

19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 543 568**

51 Int. Cl.:

**C12N 9/00** (2006.01)  
**C12N 9/02** (2006.01)  
**C12P 17/04** (2006.01)  
**C12N 1/15** (2006.01)  
**C12N 1/21** (2006.01)  
**C12N 5/10** (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **07.03.2012 E 12709978 (6)**

97 Fecha y número de publicación de la concesión europea: **20.05.2015 EP 2683815**

54 Título: **Célula genéticamente modificada y procedimiento de uso de dicha célula**

30 Prioridad:

**08.03.2011 NL 2006359**

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

**20.08.2015**

73 Titular/es:

**PURAC BIOCHEM B.V. (100.0%)  
Arkelsedijk 46  
4206 AC Gorinchem, NL**

72 Inventor/es:

**WIERCKX, NICK JOHANNES PETRUS;  
ELINK SCHUURMAN, TOM DANIËL;  
KUIJPER, SIPKO MAARTEN y  
RUIJSSENAARS, HARALD JOHAN**

74 Agente/Representante:

**CARPINTERO LÓPEZ, Mario**

**ES 2 543 568 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

Célula genéticamente modificada y procedimiento de uso de dicha célula

### Campo de la invención

5 La presente invención en general se refiere al campo de la biotransformación de compuestos furánicos. Dichas biotransformaciones encuentran utilidad en la producción de ácido furan-2,5-dicarboxílico (FDCA) y el procesamiento de material que contiene lignocelulosa para, por ejemplo, la producción de biocombustibles y productos bioquímicos. Más en particular, la presente invención se refiere a nuevas células modificadas genéticamente con mejores características para la transformación biocatalítica de compuestos furánicos. Un aspecto adicional de la invención se refiere a un vector adecuado para la modificación genética de una célula huésped para mejorar sus características y la biotransformación de compuestos furánicos. Otros aspectos de la invención se refieren a procesos para la biotransformación de ácido 5-(hidroximetil)furan-2-carboxílico (ácido HMF) y sus precursores con el uso de la célula de acuerdo con la invención.

### Antecedente de la invención

15 La biotransformación de compuestos furánicos está recibiendo una atención cada vez mayor. Esto es tanto en cuanto a la bioproducción de ácido furan-2,5-dicarboxílico (FDCA), que es un prometedor producto químico de valor añadido de la biomasa (Werpy y col., (2004)) y en cuanto a su papel negativo en la producción fermentativa de biocombustibles y productos bioquímicos a partir de materiales que contienen lignocelulosa (Almeida y col., (2009)).

20 Recientemente se ha aislado un compuesto furánico utilizando un organismo, *Cupriavidus basilensis* HMF14, (Koopman y col., (2010a)). Este organismo es capaz de metabolizar furfural y 5-(hidroximetil)furan-2-carbaldehído (HMF). La ruta de degradación de furfural y HMF de *Cupriavidus basilensis* HMF14 se ha divulgado en Koopman y col., (2010a) junto con los genes implicados.

25 La introducción funcional del gen *hmfH* de *Cupriavidus basilensis* HMF 14 en *Pseudomonas putida* S12 se divulga en Koopman y col., (2010b). La cepa resultante tiene buenas capacidades de producción de FDCA usando HMF como sustrato. No obstante, la acumulación observada de ácido 5-(hidroximetil)furan-2-carboxílico (ácido-HMF) requeriría tiempos de procesamiento prolongados o medidas alternativas para eliminar este subproducto. La eliminación suficiente del subproducto de ácido-HMF es deseable para muchas de las aplicaciones para las cuales el FDCA puede producirse y, en ocasiones, es incluso esencial.

30 En búsqueda de una solución al problema de la acumulación de ácido HMF, los inventores de la presente invención han encontrado ahora sorprendentemente que la expresión de ciertos polipéptidos en el sistema de producción de *Pseudomonas putida* S12 FDCA reduce eficazmente la acumulación de ácido HMF.

### Sumario de la invención

35 Los presentes hallazgos de los inventores han dado como resultado el concepto generalizado de que la expresión en un huésped de un polipéptido que tiene una secuencia de aminoácidos de SEC ID NO. 1 o 2 o sus análogos/homólogos (tales como las SEC ID N° 3 o 4) junto con uno o más polipéptidos capaces de la conversión de ácido 5- (hidroximetil)furan-2-carboxílico (ácido HMF), tiene como resultado la bioconversión eficaz del ácido HMF. La mejora de la bioconversión de ácido HMF es beneficiosa para la eliminación de ácido HMF y sus precursores furánicos a partir de materias primas, en las que los compuestos furánicos se consideran perjudiciales, tales como materias primas para fermentaciones etanológicas para la producción de, por ejemplo, biocombustibles o para fermentaciones para la producción biológica de productos químicos. En otras aplicaciones, la mejora de la bioconversión de ácido HMF mejorará la bioproducción de una sustancia química en la que el ácido HMF es un material de partida o un intermedio, tal como en la bioproducción de FDCA.

40 En consecuencia, un primer objeto de la invención es una célula modificada genéticamente que comprende una primera secuencia de polinucleótidos que codifica un primer polipéptido que tiene una secuencia de aminoácidos de SEC ID N° 1, 2, 3 o 4 o sus análogos / homólogos y una segunda secuencia de polinucleótidos que codifica un segundo polipéptido que tiene actividad convertidora de ácido HMF. El polipéptido convertidor de ácido HMF puede ser la oxidorreductasa codificada por el gen *hmfH* de *Cupriavidus Basilensis* HMF14 descrito previamente (Koopman y col., 2010a y Koopman y col., (2010b)). De acuerdo con determinadas realizaciones se prefiere que la célula modificada genéticamente comprenda una tercera secuencia de polinucleótidos que codifica un tercer polipéptido que tiene una secuencia de aminoácidos de SEC ID N° 19, 20, 21, 22, 23, 24, 25 o sus análogos / homólogos. La expresión funcional de las terceras secuencias de aminoácidos tiene como resultado actividad aldehído deshidrogenasa capaz de convertir aldehídos furánicos.

50 Si el segundo polipéptido es una oxidorreductasa, la coexpresión del primer polipéptido simultáneamente con la oxidorreductasa también puede mejorar la calidad de un biocatalizador de célula completa que comprende la oxidorreductasa con respecto a la producción biocatalítica FDCA.

La célula de acuerdo con la invención es modificada genéticamente mediante la introducción funcional de, al menos, la primera secuencia de polinucleótidos. Preferentemente, la célula se modifica genéticamente mediante la introducción funcional tanto de la primera como la segunda secuencia de polinucleótidos. Como alternativa, la célula se modifica genéticamente mediante la introducción funcional de la primera y la segunda secuencia de polinucleótidos. La introducción funcional de los tres del primero, el segundo y el tercer polinucleótido es una alternativa adicional.

Un aspecto adicional de la invención se refiere a un vector adecuado para la modificación genética de una célula huésped. El vector comprende una primera secuencia de polinucleótidos que codifica un primer polipéptido que tiene una secuencia de aminoácidos de SEC ID N° 1, 2, 3 o 4 o sus análogos / homólogos y una segunda secuencia de polinucleótidos que codifica un segundo polipéptido que tiene actividad convertidora de ácido 5-(hidroximetil)furan-2-carboxílico (ácido HMF). Opcionalmente, el vector puede comprender una tercera secuencia de polinucleótidos que codifica un tercer polipéptido que tiene una secuencia de aminoácidos de SEC ID N° 19, 20, 21, 22, 23, 24, 25 o sus análogos / homólogos. Tal vector es adecuado para obtener una célula modificada genéticamente de acuerdo con la invención.

Otros aspectos de la invención se refieren a un proceso de conversión de ácido 5- (hidroximetil)furan-2-carboxílico (ácido HMF). Este proceso hace uso de la célula de acuerdo con la invención. De acuerdo con realizaciones preferidas, este proceso es adecuado para la producción de FDCA.

Un aspecto adicional de la presente invención está dirigido al uso de una célula modificada genéticamente de acuerdo con la invención, para la biotransformación de precursores furánicos a FDCA.

#### **Breve descripción de las figuras**

La Figura 1 divulga una representación esquemática de las reacciones oxidativas de compuestos furánicos a ácido 2,5-furandicarboxílico. Se presentan los siguientes compuestos furánicos: 1, alcohol HMF; 2, HMF; 3, ácido HMF; 4, FFA; 5, FDCA.

La Figura 2 divulga la producción de FDCA y la acumulación de ácido HMF en cultivos de alimentación de HMF de *P. putida* S12\_2642 (figura 2a) y B38 (figura 2b) (◆, FDCA (mM); ■, ácido HMF (mM); O, peso seco de la célula (PSC; g/l); línea discontinua: Velocidad de alimentación de HMF (ml de solución 4 M/h))

La Figura 3 divulga la productividad específica de FDCA en diversos puntos temporales durante los procesos de alimentación discontinua presentados en la figura 2 (Oscuro: *P. putida* S12\_2642; claro: *P. putida* S12\_B38).

La Figura 4 divulga la producción de FDCA y la acumulación de ácido HMF en experimentos en matraz de agitación con *P. putida* S12\_B38 (figura 4a) y B51 (figura 4b) (A, HMF (mM); ◆, FDCA (mM); ■, ácido HMF (mM); O, peso seco de la célula (PSC; g/l));

La Figura 5 divulga la producción de FDCA y la acumulación de ácido HMF en cultivo de alta densidad celular de alimentación de HMF de *P. putida* S12\_B38 que comienza con densidad celular alta (□, HMF; ◆, FDCA (mM); ■, ácido HMF (mM); O, peso seco de la célula (PSC g/l); línea discontinua: Velocidad de alimentación de HMF (ml de solución 4 M/h))

La Figura 6 divulga la acumulación de FDCA, FFA y ácido HMF en cultivos en matraces de agitación que contienen HMF de *P. putida* S12\_B38 (HmfH y HmfT1 co-expresados; A); S12\_B97 (HmfH, HmfT1 y aldehído deshidrogenasa co-expresados; B); y S12\_B101 (HmfH y aldehído deshidrogenasa co-expresados; C). ▲, HMF (mM); ◆, FDCA (mM); ■, ácido HMF (mM); ●, FFA (mM); O, peso seco de la célula (PSC; g/l).

#### **Breve descripción de las secuencias**

La SEC ID N° 1 establece la secuencia de aminoácidos de HmfT1 de *Cupriavidus basilensis* HMF1. La secuencia tiene un número de acceso en GenBank ADE20411.1.

La SEC ID N° 2 establece la secuencia de aminoácidos de HmfT2 de *Cupriavidus basilensis* HMF14. La secuencia tiene un número de acceso en GenBank ADE20404.1.

La SEC ID N° 3 establece la secuencia de aminoácidos del producto proteico del gen con el locus *mrad2831\_4728* de *Methylobacterium radiotolerans* JCM 2831 (= ATCC 27329 = DSM 1819). La secuencia tiene un número de acceso en GenBank ACB26689.1.

La SEC ID N° 4 establece la secuencia de aminoácidos del producto proteico del gen *saci\_2058* de *Sulfolobus acidocaldarius* DSM 639. La secuencia tiene un número de acceso en GenBank AAY81352.1.

La SEC ID N° 5 establece la secuencia de aminoácidos de HmfH de *Cupriavidus basilensis* HMF14. La secuencia tiene un número de acceso en GenBank ADE20408.1.

## ES 2 543 568 T3

- La SEC ID N° 6 establece la secuencia de aminoácidos del producto proteico del gen *blr0367* de *Bradyrhizobium japonicum* USDA 110. La secuencia tiene un número de acceso en GenBank BAC45632.1.
- La SEC ID N° 7 establece la secuencia de codificación de *hmfT1* de *Cupriavidus basilensis* HMF14.
- La SEC ID N° 8 establece la secuencia de codificación de *hmfT2* de *Cupriavidus basilensis* HMF14.
- 5 La SEC ID N° 9 establece la secuencia de codificación del gen con el locus *mrad2831\_4728* de *Methylobacterium radiotolerans* JCM 2831 (= ATCC 27329 = DSM 1819).
- La SEC ID N° 10 establece la secuencia de codificación del gen *saci\_2058* de *Sulfolobus acidocaldarius* DSM 639.
- La SEC ID N° 11 establece la secuencia de codificación de *hmfH* de *Cupriavidus basilensis* HMF14.
- La SEC ID N° 12 establece la secuencia de codificación del gen *blr0367* de *Bradyrhizobium japonicum* USDA 110.
- 10 La SEC ID N° 13-18 establecen las secuencias de varios cebadores sintéticos. Se indican los lugares de restricción (subrayados), los codones de iniciación y terminación (complemento inverso) (en cursiva) y los supuestos sitios de unión al ribosoma (en minúscula). El cebador FN23 se diseñó justo aguas arriba del codón de iniciación de *hmfH*.
- La SEC ID N° 13 establece la secuencia de nucleótidos del cebador de ADN sintético *hmfT1* (f)
- 5'- ACGAATTCAAaggagACAACAATGGAAG-3'
- 15 La SEC ID N° 14 establece la secuencia de nucleótidos del cebador de ADN sintético *hmfT1* (r)
- 5'- AAGCTAGCTGAGCAGTCACCC<sup>T</sup>CACTC-3'
- La SEC ID N° 15 establece la secuencia de nucleótidos del cebador de ADN sintético FN23.
- 5'-CGGAATTCCACATGACAaggagACCG-3'
- La SEC ID N° 16 establece la secuencia de nucleótidos del cebador de ADN sintético FN24.
- 20 5'-CGGAATTCGCTTCGGTCTTCAACTCGGATG-3'
- La SEC ID N° 17 establece la secuencia de nucleótidos del cebador de ADN sintético *mrad* (f)
- 5'- ACGAATTCggaggAAATCTATGCAGACC -3'
- La SEC ID N° 18 establece la secuencia de nucleótidos del cebador de ADN sintético *mrad* (r).
- 5'- AAGCTAGCGCAGAACCGTATCGTCAG -3'
- 25 La SEC ID N° 19 establece la secuencia de aminoácidos de la aldehído deshidrogenasa *Adh* de *Cupriavidus basilensis* HMF14.
- La SEC ID N° 20 establece la secuencia de aminoácidos que tiene el número de acceso de Genbank: YP\_003609156.1.
- 30 La SEC ID N° 21 establece la secuencia de aminoácidos que tiene el número de acceso de Genbank: ZP\_02881557.1.
- La SEC ID N° 22 establece la secuencia de aminoácidos que tiene el número de acceso de Genbank: YP\_003451184.1.
- La SEC ID N° 23 establece la secuencia de aminoácidos que tiene el número de acceso de Genbank: ACA09737.1.
- La SEC ID N° 24 establece la secuencia de aminoácidos que tiene el número de acceso de Genbank: YP\_530742.1.
- 35 La SEC ID N° 25 establece la secuencia de aminoácidos que tiene el número de acceso de Genbank: YP\_001541929.1.
- La SEC ID N° 26 establece la secuencia de polinucleótidos de *adh* que codifica la aldehído deshidrogenasa *Adh* de *Cupriavidus basilensis* HMF14.
- 40 La SEC ID N° 27 establece la secuencia de polinucleótidos que codifica la secuencia de aminoácidos que tiene el número de acceso de Genbank: YP\_003609156.1.
- La SEC ID N° 28 establece la secuencia de polinucleótidos que codifica la secuencia de aminoácidos que tiene el número de acceso de Genbank: ZP\_02881557.1.

La SEC ID N° 29 establece la secuencia de polinucleótidos que codifica la secuencia de aminoácidos que tiene el número de acceso de Genbank: YP\_003451184.1.

La SEC ID N° 30 establece la secuencia de polinucleótidos que codifica la secuencia de aminoácidos que tiene el número de acceso de Genbank: ACA09737.1.

- 5 La SEC ID N° 31 establece la secuencia de polinucleótidos que codifica la secuencia de aminoácidos que tiene el número de acceso de Genbank: YP\_530742.1.

La SEC ID N° 32 establece la secuencia de polinucleótidos que codifica la secuencia de aminoácidos que tiene el número de acceso de Genbank: YP\_001541929.1.

- 10 La SEC ID N° 33 establece la secuencia de nucleótidos del cebador de ADN sintético FN13: 5'-ATGCGGCCGCAACAaggagAAGATGGAATGAACG- 3' (secuencia subrayada: sitio de restricción *NotI*; codón de iniciación (ATG) del gen que codifica la aldehído deshidrogenasa en cursiva; el supuesto sitio de unión al ribosoma (RBS) en minúsculas)

- 15 La SEC ID N° 34 establece la secuencia de nucleótidos del cebador de ADN sintético FN14: 5'-ATGCGGCCGCGCGTTCGGGGTCGGTGCTA- 3' (secuencia subrayada: sitio de restricción *NotI*; codón de terminación (hebra complementaria inversa) en cursiva).

### Descripción detallada de la invención

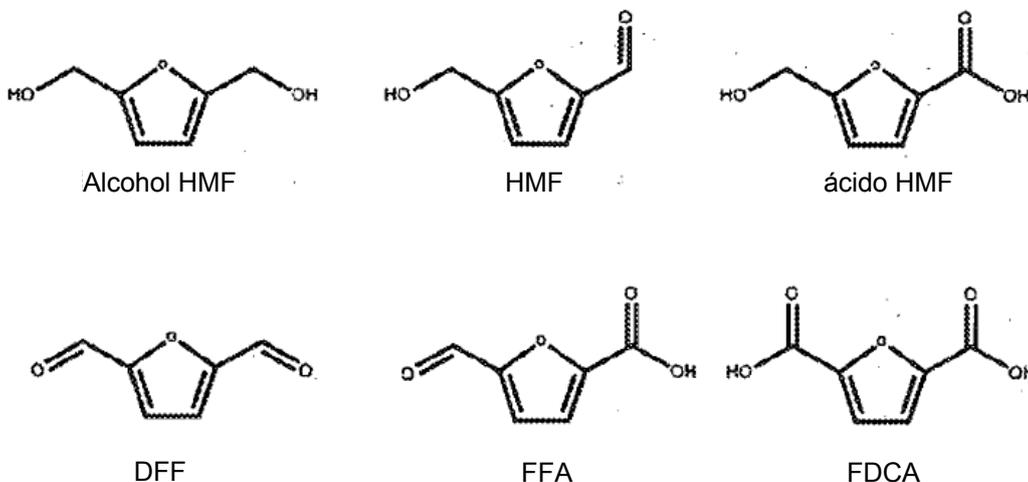
#### Definiciones generales

- 20 A lo largo de la presente memoria descriptiva y las reivindicaciones adjuntas, las palabras "comprenden" e "incluyen" y variaciones tales como "comprende", "que comprende", "incluye" y "que incluye" deben interpretarse de forma incluyente. Es decir, estas palabras tienen como objetivo transmitir la posible inclusión de otros elementos o números enteros no citados específicamente, cuando el contexto lo permita.

Los artículos "un/uno" y "una" se usan en el presente documento para referirse a uno o a más de uno (es decir, a uno o a al menos uno) del objeto gramatical del artículo. A modo de ejemplo, "un elemento" puede significar un elemento o más de un elemento.

- 25 El término "una serie de" debe entenderse que tiene el significado de uno o más.

- 30 En el presente documento se entiende que los compuestos furánicos son cualquier compuesto que tiene un anillo de furano. Preferentemente, el compuesto furánico es un compuesto que puede oxidarse a ácido 2,5-furandicarboxílico. Los compuestos furánicos relevantes en el contexto de la presente invención incluyen [5-(hidroximetil)furan-2-il]metanol ("alcohol HMF"), 5-(hidroximetil) furan-2-carbaldehído ("HMF"), ácido 5-(hidroximetil)furan-2-carboxílico ("ácido HMF"), ácido 5-formilfuran-2-carboxílico ("FFA"), furano-2,5-dicarbaldéhído (DFF) y ácido furano-2,5-dicarboxílico ("FDCA"). El anillo de furano o cualquiera de sus grupos laterales sustituibles pueden estar sustituidos, por ejemplo, con OH, alquilo C1-C10, alquilo, alilo, arilo o RO- resto éter, incluyendo grupos cíclicos, en cualquier posición disponible en el anillo furano. Las estructuras químicas de un número de compuestos furánicos relevantes se presentan a continuación.



- 35 El término "polinucleótido" incluye ácidos polidesoxirribonucleicos (ADN) y ácidos polirribonucleicos (ARN) y el término puede referirse a ADN o ARN. El experto será consciente de las diferencias en la estabilidad de las

moléculas de ADN y ARN. Así, el experto en la técnica será capaz de entender a partir del contexto del uso del término "polinucleótido", cuál de las formas de polinucleótido (ADN y / o ARN) es adecuada.

El término "similitud" secuencia tal como se utiliza en el presente documento se refiere a la medida en que las secuencias de polinucleótidos o de proteínas individuales son iguales. El grado de similitud entre dos secuencias se basa en la medida de la identidad combinada con la medida de los cambios conservadores. El porcentaje de "similitud de secuencia" es el porcentaje de aminoácidos o nucleótidos que es o bien idéntico o con cambios conservadores a saber "similitud de secuencia" = (% de identidad de secuencia) + (% cambios conservadores).

Para el propósito de la presente invención "cambios conservadores" e "identidad" se consideran especies del término más amplio "similitud". De esta manera, cada vez que se utiliza el término "similitud" de secuencia, abarca la "identidad" de secuencia y los "cambios conservadores".

La expresión "identidad de secuencia" es conocida para el experto en la técnica. Con el fin de determinar el grado de identidad de secuencia compartida por dos secuencias de aminoácidos o de dos secuencias de ácido nucleico, se alinean las secuencias para fines comparativos óptimos (p. ej., se pueden introducir espacios en la secuencia de una primera secuencia de aminoácidos o de ácido nucleico para una alineación óptima con una segunda secuencia de aminoácidos o de ácido nucleico). Dicha alineación puede llevarse a cabo sobre las longitudes completas de las secuencias que se están comparando. Como alternativa, la alineación puede llevarse a cabo sobre una longitud más corta de comparación, por ejemplo por encima de aproximadamente 20, aproximadamente 50, aproximadamente 100 o más ácidos nucleicos / bases o aminoácidos.

Después se comparan los residuos de aminoácidos o nucleótidos en las correspondientes posiciones de aminoácidos o posiciones de nucleótidos. Cuando una posición en la primera secuencia está ocupada por el mismo residuo de aminoácidos o nucleótido en la correspondiente posición en la segunda secuencia, las moléculas son idénticas en dicha posición. El grado de identidad compartida entre las secuencias normalmente se expresa en términos del porcentaje de identidad entre las dos secuencias y es función del número de posiciones idénticas compartidas por residuos idénticos en las secuencias (es decir, % de identidad= número de residuos idénticos en las posiciones correspondientes/número total de posiciones x 100). Preferentemente, las dos secuencias que se están comparando son de la misma o sustancialmente la misma longitud.

