

19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 463 486**

51 Int. Cl.:

**C12N 9/90** (2006.01)

**C12P 7/10** (2006.01)

**C12N 15/81** (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **20.03.2008 E 08734708 (4)**

97 Fecha y número de publicación de la concesión europea: **12.03.2014 EP 2139998**

54 Título: **Vector con genes de codones optimizados para una vía metabólica de arabinosa para la conversión de arabinosa en levadura para la producción de etanol**

30 Prioridad:

**05.04.2007 DE 102007016534**

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

**28.05.2014**

73 Titular/es:

**BUTALCO GMBH (100.0%)  
Mettlenstrasse 14  
6363 Fürigen , CH**

72 Inventor/es:

**BOLES, ECKHARD y  
ROTHER, BEATE**

74 Agente/Representante:

**VALLEJO LÓPEZ, Juan Pedro**

**ES 2 463 486 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

Vector con genes de codones optimizados para una vía metabólica de arabinosa para la conversión de arabinosa en levadura para la producción de etanol

5

**1. Resumen**

La presente invención se refiere a células hospedadoras que contienen nuevos casetes de expresión y vectores de expresión, que comprenden tres secuencias de ácido nucleico para araA, araB y araD, que codifican cada una un polipéptido de una vía metabólica bacteriana de L-arabinosa. La invención particularmente se refiere a células hospedadoras que contienen casetes de expresión y vectores de expresión, que comprenden secuencias de ácido nucleicos de codones optimizados para araA, araB y araD. Las células hospedadoras son, en particular, cepas modificadas de levadura que expresan los polipéptidos para la vía metabólica bacteriana de L-arabinosa. Cuando se usan estas células hospedadoras modificadas, la arabinosa se fermenta de forma más eficaz por estas células, en particular en etanol. La presente invención por lo tanto es relevante, entre otras cosas, en relación a la producción de agentes bioquímicos a partir de biomasa, tal como bioetanol, por ejemplo.

10

15

**2. Antecedentes de la invención**

La levadura de la cerveza, del vino y del pan *Saccharomyces cerevisiae* se ha usado ya durante siglos para la producción de pan, vino y cerveza debido a su característica de fermentar azúcar en etanol y dióxido de carbono. En biotecnología, *S. cerevisiae* se usa particularmente en la producción de etanol para fines industriales, además de la producción de proteínas heterólogas. El etanol se usa en numerosas ramas de la industria como sustrato inicial para síntesis. El etanol está adquiriendo una importancia creciente como combustible alternativo, debido a la presencia cada vez más escasa de petróleo, los crecientes precios del petróleo y la necesidad continuamente creciente de combustible en todo el mundo.

20

25

Para producir bioetanol de forma barata y eficaz, se plantea en sí mismo el uso de biomasa que contiene lignocelulosa, tal como por ejemplo paja, residuos de la industria maderera y la agricultura y el componente orgánico de los residuos domésticos diarios, como sustrato inicial. En primer lugar, dicha biomasa es muy conveniente y en segundo lugar está presente en grandes cantidades. Los tres componentes principales de la lignocelulosa son la lignina, la celulosa y la hemicelulosa. La hemicelulosa, que es el segundo polímero que existe de forma más frecuente después de la celulosa, es un heteropolímero muy ramificado. Consta de pentosas (L-arabinosa, D-xilosa), ácidos urónicos (ácido 4-O-metil-D-glucurónico, ácido D-galacturónico) y hexosas (D-manosa, D-galactosa, L-ramnosa, D-glucosa) (véase la Figura 1). Aunque la hemicelulosa puede hidrolizarse más fácilmente que la celulosa, sin embargo contiene las pentosas L-arabinosa y D-xilosa, que normalmente no pueden convertirse por la levadura *S. cerevisiae*.

30

35

Para ser capaces de usar pentosas para fermentaciones, primero deben penetrar en la célula a través de la membrana plasmática. Aunque *S. cerevisiae* no es capaz de metabolizar D-xilosa, puede captar D-xilosa en el interior de la célula. Sin embargo, *S. cerevisiae* no tiene un transportador específico. El transporte tiene lugar mediante numerosos transportadores de hexosa. La afinidad de los transportadores por D-xilosa es, sin embargo, claramente inferior que por D-glucosa (Kotter y Ciriacy, 1993). En levaduras que son capaces de metabolizar D-xilosa, tales como por ejemplo *P. stipitis*, *C. shehatae* o *P. tannophilus* (Du Preez et al., 1986), existen tanto transportadores de baja afinidad no específicos, que transportan D-glucosa, como también simportadores de protones de alta afinidad específicos solamente para D-xilosa (Hahn-Hagerdal et al., 2001).

40

45

En experimentos anteriores, se descubrieron algunas levaduras, tales como por ejemplo *Candida tropicalis*, *Pachysolen tannophilus*, *Pichia stipitis*, *Candida shehatae*, que de forma natural fermentan L-arabinosa o al menos pueden asimilarla. Sin embargo, estas levaduras carecen completamente de la capacidad de fermentar L-arabinosa en etanol, o solamente obtienen un rendimiento muy bajo de etanol (Dien et al., 1996).

50

*Conversión de L-arabinosa*

Para que la pentosa L-arabinosa se metabolice por *S. cerevisiae*, debe entrar en la célula mediante proteínas de transporte y convertirse en el metabolito D-xilulosa-5-fosfato en tres etapas enzimáticas. Estas tres etapas enzimáticas pueden quedar a disposición de la levadura mediante genes expresados de forma heteróloga. La D-xilulosa-5-fosfato funciona como intermedio de la vía de pentosa fosfato y puede descomponerse adicionalmente para producir etanol en condiciones anaeróbicas en la célula (véase la figura 2).

60

Becker y Boles (2003) describen el diseño por ingeniería y la selección de una cepa de laboratorio de *S. cerevisiae* que es capaz de usar L-arabinosa para el crecimiento y para fermentarla en etanol. Esto fue posible debido a la sobre-expresión de una vía metabólica bacteriana de L-arabinosa, que consta de AraA de *Bacillus subtilis* y AraB y AraD de *Escherichia coli* y la sobre-expresión simultánea de la permeasa de levadura de galactosa que transporta L-arabinosa en la cepa de levadura. El análisis molecular de la cepa seleccionada mostró que la precondición predeterminante para el uso de L-arabinosa es una menor actividad de L-ribulocinasa. Sin embargo, entre otras

65

cosas, se ha constatado un crecimiento muy lento a partir de esta cepa de levadura (véase la Figura 2).

El documento WO 2006/096130 A1 (Boles et al.) describe cepas de *S. cerevisiae* que fermentan arabinosa y xilosa, que contienen los genes de la vía metabólica de xilosa y arabinosa, donde la vía metabólica de arabinosa consta de los genes para la L-arabinosa isomerasa (araA) de *B. subtilis*, la L-ribuloquinasa (araB) de *E. coli* y la L-ribulosa-5-P-4-epimerasa (araD) de *E. coli*.

El documento WO 98/50524 A1 describe el desarrollo de plásmidos que contienen los genes de *E. coli* de la vía metabólica de xilosa y arabinosa, en particular la xilosa isomerasa, xiluloquinasa, L-arabinosa isomerasa, L-ribuloquinasa y L-ribulosa-5-P-4-epimerasa, transaldolasa y transcetolasa. Los plásmidos se usan para la transformación de cepas de *Zymomonas*, para posibilitar que estas cepas transformadas crezcan en xilosa y arabinosa y las conviertan en etanol.

Hasta ahora, solamente era posible expresar los genes nativos de las vías metabólicas bacterianas de arabinosa que son esenciales para metabolizar la arabinosa en *S. cerevisiae* en plásmidos individuales o integrarlos individualmente en el genoma de levadura, respectivamente (Karhumaa et al., 2006). Esto significa que cada transformante de levadura con una vía metabólica funcional de arabinosa contenía al menos tres plásmidos o los genes integrados en el locus de ADNr (Becker y Boles, 2003; Karhumaa et al., 2006).

La presencia de los genes en diferentes plásmidos está asociada con varias desventajas. Por un lado, los plásmidos que están presentes simultáneamente representan un estrés adicional para las células de levadura ("Plasmid stress", Review of *E. coli* por Bailey (1993)). Por otro lado, los plásmidos usados tienen fuertes homologías en sus secuencias, que pueden conducir a la pérdida de información dentro de los plásmidos debido a recombinación homóloga (Wiedemann, 2005). Sin embargo, las desventajas principales asociadas con el uso de plásmidos recaen en el hecho de que permanecen inestables en las cepas sin presión selectiva y que no son adecuados para uso industrial.

Además, sería ideal para aplicaciones industriales que el microorganismo usado fuera capaz de metabolizar todos los azúcares presentes en el medio. Como las levaduras actualmente usadas de forma industrial no son capaces de metabolizar la arabinosa en el medio, sería muy ventajoso proporcionar a cepas esta capacidad adicional de una forma estable.

El documento WO 03/008593 A2 muestra la secuencia de aminoácidos de una L-Arabinosa isomerasa (araA) de *Clostridium acetobutylicum* (SEC ID N° 9). Noelling et al. (2001) describen la secuencia genómica y el análisis comparativo de la bacteria productora de disolvente *Clostridium acetobutylicum*, a partir de la cual se clonó araA genómica.

En la base de datos NCBI, el N° de acceso AAU41894 muestra la secuencia de aminoácidos de una L-Arabinosa isomerasa (araA) de *Bacillus licheniformis* ATCC 14580. Veith et al. (2004) describen la secuencia completa de ácido nucleico de *Bacillus licheniformis* DSM13, a partir de la cual se clonó araA genómica.

El objetivo de la presente invención, por lo tanto, es proporcionar medios que superen las desventajas conocidas de la técnica anterior para introducir genes de una vía metabólica bacteriana de L-arabinosa en células hospedadoras individualmente, y que en particular puedan ser útiles para cepas industriales de levadura.

El objetivo se consigue de acuerdo con la invención reivindicada en las reivindicaciones.

El objetivo se consigue de acuerdo con la invención proporcionando células hospedadoras que contienen moléculas de ácido nucleico que comprenden tres secuencias de ácido nucleico, cada una de las cuales codifica un polipéptido de una vía metabólica bacteriana de L-arabinosa.

Una célula hospedadora de acuerdo con la invención es una célula fúngica.

Una célula hospedadora de acuerdo con la invención es preferiblemente una célula de levadura, tal como especies de *Saccharomyces*, *Kluyveromyces* sp., *Hansenula* sp., *Pichia* sp. o *Yarrowia* sp.

Una molécula de ácido nucleico, que está contenida en una célula hospedadora de acuerdo con la invención, es una molécula de ácido nucleico recombinante. Además, las moléculas de ácido nucleico comprenden ADNbc, ADNmc, APN, ACN, ARN o ARNm, o combinaciones de los mismos.

La "vía metabólica de L-arabinosa" o "vía metabólica bacteriana de L-arabinosa", tal como existe en *E. coli*, se muestra en la figura 2. Esta vía metabólica contiene 3 enzimas: L-arabinosa isomerasa, L-ribuloquinasa y L-ribulosa-5-P-4-epimerasa. Los genes que codifican estas enzimas se llaman araA, araB y araD. La L-arabinosa isomerasa convierte L-arabinosa en L-ribulosa, que se metaboliza adicionalmente en L-ribulosa-5-fosfato mediante la L-ribuloquinasa. Finalmente, la L-ribulosa-5-P-4-epimerasa convierte L-ribulosa-5-fosfato en D-xilulosa-5-fosfato. El metabolito intermedio D-xilulosa-5-fosfato se forma por los genes expresados de forma heteróloga de la vía

metabólica de L-arabinosa, particularmente la vía metabólica bacteriana de L-arabinosa, en la célula de levadura. La D-xilulosa-5-fosfato funciona como intermedio de la vía de pentosa fosfato y puede descomponerse adicionalmente en etanol en condiciones anaeróbicas en una célula de levadura. Las enzimas de la vía metabólica de xilosa también se encuentran en hongos, y estas y otras enzimas aisladas de eucariotas también pueden usarse como enzimas para la vía metabólica de L-arabinosa.

Las tres secuencias de ácido nucleico de las moléculas de ácido nucleico contenidas en una célula hospedadora de acuerdo con la invención, cada una de las cuales codifica un polipéptido de una vía metabólica de L-arabinosa, son araA (L-arabinosa isomerasa), araB (L-ribuloquinasa) y araD (L-ribulosa-5-P-4-epimerasa), donde la secuencia de ácido nucleico para araA se obtiene de *Bacillus licheniformis* o *Clostridium acetobutylicum*.

Las moléculas de ácido nucleico contenidas en una célula hospedadora de acuerdo con la invención comprenden secuencias de ácido nucleico que son idénticas a la secuencia de ácido nucleico de origen natural o que se han optimizado en los codones para su uso en una célula hospedadora.

Cada aminoácido está codificado por un codón. Sin embargo, existen varios codones diferentes que codifican un aminoácido individual. El código genético está, por tanto, degenerado. La selección de codones preferidos para un aminoácido correspondiente varía de un organismo a otro. Por ejemplo, pueden surgir problemas en genes expresados de forma heteróloga si el organismo hospedador o célula hospedadora tiene un uso de codones muy diferente. El gen solamente puede expresarse muy lentamente, si acaso. El diferente uso de codones incluso puede observarse en genes de diferentes vías metabólicas dentro del mismo organismo. Se sabe que los genes de glucólisis de *S. cerevisiae* se expresan con gran fuerza. Tienen un uso de codones altamente restrictivo. La adaptación del uso de codones de los genes bacterianos de la vía metabólica de arabinosa al uso de codones de los genes de la glucólisis de *S. cerevisiae* conduce a un metabolismo mejorado de arabinosa en levaduras.

Para la optimización de codones, los inventores no se basaron en las plataformas habituales de diseñadores de genes sintéticos para la expresión heteróloga (tal como Synthetic Gene Designer como se describe en Wu *et al.* 2006), en su lugar adaptaron la optimización de codones específicamente al uso de codones de los genes de glucólisis en la levadura. Los genes de glucólisis tienen un uso de codones altamente restrictivo, que está alineado con la frecuencia del correspondiente ARNt. Los genes de glucólisis usan principalmente codones para los cuales existen altas concentraciones del correspondiente ARNt, que a su vez provoca una mayor eficacia de traducción y expresión génica (Bennetzen y Hall, 1982, Hoekema *et al.*, 1987). Por el contrario, los diseñadores habituales de genes sintéticos están adaptados más al uso promedio de codones de todos los genes en un organismo, no justo aquellos que se expresan altamente, y también tienen en cuenta otros factores, tales como la estabilidad. Por consiguiente, la optimización de codones con la ayuda de dicha plataforma electrónica, tal como la descrita en Wu *et al.* 2006, produce una secuencia de ácido nucleico que es completamente diferente de la descrita en esta memoria de patente.

De acuerdo con la invención, al menos dos de las tres secuencias de ácido nucleico contenidas en una célula hospedadora de la invención, y preferiblemente las tres secuencias de ácido nucleico, se han optimizado en los codones para su uso en una célula hospedadora.

La secuencia de ácido nucleico para araB (L-ribuloquinasa) y la secuencia de ácido nucleico para araD (L-ribulosa-5-P-4-epimerasa) se obtienen preferiblemente de *E. coli*. De ese modo, la secuencia de ácido nucleico para araB preferiblemente comprende una secuencia de ácido nucleico con la SEC ID N° 1 y la secuencia de ácido nucleico para araD preferiblemente comprende una secuencia de ácido nucleico con la SEC ID N° 2.

La secuencia de ácido nucleico con la SEC ID N° 1 es la secuencia génica de la fase de lectura abierta (ORF) de *araB<sup>mut</sup>* de *E. coli* en una forma de codones optimizados.

La secuencia de ácido nucleico con la SEC ID N° 2 es la secuencia génica de la fase de lectura abierta (ORF) de *araD* de *E. coli* en una forma de codones optimizados.

La secuencia de ácido nucleico para araA (L-arabinosa isomerasa) se obtiene de *Bacillus licheniformis* o *Clostridium acetobutylicum*.

