

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 607 894**

51 Int. Cl.:

C12P 7/16 (2006.01)

C12Q 1/32 (2006.01)

C07K 14/33 (2006.01)

C12N 9/02 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **11.07.2008 PCT/EP2008/059119**

87 Fecha y número de publicación internacional: **29.01.2009 WO09013159**

96 Fecha de presentación y número de la solicitud europea: **11.07.2008 E 08775033 (7)**

97 Fecha y número de publicación de la concesión europea: **28.09.2016 EP 2173881**

54 Título: **Enzimas productoras de acetil-CoA en levadura**

30 Prioridad:

23.07.2007 EP 07112956

23.07.2007 US 935031 P

21.12.2007 EP 07123976

19.02.2008 EP 08101747

19.02.2008 US 64120

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

04.04.2017

73 Titular/es:

DSM IP ASSETS B.V. (100.0%)

Het Overloon 1

6411 TE Heerlen, NL

72 Inventor/es:

MUELLER, ULRIKE, MARIA;

WU, LIANG;

RAAMSDONK, LOURINA, MADELEINE y

WINKLER, AARON ADRIAAN

74 Agente/Representante:

LEHMANN NOVO, María Isabel

ES 2 607 894 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

Enzimas productoras de acetil-CoA en levadura

Campo de la invención

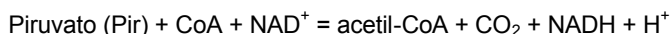
5 La presente invención está en el campo de la producción de metabolitos en levadura usando sistemas de expresión heterólogos. En particular, la presente invención se refiere a la ingeniería metabólica de cepas de levadura capaces de producir metabolitos que requieren acetil-CoA citosólica como precursor, tales como cepas de levadura productoras de butanol. La presente invención se refiere a un sistema de ensayo para identificar enzimas heterólogas capaces de convertir piruvato, acetaldehído o acetato en acetil-CoA citosólica cuando se expresa en el citosol en levadura.

Antecedentes de la invención

10 La acetil-coenzima A (CoA) es un producto intermedio esencial en numerosas vías metabólicas, y es un precursor clave en la síntesis de muchos compuestos relevantes industriales, tales como ácidos grasos, carotenoides, isoprenoides, vitaminas, aminoácidos, lípidos, ésteres de cera, (poli)sacáridos polihidroxicanoatos, estatinas, policétidos y ésteres acéticos (tales como acetato de etilo y acetato de isoamilo). En particular, la acetil-CoA también es el precursor del producto químico a granel industrialmente importante 1-butanol.

15 En comparación con las bacterias, tales como *E. coli*, las células de levadura proporcionan una alternativa muy adecuada para producir los productos derivados de acetil-CoA anteriormente mencionados, porque la levadura no es susceptible a fago u otra infección ya que los procesos basados en levadura pueden ejecutarse a un pH bajo. Por tanto, el uso de levadura no requiere un proceso estéril, reduciendo así el precio del coste del producto de interés.

20 Cuando la levadura natural (no mutante) no es capaz de producir el producto derivado de acetil-CoA de interés, el uso de ingeniería metabólica puede proporcionar que las células de levadura expresen genes heterólogos que podrían soportar un proceso tal. En tales casos, los productos de gen heterólogo se dirigen normalmente al compartimento citosólico de la levadura. Como la biosíntesis del producto derivado de acetil-CoA tendrá lugar completa o parcialmente en el citosol, el suministro de cantidades suficientes de la acetil-CoA precursora en el compartimento citosólico es crucial. En *Saccharomyces cerevisiae*, la biosíntesis de acetil-CoA tiene lugar en dos compartimentos separados. En mitocondrias, la acetil-CoA se sintetiza por descarboxilación oxidativa de piruvato catalizada por el complejo de piruvato deshidrogenasa (PDH), con la siguiente estequiometría de reacción global:



30 En el citosol, la acetil-CoA se sintetiza mediante la desviación de la piruvato deshidrogenasa (PDH), que implica las enzimas piruvato descarboxilasa (PDC), acetaldehído deshidrogenasa (ALD) y acetil-CoA sintetasa (ACS), con la siguiente estequiometría de reacción global:



35 El mutante negativo para piruvato-decarboxilasa (Pdc-) de la levadura *S. cerevisiae* no tiene una desviación de PDH funcional, y no puede crecer sobre medio mínimo con glucosa como única fuente de carbono debido a la incapacidad de suministrar (suficiente) acetil-CoA citosólica para el crecimiento (Flikweert et al., (1996) Yeast 12:247-57). La desviación de PDH es, por tanto, esencial en proporcionar acetil-CoA en el compartimento citosólico. Sin embargo, la desviación de PDH en levadura no es óptima con respecto al equilibrio de energía, como puede apreciarse de la estequiometría de reacción global: se necesitan 2 moles de ATP por acetil-CoA sintetizada mediante la desviación de PDH ya que en la reacción de acetil-CoA sintetasa se hidroliza ATP a AMP. A diferencia, la vía mitocondrial mediante PDH no requiere ATP. El requisito de ATP adicional de la desviación de PDH puede presentar un problema para sintetizar el producto de interés a partir del precursor de acetil-CoA citosólica, ya que necesita que más fuente de carbono sea desviada para la generación de ATP mediante, por ejemplo, fosforilación oxidativa y/o fosforilación de sustrato (por ejemplo, glicólisis), reduciendo así el rendimiento global del producto en carbono.

45 Cuando la levadura se manipula metabólicamente para producir 1-butanol, pueden expresarse genes biosintéticos heterólogos de 1-butanol en el citosol en células de levadura (documento WO 2007/041269). En general, 1 mol de glucosa da lugar a 2 moles de acetil-CoA mediante glicólisis, que es el precursor de 1 mol de butanol; por lo tanto, puede sintetizarse un máximo de 1 mol de butanol por mol de glucosa si no se considera crecimiento celular y mantenimiento. Sin embargo, cuando se usa la desviación de PDH en combinación con la biosíntesis de butanol, este rendimiento teórico máximo no puede lograrse debido al desequilibrio de energía: mientras que se generan 2 moles de ATP por mol de glucosa convertida en glicólisis, se necesitan un total de 4 moles (2 veces 2 moles) de ATP en la desviación de PDH para formar 2 moles de acetil-CoA, que se convierten en 1 mol de butanol. Así, hay una escasez neta de ATP si se usa la desviación de PDH para sintetizar 1 mol de 1-butanol a partir de 1 mol de glucosa.

Así, existe la necesidad de la identificación de posibles vías metabólicas alternativas para producir acetil-CoA citosólica en levadura, para la producción de productos derivados de acetil-CoA, en particular butanol, en la que no se requiere la desviación de PDH.

5 El butanol es un producto químico industrial importante y es adecuado como combustible para motores alternativo que tiene propiedades mejoradas con respecto al etanol. El butanol también encuentra uso como disolvente para una amplia variedad de procesos químicos y textiles, en la síntesis orgánica de plásticos, como producto intermedio químico y como disolvente en la industria de los recubrimientos y de alimentos y aromas. El butanol puede producirse a partir de biomasa (biobutanol), además de combustibles fósiles (petrobutanol).

10 La síntesis química del butanol en uno de sus isómeros puede llevarse a cabo mediante una variedad de métodos disponibles conocidos en la técnica (véase, por ejemplo, Ullmann's Encyclopedia of Industrial Chemistry, 6ª edición, 2003, Wiley-VCH Verlag GmbH y Co., Weinheim, Alemania, Vol. 5, pp. 716-719). Estos procesos tienen la desventaja de que se basan en el uso de derivados petroquímicos, son generalmente caros y no son respetuosos con el medioambiente.

15 La síntesis biológica del butanol puede lograrse por fermentación usando el proceso de acetona-butanol-etanol (ABE) llevado a cabo por las bacterias *Clostridium acetobutylicum* u otras especies de *Clostridium*. Una desventaja importante del proceso de ABE, sin embargo, es que produce una mezcla de acetona, 1-butanol y etanol. Además, el uso de bacterias requiere condiciones de proceso estériles y generalmente hace que el proceso sea susceptible a infección por bacteriófagos. Así, las células de levadura proporcionan una alternativa muy adecuada como se ha descrito anteriormente.

20 **Sumario de la invención**

Los presentes inventores han identificado ahora vías metabólicas alternativas para aumentar la producción de acetil-CoA citosólica en levadura que puedan vencer los problemas de la desviación de PDH.

25 Una posible vía incluye la conversión directa de acetaldehído en acetil-CoA sin consumo de ATP, usando una acetaldehído deshidrogenasa acetilante (E.C. 1.2.1.10) (véase la Figura 2, reacción A, ACDH). Otra vía incluye la conversión directa de piruvato en acetil-CoA por una enzima o un complejo multienzimático sin consumo de ATP, por ejemplo, por uso de una piruvato:NADP oxidorreductasa (E.C. 1.2.1.51) véase la Figura 2, reacción C, PNO). En estas dos posibles vías, la formación de 1 mol de butanol por mol de glucosa produciría la formación de 2 moles de ATP. Otra vía más incluye la conversión de acetato en acetil-CoA con 1 ATP consumido por acetil-CoA formada por una enzima alternativa o una combinación de enzimas, por ejemplo, por uso de acetato:CoA ligasa (formadora de ADP, E.C. 6.2.1.13), o por uso de ATP:acetato fosfotransferasa (E.C. 2.7.2.1) en combinación con acetil-CoA:Pi acetiltransferasa (E.C. 2.3.1.8). En esta vía, la formación de 1 mol de butanol por mol de glucosa está en equilibrio con ATP, es decir, no se formará ATP. Los presentes inventores han encontrado ahora que una alternativa tal a la desviación de PDH puede producir la síntesis de acetil-CoA en el citosol de la levadura, y que tal acetil-CoA puede usarse biosintéticamente para producir mayores cantidades de productos de fermentación deseables, tales como butanol.

En un primer aspecto, la presente invención proporciona un método de identificación de un polipéptido heterólogo que tiene actividad enzimática para convertir piruvato, o acetaldehído en acetil-CoA en una única etapa de conversión en (el citosol de) una célula de levadura que comprende:

- 40 - proporcionar una célula de levadura mutada, en la que dicha mutación comprende una inactivación de al menos un gen de la desviación de piruvato deshidrogenasa (PDH), seleccionado de los genes que codifican las enzimas piruvato descarboxilasa (PDC; E.C. 4.1.1.1), acetaldehído deshidrogenasa (ALDY y acetil-CoA sintetasa (ACS; E.C. 6.2.1.1);
- 45 - transformar dicha célula de levadura mutada con un vector de expresión que comprende al menos una secuencia de nucleótidos heteróloga operativamente unida a un promotor funcional en levadura y teniendo dicha al menos una secuencia de nucleótidos heteróloga que codifica al menos un polipéptido candidato posible actividad enzimática para convertir piruvato o acetaldehído en acetil-CoA;
- probar dicha célula de levadura mutada recombinante para su capacidad para crecer sobre medio mínimo que contiene glucosa como única fuente de carbono, e
- 50 - identificar dicho polipéptido candidato como un polipéptido heterólogo que tiene actividad enzimática para convertir piruvato o acetaldehído en acetil-CoA en una única etapa de conversión en (el citosol de) dicha célula de levadura cuando se observa el crecimiento de dicha célula.

En una realización preferida de dicho método, la célula de levadura es una célula de *Saccharomyces cerevisiae* y la secuencia de nucleótidos heteróloga está optimizada en (el par de) codones para la expresión en *Saccharomyces cerevisiae*.

En otra realización preferida, dicha mutación comprende una inactivación del gen para acetil-CoA sintetasa isoforma 2 (acs2). En otra realización preferida, dicho al menos un polipéptido candidato que tiene actividad enzimática para convertir acetaldehído en acetil-CoA es una (supuesta) acetaldéhido deshidrogenasa acetilante.

5 En otro aspecto, la presente invención proporciona un vector de integración para la integración en un genoma de levadura de una secuencia de nucleótidos heteróloga que codifica un polipéptido que tiene actividad enzimática para convertir piruvato, acetaldehído o acetato en acetil-CoA, y la posterior expresión del polipéptido heterólogo del mismo.

10 En otro aspecto, la presente invención proporciona un vector de expresión que expresa polipéptidos heterólogos en levadura, comprendiendo dicho vector de expresión una secuencia de nucleótidos heteróloga operativamente unida a un promotor funcional en levadura y codificando dicha secuencia de nucleótidos heteróloga un polipéptido que tiene actividad enzimática para convertir piruvato, acetaldehído o acetato en acetil-CoA en (el citosol de) dicha célula de levadura.

En otra realización preferida, dicho polipéptido para convertir acetaldehído en acetil-CoA en una única etapa de conversión está seleccionado de SEQ ID NO: 19, 22, 25, 28 y 52 y homólogos funcionales de las mismas.

15 En otra realización preferida, dicho vector de expresión es para la expresión en *Saccharomyces cerevisiae*, en el que dicha secuencia de nucleótidos heteróloga está optimizada en (el par de) codones para la expresión en *Saccharomyces cerevisiae*.

En otra realización preferida, dicha secuencia de nucleótidos heteróloga está seleccionada de SEQ ID NO: 20, 23, 26 y 29.

20 En otro aspecto, la presente invención proporciona una célula de levadura recombinante que comprende una secuencia de nucleótidos heteróloga que codifica un polipéptido que tiene actividad enzimática para la conversión directa de acetaldehído en acetil-CoA en (el citosol de) dicha célula de levadura, en la que el polipéptido es una acetaldéhido deshidrogenasa (EC 1.2.1.10) acetilante, en la que dicho polipéptido tiene más del 70 %, 80 %, 90 % o 95 % de identidad de secuencia con la secuencia de aminoácidos seleccionada de SEQ ID NO: 19, 22, 25, 28 y 52.

25 En una realización preferida, la célula de levadura recombinante comprende además una inactivación de al menos un gen de la desviación de (PDH), seleccionado de los genes que codifican las enzimas piruvato descarboxilasa (PDC), acetaldéhido deshidrogenasa (ALD) y acetil-CoA sintetasa (ACS).

Preferentemente, una célula de levadura según la presente invención comprende una inactivación de un gen que codifica una acetil-CoA sintasa

30 En otra realización preferida, la célula de levadura recombinante comprende además una inactivación de un gen (secuencia de nucleótidos) que codifica una enzima capaz de catalizar la conversión de acetaldehído en etanol, preferentemente un gen que codifica una alcohol deshidrogenasa.

35 Como se usa en el presente documento, la inactivación de un gen (secuencia de nucleótidos) que codifica una enzima puede lograrse por mutación, delección o alteración de (parte de) un gen o secuencia de nucleótidos que codifica una enzima.

Preferentemente, una célula de levadura según la presente invención muestra crecimiento sobre medio mínimo que contiene glucosa como única fuente de carbono.

40 En otra realización preferida de una célula de levadura de la invención, dicha célula de levadura comprende además uno o más genes introducidos que codifican una vía recombinante para la formación de 1-butanol a partir de acetil-CoA citosólica. Vías recombinantes adecuadas de acetil-CoA a 1-butanol se conocen en la técnica. Tales vías son, por ejemplo, conocidas del documento WO 2007/041269. Preferentemente, dicho uno o más genes introducidos codifican enzimas que producen acetoacetil-CoA, 3-hidroxi-butiril-CoA, crotonil-CoA, butiril-CoA, butilaldehído y/o 1-butanol. Dichas enzimas pueden ser:

- 45 - acetil-CoA acetiltransferasa (E.C. 2.3.1.9 [Enzyme Nomenclature 1992, Academic Press, San Diego]; aunque también serán funcionales enzimas con un intervalo de sustrato más ancho (E.C. 2.3.1.16)), que convierte 2 moles de acetil-CoA en acetoacetil-CoA;
- 3-hidroxi-butiril-CoA deshidrogenasa dependiente de NADH o dependiente de NADPH (E.C. 1.1.1.35 o E.C. 1.1.1.30, resp. E.C. 1.1.1.157 o E.C. 1.1.1.36), que convierte acetoacetil-CoA en 3-hidroxi-butiril-CoA;
- 50 - 3-hidroxi-butiril-CoA deshidratasa (también llamada crotonasa; E.C. 4.2.1.17 o E.C. 4.2.1.55), que convierte 3-hidroxi-butiril-CoA en crotonil-CoA;
- butiril-CoA deshidrogenasa dependiente de NADH o dependiente de NADPH (E.C. 1.3.1.44 resp. E.C. 1.3.1.38 o E.C. 1.3.99.2), que convierte crotonil-CoA en butiril-CoA;

- aldehído deshidrogenasa monofuncional dependiente de NADH o dependiente de NADPH (E.C. 1.2.1.10, o 1.2.1.57), que convierte butiril-CoA en butiraldehído, y
 - butanol deshidrogenasa dependiente de NADH o dependiente de NADPH (E.C. 1.1.1.-), que convierte butiraldehído en 1-butanol, o
- 5 - aldehído/alcohol deshidrogenasa bifuncional dependiente de NADH o dependiente de NADPH (E.C. 1.1.1.1/1.2.1.10), que convierte butiril-CoA en 1-butanol mediante butiraldehído

En otra realización preferida de la invención, una célula de levadura es una *Saccharomyces cerevisiae*.

En otro aspecto, la presente invención proporciona un método de producción de butanol, que comprende las etapas de fermentar un sustrato de carbono adecuado con una célula de levadura según la presente invención y recuperar el butanol producido durante dicha fermentación.

Breve descripción de los dibujos

La Figura 1 es una presentación esquemática de la desviación de PDH que muestran las enzimas piruvato descarboxilasa (PDC; E.C. 4.1.1.1), acetaldehído deshidrogenasa (ALD; E.C. 1.2.1.3, E.C. 1.2.1.4 y E.C. 1.2.1.5) y acetil-CoA sintetasa (ACS; E.C. 6.2.1.1).

15 La Figura 2 muestra una vía metabólica esquemática para la producción de butanol en *Saccharomyces cerevisiae*. Las reacciones 1-6 son las etapas de biosíntesis de butanol a partir de *Clostridium acetobutylicum* introducido en levadura. A, B y C indican reacciones alternativas para la biosíntesis de acetil-CoA en el citosol. B indica parte de la desviación de piruvato deshidrogenasa (pdc, aid y acs), la fuente natural de acetil-CoA citosólica en levadura. Glc, glucosa; EtOH, etanol; Pir, Piruvato; AA, acetaldehído; ACT, acetato; AcCoA, acetil-CoA; AACoA, acetoacetil-CoA; BuCoA, butiril-CoA; Bual, butiraldehído; BuOH, butanol; NAD(P)(H), nicotinamida adenina dinucleótido (fosfato) (en forma reducida); ATP, adenosín trifosfato; AMP, adenosín monofosfato; ciclo de TCA, ciclo de ácido tricarbónico; PDH, piruvato deshidrogenasa; pdc, piruvato descarboxilasa; adh, alcohol deshidrogenasa; acdh, acetaldehído deshidrogenasa acetilante; aid, acetaldehído deshidrogenasa; acs, acetil-CoA sintetasa; pno, piruvato:NADP oxidoreductasa. Las conversiones enzimáticas indicadas mediante la reacción 1-6 indican una vía de butanol heteróloga de *Clostridium acetobutylicum*: thlB (o ThL) que codifica acetil-CoA acetiltransferasa o tiolasa [E.C. 2.3.1.9] (SEQ ID NO: 30); hbd, 3-hidroxi-butiril-CoA deshidrogenasa [E.C. 1.1.1.157] (SEQ ID NO: 31); crt, 3-hidroxi-butiril-CoA deshidratasa [E.C. 4.2.1.55] (SEQ ID NO: 32); ter, trans-enoil CoA reductasa; bed, butiril-CoA deshidrogenasa [E.C. 1.3.99.2] (SEQ ID NO: 33); etf $\alpha\beta$, flavoproteína heterodimérica de transferencia electrónica (etf α y etf β , SEQ ID NO: 38 y SEQ ID NO: 39, respectivamente); adhE/adhE1, aldehído/alcohol deshidrogenasa E y E1 [E.C. 1.1.1.1/1.2.1.10] (SEQ ID NO: 34 y 35, respectivamente); bdhA/bdhB, butanol deshidrogenasa A y B dependiente de NAD(P)H [E.C.:1.1.1.-] (SEQ ID NO: 36 y 37, respectivamente).

La Figura 3 muestra el mapa del plásmido YEplac112PtdhTadh. La secuencia de este plásmido se proporciona en la SEQ ID NO: 40.

La Figura 4 muestra un ejemplo de un árbol de similitud basado en secuencias de aminoácidos de proteínas de los tipos 1 a 4 como se describe en el Ejemplo 2 e indica las ramificaciones.

Descripción detallada de la invención

Definiciones

El término "butanol" se refiere a n-butanol, o 1-butanol.

El término "levadura" se refiere a un grupo filogenéticamente diverso de hongos unicelulares, la mayoría de los cuales están en la división de Ascomycota y Basidiomycota. Las levaduras de gemación ("levaduras verdaderas") se clasifican en el orden Saccharomycetales, con *Saccharomyces cerevisiae* como la especie más conocida.

El término "levadura recombinante", como se usa en el presente documento, se define como una célula que contiene una secuencia de nucleótidos y/o proteína, o se transforma o modifica genéticamente con una secuencia de nucleótidos que no se produce naturalmente en la levadura, o contiene copia o copias adicionales de una secuencia de ácidos nucleicos endógena (o proteínas), o contiene una mutación, delección o alteración de una secuencia de ácidos nucleicos endógena.

El término "mutado", como se usa en el presente documento con respecto a proteínas o polipéptidos, significa que al menos un aminoácido en la secuencia de proteínas o de polipéptidos no mutante o que existe de forma natural se ha sustituido con un aminoácido diferente, o delecionado de la secuencia mediante mutagénesis de ácidos nucleicos que codifican estos aminoácidos. La mutagénesis es un método muy conocido en la materia, e incluye, por ejemplo, mutagénesis dirigida al sitio por medio de PCR o mediante mutagénesis mediada por oligonucleótidos como se describe en Sambrook et al., *Molecular Cloning-A Laboratory Manual*, 2ª ed., Vol. 1-3 (1989). El término "mutado", como se usa en el presente documento con respecto a genes, significa que al menos un nucleótido en la secuencia de nucleótidos de ese gen o una secuencia reguladora del mismo se ha sustituido con un nucleótido diferente, o se

ha delecionado de la secuencia mediante mutagénesis, produciendo la transcripción de una secuencia de proteínas no funcional o la inactivación de ese gen.

5 El término "gen", como se usa en el presente documento, se refiere a una secuencia de ácidos nucleicos que contiene un molde para una ácido nucleico polimerasa, en eucariotas, ARN polimerasa II. Los genes se transcriben en ARNm que luego se traducen en proteína.

10 El término desviación de piruvato deshidrogenasa (PDH) se refiere a la cascada enzimática de piruvato a acetil-CoA en el citosol de la levadura, y que consiste en las siguientes enzimas: piruvato descarboxilasa (PDC; E.C. 4.1.1.1) que convierte piruvato en acetaldehído; acetaldehído deshidrogenasa (ALD; E.C. 1.2.1.3, E.C. 1.2.1.4 y E.C. 1.2.1.5), que convierte acetaldehído en acetato; y acetil-CoA sintetasa (ACS; E.C. 6.2.1.1), que convierte acetato en acetil-CoA.

15 El término "ácido nucleico", como se usa en el presente documento, incluye referencia a un polímero de desoxirribonucleótido o ribonucleótido, es decir, un polinucleótido, en tanto forma mono como bicatenaria, y a menos que se limite de otro modo, engloba análogos conocidos que tienen la naturaleza esencial de nucleótidos naturales en los que se hibridan con ácidos nucleicos monocatenarios de un modo similar a nucleótidos que existen de forma natural (por ejemplo, ácidos nucleicos peptídicos). Un polinucleótido puede ser de longitud completa o una subsecuencia de un gen estructural o regulador nativo o heterólogo. A menos que se indique lo contrario, el término incluye referencia a la secuencia especificada, además de la secuencia complementaria de la misma. Así, ADN's o ARN's con esqueletos modificados por estabilidad o por otros motivos son "polinucleótidos" como el término está previsto en el presente documento. Además, los ADN's o ARN's que comprenden bases poco usuales, tales como inosina, o bases modificadas, tales como bases tritiladas, solo por nombrar dos ejemplos, son polinucleótidos como el término se usa en el presente documento. Se apreciará que se ha hecho una gran variedad de modificaciones al ADN y ARN que sirven a muchos fines útiles conocidos para aquellos expertos en la materia. El término polinucleótido, como se emplea en el presente documento, engloba tales formas químicamente, enzimáticamente o metabólicamente modificadas de polinucleótidos, además de las formas químicas de ADN y ARN características de virus y células, que incluyen, entre otras cosas, células simples y complejas.

