agcggcaccg      480
gcagcgaggt gaccatgttc ggcgtgtaca gcgactggca caaccacgtg aagtgaccg      540
tgaccagccc gcacatggcc ccgaccatcg ccctggtgga cccggccctg accgtgagcc      600
tgccggccaa gatgaccgcc gccagcggca tcgacgccct ggcccacggc atcgagacct      660
tcttcagcgt gcgcagccgc ccggccagcg acgccctggc catggaggcc atcgccgccg      720
tgaacgcccc cctgcgccgc gccgtgcacg acggcagcga cgtggaggcc cgcatcggca      780
tgagccacgg cagcctgctg gccggcatgg ccttcaccaa cggcttcctg ggcttgcccc      840
acgccatcgg cagcgccttg agcggccact gccacgtgcc gcacggcatc gccatcggcc      900
tgctgctgcc gcacgtggtg gccttcaacg ccccggcccc cccggacaag gccgccagc      960
tggcccgcct gctgggcgtg gaggccaacc cgcgcgagga gcgcggcgag gagaccagcg     1020
ccgccgtggc ccgcatggtg gccgacatcg gcctgccgac ccgcctgcgc gacgtgggcg     1080
tgccggagga gaagctgccg gccatcgcca aggacgcctt caagagcggc atgatgacct     1140
gcaaccocgc ccagccgacc gagcaggagg tgcgcgagct gctgcgccgc gccttctgag     1200
ctagc                                             1205

```

<210> 37  
<211> 1228  
<212> ADN  
<213> Artificial

10

<220>  
<223> Fragmento de KpnI-NheI sintético para la expresión de alcohol deshidrogenasa de Brevibacillus  
panacihumi

15

<400> 37

ES 2 817 001 T3

ggtaccgaat tccacatgac aaggggagac cgcagagcg caaatcagag cgttcaggg 60  
 attgaaagcc cgtttagctt tcatctgccg accaatgttc agtttgggtg tggtagcgca 120  
 agccgtctgg gtgaaatgct gctgagcatg ggtgttcgtc gtgttttct ggttaccgat 180  
 caggggtgtgc gtcaggcagg tctgctggat gaagttattc atagcctgga agaaaaaggc 240  
 ctgcactttc agatttatgc agatgttgaa ccggatccga gcctggaaac cattcaggca 300  
 ggcgagcaaa tgtttcagca gcagagcttt gattgtatgg ttgcaattgg tgggtggtagt 360  
 ccgattgata ccgcaaaagg tattcgtgtt ctggcagcaa atggtggcgg tattggtcag 420  
 tatgccggtg ttaatcgcgt tccggcagca agcgcattc cgctgattgc aattccgacc 480  
 accagtggca ccggtagcga agttaccatt tttgggtgtt atagcgattg ggagaaccac 540  
 gtgaaaatta ccgttaccag tccgcatatg gcaccgagca ccgcaactgat tgatccggca 600  
 ctgaccctga gcctgcctgc aaaaatgacc gcagcaaccg gtattgatgc actggcacat 660  
 ggcattgaaa ccttttttag cctgcgtagc agtccggcaa gtgatgccct ggcaattcat 720  
 gcaatgaaaa tgattgcacc gcatctgcgt cgtgcagttc gtgatggtgc agatatggaa 780  
 gcacgtattg gtatgagcca gggtagcgtg ctggcaggta tggcatttaa caatggtttt 840  
 ctgggtctgg cccatgccat tggtagtgca ctgagcggtc attgtcatgt tccgcatggt 900  
 gttgcgattg gcctgctgct gccgcatgtg gttgcattta atacaccggt tcgtccggaa 960  
 aaagcagaac tgattgccga tgttctgggt agcgttcaga aagaaaccgg caccgcagcc 1020  
 gaactggttg gtcagctggt tcaggatatt ggtctgccgc agcgtctgca agaagttggc 1080  
 gttccggaag cgaaactggt tgatattgca aaagatagct ttaaaagcgg catgatgaaa 1140  
 tggaatccgc gtctgccgac agaacaagaa gttctggaac tgctgcagaa agccttttga 1200  
 agaccgaagc gaattcctcg agtctaga 1228

- 5 <210> 38
- <211> 1184
- <212> ADN
- <213> Artificial

- <220>
- 10 <223> Fragmento de KpnI-NheI sintético para la expresión de alcohol deshidrogenasa de Bacillus sp. FJAT-14578

<400> 38

ES 2 817 001 T3

ggtacctagg aaaggaagat taacctatgt acccgagctt cgagttccac ctgccgacca 60  