El porcentaje de "cambios conservadores" se puede determinar de forma similar al porcentaje de identidad de secuencia. Sin embargo, en este caso, los cambios en un lugar específico de una secuencia de aminoácidos o nucleótidos que son propensos a conservar las propiedades funcionales de los residuos originales se puntúan como si se hubiera producido ningún cambio.

Para las secuencias de aminoácidos, las propiedades funcionales relevantes son las propiedades fisicoquímicas de los aminoácidos. Una sustitución conservadora para un aminoácido en un polipéptido de la invención se puede seleccionar de otros miembros de la clase a la que pertenece el aminoácido. Por ejemplo, es bien conocido en la técnica de la bioquímica de las proteínas que un aminoácido que pertenece a un grupo de aminoácidos que tienen un tamaño o característica particular (tales como carga, hidrofobicidad e hidrofiliidad) puede ser sustituido por otro aminoácido sin alterar la actividad de una proteína, particularmente en regiones de la proteína que no están directamente asociadas con la actividad biológica. Por ejemplo, los aminoácidos no polares (hidrófobos) incluyen alanina, leucina, isoleucina, valina, prolina, fenilalanina, triptófano y tirosina. Los aminoácidos neutros polares incluyen glicina, serina, treonina, cisteína, tirosina, asparagina y glutamina. Los aminoácidos con carga positiva (básicos) incluyen arginina, lisina e histidina. Los aminoácidos con carga negativa (ácidos) incluyen ácido aspártico y ácido glutámico. Las sustituciones conservadoras incluyen, por ejemplo, Lys por Arg y viceversa para mantener una carga positiva; Glu por Asp y viceversa para mantener una carga negativa; Ser por Thr de manera que se mantenga un -OH libre; y Gln por Asn para mantener un -NH<sub>2</sub> libre.

Para las secuencias de nucleótidos, las propiedades funcionales pertinentes es principalmente la información biológica que un determinado nucleótido lleva dentro del marco de lectura abierto de la secuencia en relación con la transcripción y / o maquinaria de traducción. Es de conocimiento común que el código genético tiene degeneración (o redundancia) y que múltiples codones pueden llevar a la misma información con respecto al aminoácido que codifican. Por ejemplo, en ciertas especies del aminoácido leucina está codificado por los codones UUA, UUG, CUU, CUC, CUA, los CUG (o TTA, TTG, CTT, CTC, CTA, CTG para el ADN) y el aminoácido serina está especificado por UCA, UCG, UCC, UCU, AGU, AGC (o TCA, TCG, TCC, TCT, AGT, AGC para el ADN). Los cambios de nucleótidos que no alteran la información traducida se consideran cambios conservadores.

El experto en la técnica conocerá el hecho de que varios programas de ordenador diferentes, utilizando diferentes algoritmos matemáticos, están disponibles para determinar la identidad entre dos secuencias. Por ejemplo, se puede hacer uso de un programa de ordenador que emplea el algoritmo de Needleman y Wunsch (Needleman y col., (1970)). De acuerdo con una forma de realización, el programa de ordenador es el programa GAP en el paquete de software GCG Accelerlys (Accelerlys Inc., San Diego EE.UU.). Las matrices de sustitución que se pueden usar son, por ejemplo una matriz BLOSUM 62 o una matriz PAM250, con un peso de hueco de 16, 14, 12, 10, 8, 6 o 4 y un peso de longitud de 1, 2, 3, 4, 5 o 6. El experto en la técnica apreciará que todos estos parámetros diferentes

producirán resultados ligeramente diferentes, pero que el porcentaje de identidad global de dos secuencias no se altera significativamente cuando se utilizan algoritmos diferentes.

5 De acuerdo con una realización, el porcentaje de identidad entre dos secuencias de nucleótidos se determina usando el programa GAP en el paquete de software GCG Accelrys (Accelrys Inc., San Diego EE.UU.). Se usa una matriz NWSgapdna CMP y un peso de hueco de 40, 50, 60, 70, u 80 y un peso de longitud de 1, 2, 3, 4, 5 o 6.

10 En otra realización, el porcentaje de identidad de dos secuencias de aminoácidos o de nucleótidos se determina usando el algoritmo de E. Meyers y W. Miller (Meyers y col., (1989)) que se ha incorporado en el programa ALIGN (versión 2.0) (disponible en la consulta ALIGN utilizando datos de la secuencia del servidor Genestream IGH Montpellier Francia <http://vega.igh.cnrs.fr/bin/align-guess.cgi>) usando una tabla de residuos del peso de PAM120, una penalización de la longitud del hueco de 12 y una penalización por hueco de 4.

Para la presente invención se prefiere más utilizar BLAST (Herramienta de Alineamiento Local Básico) para determinar el porcentaje de identidad y / o similitud entre secuencias de nucleótidos o de aminoácidos.

15 Las consultas que utilizan los programas de BLASTn, BLASTp, BLASTX, tBLASTn y tBLASTx Altschul y col., (1990) se pueden enviar a través de las versiones online de BLAST accesible a través de <http://www.ncbi.nlm.nih.gov/>. Como alternativa se puede usar una versión independiente de BLAST (por ejemplo, la versión 2.2.24 (publicada el 23 de agosto 2010)) descargable también a través de la página web del NCBI. Preferentemente, las consultas BLAST se realizan con los siguientes parámetros. Para determinar el porcentaje de identidad y / o similitud entre secuencias de aminoácidos: algoritmo: blastp; tamaño de la palabra: 3; matriz de puntuación: BLOSUM62; costes del hueco: Existencia: 11, Extensión: 1; ajustes de composición: ajuste de matriz de puntuación de composición condicional; filtro: desactivado; máscara: desactivado. Para determinar el porcentaje de identidad y / o similitud entre secuencias de nucleótidos: algoritmo: blastp; tamaño de la palabra: 11; coincidencias máximas en el intervalo de consultas: 0; puntuaciones por apareamientos/apareamientos erróneos: 2, -3; costes del hueco: Existencia: 5, Extensión: 2; filtro: regiones de baja complejidad; máscara: máscara para la tabla de búsqueda solamente.

25 El porcentaje de "cambios conservadores" puede determinarse que es similar al porcentaje de identidad de secuencia con la ayuda de los algoritmos indicados y programas de ordenador. Algunos programas de ordenador, por ejemplo, BLASTp, presentan el número / porcentaje de positivos (= similitud) y el número / porcentaje de identidad. El porcentaje de cambios conservadores se puede derivar de la misma restando el porcentaje de identidad del porcentaje de positivos / similitud (porcentaje de cambios conservadores = porcentaje de similitud - porcentaje de identidad).

30 Las técnicas de biología molecular generales, tales como experimentos de hibridación, experimentos de PCR, digestiones con enzimas de restricción, transformación de los huéspedes etcétera pueden realizarse de acuerdo con la práctica estándar conocida para el experto como se divulga en Sambrook y col., 1989, Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, N.Y.; y Ausubel y col., (eds.), 1995, Current Protocols in Molecular Biology, (John Wiley & Sons, N.Y.).

### 35 **Primer polinucleótido y polipéptido**

La célula modificada genéticamente de acuerdo con la invención comprende un primer polinucleótido que codifica un primer polipéptido. El primer polipéptido comprende una secuencia de aminoácidos que tiene al menos 45 %, preferentemente al menos 60 %, tal como al menos 70 %, más preferentemente al menos 80 %, tal como 90 %, más preferentemente al menos 95 % de similitud de secuencia con una secuencia de aminoácidos de la SEC ID N° 1, 2, 3 o 4.

40 Como alternativa, la similitud de secuencia se puede expresar como al menos 45 %, 50 %, 55 %, 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 % o al menos 99 % de similitud. De acuerdo con una realización, los porcentajes de similitud indicados pueden ser porcentajes de identidad. En una realización concreta, el primer polipéptido puede comprender la secuencia de aminoácidos tal como se expone en cualquiera de las SEC ID N° 1, 2, 3 o 4.

Tal polipéptido es capaz de proporcionar una mejora con respecto a la bioconversión en ácido HMF.

Sin pretender quedar ligado a teoría alguna, se cree que dicho polipéptido tiene capacidades de transporte de ácido HMF. Mediante el transporte de ácido HMF al interior de la célula por el primer polipéptido, mejora su disponibilidad para la conversión intracelular. Por lo tanto, la bioconversión en ácido HMF puede mejorar.

50 Tal mejora en la bioconversión en ácido HMF se ha demostrado en los ejemplos para HmfT1 (SEC ID N° 1) y el producto proteico del gen con el locus *mrad2831\_4728* de *Methylobacterium radiotolerans* JCM 2831 (= ATCC 27329 = DSM 1819) (SEC ID N° 3). Sobre la base del nivel de similitud / identidad de secuencia con estas secuencias está justificado esperar que HmfT2 (SEC ID N° 2) y el producto proteico del gen *Saci\_2058* de *Sulfolobus acidocaldarius* DSM 639 tendrá efectos similares.

55 HmfT2 (SEC ID N° 2) tiene más de 90 % de similitud con HmfT1 (SEC ID N° 1). El nivel de identidad es 87 %.

Una funcionalidad similar para el producto proteico del gen *saci\_2058d* e S DSM 639 (SEC ID N° 4) se puede esperar en base a una alineación de múltiples secuencias CLUSTALW2 con los transportistas de diferentes familias funcionales. El transportador *Sulfolobus* forma un clúster con HmfT1 (SEC ID N° 1) y Mrad2831\_4728 (SEC ID N° 3). Además, el análisis del genoma de *S. acidocaldarius* ha demostrado que el gen transportador *saci\_2058* está flanqueado por genes que codifican funcionalidades similares a las de los grupos de genes *hmf*, respectivamente, el tipo de actividades que se espera en la degradación de compuestos similares a HMF. Ejemplos: *Saci\_2057*, alcohol deshidrogenasa; *Saci\_2059/2060*, anillo aromático dioxigenasa; *Saci\_2062*, acil-CoA sintetasa (funcionalmente comparable a *hmfD*); *Saci\_2063*, enoil-CoA hidratasa (funcionalmente comparable con *hmfE*); *Saci\_2064*, aldehído oxidasa / xantina deshidrogenasa (funcionalmente comparable con *hmfABC*). Sobre la base de este análisis está justificado esperar que el producto proteico del gen *Saci\_2058* de *Sulfolobus acidocaldarius* DSM 639 (SEC ID N° 4) tenga efectos similares a los de HmfT1 (SEC ID N° 1) y el producto proteico del gen con el locus *mrad2831\_4728* de *Methylobacterium radiotolerans* JCM 2831 (= ATCC 27329 = DSM 1819) (SEC ID N° 3).

El primer polipéptido puede estar codificado por una primera secuencia de polinucleótidos que tiene al menos 45 %, preferentemente al menos 60 %, tal como al menos 70 %, más preferentemente al menos 80 %, tal como 90 %, más preferentemente al menos 95 % de similitud de secuencia con una secuencia de polinucleótidos expuesta en la SEC ID N° 7, 8, 9 o 10. Niveles alternativos adecuados de similitud del primer polinucleótido con una secuencia expuesta en la SEC ID N° 7, 8, 9 o 10 pueden ser al menos 66 %, 70 %, 75 %, 80 %, 85 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 % o al menos 99 % de similitud. De acuerdo con una realización, los porcentajes de similitud indicados pueden ser porcentajes de identidad. En una realización concreta, el primer polipéptido puede estar codificado por una secuencia de polinucleótidos tal como se expone en las SEC ID N° 7, 8, 9 o 10.

El primer polipéptido o el polinucleótido que codifica el primer polipéptido se pueden aislar de un organismo, preferentemente un microorganismo que exprese el primer polipéptido en ciertas condiciones de crecimiento. El microorganismo puede ser capaz de utilizar ácido HMF o sustancias furánicas relacionadas, tales como HMF o alcohol HMF, como fuente de carbono, pero esto no es necesario.

En un enfoque típico, se pueden cribar las bibliotecas de genes para aislar polinucleótidos alternativos que son adecuados. Las bibliotecas se pueden construir a partir de microorganismos del superreino de bacterias. Estos microorganismos pueden pertenecer al filo de las proteobacterias, más específicamente a la clase de Alphaproteobacterias o Betaproteobacterias. Las alphaproteobacterias pueden pertenecer al orden de Rhizobiales, a las familias de Bradyrhizobiaceae o Methylobacteriaceae. Bradyrhizobiaceae puede pertenecer al género de *Bradyrhizobium*, por ejemplo *Bradyrhizobium japonicum* o al género de *Afixia*. La familia Methylobacteriaceae puede pertenecer al género de *Methylobacterium*, por ejemplo *Methylobacterium nodulans* o *Methylobacterium radiotolerans*. Las Betaproteobacterias pueden pertenecer al orden de Burkholderiales, más específicamente, la familia de Burkholderiaceae. Pueden pertenecer al género *Cupriavidus*, por ejemplo *Cupriavidus basilensis*; o al género *Ralstonia*, por ejemplo *Ralstonia eutropha*; o al género *Burkholderia*, por ejemplo *Burkholderia phymatum*, *Burkholderia phytofirmans*, *Burkholderia xenovorans* o *Burkholderia graminis*. Las bacterias también pueden pertenecer al filo de Firmicutes, más concretamente a la clase de Bacilos, más específicamente el orden de Bacillales. El orden Bacillales puede pertenecer a la familia de Bacillaceae, más específicamente al género *Geobacillus*, por ejemplo, *Geobacillus kaustophilus*. Como alternativa, los microorganismos pueden pertenecer al superreino de Archaea, más concretamente al filo de Euryarchaeota o al filo de Crenarchaeota. El filo Euryarchaeota puede pertenecer a un género sin clasificar, por ejemplo *Cand. Parvarchaeum acidiphilum* o a la clase de Thermoplasmata, más específicamente al orden de Thermoplasmatales. El orden Thermoplasmatales puede pertenecer a la familia de Thermoplasmataceae, más específicamente al género *Thermoplasma*, por ejemplo *Thermoplasma acidophilum* o *Thermoplasma volcanium*. Las Crenarchaeota pueden pertenecer a la clase de Thermoprotei, más específicamente al orden de Sulfolobales. Los Sulfolobales pueden pertenecer a la familia de Sulfolobaceae, más específicamente al género *Sulfolobus*, por ejemplo, *Sulfolobus acidocaldarius*, *Sulfolobus islandicus*, *solfataricus Sulfolobus* o *Sulfolobus tokodaii*; o al género de *Metallosphaera*, por ejemplo, *Metallosphaera sedula*. Los Thermoprotei también pueden pertenecer al orden de Thermoproteales, familia de Thermoproteaceae. La familia Thermoproteaceae puede pertenecer al género *Vulcanisaeta*, por ejemplo *Vulcanisaeta distributa*; o al género *Caldivirga*, por ejemplo *Caldivirga maquilingsis*.

Preferentemente, el primer polipéptido y / o el primer polinucleótido que codifica el primer polipéptido se aísla de *Cupriavidus Basilensis* HMF14 (Wierckx y col., 2010). De acuerdo con una realización alternativa, el primer polipéptido y / o el primer polinucleótido que codifica el primer polipéptido se pueden aislar de *Methylobacterium radiotolerans*.

En base a las secuencias de aminoácidos proporcionadas en las SEC ID N° 1, 2, 3 o 4 y/o las secuencias de nucleótidos proporcionadas en las SEC ID N° 7, 8, 9 o 10, el experto en la técnica será capaz de construir sondas y / o cebadores adecuados para aislar una secuencia de nucleótidos que codifica el primer polipéptido.

Como alternativa, en base a las secuencias de aminoácidos proporcionadas en las SEC ID N° 1, 2, 3 o 4 y/o las secuencias de nucleótidos proporcionadas en las SEC ID N° 7, 8, 9 o 10, el experto en la técnica puede obtener secuencias sintetizadas que codifican el primer polipéptido a partir de fuentes comerciales, ya que la síntesis de genes está cada vez más disponible. Las secuencias sintéticas se pueden adquirir de, por ejemplo, Geneart A.G. (Regensburg, Alemania) o de Genscript USA Inc. (Piscataway, NJ,EE.UU.) por citar algunos.

La célula de acuerdo con la invención se modifica genéticamente mediante la introducción funcional del primer polinucleótido. Con la introducción funcional de un polinucleótido se entiende una introducción de dicho polinucleótido en una célula, de tal manera que dicha célula adquiere la posibilidad de expresar un producto polipeptídico funcional del polinucleótido. Los procedimientos y técnicas para la introducción funcional de polinucleótidos en células huésped están dentro del conocimiento general del experto en la técnica.

### Polipéptido convertidor de ácido HMF

La célula modificada genéticamente de acuerdo con la invención comprende un segundo polinucleótido que codifica un segundo polipéptido. El segundo polipéptido tiene la actividad de conversión de ácido HMF y se puede seleccionar de cualquier polipéptido capaz de convertir ácido HMF en un producto. Los inventores han observado que la bioconversión de ácido HMF por el segundo polipéptido convertidor de ácido HMF mejora de manera efectiva mediante la expresión del primer polipéptido en una célula.

El segundo polinucleótido que codifica el segundo polipéptido puede ser un componente natural de la célula de acuerdo con la invención, es decir la célula no necesita ser modificada genéticamente en relación con el segundo polinucleótido. Sin embargo, de acuerdo con ciertas realizaciones de la invención, la célula de acuerdo con la invención se modifica genéticamente en relación con el segundo polinucleótido, mediante la introducción funcional del segundo polinucleótido. la expresión "introducción funcional" ya se ha explicado anteriormente en conexión con el primer polipéptido y el primer polinucleótido.

De acuerdo con una realización preferida de la invención, el segundo polipéptido es una oxidoreductasa convertidora de ácido HMF. Esta oxidoreductasa convertidora de ácido HMF puede comprender una secuencia de aminoácidos tal como se expone en la SEC ID N° 5 o 6 o un polipéptido variante de la misma que tiene al menos 45 %, preferentemente al menos 60 %, tal como al menos 70 %, más preferentemente al menos 80 %, tal como 90 %, más preferentemente al menos 95 % de similitud de secuencia con la secuencia de aminoácidos expuesta en la SEC ID N° 5 o 6.

Como alternativa, la similitud de secuencia se puede expresar como al menos 45 %, 50 %, 55 %, 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 % o al menos 99 % de similitud. De acuerdo con una realización, los porcentajes de similitud indicados pueden ser porcentajes de identidad. En una realización concreta, el segundo polipéptido puede comprender la secuencia de aminoácidos tal como se expone en cualquiera de las SEC ID N° 5 o 6.

La oxidoreductasa convertidora de ácido HMF de la SEC ID N° 5 se divulgó previamente en Koopman y col., (2010a) y Koopman y col., (2010b) y se designó HmfH. La oxidoreductasa convertidora de ácido HMF de la SEC ID N° 6 se puede aislar de *Bradyrhizobium japonicum* USDA 110 y corresponde al producto proteico traducido del gen *blr0367*. *B. japonicum* USDA 110 contiene homólogos para todos los genes de utilización de HMF/furfural de *C. basilensis* HMF14 en su genoma (Koopman y col., 2010a). El producto traducido del gen *blr0367* es la proteína de *B. japonicum* USDA 110 que mostró la homología más alta con HmfH, de *C. basilensis* HMF14. En vista del hecho de que se ha demostrado que *B. japonicum* USDA 110 utiliza HMF como la única fuente de carbono (Koopman y col., 2010a) debe albergar un homólogo funcional de HmfH. Por tanto, está justificado esperar que una actividad de HmfH similar surja de *blr0367*.

Aunque la invención se ejemplifica con referencia a un número de oxidoreductasas convertidoras de ácido HMF, cabe señalar que dentro de la invención está permitido expresamente que el segundo polipéptido es una oxidoreductasa convertidora de ácido HMF diferente o aún un polipéptido convertidor de ácido HMF diferente no que tiene actividad de oxidoreductasa.

El segundo polipéptido puede estar codificado por una segunda secuencia de polinucleótidos que tiene al menos 45 %, preferentemente al menos 60 %, tal como al menos 70 %, más preferentemente al menos 80 %, tal como 90 %, más preferentemente al menos 95 % de similitud de secuencia con una secuencia de polinucleótidos expuesta en la SEC ID N° 11 o 12. Niveles alternativos adecuados de similitud del primer polinucleótido con una secuencia expuesta en la SEC ID N° 7, 8, 11 o 12 pueden ser al menos 66 %, 70 %, 75 %, 80 %, 85 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 % o al menos 99 % de similitud. En una realización, los porcentajes de similitud indicados pueden ser porcentajes de identidad. En una realización concreta, el segundo polipéptido puede estar codificado por una secuencia de polinucleótidos tal como se expone en las SEC ID N° 7, 8, 11 o 12.

El aislamiento de un polinucleótido adecuado que codifica el segundo polipéptido que tiene actividad oxidoreductasa de *Cupriavidus Basilensis* se ha divulgado en Koopman y col., (2010a). Este gen se designa *hmfH*.