Estas L-arabinosa isomerasas son ventajosas para el crecimiento de los transformantes de levadura en un medio de arabinosa. El Ejemplo 1 muestra (véase también la figura 4) que, en comparación con la isomerasa de *B. subtilis*, particularmente la expresión de la L-arabinosa isomerasa de *C. acetobutylicum* y *B. licheniformis* mejoraba significativamente el crecimiento de los transformantes de levadura en medio de arabinosa.

De ese modo, la secuencia de ácido nucleico para araA preferiblemente comprende una secuencia de ácido nucleico con la SEC ID N° 3, 4 ó 5.

La secuencia de ácido nucleico con la SEC ID N° 3 es la secuencia génica de la fase de lectura abierta (ORF) de *araA* de *Bacillus licheniformis* en una forma de codones optimizados.

La secuencia de ácido nucleico con la SEC ID N° 4 es la secuencia génica de la fase de lectura abierta (ORF) de *araA* de *Bacillus licheniformis*.

5 La secuencia de ácido nucleico con la SEC ID N° 5 es la secuencia génica de la fase de lectura abierta (ORF) de *araA* de *Clostridium acetobutylicum*.

Por consiguiente, las secuencias de ácido nucleico con las SEC ID N° 4 y 5 son secuencias de ácido nucleico de origen natural.

10 En una realización particularmente preferida, una molécula de ácido nucleico contenida en una célula hospedadora de acuerdo con la invención comprende la secuencia de ácido nucleico con la SEC ID N° 1, la secuencia de ácido nucleico con la SEC ID N° 2 y la secuencia de ácido nucleico con la SEC ID N° 3, 4 ó 5. Más preferiblemente, una molécula de ácido nucleico contenida en una célula hospedadora de acuerdo con la invención comprende la secuencia de ácido nucleico con la SEC ID N° 1, la secuencia de ácido nucleico con la SEC ID N° 2, y la secuencia de ácido nucleico con la SEC ID N° 3.

15 Los transformantes de levadura que tienen los dos genes con codones optimizados de la quinasa (*araB*, SEC ID N° 1) y la epimerasa (*araD*, SEC ID N° 2), y los transformantes de levadura en que los tres genes tienen codones optimizados (*araB*: SEC ID N° 1, *araD*: SEC ID N° 2 y *araA*: SEC ID N° 3), muestran una ventaja considerable de crecimiento en un medio que contiene arabinosa en comparación con transformantes de levadura que tienen solamente un gen con codones optimizados. Las cepas muestran una fase de demora considerablemente más corta y crecen hasta su densidad óptica máxima de forma considerablemente más rápida (véase el ejemplo 2). La combinación de tres genes con codones optimizados posibilita que las células recombinantes de *S. cerevisiae* conviertan L-arabinosa de forma considerablemente más eficaz.

20 En una realización particularmente preferida, la célula hospedadora de la invención contiene la molécula de ácido nucleico integrada de un modo genómicamente estable.

30 En una realización, la molécula de ácido nucleico contenida en una célula hospedadora de acuerdo con la invención está comprendida en un casete de expresión o en un vector de expresión.

Además, los casetes de expresión preferiblemente comprenden secuencias promotoras y terminadoras.

35 Las secuencias promotoras pueden seleccionarse entre HXT7, HXT7 truncado, PFK1, FBA1, PGK1, ADH1 y TDH3.

Las secuencias terminadoras pueden seleccionarse entre CYC1, FBA1, PGK1, PFK1, ADH1 y TDH3.

40 De ese modo, es preferible que diferentes pares de secuencias promotoras y terminadoras controlen cada una de las tres secuencias de ácido nucleico. Esto es necesario para evitar la posible recombinación homóloga entre las regiones/secuencias promotoras y/o terminadoras.

45 Los pares de secuencias promotoras y terminadoras pueden seleccionarse entre un promotor HXT7 o HXT7 truncado y un terminador CYC1, un promotor PFK1 y un terminador FBA1, y un promotor FBA1 y un terminador PGK1.

Una secuencia de ácido nucleico para *araA* puede controlarse mediante el promotor HXT7 o HXT7 truncado y el terminador CYC1.

50 Una secuencia de ácido nucleico para *araB* puede controlarse mediante el promotor PFK1 y el terminador FBA1.

Una secuencia de ácido nucleico para *araD* puede controlarse mediante el promotor FBA1 y el terminador PGK1.

Para detalles adicionales, véase también el ejemplo 3.

55 Los casetes de expresión pueden comprender adicionalmente secuencias de reconocimiento 5' y/o 3', tales como secuencias de reconocimiento de las enzimas *PacI* y *AsclI*.

Los vectores de expresión pueden comprender adicionalmente un marcador de selección.

60 El marcador de selección puede seleccionarse entre un marcador de leucina, un marcador de uracilo o un marcador de antibiótico dominante. Un marcador de antibiótico dominante puede seleccionarse entre geneticina, higromicina y nourseotricina.

Ejemplos de vectores de expresión son p425H7synthAra, pRS303X, p3RS305X o p3RS306X.

65 Para detalles adicionales, véase también el ejemplo 3.

Para aplicaciones industriales, sería ideal que el microorganismo usado fuera capaz de metabolizar todos los azúcares presentes en el medio. Como las levaduras que se usan actualmente no son capaces de metabolizar la arabinosa en el medio, sería muy ventajoso proporcionar a las cepas esta capacidad adicional de forma estable. Para conseguir esto, se muy beneficioso un vector de expresión con genes de una vía metabólica de arabinosa. Este vector de expresión después puede integrarse de forma genómica de una forma estable y puede permitir la metabolización de arabinosa en cepas industriales.

Esta invención consiguió (véanse también los Ejemplos) construir un vector que codifica un casete de expresión con tres genes de una vía metabólica de arabinosa, particularmente una vía metabólica bacteriana. De este modo, es posible evitar los problemas que pueden surgir cuando están presentes varios plásmidos en la misma célula al mismo tiempo ("Plasmid stress", Review of *E. coli* por Bailey (1993)). Además, se posibilita la integración genómica estable de los genes de la vía metabólica de arabinosa. Los problemas asociados con la construcción de un casete de expresión de los genes de la vía metabólica de arabinosa y su integración de un modo que es genómicamente estable ya se han mostrado por Becker (2003) y Wiedemann (2005).

Mediante la selección de los promotores y terminadores en combinación con el uso de la L-arabinosa isomerasa mejorada y las versiones con codones optimizados de los genes implicados, los inventores consiguieron construir células hospedadoras que contenían este casete de expresión funcional.

El casete de expresión construido con los tres genes y contenido en una célula hospedadora de acuerdo con la invención representa un excelente punto de partida para una integración genómica directa y también posibilita la subclonación en los plásmidos de integración de la serie pRS303X, pRS305X y pRS306X (Taxis y Knop, 2006).

Además, tenía que superarse una pluralidad de obstáculos y dificultades experimentales en el proceso para clonar los tres genes con los diferentes promotores y terminadores, y estos se presentan en gran detalle en los ejemplos y figuras.

- Hallazgo de una L-arabinosa isomerasa que funciona mejor, tal como que es más eficaz, en levadura.
- Clonación de la isomerasa que ha demostrado ser difícil y consumir mucho tiempo.
- El vector contenido en una célula hospedadora de acuerdo con la invención es el primer vector descrito que contiene todos los genes esenciales para convertir arabinosa en levadura.
- El vector contiene todos los genes en forma funcional y posibilita a la levadura recombinante un buen crecimiento en arabinosa. La funcionalidad así como un muy buen crecimiento en arabinosa no se esperaban en absoluto.

Para aplicaciones industriales, sería ideal que el microorganismo usado fuera capaz de metabolizar todos los azúcares presentes en el medio. Como las levaduras que se usan actualmente no son capaces de metabolizar la arabinosa en el medio, sería muy ventajoso proporcionar a las cepas esta capacidad adicional de forma estable. Para conseguir esto, de acuerdo con la invención, puede integrarse de forma genómica una molécula de ácido nucleico definida en este documento, un casete de expresión definido en este documento o un vector de expresión definido en este documento de forma estable y puede permitir la metabolización de arabinosa en cepas industriales. Usando las moléculas de ácido nucleico usadas de acuerdo con la invención se asegura una conversión muy eficaz de arabinosa en cepas industriales. Anteriormente, la práctica de introducir los genes de la vía metabólica bacteriana de L-arabinosa individualmente estaba asociada con la dificultad de que los genes no estaban presentes en una proporción óptima entre sí. Las transformaciones consumían mucho tiempo y el metabolismo de arabinosa resultante a menudo no era tan eficaz como se deseaba. Además, las propiedades proporcionadas a menudo no eran estables. Por el contrario, el casete de expresión usado de acuerdo con la invención o el vector de expresión usado de acuerdo con la invención, respectivamente, posibilita que la vía metabólica bacteriana de L-arabinosa se introduzca rápidamente y de forma funcional. Con la selección de los promotores, fue posible combinar los genes juntos en una molécula de ácido nucleico, un casete de expresión o un vector de expresión. La integración de la molécula de ácido nucleico definida en este documento, el casete de expresión definido en este documento o el vector de expresión definido en este documento, respectivamente, garantiza adicionalmente una conversión eficaz de arabinosa.

Una célula hospedadora de acuerdo con la invención es una célula fúngica, y preferiblemente una célula de levadura, tal como especies de *Saccharomyces*, *Kluyveromyces* sp., *Hansenula* sp., *Pichia* sp. o *Yarrowia* sp.

Una célula hospedadora de acuerdo con la invención puede seleccionarse entre BWY1, CEN.PK113-7D, Red Star Ethanol Red y Fermiol.

El objetivo se consigue adicionalmente de acuerdo con la invención proporcionando métodos para producir bioetanol. Un método de acuerdo con la invención comprende la expresión de una molécula de ácido nucleico definida en este documento, un casete de expresión definido en este documento, o un vector de expresión definido en este documento en una célula hospedadora.

De ese modo, el método se realiza en una célula hospedadora de acuerdo con la invención.

El objetivo se consigue adicionalmente de acuerdo con la invención mediante el uso de una célula hospedadora de acuerdo con la invención para producir bioetanol.

5 El objetivo se consigue adicionalmente de acuerdo con la invención mediante el uso de una célula hospedadora de acuerdo con la invención para la fermentación recombinante de biomaterial que contiene pentosa.

10 Para los métodos y usos, véanse los ejemplos y figuras. Los resultados de la fermentación registrados en el ejemplo 2 muestran que especialmente los genes con codones optimizados de *araA*, *araB* y *araD* posibilitan que los transformantes de levadura metabolicen la arabinosa de forma más eficaz. El resultado de esto es la conversión más rápida del azúcar y un rendimiento significativamente mayor de etanol.

Además se describe un polipéptido que se selecciona entre el grupo de

- 15 a. un polipéptido que es al menos un 70 %, tal como al menos un 80 % idéntico a la secuencia de aminoácidos que está codificada por la SEC ID N° 3, 4 ó 5, y tiene una función pentosa isomerasa *in vitro* y/o *in vivo*,  
 b. una variante de origen natural de un polipéptido que incluye la secuencia de aminoácidos que está codificada por la SEC ID N° 3, 4 ó 5, que tiene una función pentosa isomerasa *in vitro* y/o *in vivo*,  
 c. un polipéptido que es idéntico a la secuencia de aminoácidos que está codificada por la SEC ID N° 3, 4 ó 5, y tiene una función pentosa isomerasa *in vitro* y/o *in vivo*, y  
 20 d. un fragmento del polipéptido de a., b. o c., que comprende un fragmento de al menos 100, 200 ó 300 aminoácidos continuos de la secuencia de aminoácidos que está codificada por la SEC ID N° 3, 4 ó 5,

Dicho polipéptido puede seleccionarse entre el grupo de

- 25 a. un polipéptido que es al menos un 70 %, tal como al menos un 80 % idéntico a la secuencia de aminoácidos de acuerdo con SEC ID N° 6 ó 7, y tiene una función pentosa isomerasa *in vitro* y/o *in vivo*,  
 b. una variante de origen natural de un polipéptido que comprende la secuencia de aminoácidos de acuerdo con la SEC ID N° 6 ó 7, que tiene una función pentosa isomerasa *in vitro* y/o *in vivo*,  
 30 c. un polipéptido que es idéntico a la secuencia de aminoácidos de acuerdo con la SEC ID N° 6 ó 7, y tiene una función pentosa isomerasa *in vitro* y/o *in vivo*, y  
 d. un fragmento del polipéptido de a., b. o c., que comprende un fragmento de al menos 100, 200 ó 300 aminoácidos continuos de acuerdo con la SEC ID N° 6 ó 7.

35 Dicho polipéptido puede comprender un polipéptido que es al menos un 90 %, tal como un 95 % idéntico a la secuencia de aminoácidos que está codificada por la SEC ID N° 3, 4 ó 5, y tiene una función pentosa isomerasa *in vitro* y/o *in vivo*.

40 Dicho polipéptido puede comprender un polipéptido que es al menos un 90 %, tal como un 95 % idéntico a la secuencia de aminoácidos de acuerdo con la SEC ID N° 6 ó 7, y tiene una función pentosa isomerasa *in vitro* y/o *in vivo*.

45 La secuencia de aminoácidos con la SEC ID N° 6 es la secuencia de aminoácidos de L-arabinosa isomerasa (*araA*) de *Bacillus licheniformis*. Esta secuencia de aminoácidos está codificada por las secuencias de ácido nucleico con las SEC ID N° 3 ó 4.

La secuencia de aminoácidos con la SEC ID N° 7 es la secuencia de aminoácidos de L-arabinosa isomerasa (*araA*) de *Clostridium acetobutylicum*. Esta secuencia de aminoácidos está codificada por la secuencia de ácido nucleico con la SEC ID N° 5.

50 La pentosa es arabinosa, en particular L-arabinosa.

El polipéptido es originario de una bacteria, en particular de *Bacillus licheniformis* o *Clostridium acetobutylicum*.

55 Estas L-arabinosa isomerasas son ventajosas para el crecimiento de transformantes de levadura en medio de arabinosa. Varios experimentos diferentes indicaron que la L-arabinosa isomerasa de *B. subtilis* que se había usado previamente representa un etapa limitante en la descomposición de arabinosa en levadura (Becker y Boles, 2003; Wiedemann, 2003; Karhumaa et al., 2006; Sedlak y Ho, 2001). El Ejemplo 1 muestra (véase también la figura 4) que el crecimiento de transformantes de levadura en medio de arabinosa está significativamente mejorado particularmente por la expresión de L-arabinosa isomerasa de *C. acetobutylicum* y de *B. licheniformis*, en comparación con la isomerasa de *B. subtilis*.  
 60

Además, se describe una molécula de ácido nucleico aislada que codifica un polipéptido.

Además, se describe una célula hospedadora que contiene dicha molécula de ácido nucleico aislada.

65

El polipéptido, la molécula de ácido nucleico aislada y la célula hospedadora son adecuados en la producción de bioetanol y para la fermentación recombinante de biomaterial que contiene pentosa.

5 Además, se describen células hospedadoras que contienen una o más modificaciones, tales como moléculas de ácido nucleico.

Una modificación adicional de dicho tipo es una célula hospedadora que sobre-expresa un gen TAL1 (transaldolasa), tal como se describe por los inventores en el documento EP 1 499 708 B1, por ejemplo.

10 Una modificación adicional más de éstas es una célula hospedadora que contiene un ácido nucleico que codifica un gen transportador de L-arabinosa específico (araT), particularmente tal como un gen transportador de L-arabinosa específico del genoma de *P. stipitis*, tal como se describe por lo inventores en la solicitud de patente alemana DE 10 1006 060 381.8, presentada el 20 de diciembre de 2006.

15 Biomasa adicional con cantidades significativas de arabinosa (fuente de los datos: U.S. Department of Energy <http://www.eere.energy.gov/biomass/progs/search1.cgi>):

	Tipo de biomasa	L-arabinosa [%]
20	Pasto varilla	3,66
	Bothriochloa grande	3,55
	Festuca alta	3,19
	Robinia	3
	Rastrojo de maíz	2,69
	Paja de trigo	2,35
25	Bagazo de caña de azúcar	2,06
	Lespedeza chino	1,75
	Sorghum bicolor	1,65

30 Los ácidos nucleicos, casetes de expresión, vectores de expresión y células hospedadoras de acuerdo con la invención también son de gran importancia para su utilización.