 agatccactt cggctacaac accatcaagc agctggacca cctgccgttc gagatcaagc 120  
 gcgccttcat cgtgaccgac cagggcgtgc tgaacagcgg cctggtggag aacgtgacca 180  
 acatcctgaa ggaccaccag atcagctacg tgatctacag cgaggtggag ccggacccga 240  
 gcgtggagac cgtggacaag gccgcccaga tgttccagcg cgaggaggcc gacgccctga 300  
 tcgccatcgg cggcggcagc ccgatcgaca ccgccaaggg cgtgcgcgtg atcgccggca 360  
 acggcggcag catccgcgac tacgccggcg tgaacctgat caagcagaag agcaacatcc 420  
 cgctgatcgc catcccgacc accagcggca ccggcagcga ggtgaccatc ttcgccgtgt 480  
 tcagcgactg ggaggagaac cgcaagggtga ccgtgaccag cccgttcctg gccccggaca 540  
 tcagcatcgt ggaccogaag atgaccatga ccgccccgcc ggccatcacc gccgccagcg 600  
 gcttcgacgc cttcgcccac ggcgcccaga ccttcgtgag ccgcgccagc cagccggcca 660  
 gcgacgtgct ggccttcagc gccatgagca ccgtgagcaa gtacctgcgc cgcgcctgtgt 720  
 acaacggcga ggacgtggag gcccgcatca agatggccga ggccagcctg ctggccggca 780  
 tggccttcaa ccagagctac ctgggcctga cccacgccat cggcagcggc ctgagcggcc 840  
 acgcccacgt gagccacggc gtggccatcg gcctgctgct gccgggcgtg atccgctaca 900  
 acagcatcag ccgcatggac aagcacatcg agatggccgg cgccttcgcg gagatcgacc 960  
 gcagcctgag cgactgggag atcatcgacc agctgatcga ggacgtgagc cgcctgcgcg 1020  
 acgacatcgg cctgccgagc cgcctgcagc aggtgggcgt gaaggaggac cagctgaaga 1080  
 tgatcgccgc cgacagcgtg aagagcggca tgtggaagtt caaccgcgcg caggccagcg 1140  
 aggaggagat cctggagctg ctgaaggagc tgtactgagc tagc 1184

- 5 <210> 39
- <211> 1184
- <212> ADN
- <213> Artificial

- <220>
- 10 <223> Fragmento de KpnI-NheI sintético para la expresión de alcohol deshidrogenasa de Bacillus sp. L1(2012)

<400> 39

ES 2 817 001 T3

ggtacctagg aaaggaagat taacccatgt acaccagctt caacttccac ctgccgaccc 60  
 gcatccagtt cggctacgag aaggtgaagg agctgaagaa cctgccgttc caggccaacc 120  
 gcgccttcat cgtgaccgac aagggcgtgg agaaggccgg cctgctgaac gacgtgatcg 180  
 acgccatcaa gcaggccaac atgacctaca agatctaccg cgacgtggag ccggacccga 240  
 gcgtggagac cgtggacaag gccgccaagg ccttcgccga ggccgagtgc gacctgctga 300  
 tcgccgtggg cggcggcagc ccgatcgaca ccgccaaggg cgtgcgcgtg gtggccagca 360  
 acggcggcag catccgcaac tacagcggcg tgaacctggt gaaggaggcc ccgagcgtgc 420  
 cgctggtggc catcccgacc accgccggca ccggcagcga ggtgaccatc ttcgccgtgt 480  
 tcagcgacga caaggagaac cgcaaggtga ccgtgaccag cagccacctg agcccggacg 540  
 tgagcatcat cgacccgaag ctgaccctga ccgccccgcc gagcatcacc gccgccgccg 600  
 gcttcgacgc cttcgcccac gccgccgagg ccttcgtgag ccgcatcagc cagccgccga 660  
 gcgacgcctt ggccctgagc gccatgaaga ccgtgcacac ctacctgcgc cgcgccgtgt 720  
 acaacggcga cgacatcgag gcccgcatga agatggccga ggccagcctg ctggccggca 780  
 tggccttcaa ccagagctac ctgggcctgg cccacgccat cggcagcgcc atcagcgtgc 840  
 acgcccacgt gagccacggc gtggtgatcg gcctgctgct gccgaaggtg atcgagtaca 900  
 acctggtggc caagatcgac aagtacgccg aggccggcaa gtacatcgag cagagcagcc 960  
 acggcctgag caactacgag gccgccgccc tgttcagcga gaccgtgacc cagctgcgca 1020  
 acgacatcgg cctgccgaag cagctgcgcg aggtgaacgt gaaggaggcc cagctggagg 1080  