20 Los términos "polipéptido", "péptido" y "proteína" se usan indistintamente en el presente documento para referirse a un polímero de restos de aminoácidos. Los términos se aplican a polímeros de aminoácido en los que uno o más restos de aminoácidos es un análogo químico artificial de un aminoácido que existe de forma natural correspondiente, además de a polímeros de aminoácido que existen de forma natural. La naturaleza esencial de tales análogos de aminoácidos que existen de forma natural es que, cuando se incorporan en una proteína, esa proteína es específicamente reactiva con anticuerpos provocados para la misma proteína, pero que consisten enteramente en aminoácidos que existen de forma natural. Los términos "polipéptido", "péptido" y "proteína" también son incluyentes de modificaciones que incluyen, pero no se limitan a, glucosilación, unión de lípidos, sulfatación, gamma-carboxilación de restos de ácido glutámico, hidroxilación y ribosilación de ADP.

30 La identidad de secuencia se define en el presente documento como una relación entre dos o más secuencias de aminoácidos (polipéptidos o proteínas) o dos o más secuencias de ácidos nucleicos (polinucleótidos), como se determina comparando las secuencias. Normalmente, las identidades de secuencia se comparan con respecto a la longitud completa de las secuencias comparadas. En la materia, "identidad" también significa el grado de vinculación de secuencias entre las secuencias de aminoácidos o de ácidos nucleicos, según sea el caso, como se ha determinado por la correspondencia entre hebras de tales secuencias.

35 Métodos preferidos de determinación de la identidad se diseñan para dar la mayor correspondencia entre las secuencias probadas. Métodos de determinación de la identidad están codificados en programas informáticos públicamente disponibles. Métodos de programas informáticos preferidos para determinar la identidad y similitud entre dos secuencias incluyen BLASTP, BLASTN (Altschul, S. F. et al., J. Mol. Biol. 215:403-410 (1990), públicamente disponible de NCBI y otras fuentes (BLAST Manual, Altschul, S., et al., NCBI NLM NIH Bethesda, MD 20894). Parámetros preferidos para la comparación de secuencias de aminoácidos usando BLASTP son hueco abierto 11,0, extensión por hueco 1, matriz Blosum 62.

40 Cada secuencia de ácidos nucleicos en el presente documento que codifica un polipéptido, por tanto, por referencia al código genético, describe cada posible variación silenciosa del ácido nucleico. El término "variantes conservativamente modificadas" se aplica a tanto secuencias de aminoácidos como de ácidos nucleicos. Con respecto a secuencias de ácidos nucleicos particulares, variantes conservativamente modificadas se refiere a aquellos ácidos nucleicos que codifican variantes idénticas o conservativamente modificadas de las secuencias de aminoácidos debido a la degeneración del código genético. El término "degeneración del código genético" se refiere al hecho de que un gran número de ácidos nucleicos funcionalmente idénticos codifican cualquier proteína dada. Por ejemplo, los codones GCA, GCC, GCG y GCU todos codifican el aminoácido alanina. Así, en cada posición donde una alanina esté especificada por un codón, el codón puede alterarse a cualquiera de los codones correspondientes descritos sin alterar el polipéptido codificado. Tales variaciones de ácido nucleico son "variaciones silenciosas" y representan una especie de variación conservativamente modificada.

"Expresión" se refiere a la transcripción de un gen en ARN estructural (ARNr, ARNt) o ARN mensajero (ARNm) con posterior traducción en una proteína.

5 Como se usa en el presente documento, "heterólogo" en referencia a un ácido nucleico o proteína es un ácido nucleico o proteína que se origina a partir de una especie extraña o, si es de la misma especie, está sustancialmente modificada de su forma nativa en composición y/o locus genómico por intervención humana deliberada. Por ejemplo, un promotor operativamente unido a un gen estructural heterólogo es de una especie diferente de aquél del gen estructural del que se derivó, o, si es de la misma especie, uno o ambos están sustancialmente modificados de su forma original. Una proteína heteróloga puede originarse a partir de una especie extraña o, si es de la misma especie, está sustancialmente modificada de forma original por intervención humana deliberada.

10 Como se usa en el presente documento, "promotor" es una secuencia de ADN que dirige la transcripción de un gen (estructural). Normalmente, un promotor está localizado en la región 5' de un gen, proximal al sitio de inicio transcripcional de un gen (estructural). Las secuencias promotoras pueden ser constitutivas, inducibles o represibles. Si un promotor es un promotor inducible, entonces la tasa de transcripción aumenta en respuesta a un agente inductor.

15 El término "vector", como se usa en el presente documento, incluye referencia a un vector de expresión autosómico y a un vector de integración usado para la integración en el cromosoma.

20 El término "vector de expresión" se refiere a una molécula de ADN, lineal o circular, que comprende un segmento que codifica un polipéptido de interés bajo el control de (es decir, operativamente unido a) segmentos de ácidos nucleicos adicionales que proporcionan su transcripción. Tales segmentos adicionales pueden incluir secuencias de promotor y de terminador, y pueden opcionalmente incluir uno o más orígenes de replicación, uno o más marcadores de selección, un potenciador, una señal de poliadenilación, y similares. Los vectores de expresión se derivan generalmente de ADN de plásmido o viral, o pueden contener elementos de ambos. En particular, un vector de expresión comprende una secuencia de nucleótidos que comprende en la dirección 5' a 3' y operativamente unida a:
25 (a) una región de transcripción y de iniciación de la traducción reconocida por levadura, (b) una secuencia codificante de un polipéptido de interés, y (c) una región de transcripción y de terminación de la traducción reconocida por levadura. "Plásmido" se refiere a ADN extracromosómico que se replica autónomamente que no está integrado en un genoma de microorganismo y es normalmente circular en la naturaleza.

30 Un "vector de integración" se refiere a una molécula de ADN, lineal o circular, que puede incorporarse en el genoma de un microorganismo y proporciona herencia estable de un gen que codifica un polipéptido de interés. El vector de integración generalmente comprende uno o más segmentos que comprenden una secuencia de gen que codifica un polipéptido de interés bajo el control de (es decir, operativamente unido a) segmentos de ácidos nucleicos adicionales que proporcionan su transcripción. Tales segmentos adicionales pueden incluir secuencias de promotor y de terminador, y uno o más segmentos que conducen a la incorporación del gen de interés en el genoma de la célula diana, normalmente por el proceso de recombinación homóloga. Normalmente, el vector de integración será
35 uno que puede transferirse en la célula diana, pero que tiene un replicón que no es funcional en ese organismo. La integración del segmento que comprende el gen de interés puede seleccionarse si se incluye un marcador apropiado dentro de este segmento.

40 Como se usa en el presente documento, el término "operativamente unido" se refiere a una yuxtaposición en la que los componentes así descritos están en una relación que les permite funcionar de su manera prevista. Una secuencia de control "operativamente unida" a otra secuencia de control y/o a una secuencia codificante se une de tal forma que la transcripción y/o expresión de la secuencia codificante se logre en condiciones compatibles con la secuencia de control. Generalmente, operativamente unido significa que las secuencias de ácidos nucleicos que están unidas son contiguas y, donde sea necesario unir dos regiones codificantes de proteína, contiguas y en el mismo marco de lectura.

45 Por "célula huésped" se indica una célula que contiene un vector y soporta la replicación y/o expresión del vector. Las células huésped pueden ser células procariontas tales como *E. coli*, o células eucariotas tales como levadura, células de insecto, anfibio, o de mamífero. Preferentemente, las células huésped son células del orden de los Actinomycetales, lo más preferentemente células de levadura, lo más preferentemente células de *Saccharomyces cerevisiae*.

50 "Transformación" y "transformar", como se usan en el presente documento, se refieren a la inserción de un polinucleótido exógeno en una célula huésped, independientemente del método usado para la inserción, por ejemplo, captación directa, transducción, apareamiento f o electroporación. El polinucleótido exógeno puede mantenerse como un vector no integrado, por ejemplo, un plásmido, o alternativamente, puede integrarse en el genoma de la célula huésped.

55 El término "oligonucleótido" se refiere a una secuencia corta de monómeros de nucleótido (normalmente 6 a 100 nucleótidos) unidos por enlaces de fósforo (por ejemplo, fosfodiéster, alquil y aril-fosfato, fosforotioato, fosfotiéster), o enlaces no de fósforo (por ejemplo, péptido, sulfamato y otros). Un oligonucleótido puede contener nucleótidos modificados que tienen bases modificadas (por ejemplo, 5-metilcitosina) y grupos de azúcar modificados (por

ejemplo, 2'-O-metilribosilo, 2'-O-metoxietilribosilo, 2'-fluororribosilo, 2'-aminorribosilo, y similares). Los oligonucleótidos pueden ser moléculas que existen de forma natural o sintéticas de ADN bi- y monocatenario y ARN bi- y monocatenario con formas circulares, ramificadas o lineales y que opcionalmente incluyen dominios capaces de formar estructuras secundarias estables (por ejemplo, estructuras de tallo y bucle y bucle-tallo-bucle).

- 5 El término "polinucleótido", como se usa en el presente documento, se refiere a una forma polimérica de nucleótidos de cualquier longitud, tanto ribonucleótidos como desoxirribonucleótidos. Así, este término incluye ADN y ARN bi- y monocatenario.

El término "polinucleótido recombinante", como se usa en el presente documento, se destina a un polinucleótido de origen genómico, de ADNc, semisintético o sintético, que, en virtud de su origen o manipulación: (1) no está asociado a todo o una porción de un polinucleótido con el que está asociado en la naturaleza; o (2) está unido a un polinucleótido distinto de aquel con el que está unido en la naturaleza; o (3) no se produce en la naturaleza.

El término "medio mínimo", como se usa en el presente documento, se refiere a un medio químicamente definido, que incluye solo los nutrientes que son requeridos por las células para sobrevivir y proliferar en cultivo. Normalmente, el medio mínimo está libre de extractos biológicos, por ejemplo, factores de crecimiento, suero, extracto pituitario, u otras sustancias, que no son necesarias para soportar la supervivencia y proliferación de una población de células en cultivo. Por ejemplo, el medio mínimo generalmente incluye como sustancias esenciales: al menos una fuente de carbono, tal como glucosa; al menos una fuente de nitrógeno, tal como amonio, sulfato de amonio, cloruro de amonio, nitrato de amonio o urea; sales inorgánicas, tales como hidrogenofosfato de dipotasio, dihidrogenofosfato de potasio y sulfato de magnesio; y otros nutrientes, tales como biotina y vitaminas.

20 **Descripción de las realizaciones preferidas**

Un método de la presente invención proporciona un método de identificación de enzimas heterólogas capaces de producir acetil-CoA en el citosol de una célula de levadura. La enzima heteróloga puede producir la acetil-CoA usando piruvato o acetaldehído como sustrato en una única etapa de conversión. Preferentemente, la enzima heteróloga produce la acetil-CoA a partir de acetaldehído. Una enzima capaz de catalizar dicha reacción es la acetil-CoA deshidrogenasa (acdh; E.C. 1.2.1.10) acetilante también denominada acetil-CoA:NAD⁺ oxidoreductasa (CoA acetilante). La conversión de acetaldehído en acetil-CoA por la acetil-CoA deshidrogenasa acetilante es reversible y transcurre en la dirección de acetil-CoA cuando el acetaldehído se acumula en el citosol. Una acumulación tal puede, por ejemplo, lograrse por delección de alcohol deshidrogenasa (adh; E.C. 1.1.1.1).

La enzima heteróloga puede también producir la acetil-CoA a partir de piruvato. Una enzima capaz de catalizar dicha reacción es una piruvato:NADP oxidoreductasa (pno; E.C. 1.2.1.51). La reacción es estequiométricamente idéntica a la de la piruvato deshidrogenasa mitocondrial, excepto que *pno* usa NADPH como cofactor en comparación con PDH que usa NADH. En comparación con *acdh*, una desventaja importante del sistema enzimático de *pno* es que *pno* es sensible al oxígeno, y que es una enzima multimérica grande, y por lo tanto, su incorporación genética satisfactoria (un gen de 5-6 kb) es mucho más difícil que la de *acdh*. Por este motivo, se prefiere el uso de *acdh* en las realizaciones de la presente invención.

Una característica importante de una célula de prueba capaz de revelar la actividad enzimática deseada de un polipéptido de prueba es que la célula es prototrófica como resultado del polipéptido introducido. Con esto se indica que los requisitos nutricionales de la célula no superan a aquellos de la cepa no mutante correspondiente y que proliferará sobre medio mínimo (a diferencia del auxótrofo). En realidad, la producción de acetil-CoA como se soporta por el polipéptido de prueba cancelará el efecto de la delección de dicho al menos un gen de la desviación de PDH, producido por la delección del gen para la piruvato descarboxilasa (*pdh*; E.C. 4.1.1.1), acetil-CoA deshidrogenasa (*ald*; E.C. 1.2.1.3, E.C. 1.2.1.4 o E.C. 1.2.1.5) o acetil-CoA sintetasa (*acs*; E.C. 6.2.1.1). Tales ensayos de complementación son muy conocidos en la técnica. En aspectos de la presente invención, el ensayo se usa para identificar fuentes de enzimas heterólogas adecuadas capaces de sostener la producción de acetil-CoA citosólica en células de levadura.

El ensayo de complementación se basa en la provisión de vías alternativas para vencer la actividad enzimática delecionada de la desviación de PDH. Métodos para efectuar la delección de genes en levadura son muy conocidos en la técnica, y pueden, por ejemplo, lograrse por mutagénesis mediada por oligonucleótidos. Pueden obtenerse buenos resultados con el plásmido pUG6 que lleva el casete de alteración del gen *loxP*-kanMX-*loxP* (Güldener et al. [1996] Nucleic Acids Res. 24(13):2519-24; acceso de GenPept N.º P30114). Así, el experto será capaz de proporcionar una cepa de levadura que tiene un gen acetil-CoA deshidrogenasa y/o acetil-CoA sintetasa delecionado para bloquear la desviación de PDH en él.

Saccharomyces cerevisiae comprende dos isoformas de acetil-CoA sintetasa, *Acs1p* y *Acs2p*. Ambas son la fuente nuclear de acetil-CoA para la acetilación de histonas. La producción de acetil-CoA citosólica también se requiere para la producción de lípidos. La actividad de *Acs* es esencial, ya que un mutante nulo doble de *acs1 acs2* no es viable. Un mutante nulo de *acs1* puede crecer en etanol como la única fuente de carbono. La célula de levadura mutada usada en aspectos de la presente invención tiene preferentemente una inactivación del gen *acs2*.

Los mutantes de *Saccharomyces cerevisiae* que llevan una inactivación del gen *acs2* no son capaces de crecer sobre glucosa como única fuente de carbono, debido a que ACS1 es reprimido y la proteína se degrada activamente. La complementación de un mutante de delta *acs2* tal con un gen *acs* basado en plásmido restaurará la capacidad de la célula para crecer sobre glucosa como única fuente de carbono. Además, el crecimiento de un mutante tal se complementa por la expresión de genes que soportan vías alternativas para la producción de suficiente acetil-CoA citosólica. Así, la transformación del mutante de delta *acs2* con un plásmido del que puede expresarse un *acdh* o *pno* (heterólogo) funcional restaurará la capacidad del mutante para crecer sobre glucosa como única fuente de carbono. Debe entenderse que además de la eliminación del locus ACS2, también puede eliminarse el locus ACS1. Aunque se cree que esto puede en algunos casos prevenir la aparición de revertantes (mutaciones en el locus ACS1 que conducen a la reversión del fenotipo de delta *acs2*), sin embargo, no se encontró que esto fuera esencial. Mutantes dobles (cepas *acs1/acs2Δ*) serían completamente dependientes del gen *acdh* o *pno* introducido para la producción de acetil-CoA citosólica.

Una ventaja importante de un ensayo de complementación de la presente invención es que puede realizarse como ensayo de cribado en placa en el que se observa la complementación satisfactoria como crecimiento de colonias. Esto es mucho más rápido que los experimentos que requieren el análisis para la producción de un producto metabólico deseado.

Para la complementación de la mutación, la célula de levadura que tiene el gen *aid* y/o *acs* inactivado se transforma entonces con un vector de expresión adecuado que comprende una secuencia de nucleótidos de un polipéptido de prueba heterólogo.

Están ampliamente disponibles vectores de expresión en levadura de una variedad de proveedores comerciales. Hasta la fecha, la complementación funcional de mutaciones de levadura por homólogos extraños se ha convertido en una práctica estándar en la manipulación de *Saccharomyces cerevisiae*. Vectores de expresión adecuados para la expresión de genes heterólogos pueden basarse en promotores inducibles artificiales tales como el promotor GAL, pero se basan preferentemente en promotores constitutivos tales como el promotor *TDH3*. Sistemas adecuados se ejemplifican en los ejemplos más adelante. En ciertos sistemas de producción puede preferirse el uso de un promotor inducible, ya que permitiría la separación temporal de etapas para la producción de biomasa (promotor no inducido) y producción de productos de fermentación (promotor inducido). En otra realización altamente preferida en ciertos sistemas de producción, el vector está en el vector de integración para la integración estable de los genes heterólogos en el genoma de la cepa de producción de levadura.

Con el fin de lograr la expresión óptima en levadura, puede optimizarse el uso de codones (par) del gen heterólogo usando una cualquiera de una variedad de paquetes de software de diseño de genes sintéticos, por ejemplo, GeneOptimizer® de Geneart AG (Regensburg, Alemania) para la optimización del uso de codones o la optimización del uso del par de codones como se describe en el documento WO2008/000632. Tal adaptación del uso de codones garantiza que los genes heterólogos, que son, por ejemplo, de origen bacteriano, sean eficazmente procesados por la maquinaria de transcripción y traducción de levadura. La optimización del uso del par de codones producirá expresión potenciada de proteínas en la célula de levadura.

Las secuencias optimizadas pueden, por ejemplo, clonarse en un plásmido de expresión de levadura de copias altas, unido operativamente a un promotor (preferentemente constitutivo) funcional en levadura. Se han obtenido buenos resultados con el plásmido YEplac112 (2 μ TRP1) (Gietz & Sugino [1988] Gene 74(2):527-34).

Genes heterólogos que codifican un polipéptido candidato que tiene posible actividad enzimática para convertir piruvato o acetaldehído en acetil-CoA pueden identificarse por ordenador. Enzimas adecuadas descritas por poseer la capacidad de convertir acetaldehído en acetil-CoA son acetaldehído deshidrogenasas acetilantes (E.C. 1.2.1.10). Las secuencias de nucleótidos y de aminoácidos de más de 200 de estas enzimas de una variedad de orígenes microbianos se describen en diversas bases de datos (por ejemplo, la base de datos KEGG (Kyoto Encyclopedia of Genes and Genomes)).

Los presentes inventores han seleccionado varias acetaldehído deshidrogenasas acetilantes y probado éstas en el sistema de ensayo basado en el mutante delta *acs2* de la presente invención. Muchas de estas, aunque no todas, fueron funcionales en *S. cerevisiae* cuando se optimizó el uso del par de codones.

También pueden usarse homólogos funcionales para estas proteínas en aspectos de la presente invención. El término "homólogos funcionales", como se usa en el presente documento, se refiere a una proteína que comprende la secuencia de aminoácidos de SEQ ID NO: 19, 22, 25 o la parte de acetaldehído deshidrogenasa de SEQ ID NOs: 28 y 52 en la que uno o más aminoácidos están sustituidos, delecionados, añadidos y/o insertados, y esa proteína tiene la misma funcionalidad enzimática para la conversión de sustrato, por ejemplo, un homólogo de acetaldehído deshidrogenasa acetilante es capaz de convertir acetaldehído en acetil-CoA. Esta funcionalidad puede probarse por el uso de un sistema de ensayo que comprende una célula de levadura recombinante que comprende un vector de expresión para la expresión del homólogo en levadura, comprendiendo dicho vector de expresión una secuencia de nucleótidos heteróloga operativamente unida a un promotor funcional en levadura y codificando dicha secuencia de nucleótidos heteróloga el polipéptido homólogo del cual va a probarse la actividad enzimática para convertir piruvato, acetaldehído o acetato en acetil CoA en (el citosol de) dicha célula de levadura, y realizar un método de

identificación de un polipéptido heterólogo que tiene actividad enzimática para convertir piruvato, acetaldehído o acetato en acetyl-CoA en (el citosol de) una célula de levadura como se describe en el presente documento usando dicho sistema de ensayo. Homólogos candidatos pueden identificarse usando análisis de similitud por ordenador. Un ejemplo detallado de un análisis tal se describe en el Ejemplo 2 más adelante. El experto será capaz de derivar del mismo cómo pueden encontrarse homólogos candidatos adecuados y, opcionalmente tras la optimización de (el par de) codones, serán capaces de probar la funcionalidad requerida de tales homólogos candidatos usando el sistema de ensayo de la presente invención como se ha descrito anteriormente. Un homólogo adecuado representa un polipéptido que tiene una identidad de secuencias de aminoácidos con una acetaldeshidrogenasa acetilante superior al 50 %, preferentemente superior al 60 %, más preferentemente superior al 70 %, 80 %, 90 % o más, por ejemplo, que tiene una identidad de secuencias de aminoácidos tal con SEQ ID NOs:19, 22, 25, o la parte de acetaldeshidrogenasa de SEQ ID NOs:28 y 52 y que tiene la funcionalidad enzimática requerida para convertir acetaldeshidrogenasa en acetyl-CoA. Similarmente, también pueden usarse enzimas descritas para la conversión directa de piruvato en acetyl-CoA y los homólogos funcionales de las mismas, además de enzimas descritas para la conversión de acetato en acetyl-CoA y los homólogos funcionales de las mismas, similar a como se ha descrito para la acetaldeshidrogenasa acetilante anteriormente.

Un método de la presente invención comprende además la etapa de probar la capacidad de la célula de levadura mutada y transformada con la proteína de prueba para crecer sobre medio mínimo que contiene glucosa como única fuente de carbono. Como se ha establecido anteriormente, esto puede producirse adecuadamente sobre medio sólido (agar) en platos (placas) de Petri donde el crecimiento puede observarse como el crecimiento de una colonia, sin embargo, medios líquidos son igualmente adecuados y el crecimiento puede detectarse por turbidez. También pueden usarse otros métodos de determinación del crecimiento de la célula de levadura mutada y transformada con la proteína de prueba sobre medio mínimo que contiene glucosa como única fuente de carbono.

Cuando la célula de levadura mutada y transformada con la proteína de prueba es capaz de crecer sobre medio mínimo con glucosa, el polipéptido candidato se identifica satisfactoriamente como un polipéptido heterólogo que tiene actividad enzimática para convertir piruvato, acetaldeshidrogenasa o acetato en acetyl-CoA en (el citosol de) dicha célula de levadura. El crecimiento puede observarse adecuadamente como la formación de colonias sobre medios sólidos de crecimiento, en particular medio mínimo que contiene glucosa.

Un vector de expresión para la expresión de polipéptidos heterólogos en levadura, según la presente invención, puede ser cualquier vector de expresión adecuado para transformar levadura. Están disponibles en la materia innumerables ejemplos que pueden usarse adecuadamente para expresar secuencias de nucleótidos heterólogos en levadura. Un vector muy adecuado en los aspectos de la invención es un plásmido. Un plásmido altamente preferido es YEplac112PtdhTadh (SEQ ID NO: 40).

Generalmente, la secuencia de nucleótidos heteróloga que codifica el polipéptido que tiene actividad enzimática para convertir piruvato o acetaldeshidrogenasa en acetyl-CoA en (el citosol de) dicha célula de levadura se pondrá bajo el control de un promotor funcional en levadura. Preferentemente, el promotor es un promotor constitutivo. El promotor en el plásmido YEplac112PtdhTadh es el promotor *TDH3*.

Las secuencias de nucleótidos heterólogas incorporadas en el vector de expresión de la presente invención pueden ser cualquier enzima pno, acdh u otra enzima capaz de convertir piruvato o acetaldeshidrogenasa (respectivamente) en acetyl-CoA en el citosol de la levadura. Secuencias de nucleótidos preferidas son aquellas como se identificaron en el presente documento, concretamente las secuencias de nucleótidos que codifican:

- la proteína de utilización de etanolamina EutE de HS de *E. coli* (secuencias de nucleótidos con SEQ ID NO: 18);
- la proteína hipotética Lin1129 de *Listeria innocua* similar a la proteína de utilización de etanolamina EutE (secuencias de nucleótidos con SEQ ID NO: 21)
- la acetaldeshidrogenasa EDK33116 de DSM 555 de *Clostridium kluyveri* (secuencias de nucleótidos con SEQ ID NO: 24); y
- el homólogo de adhE de *S. aureus* (secuencias de nucleótidos con SEQ ID NO: 27) que codifica una acetaldeshidrogenasa bifuncional en *Staphylococcus aureus* subsp. *aureus* N315, o la parte funcional de acetaldeshidrogenasa de la misma
- el homólogo de adhE de *Piromyces* sp. E2 (secuencia de nucleótidos SEQ ID NO: 51) que codifica una acetaldeshidrogenasa bifuncional, o la parte de acetaldeshidrogenasa de la misma.

También son adecuados homólogos funcionales de estas secuencias de nucleótidos, o de los polipéptidos que codifican. Con este término se indica una secuencia de ácidos nucleicos que tiene más del 80 %, 90 % o 95 % de identidad de secuencia con las secuencias de nucleótidos que codifican las enzimas acdh anteriores, o que tiene más del 50 %, preferentemente más del 60 %, 70 %, 80 %, 90 %, o 95 % de identidad de secuencia con la secuencia de aminoácidos de las enzimas acdh anteriores, con la condición de que los polipéptidos codificados por las secuencias homólogas presenten actividad enzimática de acdh funcional.