Los posibles usos de los ácidos nucleicos, casetes de expresión, vectores de expresión y células hospedadoras de acuerdo con la invención incluyen tanto la producción de bioetanol como la fabricación de productos precursores de alta calidad para la posterior síntesis química.

35 La siguiente lista es originaria del estudio "Top Value Added Chemicals From Biomass" (véase [www1.eere.energy.gov/biomass/pdfs/35523.pdf](http://www1.eere.energy.gov/biomass/pdfs/35523.pdf)). Aquí, se clasificaron 30 agentes químicos como particularmente valiosos, que pueden producirse a partir de biomasa.

Cantidad de átomos de C	Principales 30 candidatos
1	hidrógeno, monóxido de carbono
2	
3	glicerol, ácido 3-hidroxipropiónico, ácido láctico, ácido malónico, ácido propiónico, serina
4	acetoina, ácido asparágico, ácido fumárico, 3-hidroxitiro lactona, ácido málico, ácido succínico, treonina
5	arabitol, furfural, ácido glutámico, ácido itacónico, ácido levulínico, prolina, xilitol, ácido xilónico
6	ácido aconítico, citrato, ácido 2,5-furanodicarboxílico, ácido glucárico, lisina, levoglucosano, sorbitol

40 Es importante tener los ácidos nucleicos, casetes de expresión, vectores de expresión y células hospedadoras de acuerdo con la invención disponibles tan pronto como estos agentes químicos se produzcan a partir de lignocelulosa por bioconversión (por ejemplo, fermentaciones con levaduras).

45 La presente invención se explicará en mayor detalle en las siguientes figuras, secuencias y ejemplos, sin limitación a los mismos. Las referencias citadas se incorporan completamente en este documento por referencia a las mismas. En las secuencias y figuras se muestran:

50 La SEC ID N° 1 muestra la secuencia génica de la fase de lectura abierta (ORF) de *araB<sup>mut</sup>* de *E. coli* en una forma de codones optimizados.

La SEC ID N° 2 muestra la secuencia génica de la fase de lectura abierta (ORF) de *araD* de *E. coli* en una forma de codones optimizados.

La SEC ID N° 3 muestra la secuencia génica de la fase de lectura abierta (ORF) de *araA* de *B. licheniformis* en una forma de codones optimizados.

La SEC ID N° 4 muestra la secuencia génica de la fase de lectura abierta (ORF) de *araA* de *B. licheniformis*.

La SEC ID N° 5 muestra la secuencia génica de la fase de lectura abierta (ORF) de *araA* de *C. acetobutylicum*.

La SEC ID N° 6 muestra la secuencia de aminoácidos de la L-arabinosa isomerasa (*araA*) de *Bacillus licheniformis*. Esta secuencia de aminoácidos está preferiblemente codificada por las secuencias de ácido nucleico con la SEC ID N° 3 ó 4.

La SEC ID N° 7 muestra la secuencia de aminoácidos de la L-arabinosa isomerasa (*araA*) de *Clostridium acetobutylicum*. Esta secuencia de aminoácidos está preferiblemente codificada por la secuencia de ácido nucleico con la SEC ID N° 5.

**Figura 1** *Composición de la biomasa.*

La biomasa consta de celulosa, hemicelulosa y lignina. La hemicelulosa que es la segunda que existe más frecuentemente es un polímero muy ramificado que consta de pentosas, ácidos urónicos y hexosas. La hemicelulosa consta de una gran proporción de las pentosas xilosa y arabinosa.

**Figura 2** *Esquema del metabolismo de L-arabinosa en S. cerevisiae recombinante mediante la integración de una vía metabólica bacteriana de L-arabinosa.*

**Figura 3** *Vectores usados y su construcción.*

El plásmido inicial para la construcción del vector p425H7synthAra (**Figura 3 A**) era el plásmido p425HXT7-6HIS (**Figura 3 B**). Las fases de lectura abiertas de los genes de codones optimizados para *araA* de *B. licheniformis* y *araB<sup>mut</sup>* y *araD* de *E. coli* se amplificaron y clonaron en el plásmido p425HXT7-6HIS tras diversos promotores y terminadores. Los cebadores se seleccionaron de tal modo que el casete de expresión resultante estuviera flanqueado por los sitios de restricción de las enzimas *PacI* y *Ascl*. De ese modo, se produjo el plásmido p425H7synthAra, que tiene un marcador de leucina.

**Figura 4** *Crecimiento en arabinosa usando diversos genes de L-arabinosa isomerasa.*

Curvas de crecimiento de cepas recombinantes de *S. cerevisiae* que contienen el metabolismo bacteriano de L-arabinosa con diversas L-arabinosa isomerasas. Se realizaron ensayos de crecimiento en 5 ml de medio SM con arabinosa al 2 % en condiciones aeróbicas. Se ensayaron las L-arabinosa isomerasas de *C. acetobutylicum*, *B. licheniformis*, *P. pentosaceus*, *L. plantarum* y *L. mesenteroides*. Se usaron la L-arabinosa isomerasa de *B. subtilis* y el vector p423HXT7-6HIS vacío como controles.

**Figura 5** *Crecimiento en arabinosa usando genes con codones optimizados de la vía metabólica de arabinosa.*

Curvas de crecimiento de cepas recombinantes de *S. cerevisiae* que contienen el metabolismo bacteriano de L-arabinosa con diferentes combinaciones de genes con codones optimizados y los genes con secuencias originales. Se realizaron ensayos de crecimiento en 5 ml de medio SM con arabinosa al 2 % en condiciones aeróbicas. Se ensayó cada una de las combinaciones que contenía uno de los genes con codones optimizados respectivamente, y la combinación que contenía los tres genes con codones optimizados. Además, también se ensayó la combinación en que estaban presentes los genes con codones optimizados de quinasa y epimerasa. Se usó una cepa de levadura recombinante con los cuatro genes que tenían las secuencias originales como control.

**Figura 6** *Formación de etanol usando genes con codones optimizados de la vía metabólica de arabinosa.*

La figura muestra los resultados de análisis de HPLC de los sobrenadantes del medio de dos fermentaciones. Una fermentación se realizó con la cepa BWY1, que posee los plásmidos p423H7synthIso, p424H7synthKin, p425H7synthEpi y pHL125<sup>re</sup> (3xsynth). En la otra fermentación, se ensayó la cepa BWY1, que contenía los plásmidos p423H7araABs<sup>re</sup>, p424H7araB<sup>re</sup>, p425H7araD<sup>re</sup> y pHL125<sup>re</sup> (3xre). Las fermentaciones se realizaron en medio SFM con L-arabinosa al 3 %. Las cepas se cultivaron hasta una elevada densidad óptica en el fermentador. Después, la fermentación se cargó en condiciones anaeróbicas (después de 48 horas). Los diagramas muestran el consumo de arabinosa y la producción de etanol.

**Figura 7** *Crecimiento en arabinosa usando el plásmido de expresión construido p425H7-synthAra.*

Curvas de crecimiento de cepas recombinantes de *S. cerevisiae* que contienen metabolismo bacteriano de L-arabinosa en forma del vector p425H7-synthAra. Se realizaron ensayos de crecimiento en 5 ml de medio SC con arabinosa al 2 % en condiciones aeróbicas. Se usó una cepa recombinante de levadura con los plásmidos p423H7araABs<sup>re</sup>, p424H7araB<sup>re</sup>, p425H7araD<sup>re</sup> y pHL125<sup>re</sup>, que se habían ensayado en 5 ml de medio SM con arabinosa al 2 %, como control.

**Ejemplo****Métodos**5 **1. Cepas y medios****-Bacterias**

- 10 - *E. coli* SURE (Stratagene)  
 - *E. coli* DH5 $\alpha$  (Stratagene)  
 - *Bacillus licheniformis* (DSMZ)  
 - *Clostridium acetobutylicum* (DSMZ)  
 - *Leuconostoc mesenteroides* (DSMZ)  
 - *Pediococcus pentosaceus* (DSMZ)  
 15 - *Lactobacillus plantarum* (DSMZ)

Medio completo LB con triptona al 1 %, extracto de levadura al 0,5 %, NaCl al 0,5 %, pH 7,5 (véase Maniatis, 1982). Se añadieron 40  $\mu$ g/ml de ampicilina al medio después de tratar en autoclave para selección basada en resistencia a antibióticos codificada en el plásmidos. El medio de cultivo sólido también contenía agar al 2 %. El cultivo se realizó a 37 °C.

20

**-Levadura**Cepa BWY1:

25 BWY1 se basa en la cepa JBY25 (*MATa leu2-3,112 ura3-52 trp1-289 his3- $\Delta$ 1MAL2-8c SUC2+* mutaciones desconocidas para un mejor crecimiento en arabinosa); la cepa JBY25 se seleccionó adicionalmente y posee mutaciones adicionales para un crecimiento mejorado en L-arabinosa en condiciones reducidas de oxígeno (Wiedemann, 2005)

30 - medio selectivo completo sintético SC:

base de nitrógeno de levadura al 0,67 % sin aminoácidos, pH 6,3, solución de aminoácidos/nucleobases, fuente de carbono a la concentración indicada en cada caso

35 - medio selectivo mínimo sintético SM:

base de nitrógeno de levadura al 0,16 % sin aminoácidos y sulfato de amonio, sulfato de amonio al 0,5 %, dihidrogenofosfato de potasio 20mM, pH 6,3, fuente de carbono a la concentración indicada en cada caso

40 - medio de fermentación sintético (medio mineral) SFM:

(Verduyn *et al.*, 1992), pH 5,5

Sales: (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 5 g/l; KH<sub>2</sub>PO<sub>4</sub>, 3 g/l; MgSO<sub>4</sub>\*7H<sub>2</sub>O, 0,5 g/l

Elementos traza: EDTA, 15 mg/l; ZnSO<sub>4</sub>\*4,5 mg/l; MnCl<sub>2</sub>\*4H<sub>2</sub>O, 0,1 mg/l;

45 CoCl<sub>2</sub>\*6H<sub>2</sub>O, 0,3 mg/l; CuSO<sub>4</sub>, 0,192 mg/l; Na<sub>2</sub>MoO<sub>4</sub>\*2H<sub>2</sub>O, 0,4 mg/l;

CaCl<sub>2</sub>\*2H<sub>2</sub>O, 4,5 mg/l; FeSO<sub>4</sub>\*7H<sub>2</sub>O, 3 mg/l; H<sub>3</sub>BO<sub>3</sub>, 1 mg/l; KI, 0,1 mg/l

Vitaminas: Biotina, 0,05 mg/l; ácido p-aminobenzoico, 0,2 mg/l; ácido nicotínico, 1 mg/l;

Pantotenato cálcico, 1 mg/l; piridoxina-HCl, 1 mg/l; tiamina-HCl, 1 mg/l;

M inositol, 25 mg/l

50

Concentración de aminoácidos y nucleobases en el medio completo sintético (basadas en Zimmermann, 1975): Adenina (0,08 mM), arginina (0,22 mM), histidina (0,25 mM), isoleucina (0,44 mM), leucina (0,44 mM), lisina (0,35 mM), metionina (0,26 mM), fenilalanina (0,29 mM), triptófano (0,19 mM), treonina (0,48 mM), tirosina (0,34 mM), uracilo (0,44 mM), valina (0,49 mM). Se usaron L-arabinosa y D-glucosa como fuente de carbono.

55 Los medios completo sólido y selectivo también contenían agar al 1,8 %. Las células de levadura se cultivaron a 30 °C. El medio mineral sintético usado para las fermentaciones contenía sales, metales traza y vitaminas en las concentraciones enumeradas anteriormente y L-arabinosa como fuente de carbono. Se prepararon soluciones madres de los metales traza y de las vitaminas. Ambas soluciones se filtraron a esterilidad. Ambas se almacenaron a 4 °C. El valor de pH es sumamente importante para la preparación de la solución de metales traza. Los diversos

60 elementos traza tuvieron que disolverse completamente en agua uno tras otro en el orden dado anteriormente. Después de cada adición, el valor de pH tuvo que ajustarse a 6,0 con KOH antes de poder añadir el siguiente elemento traza. Finalmente, el valor de pH se ajustó a 4,0 con HCl. Se añadieron 200  $\mu$ l agente antiespumante (Antifoam2004, Sigma) al medio para evitar la formación de espuma. Como los ensayos se realizaron en condiciones anaeróbicas, tuvieron que añadirse 2,5 ml/l de una solución de Tween80-Ergosterol al medio después

65 de tratamiento en autoclave. Ésta consta de 16,8 g de Tween80 y 0,4 g de Ergosterol, que se cargaron hasta 50 ml con etanol y se disolvieron en el mismo. La solución se filtró a esterilidad. Las sales y el agente antiespumante se

trataron en autoclave junto con el fermentador completo. La arabinosa se trató en autoclave por separado del resto del medio. Después de enfriar el medio, se añadieron los elementos traza y las vitaminas al mismo.

## 2. Plásmidos

5

Plásmido	Fuente/Referencia	Descripción
p423HXT7-6HIS (=p423H7)	Becker y Boles, 2003	plásmido de expresión 2μ para la sobre-expresión de diversos genes y para fusionar la L-arabinosa isomerasa de <i>E. coli</i> con una epítipo His <sub>6</sub> ; gen marcador de selección <i>HIS3</i> , promotor <i>HXT7</i> acortado y terminador <i>CYC1</i> (Hamacher <i>et al.</i> , 2002)
p424HXT7-6HIS (=p424H7)	Becker y Boles, 2003	plásmidos de expresión 2μ para la sobre-expresión de diversos genes y para fusionar la L-ribuloquinasa mutada y de tipo silvestre de <i>E. coli</i> con un epítipo His <sub>6</sub> ; gen marcador de selección <i>TRP1</i> , promotor <i>HXT7</i> acortado y terminador <i>CYC1</i> (Hamacher <i>et al.</i> , 2002)
p425HXT7-6HIS (=p425H7)	Becker y Boles, 2003	plásmido de expresión 2μ para la sobre-expresión de diversos genes; gen marcador de selección <i>LEU2</i> , promotor <i>HXT7</i> acortado y terminador <i>CYC1</i> (Hamacher <i>et al.</i> , 2002)
p426HXT7-6HIS (=p426H7)	Hamacher <i>et al.</i> , 2002	plásmidos de expresión 2μ para la sobre-expresión de genes que producen un epítipo His <sub>6</sub> ; gen marcador de selección <i>URA3</i> , promotor <i>HXT7</i> acortado y terminador <i>CYC1</i>
p423H7araABs <sup>te</sup>	Becker y Boles, 2003	<i>araA</i> de <i>B. subtilis</i> en p423HXT7-His, re-aislado de JBY25-4M
p424H7araB	Becker y Boles, 2003	<i>araB</i> de <i>E. coli</i> en p423HXT7-His
p424H7araB <sup>te</sup>	Becker y Boles, 2003	<i>araB</i> de <i>E. coli</i> en p423HXT7-His; re-aislado de JBY25-4M, mutación en <i>araB</i> , que posibilita el crecimiento en arabinosa
p425H7araD <sup>te</sup>	Becker y Boles, 2003	<i>araD</i> de <i>E. coli</i> en p425HXT7-His; re-aislado de JBY25-4M
p423H7-synthIso		<i>araA</i> de <i>B. licheniformis</i> con codones optimizados en p423HXT7-His
p424H7-synthKin		<i>araB</i> de <i>E. coli</i> con codones optimizados en p424HXT7-His, con mutación en <i>araB</i>
p425H7-synthEpi		<i>araD</i> de <i>E. coli</i> con codones optimizados en p425HXT7-His
p425H7-synthAra		plásmido 2μ con los genes <i>araA</i> , <i>araB<sup>mut</sup></i> y <i>araD</i> con codones optimizados; <i>araA</i> bajo el control del promotor <i>FBA1</i> y el terminador <i>PGK1</i> , <i>araB<sup>mut</sup></i> bajo el control del promotor <i>PFK1</i> y el terminador <i>FBA1</i> , y <i>araD</i> bajo el control del promotor <i>HXT7</i> acortado y el terminador <i>CYC1</i> , gen marcador de selección <i>LEU2</i>
pHL125 <sup>te</sup>	Liang y Gaber, 1996	plásmido 2μ con el gen <i>GAL2</i> expresado tras el promotor <i>ADH1</i> , gen marcador de selección <i>URA3</i> ; re-aislado de JBY25-4M

## 3. Transformación:

### - Transformación de *E. coli*

10 Las células de *E. coli* se transformaron mediante el método de electroporación descrito por Dower *et al.* (1988) y Wirth (1993) usando un dispositivo Easyject prima (EQUIBO).