 ccatcagcaa ggacagcatc aagagcggca tgtggcagtt caaccgcgc cgcgccagcg 1140  
 agcaggacgt gtaccagatg ctgcgcgaga tgctgtgagc tagc 1184

5 <210> 40  
 <211> 27  
 <212> ADN  
 <213> Artificial

<220>  
 10 <223> sitio de unión al ribosoma que contiene el espaciador n.º 2

<400> 40  
 gaattccaca tgacaagggg agaccgc 27

## REIVINDICACIONES

1. Proceso para oxidar ácido 5-hidroximetil-2-furancarboxílico (HMFCa) a ácido 5-formil-2-furoico (FFA), donde el proceso comprende la etapa de incubar una célula en presencia de HMFCa, preferiblemente en condiciones propicias para la oxidación de HMFCa por la célula, donde la célula comprende un constructo de expresión para la expresión de una secuencia de nucleótidos que codifica una HMFCa deshidrogenasa que tiene una secuencia de aminoácidos con al menos un 45 % de identidad con cualquiera de las secuencias de aminoácidos de la SEC ID N.º: 11, y donde el constructo de expresión se puede expresar en la célula y la expresión de la deshidrogenasa confiere o aumenta en la célula la capacidad de oxidar HMFCa a FFA, en comparación con una célula de tipo salvaje correspondiente que carece del constructo de expresión.
2. Proceso para producir FDCA, que comprende el paso de incubar una célula en un medio que comprende uno o más precursores furánicos de FDCA, preferiblemente en condiciones propicias para la oxidación de los precursores furánicos de FDCA por la célula a FDCA y, opcionalmente, recuperar el FDCA, donde la célula comprende un constructo de expresión para la expresión de una secuencia de nucleótidos que codifica una HMFCa deshidrogenasa que tiene una secuencia de aminoácidos con al menos un 45 % de identidad con cualquiera de las secuencias de aminoácidos de las SEC ID N.º: 1 a 11, donde el constructo de expresión se puede expresar en la célula y la expresión de la HMFCa deshidrogenasa confiere o aumenta en la célula la capacidad de oxidar HMFCa a FFA, en comparación con una célula de tipo salvaje correspondiente que carece del constructo de expresión, donde preferiblemente, al menos un precursor furánico de FDCA se selecciona del grupo que consta de HMF, 2,5-dihidroximetil furano (DHF), HMFCa, FFA y 2,5-diformil furano (DFF), de los cuales HMF es el preferido, y donde preferiblemente los precursores furánicos de FDCA se obtienen a partir de uno o más azúcares de hexosa, preferiblemente uno o más azúcares de hexosa obtenidos de biomasa lignocelulósica, preferiblemente por deshidratación catalizada por ácido.
3. Proceso para producir FDCA según la reivindicación 2, en el que el FDCA se recupera del medio mediante un proceso que comprende precipitación ácida o salina seguida de cristalización por enfriamiento y/o extracción con disolvente.
4. Proceso para producir un polímero a partir de uno o más monómeros de FDCA, que comprende las etapas de:
- a) preparar un monómero de FDCA en un proceso según las reivindicaciones 2 o 3; y
  - b) producir un polímero a partir del monómero de FDCA obtenido en a).