En un enfoque para el aislamiento del segundo polinucleótido, se pueden cribar bibliotecas génicas para aislar polinucleótidos que son adecuados. Las bibliotecas se pueden construir a partir de microorganismos del superreino de bacterias. Estos microorganismos pueden pertenecer al filo de las proteobacterias, más específicamente a la clase de Alfaproteobacterias, Betaproteobacterias o Gammaproteobacterias. Las alfaproteobacterias pueden pertenecer al orden de Rhizobiales o al orden Sphingomonadales. El orden Rhizobiales puede pertenecer a la familia de Methylobacteriaceae, por ejemplo, un organismo del género *Methylobacterium* como *M. nodulans* o *M. radiotolerans* o un organismo de la familia de Rhizobiaceae. La familia Rhizobiaceae puede pertenecer al grupo de

Rhizobium/Agrobacterium, más específicamente al género *Rhizobium*, tal como *R. leguminosarum* o *R. leguminosarum* *bv. trifolii*. También pueden pertenecer al género *Agrobacterium*, tal como *A. radiobacter*. Rhizobiaceae también puede pertenecer a la familia de Bradyrhizobiaceae, más específicamente al género de Bradyrhizobium, tal como *B. japonicum*. Sphingomonadales puede pertenecer a la familia de Sphingomonadaceae, más específicamente al género de *Sphingomonas*, tales como *S. wittichii* o *S. chlorophenolicum*. Las betaproteobacterias pueden pertenecer al orden de Methylophilales o al orden Burkholderiales. El orden Methylophilales puede pertenecer a la familia de Methylophilaceae, por ejemplo, un organismo del género *Methylovorus*. Burkholderiales puede pertenecer a la familia de Burkholderiaceae, por ejemplo, un organismo del género *Cupriavidus*, tal como *Cupriavidus basilensis*. También pueden pertenecer al género *Burkholderia*, tal como *Burkholderia phytotransformans*, *B. phymatum*, *B. graminis*, *B. xenovorans* o *B. cenocepacia* o a la familia de Oxalobacteraceae, género de *Janthinobacterium*. Las Gammaproteobacterias pueden pertenecer al orden de Enterobacteriales, familia de Enterobacteriaceae, género de *Yersinia* tal como *Yersinia ruckeri*. Los microorganismos pueden además ser bacterias del filo de Actinobacteria, clase de Actinobacteria, subclase de Actinobacteridae orden de Actinomycetales. El orden Actinomycetales puede pertenecer al suborden de Streptomycineae, Pseudonocardineae o Micromonosporineae. El suborden de Sphingomonadales puede pertenecer a la familia de Streptomycetaceae, más específicamente al género de *Streptomyces*, tales como *S. violaceusniger*, *S. hygrosopicus* o *S. clavuligerus*. Pseudonocardineae puede pertenecer a la familia de Pseudonocardiaceae, más específicamente al género de *Saccharopolyspora* tales como *S. erythraea*; o a la familia de Actinosynnemataceae, más específicamente al género de *Saccharothrix* tales como *S. mutabilis* o *S. mutabilis* subsp. *capreolus* o al género *Actinosynnema*, tal como *A. Mirum*. Las Micromonosporineae pueden pertenecer a la familia de Micromonosporaceae, más específicamente al género de *Micromonospora*.

En base a las secuencias de aminoácidos proporcionadas en las SEC ID N° 5 o 6 y/o las secuencias de nucleótidos proporcionadas en las SEC ID N° 11 o 12, el experto en la técnica será capaz de construir sondas y / o cebadores adecuados para aislar una secuencia de nucleótidos que codifica el segundo polipéptido.

Como alternativa, en base a las secuencias de aminoácidos proporcionadas en las SEC ID N° 5 o 6 y/o las secuencias de nucleótidos proporcionadas en las SEC ID N° 7, 8, 11 o 12, el experto en la técnica puede obtener secuencias sintetizadas que codifican el segundo polipéptido a partir de fuentes comerciales como ya se ha indicado en la sección que trata del primer polipéptido.

### Tercer polinucleótido y polipéptido

De acuerdo con una realización alternativa, la célula modificada genéticamente de acuerdo con la invención comprende un tercer polinucleótido que codifica un tercer polipéptido. El tercer polipéptido comprende una secuencia de aminoácidos que tiene al menos 45 %, preferentemente al menos 60 %, tal como al menos 70 %, más preferentemente al menos 80 %, tal como 90 %, más preferentemente al menos 95 % de similitud de secuencia con una secuencia de aminoácidos de la SEC ID N° 19, 20, 21, 22, 23, 24 o 25.

Como alternativa, la similitud de secuencia se puede expresar como al menos 45 %, 50 %, 55 %, 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 % o al menos 99 % de similitud. De acuerdo con una realización, los porcentajes de similitud indicados pueden ser porcentajes de identidad. En una realización concreta, el primer polipéptido puede comprender la secuencia de aminoácidos tal como se expone en cualquiera de las SEC ID N° 19, 20, 21, 22, 23, 24 o 25. La secuencia de aminoácidos de SEC ID N° 19 es una selección preferida del tercer polipéptido. Esta secuencia de aminoácidos se publicó recientemente en el documento WO2011 / 02690 (SEC ID N° 15).

La expresión funcional de dicho tercer polipéptido tiene como resultado una actividad de aldehído deshidrogenasa (Adh) capaz de convertir aldehídos furánicos y proporciona una mejora adicional con respecto a la bioconversión de ácido HMF y / o la producción de FDCA.

Los efectos asociados con la expresión del tercer polipéptido se han demostrado en los ejemplos para la secuencia de aminoácidos de SEC ID N° 19. Sobre la base del nivel de similitud / identidad de secuencia está justificado esperar que los polipéptidos de las SEC ID 20-25 y sus análogos / homólogos tendrán efectos similares.

El tercer polipéptido puede estar codificado por una tercera secuencia de polinucleótidos que tiene al menos 45 %, preferentemente al menos 60 %, tal como al menos 70 %, más preferentemente al menos 80 %, tal como 90 %, más preferentemente al menos 95 % de similitud de secuencia con una secuencia de polinucleótidos expuesta en la SEC ID N° 26, 27, 28, 29, 30, 31 o 32. Preferentemente, el tercer polipéptido está codificado por una secuencia de polinucleótidos expuesta en la SEC ID N° 26 o un homólogo que tiene la similitud de secuencia indicada con la secuencia de polinucleótidos expuesta en la SEC ID N° 26. La secuencias de polinucleótidos de SEC ID N° 26 se publicó recientemente en el documento WO2011 / 026906 (SEC ID N° 16). Niveles alternativos adecuados de similitud del tercer polinucleótido con una secuencia expuesta en la SEC ID N° 27, 28, 29 o 32 pueden ser al menos 66 %, 70 %, 75 %, 80 %, 85 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 % o al menos 99 % de similitud. De acuerdo con una realización, los porcentajes de similitud indicados pueden ser porcentajes de identidad. En una realización concreta, el tercer polipéptido puede estar codificado por una secuencia de polinucleótidos tal como se expone en las SEC ID N° 26, 27, 28, 29, 30, 31 o 32.

El aislamiento y manipulación adicional del la tercer polipéptido y el correspondiente tercero polinucleótido se pueden realizar, en general, como se ha expuesto anteriormente y de aquí en adelante para el primer polipéptido y el primer polipéptido correspondiente.

### Vectores

5 Otro aspecto de la invención se refiere a vectores, incluyendo vectores de clonación y de expresión, que comprenden el primero y segundo polinucleótido o un equivalente funcional de los mismos y procedimientos de crecimiento, que transforman o transfectan tales vectores en una célula huésped adecuada, por ejemplo en condiciones en las que se produce la expresión de un polipéptido de la invención. Como se usa en el presente documento, el término "vector" se refiere a una molécula de ácido nucleico capaz de transportar otro ácido nucleico al que se ha unido.

10 El primero y segundo polinucleótido y opcionalmente, el tercer polinucleótido, se pueden incorporar en un vector replicable recombinante, por ejemplo un vector de clonación o expresión. El vector se puede usar para replicar el ácido nucleico en una célula huésped compatible. Por tanto, en una realización adicional, la invención proporciona un procedimiento de fabricación de polinucleótidos de la invención introduciendo un polinucleótido de la invención en un vector replicable, introduciendo el vector en una célula huésped compatible y cultivando la célula huésped en condiciones que produzcan la replicación del vector. El vector puede recuperarse de la célula huésped. Las células huésped adecuadas se describen a continuación.

15 El vector en el que se inserta el casete de expresión o polinucleótido de la invención puede ser cualquier vector que pueda someterse de forma conveniente a procedimientos de ADN recombinante y la elección del vector a menudo dependerá de la célula huésped en la que tiene que introducirse.

Un vector de acuerdo con 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, por ejemplo, un plásmido. Como alternativa, el vector puede ser uno que, cuando se introduce en una célula huésped, se integra en el genoma de la célula huésped y se replica junto con el(los) cromosoma(s) en que se ha integrado.

25 Un tipo de vector es un "plásmido", que se refiere a un bucle de ADN circular bicatenario en el que se pueden ligar segmentos adicionales de ADN. Otro tipo de vector es un vector viral, En el que se pueden ligar en el genoma viral segmentos adicionales de ADN. Ciertos vectores son capaces de una replicación autónoma en una célula huésped en la que se introducen (p. ej., vectores bacterianos que tienen un origen de replicación bacteriano y vectores episomales de mamífero). Otros vectores (por ejemplo, vectores no episomales de mamífero) se integran en el genoma de una célula huésped tras la introducción en la célula huésped y, de este modo, se replican junto con el genoma del huésped. Además, ciertos vectores son capaces de dirigir la expresión de los genes a los que están unidos operativamente. Dichos vectores se denominan en el presente documento "vectores de expresión". En general, los vectores de expresión de utilidad en técnicas de ADN recombinante suelen estar en forma de plásmidos. Los términos "plásmido" y "vector" se pueden usar de forma intercambiable en el presente documento, ya que el plásmido es la forma de vector usada con más frecuencia. No obstante, la invención está destinada a incluir esas otras formas de vectores de expresión, tales como cósmidos, vectores virales (p. ej., retrovirus defectuosos en la replicación, adenovirus y virus adenoasociados), vectores fagos y transposones y plasposones que sirven funciones equivalentes.

40 El experto será capaz de construir los vectores de acuerdo con la invención basándose en las secuencias de aminoácidos y polinucleótidos proporcionadas, su conocimiento de la técnica y los medios disponibles comercialmente.

De acuerdo con una realización preferida, la primera y segunda secuencia de polinucleótidos se localizan en un único vector. Este vector comprende opcionalmente además la tercera secuencia de polinucleótidos. Como el experto en la técnica comprenderá, el uso de un vector que comprende la primera y segunda secuencia de polinucleótidos (y que opcionalmente comprende además la tercera secuencia de polinucleótidos) simplifica en gran medida la construcción de células genéticamente modificadas que expresan funcionalmente el primero y segundo polipéptido (opcionalmente junto con el tercer polipéptido). Sin embargo, como el experto en la materia también entenderá, las células modificadas genéticamente que expresan funcionalmente el primero y segundo polipéptido se pueden obtener a través de varios otros esquemas de transformación que implican vectores alternativos. A este respecto, hay que señalar que de acuerdo con ciertas realizaciones, la introducción funcional de la segunda secuencia de polinucleótidos no es un requisito. También de acuerdo con ciertas formas de realización alternativas adicionales, la introducción funcional de la tercera secuencia de polinucleótidos no es un requisito.

### Célula huésped

55 La célula modificada genéticamente de acuerdo con la invención se puede construir a partir de cualquier célula huésped adecuada. La célula huésped puede ser una célula no modificada o puede estar ya modificada genéticamente. La célula puede ser una célula procariota, una célula eucariota, una célula vegetal o una célula animal. En una célula de este tipo, uno o más genes pueden delecionarse, inactivarse o romperse en su totalidad o en parte, en la que opcionalmente uno o más genes codifican la proteasa. De acuerdo con una realización, la célula

huésped de acuerdo con la invención es una célula huésped eucariota. Preferentemente, la célula eucariota es una célula de mamífero, insecto, vegetal, fúngica o de alga. Las células de mamífero preferidas incluyen, por ejemplo, células de ovario de hámster chino (CHO), células COS, células 293, células PerC6, e hibridomas. Las células de insecto preferidas incluyen, por ejemplo, células Sf9 y Sf21 y derivadas de las mismas. Más preferentemente, la célula eucariota es una célula fúngica, es decir, una célula de levadura, tal como cepas de *Candida*, *Hansenula*, *Kluyveromyces*, *Pichia*, *Saccharomyces*, *Schizosaccharomyces* o *Yarrowia*. Más preferentemente, la célula eucariota es *Kluyveromyces lactis*, *Saccharomyces cerevisiae*, *Hansenula polymorpha*, *Yarrowia lipolytica*, *Pichia stipitis* y *Pichia pastoris* o una célula fúngica filamentosa. En ciertas formas de realización, la célula eucariota huésped es una célula fúngica filamentosa.

Los "hongos filamentosos" incluyen todas las formas filamentosas de la subdivisión *Eumycota* y *Oomycota* (como definen Hawksworth y col., (1995)). Los hongos filamentosos se caracterizan por una pared micelial compuesta por quitina, celulosa, glucano, quitosano, manano y otros polisacáridos complejos. El crecimiento vegetativo es mediante alargamiento de la hifa y el catabolismo del carbono es estrictamente aeróbico. Las cepas de hongos filamentosos incluyen, pero no se limitan a, las cepas de *Acremonium*, *Agaricus*, *Aspergillus*, *Aureobasidium*, *Chrysosporium*, *Coprinus*, *Cryptococcus*, *Filibasidium*, *Fusarium*, *Humicola*, *Magnaporthe*, *Mucor*, *Myceliophthora*, *Neocallimastix*, *Neurospora*, *Paecilomyces*, *Penicillium*, *Piromyces*, *Phanerochaete*, *Pleurotus*, *Schizophyllum*, *Talaromyces*, *Thermoascus*, *Thielavia*, *Tolypocladium* *Trichoderma*.

Las células de hongos filamentosos preferidas pertenecen a una especie de un género *Aspergillus*, *Chrysosporium*, *Penicillium*, *Talaromyces* o *Trichoderma* y, lo más preferentemente, una especie seleccionada de *Aspergillus niger*, *Aspergillus awamori*, *Aspergillus foetidus*, *Aspergillus sojae*, *Aspergillus fumigatus*, *Talaromyces emersonii*, *Aspergillus oryzae*, *Chrysosporium lucknowense*, *Trichoderma reesei* o *Penicillium chrysogenum*.

De acuerdo con otra realización, la célula huésped de acuerdo con la invención es una célula huésped procariota. Preferentemente, la célula huésped procariota es una célula bacteriana. El término "célula bacteriana" incluye microorganismos tanto gramnegativos como grampositivos. Las bacterias adecuadas pueden seleccionarse de, por ejemplo, los géneros *Escherichia*, *Anabaena*, *Caulobacter*, *Gluconobacter*, *Rhodobacter*, *Pseudomonas*, *Paracoccus*, *Bacillus*, *Brevibacterium*, *Corynebacterium*, *Rhizobium* (*Sinorhizobium*), *Bradyrhizobium*, *Flavobacterium*, *Klebsiella*, *Enterobacter*, *Lactobacillus*, *Lactococcus*, *Methylobacterium*, *Staphylococcus*, *Streptomyces*, *Zymomonas*, *Acetobacter*, *Streptococcus*, *Bacteroides*, *Selenomonas*, *Megasphaera*, *Burkholderia*, *Cupriavidus*, *Ralstonia*, *Methylobacterium*, *Methylovorus*, *Rhodopseudomonas*, *Acidiphilium*, *Dinoroseobacter*, *Agrobacterium*, *Sulfolobus* o *Sphingomonas*. Preferentemente, la célula bacteriana se selecciona del grupo que consiste en *Bacillus subtilis*, *Bacillus amyloliquefaciens*, *Bacillus licheniformis*, *Bacillus pumilus*, *Bacillus megaterium*, *Bacillus halodurans*, *Bacillus pumilus*, *Gluconobacter oxydans*, *Caulobacter crescentus*, *Methylobacterium extorquens*, *Methylobacterium radiotolerans*, *Methylobacterium nodulans*, *Rhodobacter sphaeroides*, *Pseudomonas zeaxanthinifaciens*, *Pseudomonas putida*, *Pseudomonas putida* S12, *Paracoccus denitrificans*, *Escherichia coli*, *Corynebacterium glutamicum*, *Staphylococcus carnosus*, *Streptomyces lividans*, *Sinorhizobium meliloti*, *Bradyrhizobium japonicum*, *Rhizobium radiobacter*, *Rhizobium leguminosarum*, *Rhizobium leguminosarum* bv. *trifolii*, *Agrobacterium radiobacter*, *Cupriavidus basilensis*, *Cupriavidus necator* (*Ralstonia eutropha*), *Ralstonia pickettii*, *Burkholderia phytofirmans*, *Burkholderia phymatum*, *Burkholderia xenovorans*, *Burkholderia graminis*, *Rhodopseudomonas palustris*, *Acidiphilium cryptum*, *Dinoroseobacter shibae*, *Sulfolobus acidocaldarius*, *Sulfolobus islandicus*, *Sulfolobus solfataricus*, *Sulfolobus tokodaii*.

Una célula huésped altamente preferida es *Pseudomonas putida* S12. En esta cepa, se ha demostrado que la expresión funcional del gen *hmfH* de *Cupriavidus basilensis* HMF14 es eficaz para introducir capacidad oxidativa de HMF, lo que tiene como resultado la producción de FDCA a partir de este sustrato.

Para usos específicos de la célula de acuerdo con la invención, la selección de la célula huésped puede hacerse de acuerdo con dicho uso. Son particularmente preferidos los huéspedes que son adecuados para la conversión de materias primas lignocelulósicas y las que son resistentes a las condiciones preferidas para la producción de compuestos furánicos, tales como FDCA. El experto en la técnica tendrá que, según sus medios y procedimientos adecuados de disponibilidad, introducir funcionalmente el primero y opcionalmente el segundo polinucleótido en cualquiera de las células huésped mencionadas.

## 50 Biotransformación de HMF ácido

La célula manipulada genéticamente de acuerdo con la invención tiene una biotransformación de ácido HMF mejorada. La mejora de la bioconversión de ácido HMF es beneficiosa para la eliminación de ácido HMF y sus precursores furánicos a partir de materias primas, en las que los compuestos furánicos se consideran perjudiciales, tales como materias primas para fermentaciones etanológicas para la producción de biocombustibles y productos bioquímicos. En otras aplicaciones, la mejora de la bioconversión de ácido HMF mejorará la bioproducción de una sustancia química en la que el ácido HMF es un material de partida o un intermedio, tal como en la bioproducción de FDCA.

Si la enzima convertidora de ácido HMF es oxidorreductasa convertidora de ácido HMF, la célula de acuerdo con la invención será capaz de realizar una reacción de oxidación biológica. La reacción de oxidación es, en el presente documento, una o más reacciones de un oxidante con ácido HMF en presencia de la oxidorreductasa.

5 Una reacción de oxidación preferida es la producción de FDCA, en la que el ácido HMF se convierte en FDCA, mediante la reacción con un oxidante en presencia de una oxidorreductasa convertidora de ácido HMF. Las bioconversiones de compuestos furánicos a FDCA, en las que el ácido HMF es un compuesto intermedio, se han divulgado en la técnica anterior, por ejemplo usando HMF como material de partida (véase Koopman y col., 2010a y Koopman y col., (2010b). Tales bioconversiones se mejorarán si las realizan una célula de acuerdo con la invención.

10 El ácido HMF se puede generar *in situ* a partir de uno o más precursores furánicos por la célula de la invención o cualquier otra célula presente. Con la generación *in situ* se entiende que el ácido HMF no se añade desde fuera del sistema. En su lugar, se genera ácido HMF dentro del sistema a través de una o más bioconversiones que convierten los precursores furánicos en ácido HMF.

15 El precursor furánico de ácido HMF puede seleccionarse del grupo que consiste en 5- (hidroximetil)furan-2-carbaldehído (HMF), furan-2,5-dicarbaldehído (DFF) y [5- (hidroximetil)furan-2- il]metanol (alcohol HMF) y preferentemente el precursor furánico es HMF,

HMF se puede obtener a partir de uno o más azúcares hexosa mediante deshidratación catalizada por ácido, como se conoce en la técnica. Los azúcares hexosa pueden obtenerse a partir de biomasa, preferentemente biomasa lignocelulósica.

20 La reacción de oxidación puede comprender un número de etapas consecutivas de reacción de oxidación que tienen como resultado un producto, por ejemplo la oxidación de ácido HM en FFA y la posterior oxidación de FFA en FDCA. Ejemplos de las reacciones de oxidación se proporcionan en la Fig. 1.

25 Las reacciones de oxidación se llevan a cabo preferentemente a temperatura relativamente suave, es decir, a 10-80 °C, más preferentemente 20-45 °C, lo más preferentemente a aproximadamente 25-40 °C. Se prefiere llevar a cabo la reacción a un pH en el que FDCA está en una forma neutra o en una forma completamente disociada, de manera que la formación de sales puede controlarse. En vista de la presencia de dos restos de ácido en FDCA, hay dos intervalos de pH preferidos separados. El pH durante la reacción puede ser desde un pH de 1 hasta 6, preferentemente desde un pH de 1 a 4, lo más preferentemente desde un pH de 1 a 3. Como alternativa, el pH durante la reacción puede ser desde un pH de 5 hasta 9, preferentemente desde un pH de 1 a 8, lo más preferentemente desde un pH de 1 a 7. El experto en la técnica entenderá que los requisitos de la célula huésped también influirán en la selección de un valor de pH adecuado para el proceso. La selección de valores de pH que son adecuados para las diversas células huésped que son adecuadas dentro de la presente invención está dentro del ámbito del experto en la técnica y se puede derivar de libros de texto estándar. Para *Pseudomonas putida*, incluida *Pseudomonas putida* S12, el intervalo de pH preferido es de pH 5 a 7.

35 El tiempo de reacción puede ser de 6 a 150 horas, con la adición de oxígeno a partir de una fuente de oxígeno, tal como oxígeno molecular o agua o una fuente diferente de oxígeno en función de los requisitos de la enzima oxidante de furánico. Se puede usar aire convenientemente como una fuente de oxígeno molecular.

El reactor puede ser cualquier biorreactor (aireado) adecuado. Se puede funcionar en operación por lotes, de forma continua o discontinua.

40 Después de la biotransformación, las células pueden separarse del caldo por procedimientos establecidos y reutilizarse. Los productos de oxidación tales como FDCA, ácido HMF, etc., pueden recuperarse de la mezcla de reacción mediante precipitación (en ácido) y la posterior cristalización por enfriamiento y separación del producto de oxidación cristalizado, por ejemplo, FDCA cristalizado. Sin embargo otros métodos de recuperación son adecuados, tales como, pero sin limitaciones, precipitación ácida y extracción con disolvente, como se conoce en la técnica.

45 Para muchas aplicaciones, tales como la eliminación de ácido HMF partir de materias primas lignocelulósicas, la forma exacta de conversión de HMF es irrelevante. Lo que es importante es que el ácido HMF se convierte de forma eficaz a fin de eliminarlo como tal o para prevenir su acumulación si se forma a partir de precursores furánicos. Para este tipo de aplicaciones, el polipéptido convertidor de ácido HMF puede ser cualquier polipéptido que tenga actividad convertidora de ácido HMF actualmente conocido o aún por descubrir.

50 Un aspecto adicional de la presente invención está dirigido al uso de una célula modificada genéticamente de acuerdo con la invención, para la biotransformación de precursores furánicos a FDCA. Los precursores furánicos pueden seleccionarse en particular de 5- (hidroximetil)furan-2-carbaldehído (HMF), [5-(hidroximetil)furan-2-il]metanol (alcohol HMF), furan-2,5- dicarbaldehído (DFF), ácido 5-(hidroximetil)furan-2-carboxílico (ácido HMF) o ácido 5-formilfuran-2-carboxílico (FFA). Preferentemente, HMF es un precursor furánico seleccionado. El ácido HMF puede ser un intermedio en la bioconversión de HMF a FDCA.