### - Transformación de *S. cerevisiae*

15 Las cepas de *S. cerevisiae* se transformaron con ADN plasmídico o fragmentos de ADN usando el método de acetato de litio de Gietz y Woods (1994).

## 4. Preparación de ADN

20

### - Aislamiento de ADN plasmídico de *E. coli*

25 El ADN plasmídico se aisló de *E. coli* con el procedimiento de lisis alcalina desarrollado por Birnboim y Doly (1979), modificado de acuerdo con Maniatis *et al.* (1982), o como alternativa con el "QIAprep Spin Miniprep Kit" fabricado por Qiagen.

Se preparó ADN plasmídico muy puro para secuenciación con el "Plasmid Mini Kit" fabricado por Qiagen de acuerdo con las instrucciones del fabricante.

#### 5 - Aislamiento de ADN plasmídico de *S. cerevisiae*

Se recogieron las células de un cultivo estacionario de levadura (5 ml) por centrifugación, se lavaron y se resuspendieron en 400 µl de tampón P1 (Plasmid Mini Kit, Qiagen). Después de la adición de 400 µl de tampón P2 y  $\frac{2}{3}$  volumen de perlas de vidrio ( $\varnothing$  0,45 mm, la alteración celular se realizó agitando durante 5 minutos en un Vibrax (Vibrax-VXR fabricado por Janke & Kunkel o IKA). El residuo se cargó con  $\frac{1}{2}$  volumen de tampón P3, se mezcló y se incubó en hielo durante 10 min. Después de centrifugar durante 10 minutos a 13000 rpm, el ADN plasmídico se precipitó a temperatura ambiente añadiendo 0,75 ml de isopropanol al residuo. El ADN se sedimentó centrifugando a 13000 rpm durante 30 min. y se lavó con etanol al 70 %, se secó y se resuspendió en 20 µl de agua. Se usó 1 µl del ADN para la transformación en *E. coli*.

#### 15 - Determinación de la concentración de ADN

La concentración de ADN se midió por espectrofotometría en un intervalo de longitud de onda de 240-300 nm. Si la pureza del ADN, determinada mediante el cociente  $E_{260nm}/E_{280nm}$ , es 1,8, la extinción  $E_{260nm}=1,0$  corresponde a una concentración de ADN de 50 µg de ADNbc/ml (Maniatis *et al.*, 1982).

#### 20 - Amplificación de ADN usando PCR

Uso del sistema Phusion™ High Fidelity

La reacción en cadena de la polimerasa se realizó en un volumen total de 50 µl con el sistema de PCR "Phusion™ High Fidelity" fabricado por la compañía Finnzymes de acuerdo con las instrucciones del fabricante. Cada solución madre constaba de 1-10 ng de ADN o 1-2 colonias de levadura como modelo de síntesis, mezcla de dNTP 0,2 mM, tampón 2 1x (contiene MgCl<sub>2</sub> 1,5 mM), 1 U de polimerasa, y 100 pmol de cada uno de los correspondientes cebadores oligonucleotídicos. La reacción de PCR se realizó en un termociclador fabricado por la compañía Techne y se seleccionaron las siguientes condiciones de PCR de acuerdo con las necesidades:

1.	1x	30 s, 98 °C	Desnaturalización de ADN
2.	30x	10 s, 98 °C	Desnaturalización de ADN
		30 s, 56-62 °C	Hibridación/unión de oligonucleótidos al ADN
		0,5-1 min., 72 °C	Síntesis/elongación de ADN
3.	1x	7 min., 72 °C	Síntesis/elongación de ADN

La polimerasa se añadió después de la primera etapa de desnaturalización ("PCR de inicio en caliente"). La cantidad de etapas de síntesis, la temperatura de hibridación y el tiempo de elongación se adaptaron a las temperaturas de fusión específicas de los oligonucleótidos usados y el tamaño del producto esperado. Los productos de PCR se ensayaron con electroforesis en gel de agarosa y después se limpiaron.

#### 35 - Purificación de ADN de los productos de PCR

Los productos de PCR se purificaron con el "QIAquick PCR Purification Kit" fabricado por Qiagen de acuerdo con las instrucciones del fabricante.

#### 40 - Separación electroforética en gel de los fragmentos de ADN

Los fragmentos de ADN que tenían un tamaño de 0,15-20 kb se separaron en geles de agarosa al 0,5-1 % con 0,5 µg/ml de bromuro de etidio. Se usó tampón TAE 1x (Tris 40 mM, ácido acético 40 mM, EDTA 2 mM) como tampón de gel y tampón de procesamiento (Maniatis *et al.*, 1982). Se usó un ADN de fago lambda diferido con las endonucleasas de restricción *EcoRI* y *HindIII* como patrón de tamaño. Antes de la carga, se añadió 1/10 volumen de marcador azul (tampón TAE 1x, glicerina al 10 %, azul de bromofenol al 0,004 %) a las muestras de ADN, que se hicieron visibles tras la separación por irradiación con luz UV (254 nm).

#### 50 - Aislamiento de fragmentos de ADN de geles de agarosa

El fragmento de ADN deseado se cortó del gel de TAE agarosa bajo luz UV de onda larga (366 nm) y se aisló con el "QIAquick Gel Extraction Kit" fabricado por Qiagen de acuerdo con las instrucciones del fabricante.

### 55 5. Modificación enzimática del ADN

#### Restricción del ADN

60 La escisión específica de secuencia del ADN con endonucleasas de restricción se realizó durante 1 hora con enzima 2-5U por µg de ADN en las condiciones de incubación recomendadas por el fabricante.

**Ejemplo 1: Selección de las mejores L-arabinosa isomerasas****A) Rendimiento de la selección**

- 5 Varios experimentos indicaron que la L-arabinosa isomerasa de *B. subtilis* representa una etapa limitante en la descomposición de arabinosa en levadura (Becker y Boles, 2003; Wiedemann, 2003; Karhumaa et al., 2006; Sedlak y Ho, 2001). Para mejorar la vía metabólica de arabinosa, se ensayaron cinco L-arabinosa isomerasas de diferentes organismos.
- 10 Para esto, se aisló el ADN genómico de los organismos *C. acetobutylicum*, *B. licheniformis*, *P. pentosaceus*, *L. plantarum* y *L. mesenteroides* (véase "Isolation of plasmid DNA from *S. cerevisiae*"). Las células se cultivaron, recogieron y absorbieron en el tampón. La alteración celular se realizó usando perlas de vidrio. Después, se precipitó el ADN, se lavó, y se usó para la PCR. La fase de lectura abierta (ORF) de *araA* de los organismos
- 15 enumerados se amplificó con cebadores, que también tenían áreas homólogas al promotor *HXT7* y el terminador *CYC1*. Los productos de PCR obtenidos se transformaron en levadura junto con el vector p423HXT7-6His linealizado con *EcoRI/BamHI* y se clonaron por recombinación *in vivo* en el plásmido entre el promotor *HXT7* y el terminador *CYC1*. La secuencia de los plásmidos obtenidos se verificó por análisis de restricción. La funcionalidad de las nuevas isomerasas y su efecto sobre el metabolismo de arabinosa también tuvo que estudiarse.
- 20 Para este fin, se produjeron cepas recombinantes de levadura, que contenían una de las nuevas isomerasas y el resto de los genes de la vía metabólica bacteriana de arabinosa (p424H7araB<sup>re</sup>, p425H7araD<sup>re</sup> y pHL125<sup>re</sup>).

**B) Comportamiento de crecimiento**

- 25 El crecimiento de las cepas se ensayó en condiciones aeróbicas en un medio que contenía arabinosa. La cepa recombinante de levadura que contenía la isomerasa de *B. subtilis* se usó como control. Se construyó una cepa de levadura con el vector p423HXT7-6HIS vacío como control negativo.
- 30 Las cepas con los diversos plásmidos de isomerasa se cultivaron en medio SM con arabinosa al 2 % y se inocularon con un DO<sub>600nm</sub> = 0,2 en 5 ml de medio SM con arabinosa al 2 %. Esto se incubó en tubos de ensayo en un frasco de agitación en condiciones aeróbicas a 30 °C. Se recogieron muestras regularmente para determinar la densidad óptica.
- 35 Los resultados se muestran en la figura 4. Se demostró que, en comparación con la isomerasa de *B. subtilis*, particularmente la expresión de L-arabinosa isomerasa de *C. acetobutylicum* y de *B. licheniformis* mejoraba significativamente el crecimiento de los transformantes de levadura en medio de arabinosa.

**Ejemplo 2: Optimización de codones de genes para la descomposición de arabinosa en levadura****A) Optimización de codones de genes de acuerdo con el uso de codones de los genes de glucólisis de *S. cerevisiae***

- 40 Se calculó el uso de codones preferidos de los genes de glucólisis de *S. cerevisiae* y se enumera en la tabla 1. Las ORF de los genes *araA* y *araB<sup>mut</sup>* de *E. coli* se optimizaron en los codones así como la ORF del gen *araA* de *B. licheniformis*. Esto implica que las secuencias de las fases de lectura abiertas se adaptaron al uso de codones preferidos enumerados a continuación. La secuencia proteica de las enzimas permaneció inalterada. Los genes se sintetizaron en las instalaciones de una compañía independiente suministros en forma seca en vectores propios de la compañía.

- 50 Hay más información detallada acerca de la síntesis de genes disponible en <http://www.sloning.com/>.

Tabla 1: *Uso de codones preferidos de genes de glucólisis de S. cerevisiae.*

	Aminoácido	codón preferido
55	Ala	GCT
	Arg	AGA
	Asn	AAC
	Asp	GAC, (GAT)
	Cys	TGT
60	Gln	CAA
	Glu	GAA
	Gly	GGT
	His	CAC
	Ile	ATT, (ATC)
65	Leu	TTG
	Lys	AAG

	Met	ATG
	Phe	TTC
	Pro	CCA
	Ser	TCT, (TCC)
5	Thr	ACC, (ACT)
	Trp	TGG
	Tyr	TAC
	Val	GTT, (GTC)
10	Parada	TAA

## B) Introducción de genes con codones optimizados en la cepa BWY1

Para introducir por transformación los tres genes con codones optimizados en la cepa BWY1 y ensayarlos, los genes tuvieron que subclonarse en vectores de levadura. Para este fin, se amplificaron la ORF de *araA*, la ORF de *araB<sup>mut</sup>* y la ORF de *araD* con codones optimizados con cebadores, de modo que se crearon proyecciones homólogas al promotor *HXT7* acortado y el terminador *CYC1*. Los plásmidos de expresión 2 $\mu$  p423HXT7-6HIS, p424HXT7-6HIS, p425HXT7-6HIS se linealizaron con endonucleasas de restricción en el intervalo entre el promotor *HXT7* y el terminador *CYC1*. El producto de PCR de *araA* se introdujo por transformación en levadura con el p423HXT7-6HIS linealizado y se clonó en el plásmido p423H7-synthIso por recombinación *in vivo*. Se siguió el mismo procedimiento con el producto de PCR *araB<sup>mut</sup>* y el vector p424HXT7-6HIS linealizado. Esto produjo el plásmido p424H7synthKin. El plásmido p425H7-synthIso se produjo por recombinación *in vivo* del producto de PCR *araD* y el vector p425HXT7-6HIS linealizado en levadura. Los plásmidos se aislaron de la levadura y se amplificaron en *E. coli*. Después de aislar los plásmidos de *E. coli*, se examinaron los plásmidos por análisis de restricción. Uno de cada uno de los plásmidos con los genes con codones optimizados se introdujo por transformación en la cepa de levadura BWY1 junto con los tres plásmidos re-aislados originales, para ensayar la funcionalidad y para análisis adicional, de modo que todas las cepas recombinantes producidas contenían una vía metabólica completa de arabinosa. Además, se ensayó la combinación p424H7synthKin y p42457synthEpi con los plásmidos reaislados originales así como un lote en que el transformante de levadura poseía los tres nuevos plásmidos. La transformación con los cuatro plásmidos en cada caso tuvo lugar al mismo tiempo. Los transformantes se sembraron en medio SM con glucosa al 2 %. Después de dos días, las colonias obtenidas se sembraron en estrías en medio SM con arabinosa al 2 %. Se usó una cepa de levadura que contenía los cuatro plásmidos reaislados originales como control positivo.

## C) Comportamiento de crecimiento

Se examinó el crecimiento de la cepa BWY1 con las diversas combinaciones de plásmido de genes con codones optimizados y genes originales en ensayos de crecimiento en medio que contenía arabinosa en condiciones aeróbicas.

Las cepas con las diversas combinaciones de plásmido se cultivaron en medio SM con L-arabinosa al 2 % y se inocularon con una  $DO_{600nm} = 0,2$  en 5 ml de medio SM con L-arabinosa al 2 %. La incubación tuvo lugar en tubos de ensayo en condiciones aeróbicas a 30 °C. Se recogieron muestras regularmente para determinar la densidad óptica.

Los resultados de la curva de crecimiento aeróbico se muestran en la figura 5. Puede observarse claramente que las cepas recombinantes de levadura que poseen solamente uno de los genes optimizados muestran poca o ninguna ventaja de crecimiento en comparación con la cepa con los cuatro plásmidos originales en un medio que contiene arabinosa. Sin embargo, los transformantes de levadura con los dos genes optimizados de quinasa y epimerasa y los transformantes de levadura con tres genes optimizados mostraron una clara ventaja de crecimiento en un medio que contiene arabinosa. Las cepas manifestaron una fase de demora significativamente más corta y crecieron hasta su densidad óptica máxima de forma considerablemente más rápida.

Esto demuestra que la combinación de los tres genes con codones optimizados posibilita que las células recombinantes de *S. cerevisiae* conviertan L-arabinosa de forma significativamente más eficaz.

## D) Producción de etanol

La Figura 6 (A) y (B) muestra los resultados de análisis de HPLC de dos fermentaciones. Una cepa recombinante de levadura contiene los plásmidos p423H7synthIso, p424H7synthKin, p425H7synthEpi y pHL125<sup>te</sup>, la otra contiene los plásmidos p423H7araABs<sup>te</sup>, p424H7araB<sup>te</sup>, p425H7araD<sup>te</sup> y pHL125<sup>te</sup>. Las fermentaciones se realizaron en medio SFM con L-arabinosa al 3 %. La Figura 6 (A) muestra el consumo de arabinosa y el peso seco de ambas cepas. La Figura 6 (B) ilustra la producción de etanol de las dos cepas.

Las cepas se cultivaron en el fermentador de forma aeróbica hasta que alcanzaron un peso seco de aproximadamente 2,8 g/l. Cuando hubo presente suficiente masa celular, las fermentaciones se cambiaron a condiciones anaeróbicas. La figura muestra los diagramas de metabolismo de arabinosa y producción de etanol. Los subproductos producidos, arabitol, acetato y glicerina, no se han enumerado porque se produjeron en cantidades comparables por ambas cepas.

Como muestran los diagramas, la producción de etanol comienza inmediatamente después del cambio a condiciones anaeróbicas para ambas cepas (el cambio a condiciones anaeróbicas se muestra en las figuras 6 (A) y (B) mediante una flecha). El etanol que ya estaba presente en el medio al inicio de la fermentación no se produjo por las levaduras, se originó a partir de la solución de Tween80/Ergosterol. En las condiciones aeróbicas que predominaban en el comienzo, el etanol lo descompuso la levadura mediante respiración.

Después de aproximadamente 80 horas, la cepa que tenía los genes de la vía metabólica de arabinosa en forma de codones optimizados muestra un metabolismo significativamente mejorado de arabinosa y una producción aumentada de etanol. La arabinosa presente en el medio se ha consumido completamente justo después de 150 horas. Por el contrario, incluso después de 180 horas aún existe arabinosa en el medio con la cepa con los plásmidos reaislados originales.

Los resultados de fermentación muestran que los genes con codones optimizados posibilitan que los transformantes de levadura metabolicen arabinosa de forma más eficaz. El resultado de esto es que el azúcar se metaboliza más rápido y se obtiene un rendimiento significativamente mayor de etanol.