5. Uso de una célula para la biotransformación a FDCA de uno o más precursores furánicos, donde preferiblemente, al menos un precursor furánico de FDCA se selecciona del grupo que consta de HMF, DHF, HMFCa, FFA y DFF, de los cuales HMF es el preferido, donde la célula comprende un constructo de expresión para la expresión de una secuencia de nucleótidos que codifica una HMFCa deshidrogenasa que tiene una secuencia de aminoácidos con al menos un 45 % de identidad con cualquiera de las secuencias de aminoácidos de las SEC ID N.º: 1 a 11, y donde el constructo de expresión se puede expresar en la célula y la expresión de la HMFCa deshidrogenasa confiere o aumenta en la célula la capacidad de oxidar HMFCa a FFA, en comparación con una célula de tipo salvaje correspondiente que carece del constructo de expresión.
6. Célula que comprende un constructo de expresión para la expresión de una secuencia de nucleótidos que codifica una HMFCa deshidrogenasa que tiene una secuencia de aminoácidos con al menos un 81.65 % de identidad con la secuencia de aminoácidos de la SEC ID N.º: 1, donde el constructo de expresión se puede expresar en la célula y la expresión de la HMFCa deshidrogenasa confiere o aumenta en la célula la capacidad de oxidar HMFCa a FFA, en comparación con una célula de tipo salvaje correspondiente que carece del constructo de expresión.
7. Célula según la reivindicación 6, donde la célula tiene además al menos uno de los siguientes:
- a) una actividad de aldehído deshidrogenasa que oxida los aldehídos furánicos a los ácidos carboxílicos furánicos correspondientes, donde preferiblemente la célula comprende un segundo constructo de expresión para la expresión de una secuencia de nucleótidos que codifica un aldehído deshidrogenasa que comprende una secuencia de aminoácidos con al menos un 45 % de identidad con cualquiera de las secuencias de aminoácidos SEC ID N.º: 24, 25, 26, 27, 28, 29 y 30, donde el segundo constructo de expresión es expresable en la célula y la expresión de la aldehído deshidrogenasa confiere o aumenta en la célula al menos una de las capacidades siguientes: i) oxidar 5-hidroximetilfurfural (HMF) a HMFCa, ii) oxidar DFF a FFA y iii) oxidar FFA a FDCA, en comparación con una célula de tipo salvaje correspondiente que carece del segundo constructo de expresión; y
  - b) la capacidad de transportar compuestos furánicos hacia dentro y/o hacia fuera de la célula, donde preferiblemente la célula comprende un tercer constructo de expresión para la expresión de una secuencia de nucleótidos que codifica un polipéptido que tiene la capacidad de transportar al menos HMFCa al interior de la célula, cuyo polipéptido comprende una secuencia de aminoácidos con al menos un 45 % de

identidad con cualquiera de las secuencias de aminoácidos SEC ID N.º: 17, 31, 32, 33 y 34, donde el tercer constructo de expresión es expresable en la célula y la expresión del polipéptido confiere o aumenta en la célula la capacidad de transportar al menos HMFCa al interior de la célula, en comparación con una célula de tipo salvaje correspondiente que carece del tercer constructo de expresión.