55 La invención se ilustrará adicionalmente con referencia a los ejemplos siguientes.

## Ejemplos

### Metodología general

**Cepas y plásmidos** *Pseudomonas putida* S12 (ATCC 700801) se utilizó como huésped para la expresión de genes de *Cupriavidus basilensis* HMF14 (Wierckx y col., 2010; Koopman y col., 2010a) y *Methylobacterium radiotolerans* JCM 2831 (= ATCC 27329 = DSM 1819; secuencia del genoma disponible en <http://genome.jgi-psf.org/metra/metra.home.html>). Las cepas DH5 $\alpha$  o TOP10 (Invitrogen) de *Escherichia coli* se usaron para fines de clonación generales.

Para la expresión episómica de los genes de *C. basilensis* o *M. radiotolerans* se usaron pJT'mcs derivado de pUCP22 (Koopman y col., 2010a) o pJNNmcs(t) (Wierckx y col., 2008) o pBT'mcs derivado de pBBRIMCS (Koopman y col., 2010a). En pJT'mcs y pBT'mcs, la expresión del gen diana está dirigida por el promotor *tac* constitutivo. En pJNNmcs(t), la expresión está dirigida por el casete de expresión NagR/P<sub>nagAa</sub> inducible por salicilato.

**Medios y condiciones de cultivo** Se usó medio con sales minerales (MM) como medio definido. El MM contenía lo siguiente (por litro de agua desmineralizada): 3,88 g de K<sub>2</sub>HPO<sub>4</sub>, 1,63 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 como se ha especificado. caldo Luria (caldo L: 10 g/l de Bacto trypton (Difco), 5 g/l de extracto de levadura (Difco), 5 g/l de NaCl) se usaron como medio completo para la propagación de *P. putida* S12 y cepas derivadas, *C. basilensis*, *M. radiotolerans* y *E. coli* DH5 $\alpha$  y derivados. Para el cultivo de placas, el caldo L se solidificó con 1,5 % (p/v) de agar (Difco). Se añadió ampicilina (Amp) al medio hasta 100  $\mu$ g/ml para la selección de transformantes de *E. coli* portadores de plásmidos derivados de pJT'mcs o pJNNmcs(t). Se añadió gentamicina (Gm) a 30  $\mu$ g/ml en caldo Luria y 10  $\mu$ g/ml en medio con sales minerales para la selección de transformantes de *P. putida* S12 portadores de plásmidos derivados de pJT'mcs o pJNNmcs(t). Para la selección de transformantes de *E. coli* o *P. putida* S12 portadores de plásmidos derivados de pBT'mcs se añadió al medio 50  $\mu$ g/ml de kanamicina (Km). Los antibióticos se adquirieron en Sigma-Aldrich. *P. putida*, *C. basilensis* y *M. radiotolerans* se cultivaron a 30 °C; *E. coli* se cultivó a 37 °C

Los experimentos de alimentación discontinua con cepas derivadas de *P. putida* S12 -se realizaron en vasos de 2 l controlados por un sistema de biorreactor de Labfors 4 (Infors Benelux BV) o un controlador BioFlo110 (New Brunswick Scientific). Se suministró aire presurizado u oxígeno puro en el espacio aéreo o se roció a través del caldo. La temperatura se controló a 30 °C y el pH se mantuvo a 7,0 mediante la adición automática de NH<sub>4</sub>OH, NaOH o KOH. La fase discontinua se realizó en medio 2x MM suplementado con antibióticos específicos de la cepa y 40 g/l de glicerol. Para los cultivos de alta densidad, el medio de la fase discontinua se suplementó adicionalmente con 10 g / l de extracto de levadura (YE). Después del agotamiento del glicerol inicial, la alimentación (4 o 8 M de glicerol en MgCl<sub>2</sub> 100 , suplementado con 1 mM de salicilato sódico cuando se requiera) se inició y controló para permitir el crecimiento al tiempo que mantenía el glicerol como el sustrato limitante en el cultivo. La alimentación de HMF (4 M en agua desmineralizada) se introdujo a través de una bomba de alimentación separada; la velocidad de alimentación se ajustó dependiendo de la cepa empleada y de la afección estudiada. La tensión de oxígeno disuelto (DO) se monitorizó continuamente y la velocidad de agitación se ajustó para mantener suficiente aireación.

**Ensayos y procedimientos analíticos** Medición del peso seco de la célula (PSC) El contenido de PSC de cultivos de bacterias se determinó midiendo la densidad óptica a 600 nm (DO<sub>600</sub>) utilizando un medidor de la densidad celular Biowave (WPA Ltd) o un espectrofotómetro de microplaca universal  $\mu$  Quant MQX200 (Biotek), utilizando microplacas de 96 pocillos de fondo plano (Greiner). Una DO<sub>600</sub> de 1,0 corresponde a 0,56 g de PSC / l (Biowave) o 1,4 g PSC / l ( $\mu$ Quant) para *P. putida*.

Análisis por HPLC: Los compuestos de furano (FDCA, HMF, alcohol HMF, ácido HMF y FFA) se analizaron por RPHPLC según lo descrito por Koopman y col., (2010a). Azúcares, alcoholes y ácidos orgánicos también se analizaron mediante HPLC (sistema Agilent 1100) utilizando un detector del índice de refracción (RI). La columna usada fue Bio-Rad Aminex HPX-87H (300 x 7,8 mm, forma de hidrógeno, tamaño de partícula 9  $\mu$ m, 8 % de reticulación, intervalo de pH 1-3) con 5 mM H<sub>2</sub>SO<sub>4</sub> 5 mM como eluyente a un caudal de 0,6 ml / min.

### Productos químicos

HMF se adquirió en Sigma, Eurolabs Ltd (Poynton, Reino Unido) o Yore Chemipharm Co. Ltd. (Ningbo, China). Los patrones analíticos de FDCA y ácido 5-hidroximetil-furoico (ácido HMF) se adquirieron en Immunosource B.V. (Halle-Zoersel, Bélgica), respectivamente, Matrix Scientific (Columbia SC, EE.UU.). Todos los otros productos químicos se adquirieron en Sigma-Aldrich Chemie B.V. (Zwijndrecht, Países Bajos).

**Técnicas moleculares y genéticas:** El ADN genómico se aisló de *C. basilensis* HMF14 y *M. radiotolerans* JCM 2831 usando el kit tisular DNeasy (QIAGEN). El ADN del plásmido se aisló con el kit de minipreparación QIAprep spin (QIAGEN). Se aislaron fragmentos de ADN atrapados en agarosa con el kit de extracción en gel de QIAEXII (QIAGEN).

Las reacciones de PCR se realizaron con la polimerasa AccuPrime Pfx (Invitrogen) de acuerdo con las instrucciones de los fabricantes. Los cebadores de oligonucleótidos (especificados en los ejemplos) fueron sintetizados por MWG Biotech AG (Alemania). El ADN del plásmido se introdujo en células electrocompetentes utilizando un dispositivo de electroporación Gene Pulser (BioRad). Otras técnicas de biología molecular estándar se realizaron de acuerdo con Sambrook y Russell (2001).

#### **Ejemplo I: La coexpresión de HmfH y HmfT1 mejora la producción de FDCA en *P. putida* S12**

El gen *hmfT1* (antes llamado *mfs1* (Koopman y col., 2010a); SEC ID N° 7) se amplificó a partir de ADN genómico de *Cupriavidus basilensis* HMF14 mediante PCR usando cebadores *hmfT1*(f) (SEC ID N° 13) y *hmfT1*(r) (SEC ID N°: 14). El producto de PCR se introdujo como un fragmento de 1359 pb *EcoRI-NheI* en *pJNNmcs*(t), dando *pJNNhmfT1*(t). El gen *hmfH* (SEC ID N° 11) incluyendo su sitio de unión al ribosoma nativo (RBS) se amplificó mediante PCR del ADN genómico de *C. basilensis* HMF14 usando los cebadores FN23 (SEC ID N° 15) y FN24 (SEC ID N° 16). El producto de PCR se clonó como un fragmento de 1777 pb de *EcoRI-NheI*, dando el plásmido *pJNNhmfT*. Los plásmidos *pBT<sup>+</sup>hmfH* y *pJNNhmfT1*(t) se introdujeron sucesivamente en *P. putida* S12, dando *P. putida* S12\_B38.

*P. putida* S12\_B38 se cultivó en lotes de alimentación como se describe en la sección de metodología general. En la fase discontinua, se alcanzó una densidad celular de aproximadamente 3 g CDW / l, tras lo cual se inició la alimentación con glicerol y la alimentación con HMF. Se realizó un cultivo de alimentación discontinua control con *P. putida* S12\_2642 (similar a *P. putida* S12\_hmfH (Koopman y col., 2010b)) que no expresa HmfT1.

La Fig. 2 muestra las concentraciones de FDCA y ácido HMF en los cultivos alimentados con HMF de las cepas de *P. putida* S12\_2642 y S12\_B38. La acumulación extensa de ácido HMF es evidente para *P. putida* S12\_2642. Por el contrario, la acumulación de ácido HMF fue insignificante en el cultivo de *P. putida* S12\_B38. La reducida acumulación de ácido HMF permitió adicionalmente incrementos de las velocidades de alimentación de HMF, lo que tuvo como resultado títulos de FDCA más altos en un tiempo de proceso considerablemente más corto. Esto también se reflejó claramente en la productividad de FDCA específica para las cepas analizadas (Fig. 3).

#### **Ejemplo II La coexpresión de HmfH y un polipéptido de *Methylobacterium radiotolerans* JCM 2831 mejora la producción de FDCA en *P. putida* S12**

El gen con el locus tag *mrad2831\_4728* (SEC ID: 9) se amplificó a partir de ADN genómico de *Methylobacterium radiotolerans* JCM2831 mediante PCR usando los cebadores *Mrad*(f) (SEC ID N° 17) y *Mrad*(r) (SEC ID N° 18). El producto de PCR se introdujo como un fragmento de 1381 pb *EcoRI-NheI* en *pJNNmcs*(t), dando *pJNN\_Mrad*(t). Los plásmidos *pBT<sup>+</sup>hmfH* (véase el Ejemplo I) y *pJNN\_Mrad*(t) se introdujeron sucesivamente en *P. putida* S12, dando *P. putida* S12\_B51.

*P. putida*\_B51 se cultivó en matraces de agitación en medio con sales minerales (4 x concentración del tampón) suplementado con 1 g/l de extracto de levadura, glicerol 80 mM, glucosa 2 mM, salicilato sódico 100 µM, 50 µg/ml de kanamicina y 10 µg/ml de gentamicina.

Como cepa control se usó *P. putida* S12\_B38 (véase el Ejemplo I). Después de cultivar durante la noche, los cultivos se suplementaron con glicerol 80 mM y salicilato sódico 100 µM. Después se añadió HMF a aproximadamente 10 mM y se evaluó la producción de FDCA.

La Fig. 4 muestra que la acumulación de ácido HMF durante la producción de FDCA fue insignificante para ambas cepas, lo que confirma que HmfT1 y el polipéptido *Mrad2831\_4728* exhibían una funcionalidad similar. Para *P. putida* S12\_B38, la producción de FDCA mostró una fase log más prolongada y una velocidad algo más lenta, lo que se pudo atribuir a la menor densidad inicial de biomasa (Fig. 4A). No obstante, la productividad máxima específica de FDCA fue idéntica en ambas cepas, es decir 2,36 mmol de FDCA/(g PSC, h), lo que indica que los polipéptidos *Mrad2831\_4728* y HmfT1 eran igualmente eficaces en la minimización de la acumulación de ácido HMF y maximización de la producción de FDCA.

#### **Ejemplo III: Capacidad de conversión de ácido HMF de niveles altos por *P. putida* S12 que coexpresa HmfH y HmfT1**

Como se demuestra en el Ejemplo I, la coexpresión de HmfH y HmfT1 en *P. putida* S12 mejoró considerablemente la capacidad específica para oxidar HMF en FDCA. Para realizar un uso óptimo de esta capacidad mejorada se realizó un experimento de alimentación discontinua con *P. putida* S12\_B38 comenzando a una densidad de biomasa elevada.

La alimentación de HMF se inició a una velocidad alta (20 ml/h; solución de alimentación 4 M de HMF) con el fin de saturar la capacidad de oxidación de *P. putida* S12\_B38 y provocar la acumulación de HMF y ácido HMF (Fig. 5). Cuando se hubo acumulado ácido HMF hasta aproximadamente 20 mM, la velocidad de alimentación de HMF se redujo hasta 5 ml/h y se vigilaron las concentraciones de compuestos furánicos. Inicialmente, la concentración de ácido HMF aumentó hasta aproximadamente 37 mM debido a la oxidación de HMF acumulado residual, tras lo cual disminuyó hasta menos de 2 mM en 5 horas, a una velocidad de alimentación de 0,72 mmol/(g PSC.h).

Estos resultados demuestran claramente que la capacidad de oxidación de ácido HMF se había mejorado por la coexpresión de HmfH y HmfT1. Los resultados de Koopman y col., (2010b), mostraron que *P. putida* S12\_hmfH (que carece de HmfT1) requerían más de 50 horas para reducir la concentración de HMF ácido desde aproximadamente 50 mM a menos de 5 mM, a una velocidad de alimentación de HMF mucho menor (0,09 mmol/(g PSC.h)). La capacidad de oxidación de HMF ácido mejorada tuvo como resultado un título de FDCA final mucho más alto (152 g/l frente a 30,1 g/l por Koopman y col., (2010b)) que también se alcanzó en un tiempo de proceso más corto (94 h frente a 115 h por Koopman y col., (2010b)).

#### **Ejemplo IV: La coexpresión de HmfH, HmfT1 y una aldehído deshidrogenasa de *C. basilensis* mejora la producción de FDCA en *P. putida* S12**

El gen que codifica una aldehído deshidrogenasa (SEC ID 26; secuencia de aminoácidos traducida: SEC ID 19) asociada con el operón de degradación de HMF en *Cupriavidus basilensis* HMF14 (Wierckx y col., 2011) se amplificó mediante PCR usando los cebadores FN13 (SEC ID 33) y FN14 (SEC ID 34). El producto de la PCR se introdujo como un fragmento *NotI* de 153 pb en pBT<sup>+</sup>hmfH digerido con *Bsp120I* (compatible con *NotI*) (véase el ejemplo I), dando pBT<sup>+</sup>hmfH-adh. La variante del plásmido en la que el gen codificante de la aldehído deshidrogenasa estaba presente en orientación directa (f) (pBT<sup>+</sup>hmfH-adh(f)) y pJNNhmfT1(t) (véase el Ejemplo I) se introdujeron sucesivamente en *P. putida* S12. La cepa resultante *P. putida* S12\_B97, coexpresó HmfH, HmfT1 y la aldehído deshidrogenasa. Como cepa control para la coexpresión de la HmfH oxidorreductasa y la aldehído deshidrogenasa (es decir, sin el transportador de HMF ácido HmfT1), se construyó *P. putida* S12\_B101 que solo contenía pBT<sup>+</sup>hmfH-adh(f). *P. putida* S12-B38 (véase el Ejemplo I) se usó como cepa control para la coexpresión de HmfT1 y HmfH sin la aldehído deshidrogenasa.

Las cepas de *P. putida* S12\_B38, S12\_B97 y S12\_B101 se cultivaron en matraces de agitación en medio con sales minerales (4x la concentración del tampón) suplementado con 1 g/l de extracto de levadura, glicerol 80 mM, glucosa 2 mM, kanamicina 50 µg/ml y 10 µg/µl de gentamicina (nota: para la cepa B101 solo se añadió kanamicina). Se añadió salicilato sódico (1 µM) para la inducción de *hmfT1* solo en los precultivos. Tras la adición de HMF aproximadamente 10 mM se evaluó la acumulación de FDCA y ácido HMF.

En la cepa que coexpresó HmfH (oxidorreductasa) y HmfT1 (transportador de ácido HMF) (cepa S12\_B38; Fig. 6A), la producción de FDCA solo comenzó después de haber acumulado FFA hasta un nivel sustancial. El ácido HMF se acumuló de forma transitoria a cantidades bajas, como se ha observado anteriormente (véase el Ejemplo II). Cuando la aldehído deshidrogenasa se coexpresó con HmfH y HmfT1 (cepa S12\_B97; Fig. 6B), la formación de FDCA comenzó sin retraso y se observaron tanto FFA como ácido HMF únicamente en cantidades de restos. La coexpresión de la aldehído deshidrogenasa y HmfH sin HmfT1 (cepa S12\_B101; Fig. 6C), tuvo como resultado una amplia acumulación de ácido HMF, mientras que solo se produjeron cantidades pequeñas de FFA y FDCA.

Los resultados demostraron que la oxidación de HMF en ácido HMF se ve significativamente potenciada por la expresión de la aldehído deshidrogenasa. No obstante, HmfT1 debe coexpresarse para permitir una biotransformación eficiente del ácido HMF producido. La aldehído deshidrogenasa mejoró aún más la oxidación del producto intermedio FFA en FDCA. Por tanto, la expresión simultánea de la aldehído deshidrogenasa y HmfT1 mejora considerablemente el potencial global y la velocidad de, la oxidación de HMF a través de ácido HMF hasta el producto final.

#### **REFERENCIAS**

- Almeida y col.,(2009) Metabolic effects of furaldehydes and impacts on biotechnological processes. Applied Microbiology and Biotechnology , 82 (4): 625-638.
- Altschul y col., (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res, 25 (17): 3389-3402.
- Altschul, y col., (1990) Basic local alignment search tool. J. Mol. Biol., 215: 403-10. Ausubel y col., (eds.) (1995), Current Protocols in Molecular Biology, (John Wiley & Sons, N.Y.
- Hawksworth y col., (1995) In: Ainsworth and Bisby's Dictionary of The Fungi, 8ª edición, CAB International, University Press, Cambridge, UK.
- Koopman y col., (2010a) Identification and characterization of the furfural and 5-(hydroxymethyl)furfural degradation pathways of *Cupriavidus basilensis* HMF14. PNAS (2010a), Vol. 107, p. 4919-4924.
- Koopman y col., (2010b) Efficient whole-cell biotransformation of 5-(hydroxymethyl)furfural into FDCA, 2,5-furandicarboxylic acid. Bioresour. Technol. (2010b), Vol. 101, p. 6291-6296.
- Meyers y col., (1988) Optimal alignments in linear space CABIOS, 4:11-17.
- Needleman y col., (1970) A general method applicable to the search for similarities in the amino acid sequences of two proteins. J. Mol. Biol. (48): 444-453.

Nichols NN, Mertens JA. Identification and transcriptional profiling of *Pseudomonas putida* genes involved in furoic acid metabolism (2008) FEMS Microbiol Lett 284: 52-57.

Sambrook y col., (1989), Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, N.Y.

5 Sambrook y col., (2001) Molecular cloning - a laboratory manual. Tercera edición. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Werpy y col., (2004) Top Value-Added Chemicals from Biomass, Volume I - Results of screening for potential Candidates from Sugars and Synthesis gas.

Wierckx y col., (2008) Transcriptome analysis of a phenol producing *Pseudomonas putida* S12 construct: genetic and physiological basis for improved production. J Urol 190: 2822-2830.

10 Wierckx y col., (2010) Isolation and characterization of *Cupriavidus basilensis* HMF14 for biological removal of inhibitors from lignocellulosic hydrolysate. Microb Biotechnol 3: 336-343.

Wierckx y col., (2011) Microbial degradation of furanic compounds: biochemistry, genetics, and impact. Appl Microbiol Biotechnol 92:1095-1105

#### LISTADO DE SECUENCIAS

15 <110> Bird Engineering B.V.