### Ejemplo 3: Construcción de un casete de expresión con tres genes para la vía metabólica de arabinosa

El vector con el casete de expresión con tres genes para la vía metabólica de arabinosa se construyó tanto para evitar los problemas que pueden surgir cuando están presentes varios plásmidos en la misma célula al mismo tiempo ("Plasmid stress", Review of *E. coli* por Bailey (1993)), como para posibilitar la integración genómica estable de los genes de la vía metabólica de arabinosa. Las cuestiones asociadas con la construcción de un casete de expresión de los genes de la vía metabólica de arabinosa y su integración individualmente de un modo que sea genómicamente estable ya se han mostrado por Becker (2003) y Wiedemann (2005). El casete de expresión con los tres genes que ahora se ha construido representa un excelente punto de partida para la integración genómica directa y posibilita la subclonación en los plásmidos de integración de la serie pRS303X, pRS305X y pRS306X (Taxis y Knop, 2006).

#### A) Construcción del casete de expresión

El punto de partida para construir el casete de expresión fue el plásmido p425H7-synthEpi, en que la forma de codones optimizados de epimerasa se expresó a partir de *E. coli* detrás del promotor *HXT7* acortado y delante del terminador *CYC1*. Para evitar la posible recombinación homóloga entre regiones idénticas del promotor o el terminador, la ORF de *araB<sup>mut</sup>* de codones optimizados debe expresarse a partir de *E. coli* entre el promotor *PFK1* y el terminador *FBA1*, la ORF de *araA* de codones optimizados de *B. licheniformis* entre el promotor *FBA1* y el terminador *PGK1*. El plásmido p425H7-synthEpi se abrió antes del promotor *HXT7* con la endonucleasa de restricción *SacI*, se sembró en estrías en un gel de agarosa, y se eluyó del gel. La ORF de *araB<sup>mut</sup>* se amplificó por PCR. El promotor *PFK1* y el terminador *FBA1* se amplificaron a partir del ADN genómico de *S. cerevisiae*, habiéndose seleccionado los cebadores de modo que se sintetizaran una secuencia de 500 pb de longitud del promotor y una secuencia de 300 pb de longitud del terminador y se produjeran proyecciones homólogas al plásmido p425H7-synthEpi y a la ORF de *araB<sup>mut</sup>* al mismo tiempo. El cebador que amplificó el promotor *PFK1* con las regiones homólogas a p425H7-synthEpi también contenía una secuencia para un sitio de restricción *PacI*. Los tres productos de PCR se introdujeron por transformación en levadura junto con el vector linealizado y se clonaron en el plásmido mediante recombinación *in vivo*. Se usó análisis de restricción para verificar que el plásmido p425H7synthEpisynthKin producido se había reconstruido satisfactoriamente. Se ensayó la funcionalidad del vector. Para hacer esto, se prepararon transformantes de levadura que contenían los plásmidos p425H7synthEpisynthKin y p423H7araABs<sup>re</sup>. Los transformantes se ensayaron para el crecimiento en arabinosa. La cepa era capaz de crecer en un medio que contenía arabinosa. Se usó una cepa de levadura que contenía los vectores p424H7synthEpi y p423HXT7-6HIS como control negativo. Esta cepa no fue capaz de crecer en el medio.

En la siguiente etapa, la forma de codones optimizados de la isomerasa de *B. licheniformis* se integró en el vector. Para esto, se linealizó el plásmido p425H7synthEpisynthKin con *NgoMVI* tras el terminador *CYC1*, se sembró en estrías en un gel de agarosa y se eluyó del gel. Se amplificó una secuencia de 500 pb de longitud del promotor *FBA1* del ADN genómico de *S. cerevisiae*, y los cebadores se seleccionaron de modo que se produjeran proyecciones homólogas al plásmido p425H7synthEpisynthKin en el terminador *CYC1* y a la ORF del *araA* de codones optimizados. También se amplificó una secuencia de 300 pb de longitud del terminador *PGK1* a partir del ADN genómico de *S. cerevisiae*, en que un cebador tenía proyecciones a la ORF del *araA* de codones optimizados y el otro cebador incluía proyecciones homólogas al plásmido p425H7synthEpisynthKin y un sitio de restricción *AscI*.

Se usó de nuevo análisis de restricción para verificar la construcción satisfactoria del plásmido p425H7synthAra, y se ensayó su funcionalidad. El ensayo para la funcionalidad se realizó para el crecimiento en arabinosa. Los transformantes de levadura que contenían el plásmido p425H7synthAra mostraron crecimiento en un medio que contenía arabinosa. Se registraron curvas de crecimiento en 5 ml de medio SC con arabinosa al 2 %. La Figura 7 muestra que los transformantes con el vector p425H7synthAra muestran crecimiento comparable a una cepa con los cuatro plásmidos reaislados originales.

**B) Papel de los promotores y terminadores**

Para evitar la posible recombinación homóloga entre las regiones promotora y terminadora, los tres genes se clonaron detrás de diferentes promotores y terminadores. En este contexto, la selección de los promotores fue particularmente importante. Se había descubierto en previas investigaciones (Becker y Boles, 2003) que la dosis de gen de los tres genes relativa unos a los otros era sumamente importante. Además, los tres genes tenían que expresarse de forma potente. Por estas razones, se tomó la decisión de usar el promotor HXT7 acortado, que se expresa de forma potente y constitutiva, y los promotores PFK1 y FBA1, que se sabe de ambos que promueven una potente expresión de genes.

**C) Ejemplos de vectores para el casete de expresión**

El plásmido de partida para la construcción de p425H7synthAra fue el plásmido p425H7synthEpi, que se basa en el plásmido p425HXT7-6HIS. El vector es un plásmido de expresión  $2\mu$  que posee un marcador de leucina.

Los tres genes de la vía metabólica de arabinosa se clonaron en un vector uno tras otro bajo el control de diversos promotores y terminadores. El casete de expresión está flanqueado por las secuencias de reconocimiento de las enzimas *Pacl* y *Ascl*.

Otros posibles vectores de expresión provienen de las series pRS303X, p3RS305X y p3RS306X. Éstos son vectores de integración que tienen un marcador de antibiótico dominante. Se proporciona más información acerca de estos vectores en Taxis y Knop (2006).

**Referencias**

- Becker, J. (2003)  
Konstruktion und Charakterisierung eines L-Arabinose fermentierenden *Saccharomyces cerevisiae* Hefestammes.  
Dissertation, Heinrich-Heine-Universität Düsseldorf
- Becker, J. y Boles, E. (2003)  
A modified *Saccharomyces cerevisiae* strain that consumes L-arabinose and produces ethanol.  
*Appl. Environ. Microbiol.* 69:4144-4150
- Bailey, J.E. (1993)  
Host-vector interactions in *Escherichia coli*.  
*Adv. Biochem Eng.* 48:29-52
- Bennetzen, J.L. y Hall, B.D. (1982)  
Codon selection in yeast.  
*J Biol Chem.* 257(6):3026-2031.
- Birnboim, H.C. y J. Doly (1979)  
A rapid alkaline extraction procedure for screening recombinant plasmid DNA.  
*Nucl. Acids Res.* 7: 1513-1523
- Dien, B.S., Kurtzman, C.P., Saha, B.C. y Bothast, R.J. (1996)  
Screening for L-arabinose fermenting yeasts.  
*Appl. Biochem. Biotechnol.* 57-58, 233-42.
- Dower, W.J., Miller, J.F. y Ragsdale, C.W. (1988)  
High efficiency transformation of *E. coli* by high voltage electroporation.  
*Nucl. Acids Res.* 16: 6127-6145
- Du Preez, J.C., Bosch, M., y Prior, B.A. (1986)  
The fermentation of hexose and pentose sugars by *Candida shehatae* and *Pichia stipitis*.  
*Appl. Microbiol. Biotechnol.* 23: 228–233.
- Gietz, R.D. y Woods, R.A. (1994)  
High efficiency transformation in yeast.  
En: *Molecular Genetics of Yeast: Practical Approaches*, J.A. Johnston (Ed.).  
Oxford University Press pág. 121-134
- Hahn-Hägerdal, B., Wahlbom, C.F., Gárdonyi, M., van Zyl, W., Otero, R.R.C. y Jönsson, L.J. (2001)  
Metabolic engineering of *Saccharomyces cerevisiae* for xylose utilization.  
*Adv. Biochem. Eng. Biotechnol.* 73:53-84.

- 5 Hamacher, T., Becker, J., Gárdonyi, M., Hahn-Hägerdal, B. y Boles., E. (2002)  
 Characterization of the xylose-transporting properties of yeast hexose transportes and their influence on xylose utilization.  
*Microbiology* 148:2783-2788.
- 10 Hoekema A, Kastelein RA, Vasser M, de Boer HA. (1987)  
 Codon replacement in the PGK1 gene of *Saccharomyces cerevisiae*: experimental approach to study the role of biased codon usage in gene expression.  
*Mol Cell Biol.* 7(8):2914-2924.
- 15 Karhumaa, K., Wiedemann, B., Hahn-Hägerdal, B., Boles, E. y Gorwa-Grauslund, MF. (2006)  
 Co-utilisation of L-arabinose and D-xylose by laboratory and industrial *Saccharomyces cerevisiae* strains.  
*Microbial Cell Factories* 5(1):18
- 20 Kötter, P. y Ciriacy, M. (1993)  
 Xylose fermentation by *Sacharomyces cerevisiae*.  
*Appl. Microbiol. Biotechnol.* 38: 776-783.
- 25 Liang, H. y Gaber, R.F. (1996)  
 A novel signal transduction pathway in *Saccharomyces cerevisiae* defined by Snf3-regulated expression of *HXT6*.  
*Mol. Biol. Cell* 7: 1953-1966.
- 30 Maniatis T, Fritsch, E.F y Sambrook, J. (1982)  
 Molecular cloning. A laboratory manual.  
 Cold Spring Harbor Laboratory, Nueva York.
- 35 Noelling J, Breton G, Omelchenko MV, Makarova KS, Zeng Q, Gibson R, Lee HM, Dubois J, Qiu D, Hitti J, Wolf YI, Tatusov RL, Sabathe F, Doucette-Stamm L, Soucaille P, Daly MJ, Bennett GN, Koonin EV, Smith DR. (2001)  
 Genome sequence and comparative analysis of the solvent-producing bacterium *Clostridium acetobutylicum*.  
*J Bacteriol.* 183(16):4823-38.
- 40 Sedlak, M. y Ho, N.W.Y. (2001)  
 Expression of *E. coli* araBAD operon encoding enzymes for metabolizing L-arabinose in *Saccharomyces cerevisiae*.  
*Enz. Microbiol.* 28:16-24
- 45 Taxis, C. y Knop, M. (2006)  
 System of centromeric, episomal, and integrative vectors based on drug resistance markers for *Saccharomyces cerevisiae*.  
*BioTechniques* 40, N° 1
- 50 Veith B, Herzberg C, Steckel S, Feesche J, Maurer KH, Ehrenreich P, Bäumer S, Henne A, Liesegang H, Merkl R, Ehrenreich A, Gottschalk G.  
 The complete genome sequence of *Bacillus licheniformis* DSM13, an organism with great industrial potential.  
*J Mol Microbiol Biotechnol.* 7(4):204-11.
- 55 Verduyn, C., Postma, E., Scheffers, W.A. y Van Dijken, J.P. (1992)  
 Effect of benzoic acid on metabolic fluxes in yeasts: a continuous-culture study on the regulation of respiration and alcoholic fermentation.  
*Yeast* 8 (7), 501-17
- 60 Wiedemann, B.(2005)  
 Molekulargenetische und physiologische Charakterisierung eines rekombinanten Pentose-vergärenden Hefestammes. Diplomarbeit. Johann Wolfgang Goethe-Universität, Frankfurt am Main.
- 65 Wirth;R. (1993)  
 Elektroporation: Eine alternative Methode zur Transformation von Bakterien mit Plasmid-DNA.  
*Forum Mikrobiologie* 11 (507-515).
- 70 Wu G, Bashir-Bello N, Freeland SJ. (2006)  
 The Synthetic Gene Designer: a flexible web platform to explore sequence manipulation for heterologous expression.  
*Protein Expr Purif.* 47(2):441-445.
- 75 Zimmermann, F. K. (1975)  
 Procedures used in the induction of mitotic recombination and mutation in the yeast *Saccharomyces cerevisiae*.

# ES 2 463 486 T3

*Mutation Res.* 31:71-81

## LISTA DE SECUENCIAS

- 5 <110> Butalco GmbH
- <120> Vector con genes de codones optimizados para una vía metabólica de arabinosa para la conversión de arabinosa en levadura para la producción de etanol
- 10 <130> FB22179
- <160> 7
- <170> PatentIn versión 3.2
- 15 <210> 1  
<211> 1704  
<212> ADN  
<213> Artificial
- 20 <220>  
<223> Secuencia génica de la fase de lectura abierta (ORF) de araBmut de E. coli en forma de codones optimizados
- 25 <400> 1
- |    |  |      |
|----|--|------|
|    | atggctattg ctattggttt ggacttcggt tctgactctg ttagagcttt ggctgttgac  | 60   |
|    | tgtgcttctg gtgaagaaat tgctacctct gttgaatggt acccaagatg gcaaaagggg  | 120  |
| 30 | caattctgtg acgctccaaa caaccaattc agacaccacc caagagacta cattgaatct  | 180  |
|    | atggaagctg ctttgaagac cgttttggct gaattgtctg ttgaacaaag agctgctggt  | 240  |
| 35 | gttggtattg gtgttgactc taccggttct accccagctc caattgacgc tgacggtaac  | 300  |
|    | gttttggctt tgagaccaga attcgcgtgaa aaccctaacg ctatgttcgt tttgtggaag | 360  |
|    | aaccacaccg ctggtgaaga agctgaagaa attaccagat tgtgtcacgc tccaggtaac  | 420  |
| 40 | gttgactact ctagatacat tgggtggtatt tactcttctg aatggttctg ggctaagatt | 480  |
|    | ttgcacgtta ccagacaaga ctctgctggt gctcaatctg ctgcttcttg gattgaattg  | 540  |
| 45 | tgtgactggg ttccagcttt gttgtctggt accaccagac cacaagacat tagaagaggt  | 600  |
|    | agatgttctg ctggtcacaa gtctttgtgg cacgaatctt ggggtggttt gccaccagct  | 660  |
|    | tctttcttcg acgaattgga cccaattttg aacagacact tgccatctcc attgttcacc  | 720  |
| 50 | gacacctgga ccgctgacat tccagttggt accttgtgtc cagaatgggc tcaaagattg  | 780  |
|    | ggtttgccag aatctgttgt tatttctggt ggtgctttcg actgtcacat gggtgctggt  | 840  |
| 55 | ggtgctggtg ctcaacccaaa cgctttgggt aaggttattg gtacctctac ctgtgacatt | 900  |
|    | ttgattgctg acaagcaatc tgttggtgaa agagctgtta agggatattg tggtaagtt   | 960  |
|    | gacggttctg ttgttccagg tttcattggt ttggaagctg gtcaatctgc tttcgggtgac | 1020 |
| 60 | atttacgctt ggttcggtag agttttgtct tggccattgg aacaattggc tgctcaacac  | 1080 |