Como se ha establecido anteriormente, estas secuencias de nucleótidos pueden optimizarse para la expresión en *Saccharomyces cerevisiae* por optimización del uso del par de codones muy conocida en la técnica. Secuencias optimizadas en el par de codones para SEQ ID NO: 18, 21, 24 y 27 se proporcionan en SEQ ID NO: 20, 23, 26 y 29, respectivamente.

- 5 El vector de expresión de la invención puede usarse para transformar una célula de levadura. Métodos de transformación incluyen electroporación, transformación en perlas de vidrio y biolística, todos los cuales son muy conocidos en la técnica y, por ejemplo, se describen en Sambrook et al., *Molecular Cloning-A Laboratory Manual*, 2ª ed., Vol. 1-3 (1989).

10 Una célula de levadura según la presente invención comprende una secuencia de nucleótidos heteróloga que codifica un polipéptido que tiene actividad enzimática para convertir piruvato o acetaldehído en acetil-CoA en (el citosol de) dicha célula de levadura. Preferentemente, una célula de levadura de la invención comprende un *acdh* o *pno* heterólogo. La ventaja de una célula de levadura tal es que puede producir acetil-CoA por una vía metabólica en la que no se requiere la desviación de PDH. Esto es energéticamente más favorable en condiciones anaerobias, y puede formar la base de cualquier proceso de síntesis biológica usando células de levadura en condiciones anaerobias donde la acetil-CoA es un producto intermedio. Además de comprender *acdh* o *pno* heterólogo, la célula de levadura de la invención puede comprender diversas delecciones de genes o suplementaciones de genes, dependiendo del uso previsto de la levadura.

20 Preferentemente, una célula de levadura según la presente invención comprende una inactivación de una secuencia de nucleótidos (gen) que codifica una enzima capaz de catalizar la conversión de acetaldehído en etanol, preferentemente una alcohol deshidrogenasa, por ejemplo, para optimizar la acumulación de acetaldehído en la célula de levadura.

25 Si se usa en un método de cribado de enzimas heterólogas según un método de la invención, la célula de levadura comprende una delección de al menos un gen de la desviación (PDH), seleccionado de los genes que codifican las enzimas piruvato descarboxilasa (PDC), acetaldehído deshidrogenasa (ALD) y acetil-CoA sintetasa (ACS), preferentemente acetil-CoA sintetasa, lo más preferentemente *acs2*.

30 Si se usa en un método de producción de un producto de fermentación, la célula de levadura puede comprender opcionalmente varias suplementaciones de genes (heterólogos) que soportan la vía metabólica de acetil-CoA a dicho butanol. Una vía tal puede consistir solo de productos del gen heterólogo, o puede hacer uso de una mezcla de productos de genes heterólogos y endógenos. En el caso de que el producto de fermentación sea butanol, puede hacerse uso de una levadura que comprende genes que codifican enzimas para la vía de butanol de, por ejemplo, *Clostridium acetobutylicum* como se describe en el presente documento y en la Figura 2. En el caso de que la célula de levadura según la presente invención comprenda genes que codifican enzimas para la producción de butanol, la levadura comprende preferentemente una secuencia de nucleótidos que codifica una butiril-CoA deshidrogenasa y al menos una secuencia de nucleótidos que codifica una flavoproteína heteróloga de transferencia de electrones (ETF).
35 Se encontró que una célula de levadura que comprende una ETF, además de genes de la vía del butanol, produce una elevada cantidad de butanol.

40 Una flavoproteína heteróloga de transferencia de electrones en la célula eucariota según la presente invención puede ser una única proteína o la ETF puede comprender dos o más subunidades, por ejemplo, una subunidad alfa y una beta. Preferentemente, la ETF comprende una ETF alfa (SEQ ID NO: 38) y una ETF beta (SEQ ID NO: 39). La flavoproteína de transferencia de electrones puede derivarse de cualquier origen adecuado. Preferentemente, la ETF se deriva del mismo origen que la butiril-CoA deshidrogenasa. Preferentemente, la ETF se deriva de origen procariota, preferentemente de un *Clostridium sp.*, preferentemente un *Clostridium acetobutylicum* o un *Clostridium beijerinckii*.

45 Un método de producción de un producto de fermentación según la presente invención comprende preferentemente cultivar una levadura en condiciones anaerobias sobre una fuente de carbono y de energía adecuada. Fuentes adecuadas de carbono y de energía son azúcares C5 y C6 (monosacáridos) tales como glucosa y polisacáridos tales como almidón. También son adecuados otros materiales de partida tales como caña de azúcar, maíz, trigo, cebada, remolachas azucareras, semilla de colza y girasol. En algunos casos, el material de partida puede ser previamente digerido por tratamiento enzimático. Lo más preferentemente, la fuente de carbono es lignocelulosa, que está compuesta por principalmente celulosa, hemicelulosa, pectina y lignina. La lignocelulosa se encuentra, por ejemplo, en los tallos, hojas, cáscaras, cortezas y mazorcas de las plantas. La hidrólisis de estos polímeros por tratamiento enzimático específico libera una mezcla de azúcares neutros que incluyen glucosa, xilosa, manosa, galactosa y arabinosa. Pueden usarse materiales lignocelulósicos, tales como madera, material herbáceo, residuos agrícolas, fibra de maíz, papel usado, residuos de la molienda de la pulpa y el papel para producir butanol. Enzimas hidrolizantes son, por ejemplo, glucanos unidos en beta para la hidrólisis de celulosa (estas enzimas incluyen endoglucanasas, celobiohidrolasas, glucohidrolasas y beta-glucosidasas); las beta-glucosidasas hidrolizan celobiosa; las hemicelulasas que actúan en endo y exo para la hidrólisis de hemicelulosa, y acetilsterasas y esterazas que hidrolizan enlaces glucósido de la lignina. Estos y otros métodos para la hidrólisis de lignocelulosa son muy conocidos en la técnica.

Son posibles variaciones y modificaciones de las realizaciones desveladas en el presente documento, y alternativas prácticas a y equivalentes de los diversos elementos de las realizaciones serían entendidos por aquellos expertos habituales en la materia tras el estudio de este documento de patente. Estas y otras variaciones y modificaciones de las realizaciones desveladas en el presente documento pueden hacerse sin apartarse del alcance y espíritu de la invención.

La invención se ilustrará ahora a modo de los siguientes ejemplos no limitantes.

EJEMPLOS

Los siguientes ejemplos ilustran la provisión de una cepa de *Saccharomyces cerevisiae* útil en los ensayos y métodos de la presente invención, por ejemplo, en métodos de identificación de enzimas heterólogas capaces de formar acetil-CoA citosólica en *S. cerevisiae*. Tales métodos son útiles en la identificación de vías/enzimas que permiten el suministro citosólico de acetil-CoA en *S. cerevisiae* en condiciones anaerobias.

Con el fin de potenciar la formación de acetil-CoA citosólica en la cepa de producción de butanol de los presentes inventores, se configuró un método de selección para identificar enzimas heterólogas que formaban acetil-CoA citosólica en *S. cerevisiae*. El sistema de prueba se basa en un mutante de levadura delta *acs2* deficiente en la biosíntesis de acetil-CoA citosólica en glucosa, una cepa tal es incapaz de crecer sobre glucosa como única fuente de carbono a menos que se complemente la formación de acetil-CoA citosólica. Estudios de complementación en una cepa tal pueden revelar qué enzimas heterólogas son adecuadas para su uso en cepas productoras de butanol de *Saccharomyces cerevisiae*.

Se identificó acetaldehído deshidrogenasa acetilante por ser un buen candidato para el suministro de acetil-CoA citosólica con respecto a la desviación de PDH homóloga debido a que no se disipa ATP. Se sintetizaron doce supuestas acetaldehído deshidrogenasas acetilantes, identificadas basándose en homología de secuencias, y se comprobaron para la complementación de la levadura delta *acs2*.

Los genes optimizados en el par de codones de los homólogos de *eutE* de *E. coli*, *L. innocua* y *C. kluyveri* y el homólogo de *adhE* de *S. aureus* fueron capaces de complementar los mutantes de levadura *acs2* (4 de los 7), produciendo el crecimiento del huésped de *S. cerevisiae acs2Δ*. El objetivo es mejorar la biosíntesis de butanol en levadura por la expresión de uno o más genes así identificados.

Con el fin de probar si estas vías heterólogas para el suministro de acetil-CoA citosólica funcionan en *S. cerevisiae*, se desarrolló un sistema de cribado basado en mutantes de *Saccharomyces cerevisiae* que llevan una delección del gen *acs2*. Estas células no son capaces de crecer sobre glucosa como única fuente de carbono, a menos que el mutante delta *acs2* se complemente con un gen *acs* basado en plásmido o complementado con la expresión de cualquier otro gen que genera suficiente acetil-CoA citosólica. Así, si fuera a transformarse con un plásmido que conduce a la expresión activa de *acdH* o *pno*, un mutante tal debe ser capaz de crecer otra vez con glucosa como única fuente de carbono. Los estudios de complementación se realizaron en placas. Los siguientes experimentos se realizaron para configurar y evaluar el sistema de prueba.

Ejemplo 1

Construcción de la cepa delta *acs2*

Se produjo la cepa delecionada en *acs2* de *S. cerevisiae* (cepa *acs2Δ*) realizando primero una PCR en el plásmido pUG6 (Güldener et al., 1996, arriba) con los siguiente oligonucleótidos:

5'*acs2*::Kanlox 5'-tacacaacagaatacaggaaagtaaatcaatacaataataaaacagctgaagcttctgtacgc-3'

3'*acs2*::Kanlox 5'-tctcattacgaaattttctcatttaagtattttctttttgaggcataggccactagtgatctg-3'.

Se usó el fragmento de 1,4 kb resultante, que contenía el marcador KanMX que confiere resistencia a G418, para transformar CEN.PK113-3C de *S. cerevisiae* (*MATA trp1-289*). Después de la transformación, la cepa se sembró en YPD (10 g l⁻¹ de extracto de levadura (BD Difco), 20 g l⁻¹ de peptona (BD Difco)), 10 g l⁻¹ de glucosa) con 200 mg/ml de genética (G418). En transformantes resistentes, la integración correcta se verificó por PCR usando los oligonucleótidos:

5'ACS2: 5'-gatattcggtagccgattcc-3'

3'ACS2: 5'-ccgtaaccttctgtaatgc-3'

ACS2internal: 5'-cggattcgtcatcagcttca-3'

KanA: 5'-cgcacgtcaagactgtcaag-3'

KanB: 5'-tcgtatgtgaatgctggtcg-3'

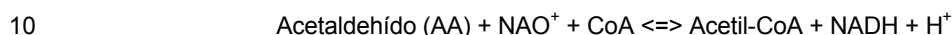
El fenotipo se verificó probando el crecimiento sobre YP con 1 % de glucosa (YPD) o 1 % de etanol+ 1 % de glicerol (YPEG) como fuente de carbono.

Se recogió un transformante que tenía las bandas de PCR correctas y que no creció en YP con glucosa, pero creció en YP con etanol y glicerol como fuentes de carbono, y se llamó RWB060 (*MATA trp1-289 acs2::Kanlox*).

5 Ejemplo 2

Identificación por ordenador de supuestas acetaldehído deshidrogenasas acetilantes para la conversión directa de acetaldehído en acetil-CoA

Las enzimas descritas para la conversión de acetaldehído en acetil-CoA son las llamadas acetaldehído deshidrogenasas acetilantes (ACDH) (E.C. 1.2.1.10) que catalizan la siguiente reacción:



A partir de la bibliografía se han descrito cuatro tipos de proteínas que tienen esta actividad:

15 1) Proteínas bifuncionales que catalizan la conversión reversible de acetil-CoA en acetaldehído, y la posterior conversión reversible de acetaldehído en etanol. Un ejemplo de este tipo de proteínas es la proteína AdhE en *E. coli* (GenBank No: NP_415757). AdhE parece ser el producto evolutivo de una fusión génica. La región del extremo NH₂ de la proteína AdhE es altamente homóloga a las aldehído:NAD⁺ oxidorreductasas, mientras que la región del extremo COOH es homóloga a una familia de etanol:NAD⁺ oxidorreductasas dependientes de Fe²⁺ (Membrillo-Hernández et al., (2000) J. Biol. Chem. 275: 33869-33875). AdhE de *E. coli* se somete a oxidación catalizada por metal y, por tanto, es sensible a oxígeno (Tamarit et al. (1998) J. Biol. Chem. 273:3027-32).

20 2) Proteínas que catalizan la conversión reversible de acetil-CoA en acetaldehído en microorganismos estrictamente anaerobios o facultativos, pero no poseen actividad de alcohol deshidrogenasa. Un ejemplo de este tipo de proteínas se ha informado en *Clostridium kluyveri* (Smith et al. (1980) Arch. Biochem. Biophys. 203: 663-675). Se ha anotado una acetaldehído deshidrogenasa acetilante en el genoma de *Clostridium kluyveri* DSM 555 (GenBank N.º: EDK33116). Se identifica una proteína AcdH homóloga en el genoma de *Lactobacillus plantarum* (GenBank No: NP_784141). Otro ejemplo de este tipo de proteínas es el producto génico *ald* en *Clostridium beijerinckii* NRRL B593 (Toth et al. (1999) Appl. Environ. Microbiol. 65: 4973-4980, GenBank No: AAD31841).

30 3) Proteínas que participan en el catabolismo de la etanolamina. La etanolamina puede ser utilizada tanto como fuente de carbono como de nitrógeno por muchas enterobacterias (Stojiljkovic et al. (1995) J. Bacteriol. 177: 1357-1366). La etanolamina se convierte primero por la etanolamina amoniaco liasa en amoniaco y acetaldehído, posteriormente, el acetaldehído se convierte por acetaldehído deshidrogenasa acetilante en acetil-CoA. Un ejemplo de este tipo de acetaldehído deshidrogenasa acetilante es la proteína EutE en *Salmonella typhimurium* (Stojiljkovic et al. (1995) J. Bacteriol. 177: 1357-1366, GenBank No: AAL21357). *E. coli* también es capaz de utilizar etanolamina (Scarlett et al. (1976) J. Gen. Microbiol. 95:173-176) y tiene una proteína EutE (GenBank No: AAG57564) que es homóloga a la proteína EutE en *S. typhimurium*.

40 4) Proteínas que son parte de un complejo bifuncional de aldolasa-deshidrogenasa implicado en el catabolismo de 4-hidroxi-2-cetovalerato. Tales enzimas bifuncionales catalizan las dos etapas finales de la vía de meta-escisión para el catecol, un producto intermedio en muchas especies bacterianas en la degradación de fenoles, toluatos, naftaleno, bifenilos y otros compuestos aromáticos (Powlowski y Shingler (1994) Biodegradation 5, 219-236). El 4-hidroxi-2-cetovalerato es primero convertido por la 4-hidroxi-2-cetovalerato aldolasa en piruvato y acetaldehído, posteriormente el acetaldehído es convertido por la acetaldehído deshidrogenasa acetilante en acetil-CoA. Un ejemplo de este tipo de acetaldehído deshidrogenasa acetilante es la proteína DmpF en CF600 de *Pseudomonas sp.* (GenBank No: CAA43226) (Shingler et al. (1992) J. Bacteriol. 174:711-24). *E. coli* tiene una proteína MphF homóloga (Ferrandez et al. (1997) J. Bacteriol. 179: 2573-2581, GenBank No: NP_414885) a la proteína DmpF en CF600 de *Pseudomonas sp.*

50 Para identificar los miembros de la familia de las proteínas de la acetaldehído deshidrogenasa acetilante, las secuencias de aminoácidos de la proteína AdhE bifuncional de *E. coli* (GenBank No: NP_415757), la proteína AcdH de *L. plantarum* (acetilante) (GenBank No: NP_784141), la proteína EutE de *E. coli* (GenBank No: AAG57564) y la proteína MphF de *E. coli* (GenBank No: NP_414885) se ejecutaron cada uno como una secuencia de búsqueda en una búsqueda de BLASTp contra la base de datos de proteínas no redundantes GenBank usando parámetros por defecto. Se extrajeron secuencias de aminoácidos con un valor de E más pequeño o igual a 1e-20. Se eliminaron secuencias redundantes y las secuencias restantes se alinearon y se construyó un árbol de similitud usando el software analizador de proteínas Genedata Physolopher, versión 6.5.2. Un árbol de similitud proporciona información sobre la similitud de secuencias del organismo. El árbol se crea independientemente del algoritmo ClustalW por comparación por parejas de las secuencias de aminoácidos por posición de residuo. En cada posición, la similitud se

evalúa y se suma hasta una puntuación global para cada par de secuencias. Basándose en estas puntuaciones por parejas se realiza una agrupación jerárquica, que dispone las secuencias en un árbol. Obsérvese que el producto génico *ald* de *C. beijerinckii* (GenBank no: AAD31841) se agrupó junto con las proteínas EutE de *E. coli* y *S. typhimurium*. A partir de este árbol de similitud podrían definirse cuatro ramas principales, cada rama contiene una secuencia de aminoácidos que se usó como búsqueda para la búsqueda de BLASTp. La Figura 4 muestra un ejemplo de un árbol de similitud tal, que contiene todas las secuencias que se mencionan en este ejemplo.

Se seleccionó al menos una secuencia de aminoácidos de cada rama para las pruebas de complementación en delta *acs2* de *S. cerevisiae*. Preferentemente, las secuencias de aminoácidos seleccionadas tienen evidencia experimental de su función bioquímica como la acetaldehído deshidrogenasa acetilante. Tales evidencias pueden encontrarse en bases de datos públicas, tales como en las bases de datos BRENDA, UniProt y NCBI Entrez.

Ejemplo 3

Construcción de plásmidos de expresión y prueba de complementación

Para probar si las acetaldehído deshidrogenasas acetilantes (ACDH) podrían complementar la delección de *ACS2* en *S. cerevisiae*, se eligieron varios genes que codifican una (supuesta) ACDH de una variedad de bases de datos como se ha descrito anteriormente.

Para lograr la expresión óptima en levadura, el uso de codones de todos los genes se adaptó por optimización del par de codones. Estas secuencias se sintetizaron en Geneart AG (Regensburg, Alemania).

Las secuencias optimizadas se clonaron en el plásmido de expresión de levadura de copias altas YEplac112PtdhTadh (SEQ ID NO: 40; basado en YEplac112 (2 μ TRP1) (Gietz & Sugino [1988] Gene 74(2):527-34), permitiendo la expresión constitutiva del promotor *TDH3*.

Se preparó YEplac112PtdhTadh clonando un fragmento *KpnI-SacI* de p426GPD (Mumberg et al. [1995] Gene. 156(1):119-22), que contenía el promotor *TDH3* y el terminador *CYC1*, en YEplac112 cortado con *KpnI-SacI*. El plásmido resultante se cortó con *KpnI* y *SphI* y los extremos se hicieron romos, luego se unieron dando YEplac112TDH. Para obtener YEplac112PtdhTadh, YEplac112TDH se cortó con *PstI-HindIII* y se unió a un fragmento de PCR *PstI-HindIII* de 345 pb que contenía el terminador *ADH1* (Tadh), sustituyendo así al terminador *CYC1* y cambiando el policonector entre el promotor y el terminador. Se generó el fragmento de PCR Tadh usando los siguientes oligonucleótidos:

MCS-5'Tadh: 5'-aaggtacctagactagtagccgggctgcagtcgactcgagcgaattcttatgatttatgatt-3'

Tadh1-Hind: 5'-aggaagcttaggcctgtgtggaagaacgattacaacagg-3'

Se hizo PCR con la ADN polimerasa Vent^R, según las especificaciones del fabricante.

Las construcciones sintéticas que contenían los genes ACDH se cortaron con *SpeI-PstI* y se unieron en YEplac112PtdhTadh digerido con las mismas enzimas, produciendo pBOL058 a pBOL068 y pBOL082. Los nombres de los plásmidos finales y los genes que contienen se dan en la Tabla 1.

Tabla 1: Visión general de las supuestas acetaldehído deshidrogenasas acetilantes probadas para la complementación de la cepa de *S. cerevisiae* delta *acs2*. Los genes que produjeron complementación se dan en negrita. Se proporcionan SEQ ID NOs para la secuencia de ADN del gen no mutante, la proteína expresada a partir del mismo y la secuencia de ADN optimada en el par de codones.

Tabla 1

Organismos	Nombre	Grupo*	Tamaño (kb)	SEQ ID NO. ADN/PRT/OPT
<i>Escherichia coli</i>	<i>adhE</i>	1	2,6	
<i>Entamoeba histolytica</i>	<i>adh2</i>	1	2,6	48/50/49
<i>Staphylococcus aureus</i>	<i>adhE</i>	1	2,6	27 / 28 / 29
<i>Piromyces sp. E2</i>	<i>adhE</i>	1	2,6	51/52
<i>Clostridium kluyveri</i>	<i>EDK33116</i>	2	1,5	24 / 25 / 26
<i>Lactobacillus plantarum</i>	<i>acdH</i>	2	1,4	
<i>Escherichia coli</i>	<i>EutE</i>	3	1,4	18 / 19 / 20
<i>Listeria innocua</i>	<i>Lin1129</i>	3	1,4	21 / 22 / 23

Organismos	Nombre	Grupo*	Tamaño (kb)	SEQ ID NO. ADN/PRT/OPT
<i>Pseudomonas putida</i>	YP 001268189	4	1,0	

* Grupo se refiere al grupo de proteínas que tiene actividad de ACDH como se define en el Ejemplo 2. Grupo 1: similar a *AdhE* de *E. coli* bifuncional (tipo *AdhE* de proteínas); Grupo 2: proteínas que tienen similitud con *AcdH* de *Lactobacillus plantarum* (tipo *AcdH* de proteínas); Grupo 3: similar a *EutE* de *E. coli* (tipo *EutE* de proteínas); Grupo 4: similar a *MhpF* de *E. coli* (tipo *MhpF* de proteínas).

5 Todos los plásmidos se usaron para transformar la cepa de levadura de delta *acs2* RWB060. Como control negativo, se usó el vector vacío YEplac112. Los transformantes se sembraron sobre medio mineral (Verduyn et al. [1992] Yeast 8 (1992), pp. 501-517) que contenía tanto 1 % de glucosa (MYD) como 1 % de etanol + 1 % de glicerol (MYEG) como única fuente de carbono.

10 Aunque para todas las construcciones podrían seleccionarse varios transformantes sobre medio mínimo con etanol/glicerol, éste no fue el caso en las placas que contenían glucosa.

Tabla 2: Resultado de un experimento de complementación para supuestas acetaldehído deshidrogenasas acetilantes en la cepa de *S. cerevisiae* de delta *acs2* RWB060. Los genes que producen complementación se dan en negrita. Las columnas MYEG y MYD indican el número de transformantes en placas MYEG (etanol/glicerol) y MYG (glucosa).

Organismos	Gen (acceso de GenPept)	plásmido	MYEG	MYD
	<i>ninguno</i>	YEplac112	75	0
<i>Escherichia coli</i>	<i>adhE</i>	pBOL059	6	0
<i>Entamoeba histolytica</i>	<i>adh2</i>	pBOL061	54	0
<i>Staphylococcus aureus</i>	<i>adhE (BAB41363)</i>	pBOL064	36	39
<i>Piromyces sp. E2</i>	<i>adhE</i>	pBOL139	32	3
<i>Clostridium kluyveri</i>	<i>EDK33116(EDK33116)</i>	pBOL065	21	8
<i>Lactobacillus plantarum</i>	<i>acdH</i>	pBOL058	6	0
<i>Escherichia coli</i>	<i>EutE(ABV06849)</i>	pBOL066	24	18
<i>Listeria innocua</i>	<i>Lin1129(CAC96360)</i>	pBOL067	28	8
<i>Pseudomonas putida</i>	YP 001268189	pBOL068	32	0

15 En las placas que contienen glucosa, los transformantes solo pudieron seleccionarse para los plásmidos pBOL064, pBOL065, pBOL066 y pBOL067, no el vector vacío. También hubo una clara diferencia en el tamaño de colonias, dependiendo del plásmido usado. Mientras que la construcción pBOL066 (*eutE* de *E. coli*) produjo las mayores colonias, las colonias de pBOL067 (*lin1129* de *L. innocua*) parecieron un poco más pequeñas y pBOL065 (*edk3116* de *C. kluyveri*) mostró las colonias más pequeñas. El plásmido pBOL064 (*adhE* de *S. aureus*) y el plásmido pBOL139 (*Piromyces sp. E2, adhE*) se hicieron en una fecha posterior, por lo que no pudieron compararse directamente. Las colonias que contienen pBOL064 parecieron ser similares a las colonias que comprenden pBOL066 y las colonias que comprenden pBOL139 parecieron ser similares a las colonias que comprenden pBOL065.

20 Para garantizar que estos resultados no surgieron de revertantes espontáneos, se repitieron experimentos de transformación para algunos de los plásmidos, dando los mismos resultados. Además, para casi todos los plásmidos se seleccionaron cuatro transformantes al azar de las placas de MYEG y se volvieron a depositar en manchas sobre placas MYD y MYEG.