<120> CÉLULA GENÉTICAMENTE MODIFICADA Y PROCEDIMIENTO PARA USAR DICHA CÉLULA

<130> 1.848.001 PTWO

<150> NL2006359

<151> 2011-03-08

20 <160> 34

<170> PatentIn versión 3.5

<210> 1

<211> 441

<212> PRT

25 <213> *Cupriavidus basilensis* HMF14

<400> 1

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  
 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

ES 2 543 568 T3

					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				
Pro	Ile	Leu	Ile	Phe	Leu	Asn	Glu	Arg	Val	Ser	His	Gln	His	Pro	Cys	
	370					375					380					
Tyr	Gly	Asn	Trp	Pro	Val	Met	Glu	Tyr	Arg	Leu	Cys	His	Arg	Arg	His	
385					390					395					400	
Asp	Ala	Thr	Phe	Ala	Ser	Leu	Cys	Ala	Ala	Pro	Pro	Arg	Leu	Pro	Lys	
				405					410					415		

ES 2 543 568 T3

Cys Trp Gly Ser Arg Arg Gly Val Lys Ala Phe Thr Ala Gly Ala Ala  
 420 425 430

Ile Val Trp Asn Ala Pro Leu Gly Glu  
 435 440

<210> 2

<211> 449

<212> PRT

5 <213> Cupriavidus basilensis HMF14

<400> 2

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

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  
 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> 3

<211> 449

<212> PRT

5 <213> Methylobacterium radiotolerans JCM2831

<400> 3

```

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

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	
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															

ES 2 543 568 T3

<210> 4

<211> 443

<212> PRT

<213> Sulfolobus acidocaldarius DSM 639

5 <400> 4

```

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

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  
 420 425 430

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

<210> 5

<211> 579

5 <212> PRT

<213> Cupriavidus basilensis HMF14

<400> 5

Met Asp Thr Pro Arg Glu Arg Phe Asp Tyr Val Ile Val Gly Gly Gly  
 1 5 10 15

Ser Ala Gly Cys Val Leu Ala Asn Arg Leu Ser Gln Asp Pro Ala Ile  
 20 25 30

Arg Val Ala Leu Ile Glu Ala Gly Val Asp Thr Pro Pro Asp Ala Val  
 35 40 45

Pro Ala Glu Ile Leu Asp Ser Tyr Pro Met Pro Leu Phe Phe Gly Asp  
 50 55 60

Arg Tyr Ile Trp Pro Ser Leu Gln Ala Arg Ala Val Ala Gly Gly Arg  
 65 70 75 80

Ser Lys Val Tyr Glu Gln Gly Arg Val Met Gly Gly Gly Ser Ser Ile  
 85 90 95

Asn Val Gln Ala Ala Asn Arg Gly Leu Pro Arg Asp Tyr Asp Glu Trp  
 100 105 110

Ala Ala Ser Gly Ala Ser Gly Trp Ser Trp Gln Asp Val Leu Pro Tyr  
 115 120 125

Phe Arg His Leu Glu Arg Asp Val Asp Tyr Gly Asn Ser Pro Leu His  
 130 135 140

Gly Ser His Gly Pro Val Pro Ile Arg Arg Ile Leu Pro Gln Ala Trp  
 145 150 155 160

Pro Pro Phe Cys Thr Glu Phe Ala His Ala Met Gly Arg Ser Gly Leu  
 165 170 175

Ser Ala Leu Ala Asp Gln Asn Ala Glu Phe Gly Asp Gly Trp Phe Pro  
 180 185 190

Ala Ala Phe Ser Asn Leu Asp Asp Lys Arg Val Ser Thr Ala Ile Ala  
 195 200 205

Tyr Leu Asp Ala Asp Thr Arg Arg Arg Ala Asn Leu Arg Ile Tyr Ala  
 210 215 220

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

Val Ile Ala Met Arg Ala Asp Gly Ser Arg Leu Ala Leu Asp Ala Gly

				245					250					255	
Glu	Val	Ile	Val	Ser	Ala	Gly	Ala	Leu	Gln	Ser	Pro	Ala	Ile	Leu	Met
			260					265					270		
Arg	Ala	Gly	Ile	Gly	Asp	Ala	Gly	Ala	Leu	Gln	Ala	Leu	Gly	Ile	Glu
		275					280					285			
Val	Val	Ala	Asp	Arg	Pro	Gly	Val	Gly	Arg	Asn	Leu	Gln	Asp	His	Pro
	290					295					300				
Ala	Leu	Thr	Phe	Cys	Gln	Phe	Leu	Ala	Pro	Gln	Tyr	Arg	Met	Pro	Leu
305					310					315					320
Ser	Arg	Arg	Arg	Ala	Ser	Met	Thr	Ala	Ala	Arg	Phe	Ser	Ser	Gly	Val
				325					330					335	
Pro	Gly	Gly	Glu	Ala	Ser	Asp	Met	Tyr	Leu	Ser	Ser	Ser	Thr	Arg	Ala
			340					345					350		
Gly	Trp	His	Ala	Leu	Gly	Asn	Arg	Leu	Gly	Leu	Phe	Phe	Leu	Trp	Cys
		355					360					365			
Asn	Arg	Pro	Phe	Ser	Arg	Gly	Gln	Val	Ser	Leu	Ala	Gly	Ala	Gln	Pro
	370					375					380				
Asp	Val	Pro	Pro	Met	Val	Glu	Leu	Asn	Leu	Leu	Asp	Asp	Glu	Arg	Asp
385					390					395					400
Leu	Arg	Arg	Met	Val	Ala	Gly	Val	Arg	Lys	Leu	Val	Gln	Ile	Val	Gly
			405						410					415	
Ala	Ser	Ala	Leu	His	Gln	His	Pro	Gly	Asp	Phe	Phe	Pro	Ala	Thr	Phe
			420					425					430		
Ser	Pro	Arg	Val	Lys	Ala	Leu	Ser	Arg	Val	Ser	Arg	Gly	Asn	Val	Leu
		435					440					445			
Leu	Thr	Glu	Leu	Leu	Gly	Ala	Val	Leu	Asp	Val	Ser	Gly	Pro	Leu	Arg
	450					455					460				
Arg	Ser	Leu	Ile	Ala	Arg	Phe	Val	Thr	Gly	Gly	Ala	Asn	Leu	Ala	Ser
465					470					475					480
Leu	Leu	Thr	Asp	Glu	Ser	Ala	Leu	Glu	Gly	Phe	Val	Arg	Gln	Ser	Val
				485					490					495	

Phe Gly Val Trp His Ala Ser Gly Thr Cys Arg Met Gly Ala His Ala  
 500 505 510

Asp Arg Ser Ala Val Thr Asp Ala Ala Gly Arg Val His Asp Val Gly  
 515 520 525

Arg Leu Arg Val Ile Asp Ala Ser Leu Met Pro Arg Leu Pro Thr Ala  
 530 535 540

Asn Thr Asn Ile Pro Thr Ile Met Leu Ala Glu Lys Ile Ala Asp Thr  
 545 550 555 560

Met Gln Ala Glu Arg Arg Ala Val Arg Pro Ala Ser Ser Glu Val Ala  
 565 570 575

His Pro Ser

<210> 6

5 <211> 564

<212> PRT

<213> Bradyrhizobium japonicum USDA 110

<400> 6

Met Tyr Asp Tyr Ile Ile Val Gly Gly Gly Ser Ala Gly Ser Val Leu  
 1 5 10 15

Ala His Arg Leu Ser Ala Lys Ser Ala Asn Lys Val Leu Leu Cys Glu  
 20 25 30

Ala Gly Gln Asp Thr Pro Pro Gly Asn Glu Pro Ala Glu Ile Arg Asp  
 35 40 45

Ser Tyr Pro Gly Thr Ala Tyr Phe Asp Pro Arg Phe His Trp Thr Glu  
 50 55 60

Leu Lys Val Thr Thr Gln Val Val Ser His Asn Asn Pro Thr Glu Ala  
 65 70 75 80

Arg Pro Pro Leu Arg Lys Tyr Glu Gln Ala Arg Val Leu Gly Gly Gly  
 85 90 95

Ser Ser Ile Asn Gly Gln Met Ala Asn Arg Gly Ala Pro Thr Asp Tyr  
 100 105 110

Asp Glu Trp Asp Ala Arg Gly Ala Glu Gly Trp Thr Trp Asn Asp Val  
 115 120 125

Leu	Pro	Phe	Phe	Lys	Lys	Val	Glu	Arg	Asp	Leu	Asp	Phe	Asp	Gly	Pro
	130					135					140				
Tyr	His	Gly	Lys	Asp	Gly	Arg	Ile	Pro	Val	Arg	Arg	Ile	Pro	Arg	Glu
145					150					155					160
His	Trp	Thr	Arg	His	Ser	Gln	Ala	Phe	Ala	Asp	Ala	Phe	Gln	Gln	Ala
				165					170					175	
Gly	His	Gln	Phe	Val	Ala	Asp	Gln	Asn	Gly	Glu	Phe	Val	Asp	Gly	Tyr
			180					185					190		
Phe	Ala	Val	Thr	His	Ser	Asn	Gln	Ala	Glu	Gln	Arg	Val	Ser	Ala	Ala
		195					200						205		
Met	Gly	Tyr	Leu	Asp	Arg	Asp	Thr	Arg	Lys	Arg	Ala	Asn	Leu	Thr	Ile
	210					215					220				
Ser	Thr	Asn	Thr	Gln	Val	Arg	Glu	Leu	Leu	Phe	Glu	Gly	Thr	Gln	Cys
225					230					235					240
Val	Gly	Val	Lys	Ala	Arg	Val	Asp	Gly	Arg	Glu	Gln	Glu	Phe	Arg	Gly
				245					250					255	
Arg	Glu	Ile	Ile	Leu	Ser	Ser	Gly	Ala	Ile	His	Ser	Pro	Ala	His	Leu
			260					265					270		
Leu	Arg	Ala	Gly	Ile	Gly	Pro	Val	Gly	His	Leu	Lys	Asp	Met	Gly	Ile
		275					280					285			
Pro	Val	Leu	Thr	Gly	Leu	Pro	Gly	Val	Gly	Gln	Arg	Leu	Met	Asp	His
	290					295					300				
Pro	Ser	Ile	Ser	Leu	Ser	Ser	Phe	Val	Arg	Arg	Gly	Ala	Arg	Met	Asn
305					310					315					320
Glu	His	Thr	Arg	Arg	His	Met	Gln	Leu	Gly	Leu	Arg	Tyr	Ser	Ser	Gly
				325					330					335	
Leu	Ser	Gly	Val	Pro	Lys	Gly	Asp	Met	Phe	Val	Val	Val	Leu	Ser	Lys
			340					345					350		
Ser	Ala	Trp	His	Ala	Val	Gly	Ala	Gln	Ile	Gly	Ser	Leu	Leu	Thr	Phe
		355					360					365			
Val	Asn	Lys	Thr	Tyr	Ser	Glu	Thr	Gly	Gln	Val	Lys	Leu	Ala	Ser	Arg
	370					375					380				

ES 2 543 568 T3

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

Arg Asp Leu Asp Arg Leu Met Ser Gly Phe Arg Lys Met Ala Ala Val  
 405 410 415

Gln Met Ser Glu Ile Val Arg Lys Val Thr Asp Lys Pro Phe Pro Ala  
 420 425 430

Ala Tyr Thr Asp Lys Val Arg Lys Ile Gly Val Val Asn Thr Gly Asn  
 435 440 445

Lys Ile Leu Thr Arg Val Ala Ala Ala Leu Met Asp Gly Pro Ala Ala  
 450 455 460

Leu Arg His Tyr Leu Ile Asp Asn Phe Val Val Glu Gly Phe Thr Phe  
 465 470 475 480

Asp Asp Val Met Asn Asp Asp Glu Ala Leu Glu Ala Phe Val Arg Lys  
 485 490 495

Ala Thr Ile Gly Val Trp His Ala Ser Cys Ser Cys Arg Met Gly Arg  
 500 505 510

Ala Asp Asp Pro Met Ala Val Val Asp Asn Gln Gly Arg Val Arg Gly  
 515 520 525

Val Gln Gly Leu Arg Val Val Asp Ala Ser Ile Phe Pro Val Val Pro  
 530 535 540

Cys Ala Asn Thr Asn Phe Pro Val Leu Met Ser Ala Glu Lys Ile Ala  
 545 550 555 560

Ala Thr Met Gln

<210> 7

<211> 1326

5 <212> ADN

<213> Cupriavidus basilensis HMF14

<400> 7

ES 2 543 568 T3

atggaagccg tagcaaagaa gcgtacagag acgatcagcg aggcgctgcc agcggcgacc 60  
 aatcgccagg tgtttggcgc cgtgacggcg tcgtgcatgg gatgggcgct ggacctgttc 120  
 gacctgttca tcctgctggt cgtggcgccc gtgatcggca ggctgttttt cccgtcggag 180  
 cacgccatgc tgcgctggc ggcggtgtat gcgtcgttg ccgtgacgct gctgatgcgg 240  
  
 ccgctoggct cggcgatctt cggcacttat gccgaccgcc acggccgcaa gggggcgatg 300  
 gtagttgccg tcaactggcg tggcttgtcc acggcggcgt tcggcctgct gcctacgggt 360  
 ggtcaggtgg ggctgcttgc gccagccttg tttatcctgc tgcggctggt gcagggcatc 420  
 ttcgtaggtg gcgtggtggc atccacccac accatcggta ccgaatcggg gcccccgcc 480  
 tggcgcggcg ccgtctccgg gctggtcggg ggcggggcg cgggcatcgg ggcactgctg 540  
 gcttccatta cctacatggc gatgaccgcg ctgtttccgg ggggaagcgtt cgatgcctgg 600  
 ggttggcgct gcattgttctt ctccggcacc atcagctcgg tgctcggcct gttcatcttc 660  
 aactcgtcgg aggagtctcc gctgtggaag cagttgcagg cggccaaggg gcacgcgcgg 720  
 ccggttgaga acccgtcgcg cgtgatcttc tcccgccagt accgtggcgt cctcttcgtc 780  
 aacatcctgc tcaccgtggg cggcggcagc gcctactacc tgacctccgg ctatctgccg 840  
 accttctca aggtggtggt gaaggcaccg gctggcgcat ccgcagccat cctgatggcc 900  
 agcagtgttg gcgttatcgt ggcacogata attgcgggtc acctcagcac gctgattggt 960  
 cgcaagcgag ccttctcgtc gatcggcgcc ttgaacgtgg tgctgctgcc gttgatctac 1020  
 caacggatgc ccgcggcgcc ggatgtcacc acgcttggca tttatgccgt ggcgctggcg 1080  
 atgctgggca gcaccggctt cgcctcagtc ctcatcttcc tgaacgaacg ggtttccac 1140  
 cagcatccgt gctacgggaa ctggcctgtc atggaatata ggctttgcca tcggcggcat 1200  
 gatgcgacgt ttgcgtcgtc gtgcgcagca cccccgcgac tgccaaagtg ctggggatct 1260  
 cgtcgcggtg tcaaagcatt tactgccggt gcggcgatcg tctggaacgc gccgctgggg 1320  
 gagtga 1326

<210> 8

<211> 1350

5 <212> ADN

<213> Cupriavidus basilensis HMF14

<400> 8

ES 2 543 568 T3

atggaagccg tagcaaagaa gagtgcagcg acgatcagcg aggcgctgcc agcggcgagc 60  
 aatcgccagg tgtttggtgc cgtggcggcg tcgtgcatgg gatgggcgct ggacctgttc 120  
 gacctgttca tcttctgttt cgtggcgccc gtgatcggca ggctgttttt cccgtcggag 180  
 cacgcgatgc tgtcgtctggc ggcggtgtat gcgtcgtttg ccgtgacgct gctgatgcgg 240  
 ccgctcggct cggcgatctt cggctcttat gccgaccgcc acggccgcaa gggggcgatg 300  
 gtggttgccg tcaactggcg tggcttgtcc acggcggcgt tcggcctgct gccgacggtg 360  
 ggtcagggtg ggctgcttgc gccagccttg tttatcctgc tgcggctggt gcagggcatc 420  
 ttcgtgggtg gcgtggtggc atccaccac accatcggta ccgaatcggg gcccccgctc 480  
 tggcgcggcg ccgtttccgg gctggttggg ggcggtggcg cgggtctcgg ggcgctgctg 540  
  
 gcttccatta cctacatggc gatgaccgcg ctgtttccgg ggggaagcgtt cgatgcctgg 600  
 ggttggcgct gcatgttctt ctccggcacc atcagctcgg tgcctggcct gttcatcttc 660  
 aactcgtctg aggagtctcc gctgtggaag cagttgcagg cggccaaggg gcacgccgcg 720  
 ccggttgaga acccgtctgg cgtgatcttc tcccgccagt accgtggtgt cctcttcgtc 780  
 aacatcctgc tcaccgtggg cgggtggcagc gcctactacc tgacctcggg ctatctgccg 840  
 accttctca aggtggtggt gaaggcatcg gctggcgagt ctgccgccat cctgatggcc 900  
 agcagtctgg gtgtgatcgt ggcacgatt cttgccggcc acctcagtac gatgatcggc 960  
 cgcaagcgag ccttctgtt gatcggcgcg ctgaacgtgg tagtactgcc gctgctctac 1020  
 cagtggatgc cggcggcgcc ggacaccacc acgctcggcc tgtatgotgt ggtgctgtcc 1080  
 atgctgggct gcagcggctt cgccccgatc ctcatcttc tgaacgaacg gttccccacc 1140  
 agcatccgtg ccacggggac cggcctgtca tggaatatcg gatttgccgt cgggtggcatg 1200  
 atgccgacgt ttgcttcgct gtgcgccagc acccctgccg aactgcccat ggtgctgggc 1260  
 atcttctcgg cggttgtcac catcatctac ctggtgggtg cgttcatcgt tccggagacg 1320  
 gtagggcgcc ttggcgacaa tggagcgtag 1350

<210> 9

5 <211> 1350

<212> ADN

<213> Methylobacterium radiotolerans JCM2831

<400> 9

ES 2 543 568 T3

atgcagaccg cegccacctt cgcataccgat ccccccgccc tcgccaaacc cacaggccgg 60  
 cagacgggtga cgcgcgccat ggcctcgcgtg ttccggctggg ggcctcgatct cttcgacctg 120  
 ttcatcctgc tttacgtcgc gccggtcggt ggacacgttg tcttcccggc ggacaagccg 180  
 atgctgtcgc tggccggcgc ctatgcgtcc ttccgggtca cgetgetgat ccggcccctc 240  
 ggctcggccc tgttcggctc ctatgcggac cgcctcggcc gccgtcgcgc cctcatggtg 300  
 gccgtggctg gcgtcggcat atcgaccgcg gtcttcggtc tccctgccgac ggtcggccag 360  
 atcggatggc tcgcgactgc ggtcttcctt ttcttcgcc tcgtccaggg catcttcgtc 420  
 ggcggcgtcg tcgcggcatc ccacacgatc gggacggaat ccgtcccgga gcgggtggcgg 480  
 gggctgatgt ccggggcggg cggcgggtggc ggctcggcca tcggcggcct gcttgccctc 540  
 ctggtcttct atgtcgtctc gctcatggcg ccgggagaag ccttcgccga gtggggctgg 600  
 cggctgatgt tcttctcggg tctcctgacc tcgggtgatcg gcctgatcct cttccgcaat 660  
 ctogaggaat ccccgatctt caaggaactg caggcccgga aggcggcgcct ccgggcaggc 720  
 gctccggcgg aggcctcacc gatccgctcc ctgttctcac cgtccaaccg gggaagcttc 780  
 gccgtcgcga ctctcatctc cttcgggtggg ggcgcagcct actaccttac gtccggctac 840  
  
 ctgcogacco tcctcaagct cgtgaacggc gtgccgaacg ccaccgcctc gatgatcctg 900  
 atcggcgcca acgtcgcggc ggcgatcggc gcctcgggca tgggcgaact gagccagcat 960  
 atcggcgcca aacggtcggt ccttctgatg ggcgtcatcc gcctgctcgc ttttcccgcc 1020  
 ctgttctga ccatggcgaa cacgaacgagc ctcgttgggg tcgcccgcctg cgccttctg 1080  
 ctggcgctca tcgccaacgg cagctacggg ccgctgctga tcttctcaa cgaaaaattc 1140  
 ccgacggcgg tgcgggccac ggggacgggc ctgacctgga acatcggctt cgcgctgggc 1200  
 ggcattgctgc cgacgctcgt ctcccttgtc gccgacgggc cgacgcaaat tccgatggtg 1260  
 ctggcgtca tcacgaccgg cgtcacctg gtctatcttg tcggtgctt cctgaccgac 1320  
 gagacgcagg gcaatctcga ccgggcctga 1350

<210> 10

<211> 1332

5 <212> ADN

<213> Sulfolobus acidocaldarius DSM 639

<400> 10

ES 2 543 568 T3

atgaaaaaag aagaaaaaatt tacttcaaat cattttaaat ggacttttagc aacctttttc 60  
 acctggacat tcgacttata tgacctgttc actatcctac ttgttgacc ttatatctct 120  
 tcctgtttt tccttcaag taccactttc ttgtctatag cggcaacata tgcgggggtt 180  
 gccacctcat tgatcatgag acctgtaggt gctacagtct ttgggtctag ggtatcggac 240  
 aaggtaggta gaaagagggc tattttttac ggtctaataag gtcttgtgat aacttcaca 300  
 ttacagggcg cattaccgac ttatcaagta gtaggagtga tagcaccaat attactgta 360  
 gcagtgagac tgattcaagg ggtgttcata ggaggcataa ctgcaggtag tcatgtgata 420  
 ggacctgagt ctgtcccaga gaggtataga gggattgtag gtggactagg gttttcagct 480  
 gctggagtgg cataacctcat agcagctggc tggtttttcc tgacaacct actctaccct 540  
 ggaagttcct atctggtatg gggatggaga gtgatgttct ttggcgggct actgagctta 600  
 gctgtacttg gatttgtaa ctatttagtc cctgagtctg aggtttggac taaaattaa 660  
 aagagagggg cagtgggtgaa gtctcccctc aaggagatat tttcgaaata caggtatcag 720  
 ttaggcgttg cactgctact ttcaattggc tggggtgcaa gtttttacgt tactgatggt 780  
 atactaccta cgtttttaag tagtgtcaat aaattagcca agactgaaat agcaatagta 840  
 atgataatag ggtcaattgg gatgtcgata ggaccattaa taggagggga gatatccag 900  
 ataatagggg ggaagataac gtcattaata ggtgcaatta ttgtactagc agtagttgga 960  
 cctctctttc tctcctggg ttcaactaaa tccggtgacc ttaatcagat aattcttcac 1020  
 tcttttgcaa tctgttctt agtcgatata ggaggagga tgctaatgac ttacctaat 1080  
 gagatatac cagctagcgt gaggggaaca ggggttgct tcacgtggaa tactgggtt 1140  
 gcaataggag ggacaatccc aaccataatc agcctagcag tagcttcagc aggtctctct 1200  
 gcttttccct ctataatggt ctacacctta atagtgtct ctgtaataat attagtgggt 1260  
 acagtactca cgaaggagac taagggtacc atatcaaagg aggagtatga gattcagaag 1320  
 gagactttat ag 1332

<210> 11

<211> 1740

5 <212> ADN

<213> Cupriavidus basilensis HMF14

<400> 11

atggatacgc cgaggggagcg tttagactac gtgattgttg ggggggggc cgcgggtgc 60  
gtactggcca atcgccctgtc gcaggaccgc gccatccgcg tcgcgctgat cgaggggggc 120  
gtcgatacgc cgcgggacgc tgtgcggcg gagatcctcg acagctatcc gatgcccttg 180  
ttcttgggtg accggtatat ctggccatcg ctgcaagccc gcgcogtggc agggggcagg 240  
tocaaggtct acgagcaagg gcggtcatg gggggggct ccagcatcaa cgtgcaggcg 300  
gcaaaccgcg ggctgcgcg cgactacgat gaggggccg cgtcgggcgc gtcgggatgg 360  
tcgtggcagg atgtgctgcc gtatttcgc caccttgagc gcgatgtgga ttaaggcaac 420  
agccogctgc accgcagcca cggaccggtg ccgatccgc gccatcctgcc gcaggcttgg 480  
cgcogttct gcaaggagt tgcgcacgc atgggcgcga gggcttgc cgcgctggcc 540  
gaccagaacg cggagttcgg cgatggctgg ttccggccg ccttctgaa cctggatgac 600  
aagcgggtt cgcaccgat cgcctatctc gacgggata cgcgcggcg ggccaatctg 660  
cggatctatg ccgagacaac ggtgcgcaag ctogtctat ccggccggga agcgcgtggg 720  
gtgatcgcca tgggggcga tgggtgcgg ctggcctgg acccgggga ggtcatcgtg 780  
tcggggggcg ccttgcagtc gcccgccatc ctgatgcgcg cggggatcgg cgcgcggc 840  
gcgctgcagg ccctcggcat cgaggtcgtg gcgaccgac cggcgcttgg ccgcaatctc 900  
caggatcctc ccgcgctgac gttctgccag ttctcgcgc ccagtagcg catgccctc 960  
tcgcccggc gcgctagcat gacggggcg cggttctcat cgggggtgcc aggtggcgag 1020  
gcgctggaca tgtacctgtc cagttccaca cgggcaggct ggcctgcaact cggtaatcgg 1080  
ctggcctct tcttctgtg gtgcaatcgg ccattctcgc gggggcagg gtgacctgcg 1140  
ggagcccagc cggatgtgcc gcccatggtg gagctcaacc tgctcgacga cgagcgggat 1200  
ctgcccgcga tgggtggcgg cgtacgcaag ttggtgcaga tcgtgggtgc gtcggccttg 1260  
catcagcctc ccggtgattt ctcccccgt acgttttcgc cgcgcgtcaa ggcgctgagc 1320  
cgcgtgagcc ggcgcaatgt gttgctcag gatttctgg gggcagctct tgatgtctcg 1380  
gggcgctgc gcagaagcct gatcgcgcgc ttgtcacgg ggggcgcaa cctggccagc 1440  
ctgctgacgg atgagtcgc gctagagggc ttogtgcgc agagcgtctt cggggtctgg 1500  
catgccagcg gcacttgcg gatgggcgc catgcggacc ggagcgcggt gacggatgcg 1560  
ggggccgcg ttacagatgt tggcaggctg cgcgttattg accctctct gatgcgcgg 1620  
ctgcccagcg ccaataccaa catccccacc atcatgctcg cggaaaagat tgccgacacc 1680  
atgcaagccg agcgcgcgc ggtccggccg gcctcgagcg aagttgcca tccgagttga 1740