5 8. Célula según la reivindicación 7, donde la célula es una célula microbiana, preferiblemente la célula es una levadura o una célula fúngica filamentosa seleccionada de un género del grupo que consta de *Candida*, *Hansenula*, *Kluyveromyces*, *Pichia*, *Saccharomyces*, *Schizosaccharomyces*, *Yarrowia*, *Acremonium*, *Agaricus*,  
 10 *Aspergillus*, *Aureobasidium*, *Myceliophthora*, *Chrysosporium*, *Coprinus*, *Cryptococcus*, *Filibasidium*, *Fusarium*, *Humicola*, *Magnaporthe*, *Mucor*, *Myceliophthora*, *Neocallimastix*, *Neurospora*, *Paecilomyces*, *Penicillium*, *Piromyces*, *Panerochaete*, *Pleurotus*, *Schizophyllum*, *Talaromyces*, *Thermoascus*, *Thielavia*, *Tolypocladium* y *Trichoderma*, más preferiblemente una levadura o una célula fúngica filamentosa seleccionada de una especie del grupo que consta de *Kluyveromyces lactis*, *S. cerevisiae*, *Hansenula polymorpha*, *Yarrowia lipolytica*, *Pichia pastoris*, *Aspergillus niger*, *Aspergillus awamori*, *Aspergillus foetidus*, *Aspergillus sojae*, *Aspergillus fumigatus*,  
 15 *Talaromyces emersonii*, *Aspergillus oryzae*, *Myceliophthora thermophila*, *Trichoderma reesei* y *Penicillium chrysogenum*;  
 o,  
 preferiblemente la célula es una célula bacteriana seleccionada de un género del grupo que consta de *Escherichia*, *Anabaena*, *Aeribacillus*, *Aneurinibacillus*, *Burkholderia*, *Bradyrhizobium*, *Caulobacter*, *Cupriavidus*,  
 20 *Desulfotomaculum*, *Desulfurispora*, *Gluconobacter*, *Rhodobacter*, *Pelotomaculum*, *Pseudomonas*, *Paracoccus*, *Bacillus*, *Geobacillus*, *Brevibacillus*, *Brevibacterium*, *Corynebacterium*, *Rhizobium* (*Sinorhizobium*), *Flavobacterium*, *Klebsiella*, *Enterobacter*, *Lactobacillus*, *Lactococcus*, *Methylobacterium*, *Ralstonia*, *Rhodopseudomonas*, *Staphylococcus* y *Streptomyces*, más preferiblemente una célula bacteriana seleccionada de una especie del grupo que consta de *A. pallidus*, *A. terranovensis*, *B. subtilis*, *B. amyloliquefaciens*, *B. coagulans*, *B. kribbensis*, *B. licheniformis*, *B. puntis*, *B. megaterium*, *B. halodurans*, *B. pumilus*, *B. thermoruber*, *B.*  
 25 *panacihumi*, *C. basilensis*, *D. kuznetsovii*, *D. thermophila*, *G. kaustophilus*, *Gluconobacter oxydans*, *Caulobacter crescentus* CB 15, *Methylobacterium extorquens*, *Rhodobacter sphaeroides*, *Pelotomaculum thermopropionicum*, *Pseudomonas zeaxanthinifaciens*, *Pseudomonas putida*, *Paracoccus denitrificans*, *E. coli*, *C. glutamicum*, *Staphylococcus carnosus*, *Streptomyces lividans*, *Sinorhizobium meliloti* y *Rhizobium radiobacter*.

30 9. Polipéptido que tiene actividad de HMFCa deshidrogenasa, cuyo polipéptido comprende una secuencia de aminoácidos que tiene al menos un 81.65 % de identidad de secuencia con la secuencia de aminoácidos de la SEC ID N.º: 1.

35 10. Molécula de ácidos nucleicos que comprende al menos una de las siguientes:

- a) una secuencia de nucleótidos que codifica el polipéptido como se define en la reivindicación 9;
- b) una secuencia de nucleótidos como la establecida en la SEC ID N.º: 12 o 13;
- c) una secuencia de nucleótidos cuya secuencia difiere de la secuencia de una secuencia de nucleótidos de b) debido a la degeneración del código genético; y
- 40 d) una secuencia de nucleótidos que es el complemento inverso de una secuencia de nucleótidos tal como se ha definido en los puntos a) a c),

donde, preferiblemente, la molécula de ácidos nucleicos es un vector.