25 En todos los experimentos nunca se observó crecimiento sobre glucosa con el vector vacío (YEplac112), mientras que solo pBOL065, pBOL066 y pBOL067 dieron repetidamente buen crecimiento sobre glucosa. El plásmido pBOL064 no volvió a probarse de esta forma después del muy positivo resultado inicial.

30 A partir de estos resultados, se llegó a la conclusión de que los genes optimizados en el par de codones de los homólogos de *eutE* de:

- *E. coli* (SEQ ID NO: 20) que codifica la proteína de utilización de etanolamina *EutE* de HS de *E. coli*;

- *L. innocua* (SEQ ID NO: 23) que codifica una proteína hipotética de *L. innocua* similar a la proteína de utilización de etanolamina EutE, y
- *C. kluyveri* (SEQ ID NO: 26) que codifica acetaldehído deshidrogenasa acetilante en DSM 555 de *Clostridium kluyveri*; y el gen optimizado en los pares de codones del homólogo de adhE de
- 5 - *S. aureus* (SEQ ID NO: 29) que codifica una acetaldehído/alcohol deshidrogenasa bifuncional en *Staphylococcus aureus* subsp. *aureus* N315;
y el gen no optimizado en los pares de codones del homólogo de adhE
- *Piromyces* sp. E2 (SEQ ID NO: 51) que codifica una acetaldehído/alcohol deshidrogenasa bifuncional

10 son capaces de complementar los mutantes de levadura *acs2*. Estos genes codifican una actividad enzimática que permite la formación de acetil-CoA citosólica a partir de acetaldehído en levadura.

Conclusiones

Se cree que el suministro de acetil-CoA citosólica es un cuello de botella en la producción de butanol en levadura. Con el fin de identificar genes heterólogos que codifican enzimas que forman acetil-CoA citosólica en *S. cerevisiae*, se estableció un sistema de prueba basado en un mutante de levadura delta *acs2*.

15 Debido a su deficiencia en biosíntesis de acetil-CoA citosólica sobre glucosa, la cepa *acs2Δ* es incapaz de crecer con glucosa como única fuente de carbono.

20 Se expresaron 9 supuestas acetaldehído deshidrogenasas acetilantes identificadas como candidatos para el suministro de acetil-CoA citosólica a partir de acetaldehído en la levadura *acs2Δ*. En total, 5 de estos 9 genes complementaron el crecimiento de la cepa *acs2Δ* con glucosa como única fuente de carbono. Con esto, se mostró el uso de la cepa delta *acs2* como herramienta de pre-selección para vías factibles para el suministro citosólico de acetil-CoA.

25 4 de las 5 acetaldehído deshidrogenasas acetilantes identificadas hasta la fecha, homólogos de *eutE* de *E. coli*, *L. innocua* y *C. kluyveri* y el homólogo de *adhE* de *S. aureus*, y E2 de *Piromyces* sp., se integraron satisfactoriamente en cepas productoras de butanol de *S. cerevisiae*. El efecto sobre la producción de butanol se investigó como se describe en los ejemplos más adelante.

Este sistema de prueba también puede usarse para analizar si la piruvato:NADP oxidoreductasa puede expresarse en exceso satisfactoriamente en levadura. Debido a la sensibilidad del oxígeno, esta prueba tiene que realizarse anaerobiamente.

30 Los Ejemplos 4-6 a continuación describen la prueba de 4 de los 5 genes ACDH seleccionados del Ejemplo 3 para la mejora de la producción de butanol.

Ejemplo 4

Construcción de una cepa de levadura productora de butanol e inactivación de los genes ADH1 y ADH2

35 Los seis genes de *Clostridium acetobutylicum* implicados en la biosíntesis de butanol a partir de acetil-CoA se enumeran en la Tabla 3. Los genes fueron optimizados en el par de codones para *S. cerevisiae* como se describe en el documento WO2008/000632 y se expresaron de promotores de levadura y terminadores como se enumera en la Tabla 3.

Se diseñaron dos vectores de integración de levadura (pBOL34 [SEQ ID NO: 41] y pBOL36 [SEQ ID NO: 42]), conteniendo cada uno 3 de los seis genes optimizados en el par de codones de *Clostridium acetobutylicum* implicados en la biosíntesis de butanol, y se sintetizaron en Geneart.

40 Los genes ThiL, Hbd y Crt se expresan a partir de pBOL34 que contiene un marcador de selección *AmdS*. Los tres genes finales, Bed, BdhB y AdhE, se expresaron a partir de un vector de integración con un marcador de selección *AmdS* llamado pBOL36.

Tabla 3: Genes usados para la producción de butanol en *S. cerevisiae* que incluye el promotor (1000 pb) y terminador (500 pb)

Gen	actividad	Promotor	Terminador
ThiL	acetil CoA c-acetiltransferasa [E.C. 2.3.1.9]	<i>ADH1</i>	<i>TDH1</i>
Hbd	3-hidroxibutiril-CoA deshidrogenasa [E.C. 1.1.1.157]	<i>ENO1</i>	<i>PMA1</i>
Crt	3-hidroxibutiril-CoA deshidratasa [E.C. 4.2.1.55]	<i>TDH1</i>	<i>ADH1</i>
Bcd	butiril-CoA deshidrogenasa [E.C. 1.3.99.2],	<i>PDC1</i>	<i>TDH1</i>
BdhB	butanol deshidrogenasa dependiente de NADH [E.C. 1.1.1.-],	<i>ENO1</i>	<i>PMA1</i>
<i>adhE</i>	alcohol/acetaldehído CoA deshidrogenasa [E.C.: 1.1.1.1/ 1.2.1.10]	<i>TDH1</i>	<i>ADH2</i>

5 Para la integración en el locus ADH2, se linealizó pBOL36 por una digestión con *Bsa*BI. Se transformó *S. cerevisiae* CEN.PK113-5D (MATa MAL2-8c SUC2 ura3-52) con el fragmento lineal y se cultivó en placas con YCB (Difco) y acetamida 5 mM como fuente de nitrógeno.

10 El marcador *AmdS* se eliminó por recombinación cultivando los transformantes durante 6 horas en YEPD en tubos de 2 ml a 30 °C. Las células se sembraron posteriormente sobre 1,8 % de medio de agar que contenía YCB (Difco) y fluoracetamida 40 mM y tampón fosfato 30 mM a pH 6,8 soportando el crecimiento solo a partir de células que han perdido el marcador *AmdS*. Se confirmaron la correcta integración y recombinación por PCR. La correcta integración del fragmento en la dirección 5' se confirmó con los siguientes cebadores:

P1: 5'-GAATTGAAGGATATCTACATCAAG-3' y

P2: 5'-CCCATCTACGGAACCCTGATCAAGC-3'.

La correcta integración del fragmento en la dirección 3' se confirmó con los siguientes cebadores:

15 P3: 5'-GATGGTGTACCATTACCAGGTCTAG-3' y

P4: 5'-GTTCTCTGGTCAAGTTGAAGTCCATTTTGATTGATTTGACTGTGTTATTTTGCCTG-3'.

La cepa resultante se llamó BLT021.

20 pBOL34 se linealizó por una digestión con *Psi*I y se integró en el locus ADH1 de BLT021. Los transformantes se cultivaron en placas que contenían YCB (Difco) y acetamida 5 mM. Para la eliminación del marcador de selección *AmdS*, las colonias se inocularon en YEPD y se cultivaron durante 6 horas en tubos de 2 ml a 30 °C. Las células se sembraron en YCB (Difco) y fluoracetamida 40 mM y 0,1 % de sulfato de amonio.

Se confirmaron la correcta integración y recombinación por PCR. La correcta integración del fragmento en la dirección 5' se confirmó con el siguiente conjunto de cebadores:

P5: 5'-GAACAATAGAGCGACCATGACCTTG-3' y

25 P6: 5'-GACATCAGCGTACCAGCCTTGATG-3'.

La correcta integración del fragmento en la dirección 3' se confirmó con el siguiente conjunto de cebadores:

P7: 5'-GATTGAAGGTTTCAAGAACAGGTGATG-3' y

P8: 5'-GGCGATCAGAGTTGAAAAAAAAAATG-3'.

La cepa resultante se llamó BLT057.

30 **Ejemplo 5**

Introducción de ETF α y ETF β en BLT057

Los genes ETF y los genes *Acdh* como se enumeran en la Tabla 4 se optimizaron en el par de codones para *S. cerevisiae* como se describe en el documento WO2008/000632 y se expresaron a partir de promotores y terminadores de levadura como se enumera en la Tabla 4.

35

Tabla 4: Promotores y terminadores usados para la expresión de genes ETF optimizados en el par de codones y genes *Acdh* en *S. cerevisiae*

	<i>Promotor</i>	<i>Terminador</i>
Etfa(CpO)	tef1	tdh2
Etfβ(CpO)	tdh2	tef1
Acdh64 (<i>AdhE S. aureus</i>)	tdh3	adh
Acdh65 (<i>Clostridium</i>)	tdh3	adh
Acdh66 (<i>EutE E. coli</i>)	tdh3	adh
Acdh67 (<i>lin1129 Ec</i>)	tdh3	Adh

5 Se sintetizaron los vectores de integración que expresan ETFα y ETFβ solo (pBOL113, [SEQ ID NO: 43]) o ETFα y ETFβ combinados con *Acdh64* (pBOL115, [SEQ ID NO: 44]), *Acdh65* (pBOL116, [SEQ ID NO: 45]), *Acdh66* (pBOL118, [SEQ ID NO: 46]) o *Acdh67* (pBOL120, [SEQ ID NO: 47]) por Geneart AG.

Los vectores, pBOL113, pBOL115, pBOL116, pBOL118 y pBOL120, se linealizaron con *StuI* y se integraron en el locus *ura3-52* de la cepa BLT057.

10 Los transformantes se cultivaron en YNB (Difco) sin aminoácidos + 2 % de galactosa para seleccionar cepas prototróficas de uracilo. Las cepas derivadas de la cepa BLT057 con pBOL113/115/116/118/120 integrados en el genoma se designaron las cepas: BLT071, BLT072, BLT073, BLT074 y BLT075, respectivamente.

Ejemplo 6

Producción de butanol mejorada expresando genes *Acdh* positivos

15 Las cepas BLT071 a BLT075 como se prepararon en el Ejemplo 5 se cultivaron en medio de Verduyn (Verduyn et al. (1992) *Yeast* 8: 501-517) en el que el sulfato de amonio está sustituido con 2 g/l de urea y que contiene además 4 % en peso de galactosa. Las células se cultivaron en matraces con agitación de 100 ml que contenían 50 ml de medio durante 72 horas a 30 °C a 180 rpm en un agitador rotatorio.

20 Se determinó la concentración de butanol en el sobrenadante del cultivo. Las muestras se analizaron en un HS-GC equipado con un detector de ionización de llama y un sistema de inyección automático. Columna J&W DB-1 longitud 30 m, id 0,53 mm, df 5 μm. Se usaron las siguientes condiciones: helio como gas portador con un caudal de 5 ml/min. La temperatura de la columna se estableció a 110 °C. El inyector se estableció a 140 °C y el detector funcionó a 300 °C. Los datos se obtuvieron usando el software Chromeleon. Las muestras se calentaron a 60 °C durante 20 min en el muestreador del espacio de cabeza. Se inyectó automáticamente (1) ml de los volátiles del espacio de cabeza en la columna.

25 La producción de 1-butanol de las diversas cepas fue la siguiente:

BLT057: 120 mg/l

BLT071: 450 mg/l

BLT072: 500 mg/l

BLT073: 600 mg/l

30 BLT074: 670 mg/l

BLT075: 700 mg/l

Los resultados muestran que la introducción de flavoproteínas de transferencia de electrones (ETF alfa y ETF beta) y/o la introducción de acetaldehído deshidrogenasas acetilantes como se identifica por un ensayo de complementación del Ejemplo 3 aumentan el nivel de producción de butanol.

35 LISTADO DE SECUENCIAS

<110> DSM IP ASSETS B.V.

Raamsdonk, L.M.

Mueller, U.

Wu, L.

Winkler A.A.

5 <120> Enzimas productoras de acetil CoA en levadura

<130> 26480WO

<140> 26480WO

10 <141> 2008-07-11

<160> 52

<170> PatentIn versión 3.3

15

<210> 1

<211> 64

<212> ADN

<213> Artificial

20

<220>

<223> cebador 5'acs2

<400> 1

25

atacacaac agaatacagg aaagtaaate aatacaataa taaaacagct gaagcttcgt 60

acgc 64

<210> 2

<211> 67

30

<212> ADN

<213> Artificial

<220>

<223> cebador 3'acs2

35

<400> 2

ES 2 607 894 T3

tctcattaag aaatTTTTct catttaagtt atttctTTTT ttgaggcata ggccactagt 60
ggatctg 67

<210> 3
<211> 20
5 <212> ADN
<213> Artificial

<220>
<223> sonda 5'acs2
10

<400> 3
gatattcggg agccgattcc 20

<210> 4
15 <211> 20
<212> ADN
<213> Artificial

<220>
20 <223> sonda 3'acs2

<400> 4
ccgtaacctt ctcgtaatgc 20

25 <210> 5
<211> 20
<212> ADN
<213> Artificial

30 <220>
<223> sonda ACS2internal

<400> 5
cggattcgtc atcagcttca 20
35

ES 2 607 894 T3

<210> 6
<211> 20
<212> ADN
<213> Artificial
5
<220>
<223> sonda KanA

<400> 6
10 cgcacgtcaa gactgtcaag 20

<210> 7
<211> 20
<212> ADN
15 <213> Artificial

<220>
<223> sonda KanB

20 <400> 7
tcgtatgtga atgctggtcg 20

<210> 8
<211> 64
25 <212> ADN
<213> Artificial

<220>
<223> cebador MCS-5'Tadh
30

<400> 8

aaggtacctc tagactagtc cggggtgca gtcgactcga gcgaatttct tatgatttat 60
gatt 64

35 <210> 9
<211> 39

<212> ADN

<213> Artificial

<220>

5 <223> cebador Tadh1-Hind

<400> 9

aggaagccta ggcctgtgtg gaagaacgat tacaacagg 39

10 <210> 10

<211> 24

<212> ADN

<213> Artificial

15 <220>

<223> cebador P1

<400> 10

gaattgaagg atatctacat caag 24

20

<210> 11

<211> 25

<212> ADN

<213> Artificial

25

<220>

<223> cebador P2

<400> 11

30 cccatctacg gaaccctgat caagc 25

<210> 12

<211> 26

<212> ADN

35 <213> Artificial

<220>

<223> cebador P3

<400> 12

gatggtgtca ccattaccag gtctag 26

5

<210> 13

<211> 56

<212> ADN

<213> Artificial

10

<220>

<223> cebador P4

<400> 13

15

gttctctggt caagttgaag tccatttga ttgatttgac tgtgttatt tgcgtg 56

<210> 14

<211> 25

<212> ADN

<213> Artificial

20

<220>

<223> cebador P5

<400> 14

25

gaacaataga gcgaccatga ccttg 25

<210> 15

<211> 25

<212> ADN

<213> Artificial

30

<220>

<223> cebador P6

35

<400> 15

gacatcagcg tcaccagcct tgatg 25

ES 2 607 894 T3

<210> 16
<211> 27
<212> ADN
<213> Artificial
5
<220>
<223> cebador P7

<400> 16
10 gattgaaggt ttcaagaaca ggtgatg 27

<210> 17
<211> 25
<212> ADN
15 <213> Artificial

<220>
<223> cebador P8

20 <400> 17
ggcgcacaga gttgaaaaaa aaatg 25

<210> 18
<211> 1404
25 <212> ADN
<213> Escherichia coli

<220>
< 221> CDS
30 < 222> (1)..(1404)

<400> 18

ES 2 607 894 T3

atg aat caa cag gat att gaa cag gtg gtg aaa gcg gta ctg ctg aaa	48
Met Asn Gln Gln Asp Ile Glu Gln Val Val Lys Ala Val Leu Leu Lys	
1 5 10 15	
atg caa agc agt gac acg ccg tcc gcc gcc gtt cat gag atg ggc gtt	96
Met Gln Ser Ser Asp Thr Pro Ser Ala Ala Val His Glu Met Gly Val	
20 25 30	
ttc gcg tcc ctg gat gac gcc gtt gcg gca gcc aaa gtc gcc cag caa	144
Phe Ala Ser Leu Asp Asp Ala Val Ala Ala Ala Lys Val Ala Gln Gln	
35 40 45	
ggg tta aaa agc gtg gca atg cgc cag tta gcc att gct gcc att cgt	192
Gly Leu Lys Ser Val Ala Met Arg Gln Leu Ala Ile Ala Ala Ile Arg	
50 55 60	
gaa gca ggc gaa aaa cac gcc aga gat tta gcg gaa ctt gcc gtc agt	240
Glu Ala Gly Glu Lys His Ala Arg Asp Leu Ala Glu Leu Ala Val Ser	
65 70 75 80	
gaa acc ggc atg ggg cgc gtt gaa gat aaa ttt gca aaa aac gtc gct	288
Glu Thr Gly Met Gly Arg Val Glu Asp Lys Phe Ala Lys Asn Val Ala	
85 90 95	
cag gcg cgc ggc aca cca ggc gtt gag tgc ctc tct ccg caa gtg ctg	336
Gln Ala Arg Gly Thr Pro Gly Val Glu Cys Leu Ser Pro Gln Val Leu	
100 105 110	
act ggc gac aac ggc ctg acc cta att gaa aac gca ccc tgg ggc gtg	384
Thr Gly Asp Asn Gly Leu Thr Leu Ile Glu Asn Ala Pro Trp Gly Val	
115 120 125	
gtg gct tcg gtg acg cct tcc act aac ccg gcg gca acc gta att aac	432
Val Ala Ser Val Thr Pro Ser Thr Asn Pro Ala Ala Thr Val Ile Asn	
130 135 140	
aac gcc atc agc ctg att gcc gcg ggc aac agc gtc att ttt gcc ccg	480
Asn Ala Ile Ser Leu Ile Ala Ala Gly Asn Ser Val Ile Phe Ala Pro	
145 150 155 160	
cat ccg gcg gcg aaa aaa gtc tcc cag cgg gcg att acg ctg ctc aac	528
His Pro Ala Ala Lys Lys Val Ser Gln Arg Ala Ile Thr Leu Leu Asn	
165 170 175	
cag gcg att gtt gcc gca ggt ggg ccg gaa aac tta ctg gtt act gtg	576

ES 2 607 894 T3

Gln	Ala	Ile	Val	Ala	Ala	Gly	Gly	Pro	Glu	Asn	Leu	Leu	Val	Thr	Val	
			180					185					190			
gca	aat	ccg	gat	atc	gaa	acc	gcg	caa	cgc	ttg	ttc	aag	ttt	ccg	ggt	624
Ala	Asn	Pro	Asp	Ile	Glu	Thr	Ala	Gln	Arg	Leu	Phe	Lys	Phe	Pro	Gly	
		195					200					205				
atc	ggc	ctg	ctg	gtg	gta	acc	ggc	ggc	gaa	gcg	gta	gta	gaa	gcg	gcg	672
Ile	Gly	Leu	Leu	Val	Val	Thr	Gly	Gly	Glu	Ala	Val	Val	Glu	Ala	Ala	
	210						215				220					
cgt	aaa	cac	acc	aat	aaa	cgt	ctg	att	gcc	gca	ggc	gct	ggc	aac	ccg	720
Arg	Lys	His	Thr	Asn	Lys	Arg	Leu	Ile	Ala	Ala	Gly	Ala	Gly	Asn	Pro	
225					230					235					240	
ccg	gta	gtg	gtg	gat	gaa	acc	gcc	gac	ctc	gcc	cgt	gcc	gct	cag	tcc	768
Pro	Val	Val	Val	Asp	Glu	Thr	Ala	Asp	Leu	Ala	Arg	Ala	Ala	Gln	Ser	
				245					250					255		
atc	gtc	aaa	ggc	gct	tct	ttc	gat	aac	aac	atc	att	tgt	gcc	gac	gaa	816
Ile	Val	Lys	Gly	Ala	Ser	Phe	Asp	Asn	Asn	Ile	Ile	Cys	Ala	Asp	Glu	
			260					265					270			
aag	gta	ctg	att	gtt	gtt	gat	agc	gta	gcc	gat	gaa	ctg	atg	cgt	ctg	864
Lys	Val	Leu	Ile	Val	Val	Asp	Ser	Val	Ala	Asp	Glu	Leu	Met	Arg	Leu	
		275					280					285				
atg	gaa	ggc	cag	cac	gcg	gtg	aaa	ctg	acc	gca	gaa	cag	gcg	cag	cag	912
Met	Glu	Gly	Gln	His	Ala	Val	Lys	Leu	Thr	Ala	Glu	Gln	Ala	Gln	Gln	
	290					295					300					
ctg	caa	ccg	gtg	ttg	ctg	aaa	aat	atc	gac	gag	cgc	gga	aaa	ggc	acc	960
Leu	Gln	Pro	Val	Leu	Leu	Lys	Asn	Ile	Asp	Glu	Arg	Gly	Lys	Gly	Thr	
305					310					315				320		
gtc	agc	cgt	gac	tgg	gtt	ggt	cgc	gac	gca	ggc	aaa	atc	gcg	gcg	gca	1008
Val	Ser	Arg	Asp	Trp	Val	Gly	Arg	Asp	Ala	Gly	Lys	Ile	Ala	Ala	Ala	
				325					330					335		
atc	ggc	ctt	aaa	gtt	ccg	caa	gaa	acg	cgc	ctg	ctg	ttt	gtg	gaa	acc	1056
Ile	Gly	Leu	Lys	Val	Pro	Gln	Glu	Thr	Arg	Leu	Leu	Phe	Val	Glu	Thr	
			340					345					350			
acc	gca	gaa	cat	ccg	ttt	gcc	gtg	act	gaa	ctg	atg	atg	ccg	gtg	ttg	1104
Thr	Ala	Glu	His	Pro	Phe	Ala	Val	Thr	Glu	Leu	Met	Met	Pro	Val	Leu	
		355				360							365			
ccc	gtc	gtg	cgc	gtc	gcc	aac	gtg	gcg	gat	gcc	att	gcg	cta	gcg	gtg	1152
Pro	Val	Val	Arg	Val	Ala	Asn	Val	Ala	Asp	Ala	Ile	Ala	Leu	Ala	Val	
		370				375					380					
aaa	ctg	gaa	ggc	ggt	tgc	cac	cac	acg	gcg	gca	atg	cac	tcg	cgc	aac	1200
Lys	Leu	Glu	Gly	Gly	Cys	His	His	Thr	Ala	Ala	Met	His	Ser	Arg	Asn	
385					390					395					400	
atc	gaa	aac	atg	aac	cag	atg	gcg	aat	gct	att	gat	acc	agc	att	ttc	1248
Ile	Glu	Asn	Met	Asn	Gln	Met	Ala	Asn	Ala	Ile	Asp	Thr	Ser	Ile	Phe	

ES 2 607 894 T3

	405	410	415	
gtt aag aac gga ccg tgc att gcc ggg ctg ggg ctg ggc ggg gaa ggc				1296
Val Lys Asn Gly Pro Cys Ile Ala Gly Leu Gly Leu Gly Gly Glu Gly	420	425	430	
tgg acc acc atg acc atc acc acg cca acc ggt gaa ggg gta acc agc				1344
Trp Thr Thr Met Thr Ile Thr Thr Pro Thr Gly Glu Gly Val Thr Ser	435	440	445	
gcg cgt acg ttt gtc cgt ctg cgt cgc tgt gta tta gtc gat gcg ttt				1392
Ala Arg Thr Phe Val Arg Leu Arg Arg Cys Val Leu Val Asp Ala Phe	450	455	460	
cgc att gtt taa				1404
Arg Ile Val				
465				

<210> 19

<211> 467

5

<212> PRT

<213> Escherichia coli

<400> 19

Met Asn Gln Gln Asp Ile Glu Gln Val Val Lys Ala Val Leu Leu Lys	1	5	10	15
Met Gln Ser Ser Asp Thr Pro Ser Ala Ala Val His Glu Met Gly Val	20	25	30	
Phe Ala Ser Leu Asp Asp Ala Val Ala Ala Ala Lys Val Ala Gln Gln	35	40	45	
Gly Leu Lys Ser Val Ala Met Arg Gln Leu Ala Ile Ala Ala Ile Arg	50	55	60	
Glu Ala Gly Glu Lys His Ala Arg Asp Leu Ala Glu Leu Ala Val Ser	65	70	75	80
Glu Thr Gly Met Gly Arg Val Glu Asp Lys Phe Ala Lys Asn Val Ala	85	90	95	
Gln Ala Arg Gly Thr Pro Gly Val Glu Cys Leu Ser Pro Gln Val Leu	100	105	110	
Thr Gly Asp Asn Gly Leu Thr Leu Ile Glu Asn Ala Pro Trp Gly Val	115	120	125	