## ES 2 463 486 T3

	ccagaattga aggctcaaat taacgcttct caaaagcaat tgttgccagc tttgaccgaa	1140
	gcttgggcta agaaccacatc tttggaccac ttgccagttg ttttggactg gttcaacggt	1200
5	agaagatctc caaacgctaa ccaaagattg aagggtgtta ttaccgactt gaacttggct	1260
	accgacgctc cattgttggt cggtggtttg attgctgcta ccgctttcgg tgctagagct	1320
10	attatggaat gtttcaccga ccaaggtatt gctgtcaaca acgttatggc tttgggtggt	1380
	attgctagaa agaaccaagt tattatgcaa gcttgttggt acgttttgaa cagaccattg	1440
	caaattggtg cttctgacca atgttgtgct ttgggtgctg ctattttcgc tgctggtgct	1500
15	gctaaggttc acgctgacat tccatctgct caacaaaaga tggcttctgc tgttgaaaag	1560
	accttgcaac caagatctga acaagctcaa agattcgaac aattgtacag aagataccaa	1620
20	caatgggcta tgtctgctga acaacactac ttgccaacct ctgctccagc tcaagctgct	1680
	caagctggtg ctaccttgta ataa	1704
25	<210> 2 <211> 699 <212> ADN <213> Artificial	
30	<220> <223> Secuencia génica de la fase de lectura abierta (ORF) de araD de E. coli en forma de codones optimizados	
	<400> 2	
35	atgttgggaag acttgaagag acaagttttg gaagctaact tggctttgcc aaagcacaac	60
	ttggttacct tgacctgggg taacgtttct gctgttgaca gagaaagagg tgttttcgtt	120
	attaagccat ctggtgttga ctactctggt atgaccgctg acgacatggt tgttgtttct	180
40	attgaaaccg gtgaagttgt tgaaggtacc aagaagccat cttctgacac cccaaccac	240
	agattgttgt accaagcttt cccatctatt ggtggtattg ttcacacca ctctagacac	300
45	gctaccattt gggctcaagc tgggtcaatct attccagcta ccggtaccac ccacgctgac	360
	tacttctacg gtaccattcc atgtaccaga aagatgaccg acgctgaaat taacggtgaa	420
	tacgaatggg aaaccggtaa cgttattggt gaaaccttcg aaaagcaagg tattgacgct	480
50	gctcaaatgc caggtgtttt ggttcaactct cacgggtccat tcgcttgggg taagaacgct	540
	gaagacgctg ttcacaacgc tattgttttg gaagaagttg cttacatggg tattttctgt	600
55	agacaattgg ctccacaatt gccagacatg caacaaacct tgttggacaa gcactacttg	660
	agaaagcacg gtgctaaggc ttactacggt caataataa	699
60	<210> 3 <211> 1485 <212> ADN	

## ES 2 463 486 T3

<213> Artificial

<220>

<223> Secuencia génica de la fase de lectura abierta (ORF) de araA de Bacillus licheniformis en forma de codones optimizados

5

<400> 3

10	atgattcaag ctaagaccca cgttttctgg ttcgttaccg gttctcaaca cttgtacggt	60
	gaagaagctg ttcaagaagt tgaagaacac tctaagatga tttgtaacgg tttgaacgac	120
	ggtgacttga gattccaagt tgaatacaag gctgttgcta cctctttgga cggtgtaga	180
15	aagttgttcg aagaagctaa cagagacgaa gaatgtgctg gtattattac ctggatgcac	240
	accttctctc cagctaagat gtggattcca ggtttgtctg aattgaacaa gccattggtg	300
	cacttcaca cccaattcaa cagagacatt ccatgggaca agattgacat ggacttcatg	360
20	aacattaacc aatctgctca cggtgacaga gaatacgggt tcattggtgc tagattgggt	420
	attccaagaa aggttattgc tggttactgg gaagacagag aagttaagag atctattgac	480
25	aagtggatgt ctgctgctgt tgcttacatt gaatctagac acattaaggt tgctagattc	540
	ggtgacaaca tgagaaacgt tgctgttacc gaaggtgaca agattgaagc tcaaatcaa	600
	ttggggttgg ctggtgacgg ttacggtatt ggtgacttgg ttaccgaaat taacgctggt	660
30	tctgaacaat ctttgtctga attgatttct gaatacgaag aattgtacga atggccagaa	720
	ggtgaagctg ctagagaatc tgttaaggaa caagctagaa ttgaattggg tttgaagaga	780
35	ttcttgtctt ctggtgggta caccgcttcc accaccacct tcgaagactt gcacggtatg	840
	aagcaattgc caggtttggc tgttcaaaga ttgatggctg aaggttacgg tttcgggtgt	900
	gaaggtgact ggaagaccgc tgctttgggt agaatgatga agatgatggc tggtggtgtaag	960
40	gaaacctctt tcatggaaga ctacacctac cacttcgaac caggtaacga aatgattttg	1020
	ggttctcaca tgttggaagt ttgtccatct attgctgaac acaagccaag aattgaagtt	1080
45	caccattgt ctatgggtgc taaggacgac ccagctagat tggttttcga cggatttgct	1140
	ggtccagctg tcaacgttcc tttgattgac ttgggtggta gattcagatt ggttattaac	1200
	aaggttgaag ctgttaaggt tccacacgac atgccaaact tgccagttgc tagagttttg	1260
50	tggaagccac aaccatcttt gagaacctct gctgaagctt ggattttggc tgggtggtgct	1320
	caccacacct gtttgtctta ccaattgacc gctgaacaaa tgttggactg ggctgaaatg	1380
55	tctggtattg aagctgtttt gattaacaga gacaccacca ttttgaactt gagaaacgaa	1440
	ttgaagtggc ctgaagctgc ttacagattg agaaagttct aataa	1485

60

<210> 4

<211> 1482

<212> ADN

ES 2 463 486 T3

<213> Bacillus licheniformis

<400> 4

5 atgattcagg caaagacgca tgtgttttgg tttgtgacag gcagccagca tttatatggc 60  
gaagaggcgg tacaagaggt agaagagcat tccaaaatga tctgcaacgg attaaatgac 120  
10 ggagatttaa ggtttcaagt cgagtacaaa gcggtggcca cttcgcctgga cggcgtcaga 180  
aaactgtttg aagaggcgaa ccgggacgaa gagtgcgag gcatcatcac ctggatgcat 240  
acgttttcac cggccaaaat gtggattccc ggcctttccg agctgaataa gccgctgctc 300  
15 cattttcata cccagtttaa ccgggacatt ccgtgggata aaatcgacat ggatttcatg 360  
aatattaatc agtctgcccc cggcgaccgc gaatacgggt ttatcggagc gagattgggc 420  
attcctcgaa aagtaatcgc cggatattgg gaagacagag aagtaaagcg ctcgatcgac 480  
20 aaatggatga gcgcagcggc cgcatatatt gaaagccgcc atatcaaagt cgccccgattt 540  
ggggacaaca tgcggaatgt ggcggtaaca gaaggagata agattgaagc gcagattcag 600  
25 cttggctggt ctgtcgacgg atatggaatc ggcgatctcg tcacggagat caatgctgta 660  
tcagaacaga gtctaagcga gctcatctcc gaatatgagg agctttatga gtggccggaa 720  
ggcgaagcgg caagggaatc cgtcaaggag caggcgcgga ttgagcttgg attaaagcgc 780  
30 tttctttcga gcggaggcta tacggccttt acgacaacct ttgaagacct ccacggcatg 840  
aagcagcttc cgggtttggc tgttcagcgg ctgatggctg aaggctatgg ctttggagga 900  
35 gaaggagatt ggaaaacggc tgcgcttgtc aggatgatga aaatgatggc cggcggaaaa 960  
gaaacgtcat ttatggagga ttacacgtac cattttgaac cgggcaatga aatgatcctc 1020  
ggctcccaca tgcttgaagt gtgcccttcg attgcccgaac acaaaccgag aattgaagtc 1080  
40 catccgctgt caatgggtgc aaaggatgat ccggcgcggc tcgtgtttga cggaaattgca 1140  
ggtcctgceg tcaatgtctc ccttattgat ttgggcggac gtttcaggct tgtcatcaat 1200  
45 aaagtggaag ctgtgaaggt tccgcatgat atgccgaatc ttccggtggc gcgcgtactt 1260  
tggaagcctc agccgtctct gagaacatcg gctgaagcgt ggattttggc tggcgggtgct 1320  
caccatacat gcctttcgta ccagctgaca gccgagcaga tgctcgattg ggccgaaatg 1380  
50 tcgggcattg aagcgggtgt gatcaatcgt gatacaacga ttttaaactc tcgcaatgaa 1440  
ctcaaatgga gtgaagctgc atatcggctt cgcaagtttt aa 1482

55

<210> 5

<211> 1467

<212> ADN

<213> Clostridium acetobutylicum

60

<400> 5

ES 2 463 486 T3

	atggttagaaa ataaaaagat ggaatTTTTgG tttgtagtag gaagtcaaca tttatatggt	60
	gaagaggcctt taaaagaagt aagaaaaaat tctgagacaa ttgtagatga attaaataaa	120
5	agtgctaate ttccatataa aataatattt aaagatttag caacttctgc tgataaaata	180
	aaggaaataa tgaaggaagt taactataga gatgaagtag caggagttat aacttggatg	240
10	catacgtttt ctccagctaa aatgtggata gcaggtacaa agatattaca aaaaccttta	300
	cttcattttg caactcaata taatgaaaat attccatgga aaacaataga tatggattat	360
	atgaacttac atcaaagtgC tcatggagat agagagtatg gatttattaa tgcaagactt	420
15	aaaaagcata ataaagttgt tgtaggatat tggaaggata aagaagttca aaaacaagtt	480
	tcagattgga tgaaggttgc tgcaggatat attgcaagtG aaagcataaa agttgcacgt	540
20	tttggtgata acatgcgtaa tgttgCagtt acagagggag ataaagtaga agctcaaata	600
	caattcggat ggacagtaga ttactttggt ataggtgatt tagttgctga aatggacaag	660
	gtaagccaag atgaaataaa taaaacttat gaagaattta aagatttata tatttttagat	720
25	ccaggtgaaa atgatcctgc tttctatgag aaacaagtta aagaacaaat caaaattgaa	780
	ataggattaa gaaggttctt agaaaaagga aattataatg catttacaac aaactttgaa	840
30	gatctttatg gaatgaaaca gttacctgga cttgcagtac aacgtttaaa tgctgaaggc	900
	tatggctttg caggcgaagg agactggaaa actgcagctt tagatagatt attaaaggtt	960
	atgactaata atactgctac aggttttatg gaagattaca catatgaact tagtcgtgga	1020
35	aatgagaagg cattaggagc tcatatgctt gaagttgacc caacttttgc ttcagataaa	1080
	ccaaaggtta ttgttaaacc actaggaatt ggagataaag aagatccagc acgtttaatc	1140
40	tttaatggtt caacaggaaa aggtgtagca gtttcaatgc ttgatttagg aacacattat	1200
	cgtttaataa taaacggact tacagcagtg aaaccagatg aagacatgcc aaacctacca	1260
	gttgctaaaa tggatggaa accagaacca aacttcattg aaggagttaa atcttggatt	1320
45	tatgcaggtg gcggacatca tacagtgggt tcaactagaat taacagtaga acaggtttat	1380
	gattggagtc gtatggtagg cttggaagct gtaataatag ataaggatac taaattaaga	1440
50	gatataatag aaaagacaac aaaataa	1467
	<210> 6	
	<211> 493	
	<212> PRT	
	<213> Bacillus licheniformis	
55	<400> 6	

# ES 2 463 486 T3

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

ES 2 463 486 T3

Glu Ala Asn Arg Asp Glu Glu Cys Ala Gly Ile Ile Thr Trp Met His  
 65 70 75 80  
 5 Thr Phe Ser Pro Ala Lys Met Trp Ile Pro Gly Leu Ser Glu Leu Asn  
 85 90 95  
 Lys Pro Leu Leu His Phe His Thr Gln Phe Asn Arg Asp Ile Pro Trp  
 100 105 110  
 10 Asp Lys Ile Asp Met Asp Phe Met Asn Ile Asn Gln Ser Ala His Gly  
 115 120 125  
 Asp Arg Glu Tyr Gly Phe Ile Gly Ala Arg Leu Gly Ile Pro Arg Lys  
 130 135 140  
 15 Val Ile Ala Gly Tyr Trp Glu Asp Arg Glu Val Lys Arg Ser Ile Asp  
 145 150 155 160  
 20 Lys Trp Met Ser Ala Ala Val Ala Tyr Ile Glu Ser Arg His Ile Lys  
 165 170 175  
 Val Ala Arg Phe Gly Asp Asn Met Arg Asn Val Ala Val Thr Glu Gly  
 180 185 190  
 25 Asp Lys Ile Glu Ala Gln Ile Gln Leu Gly Trp Ser Val Asp Gly Tyr  
 195 200 205  
 Gly Ile Gly Asp Leu Val Thr Glu Ile Asn Ala Val Ser Glu Gln Ser  
 210 215 220  
 30 Leu Ser Glu Leu Ile Ser Glu Tyr Glu Glu Leu Tyr Glu Trp Pro Glu  
 225 230 235 240  
 35 Gly Glu Ala Ala Arg Glu Ser Val Lys Glu Gln Ala Arg Ile Glu Leu  
 245 250 255  
 Gly Leu Lys Arg Phe Leu Ser Ser Gly Gly Tyr Thr Ala Phe Thr Thr  
 260 265 270  
 40 Thr Phe Glu Asp Leu His Gly Met Lys Gln Leu Pro Gly Leu Ala Val  
 275 280 285  
 Gln Arg Leu Met Ala Glu Gly Tyr Gly Phe Gly Gly Glu Gly Asp Trp  
 290 295 300  
 45 Lys Thr Ala Ala Leu Val Arg Met Met Lys Met Met Ala Gly Gly Lys  
 305 310 315 320

ES 2 463 486 T3

Glu Thr Ser Phe Met Glu Asp Tyr Thr Tyr His Phe Glu Pro Gly Asn  
 325 330 335  
 5 Glu Met Ile Leu Gly Ser His Met Leu Glu Val Cys Pro Ser Ile Ala  
 340 345 350  
 Glu His Lys Pro Arg Ile Glu Val His Pro Leu Ser Met Gly Ala Lys  
 355 360 365  
 10 Asp Asp Pro Ala Arg Leu Val Phe Asp Gly Ile Ala Gly Pro Ala Val  
 370 375 380  
 Asn Val Ser Leu Ile Asp Leu Gly Gly Arg Phe Arg Leu Val Ile Asn  
 385 390 395 400  
 15 Lys Val Glu Ala Val Lys Val Pro His Asp Met Pro Asn Leu Pro Val  
 405 410 415  
 Ala Arg Val Leu Trp Lys Pro Gln Pro Ser Leu Arg Thr Ser Ala Glu  
 420 425 430  
 Ala Trp Ile Leu Ala Gly Gly Ala His His Thr Cys Leu Ser Tyr Gln  
 435 440 445  
 25 Leu Thr Ala Glu Gln Met Leu Asp Trp Ala Glu Met Ser Gly Ile Glu  
 450 455 460  
 Ala Val Leu Ile Asn Arg Asp Thr Thr Ile Leu Asn Leu Arg Asn Glu  
 465 470 475 480  
 30 Leu Lys Trp Ser Glu Ala Ala Tyr Arg Leu Arg Lys Phe  
 485 490  
 <210> 7  
 35 <211> 488  
 <212> PRT  
 <213> Clostridium acetobutylicum  
 <400> 7  
 40 Met Leu Glu Asn Lys Lys Met Glu Phe Trp Phe Val Val Gly Ser Gln  
 1 5 10 15  
 His Leu Tyr Gly Glu Glu Ala Leu Lys Glu Val Arg Lys Asn Ser Glu  
 45 20 25 30  
 Thr Ile Val Asp Glu Leu Asn Lys Ser Ala Asn Leu Pro Tyr Lys Ile  
 35 40 45  
 50

ES 2 463 486 T3

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

ES 2 463 486 T3

Pro Gly Leu Ala Val Gln Arg Leu Asn Ala Glu Gly Tyr Gly Phe Ala  
 290 295 300  
 5 Gly Glu Gly Asp Trp Lys Thr Ala Ala Leu Asp Arg Leu Leu Lys Val  
 305 310 315 320  
 Met Thr Asn Asn Thr Ala Thr Gly Phe Met Glu Asp Tyr Thr Tyr Glu  
 325 330 335  
 10 Leu Ser Arg Gly Asn Glu Lys Ala Leu Gly Ala His Met Leu Glu Val  
 340 345 350  
 Asp Pro Thr Phe Ala Ser Asp Lys Pro Lys Val Ile Val Lys Pro Leu  
 355 360 365  
 15 Gly Ile Gly Asp Lys Glu Asp Pro Ala Arg Leu Ile Phe Asn Gly Ser  
 370 375 380  
 Thr Gly Lys Gly Val Ala Val Ser Met Leu Asp Leu Gly Thr His Tyr  
 385 390 395  
 20 Arg Leu Ile Ile Asn Gly Leu Thr Ala Val Lys Pro Asp Glu Asp Met  
 405 410 415  
 25 Pro Asn Leu Pro Val Ala Lys Met Val Trp Lys Pro Glu Pro Asn Phe  
 420 425 430  
 Ile Glu Gly Val Lys Ser Trp Ile Tyr Ala Gly Gly Gly His His Thr  
 435 440 445  
 30 Val Val Ser Leu Glu Leu Thr Val Glu Gln Val Tyr Asp Trp Ser Arg  
 450 455 460  
 Met Val Gly Leu Glu Ala Val Ile Ile Asp Lys Asp Thr Lys Leu Arg  
 465 470 475 480  
 35 Asp Ile Ile Glu Lys Thr Thr Lys  
 485