<210> 12

<211> 1695

<212> ADN

ES 2 543 568 T3

<213> Bradyrhizobium japonicum USDA 110

<400> 12

gtgtacgact acattatcgt gggcggcggc tcggcggggg cegtgtgtgg ccaccggctc	60
tccgcaaaga gtgccaacaa ggtcctgtct tgcgaagccg gacaggacac accaccgggc	120
aacgagccgg ccgagatcag ggacagctat ccgggcacgg cctatttoga tccgcgtctc	180
cactggacgg aattgaaggt cacgacgcag gtcgtcagcc acaacaatcc gaccgaagcg	240
cgtcctccct tgcgcaaata cgagcaggcg cegtgtgtcg gcggcggctc gtcgatcaac	300
ggccagatgg ccaaccggcg cgcgccgacc gactacgacg aatgggatgc gcggggtgcc	360
gaaggggtga cgtggaacga cgtgtgtccg ttcttcaaga aggtcgagcg cgacctgat	420
ttcgacggcc cgtatcacgg caaggacggc cggatcccgg tgcgccggat tccgcgggag	480
cactggacac ggcattcgca ggccttcgcc gatgccttc agcaggccgg tcatcaattc	540
gtggcggacc agaacggcga gttcgtcgac ggctatttgc cggtgacgca ctccaaccag	600
gccgagcagc gcgtctccgc cgcgatgggc tatctcgatc gcgacaccgg caagcgcgcc	660
aacctcacga tctccaccaa cacgcaagtt cgggagctgc tgttogaagg cacgcaatgc	720
gtcggcgtga aggccagggt cgacgggagg gagcaggaat tccgcggacg cgagatcatt	780
ctctccagcg gcgccatcca ttcgccggcg catctgtctc gcgccggtat cggcccggtc	840
ggccacctca aggacatggg gattcccgtg ctgacgggct tgcggggcgt cggccagcgc	900
ctgatggatc atccctgat ttcgtgtctg tcctttgtcc gccgcggcgc gcgcatgaac	960
gagcatacca ggccacacat gcagcttggc ctgcgctatt cgtccgggct gtcaggggtg	1020
ccgaagggcg acatgttcgt cgtcgtgtct agcaaacgg cctggcacgc ggtcggcgcg	1080
caaacggct cgctgtgtac cttcgtcaac aagacctatt ccgagaccgg acaggtcaag	1140
cttgccctgc gcgaccctc cgcagagccg atcgtcgagt tcaacctgct gtcggaccgg	1200
cgcgacctcg atcggctgat gagcggcttc cgcaagatgg cggcggtgca gatgagcgaa	1260
atcgtcagga aggtgacgga caagccgttc ccggccgctt ataccgaaa ggtccgcaag	1320
atcggcgtgg tcaacaccgg gaacaagatc ctgaccagag tcgcggcggc gctcatggac	1380

ES 2 543 568 T3

gggcccggcgg cgctgcgtca ctatctgatc gacaatttcg tggtcgaagg cttcaccttc 1440  
gatgacgtga tgaacgacga cgagggcgtt gaagccttcg tgcgcaaggc gaccatcggc 1500  
gtgtggcacc cctcgtgctc atgccgcatg ggccggggccg acgatccgat ggcgggtggtc 1560  
gacaaccagg gccgcgtcag ggggtgtccag ggtctgcggg tgcgcgacgc ctcgatcttc 1620  
ccggtgggtgc catgcgccaa caccaacttt ccggtgctga tgcgggcaga gaagatcgct 1680  
gcaaccatgc agtga 1695

<210> 13

<211> 28

5 <212> ADN

<213> sintético

<400> 13

acgaattcaa aggagacaac aatggaag 28

<210> 14

10 <211> 27

<212> ADN

<213> sintético

<400> 14

aagctagctg agcagtcacc ctactc 27

15 <210> 15

<211> 27

<212> ADN

<213> sintético

<400> 15

20 cgggaattcca catgacaagg gagaccg 27

<210> 16

<211> 30

<212> ADN

<213> sintético

25 <400> 16

cgggaattcgc ttcggtcttc aactcggatg 30

<210> 17

<211> 28

<212> ADN

30 <213> sintético

## ES 2 543 568 T3

<400> 17

acgaattcgg aggaaatcta tgcagacc 28

<210> 18

<211> 26

5 <212> ADN

<213> sintético

<400> 18

aagctagcgc agaaccgtat cgtcag 26

10 <210> 19

<211> 500

<212> PRT

<213> Cupriavidus basilensis HMF14

<400> 19

15

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

ES 2 543 568 T3

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> 20

<211> 479

5 <212> PRT

<213> Burkholderia sp. CCGE1002

<400> 20

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 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

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

ES 2 543 568 T3

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> 21

<211> 483

5 <212> PRT

<213> Burkholderia graminis C4D1M

<400> 21

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

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

ES 2 543 568 T3

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  
450 455 460

Ile Lys His Ile Cys Val Gly Ala Gly Val Leu Glu Gly Ile Glu Val  
465 470 475 480

Leu Gly Ser

<210> 22

5 <211> 483

<212> PRT

<213> Azospirillum sp. B510

<400> 22

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
			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

ES 2 543 568 T3

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

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> 23

<211> 480

5 <212> PRT

<213> Pseudomonas putida

<400> 23

ES 2 543 568 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  
 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

ES 2 543 568 T3

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

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> 24

<211> 486

5 <212> PRT

<213> Rhodopseudomonas palustris BisB18

<400> 24

ES 2 543 568 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	
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					

ES 2 543 568 T3

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

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> 25

<211> 485

5 <212> PRT

<213> Dinoroseobacter shibae DFL 12

<400> 25

ES 2 543 568 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  
 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

ES 2 543 568 T3

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
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> 26

<211> 1503

5 <212> ADN

<213> Cupriavidus basilensis HMF14

<400> 26

ES 2 543 568 T3

```

atgaacgcgc aacactggat tgccggcgcc tggaccggcg agccttcgcg cgatagcgtc      60
aaccccgcgc acgggaccct gatcgggcag ttccgggacg gccgcacctg gcaagccgaa      120
gcggccatcg ccgcgcgcgc ccattgtctc gagcgcacca cctggggcca ggatgcccgc      180
ctgcgccagg acgtgcttct agcctgggct ggtgcgctcg aggcagagcg agagcgcccg      240
gccagcctgc tcaccgcgga aaacggcaag ccggtcgcac aagcccgagg cgaggtcggc      300
gcgcgaattt cagaggtccg ctattacgcc gggctggcgc ggcacatccc gggtcacgtg      360
ctggagcccg agccaggcac gatatcgacc atcctgcgcg agcccggccg cgtcgcgcc      420
atcatcgtcc cctggaacgc gcgggggtg ctgctcgtgc gctccctcgc gccagcgctt      480
gccgcgggct gcacggcagt ggtcaaatcg gcagcgcaa ccacgctgtt cacagccgca      540
atgctgogct tgttcgagcg caaggccctg ccggccggcg ccgtcaatct ggtctgogaa      600
acgggctatg cggcagcgga ccacctggcg cgttcgcgcg acgtggacgt agtgagcttc      660

acaggatcga ccgcaaccgg caagaagatc atgatcgccg ctgoggacag cgtgaaaaaa      720
ctctcgetgg aactcggcgg gaaatcgtgc tgccctggtg tcgacgacgt cgatgogcaa      780
gcggtcgcga aacggcttgc gcttgccgcc accgtcatct cgggccagca atgcaccgcc      840
gcggggcgag tactggttca cgaagccatc ggcaccacaga tgcgcggca cctgaccgag      900
gccctcgcgc cgtcgcgcct cgggcccggc atcagcccgc acacccaat cggcccgcctg      960
atcgaccacc cgaacgcgcg gatggtgagc gcgcaagtcg agcgcgcctg cgacgaggcg      1020
gacacggctc tgctgcgcgc cacgatgccg ggcggcgcgc tagcgcgcgg cgccttcctc      1080
agccccacac tagtgaaca cagcgacccc ggtgccttct tctgccagga ggagatcttc      1140
gggccttcg tcacattcga gaccttcgcg accgaagacg aggcgctagc caaggccaac      1200
aacaccgtct tcggcctgtc cgcacgcgtc tggacgcacc acggcgagcg cgcctacgc      1260
ctagcggggg cgtcgcgcaa cggcacggtc tgggtcaacg accacaaccg cctgttcgcc      1320
gaagcggaga cgggcggcta tcggcaaagc ggccttgac ggctccacgg ttatgacgcc      1380
ctgcgggact tcaccgagtt gaagcacatc tgcacccagg cgggcctgcc gaaagggatg      1440
tcgcaggcgg gctgcaggct cagtgggta gcagcgcgcg agcggatggg agtttcgcgc      1500
tag                                                                                   1503

```

<210> 27

<211> 1440

5

<212> ADN

<213> Burkholderia sp. CCGE1002

<400> 27

ES 2 543 568 T3

atgaacgcac gacactggat tgcggcgaa tggaccggca cgcccaacat cgacagcacc 60  
 gatcccgga cgggcatgc catcgccggg ttcccgacg gcggatcgag cgaagccgat 120  
 gccgccatcg ccgccgctcg ccacgcgttc gatcgacga catggggcga ggacgcggc 180  
 ttgcgccagg acgtgctact cggctgggct tcggcgctcg aagccgaacg cgacatgctc 240  
 gcgacgctgc tcacgcgtga aaacggcaag gcgatcggc aatcgcgga cgagattgcc 300  
 ggcgcgatat cggaggtgag ctattacggg gggctcggc gccatattgc cgggcacgtg 360  
 ctggaaccgg aaccgggcac gatatcgac atgctgcgc aagcggccgg cgtcgcgcg 420  
 atcatcgtgc cgtggaatgc gcccgccgtg ctgctcgtgc gctcgtcgc gcccgcttc 480  
 gcggccggct gcacggtgat agtcaaaccg gcggcgaaa cgtcgtgct cacggccggc 540  
 atgttgcgct gcttgaaca cacggcgtt cctgaaggcg cggtgaacct agtcaacgaa 600  
 cggggatagc cggccagcca gcggctcgtc gactcgacg gggtcgacgt ggtgagcttc 660  
 acgggatcga cggcgacggg caagaagatc atggcggccg ctgccgacag catgaagaag 720  
 ctgtcgctcg aacttggcgg caagtcatgc tgcgtcgtc tcgacgatgc cgatgtcgcg 780  
 gcgatcggc cgctctcgc gcgcgcggc acgatcatc cggggcagca atgcacggc 840  
 gcgcgcggg tgctggtca tgcgtcgcg gctgcgaaa tgcgcgagca gctcgcgtc 900  
 gcgcttgcgt cgtcgcggg cggaccggc atogatcgg cgacggatat cggcgcctc 960  
 atcgacggca ccacgcgga tgcggtggca cggacgatc agcgcgcgtg cggaacggcc 1020  
 gagcgtgtgc tgctgcgagg cacttgcctg gggcacgcgt ttctgtgcc gacgctcgtc 1080  
 gagcatgacg atccgaaggc gttcttttgc caggacgaga tctttggccc gttcgtcacg 1140  
 ctggaagtct tcgagaacga aatggaagc atcgagaaag ccaacgacac ggtcttcggt 1200  
 ctctcggcga gcgatggac gcacgacgga gcgcgcgcat tgcgcgtcgc gcgtgcgctg 1260  
 cgcaacggca cgggtgtgat caacgaccac aacaagcttt tcgccgaagc cgaaacgggc 1320  
 gggtatcgcc aaagcggact gggtcggctg cacggctacg atgcgctggc ggacttcacc 1380  
 gagttgaagc acatctgcat gccggccggc gtggcggaag gtatcgccc gttgcggtga 1440

<210> 28

<211> 1452

5 <212> ADN

<213> Burkholderia graminis C4D1M

<400> 28

ES 2 543 568 T3

atggaagag acgctatgaa ttggatogcc ggcgaatggg ccgcaagcc cgtgctcgca	60
agctccgatc cttegaacgg agagacgctc ggccggttcg tcagctcgaa cacgcaagac	120
gcggacgccg ccgtctcggc tgcccgccat acgttcgatc acacaacgtg ggcacaggat	180
gcccgtcgac ggcaggatgt tctgcttcgc tgggcgcaag ctctcgagtt gtcggtggaa	240
ccgctagcgg aattgctgac gcatgaaaac ggcaagacta ttgggcaggc gcgaggagaa	300
atgcgcgccg caatttcoga agtgcgctac taocggggac tcgcgcgcca cattgccggt	360
cacgtcatcg agcccagacc aggcaccatc togacgatgt tgcgcgagge ccgccggcgtt	420
gctgcgatca tcgtgcogtg gaacgctccc gcogtgcctg tcgtgcggtc cettgcgccc	480
gccctcgtg ccggatgcac ggccatcgtc aagccggctg ccagacgctc attgatcacc	540
gcggcaatga tcogttgtct cgaccagccg gcgcttcccg ccggtgctgt caatctgctt	600
ctcgaaaacg gcgcagaagc cgctgcgcga ctggtcgagt ccgccgatgt cgatgtgatc	660
agctttacgg gttcgacgga agtcggcaag cgaatcatgc gcgccgctgc agacagcatg	720
aagcgattgt cgcttgagct gggcggcaaa tcgtgttgct tcgtcttcga ggacagcgac	780
gttaaggcca ttgcacctcg tcttgcccgg gcgcgcacca tcactcccgg acaacaatgc	840
accgctgcac ggcgaatact cgtgcatgtg tcaaaggcgg accaaatgog agacgaactc	900
gtcaaagcac tcgcaagcct caaggttggc ccaggggatcg atcccgcttc cgacattggc	960
gcgctgatcg atgctgcgct ccgcgacgct gtccagacca ccggtgaacg agcgtgtgac	1020
cttgacagatc gcgtgctctt ggcggggact tcgagoggac cgggcgcctt tctgtcaccg	1080
acgctggctg agcaocggaga acctcaagcc ttcttttgcc aggacgagat ctttgggccc	1140
ttcgtcacgt tagagacctt tgtgaccgaa aaggaagccg ttgaaaaggc aaacaacacc	1200
gtttttggcc tgtccogogag cgtgtggacc cacgactcgg ctccgggcttt ccggattgog	1260
cgtgcgctgc gtgatgggac cgtctggatc aacgaccaca acaggttggt ccgagaggcc	1320
gaaactggcg gctatcggca gagcggactg ggccggttgc acggctatga ccgctagcg	1380
gacttcacgg agatcaagca tatctgtgct ggtgcaggag tcctggaagg catcgaagtg	1440
cttggtcgt ga	1452

<210> 29

<211> 1452

5 <212> ADN

<213> Azospirillum sp. B510

<400> 29

ES 2 543 568 T3

gtgaccaact tggacagccg gcattggatc gacggcgcat gggtgcccgg caccgaccgc	60
ttcgccagca tcaaccccgc cgacggcagc gtccctggcc atgccgccga tggcggacgc	120
gccgaggccg aggccgccat cgccgcgcg catgccgcct tcaatcggcc cgactgggcg	180
caaaaccocg ggctgcgcca atccatcctg ctcggtggg ccgacaggct cgacacgcaa	240
gccgaggatc tcgcccgcct gctgacgctg gagaacggca aggccatcgc ccagtgcggg	300
ggcgagatcg ccggcgccat ttcggaatc cgctattacg gggcctcgc tcgccatgtt	360
cccggccatg tgctggaggt cgagccgggc gtgctgtcca ccatgctgcg agagccggcc	420
ggcgtggcgg cgctgatcat tcogtggat ggcggggcgg tgctgctggc gggggccatc	480
ggtccggcgc tggcctgcgg ctgcaccgtc gtggtcaagc cggcggcaca gaccaccttg	540
ctgaccgccg ccttctcgcg cgccctgtcc gaggttccca gcctgccccg cggcgtctgc	600
aacatgatca gcgagaccgg ccatgccgcc gcccccgcc tggtcgactc gccgttggtc	660
gacgtggtea gcttcaccgg ctccaccgcc accggcaagc gcatcatggt cgcggcggcg	720
gacaccatga agaagctgtc gctggagctg ggcggcaaga gctgctgctt ggtcttcccc	780
gacgccgatc cggcagagac cgcggcgcg atgccaccg ccgccaccat catctccggc	840
cagcaatgca ccgcccccg ccgggtgctg gtccatgcat cggccttcga cgggatgaag	900
acacatctgc gggcagcgtt ggcgccatg acgggtgggga acggcctcga tccggcgata	960
cggatgggac cgctgatcga ccggccggcg cgggatcagg tgcagacca ggtcgaacgg	1020
gccttcgatg cctgcgacga ggtgctgctg cgtggcgggg tgccgacgga cagccccgcc	1080
gccgcgtcgt tcctcaccoc atcgtcgtg ggcacgacg acccctccgc cttcttctgc	1140
caggacgaga ttttcggccc ctctcgtcgt ctggagcgtc tcgagaccga agcggaggcg	1200
gtggccaagg ccaacaacac cgtgttcggc ctgtggcca gcgtctggac ccgcgacggc	1260
gcccgggcgt tgccgatggc ccgggcgttg cgcaatggaa ccgtctggat caacgaccac	1320
aaccgcctgt tcgccgaggc ggaaacagc ggctaccggc aaagcggcct gggcgggtg	1380
cacggctacg atgccttcgc cgatttcaog gaactgaagc atgtttgcca gacggtggga	1440
accattggct ga	1452

<210> 30

<211> 1443

5

<212> ADN

<213> Pseudomonas putida

<400> 30

ES 2 543 568 T3

```

atgcaaagcc aacactacat cgatggacaa tggacctcaa cggatcgttg gaccgacagt      60
ctcgaccogg caagoggoga actgatogga tgctttgocg acggcggcga ggccgaagct      120
gaagcggcgg tggcagcggc cgcccagacc ttcaacgacc cgcaatgggc gcagaatcca      180
cgctgcgtc agcagctatt gctggagtgg gctgcgggcc tgaaggcggc tcaggaacaa      240
ctcgcgcaat tgctgacggc cgagaacggc aaggcattgg cgcagtcccg aggagagatc      300
ggcggggcta tttccgagat tctttactac gccggcctgg cccggcataa cccggggcac      360
atgctagaag tggcgccogg tgagttctcc agcatgctgc gogaaccogc cgggggtggcg      420
gggctgatca ttccatggaa tgcccoggcg gtgctgctgg ttcgtgctct ggcgccagcg      480
atcgcggcgg gttgtacagt ggtgatcaaa cccgcgcgc aaaccgcgct attcaatgct      540
gcgatgctcg aaccgctgtt cgcattgcca ggctgcocgg ctggggcggg caatctattc      600
gccgagagcg ggcattgcgg ggctgcgcac ttgggtggctt cggcacgggt ggacgtgctc      660
agcttcaccg gctcgactgc tacgggcccag cgcacatgc gcgactgcgc ggcgactatg      720
aagaagctct ctctggaact gggcgggaaa tcttgctgcc tggctcttga agatgcogat      780
attgcgcgca tcgcgcgaa actggcagcg gcggccacca ttatctctgg ccagcaatgc      840
actgcgcgcc ggcgcgtact ggtgcatgcc agccgtttcg ccgaaatgaa aaccgcgctg      900
agcgcgcgac tggggcagat ccgcctgggt aatggcctcg atccagcgaa caacatggga      960
ccgctgatcg actggcattc cggggacagc gttgagcggc gcattggcga ggcgctggac      1020
agctgtgacg aagtgtgctt ggccggaggt cgcocgagc gtgaaactgag caaaggcgca      1080
tttctcgcgc cgtcgtgat cgcctatcgc gacagcagcg cattcttttg tcaggaagaa      1140
atthttggcc cgttgctggt actggagtcg ttogaagacg aaaccgaggc tgtagcgcgc      1200
gccaaccata ccgagttcgg tctgtccgcg agtgtctgga ccgatcaggg tgctcgcgcc      1260
tggcgtgtcg ctogggcact ggcgaacggg acggtgtggc tcaacgacca caacagattg      1320
tttgccgagg cggaaaccgg tggctatcga aagagcggtc tggggcgctt gcatggcgtc      1380

gatgcgctgc tggatttcag cgagctcaaa cacatttacc agaacgtcgg cacgctcggg      1440
tga                                                                                   1443