10

ES 2 607 894 T3

Val Ala Ser Val Thr Pro Ser Thr Asn Pro Ala Ala Thr Val Ile Asn
 130 135 140

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

His Pro Ala Ala Lys Lys Val Ser Gln Arg Ala Ile Thr Leu Leu Asn
 165 170 175

Gln Ala Ile Val Ala Ala Gly Gly Pro Glu Asn Leu Leu Val Thr Val
 180 185 190

Ala Asn Pro Asp Ile Glu Thr Ala Gln Arg Leu Phe Lys Phe Pro Gly
 195 200 205

Ile Gly Leu Leu Val Val Thr Gly Gly Glu Ala Val Val Glu Ala Ala
 210 215 220

Arg Lys His Thr Asn Lys Arg Leu Ile Ala Ala Gly Ala Gly Asn Pro
 225 230 235 240

Pro Val Val Val Asp Glu Thr Ala Asp Leu Ala Arg Ala Ala Gln Ser
 245 250 255

Ile Val Lys Gly Ala Ser Phe Asp Asn Asn Ile Ile Cys Ala Asp Glu
 260 265 270

Lys Val Leu Ile Val Val Asp Ser Val Ala Asp Glu Leu Met Arg Leu
 275 280 285

Met Glu Gly Gln His Ala Val Lys Leu Thr Ala Glu Gln Ala Gln Gln
 290 295 300

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

Val Ser Arg Asp Trp Val Gly Arg Asp Ala Gly Lys Ile Ala Ala Ala
 325 330 335

Ile Gly Leu Lys Val Pro Gln Glu Thr Arg Leu Leu Phe Val Glu Thr
 340 345 350

ES 2 607 894 T3

Thr Ala Glu His Pro Phe Ala Val Thr Glu Leu Met Met Pro Val Leu
 355 360 365
 Pro Val Val Arg Val Ala Asn Val Ala Asp Ala Ile Ala Leu Ala Val
 370 375 380
 Lys Leu Glu Gly Gly Cys His His Thr Ala Ala Met His Ser Arg Asn
 385 390 395 400
 Ile Glu Asn Met Asn Gln Met Ala Asn Ala Ile Asp Thr Ser Ile Phe
 405 410 415
 Val Lys Asn Gly Pro Cys Ile Ala Gly Leu Gly Leu Gly Gly Glu Gly
 420 425 430
 Trp Thr Thr Met Thr Ile Thr Thr Pro Thr Gly Glu Gly Val Thr Ser
 435 440 445
 Ala Arg Thr Phe Val Arg Leu Arg Arg Cys Val Leu Val Asp Ala Phe
 450 455 460
 Arg Ile Val
 465

<210> 20

<211> 1401

5

<212> ADN

<213> Artificial

<220>

<223> secuencia optimizada

10

<400> 20

atgaaccaac aagatatcga acaagttgtc aaggctgtct tgttgaaaat gcaatcttct 60
 gacactccat ctgctgctgt ccacgaaatg ggtgttttcg cttctttgga cgacgctggt 120
 gctgctgcca aggttgctca acaaggtttg aaatctgttg ccatgagaca attggccatt 180
 gctgccatca gagaagctgg tgaaaagcat gccagagact tggctgaatt ggctgtctcc 240
 gaaaccggtg tgggtagagt tgaagacaaa ttcgctaaga acgttgctca agctagaggt 300
 actccaggtg tcgaatgttt gtctccacaa gtcttgaccg gtgataatgg tttgactttg 360
 attgaaaatg ctccatgggg tgttgttgct tcogtcaccc catctaccaa ccagctgct 420
 actgtcatca acaacgccat ctctttgatt gctgctggta actcogttat ctctcgtcca 480

ES 2 607 894 T3

```

caccagctg ccaagaaggt ttctcaaaga gccatcactc tattgaacca agccattggt      540
gctgctggtg gtccagaaaa cttgttggtc actgttgcca acccagatat cgaaactgct      600
caaagattat tcaagttccc aggtatcggt ctattagtcg tcaactggtg tgaagctggt      660
gttgaagctg ccagaaagca caccaacaag agattgattg ctgctggtgc tggtaaccct      720
cctgttggtg tcgatgaaac cgctgatttg gccagagctg ctcaatccat tgtcaagggt      780
gcttctttcg acaacaacat catctgtgct gacgaaaagg ttttgattgt tgttgactcc      840
gttgctgacg aattgatgag attgatggaa ggtcaacatg ccgtcaagtt gactgctgaa      900
caagctcaac aattgcaacc agttttggtg aagaacatcg atgaaagagg taagggtacc      960
gtctccagag actggggttg tagagatgct ggtaagattg ctgctgccat cggtttgaag    1020
gttcacaag aaaccagatt attatctgtc gaaaccaccg ctgaacaccc atttgctgtc    1080
actgaattga tgatgccagt cttaccagtt gtcctgtgtg ctaacgttgc tgacgtatt    1140
gctttggctg tcaaattgga agtggttgt caccacactg ctgccatgca ctccagaaac    1200
atcgaaaaca tgaaccaaat ggctaacgcc attgacactt ccatctttgt caagaacggt    1260
ccatgtatcg ctggtttggg tttgggtggt gaaggttga ccaccatgac catcaccacc    1320
cgaactggtg aaggtgtcac ttctgccaga actttcgtca gattacgtcg ttgtgttttg    1380
gtcgatgctt tcagaattgt t                                     1401

```

<210> 21

<211> 1410

5

<212> ADN

<213> *Listeria innocua*

<220>

< 221> CDS

10

< 222> (1)..(1410)

<400> 21

```

atg gaa tca tta gaa ctc gaa caa ctg gta aaa aaa gtt ctc tta gaa      48
Met Glu Ser Leu Glu Leu Glu Gln Leu Val Lys Lys Val Leu Leu Glu
1           5           10           15

aaa tta gca gaa caa aaa gaa gta cca aca aaa aca act aca caa ggc      96
Lys Leu Ala Glu Gln Lys Glu Val Pro Thr Lys Thr Thr Thr Gln Gly
20           25           30

gcg aaa agt ggc gtt ttt gat aca gtt gac gag gct gtt caa gca gca    144
Ala Lys Ser Gly Val Phe Asp Thr Val Asp Glu Ala Val Gln Ala Ala
35           40           45

gtt ata gcg cag aat tgc tat aaa gaa aaa tca ctt gaa gaa cgc cgc    192

```

ES 2 607 894 T3

Val	Ile	Ala	Gln	Asn	Cys	Tyr	Lys	Glu	Lys	Ser	Leu	Glu	Glu	Arg	Arg	
	50					55					60					
aat	gtt	gta	aaa	gca	att	cgt	gaa	gca	ctt	tat	cca	gaa	att	gaa	aca	240
Asn	Val	Val	Lys	Ala	Ile	Arg	Glu	Ala	Leu	Tyr	Pro	Glu	Ile	Glu	Thr	
65					70					75					80	
att	gcg	aca	aga	gca	gtt	gca	gag	act	ggt	atg	gga	aat	gtg	aca	gat	288
Ile	Ala	Thr	Arg		Val	Ala	Glu	Thr		Gly	Met	Gly	Asn	Val	Thr	Asp
				85					90					95		
aaa	att	ttg	aaa	aac	acg	tta	gca	atc	gaa	aaa	acg	cca	ggg	gta	gaa	336
Lys	Ile	Leu	Lys	Asn	Thr	Leu	Ala	Ile	Glu	Lys	Thr	Pro	Gly	Val	Glu	
			100					105					110			
gat	tta	tat	aca	gaa	gta	gct	aca	ggt	gat	aac	ggt	atg	aca	cta	tat	384
Asp	Leu	Tyr	Thr	Glu	Val	Ala	Thr	Gly	Asp	Asn	Gly	Met	Thr	Leu	Tyr	
		115					120					125				
gaa	ctc	tct	ccg	tat	ggc	gta	att	ggt	gca	gta	gcg	ccg	agc	aca	aac	432
Glu	Leu	Ser	Pro	Tyr	Gly	Val	Ile	Gly	Ala	Val	Ala	Pro	Ser	Thr	Asn	
	130					135					140					
cca	acg	gaa	aca	ttg	att	tgt	aat	tca	atc	ggt	atg	ctc	gca	gct	gga	480
Pro	Thr	Glu	Thr	Leu	Ile	Cys	Asn	Ser	Ile	Gly	Met	Leu	Ala	Ala	Gly	
145					150					155					160	
aat	gcc	gtt	ttt	tat	agc	cct	cat	cca	ggg	gca	aaa	aac	att	tca	ctg	528
Asn	Ala	Val	Phe	Tyr	Ser	Pro	His	Pro	Gly	Ala	Lys	Asn	Ile	Ser	Leu	
				165					170					175		
tgg	ttg	att	gaa	aaa	cta	aac	aca	att	gtt	cgc	gat	agt	tgt	ggt	ata	576
Trp	Leu	Ile	Glu	Lys	Leu	Asn	Thr	Ile	Val	Arg	Asp	Ser	Cys	Gly	Ile	
			180					185					190			
gat	aat	cta	att	gtc	acc	gtg	gct	aaa	cca	tcc	atc	caa	gca	gct	caa	624
Asp	Asn	Leu	Ile	Val	Thr	Val	Ala	Lys	Pro	Ser	Ile	Gln	Ala	Ala	Gln	
		195					200					205				
gaa	atg	atg	aac	cat	cca	aaa	gta	ccg	cta	ctt	gtt	att	aca	ggt	ggt	672
Glu	Met	Met	Asn	His	Pro	Lys	Val	Pro	Leu	Leu	Val	Ile	Thr	Gly	Gly	
	210					215					220					
ccg	ggc	gtt	gtt	ctc	caa	gcg	atg	caa	tca	ggt	aaa	aaa	gtg	att	gga	720
Pro	Gly	Val	Val	Leu	Gln	Ala	Met	Gln	Ser	Gly	Lys	Lys	Val	Ile	Gly	
225					230					235				240		
gca	gga	gca	ggg	aac	ccg	cct	tct	att	gtt	gac	gaa	aca	gct	aat	atc	768
Ala	Gly	Ala	Gly	Asn	Pro	Pro	Ser	Ile	Val	Asp	Glu	Thr	Ala	Asn	Ile	
				245					250					255		
gaa	aaa	gcg	gct	gct	gac	atc	gta	gac	gga	gca	tct	ttt	gac	cat	aat	816
Glu	Lys	Ala	Ala	Ala	Asp	Ile	Val	Asp	Gly	Ala	Ser	Phe	Asp	His	Asn	
			260					265					270			
att	tta	tgt	att	gct	gaa	aaa	agt	gtg	gta	gct	gtt	gat	agc	att	gct	864
Ile	Leu	Cys	Ile	Ala	Glu	Lys	Ser	Val	Val	Ala	Val	Asp	Ser	Ile	Ala	

ES 2 607 894 T3

	275	280	285	
gat ttc ttg tta ttc caa atg gaa aaa aat ggt gcc ctt cat gtt act Asp Phe Leu Leu Phe Gln Met Glu Lys Asn Gly Ala Leu His Val Thr 290 295 300				912
aat cca agt gat att caa aaa tta gaa aaa gta gcc gtt acc gat aaa Asn Pro Ser Asp Ile Gln Lys Leu Glu Lys Val Ala Val Thr Asp Lys 305 310 315 320				960
ggt gta act aat aaa aaa tta gtc gga aaa agt gca act gaa atc tta Gly Val Thr Asn Lys Lys Leu Val Gly Lys Ser Ala Thr Glu Ile Leu 325 330 335				1008
aaa gaa gca gga ata gct tgt gat ttt aca cca cgt tta atc att gtg Lys Glu Ala Gly Ile Ala Cys Asp Phe Thr Pro Arg Leu Ile Ile Val 340 345 350				1056
gaa acg gag aaa tct cat cca ttt gca aca gta gag cta tta atg cca Glu Thr Glu Lys Ser His Pro Phe Ala Thr Val Glu Leu Leu Met Pro 355 360 365				1104
atc gtt cca gtt gta agg gtg cct gat ttt gac gaa gcc ctt gaa gtg Ile Val Pro Val Val Arg Val Pro Asp Phe Asp Glu Ala Leu Glu Val 370 375 380				1152
gct att gaa ctc gaa caa ggc tta cat cat aca gca aca atg cat tca Ala Ile Glu Leu Glu Gln Gly Leu His His Thr Ala Thr Met His Ser 385 390 395 400				1200
caa aat atc tcg aga tta aac aaa gct gca aga gat atg caa act tcc Gln Asn Ile Ser Arg Leu Asn Lys Ala Ala Arg Asp Met Gln Thr Ser 405 410 415				1248
atc ttt gtc aaa aat ggt ccg tcc ttt gcg gga tta ggc ttt aga gga Ile Phe Val Lys Asn Gly Pro Ser Phe Ala Gly Leu Gly Phe Arg Gly 420 425 430				1296
gaa ggt agt act act ttc act att gca acg cct act gga gaa gga aca Glu Gly Ser Thr Thr Phe Thr Ile Ala Thr Pro Thr Gly Glu Gly Thr 435 440 445				1344
act aca gca cgt cat ttt gct aga cgc cgc cgc tgt gtt tta aca gat Thr Thr Ala Arg His Phe Ala Arg Arg Arg Arg Cys Val Leu Thr Asp 450 455 460				1392
ggt ttt tcg att cgt taa Gly Phe Ser Ile Arg 465				1410

<210> 22

<211> 469

5

<212> PRT

<213> Listeria innocua

<400> 22

ES 2 607 894 T3

Met Glu Ser Leu Glu Leu Glu Gln Leu Val Lys Lys Val Leu Leu Glu
 1 5 10 15
 Lys Leu Ala Glu Gln Lys Glu Val Pro Thr Lys Thr Thr Thr Gln Gly
 20 25 30
 Ala Lys Ser Gly Val Phe Asp Thr Val Asp Glu Ala Val Gln Ala Ala
 35 40 45
 Val Ile Ala Gln Asn Cys Tyr Lys Glu Lys Ser Leu Glu Glu Arg Arg
 50 55 60
 Asn Val Val Lys Ala Ile Arg Glu Ala Leu Tyr Pro Glu Ile Glu Thr
 65 70 75 80
 Ile Ala Thr Arg Ala Val Ala Glu Thr Gly Met Gly Asn Val Thr Asp
 85 90 95
 Lys Ile Leu Lys Asn Thr Leu Ala Ile Glu Lys Thr Pro Gly Val Glu
 100 105 110
 Asp Leu Tyr Thr Glu Val Ala Thr Gly Asp Asn Gly Met Thr Leu Tyr
 115 120 125
 Glu Leu Ser Pro Tyr Gly Val Ile Gly Ala Val Ala Pro Ser Thr Asn
 130 135 140
 Pro Thr Glu Thr Leu Ile Cys Asn Ser Ile Gly Met Leu Ala Ala Gly
 145 150 155 160
 Asn Ala Val Phe Tyr Ser Pro His Pro Gly Ala Lys Asn Ile Ser Leu
 165 170 175
 Trp Leu Ile Glu Lys Leu Asn Thr Ile Val Arg Asp Ser Cys Gly Ile
 180 185 190
 Asp Asn Leu Ile Val Thr Val Ala Lys Pro Ser Ile Gln Ala Ala Gln
 195 200 205
 Glu Met Met Asn His Pro Lys Val Pro Leu Leu Val Ile Thr Gly Gly
 210 215 220

ES 2 607 894 T3

Pro Gly Val Val Leu Gln Ala Met Gln Ser Gly Lys Lys Val Ile Gly
 225 230 235 240

Ala Gly Ala Gly Asn Pro Pro Ser Ile Val Asp Glu Thr Ala Asn Ile
 245 250 255

Glu Lys Ala Ala Ala Asp Ile Val Asp Gly Ala Ser Phe Asp His Asn
 260 265 270

Ile Leu Cys Ile Ala Glu Lys Ser Val Val Ala Val Asp Ser Ile Ala
 275 280 285

Asp Phe Leu Leu Phe Gln Met Glu Lys Asn Gly Ala Leu His Val Thr
 290 295 300

Asn Pro Ser Asp Ile Gln Lys Leu Glu Lys Val Ala Val Thr Asp Lys
 305 310 315 320

Gly Val Thr Asn Lys Lys Leu Val Gly Lys Ser Ala Thr Glu Ile Leu
 325 330 335

Lys Glu Ala Gly Ile Ala Cys Asp Phe Thr Pro Arg Leu Ile Ile Val
 340 345 350

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

Ile Val Pro Val Val Arg Val Pro Asp Phe Asp Glu Ala Leu Glu Val
 370 375 380

Ala Ile Glu Leu Glu Gln Gly Leu His His Thr Ala Thr Met His Ser
 385 390 395 400

Gln Asn Ile Ser Arg Leu Asn Lys Ala Ala Arg Asp Met Gln Thr Ser
 405 410 415

Ile Phe Val Lys Asn Gly Pro Ser Phe Ala Gly Leu Gly Phe Arg Gly
 420 425 430

Glu Gly Ser Thr Thr Phe Thr Ile Ala Thr Pro Thr Gly Glu Gly Thr
 435 440 445

Thr Thr Ala Arg His Phe Ala Arg Arg Arg Arg Cys Val Leu Thr Asp
 450 455 460

Gly Phe Ser Ile Arg
 465

<210> 23

5 <211> 1407

<212> ADN

ES 2 607 894 T3

<213> Artificial

<220>

<223> secuencia optimizada

5

<400> 23

```
atggaatctt tgggaattgga acaattagtc aagaaggttt tgttgaaaa attggctgaa    60
caaaaggaag ttccaaccaa gaccaccacc caaggtgcc agtccggtgt tttcgatacc    120
gtcgatgaag ctgtccaagc tgccgctcatt gctcaaaact gttacaagga aaaatctttg    180
gaagaagaa gaaacgttgt caaggccatc agagaagctt tataaccaga aatcgaaacc    240
attgctacca gagctgttgc tgaaacgggt atgggtaatg tcaccgataa aatcttgaag    300
aacactttag ctatcgaaaa gactccaggt gttgaagact tgtacactga agttgctacc    360
ggtgacaacg gtatgacttt atacgaatta tctccatacg gtgtcatcgg tgcgtttgct    420
ccatctacca acccaactga aactttgatc tgtaactcca tcggtatggt ggctgctggt    480
aacgccgttt tctactctcc tcaccaggt gccaaagaaca tctctttatg gttgattgaa    540
aagttgaaca ctatcgtcag agattcttgt ggtattgaca acttgattgt caccggtgcc    600
aagccatcta tccaagctgc tcaagaaatg atgaaccacc caaagggtcc attgttggtc    660
atcactgggtg gtccaggtgt tgccttgcaa gctatgcaat ctggtaagaa ggttatcggg    720
gctggtgctg gtaaccctcc atccatcgtt gacgaaaccg ctaacattga aaaggctgct    780
gctgacattg tcgacggtgc ttcctttgac cataatatct tgtgtatcgc tgaaaagtct    840
gttgttgccg ttgactccat tgcctgactc ttgtttgtcc aaatggaaaa gaacgggtgct    900
ttgcacgtca ctaaccatc tgatatccaa aaattggaaa aggttgccgt cactgacaag    960
gggtgcacca acaagaaatt ggttggtgtaag tctgccactg aaatcttgaa agaagctggt   1020
attgcttgtg atttcaccoc aagattgatc attgtcgaaa ctgaaaagtc ccaccattc   1080
gctactgttg aattgttgat gccaatgtt ccagttgtca gagttccaga cttcgatgaa   1140
gctttggaag ttgccattga attggaacaa ggtctacatc aactgctac catgcaactc   1200
caaaacatct ccagattgaa caaggctgcc cgtgacatgc aaacctccat ctttgtcaag   1260
aacgggtccat ctttcgctgg tttaggtttc agaggtgaa gttccaccac tttcaccatt   1320
gtactccaa ctggtgaagg tactaccact gcccgctcact tcgctagaag aagaagatgt   1380
gtcttgactg atggtttctc cattaga                                     1407
```

10

<210> 24

<211> 1476

<212> ADN

<213> Clostridium kluveri

ES 2 607 894 T3

<220>

< 221> CDS

< 222> (1) .. (1476)

5

<400> 24

```

atg gag ata atg gat aag gac tta cag tca ata cag gaa gta aga act      48
Met Glu Ile Met Asp Lys Asp Leu Gln Ser Ile Gln Glu Val Arg Thr
1                               5                               10                               15

ctt ata gca aaa gca aag aaa gct caa gca gaa ttt aaa aat ttt tct      96
Leu Ile Ala Lys Ala Lys Lys Ala Gln Ala Glu Phe Lys Asn Phe Ser
20                               25                               30

caa gaa gct gta aac aag gta ata gaa aaa ata gct aag gct aca gaa      144
Gln Glu Ala Val Asn Lys Val Ile Glu Lys Ile Ala Lys Ala Thr Glu
35                               40                               45

gtt gaa gct gta aaa ctt gca aaa ttg gca tat gaa gat aca gga tat      192
Val Glu Ala Val Lys Leu Ala Lys Leu Ala Tyr Glu Asp Thr Gly Tyr
50                               55                               60

gga aaa tgg gaa gat aaa gta ata aag aat aag ttt tca agt ata gta      240
Gly Lys Trp Glu Asp Lys Val Ile Lys Asn Lys Phe Ser Ser Ile Val
65                               70                               75                               80

gtt tat aac tat att aaa gat ttg aaa acg gtt gga att tta aaa gaa      288
Val Tyr Asn Tyr Ile Lys Asp Leu Lys Thr Val Gly Ile Leu Lys Glu
85                               90                               95

gac aag gaa aag aaa tta ata gat ata gct gtt cca ctt gga gtt ata      336
Asp Lys Glu Lys Lys Leu Ile Asp Ile Ala Val Pro Leu Gly Val Ile
100                              105                              110

gca gga ctt ata cct tca act aac cca act tca aca gca ata ttc aag      384
Ala Gly Leu Ile Pro Ser Thr Asn Pro Thr Ser Thr Ala Ile Phe Lys
115                              120                              125

gta tta ata gca tta aag gca gga aat gca ata gta ttc tca cca cat      432
Val Leu Ile Ala Leu Lys Ala Gly Asn Ala Ile Val Phe Ser Pro His
130                              135                              140

cca aca gca gta aga agt att aca gaa act gta aag ata atg cag aaa      480
Pro Thr Ala Val Arg Ser Ile Thr Glu Thr Val Lys Ile Met Gln Lys

```

ES 2 607 894 T3

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

ES 2 607 894 T3

ata aga gaa ttc gca ttg aag aaa cca gta tca aga ata ctt gta aat 1200
 Ile Arg Glu Phe Ala Leu Lys Lys Pro Val Ser Arg Ile Leu Val Asn
 385 390 395 400

tca cct gga tca ctt gga gga ata ggt gga gct aca aat ctt gta cca 1248
 Ser Pro Gly Ser Leu Gly Gly Ile Gly Gly Ala Thr Asn Leu Val Pro
 405 410 415

tca ctt aca tta ggc tgt gga gca gta ggt gga agt gca act tca gat 1296
 Ser Leu Thr Leu Gly Cys Gly Ala Val Gly Gly Ser Ala Thr Ser Asp
 420 425 430

aac gta gga cca gaa aac tta ttc aac ata aga aaa gta gct tat gga 1344
 Asn Val Gly Pro Glu Asn Leu Phe Asn Ile Arg Lys Val Ala Tyr Gly
 435 440 445

act acg aca gta gaa gaa ata aga gaa gct ttt ggt gta gga gca gct 1392
 Thr Thr Thr Val Glu Glu Ile Arg Glu Ala Phe Gly Val Gly Ala Ala
 450 455 460

tca tca agt gca cca gca gaa cca gaa gat aat gaa gat gta cag gct 1440
 Ser Ser Ser Ala Pro Ala Glu Pro Glu Asp Asn Glu Asp Val Gln Ala
 465 470 475 480

ata gta aaa gct ata atg gct aaa tta aat ctt taa 1476
 Ile Val Lys Ala Ile Met Ala Lys Leu Asn Leu
 485 490