Fig. 1a

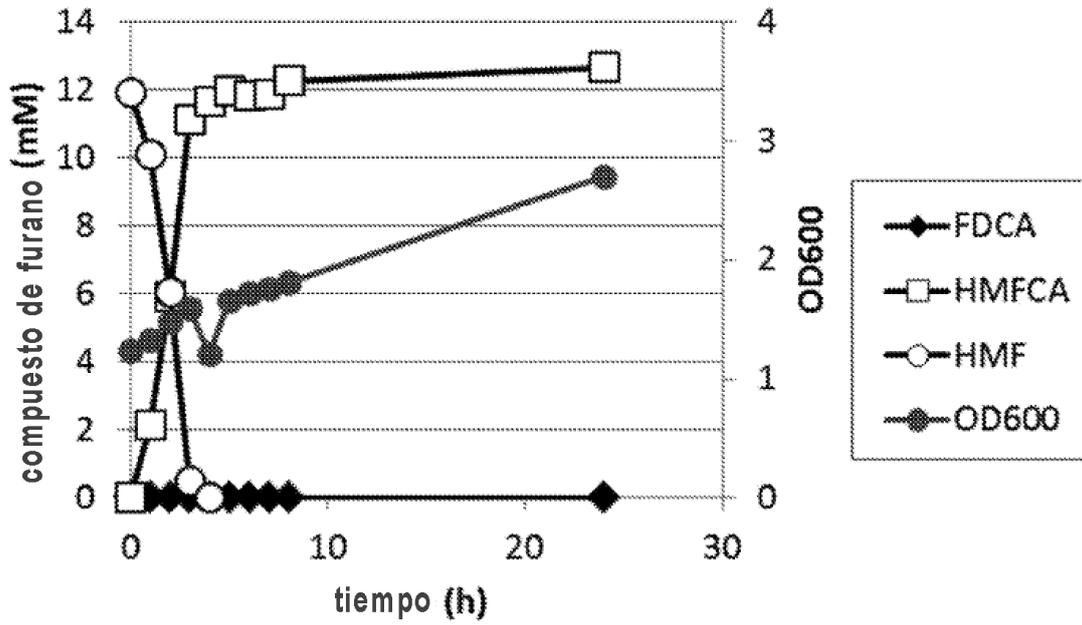


Fig. 1b

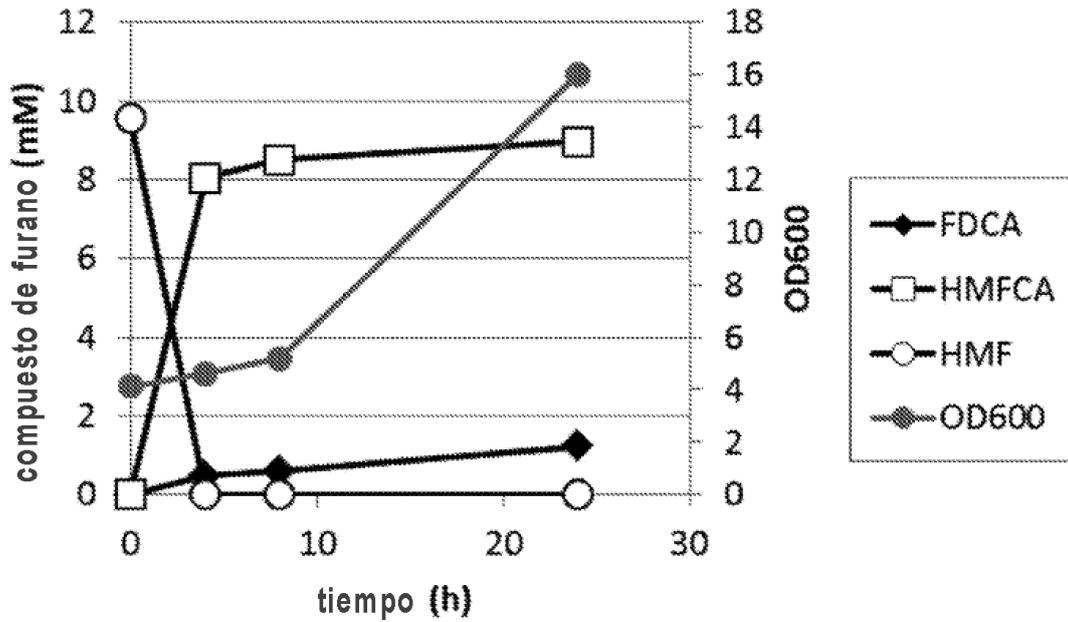