```

<210> 31

<211> 1461

5 <212> ADN

<213> Rhodopseudomonas palustris BisB18

<400> 31

ES 2 543 568 T3

```

atgggcatga ctgctctgca tgcggataat ctgatogaag ggcgctggca gccggcgcaa      60
tccggcgcca cggcgccgag ccttgatccg tcgagcggcg gcacgatcgg tggcttcggg      120
gctggtggcg cggcggatgc gcaggccgcg gtagcggcgcg cgcggcgcgc cttcgagcgg      180
ccggagtggg cgcagaacct gcggggcgcg cagatggtga tgctgcgctg ggccgaccgg      240
atggaggcgc aggcgatca gctcgcccg cctgctgacac tggagaacgg caagccgctg      300
ccgcagtcgc gggcgcaaat cgcggcgagc gtctcggaga tccgttatta cgcggggctc      360
acgcgttaca tccccggcca cgtgttcgag gtogagccgg gcagtttctc gacgttggtg      420
aaagaaccgg ccggcgtcgc cggactgatc attccgtgga acgcgccggc ggtgctgctg      480
atccgcgcgc tgacgcgggc gctcgccgcc ggotgcaccg ttgtgatcaa gccggcgccg      540
cagaccgcgc agatcacgcg cgcgatcacc aaatgcctgc acgaggtcga cgggctgccg      600
cgcggcgtgg tcaatctggt cagcggagcag gcccaccaag tggccgagca tctggtgacc      660
tcgaacgacg tcgacgtcat cagcttcacc ggctccaacg ccacgggcgc gggatcatg      720
gcggcggcgcg cgcgaccat gaagaaactg tcgctcgagc tcggcggcaa atccgcttgc      780
ctggtgttcg acgacgccga tategccgac gtggctccga agctcgccgc tgccgccacc      840
atcatcgccg gccagcaatg caccgccgcg cgcgcgctgc ttgtgcatgc ctgcgctac      900
gacgagatga aggcggcgtc gaaagccgcg ctggcgaaca tccgattgc gccgggcagc      960
gcggcggggcg ccgagatggg tccgctgatc gacgcagcct cgctggccgc cgtggcgaag     1020
cgcgccgacg aggcgatgca ggccgccgac gaggtggtgc tgccgcggcg caggcccgcc     1080
ggcgacctcg ccaacgggta ctctctgctg ccgacgctcg tggcgcacatg cgacacgtcg     1140
gcgttcttcg tccaggaaga gatttttggg ccgctggtgg tgctcgaaaa attcaggac      1200
gagaaggagg cggtaggccc cgccaatcac agcgactacg ggctgtccgc cagcgtgtgg      1260
acccatgacg gcgcccgccg gatgcgggtg gcgagggcgc tgcgcaacgg cacggtgtgg      1320
atcaacgacc acaacaagct attcgccgaa gccgagaccg gcggttatcg ccgcagcggg      1380
ctcgcccgcc tgcacggcta tgacgcgctg atcgacttcc tcgagatcaa gcacgtctat      1440
cagagctgcg gcgtagtgta g                                     1461

```

<210> 32

<211> 1458

5 <212> ADN

<213> Dinoroseobacter shibae DFL 12

<400> 32

ES 2 543 568 T3

atgacaacga cggacctgat cgcgcggcac atgatcggcg gtcctattc ggacgcgggc 60  
gacaagatcg cgtcgatcaa cccggcgacc ggggoggtcg ttgggcatgt ccgggcccgc 120  
ggcgcggcgc aggccacagc cgcattgcc gcgcgcggg ccgcattoga cacaacgctc 180  
tggccgcaat caccocgcga ccggcagatg gccctgctgc ggtgggcccga cgcgctggag 240  
gccgatctcg cacggcttgc ggaattgctg acgttgacca atggcaaacc gcttggcgcg 300  
tccaagggcg aattgggccc ggcgatatcg gaaatccgat actatgcggg cctgaccocg 360  
cacaatcccg gccacgcgat ggaagtggcg ccgggcgagc ttccggtgat gctgcgcgag 420  
ccggccgggg tcgcgggcat catcgtgccg tggaaacgcac ccgcctgct gctgatccgg 480  
tcgctcgcct cggcctggc ggtgggctgc acgaccgtga cgaaccocgc gccgcagacg 540  
gcgctgttca ccgcggcctg catggccccg ctgttcgagg atgcgcctat ccccgcgggc 600  
gtggtcaatg tggttttcga ggtcggctcat gacgcagcgc aaacgctggt cacatcgccc 660  
gatgtcgatg tcatcagctt tacccgctcg aacgcggctg gccagcggat catggcggac 720  
gccgcgccga cgatgaagaa actgtcgcctg gaactgggog gcaaatcctg ctgcatcgtg 780  
ctggatgatg cagacatcgg cgtggtggca ccgaaacttg ccgcgcagc cacgatcctc 840  
tccggccagc aatgcaccgc ggccgcctcg gttctggtgc acgaatcccg gcttgatgaa 900  
gcgaaaagcg cgttttcgc tgccttcag gctgtctcca tcggcgacgg catgtcagac 960  
ggcaactcga tggggccgct gatcgacatc cagtcgcgcg accgggtcat gcgtgattgc 1020  
ggaacggtat acgataccgc cgcagaggtc gtgctgcgcg gcggcccgct cgcaggtccg 1080  
aagggcagcg cgttcatgct gcgcggcctg gtgctgcata gcgatccgaa cgcagcttc 1140  
gtacaggacg aaattttcgg cccgcttctg gtgcttgaga cgttcgcgaga cgcagcggat 1200  
gcggtggcaa aggcaaaaa caccgtctat ggcttgtccg catcaatctg gacccatcgc 1260  
ggcgatgcgt catggcggct ggccgcctgc ctgcgcaatg gcaccgtctg gatcaacgac 1320  
cacaacaggc ttttcgcga agccgagacc ggcggctatc gccttcggg cctgggcaga 1380  
ttgcaocggt tcgacggtct gctcgatttc tgcgagctca agcaocgtta ccagaacgctc 1440  
ggtgtcgtcg gtcactga 1458

<210> 33

<211> 34

5 <212> ADN

<213> sintético

<400> 33

atgcggccc aacaaggaga agatggaatg aacg 34

<210> 34

10 <211> 28

<212> ADN

## ES 2 543 568 T3

<213> sintético

<400> 34

atgcggccgc gcgtcgggt cggtgcta 28

## REIVINDICACIONES

1. Célula modificada genéticamente que comprende:

(i) una primera secuencia de polinucleótidos que codifica un primer polipéptido que tiene capacidades de transporte de ácido 5-(hidroximetil)furan-2-carboxílico (ácido HMF), que comprende una secuencia de aminoácidos que tiene al menos 45 %, preferentemente al menos 60 %, tal como al menos 70 %, más preferentemente al menos 80 %, tal como 90 %, más preferentemente al menos 95 % de similitud de secuencia con una secuencia de aminoácidos definida en las SEC ID N° 1, 2, 3 o 4; e

(ii) una segunda secuencia de polinucleótidos que codifica un segundo polipéptido que tiene actividad convertidora de ácido HMF;

**caracterizado porque** dicha célula es modificada genéticamente mediante la introducción funcional de al menos la primera secuencia de polinucleótidos y, preferentemente, mediante la introducción funcional de la primera y la segunda secuencia de polinucleótidos.

2. La célula modificada genéticamente de acuerdo con la reivindicación 1, en la que el segundo polipéptido es un ácido 5-(hidroximetil)furan-2-carboxílico oxidorreductasa, preferentemente un polipéptido que tiene una secuencia de aminoácidos que tiene al menos 45 %, preferentemente al menos 60 %, tal como al menos 70 %, más preferentemente al menos 80 %, tal como 90 %, más preferentemente al menos 95 % de similitud de secuencia con una secuencia de aminoácidos definida en las SEC ID N° 5 o 6.

3. Una célula modificada genéticamente de acuerdo con cualquiera de las reivindicaciones 1-2, en la que la primera secuencia de polinucleótidos comprende una secuencia de nucleótidos que tiene al menos 45 %, preferentemente al menos 60 %, tal como al menos 70 %, más preferentemente al menos 80 %, tal como 90 %, más preferentemente al menos 95 % de similitud de secuencia con una secuencia de nucleótidos definida en las SEC ID N° 7, 8, 9 o 10.

4. Una célula modificada genéticamente de acuerdo con cualquiera de las reivindicaciones 1-3, en la que la segunda secuencia de polinucleótidos comprende una secuencia de nucleótidos que tiene al menos 45 %, preferentemente al menos 60 %, tal como al menos 70 %, más preferentemente al menos 80 %, tal como 90 %, más preferentemente al menos 95 % de similitud de secuencia con una secuencia de nucleótidos definida en las SEC ID N° 11 o 12.

5. Una célula modificada genéticamente de acuerdo con cualquiera de las reivindicaciones 1-4, que comprende una tercera secuencia de polinucleótidos que codifica un tercer polipéptido que comprende una secuencia de aminoácidos que tiene al menos 45 %, preferentemente al menos 60 %, tal como al menos 70 %, más preferentemente al menos 80 %, tal como 90 % más preferentemente al menos 95 % de similitud de secuencia con una secuencia de aminoácidos definida en las SEC ID N° 19, 20, 21, 22, 23, 24 o 25.

6. Una célula modificada genéticamente de acuerdo con cualquiera de las reivindicaciones 1-5, en la que la tercera secuencia de polinucleótidos comprende una secuencia de nucleótidos que tiene al menos 45 %, preferentemente al menos 60 %, tal como al menos 70 %, más preferentemente al menos 80 %, tal como 90 %, más preferentemente al menos 95 % de similitud de secuencia con una secuencia de nucleótidos definida en las SEC ID N° 26, 27, 28, 29, 30, 31 o 32.

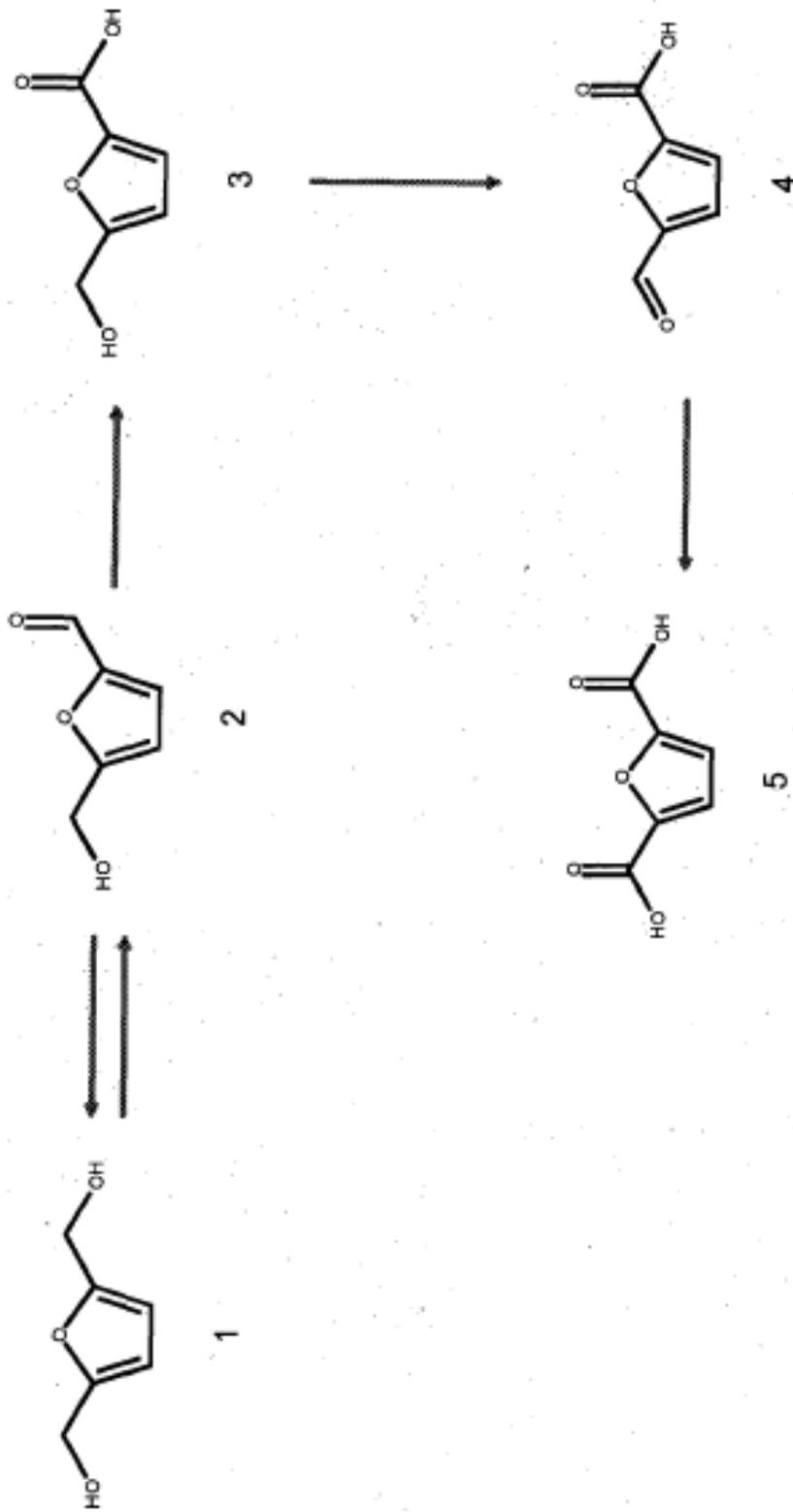
7. Una célula modificada genéticamente de acuerdo con cualquiera de las reivindicaciones 1-6, en la que la célula es una célula procarionta tal como una célula bacteriana o una célula eucariota, tal como una célula de levadura, una célula fúngica, una célula vegetal o una célula animal.

8. Una célula modificada genéticamente de acuerdo con cualquiera de las reivindicaciones 1-7, en la que la célula es una célula procarionta, preferentemente una célula bacteriana seleccionada de los géneros *Escherichia*, *Anabaena*, *Caulobacter*, *Gluconobacter*, *Rhodobacter*, *Pseudomonas*, *Paracoccus*, *Bacillus*, *Brevibacterium*, *Corynebacterium*, *Rhizobium* (*Sinorhizobium*), *Bradyrhizobium*, *Flavobacterium*, *Klebsiella*, *Enterobacter*, *Lactobacillus*, *Lactococcus*, *Methylobacterium*, *Staphylococcus*, *Streptomyces*, *Zymomonas*, *Acetobacter*, *Streptococcus*, *Bacteroides*, *Selenomonas*, *Megasphaera*, *Burkholderia*, *Cupriavidus*, *Ralstonia*, *Methylobacterium*, *Methylovorus*, *Rhodopseudomonas*, *Acidiphilium*, *Dinoroseobacter*, *Agrobacterium*, *Sulfolobus* o *Sphingomonas*, más preferentemente seleccionada del grupo que consiste en *Bacillus subtilis*, *Bacillus amyloliquefaciens*, *Bacillus licheniformis*, *Bacillus pumilus*, *Bacillus megaterium*, *Bacillus halodurans*, *Bacillus pumilus*, *Gluconobacter oxydans*, *Caulobacter crescentus*, *Methylobacterium extorquens*, *Methylobacterium radiotolerans*, *Methylobacterium nodulans*, *Rhodobacter sphaeroides*, *Pseudomonas zeaxanthinifaciens*, *Pseudomonas putida*, *Pseudomonas putida* S12, *Paracoccus denitrificans*, *Escherichia coli*, *Corynebacterium glutamicum*, *Staphylococcus carnosus*, *Streptomyces lividans*, *Sinorhizobium melliloti*, *Bradyrhizobium japonicum*, *Rhizobium radiobacter*, *Rhizobium leguminosarum*, *Rhizobium leguminosarum* bv. *trifolii*, *Agrobacterium radiobacter*, *Cupriavidus basilensis*, *Cupriavidus necator* (*Ralstonia eutropha*), *Ralstonia pickettii*, *Burkholderia phytofirmans*, *Burkholderia phymatum*, *Burkholderia xenovorans*, *Burkholderia graminis*,

*Rhodopseudomonas palustris*, *Acidiphilium cryptum*, *Dinoroseobacter shibae*, *Sulfolobus acidocaldarius*, *Sulfolobus islandicus*, *Sulfolobus solfataricus*, *Sulfolobus tokodaii*.

- 5 9. Una célula modificada genéticamente de acuerdo con cualquiera de las reivindicaciones 1-8, en la que la célula es una célula huésped eucariota, por ejemplo una célula de mamífero, de insecto, vegetal, fúngica o de alga, tal como una célula de mamífero seleccionada de células de ovario de hámster chino (CHO), células COS, células 293, células PerC6 o hibridomas, una célula de insecto seleccionada de células Sf9 o Sf21 y derivadas de los mismos, preferentemente una célula eucariota seleccionada de células fúngicas o de levadura, tales como especies de *Candida*, *Hansenula*, *Kluyveromyces*, *Pichia*, *Saccharomyces*, *Schizosaccharomyces* o *Yarrowia*, con mayor preferencia una célula eucariota seleccionada de *Kluyveromyces lactis*, *Saccharomyces cerevisiae*, 10 *Harzsenula polymorpha*, *Yarrowia lipolytica*, *Pichia stipitis* o *Pichia pastoris* o una célula de hongo filamentoso de una especie seleccionada de la subdivisión *Eumycota* y *Oomycota* o, como alternativa, una especie del género *Acremonium*, *Agaricus*, *Aspergillus*, *Aureobasidium*, *Chrysosporium*, *Coprinus*, *Cryptococcus*, *Filibasidium*, *Fusarium*, *Humicola*, *Magnaporthe*, *Mucor*, *Myceliophthora*, *Neocallimastix*, *Neurospora*, *Paecilomyces*, *Penicillium*, *Piromyces*, *Phanerochaete*, *Pleurotus*, *Schizophyllum*, *Talaromyces*, *Thermoascus*, 15 *Thielavia*, *Tolyptocladium* o *Trichoderma* tal como *Aspergillus niger*, *Aspergillus awamori*, *Aspergillus foetidus*, *Aspergillus sojae*, *Aspergillus fumigatus*, *Talaromyces emersonii*, *Aspergillus oryzae*, *Chrysosporium lucknowense*, *Trichoderma reesei* o *Penicillium chrysogenum*.
- 20 10. Una célula modificada genéticamente de acuerdo con cualquiera de las reivindicaciones 1-9, en la que la primera y la segunda secuencia de polinucleótidos se localizan en un único vector opcionalmente un vector que además comprende la tercera secuencia de polinucleótidos.
11. Un vector que comprende:
- (i) una primera secuencia de polinucleótidos que codifica un primer polipéptido que comprende una secuencia de aminoácidos que tiene al menos 45 %, preferentemente al menos 60 %, tal como al menos 70 %, más preferentemente al menos 80 %, tal como 90 %, más preferentemente al menos 95 % de similitud de secuencia con una secuencia de aminoácidos definida en las SEC ID N° 1, 2, 3 o 4; y
- 25 (ii) una segunda secuencia de polinucleótidos que codifica un segundo polipéptido que tiene actividad convertidora de ácido 5-(hidroximetil)furan-2-carboxílico (ácido HMF).
12. Un vector de acuerdo con la reivindicación 11, en la que el segundo polipéptido es una ácido 5-hidroximetilfuroico oxidoreductasa, preferentemente un polipéptido que tiene una secuencia de aminoácidos que tiene al menos 45 %, preferentemente al menos 60 %, tal como al menos 70 %, más preferentemente al menos 80 %, tal como 90 %, más preferentemente al menos 95 % de similitud de secuencia con una secuencia de aminoácidos definida en la SEC ID N° 5 o 6.
- 30 13. Un vector de acuerdo con cualquiera de las reivindicaciones 11-12, en la que la primera secuencia de polinucleótidos comprende una secuencia de nucleótidos que tiene al menos 45 %, preferentemente al menos 60 %, tal como al menos 70 %, más preferentemente al menos 80 %, tal como 90 %, más preferentemente al menos 95 % de similitud de secuencia con una secuencia de nucleótidos definida en las SEC ID N° 7, 8, 9 o 10.
- 35 14. Un vector de acuerdo con cualquiera de las reivindicaciones 11-13, en la que la segunda secuencia de polinucleótidos comprende una secuencia de nucleótidos que tiene al menos 45 %, preferentemente al menos 60 %, tal como al menos 70 %, más preferentemente al menos 80 %, tal como 90 %, más preferentemente al menos 95 % de similitud de secuencia con una secuencia de nucleótidos definida en las SEC ID N° 11 o 12.
- 40 15. Un vector de acuerdo con cualquiera de las reivindicaciones 11-14, que comprende una tercera segunda secuencia de polinucleótidos que codifica un tercer polipéptido que comprende una secuencia de aminoácidos que tiene al menos 45 %, preferentemente al menos 60 %, tal como al menos 70 %, más preferentemente al menos 80 %, tal como 90 %, más preferentemente al menos 95 % de similitud de secuencia con una secuencia de nucleótidos definida en las SEC ID N° 19, 20, 21, 22, 23, 24 o 25.
- 45 16. Un vector de acuerdo con cualquiera de las reivindicaciones 11-15, en la que la tercera secuencia de polinucleótidos comprende una secuencia de nucleótidos que tiene al menos 45 %, preferentemente al menos 60 %, tal como al menos 70 %, más preferentemente al menos 80 %, tal como 90 %, más preferentemente al menos 95 % de similitud de secuencia con una secuencia de aminoácidos definida en las SEC ID N° 26, 27, 28, 29, 30, 31 o 32.
- 50 17. Procedimientos para convertir el ácido 5-(hidroximetil)furan-2-carboxílico (ácido HMF) que comprenden incubar una célula de acuerdo con una cualquiera de las reivindicaciones 1-10 en presencia de ácido HMF en condiciones adecuadas para la conversión por dicha célula del ácido HMF.
- 55 18. Procedimiento de acuerdo con la reivindicación 17, en el que el ácido HMF es formado in situ a partir de uno o más precursores furánicos de ácido HMF en condiciones adecuadas para la conversión por dicha célula del uno o más precursores furánicos en ácido HMF.

19. Procedimiento de acuerdo con cualquiera de las reivindicaciones 17-18, en el que el segundo polipéptido de la célula es una ácido HMF oxidoreductasa y que además comprende someter la célula a condiciones adecuadas para la conversión de ácido HMF en FDCA.
- 5 20. El procedimiento de acuerdo con cualquiera de las reivindicaciones 17-19, en el que un precursor furánico es seleccionado del grupo que comprende ácido 5-(hidroximetil)furan-2-carbaldehído (HMF), furan-2,5-dicarbaldehído (DFF), [5-(hidroximetil)furan-2-il]metanol (alcohol HMF) y es, preferentemente, HMF.
21. Procedimiento de acuerdo con cualquiera de las reivindicaciones 17-20, en el que un precursor furánico es obtenido a partir de uno o más azúcares hexosa, preferentemente uno o más azúcares hexosa obtenidos de biomasa lignocelulósica, por ejemplo mediante deshidratación catalizada por ácido.
- 10 22. Procedimiento de acuerdo con cualquiera de las reivindicaciones 17-21, en el que el FDCA es recuperado de la mezcla de reacción mediante un procedimiento que comprende precipitación en ácido, seguido de cristalización por enfriamiento o, como alternativa, seguido de extracción en disolvente.
23. Uso de una célula modificada genéticamente de acuerdo con cualquiera de las reivindicaciones 1-10, para la biotransformación de una serie de precursores furánicos en FDCA.
- 15 24. Uso de acuerdo con la reivindicación 23, en el que el número de precursores furánicos es seleccionado del grupo que comprende 5-(hidroximetil) furan-2-carbaldehído (HMF), [5-(hidroximetil)furan-2-il]metanol (alcohol HMF), ácido 5-(hidroximetil) furan-2-carboxílico (ácido HMF) furan-2,5-dicarbaldehído (DFF) o ácido 5-formilfuran-2-carboxílico (FFA).