<210> 25

<211> 491

5

<212> PRT

<213> Clostridium kluveri

<400> 25

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

Leu Ile Ala Lys Ala Lys Lys Ala Gln Ala Glu Phe Lys Asn Phe Ser
 20 25 30

Gln Glu Ala Val Asn Lys Val Ile Glu Lys Ile Ala Lys Ala Thr Glu
 35 40 45

Val Glu Ala Val Lys Leu Ala Lys Leu Ala Tyr Glu Asp Thr Gly Tyr
 50 55 60

Gly Lys Trp Glu Asp Lys Val Ile Lys Asn Lys Phe Ser Ser Ile Val
 65 70 75 80

10

ES 2 607 894 T3

Val Tyr Asn Tyr Ile Lys Asp Leu Lys Thr Val Gly Ile Leu Lys Glu
85 90 95

Asp Lys Glu Lys Lys Leu Ile Asp Ile Ala Val Pro Leu Gly Val Ile
100 105 110

Ala Gly Leu Ile Pro Ser Thr Asn Pro Thr Ser Thr Ala Ile Phe Lys
115 120 125

Val Leu Ile Ala Leu Lys Ala Gly Asn Ala Ile Val Phe Ser Pro His
130 135 140

Pro Thr Ala Val Arg Ser Ile Thr Glu Thr Val Lys Ile Met Gln Lys
145 150 155 160

Ala Ala Val Glu Ala Gly Ala Pro Asp Gly Leu Ile Gln Cys Met Ser
165 170 175

Ile Leu Thr Val Glu Gly Thr Ala Glu Leu Met Lys Asn Lys Asp Thr
180 185 190

Ala Leu Ile Leu Ala Thr Gly Gly Glu Gly Met Val Arg Ala Ala Tyr
195 200 205

Ser Ser Gly Thr Pro Ala Ile Gly Val Gly Pro Gly Asn Gly Pro Cys
210 215 220

Phe Ile Glu Arg Thr Ala Asp Ile Pro Thr Ala Val Arg Lys Val Ile
225 230 235 240

Gly Ser Asp Thr Phe Asp Asn Gly Val Ile Cys Ala Ser Glu Gln Ser
245 250 255

Ile Ile Ala Glu Thr Val Lys Lys Ala Glu Ile Ile Glu Glu Phe Lys
260 265 270

Arg Gln Lys Gly Tyr Phe Leu Asn Ala Glu Glu Ser Glu Lys Val Gly
275 280 285

Lys Ile Leu Leu Arg Ala Asn Gly Thr Pro Asn Pro Ala Ile Val Gly
290 295 300

Lys Asp Val Gln Ala Leu Ala Lys Leu Ala Gly Ile Ser Ile Pro Ser

ES 2 607 894 T3

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

<210> 26

<211> 1473

5

<212> ADN

<213> Artificial

<220>

<223> secuencia optimizada

10

<400> 26

atggaaatca tggacaagga tttgcaatcc atccaagaag ttagaacttt gattgccaag 60

ES 2 607 894 T3

```

gccaagaagg ctcaagctga attcaagaac ttttccaag aagctggtta caaggtcatc 120
gaaaagatcg ccaaggctac tgaagttgaa gctgtcaaat tggccaaatt ggcttacgaa 180
gacaccggtt acggtaaatg ggaagacaag gtcacaaaga acaaattctc ctccattggt 240
gtctacaact acatcaagga tttgaagacc gttggtatct tgaaggaaga caaggaaaag 300
aaattgattg acattgctgt cccattaggt gtcattgctg gtttgattcc atctaccaac 360
ccaacttcca ctgccattht caaggtcttg attgctttga aggctggtaa cgccattgtc 420
ttctctccac acccaactgc tgtccgttcc atcactgaaa ccgttaagat catgcaaaaag 480
gctgctgttg aagctgggtc tccagatggt ttgatccaat gtatgtccat tttgaccggt 540
gaaggtactg ctgaattgat gaagaacaag gacaccgctt tgatcttggc taccggtggt 600
gaaggtatgg ttagagctgc ttactcctct ggtactccag ccacgggtgt cgggtccaggt 660
aacgggtccat gtttcatcga aagaactgct gacattccaa ctgctggttag aaaggttatc 720
ggttctgaca ctttcgacaa cgggtgtcatc tgtgcttctg aacaatccat cattgctgaa 780
accgtcaaga aggctgaaat catcgaagaa ttcaagagac aaaagggta cttcttgaat 840
gctgaagaat ctgaaaaggt tggtaagatt ctattacgtg ccaacggtac tccaaacca 900
gccatcggtg gtaaggatgt ccaagctttg gccaaattgg ctggtatttc cattccatct 960
gatgctgta tcttactatc cgaacaaacc gatgtttctc ctaaaaatcc atacgtaag 1020
gaaaaattgg ctccagtctt ggctttctac accgtcgaag actggcatga agcttgtgaa 1080
aagtctttgg ctttattgca caaccaaggt tctggtcaca ctttgatcat ccactctcaa 1140
aacgaagaaa tcattagaga atttgctttg aagaagcctg tttccagaat tttggttaac 1200
tctccaggtt ctttgggtgg taccggtggt gctaccaact tagtccatc tttgacttta 1260
ggttgtggtg ctggtggtg ttctgccacc tctgacaacg ttggtccaga aaacttgttc 1320
aacatcagaa aggttgctta cggtagacc accgtcgaag aatcagaga agctttcgggt 1380
gtcgggtgctg cttcttcttc tgctccagct gaaccagaag acaacgaaga tgttcaagcc 1440
attgttaagg ccatcatggc caaattgaac ttg 1473

```

<210> 27

<211> 2610

5 <212> ADN

<213> *Staphylococcus aureus*

<220>

< 221> CDS

10 < 222> (1)..(2610)

<400> 27

ES 2 607 894 T3

atg tta act ata cct gaa aaa gaa aat cgt gga tcg aaa gaa caa gaa	5	10	15		48
Met Leu Thr Ile Pro Glu Lys Glu Asn Arg Gly Ser Lys Glu Gln Glu					
1					
gtg gca att atg att gat gct cta gct gac aaa ggg aaa aaa gca tta	20	25	30		96
Val Ala Ile Met Ile Asp Ala Leu Ala Asp Lys Gly Lys Lys Ala Leu					
gaa gca tta tct aaa aag tca caa gaa gaa att gat cat att gtt cat	35	40	45		144
Glu Ala Leu Ser Lys Lys Ser Gln Glu Glu Ile Asp His Ile Val His					
caa atg agc tta gca gct gtt gat caa cat atg gtg cta gca aaa tta	50	55	60		192
Gln Met Ser Leu Ala Ala Val Asp Gln His Met Val Leu Ala Lys Leu					
gca cat gaa gaa act gga aga ggt ata tac gaa gat aaa gcg att aaa	65	70	75	80	240
Ala His Glu Glu Thr Gly Arg Gly Ile Tyr Glu Asp Lys Ala Ile Lys					
aat tta tac gct tct gaa tat ata tgg aat tca ata aaa gac aat aag	85	90	95		288
Asn Leu Tyr Ala Ser Glu Tyr Ile Trp Asn Ser Ile Lys Asp Asn Lys					
aca gta ggg att att ggt gaa gat aaa gaa aaa gga tta acg tat gta	100	105	110		336
Thr Val Gly Ile Ile Gly Glu Asp Lys Glu Lys Gly Leu Thr Tyr Val					
gcg gaa cca att ggt gtt att tgt ggt gtt acg cca aca aca aat cct	115	120	125		384
Ala Glu Pro Ile Gly Val Ile Cys Gly Val Thr Pro Thr Thr Asn Pro					
acg tcg aca act att ttt aaa gcg atg att gca att aag aca gga aat	130	135	140		432
Thr Ser Thr Thr Ile Phe Lys Ala Met Ile Ala Ile Lys Thr Gly Asn					
cca atc att ttt gca ttc cat cca agt gca caa gaa tcg tcg aag cgt	145	150	155	160	480
Pro Ile Ile Phe Ala Phe His Pro Ser Ala Gln Glu Ser Ser Lys Arg					
gca gca gaa gtt gta tta gaa gcg gca atg aag gca ggt gca cct aaa	165	170	175		528
Ala Ala Glu Val Val Leu Glu Ala Ala Met Lys Ala Gly Ala Pro Lys					
gat att att cag tgg att gaa gtg cct tct atc gaa gca aca aaa caa	180	185	190		576
Asp Ile Ile Gln Trp Ile Glu Val Pro Ser Ile Glu Ala Thr Lys Gln					
tta atg aat cac aaa ggt att gca tta gtt cta gca aca ggt ggt tcg	195	200	205		624
Leu Met Asn His Lys Gly Ile Ala Leu Val Leu Ala Thr Gly Gly Ser					
ggc atg gtt aag tct gca tat tca act ggc aaa ccg gca tta ggt gtg					672

ES 2 607 894 T3

Gly	Met	Val	Lys	Ser	Ala	Tyr	Ser	Thr	Gly	Lys	Pro	Ala	Leu	Gly	Val	
	210					215					220					
gga cca ggt aac gtg ccg tct tac att gaa aaa aca gca cac att aaa																720
Gly	Pro	Gly	Asn	Val	Pro	Ser	Tyr	Ile	Glu	Lys	Thr	Ala	His	Ile	Lys	
					230					235					240	
cgt gca gta aat gat atc att ggt tca aaa aca ttt gat aat ggt atg																768
Arg	Ala	Val	Asn	Asp	Ile	Ile	Gly	Ser	Lys	Thr	Phe	Asp	Asn	Gly	Met	
				245					250					255		
att tgt gct tct gaa caa gtt gta gtc att gat aaa gaa att tat aaa																816
Ile	Cys	Ala	Ser	Glu	Gln	Val	Val	Val	Ile	Asp	Lys	Glu	Ile	Tyr	Lys	
			260					265					270			
gat gtt act aat gaa ttt aaa gca cat caa gca tac ttt gtt aaa aaa																864
Asp	Val	Thr	Asn	Glu	Phe	Lys	Ala	His	Gln	Ala	Tyr	Phe	Val	Lys	Lys	
		275					280					285				
gat gaa tta caa cgc tta gaa aat gca att atg aat gaa caa aaa aca																912
Asp	Glu	Leu	Gln	Arg	Leu	Glu	Asn	Ala	Ile	Met	Asn	Glu	Gln	Lys	Thr	
	290					295					300					
ggt att aag cct gat att gtc ggt aaa tct gca gtt gaa ata gct gaa																960
Gly	Ile	Lys	Pro	Asp	Ile	Val	Gly	Lys	Ser	Ala	Val	Glu	Ile	Ala	Glu	
	305				310					315					320	
tta gca ggt ata cct gtc ccc gaa aat aca aaa ctt atc ata gcc gaa																1008
Leu	Ala	Gly	Ile	Pro	Val	Pro	Glu	Asn	Thr	Lys	Leu	Ile	Ile	Ala	Glu	
				325					330					335		
att agc ggt gta ggt tca gac tat ccg tta tct cgt gaa aaa tta tct																1056
Ile	Ser	Gly	Val	Gly	Ser	Asp	Tyr	Pro	Leu	Ser	Arg	Glu	Lys	Leu	Ser	
			340					345					350			
cca gta tta gcc tta gta aaa gcc caa tct aca aaa caa gca ttt caa																1104
Pro	Val	Leu	Ala	Leu	Val	Lys	Ala	Gln	Ser	Thr	Lys	Gln	Ala	Phe	Gln	
		355				360						365				
att tgt gaa gac aca cta cat ttt ggt gga tta gga cac aca gcc gtt																1152
Ile	Cys	Glu	Asp	Thr	Leu	His	Phe	Gly	Gly	Leu	Gly	His	Thr	Ala	Val	
		370				375						380				
atc cat aca gaa gat gaa aca tta caa aaa gat ttt gga cta aga atg																1200
Ile	His	Thr	Glu	Asp	Glu	Thr	Leu	Gln	Lys	Asp	Phe	Gly	Leu	Arg	Met	
					385					390				395		400
aaa gct tgt cgt gta ctt gta aat aca cca tca gcg gtt gga ggt att																1248
Lys	Ala	Cys	Arg	Val	Leu	Val	Asn	Thr	Pro	Ser	Ala	Val	Gly	Gly	Ile	
				405					410					415		
ggt gat atg tat aac gaa ttg att ccg tct tta aca tta ggt tgt ggt																1296
Gly	Asp	Met	Tyr	Asn	Glu	Leu	Ile	Pro	Ser	Leu	Thr	Leu	Gly	Cys	Gly	
			420					425						430		
tcg tac ggt aga aac tca att tca cat aat gtt agt gcg aca gat tta																1344
Ser	Tyr	Gly	Arg	Asn	Ser	Ile	Ser	His	Asn	Val	Ser	Ala	Thr	Asp	Leu	

ES 2 607 894 T3

435	440	445	
tta aac att aaa acg att gct aaa cga cgt aat aat act caa att ttc			1392
Leu Asn Ile Lys Thr Ile Ala Lys Arg Arg Asn Asn Thr Gln Ile Phe			
450	455	460	
aag gtg cct gct caa att tat ttt gaa gaa aat gca atc atg agt cta			1440
Lys Val Pro Ala Gln Ile Tyr Phe Glu Glu Asn Ala Ile Met Ser Leu			
465	470	475	480
aca aca atg gac aag att gaa aaa gtg atg att gtc tgt gac cct ggt			1488
Thr Thr Met Asp Lys Ile Glu Lys Val Met Ile Val Cys Asp Pro Gly			
485	490	495	
atg gta gaa ttc ggt tat aca aaa aca gtt gag aat gta tta aga caa			1536
Met Val Glu Phe Gly Tyr Thr Lys Thr Val Glu Asn Val Leu Arg Gln			
500	505	510	
aga acg gaa cag cct caa att aaa ata ttt agc gaa gtc gaa ccg aac			1584
Arg Thr Glu Gln Pro Gln Ile Lys Ile Phe Ser Glu Val Glu Pro Asn			
515	520	525	
cca tca act aat aca gta tat aaa ggt ctg gaa atg atg gtt gat ttc			1632
Pro Ser Thr Asn Thr Val Tyr Lys Gly Leu Glu Met Met Val Asp Phe			
530	535	540	
caa cca gat aca atc att gca ctt ggt ggt ggt tca gcg atg gat gct			1680
Gln Pro Asp Thr Ile Ile Ala Leu Gly Gly Ser Ala Met Asp Ala			
545	550	555	560
gca aaa gca atg tgg atg ttc ttt gaa cac cct gag aca tca ttc ttc			1728
Ala Lys Ala Met Trp Met Phe Phe Glu His Pro Glu Thr Ser Phe Phe			
565	570	575	
ggt gct aaa caa aag ttc cta gac atc ggt aaa cgt act tat aaa ata			1776
Gly Ala Lys Gln Lys Phe Leu Asp Ile Gly Lys Arg Thr Tyr Lys Ile			
580	585	590	
ggc atg cct gaa aat gcg acg ttc att tgt atc cct acg aca tca ggt			1824
Gly Met Pro Glu Asn Ala Thr Phe Ile Cys Ile Pro Thr Thr Ser Gly			
595	600	605	
aca ggt tca gaa gta aca cca ttt gca gtt atc aca gat agt gaa aca			1872
Thr Gly Ser Glu Val Thr Pro Phe Ala Val Ile Thr Asp Ser Glu Thr			
610	615	620	
aat gta aaa tat ccg ttg gct gat ttt gct tta aca cct gac gtt gca			1920
Asn Val Lys Tyr Pro Leu Ala Asp Phe Ala Leu Thr Pro Asp Val Ala			
625	630	635	640
att att gac cct caa ttt gtg atg agt gtg cca aaa agc gtt aca gca			1968
Ile Ile Asp Pro Gln Phe Val Met Ser Val Pro Lys Ser Val Thr Ala			
645	650	655	
gat aca gga atg gat gta cta acg cat gca atg gaa tca tat gta tct			2016
Asp Thr Gly Met Asp Val Leu Thr His Ala Met Glu Ser Tyr Val Ser			
660	665	670	

ES 2 607 894 T3

gta atg gct tca gac tat aca aga ggt ttg agt cta caa gcg att aaa	2064
Val Met Ala Ser Asp Tyr Thr Arg Gly Leu Ser Leu Gln Ala Ile Lys	
675	680
685	
ttg acg ttc gaa tat tta aaa tca tct gtt gaa aag ggt gat aaa gtt	2112
Leu Thr Phe Glu Tyr Leu Lys Ser Ser Val Glu Lys Gly Asp Lys Val	
690	695
700	
tca aga gag aaa atg cat aac gca tca act ttg gct ggt atg gca ttt	2160
Ser Arg Glu Lys Met His Asn Ala Ser Thr Leu Ala Gly Met Ala Phe	
705	710
715	720
gca aat gca ttc tta ggc att gca cac tca att gca cat aaa att ggt	2208
Ala Asn Ala Phe Leu Gly Ile Ala His Ser Ile Ala His Lys Ile Gly	
725	730
735	
ggc gaa tat ggt att ccg cat ggt aga gcg aat gcg ata tta cta ccg	2256
Gly Glu Tyr Gly Ile Pro His Gly Arg Ala Asn Ala Ile Leu Leu Pro	
740	745
750	
cat att atc cgt tat aat gcc aaa gac ccg caa aaa cat gca tta ttc	2304
His Ile Ile Arg Tyr Asn Ala Lys Asp Pro Gln Lys His Ala Leu Phe	
755	760
765	
cct aaa tat gag ttc ttc aga gca gat aca gat tat gca gat att gcc	2352
Pro Lys Tyr Glu Phe Phe Arg Ala Asp Thr Asp Tyr Ala Asp Ile Ala	
770	775
780	
aaa ttc tta gga tta aaa ggg aat acg aca gaa gca ctc gta gaa tca	2400
Lys Phe Leu Gly Leu Lys Gly Asn Thr Thr Glu Ala Leu Val Glu Ser	
785	790
795	800
tta gct aaa gct gtc tac gaa tta ggt caa tca gtc gga att gaa atg	2448
Leu Ala Lys Ala Val Tyr Glu Leu Gly Gln Ser Val Gly Ile Glu Met	
805	810
815	
aat ttg aaa tca caa ggt gtg tct gaa gaa gaa tta aat gaa tca att	2496
Asn Leu Lys Ser Gln Gly Val Ser Glu Glu Glu Leu Asn Glu Ser Ile	
820	825
830	
gat aga atg gca gag ctc gca ttt gaa gat caa tgt aca act gct aat	2544
Asp Arg Met Ala Glu Leu Ala Phe Glu Asp Gln Cys Thr Thr Ala Asn	
835	840
845	
cct aaa gaa gca cta atc agt gaa atc aaa gat atc att caa aca tca	2592
Pro Lys Glu Ala Leu Ile Ser Glu Ile Lys Asp Ile Ile Gln Thr Ser	
850	855
860	
tat gat tat aag caa taa	2610
Tyr Asp Tyr Lys Gln	
865	

<210> 28

<211> 869

5

<212> PRT

<213> Staphylococcus aureus

<400> 28

ES 2 607 894 T3

Met Leu Thr Ile Pro Glu Lys Glu Asn Arg Gly Ser Lys Glu Gln Glu
 1 5 10 15

Val Ala Ile Met Ile Asp Ala Leu Ala Asp Lys Gly Lys Lys Ala Leu
 20 25 30

Glu Ala Leu Ser Lys Lys Ser Gln Glu Glu Ile Asp His Ile Val His
 35 40 45

Gln Met Ser Leu Ala Ala Val Asp Gln His Met Val Leu Ala Lys Leu
 50 55 60

Ala His Glu Glu Thr Gly Arg Gly Ile Tyr Glu Asp Lys Ala Ile Lys
 65 70 75 80

Asn Leu Tyr Ala Ser Glu Tyr Ile Trp Asn Ser Ile Lys Asp Asn Lys
 85 90 95

Thr Val Gly Ile Ile Gly Glu Asp Lys Glu Lys Gly Leu Thr Tyr Val
 100 105 110

Ala Glu Pro Ile Gly Val Ile Cys Gly Val Thr Pro Thr Thr Asn Pro
 115 120 125

Thr Ser Thr Thr Ile Phe Lys Ala Met Ile Ala Ile Lys Thr Gly Asn
 130 135 140

Pro Ile Ile Phe Ala Phe His Pro Ser Ala Gln Glu Ser Ser Lys Arg
 145 150 155 160

Ala Ala Glu Val Val Leu Glu Ala Ala Met Lys Ala Gly Ala Pro Lys
 165 170 175

Asp Ile Ile Gln Trp Ile Glu Val Pro Ser Ile Glu Ala Thr Lys Gln
 180 185 190

Leu Met Asn His Lys Gly Ile Ala Leu Val Leu Ala Thr Gly Gly Ser
 195 200 205

Gly Met Val Lys Ser Ala Tyr Ser Thr Gly Lys Pro Ala Leu Gly Val

ES 2 607 894 T3

210	215	220
Gly Pro Gly Asn Val Pro Ser Tyr Ile Glu Lys Thr Ala His Ile Lys 225 230 235 240		
Arg Ala Val Asn Asp Ile Ile Gly Ser Lys Thr Phe Asp Asn Gly Met 245 250 255		
Ile Cys Ala Ser Glu Gln Val Val Val Ile Asp Lys Glu Ile Tyr Lys 260 265 270		
Asp Val Thr Asn Glu Phe Lys Ala His Gln Ala Tyr Phe Val Lys Lys 275 280 285		
Asp Glu Leu Gln Arg Leu Glu Asn Ala Ile Met Asn Glu Gln Lys Thr 290 295 300		
Gly Ile Lys Pro Asp Ile Val Gly Lys Ser Ala Val Glu Ile Ala Glu 305 310 315 320		
Leu Ala Gly Ile Pro Val Pro Glu Asn Thr Lys Leu Ile Ile Ala Glu 325 330 335		
Ile Ser Gly Val Gly Ser Asp Tyr Pro Leu Ser Arg Glu Lys Leu Ser 340 345 350		
Pro Val Leu Ala Leu Val Lys Ala Gln Ser Thr Lys Gln Ala Phe Gln 355 360 365		
Ile Cys Glu Asp Thr Leu His Phe Gly Gly Leu Gly His Thr Ala Val 370 375 380		
Ile His Thr Glu Asp Glu Thr Leu Gln Lys Asp Phe Gly Leu Arg Met 385 390 395 400		
Lys Ala Cys Arg Val Leu Val Asn Thr Pro Ser Ala Val Gly Gly Ile 405 410 415		
Gly Asp Met Tyr Asn Glu Leu Ile Pro Ser Leu Thr Leu Gly Cys Gly 420 425 430		
Ser Tyr Gly Arg Asn Ser Ile Ser His Asn Val Ser Ala Thr Asp Leu 435 440 445		

ES 2 607 894 T3

Leu Asn Ile Lys Thr Ile Ala Lys Arg Arg Asn Asn Thr Gln Ile Phe
 450 455 460

Lys Val Pro Ala Gln Ile Tyr Phe Glu Glu Asn Ala Ile Met Ser Leu
 465 470 475 480

Thr Thr Met Asp Lys Ile Glu Lys Val Met Ile Val Cys Asp Pro Gly
 485 490 495

Met Val Glu Phe Gly Tyr Thr Lys Thr Val Glu Asn Val Leu Arg Gln
 500 505 510

Arg Thr Glu Gln Pro Gln Ile Lys Ile Phe Ser Glu Val Glu Pro Asn
 515 520 525

Pro Ser Thr Asn Thr Val Tyr Lys Gly Leu Glu Met Met Val Asp Phe
 530 535 540

Gln Pro Asp Thr Ile Ile Ala Leu Gly Gly Gly Ser Ala Met Asp Ala
 545 550 555 560

Ala Lys Ala Met Trp Met Phe Phe Glu His Pro Glu Thr Ser Phe Phe
 565 570 575

Gly Ala Lys Gln Lys Phe Leu Asp Ile Gly Lys Arg Thr Tyr Lys Ile
 580 585 590

Gly Met Pro Glu Asn Ala Thr Phe Ile Cys Ile Pro Thr Thr Ser Gly
 595 600 605

Thr Gly Ser Glu Val Thr Pro Phe Ala Val Ile Thr Asp Ser Glu Thr
 610 615 620

Asn Val Lys Tyr Pro Leu Ala Asp Phe Ala Leu Thr Pro Asp Val Ala
 625 630 635 640

Ile Ile Asp Pro Gln Phe Val Met Ser Val Pro Lys Ser Val Thr Ala
 645 650 655

Asp Thr Gly Met Asp Val Leu Thr His Ala Met Glu Ser Tyr Val Ser
 660 665 670

ES 2 607 894 T3

Val Met Ala Ser Asp Tyr Thr Arg Gly Leu Ser Leu Gln Ala Ile Lys
675 680 685

Leu Thr Phe Glu Tyr Leu Lys Ser Ser Val Glu Lys Gly Asp Lys Val
690 695 700

Ser Arg Glu Lys Met His Asn Ala Ser Thr Leu Ala Gly Met Ala Phe
705 710 715 720

Ala Asn Ala Phe Leu Gly Ile Ala His Ser Ile Ala His Lys Ile Gly
725 730 735

Gly Glu Tyr Gly Ile Pro His Gly Arg Ala Asn Ala Ile Leu Leu Pro
740 745 750

His Ile Ile Arg Tyr Asn Ala Lys Asp Pro Gln Lys His Ala Leu Phe
755 760 765

Pro Lys Tyr Glu Phe Phe Arg Ala Asp Thr Asp Tyr Ala Asp Ile Ala
770 775 780

Lys Phe Leu Gly Leu Lys Gly Asn Thr Thr Glu Ala Leu Val Glu Ser
785 790 795 800

Leu Ala Lys Ala Val Tyr Glu Leu Gly Gln Ser Val Gly Ile Glu Met
805 810 815

Asn Leu Lys Ser Gln Gly Val Ser Glu Glu Glu Leu Asn Glu Ser Ile
820 825 830

Asp Arg Met Ala Glu Leu Ala Phe Glu Asp Gln Cys Thr Thr Ala Asn
835 840 845

Pro Lys Glu Ala Leu Ile Ser Glu Ile Lys Asp Ile Ile Gln Thr Ser
850 855 860

Tyr Asp Tyr Lys Gln
865

<210> 29

<211> 2607

5

<212> ADN

<213> Artificial

<220>

<223> secuencia optimizada

10

<400> 29

ES 2 607 894 T3

atgttgacca ttccagaaaa ggaaaacaga ggttccaagg aacaagaagt tgccatcatg	60
attgatgctt tagctgacaa aggtaagaag gctttggaag ctttgtccaa gaagtctcaa	120
gaagaaattg accacattgt ccaccaaagc tccttggtg ctggtgacca acacatgggt	180
ttggccaagt tggctcatga agaaaccggt agaggtatct acgaagacaa ggctatcaag	240
aacttatacg cctctgaata catctggaac tccatcaagg acaacaagac tgttggtatc	300
attggtgaag acaaagaaaa gggtttgacc tacggtgctg aaccaattgg tgtcatctgt	360
ggtgtcactc caaccaccaa cccaacttct accaccatct tcaaggctat gattgccatc	420
aagactggta acccaattat ttctgcttcc caccatctg ctcaagaatc ttccaagaga	480
gctgctgaag ttgttttggg agctgccatg aaggctgggt ctccaaagga tatcatccaa	540
tggattgaag ttccatccat tgaagctacc aagcaattga tgaaccacaa gggatttgct	600
ttagtcttgg ctaccggtgg ttctggtatg gtttaagtctg cttactccac tggtaaacca	660
gctttgggtg ttggtccagg taacgttcca tcttacatcg aaaagactgc tcatatcaag	720
cgtgctgtca acgatatcat cggttccaag actttcgata atggtatgat ctgtgcttct	780
gaacaagttg ttgtcattga caaggaaatc tacaaggatg tcaccaatga attcaaggct	840
caccaagctt acttcgtcaa gaaggacgaa ttacaaagat tagaaaacgc catcatgaac	900
gaacaaaaga ctggtatcaa gccagatata gttggttaagt ctgctggtga aattgctgaa	960
ttggccggtg tcccagttcc agaaaacacc aaattgatca ttgctgaaat ctccggtgtc	1020
ggttctgact acccattgct cagagaaaag ttgtctccag ttttggcttt agtcaaggct	1080
caatctacca agcaagcttt ccaaatctgt gaagacactt tgcacttcgg tggtttaggt	1140
cacactgctg ttatccacac tgaagacgaa actttgcaa aggatttcgg tctaagaatg	1200
aaggcttgtc gtgttttggg caaactcca tctgctgttg gtggtatcgg tgacatgtac	1260
aacgaattga ttccatcctt gactttgggt tgtggttctt acggtagaaa ctccatctcc	1320
cacaacgtct ctgctaccga tttgttgaac atcaagacca ttgccaagag aagaaacaac	1380
actcaaactc tcaaggttcc agctcaaatc tatttcgaag aaaacgctat catgtccttg	1440
accaccatgg acaagattga aaaggtcatg atcgtttgtg acccaggtat ggttgaattt	1500
ggttacacca aaaccgtcga aaacgtctta cgtcaaagaa ctgaacaacc tcaaatcaag	1560