Fig. 2

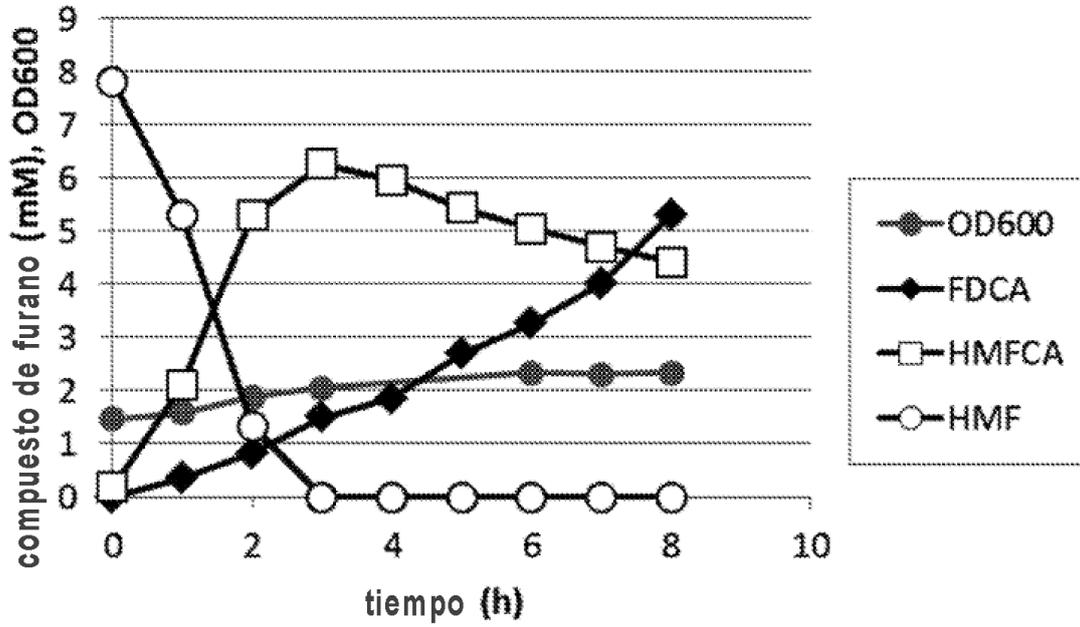


Fig. 3

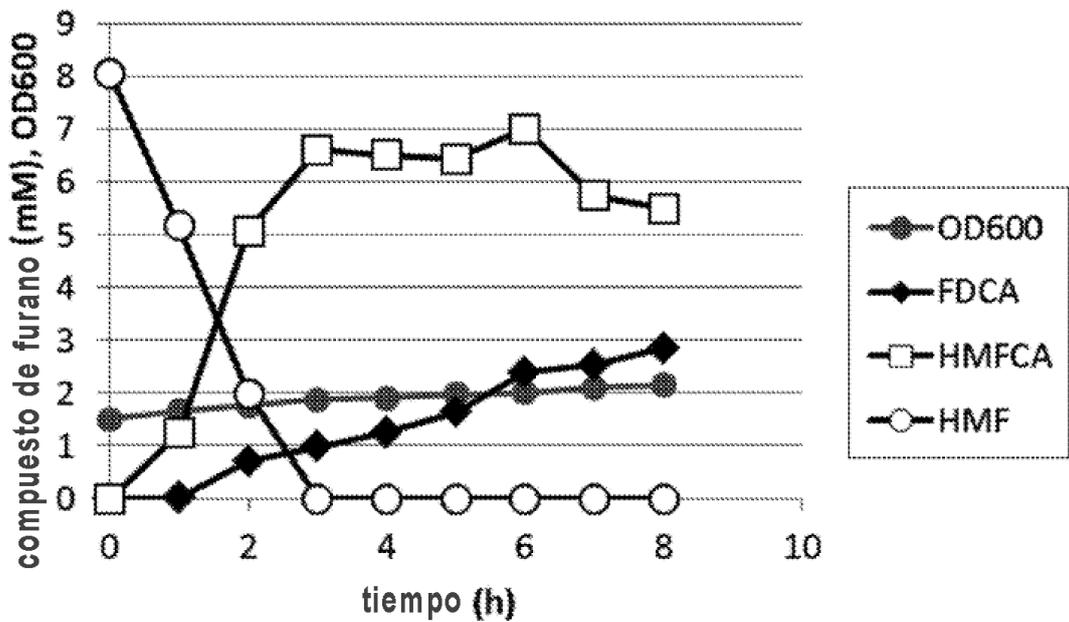


Fig. 4