**FIG. 1.**

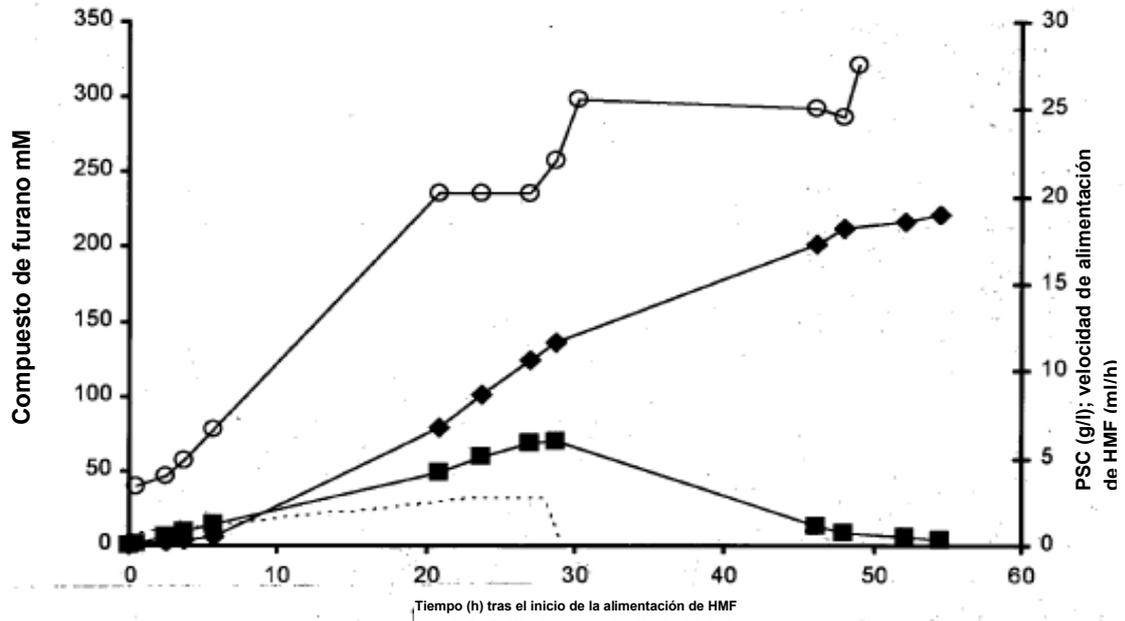


FIG. 2a

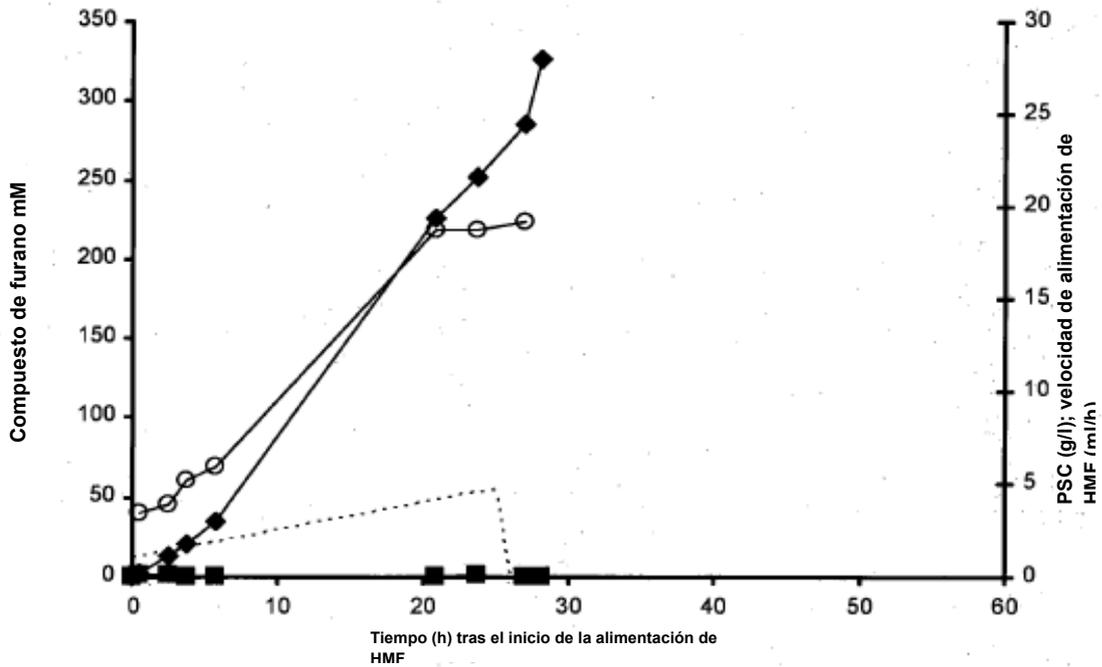
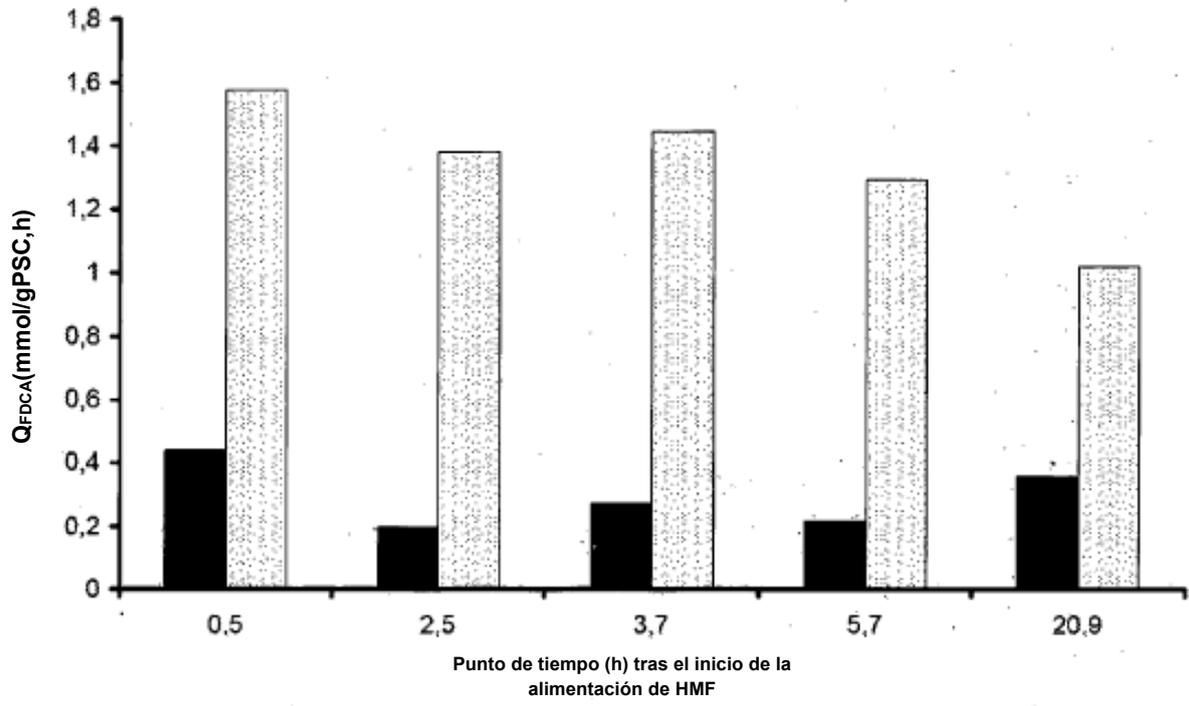


FIG. 2b



**FIG. 3.**

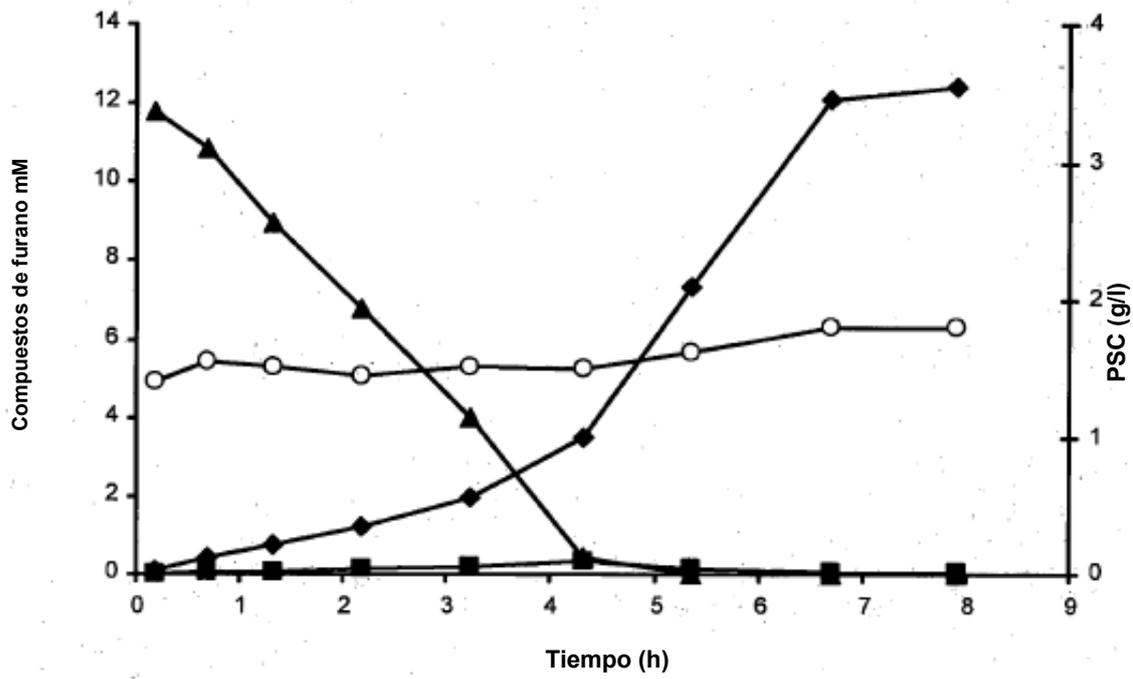


FIG. 4a

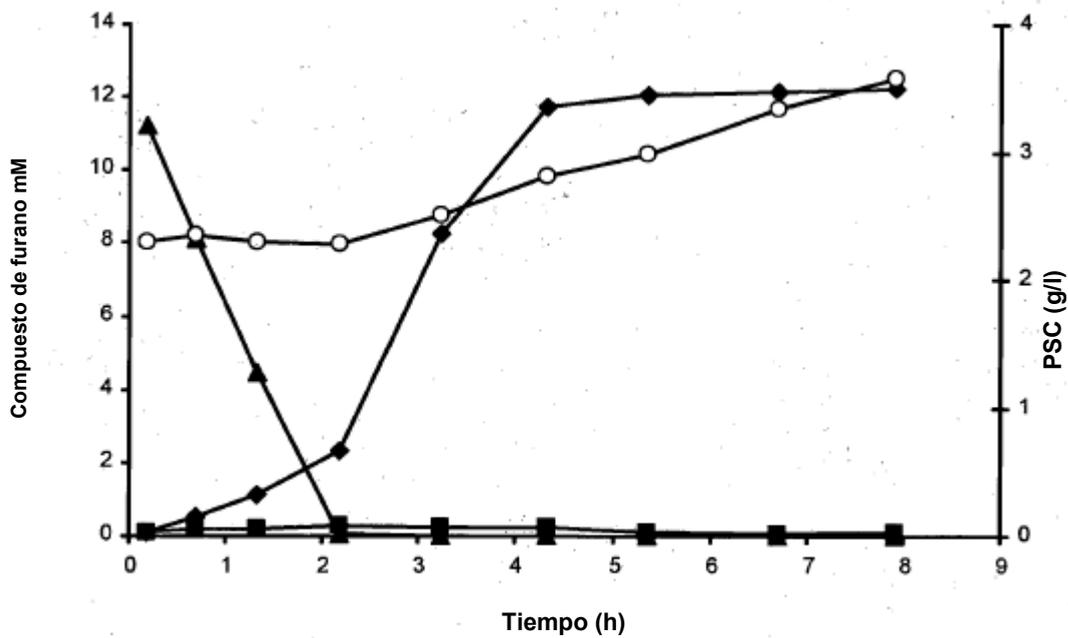


FIG. 4b.

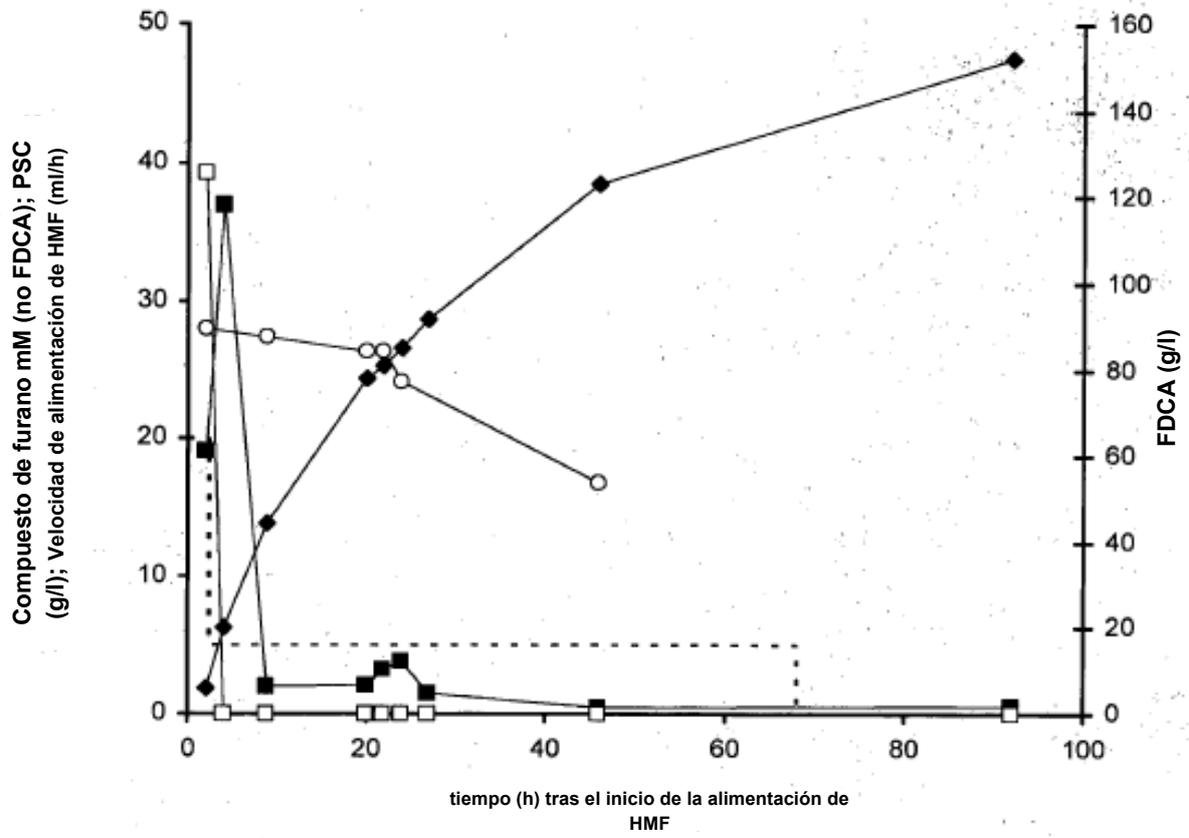


FIG. 5.

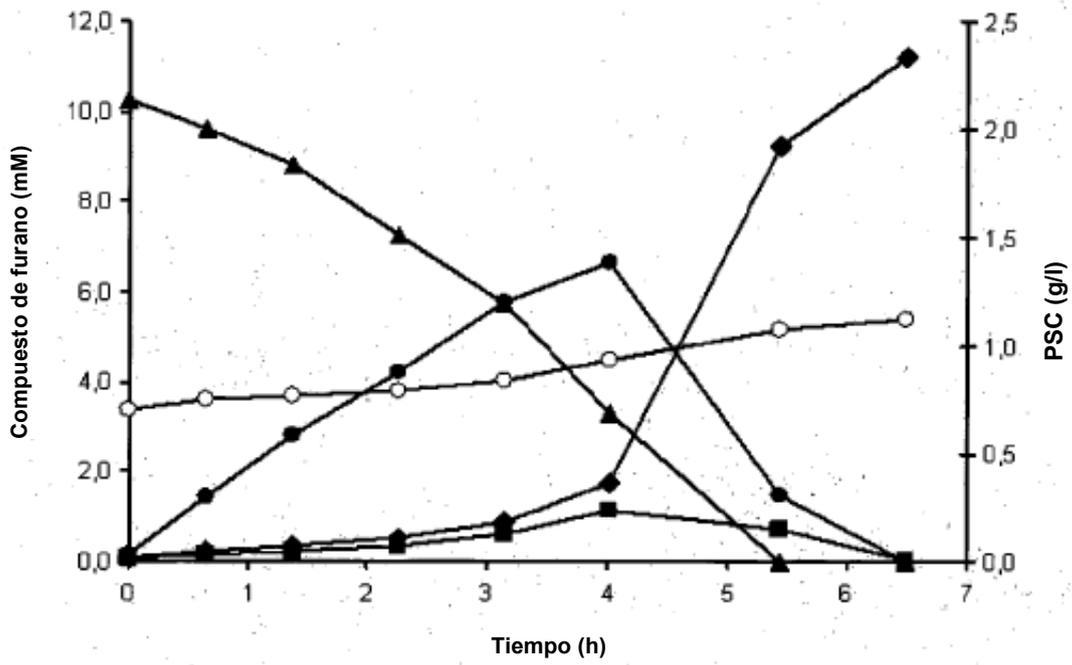


FIG. 6a

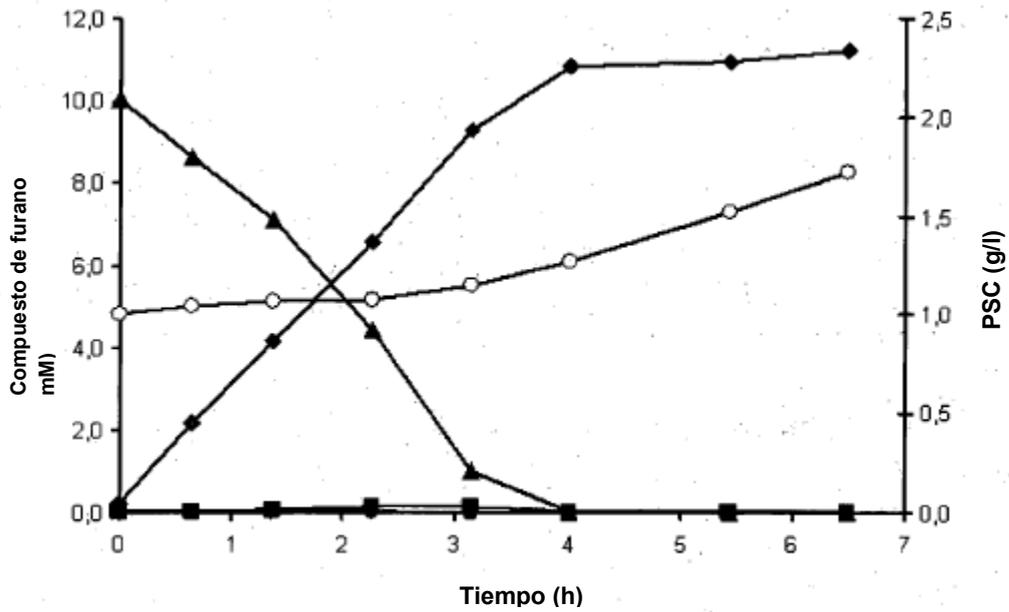


FIG. 6b.

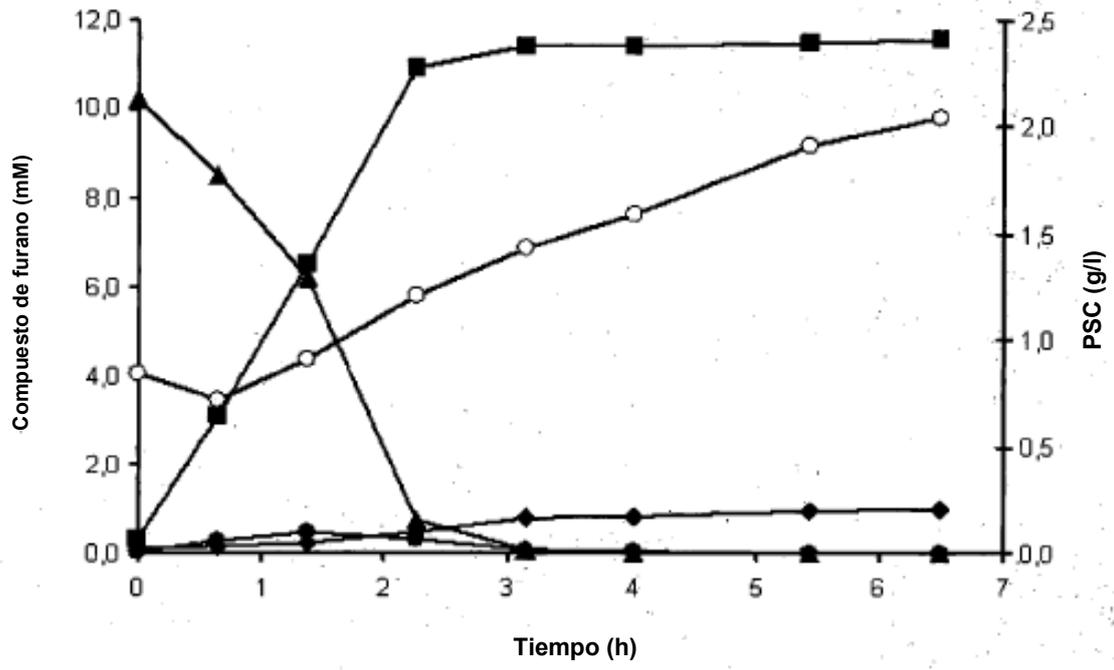


FIG. 6c