ES 2 607 894 T3

```

atcttctctg aagttgaacc aaatccatcc accaactctg tctacaaggg tttggaatg 1620
atggctcgatt tccaaccaga caccatcatt gctttgggtg gtggttctgc catggatgct 1680
gccaaggcta tgtggatggt cttcgaacat ccagaaactt ctttcttcg tgccaagcaa 1740
aaattcttgg acattggtaa gagaacctac aagattggta tgccagaaaa cgccactttc 1800
atctgtattc caaccacttc tggtagctgt tctgaagtca ctccatttgc tgttatcact 1860
gactctgaaa ccaacgtcaa ataccattg gctgatttcg ctttgactcc agatgtcgcc 1920
atcattgacc ctcaatttgt catgtccgtc caaaatctg tcaactgctga taccggtatg 1980
gacgttttga ctcacgctat ggaatcttac gtttctgtca tggcctccga ttaccacaga 2040
ggtttgtccc tacaagctat caaattgacc tttgaatact tgaaatcttc cgttgaaaaa 2100
ggtgacaagg tttccagaga aaagatgcac aacgcttcta ctttggccgg tatggccttt 2160
gctaacgctt tcttgggtat tgctcactcc attgctcaca aaattgggtg tgaatacggg 2220
attccacatg gtagagctaa cgccatcttg ttgcctcaca tcatcagata caacgccaa 2280
gacctcaaaa agcacgcttt gttcccaaag tacgaattct tcagagctga caccgattac 2340
gctgatatcg ccaagttctt aggtttgaaa ggtaacacca ctgaagcttt ggttgaatct 2400
ttggccaagg ctgtctacga attaggtcaa tctgttggtg ttgaaatgaa cttgaaatct 2460
caaggtgtct ctgaagaaga attgaacgaa tccattgaca gaatggctga attggctttc 2520
gaagaccaat gtaccactgc caacccaaag gaagctttga tttctgaaat caaggatatc 2580
atccaaactt cttacgacta caagcag 2607

```

<210> 30

<211> 392

5

<212> PRT

<213> Clostridium acetobutylicum

<400> 30

```

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

```

```

Tyr Gly Lys Ser Leu Lys Asp Val Pro Ala Val Asp Leu Gly Ala Thr
          20           25           30

```

```

Ala Ile Lys Glu Ala Val Lys Lys Ala Gly Ile Lys Pro Glu Asp Val
          35           40           45

```

10

```

Asn Glu Val Ile Leu Gly Asn Val Leu Gln Ala Gly Leu Gly Gln Asn

```


ES 2 607 894 T3

Gly Tyr Gly Pro Phe Tyr Ala Thr Lys Ala Ala Ile Glu Lys Ala Gly
 290 295 300

Trp Thr Val Asp Glu Leu Asp Leu Ile Glu Ser Asn Glu Ala Phe Ala
 305 310 315 320

Ala Gln Ser Leu Ala Val Ala Lys Asp Leu Lys Phe Asp Met Asn Lys
 325 330 335

Val Asn Val Asn Gly Gly Ala Ile Ala Leu Gly His Pro Ile Gly Ala
 340 345 350

Ser Gly Ala Arg Ile Leu Val Thr Leu Val His Ala Met Gln Lys Arg
 355 360 365

Asp Ala Lys Lys Gly Leu Ala Thr Leu Cys Ile Gly Gly Gly Gln Gly
 370 375 380

Thr Ala Ile Leu Leu Glu Lys Cys
 385 390

<210> 31

<211> 282

5

<212> PRT

<213> Clostridium acetobutylicum

<400> 31

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

Ala Gln Ala Phe Ala Ala Lys Gly Phe Glu Val Val Leu Arg Asp Ile
 20 25 30

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

Ser Lys Leu Val Lys Lys Gly Lys Ile Glu Glu Ala Thr Lys Val Glu
 50 55 60

Ile Leu Thr Arg Ile Ser Gly Thr Val Asp Leu Asn Met Ala Ala Asp
 65 70 75 80

10

ES 2 607 894 T3

Cys Asp Leu Val Ile Glu Ala Ala Val Glu Arg Met Asp Ile Lys Lys
 85 90 95

Gln Ile Phe Ala Asp Leu Asp Asn Ile Cys Lys Pro Glu Thr Ile Leu
 100 105 110

Ala Ser Asn Thr Ser Ser Leu Ser Ile Thr Glu Val Ala Ser Ala Thr
 115 120 125

Lys Arg Pro Asp Lys Val Ile Gly Met His Phe Phe Asn Pro Ala Pro
 130 135 140

Val Met Lys Leu Val Glu Val Ile Arg Gly Ile Ala Thr Ser Gln Glu
 145 150 155 160

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

Val Glu Val Ala Glu Ala Pro Gly Phe Val Val Asn Arg Ile Leu Ile
 180 185 190

Pro Met Ile Asn Glu Ala Val Gly Ile Leu Ala Glu Gly Ile Ala Ser
 195 200 205

Val Glu Asp Ile Asp Lys Ala Met Lys Leu Gly Ala Asn His Pro Met
 210 215 220

Gly Pro Leu Glu Leu Gly Asp Phe Ile Gly Leu Asp Ile Cys Leu Ala
 225 230 235 240

Ile Met Asp Val Leu Tyr Ser Glu Thr Gly Asp Ser Lys Tyr Arg Pro
 245 250 255

His Thr Leu Leu Lys Lys Tyr Val Arg Ala Gly Trp Leu Gly Arg Lys
 260 265 270

Ser Gly Lys Gly Phe Tyr Asp Tyr Ser Lys
 275 280

<210> 32

<211> 261

5

<212> PRT

<213> Clostridium acetobutylicum

<400> 32

ES 2 607 894 T3

Met Glu Leu Asn Asn Val Ile Leu Glu Lys Glu Gly Lys Val Ala Val
 1 5 10 15

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

Leu Lys Glu Met Asp Tyr Val Ile Gly Glu Ile Glu Asn Asp Ser Glu
 35 40 45

Val Leu Ala Val Ile Leu Thr Gly Ala Gly Glu Lys Ser Phe Val Ala
 50 55 60

Gly Ala Asp Ile Ser Glu Met Lys Glu Met Asn Thr Ile Glu Gly Arg
 65 70 75 80

Lys Phe Gly Ile Leu Gly Asn Lys Val Phe Arg Arg Leu Glu Leu Leu
 85 90 95

Glu Lys Pro Val Ile Ala Ala Val Asn Gly Phe Ala Leu Gly Gly Gly
 100 105 110

Cys Glu Ile Ala Met Ser Cys Asp Ile Arg Ile Ala Ser Ser Asn Ala
 115 120 125

Arg Phe Gly Gln Pro Glu Val Gly Leu Gly Ile Thr Pro Gly Phe Gly
 130 135 140

Gly Thr Gln Arg Leu Ser Arg Leu Val Gly Met Gly Met Ala Lys Gln
 145 150 155 160

Leu Ile Phe Thr Ala Gln Asn Ile Lys Ala Asp Glu Ala Leu Arg Ile
 165 170 175

Gly Leu Val Asn Lys Val Val Glu Pro Ser Glu Leu Met Asn Thr Ala
 180 185 190

Lys Glu Ile Ala Asn Lys Ile Val Ser Asn Ala Pro Val Ala Val Lys
 195 200 205

Leu Ser Lys Gln Ala Ile Asn Arg Gly Met Gln Cys Asp Ile Asp Thr
 210 215 220

Ala Leu Ala Phe Glu Ser Glu Ala Phe Gly Glu Cys Phe Ser Thr Glu
 225 230 235 240

Asp Gln Lys Asp Ala Met Thr Ala Phe Ile Glu Lys Arg Lys Ile Glu
 245 250 255

Gly Phe Lys Asn Arg
 260

<210> 33

5

<211> 379

ES 2 607 894 T3

<212> PRT

<213> Clostridium acetobutylicum

<400> 33

5

Met Asp Phe Asn Leu Thr Arg Glu Gln Glu Leu Val Arg Gln Met Val
1 5 10 15

Arg Glu Phe Ala Glu Asn Glu Val Lys Pro Ile Ala Ala Glu Ile Asp
20 25 30

Glu Thr Glu Arg Phe Pro Met Glu Asn Val Lys Lys Met Gly Gln Tyr
35 40 45

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

Asp Val Leu Ser Tyr Ile Ile Ala Val Glu Glu Leu Ser Lys Val Cys
65 70 75 80

Gly Thr Thr Gly Val Ile Leu Ser Ala His Thr Ser Leu Cys Ala Ser
85 90 95

Leu Ile Asn Glu His Gly Thr Glu Glu Gln Lys Gln Lys Tyr Leu Val
100 105 110

Pro Leu Ala Lys Gly Glu Lys Ile Gly Ala Tyr Gly Leu Thr Glu Pro
115 120 125

Asn Ala Gly Thr Asp Ser Gly Ala Gln Gln Thr Val Ala Val Leu Glu
130 135 140

Gly Asp His Tyr Val Ile Asn Gly Ser Lys Ile Phe Ile Thr Asn Gly

ES 2 607 894 T3

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

<210> 34

<211> 858

5 <212> PRT

<213> Clostridium acetobutylicum

<400> 34

ES 2 607 894 T3

Met	Lys	Val	Thr	Asn	Gln	Lys	Glu	Leu	Lys	Gln	Lys	Leu	Asn	Glu	Leu
1				5					10					15	
Arg	Glu	Ala	Gln	Lys	Lys	Phe	Ala	Thr	Tyr	Thr	Gln	Glu	Gln	Val	Asp
			20					25						30	
Lys	Ile	Phe	Lys	Gln	Cys	Ala	Ile	Ala	Ala	Ala	Lys	Glu	Arg	Ile	Asn
		35					40						45		
Leu	Ala	Lys	Leu	Ala	Val	Glu	Glu	Thr	Gly	Ile	Gly	Leu	Val	Glu	Asp
	50					55					60				
Lys	Ile	Ile	Lys	Asn	His	Phe	Ala	Ala	Glu	Tyr	Ile	Tyr	Asn	Lys	Tyr
65					70					75					80
Lys	Asn	Glu	Lys	Thr	Cys	Gly	Ile	Ile	Asp	His	Asp	Asp	Ser	Leu	Gly
				85					90					95	
Ile	Thr	Lys	Val	Ala	Glu	Pro	Ile	Gly	Ile	Val	Ala	Ala	Ile	Val	Pro
			100					105						110	
Thr	Thr	Asn	Pro	Thr	Ser	Thr	Ala	Ile	Phe	Lys	Ser	Leu	Ile	Ser	Leu
		115					120						125		
Lys	Thr	Arg	Asn	Ala	Ile	Phe	Phe	Ser	Pro	His	Pro	Arg	Ala	Lys	Lys
	130					135					140				
Ser	Thr	Ile	Ala	Ala	Ala	Lys	Leu	Ile	Leu	Asp	Ala	Ala	Val	Lys	Ala
145					150					155					160
Gly	Ala	Pro	Lys	Asn	Ile	Ile	Gly	Trp	Ile	Asp	Glu	Pro	Ser	Ile	Glu
				165					170					175	
Leu	Ser	Gln	Asp	Leu	Met	Ser	Glu	Ala	Asp	Ile	Ile	Leu	Ala	Thr	Gly
			180						185				190		

ES 2 607 894 T3

Gly Pro Ser Met Val Lys Ala Ala Tyr Ser Ser Gly Lys Pro Ala Ile
 195 200 205

Gly Val Gly Ala Gly Asn Thr Pro Ala Ile Ile Asp Glu Ser Ala Asp
 210 215 220

Ile Asp Met Ala Val Ser Ser Ile Ile Leu Ser Lys Thr Tyr Asp Asn
 225 230 235 240

Gly Val Ile Cys Ala Ser Glu Gln Ser Ile Leu Val Met Asn Ser Ile
 245 250 255

Tyr Glu Lys Val Lys Glu Glu Phe Val Lys Arg Gly Ser Tyr Ile Leu
 260 265 270

Asn Gln Asn Glu Ile Ala Lys Ile Lys Glu Thr Met Phe Lys Asn Gly
 275 280 285

Ala Ile Asn Ala Asp Ile Val Gly Lys Ser Ala Tyr Ile Ile Ala Lys
 290 295 300

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

Val Gln Ser Val Glu Lys Ser Glu Leu Phe Ser His Glu Lys Leu Ser
 325 330 335

Pro Val Leu Ala Met Tyr Lys Val Lys Asp Phe Asp Glu Ala Leu Lys
 340 345 350

Lys Ala Gln Arg Leu Ile Glu Leu Gly Gly Ser Gly His Thr Ser Ser
 355 360 365

Leu Tyr Ile Asp Ser Gln Asn Asn Lys Asp Lys Val Lys Glu Phe Gly
 370 375 380

Leu Ala Met Lys Thr Ser Arg Thr Phe Ile Asn Met Pro Ser Ser Gln
 385 390 395 400

Gly Ala Ser Gly Asp Leu Tyr Asn Phe Ala Ile Ala Pro Ser Phe Thr
 405 410 415

Leu Gly Cys Gly Thr Trp Gly Gly Asn Ser Val Ser Gln Asn Val Glu

ES 2 607 894 T3

420	425	430
Pro Lys His Leu Leu Asn Ile Lys Ser Val Ala Glu Arg Arg Glu Asn 435 440 445		
Met Leu Trp Phe Lys Val Pro Gln Lys Ile Tyr Phe Lys Tyr Gly Cys 450 455 460		
Leu Arg Phe Ala Leu Lys Glu Leu Lys Asp Met Asn Lys Lys Arg Ala 465 470 475 480		
Phe Ile Val Thr Asp Lys Asp Leu Phe Lys Leu Gly Tyr Val Asn Lys 485 490 495		
Ile Thr Lys Val Leu Asp Glu Ile Asp Ile Lys Tyr Ser Ile Phe Thr 500 505 510		
Asp Ile Lys Ser Asp Pro Thr Ile Asp Ser Val Lys Lys Gly Ala Lys 515 520 525		
Glu Met Leu Asn Phe Glu Pro Asp Thr Ile Ile Ser Ile Gly Gly Gly 530 535 540		
Ser Pro Met Asp Ala Ala Lys Val Met His Leu Leu Tyr Glu Tyr Pro 545 550 555 560		
Glu Ala Glu Ile Glu Asn Leu Ala Ile Asn Phe Met Asp Ile Arg Lys 565 570 575		
Arg Ile Cys Asn Phe Pro Lys Leu Gly Thr Lys Ala Ile Ser Val Ala 580 585 590		
Ile Pro Thr Thr Ala Gly Thr Gly Ser Glu Ala Thr Pro Phe Ala Val 595 600 605		
Ile Thr Asn Asp Glu Thr Gly Met Lys Tyr Pro Leu Thr Ser Tyr Glu 610 615 620		
Leu Thr Pro Asn Met Ala Ile Ile Asp Thr Glu Leu Met Leu Asn Met 625 630 635 640		
Pro Arg Lys Leu Thr Ala Ala Thr Gly Ile Asp Ala Leu Val His Ala 645 650 655		

ES 2 607 894 T3

Ile Glu Ala Tyr Val Ser Val Met Ala Thr Asp Tyr Thr Asp Glu Leu
 660 665 670

Ala Leu Arg Ala Ile Lys Met Ile Phe Lys Tyr Leu Pro Arg Ala Tyr
 675 680 685

Lys Asn Gly Thr Asn Asp Ile Glu Ala Arg Glu Lys Met Ala His Ala
 690 695 700

Ser Asn Ile Ala Gly Met Ala Phe Ala Asn Ala Phe Leu Gly Val Cys
 705 710 715 720

His Ser Met Ala His Lys Leu Gly Ala Met His His Val Pro His Gly
 725 730 735

Ile Ala Cys Ala Val Leu Ile Glu Glu Val Ile Lys Tyr Asn Ala Thr
 740 745 750

Asp Cys Pro Thr Lys Gln Thr Ala Phe Pro Gln Tyr Lys Ser Pro Asn
 755 760 765

Ala Lys Arg Lys Tyr Ala Glu Ile Ala Glu Tyr Leu Asn Leu Lys Gly
 770 775 780

Thr Ser Asp Thr Glu Lys Val Thr Ala Leu Ile Glu Ala Ile Ser Lys
 785 790 795 800

Leu Lys Ile Asp Leu Ser Ile Pro Gln Asn Ile Ser Ala Ala Gly Ile
 805 810 815

Asn Lys Lys Asp Phe Tyr Asn Thr Leu Asp Lys Met Ser Glu Leu Ala
 820 825 830

Phe Asp Asp Gln Cys Thr Thr Ala Asn Pro Arg Tyr Pro Leu Ile Ser
 835 840 845

Glu Leu Lys Asp Ile Tyr Ile Lys Ser Phe
 850 855

<210> 35

<211> 862

5

<212> PRT

<213> Clostridium acetobutylicum

<400> 35

ES 2 607 894 T3

Met Lys Val Thr Thr Val Lys Glu Leu Asp Glu Lys Leu Lys Val Ile
 1 5 10 15

Lys Glu Ala Gln Lys Lys Phe Ser Cys Tyr Ser Gln Glu Met Val Asp
 20 25 30

Glu Ile Phe Arg Asn Ala Ala Met Ala Ala Ile Asp Ala Arg Ile Glu
 35 40 45

Leu Ala Lys Ala Ala Val Leu Glu Thr Gly Met Gly Leu Val Glu Asp
 50 55 60

Lys Val Ile Lys Asn His Phe Ala Gly Glu Tyr Ile Tyr Asn Lys Tyr
 65 70 75 80

Lys Asp Glu Lys Thr Cys Gly Ile Ile Glu Arg Asn Glu Pro Tyr Gly
 85 90 95

Ile Thr Lys Ile Ala Glu Pro Ile Gly Val Val Ala Ala Ile Ile Pro
 100 105 110

Val Thr Asn Pro Thr Ser Thr Thr Ile Phe Lys Ser Leu Ile Ser Leu
 115 120 125

Lys Thr Arg Asn Gly Ile Phe Phe Ser Pro His Pro Arg Ala Lys Lys
 130 135 140

Ser Thr Ile Leu Ala Ala Lys Thr Ile Leu Asp Ala Ala Val Lys Ser
 145 150 155 160

Gly Ala Pro Glu Asn Ile Ile Gly Trp Ile Asp Glu Pro Ser Ile Glu
 165 170 175

Leu Thr Gln Tyr Leu Met Gln Lys Ala Asp Ile Thr Leu Ala Thr Gly
 180 185 190

Gly Pro Ser Leu Val Lys Ser Ala Tyr Ser Ser Gly Lys Pro Ala Ile
 195 200 205

Gly Val Gly Pro Gly Asn Thr Pro Val Ile Ile Asp Glu Ser Ala His

ES 2 607 894 T3

210	215	220														
Ile Lys Met Ala Val Ser Ser Ile Ile Leu Ser Lys Thr Tyr Asp Asn	230	235													240	
225																
Gly Val Ile Cys Ala Ser Glu Gln Ser Val Ile Val Leu Lys Ser Ile	245	250													255	
260																
Tyr Asn Lys Val Lys Asp Glu Phe Gln Glu Arg Gly Ala Tyr Ile Ile	265	270														
275																
Lys Lys Asn Glu Leu Asp Lys Val Arg Glu Val Ile Phe Lys Asp Gly	280	285														
290																
Ser Val Asn Pro Lys Ile Val Gly Gln Ser Ala Tyr Thr Ile Ala Ala	295	300														
305																
Met Ala Gly Ile Lys Val Pro Lys Thr Thr Arg Ile Leu Ile Gly Glu	310	315														
320																
Val Thr Ser Leu Gly Glu Glu Glu Pro Phe Ala His Glu Lys Leu Ser	325	330													335	
340																
Pro Val Leu Ala Met Tyr Glu Ala Asp Asn Phe Asp Asp Ala Leu Lys	345	350														
355																
Lys Ala Val Thr Leu Ile Asn Leu Gly Gly Leu Gly His Thr Ser Gly	360	365														
370																
Ile Tyr Ala Asp Glu Ile Lys Ala Arg Asp Lys Ile Asp Arg Phe Ser	375	380														
385																
Ser Ala Met Lys Thr Val Arg Thr Phe Val Asn Ile Pro Thr Ser Gln	390	395													400	
405																
Gly Ala Ser Gly Asp Leu Tyr Asn Phe Arg Ile Pro Pro Ser Phe Thr	410	415														
420																
Leu Gly Cys Gly Phe Trp Gly Gly Asn Ser Val Ser Glu Asn Val Gly	425	430														
435																
Pro Lys His Leu Leu Asn Ile Lys Thr Val Ala Glu Arg Arg Glu Asn	440	445														