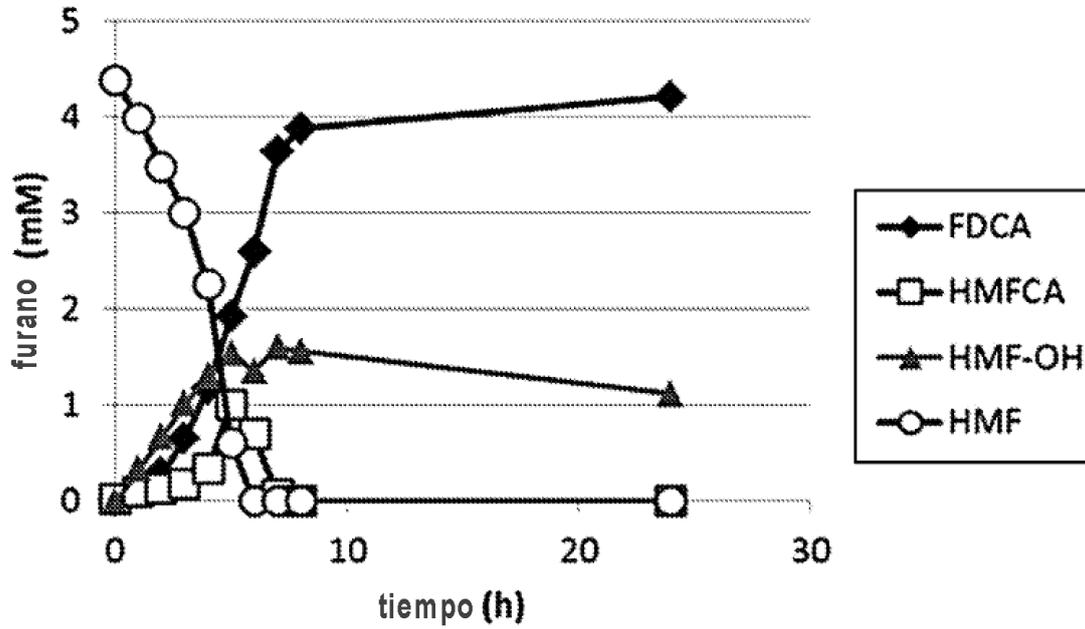


Fig. 5

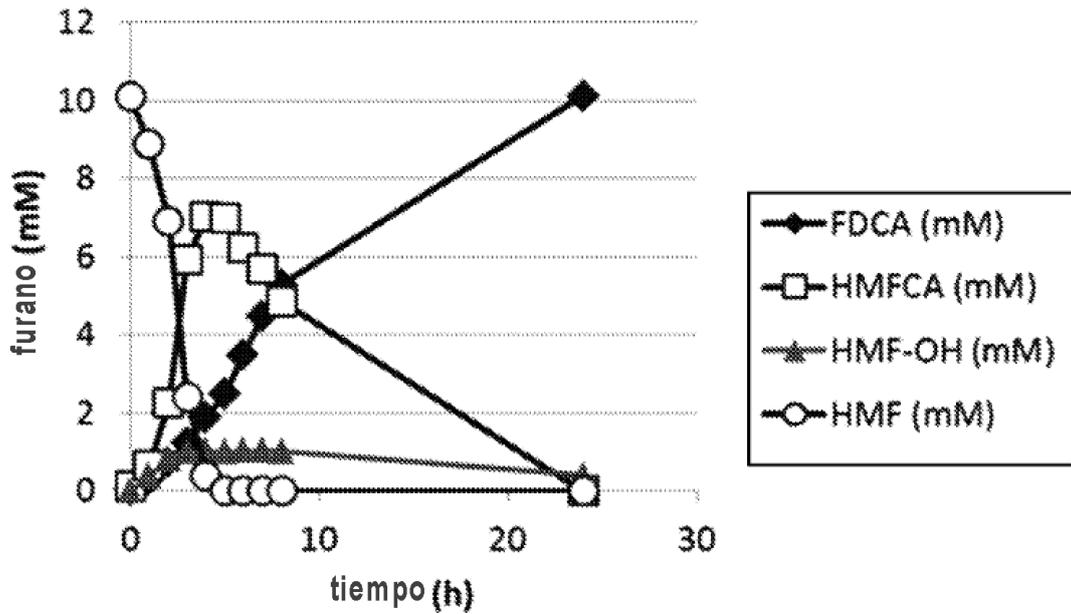


Fig. 6