## REIVINDICACIONES

- 5 1. Una célula hospedadora que contiene una molécula de ácido nucleico que comprende tres secuencias de ácido nucleico cada una de las cuales codifica un polipéptido de una vía metabólica bacteriana de L-arabinosa, donde las tres secuencias de ácido nucleico son araA (L-arabinosa isomerasa), araB (L-ribuloquinasa) y araD (L-ribulosa-5-P-4-epimerasa),  
donde la secuencia de ácido nucleico para araA se obtiene de *Bacillus licheniformis* o *Clostridium acetobutylicum*, y donde la célula hospedadora es una célula fúngica.
- 10 2. La célula hospedadora de acuerdo con la reivindicación 1, que es una célula de levadura, tal como especies de *Saccharomyces*, *Kluyveromyces* sp., *Hansenula* sp., *Pichia* sp. o *Yarrowia* sp.
- 15 3. La célula hospedadora de acuerdo con las reivindicaciones 1 ó 2, donde la secuencia de ácido nucleico para araB y la secuencia de ácido nucleico para araD son originarias de *E. coli*.
4. La célula hospedadora de acuerdo con una cualquiera de las reivindicaciones 1 a 3, donde la secuencia de ácido nucleico para araA comprende una secuencia de ácido nucleico con la SEC ID N° 3, 4 ó 5.
- 20 5. La célula hospedadora de acuerdo con una cualquiera de las reivindicaciones anteriores, donde al menos dos de las tres secuencias de ácido nucleico están optimizadas en los codones para su uso en una célula hospedadora.
- 25 6. La célula hospedadora de acuerdo con una cualquiera de las reivindicaciones anteriores, donde la secuencia de ácido nucleico para araB comprende una secuencia de ácido nucleico con la SEC ID N° 1, y/o donde la secuencia de ácido nucleico para araD comprende una secuencia de ácido nucleico con la SEC ID N° 2.
- 30 7. La célula hospedadora de acuerdo con una cualquiera de las reivindicaciones anteriores, donde la molécula de ácido nucleico comprende la secuencia de ácido nucleico con la SEC ID N° 1, la secuencia de ácido nucleico con la SEC ID N° 2 y la secuencia de ácido nucleico con las SEC ID N° 3, 4 ó 5.
- 35 8. La célula hospedadora de acuerdo con una cualquiera de las reivindicaciones anteriores, donde la célula hospedadora contiene la molécula de ácido nucleico integrada de un modo genómicamente estable.
9. La célula hospedadora de acuerdo con una cualquiera de las reivindicaciones anteriores, donde la molécula de ácido nucleico comprende ADNbc, ADNmc, APN, ACN, ARN o ARNm o combinaciones de los mismos.
- 40 10. La célula hospedadora de acuerdo con una cualquiera de las reivindicaciones anteriores, donde la molécula de ácido nucleico está comprendida en un casete de expresión o en un vector de expresión.
- 45 11. La célula hospedadora de acuerdo con la reivindicación 10, donde el casete de expresión comprende secuencias promotoras y terminadoras, o comprende diferentes pares de secuencias promotoras y terminadoras, respectivamente.
12. Un método para producir bioetanol, que comprende la expresión de una molécula de ácido nucleico, definida en una cualquiera de las reivindicaciones 1 a 10, en una célula hospedadora de acuerdo con una cualquiera de las reivindicaciones 1 a 11.
- 50 13. Uso de una célula hospedadora de acuerdo con una cualquiera de las reivindicaciones 1 a 11 para producir bioetanol o para la fermentación recombinante de biomaterial que contiene pentosa.

Figura 1

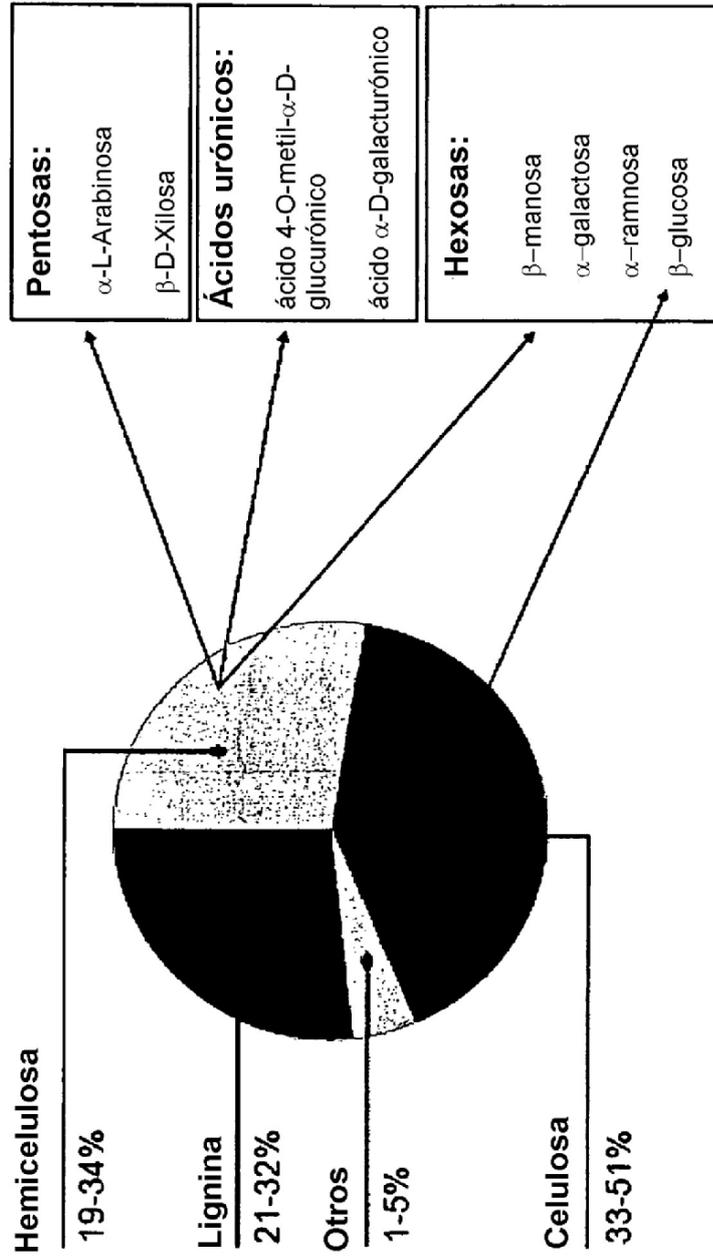
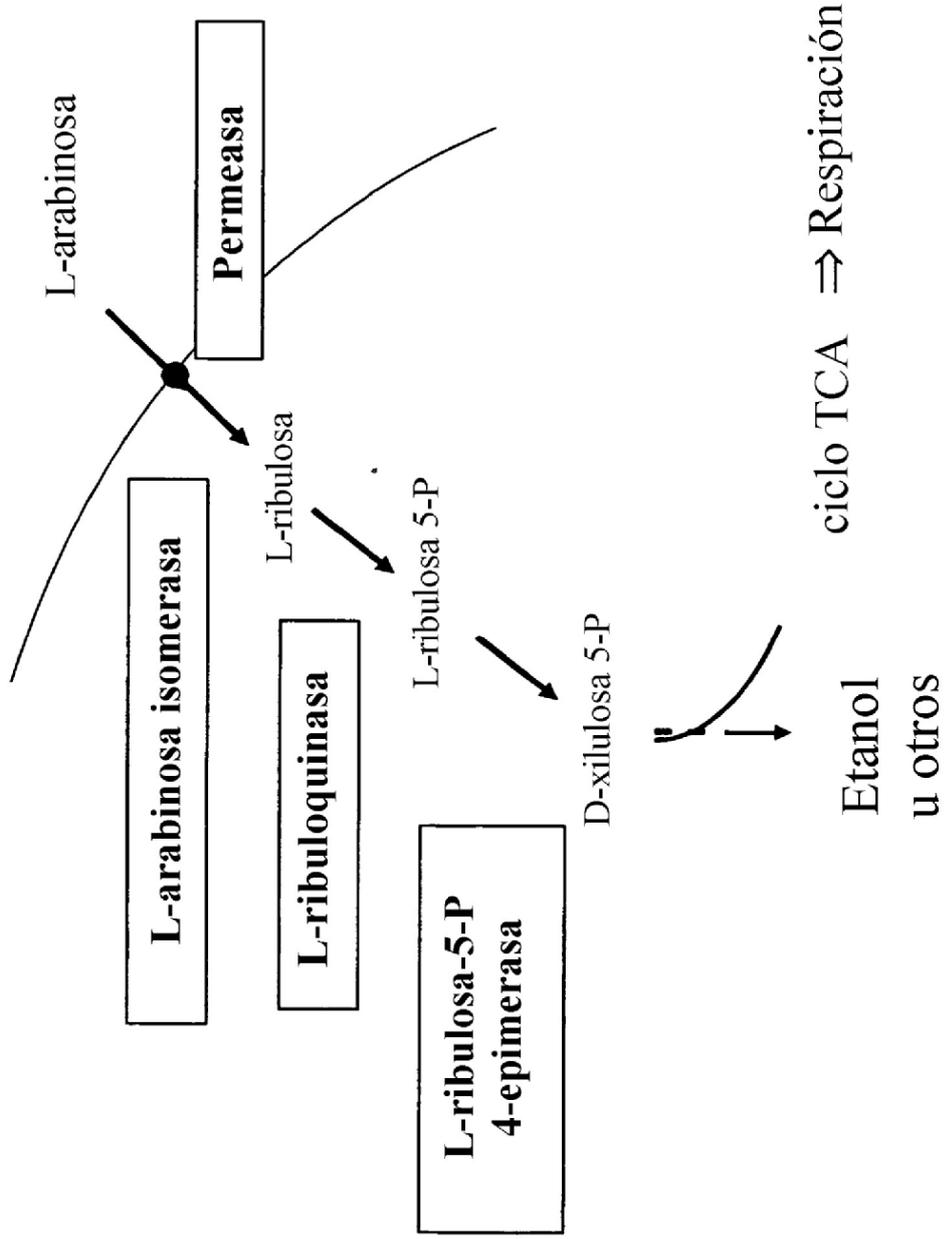


Figura 2



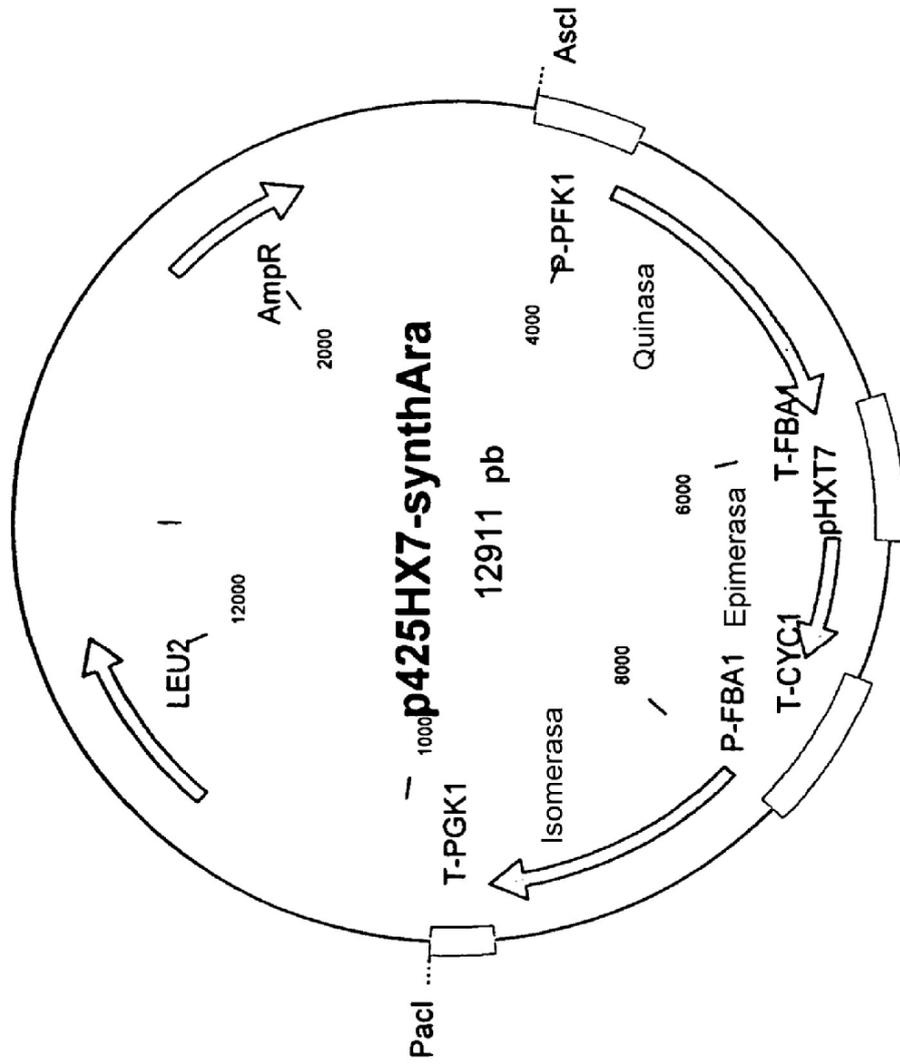


Figura 3 A

Figura 3 B

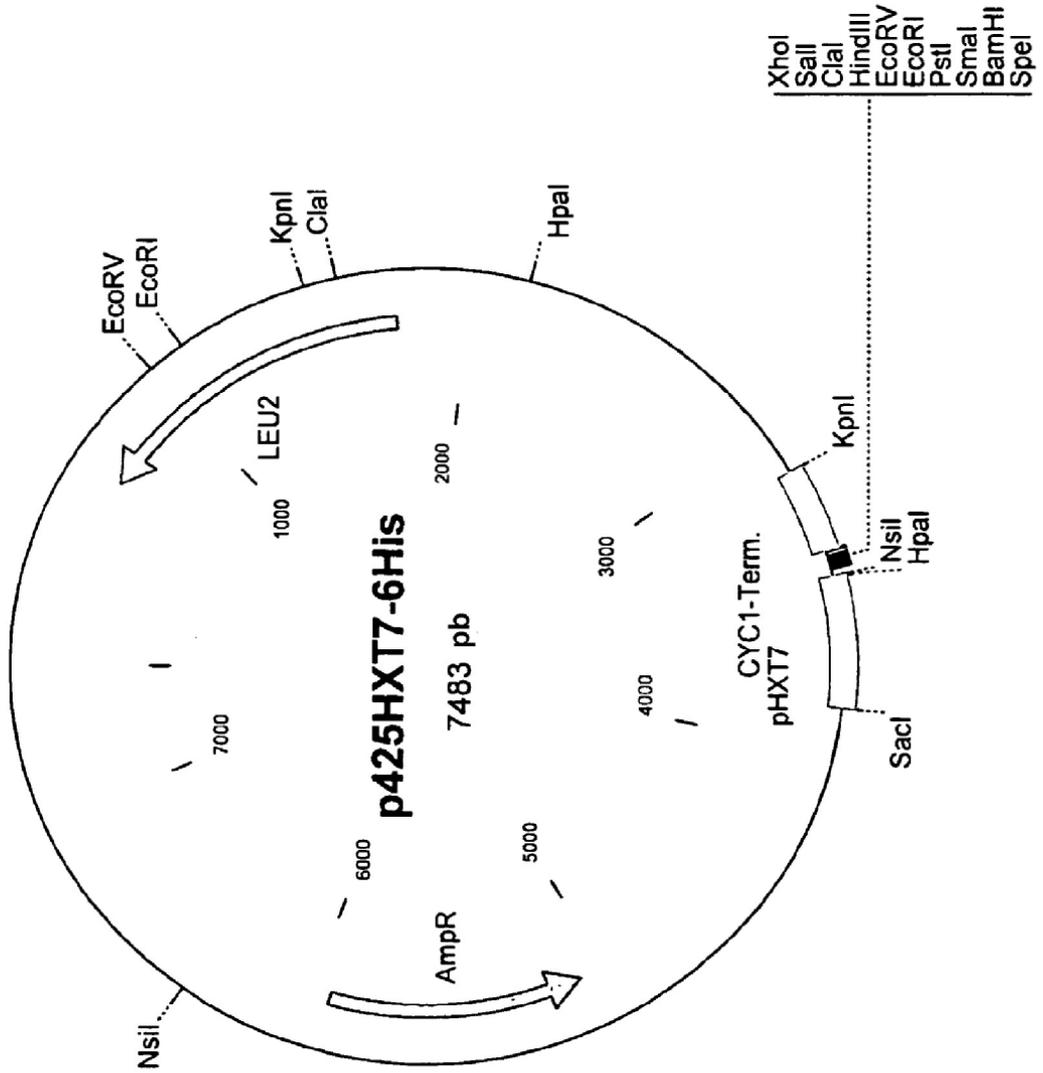


Figura 4

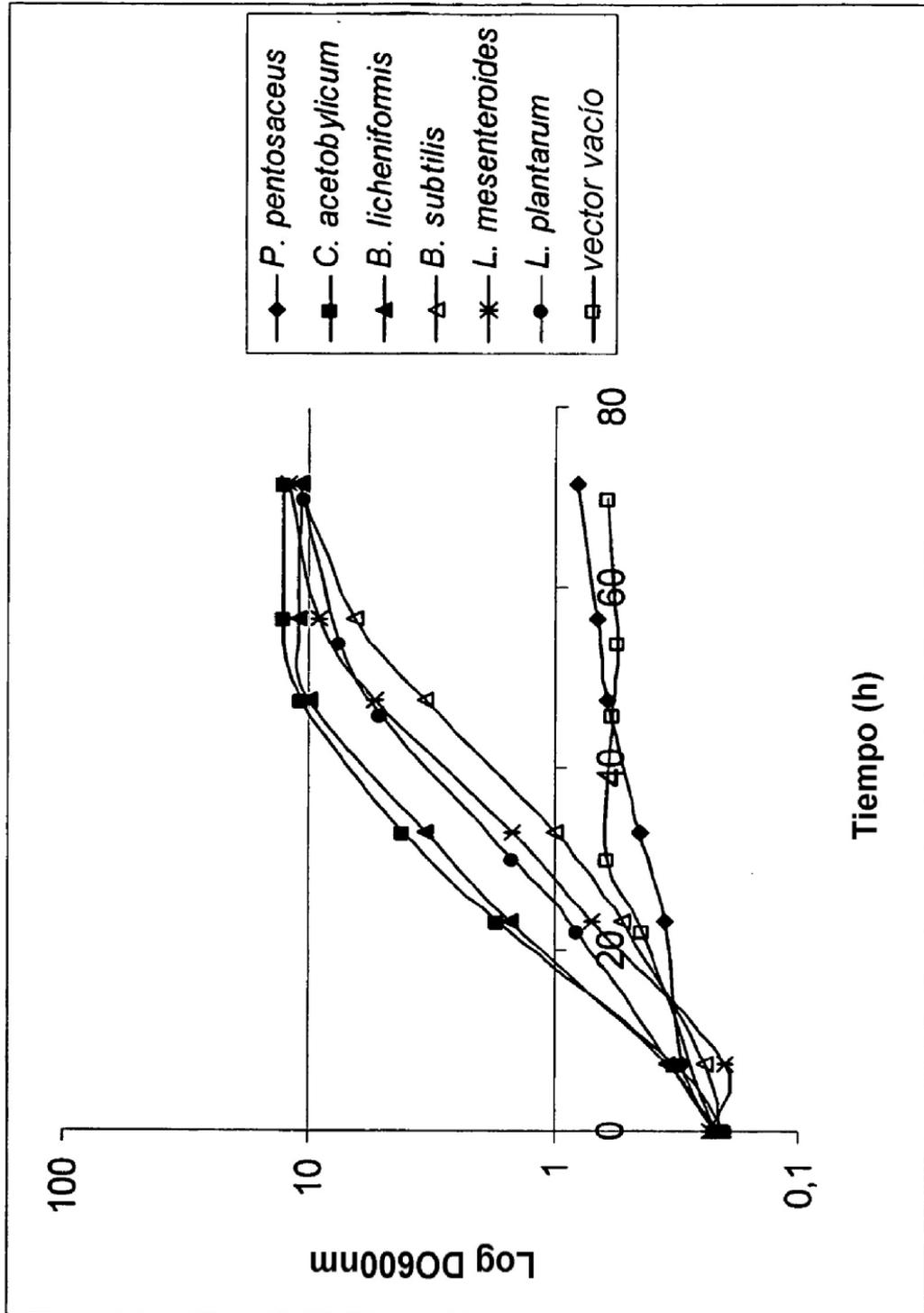


Figura 5

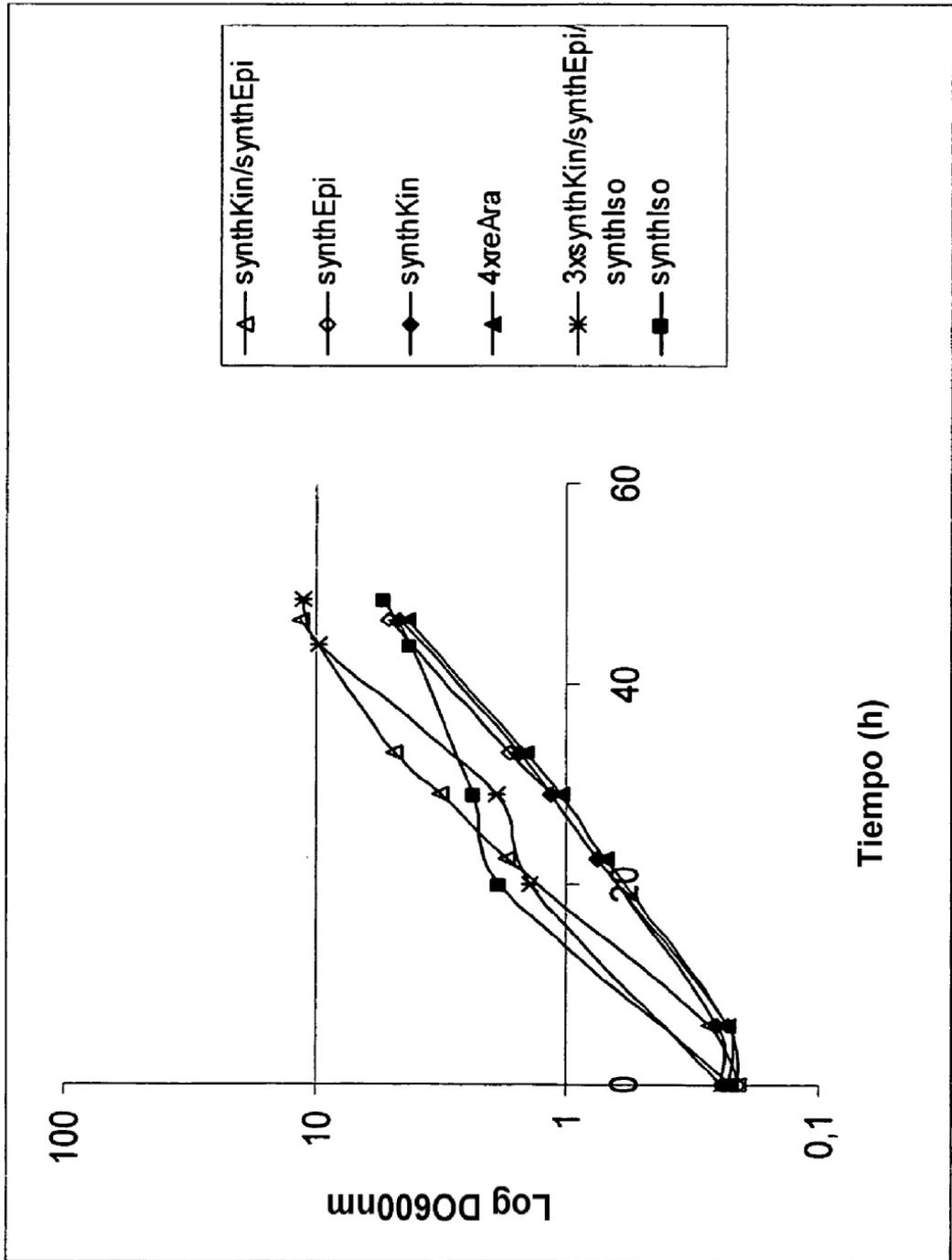


Figura 6 A

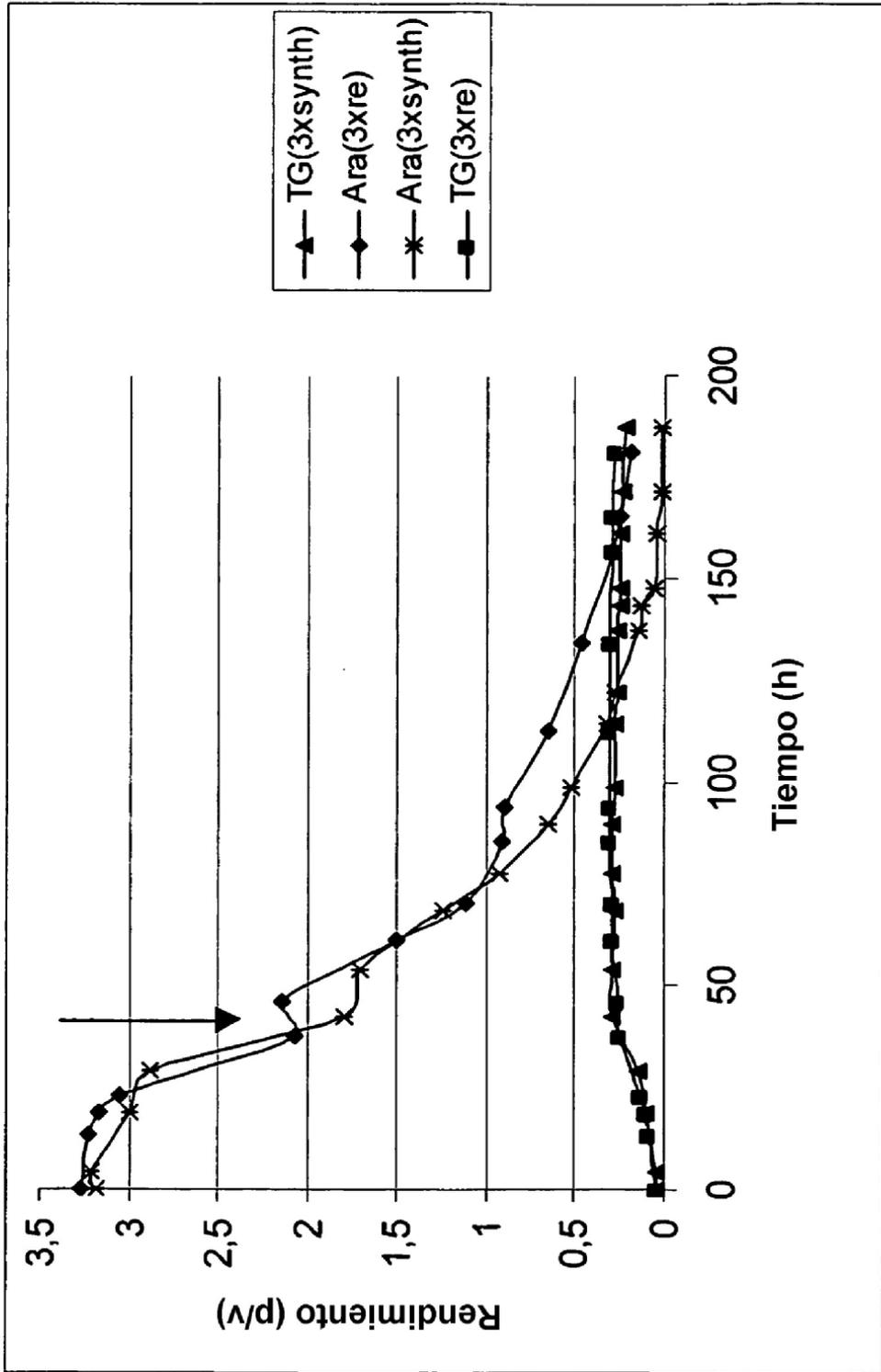


Figura 6 B

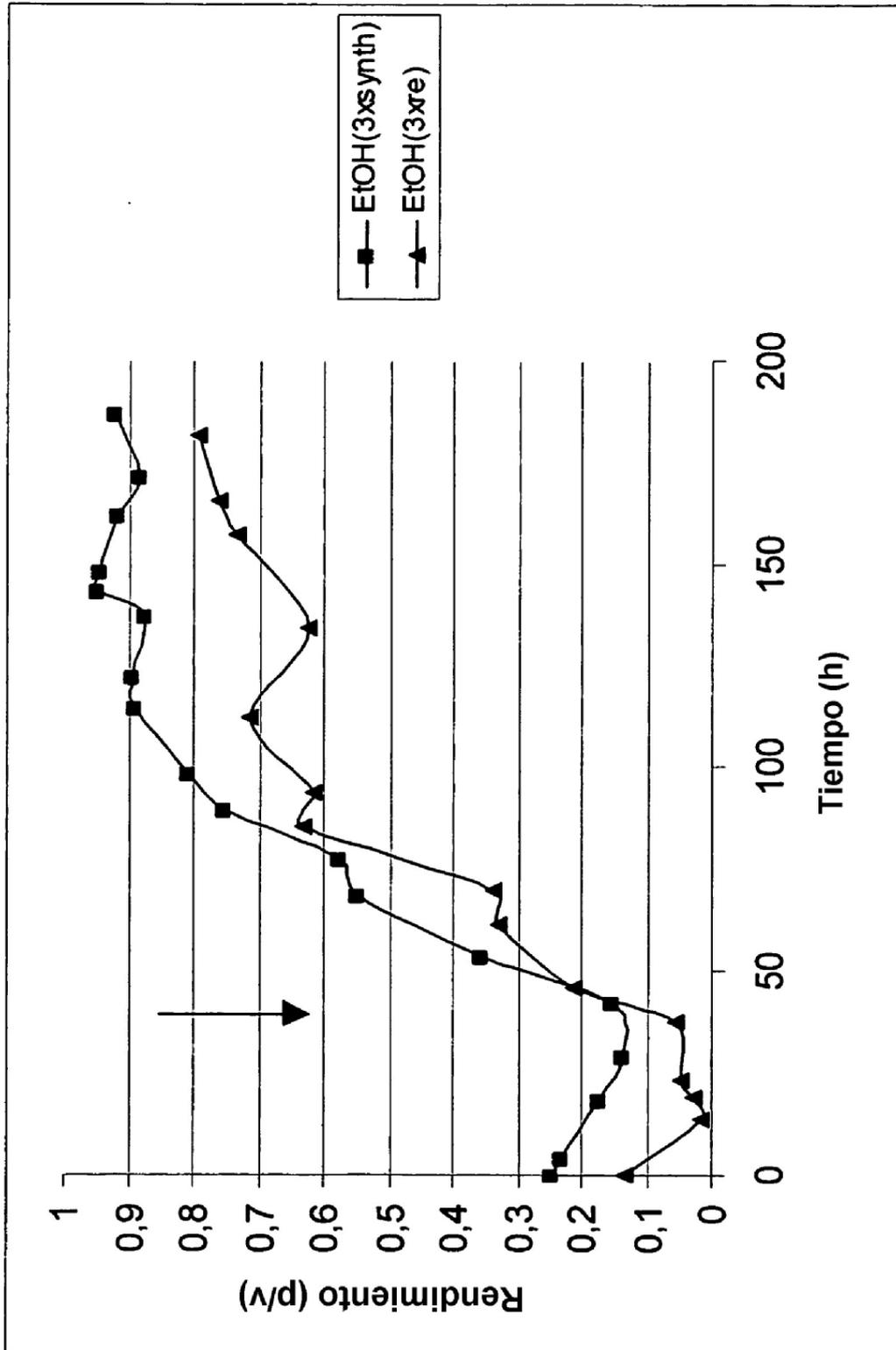


Figura 7