ES 2 607 894 T3

Met Leu Trp Phe Arg Val Pro His Lys Val Tyr Phe Lys Phe Gly Cys
 450 455 460

Leu Gln Phe Ala Leu Lys Asp Leu Lys Asp Leu Lys Lys Lys Arg Ala
 465 470 475 480

Phe Ile Val Thr Asp Ser Asp Pro Tyr Asn Leu Asn Tyr Val Asp Ser
 485 490 495

Ile Ile Lys Ile Leu Glu His Leu Asp Ile Asp Phe Lys Val Phe Asn
 500 505 510

Lys Val Gly Arg Glu Ala Asp Leu Lys Thr Ile Lys Lys Ala Thr Glu
 515 520 525

Glu Met Ser Ser Phe Met Pro Asp Thr Ile Ile Ala Leu Gly Gly Thr
 530 535 540

Pro Glu Met Ser Ser Ala Lys Leu Met Trp Val Leu Tyr Glu His Pro
 545 550 555 560

Glu Val Lys Phe Glu Asp Leu Ala Ile Lys Phe Met Asp Ile Arg Lys
 565 570 575

Arg Ile Tyr Thr Phe Pro Lys Leu Gly Lys Lys Ala Met Leu Val Ala
 580 585 590

Ile Thr Thr Ser Ala Gly Ser Gly Ser Glu Val Thr Pro Phe Ala Leu
 595 600 605

Val Thr Asp Asn Asn Thr Gly Asn Lys Tyr Met Leu Ala Asp Tyr Glu
 610 615 620

Met Thr Pro Asn Met Ala Ile Val Asp Ala Glu Leu Met Met Lys Met
 625 630 635 640

Pro Lys Gly Leu Thr Ala Tyr Ser Gly Ile Asp Ala Leu Val Asn Ser
 645 650 655

Ile Glu Ala Tyr Thr Ser Val Tyr Ala Ser Glu Tyr Thr Asn Gly Leu
 660 665 670

ES 2 607 894 T3

Ala Leu Glu Ala Ile Arg Leu Ile Phe Lys Tyr Leu Pro Glu Ala Tyr
675 680 685

Lys Asn Gly Arg Thr Asn Glu Lys Ala Arg Glu Lys Met Ala His Ala
690 695 700

Ser Thr Met Ala Gly Met Ala Ser Ala Asn Ala Phe Leu Gly Leu Cys
705 710 715 720

His Ser Met Ala Ile Lys Leu Ser Ser Glu His Asn Ile Pro Ser Gly
725 730 735

Ile Ala Asn Ala Leu Leu Ile Glu Glu Val Ile Lys Phe Asn Ala Val
740 745 750

Asp Asn Pro Val Lys Gln Ala Pro Cys Pro Gln Tyr Lys Tyr Pro Asn
755 760 765

Thr Ile Phe Arg Tyr Ala Arg Ile Ala Asp Tyr Ile Lys Leu Gly Gly
770 775 780

Asn Thr Asp Glu Glu Lys Val Asp Leu Leu Ile Asn Lys Ile His Glu
785 790 795 800

Leu Lys Lys Ala Leu Asn Ile Pro Thr Ser Ile Lys Asp Ala Gly Val
805 810 815

Leu Glu Glu Asn Phe Tyr Ser Ser Leu Asp Arg Ile Ser Glu Leu Ala
820 825 830

Leu Asp Asp Gln Cys Thr Gly Ala Asn Pro Arg Phe Pro Leu Thr Ser
835 840 845

Glu Ile Lys Glu Met Tyr Ile Asn Cys Phe Lys Lys Gln Pro
850 855 860

<210> 36

<211> 389

5 <212> PRT

<213> Clostridium acetobutylicum

<400> 36

10 Met Leu Ser Phe Asp Tyr Ser Ile Pro Thr Lys Val Phe Phe Gly Lys

ES 2 607 894 T3

1	5	10	15												
Gly	Lys	Ile	Asp	Val	Ile	Gly	Glu	Glu	Ile	Lys	Lys	Tyr	Gly	Ser	Arg
			20					25					30		
Val	Leu	Ile	Val	Tyr	Gly	Gly	Gly	Ser	Ile	Lys	Arg	Asn	Gly	Ile	Tyr
		35					40					45			
Asp	Arg	Ala	Thr	Ala	Ile	Leu	Lys	Glu	Asn	Asn	Ile	Ala	Phe	Tyr	Glu
	50					55					60				
Leu	Ser	Gly	Val	Glu	Pro	Asn	Pro	Arg	Ile	Thr	Thr	Val	Lys	Lys	Gly
65					70					75					80
Ile	Glu	Ile	Cys	Arg	Glu	Asn	Asn	Val	Asp	Leu	Val	Leu	Ala	Ile	Gly
				85					90					95	
Gly	Gly	Ser	Ala	Ile	Asp	Cys	Ser	Lys	Val	Ile	Ala	Ala	Gly	Val	Tyr
			100					105					110		
Tyr	Asp	Gly	Asp	Thr	Trp	Asp	Met	Val	Lys	Asp	Pro	Ser	Lys	Ile	Thr
		115					120					125			
Lys	Val	Leu	Pro	Ile	Ala	Ser	Ile	Leu	Thr	Leu	Ser	Ala	Thr	Gly	Ser
	130					135					140				
Glu	Met	Asp	Gln	Ile	Ala	Val	Ile	Ser	Asn	Met	Glu	Thr	Asn	Glu	Lys
145					150					155					160
Leu	Gly	Val	Gly	His	Asp	Asp	Met	Arg	Pro	Lys	Phe	Ser	Val	Leu	Asp
				165					170					175	
Pro	Thr	Tyr	Thr	Phe	Thr	Val	Pro	Lys	Asn	Gln	Thr	Ala	Ala	Gly	Thr
			180					185					190		
Ala	Asp	Ile	Met	Ser	His	Thr	Phe	Glu	Ser	Tyr	Phe	Ser	Gly	Val	Glu
		195					200					205			
Gly	Ala	Tyr	Val	Gln	Asp	Gly	Ile	Ala	Glu	Ala	Ile	Leu	Arg	Thr	Cys
	210					215					220				
Ile	Lys	Tyr	Gly	Lys	Ile	Ala	Met	Glu	Lys	Thr	Asp	Asp	Tyr	Glu	Ala
225					230					235					240

ES 2 607 894 T3

Arg Ala Asn Leu Met Trp Ala Ser Ser Leu Ala Ile Asn Gly Leu Leu
 245 250 255

Ser Leu Gly Lys Asp Arg Lys Trp Ser Cys His Pro Met Glu His Glu
 260 265 270

Leu Ser Ala Tyr Tyr Asp Ile Thr His Gly Val Gly Leu Ala Ile Leu
 275 280 285

Thr Pro Asn Trp Met Glu Tyr Ile Leu Asn Asp Asp Thr Leu His Lys
 290 295 300

Phe Val Ser Tyr Gly Ile Asn Val Trp Gly Ile Asp Lys Asn Lys Asp
 305 310 315 320

Asn Tyr Glu Ile Ala Arg Glu Ala Ile Lys Asn Thr Arg Glu Tyr Phe
 325 330 335

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

Asp Lys Leu Glu Leu Met Ala Lys Gln Ala Val Arg Asn Ser Gly Gly
 355 360 365

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

Phe Lys Lys Ser Tyr
 385

<210> 37

<211> 390

5

<212> PRT

<213> Clostridium acetobutylicum

<400> 37

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

Asp Lys Ile Asn Val Leu Gly Arg Glu Leu Lys Lys Tyr Gly Ser Lys
 20 25 30

10

ES 2 607 894 T3

Val Leu Ile Val Tyr Gly Gly Gly Ser Ile Lys Arg Asn Gly Ile Tyr
 35 40 45

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

Leu Ala Gly Val Glu Pro Asn Pro Arg Val Thr Thr Val Glu Lys Gly
 65 70 75 80

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

Gly Gly Ser Ala Ile Asp Cys Ala Lys Val Ile Ala Ala Ala Cys Glu
 100 105 110

Tyr Asp Gly Asn Pro Trp Asp Ile Val Leu Asp Gly Ser Lys Ile Lys
 115 120 125

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

Glu Met Asp Thr Trp Ala Val Ile Asn Asn Met Asp Thr Asn Glu Lys
 145 150 155 160

Leu Ile Ala Ala His Pro Asp Met Ala Pro Lys Phe Ser Ile Leu Asp
 165 170 175

Pro Thr Tyr Thr Tyr Thr Val Pro Thr Asn Gln Thr Ala Ala Gly Thr
 180 185 190

Ala Asp Ile Met Ser His Ile Phe Glu Val Tyr Phe Ser Asn Thr Lys
 195 200 205

Thr Ala Tyr Leu Gln Asp Arg Met Ala Glu Ala Leu Leu Arg Thr Cys
 210 215 220

Ile Lys Tyr Gly Gly Ile Ala Leu Glu Lys Pro Asp Asp Tyr Glu Ala
 225 230 235 240

Arg Ala Asn Leu Met Trp Ala Ser Ser Leu Ala Ile Asn Gly Leu Leu
 245 250 255

Thr Tyr Gly Lys Asp Thr Asn Trp Ser Val His Leu Met Glu His Glu

ES 2 607 894 T3

260 265 270

Leu Ser Ala Tyr Tyr Asp Ile Thr His Gly Val Gly Leu Ala Ile Leu
 275 280 285

Thr Pro Asn Trp Met Glu Tyr Ile Leu Asn Asn Asp Thr Val Tyr Lys
 290 295 300

Phe Val Glu Tyr Gly Val Asn Val Trp Gly Ile Asp Lys Glu Lys Asn
 305 310 315 320

His Tyr Asp Ile Ala His Gln Ala Ile Gln Lys Thr Arg Asp Tyr Phe
 325 330 335

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

Glu Glu Lys Leu Asp Ile Met Ala Lys Glu Ser Val Lys Leu Thr Gly
 355 360 365

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

Ile Phe Lys Lys Ser Val
 385 390

<210> 38

<211> 336

5

<212> PRT

<213> Clostridium acetobutylicum

<400> 38

Met Asn Lys Ala Asp Tyr Lys Gly Val Trp Val Phe Ala Glu Gln Arg
 1 5 10 15

Asp Gly Glu Leu Gln Lys Val Ser Leu Glu Leu Leu Gly Lys Gly Lys
 20 25 30

Glu Met Ala Glu Lys Leu Gly Val Glu Leu Thr Ala Val Leu Leu Gly
 35 40 45

His Asn Thr Glu Lys Met Ser Lys Asp Leu Leu Ser His Gly Ala Asp
 50 55 60

10

ES 2 607 894 T3

Lys Val Leu Ala Ala Asp Asn Glu Leu Leu Ala His Phe Ser Thr Asp
65 70 75 80

Gly Tyr Ala Lys Val Ile Cys Asp Leu Val Asn Glu Arg Lys Pro Glu
85 90 95

Ile Leu Phe Ile Gly Ala Thr Phe Ile Gly Arg Asp Leu Gly Pro Arg
100 105 110

Ile Ala Ala Arg Leu Ser Thr Gly Leu Thr Ala Asp Cys Thr Ser Leu
115 120 125

Asp Ile Asp Val Glu Asn Arg Asp Leu Leu Ala Thr Arg Pro Ala Phe
130 135 140

Gly Gly Asn Leu Ile Ala Thr Ile Val Cys Ser Asp His Arg Pro Gln
145 150 155 160

Met Ala Thr Val Arg Pro Gly Val Phe Glu Lys Leu Pro Val Asn Asp
165 170 175

Ala Asn Val Ser Asp Asp Lys Ile Glu Lys Val Ala Ile Lys Leu Thr
180 185 190

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

Asp Ile Ala Asp Ile Gly Glu Ala Lys Val Leu Val Ala Gly Gly Arg
210 215 220

Gly Val Gly Ser Lys Glu Asn Phe Glu Lys Leu Glu Glu Leu Ala Ser
225 230 235 240

Leu Leu Gly Gly Thr Ile Ala Ala Ser Arg Ala Ala Ile Glu Lys Glu
245 250 255

Trp Val Asp Lys Asp Leu Gln Val Gly Gln Thr Gly Lys Thr Val Arg
260 265 270

Pro Thr Leu Tyr Ile Ala Cys Gly Ile Ser Gly Ala Ile Gln His Leu
275 280 285

Ala Gly Met Gln Asp Ser Asp Tyr Ile Ile Ala Ile Asn Lys Asp Val
290 295 300

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

Asn Lys Val Val Pro Glu Leu Ile Ala Gln Val Lys Ala Ala Asn Asn
325 330 335

<210> 39

5

<211> 259

ES 2 607 894 T3

<212> PRT

<213> Clostridium acetobutylicum

<400> 39

5

```

Met Asn Ile Val Val Cys Leu Lys Gln Val Pro Asp Thr Ala Glu Val
 1                               10                15

Arg Ile Asp Pro Val Lys Gly Thr Leu Ile Arg Glu Gly Val Pro Ser
 20                25                30

Ile Ile Asn Pro Asp Asp Lys Asn Ala Leu Glu Glu Ala Leu Val Leu
 35                40                45

Lys Asp Asn Tyr Gly Ala His Val Thr Val Ile Ser Met Gly Pro Pro
 50                55                60

Gln Ala Lys Asn Ala Leu Val Glu Ala Leu Ala Met Gly Ala Asp Glu
 65                70                75                80

Ala Val Leu Leu Thr Asp Arg Ala Phe Gly Gly Ala Asp Thr Leu Ala
 85                90                95

Thr Ser His Thr Ile Ala Ala Gly Ile Lys Lys Leu Lys Tyr Asp Ile
 100               105               110

Val Phe Ala Gly Arg Gln Ala Ile Asp Gly Asp Thr Ala Gln Val Gly
 115                120                125

Pro Glu Ile Ala Glu His Leu Gly Ile Pro Gln Val Thr Tyr Val Glu
 130                135                140

Lys Val Glu Val Asp Gly Asp Thr Leu Lys Ile Arg Lys Ala Trp Glu
 145                150                155                160

```

ES 2 607 894 T3

Asp Gly Tyr Glu Val Val Glu Val Lys Thr Pro Val Leu Leu Thr Ala
 165 170 175

Ile Lys Glu Leu Asn Val Pro Arg Tyr Met Ser Val Glu Lys Ile Phe
 180 185 190

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

Val Asp Lys Ala Asn Leu Gly Leu Lys Gly Ser Pro Thr Lys Val Lys
 210 215 220

Lys Ser Ser Thr Lys Glu Val Lys Gly Gln Gly Glu Val Ile Asp Lys
 225 230 235 240

Pro Val Lys Glu Ala Ala Ala Tyr Val Val Ser Lys Leu Lys Glu Glu
 245 250 255

His Tyr Ile

<210> 40

<211> 5976

5

<212> ADN

<213> Artificial

<220>

<223> plásmido YEplac112PtdhTadh

10

<400> 40

```

gcccggggga tccactagtt ctagaatccg tcgaaactaa gttctggtgt tttaaaacta    60
aaaaaaagac taactataaa agtagaatth aagaagttha agaatagat ttacagaatt    120
acaatcaata cctaccgtct ttatatactt attagtcaag taggggaata atttcagggg    180
actggtttca accttttttt tcagcttttt ccaaatcaga gagagcagaa ggtaatagaa    240
ggtgtaagaa aatgagatag atacatgcgt ggggtcaattg ccttgtgtca tcatttactc    300
caggcaggtt gcactactcc attgaggttg tgcccgtttt ttgcctgttt gtgcccctgt    360
tctctgtagt tgcgctaaga gaatggacct atgaactgat ggttggtgaa gaaaacaata    420
ttttggtgct gggattcttt tttttctggt atgccagctt aaaagcggg ctccattata    480
tttagtgat gccaggaata aactgttcac ccagacacct acgatgttat atattctgtg    540
    
```


ES 2 607 894 T3

taaccgccc cctatTTTgG gcatgtacgg gttacagcag aattaaaagg ctaatTTTT	600
gactaaataa agttaggaaa atcactacta ttaattatTT acgtattcTT tgaaatggcg	660
agtattgata atgataaact gagctcgaat tcaactggccg tcgTTTTaca acgtcgtgac	720
tgggaaaacc ctggcgTTac ccaacttaat cgccttgCag cacatcccc tttcgccagc	780
tggcgtaata gcgaagaggc ccgcaccgat cgccttccc aacagttgCg cagcctgaat	840
ggcgaatggc gectgatgCg gtatTTTctc cttacgcate tgtgCggTat ttcacaccgc	900
atatatcgga tcgtactTgt tacccatcat tgaatTTTga acatccgaac ctgggagTtt	960
tcctgaaac agatagtata tttgaaCctg tataataata tatagtctag cgctttacgg	1020
aagacaatgt atgtatTTcg gttcctggag aaactattgc atctattgca taggtaatct	1080
tgcacgtcgc atccccggtt cattttctgc gtttccatct tgcacttcaa tagcatatct	1140
ttgttaacga agcatctgtg cttcatTTTg tagaacaaaa atgcaacgcg agagcGctaa	1200
TTTTTcaaac aaagaatctg agctgcattt ttacagaaca gaaatgcaac gcgaaagcgc	1260
tatTTTacca acgaagaatc tgtgcttcat tTTTgtaaaa caaaaatgca acgcgagagc	1320
gtaatTTTT caaacaaaga atctgagctg cattTTTaca gaacagaaat gcaacgcgag	1380
agcGctatTT taccaacaaa gaatctatac ttctTTTTg ttctacaaaa atgcatcccg	1440
agagcGctat tTTctaaaca aagcatctta gattactTTT tttctccttt gtgcGctcta	1500
taatgcagtc tcttgataac tTTTtgcact gtaggtccgt taaggTtaga agaagGctac	1560
tttggtgtct atTTTctctt ccataaaaa agcctgactc cacttcccgc gtttactgat	1620
tactagcGaa gctgcgggtg cattTTTtca agataaaggc atccccgatt atattctata	1680
ccgatgtgga ttgcgcatac tttgtgaaca gaaagtgata gcgttgatga ttcttcattg	1740
gtcagaaaaat tatgaacggt ttcttctatt ttgtctctat atactacgta taggaaatgt	1800
ttacattttc gtattgTTTT cgattcactc tatgaatagt tcttactaca atTTTTTgt	1860
ctaaagagta atactagaga taaacataaa aaatgtagag gtcgagTTta gatgcaagtt	1920
caaggagcga aaggTggatg ggtaggttat atagggatat agcacagaga tatatagcaa	1980
agagatactt ttgagcaatg tttgtggaag cggtatctgc aatatTTtag tagctcgTTa	2040
cagtccggtg cgtTTTTggt tTTTtgaag tgcgtcttca gagcGctTTT ggTTTTTcaa	2100
agcGctctga agttcctata ctttctagct agagaatagg aacttcgGaa taggaacttc	2160
aaagcgtttc cgaaaacgag cgcttccGaa aatgcaacgc gagctgcGca catacagctc	2220
actgttcacg tcgcacctat atctgcgtgt tgctgtata tatatataca tgagaagaac	2280

ES 2 607 894 T3

ggcatagtgc gtgtttatgc ttaa atgcgt acttatatgc gtctat tttat gtaggatgaa	2340
aggtagtcta gtacctcctg tgatattatc ccattccatg cggggtatcg tatgcttctt	2400
tcagcactac cctttagctg ttctatatgc tgccactcct caattggatt agtctcatcc	2460
ttcaatgcta tcatttcctt tgatattgga tcgatccgat gataagctgt caaacatgag	2520
aattgatctt ttatgcttgc ttttcaaaag gcttgcagge aagtgcacaa acaatactta	2580
aataaatact actcagtaat aacctatttc ttagcatttt tgacgaaatt tgctat ttttg	2640
ttagagtctt ttacaccatt tgtctccaca cctccgctta catcaacacc aataacgcca	2700
tttaatctaa gcgcatcacc aacattttct gccgtcagtc caccagctaa cataaaatgt	2760
aagctctcgg ggctctcttg ccttccaacc cagtcagaaa tcgagttcca atccaaaagt	2820
tcacctgtcc cacctgcttc tgaatcaaac aagggaataa acgaatgagg tttctgtgaa	2880
gctgcactga gtagtatgtt gcagcttttt ggaaatacga gtcttttaat aactggcaaa	2940
ccgaggaact cttggtatc ttgccacgac tcactctcat gcagttggac gatatcaatg	3000
ccgtaatcat tgaccagagc caaacatcc tccttaggtt gattacgaaa cagccaacc	3060
aagtatttcg gagtgcctga actattttta tatgctttta caagacttga aattttcctt	3120
gcaataaccg ggtcaattgt tctctttcta ttgggcacac atataatacc cagcaagtca	3180
gcatcggaat ctagtgcaca ttctgcggcc tctgtgctct gcaagccgca aactttcacc	3240
aatggaccag aactacctgt gaaattaata acagacatac tccaagctgc ctttgtgtgc	3300
ttaatcacgt atactcacgt gctcaatagt caccaatgcc ctcctcttg gccctctct	3360
tttctttttt cgaccgaatt aattcttgaa gacgaaaggg cctcgtgata cgcctat ttt	3420
tataggttaa tgtcatgata ataatggttt cttagacgtc aggtggcact tttcggggaa	3480
atgtgcgcgg aaccctatt tgtttat tctaaataca ttcaa atatg tatccgctca	3540
tgagacaata accctgataa atgcttcaat aatattgaaa aaggaagagt atgagtattc	3600
aacatttccg tgtcgcctt attccctttt ttgoggcatt ttgccttctt gtttttctc	3660
accagaaac gctggtgaaa gtaaaagatg ctgaagatca gttgggtgca cgagtgggtt	3720
acatcgaact ggatctcaac agcggtaaga tccttgagag ttttcgcccc gaagaacggt	3780
ttccaatgat gacactttt aaagttctgc tatgtggcgc ggtattatcc cgtattgacg	3840
ccgggcaaga gcaactcggc cgccgcatac actattctca gaatgacttg gttgagtact	3900
caccagtcac agaaaagcat cttacggatg gcatgacagt aagagaatta tgcagtctg	3960

ES 2 607 894 T3

ccataacat gagtgataac actgcgcca acttacttct gacaacgatc ggaggaccga 4020
 aggagctaac cgcttttttg cacaacatgg gggatcatgt aactcgcctt gatcgttggg 4080
 aaccggagct gaatgaagcc ataccaaacy acgagcgtga caccacgatg cctgtagcaa 4140
 tggcaacaac gttgcgcaaa ctattaactg gcgaactact tactctagct tcccggcaac 4200
 aattaataga ctggatggag gcggataaag ttgcaggacc acttctgcgc tcggcccttc 4260
 cggttggtg gtttattgct gataaatctg gagccgggtga gcgtgggtct cgcggtatca 4320
 ttgcagcact ggggccagat ggtaagccct cccgtatcgt agttatctac acgacgggga 4380
 gtcaggcaac tatggatgaa cgaaatagac agatcgctga gatagggtgc tcaactgatta 4440
 agcattggta actgtcagac caagtttact catatatact ttagattgat ttaaaacttc 4500
 atttttaatt taaaaggatc taggtgaaga tcctttttga taatctcatg accaaaatcc 4560
 cttaacgtga gttttcgttc cactgagcgt cagaccccgt agaaaagatc aaaggatctt 4620
 cttgagatec tttttttctg cgcgtaatct gctgcttga aacaaaaaaa ccaccgctac 4680
 cagcgggtgg ttgtttgccc gatcaagagc taccaactct ttttcogaag gtaactggct 4740
 tcagcagagc gcagatacca aataactgtcc ttctagtgtg gccgtagtta ggccaccact 4800
 tcaagaactc tgtagcaccg cctacatacc tcgctctgct aatcctgtta ccagtggctg 4860
 ctgccagtgg cgataagtcg tgtcctaccg ggttgactc aagacgatag ttaccggata 4920
 aggcgcagcg gtcgggctga acgggggggt cgtgcacaca gccagcttg gagcgaacga 4980
 cctacaccga actgagatac ctacagcgtg agctatgaga aagcgcacg cttcccgaag 5040
 ggagaaagc ggacaggtat ccggtaaagc gcagggtcgg aacaggagag cgcacgaggg 5100
 agcttccagg gggaaacgcc tggatcttt atagtcctgt cgggtttcgc cacctctgac 5160
 ttgagcgtcg atttttgtga tgctcgtcag gggggcggag cctatggaaa aacgccagca 5220
 acgggcctt tttacggttc ctggccttt gctggcctt tgctcacatg ttctttctg 5280
 cgttatccc tgattctgtg gataaccgta ttaccgcctt tgagttagct gataccgctc 5340
 gccgcagccg aacgaccgag cgcagcagat cagtgagcga ggaagcggaa gagcgcctaa 5400
 tacgcaaacc gcctctcccc gcgcgttggc cgattcatta atgcagctgg cagcagcgg 5460
 ttcccactg gaaagcgggc agtgagcga acgcaattaa tgtgagttag ctcaactcatt 5520
 aggcaccca ggctttacac tttatgcttc cggctcgtat gttgtgtgga attgtgagcg 5580
 gataacaatt tcacacagga aacagctatg accatgatta cgccaagctt aggcctgtgt 5640
 ggaagaacga ttacaacagc tgttgtcctc tgaggacata aaatacacac cgagattcat 5700
 caactcattg ctggagttag catatctaca attgggtgaa atggggagcg atttgcaggc 5760
 atttgcctcg catgcogga gaggtgtggt caataagagc gacctcatgc tatacctgag 5820
 aaagcaacct gacctacagc aaagagttac tcaagaataa gaattttcgt tttaaaacct 5880
 aagagtcact ttaaaatttg tatacactta ttttttttat aacttattta ataataaaaa 5940
 tcataaatca taagaaattc gctcagctcg actgca 5976

ES 2 607 894 T3

<211> 13286

<212> ADN

<213> Artificial

5

<220>

<223> pBOL34

<400> 41

```
aagcttgcac gectgcaggt cgaecggcgcg ccgggcccgt ttaaaccggcc ggccaagggtg      60
agacgcgcac aaccgctaga gtactttgaa gaggaaacag caataggggtt gctaccagta      120
taaatagaca ggtacataca aacttggaat tgggtgtctg tttgagtacg ctttcaattc      180
atctgggtgt gcactttatt atgttacaat atggaagga actttacact tctcctatgc      240
acatatatta attaaagtcc aatgctagta gagaaggggg gtaacacccc tccgcgctct      300
tttccgattt ttttctaaac cgtggaatat ttcggatata cttttggtgt ttccgggtgt      360
acaatatgga cttcctcttt tctggcaacc aaaccatac atcgggatcc ctataatacc      420
ttcgttggtc tccttaacat gtaggtggcg gaggggagat atacaataga acagatacca      480
gacaagacat aatgggctaa acaagactac accaattaca ctgcctcatt gatggtggta      540
cataacgaac taatactgta gccctagact tgatagccat catcatatcg aagtttcaact      600
accctttttc catttgccat ctattgaagt aataatagge gcatgcaact tcttttcttt      660
ttttttcttt tctctctccc ccgttggtgt ctcaccatat ccgcaatgac aaaaaaatga      720
tggaagacac taaaggaaaa aattaacgac aaagacagca ccaacagatg tcgttggtcc      780
agagctgatg aggggtatct cgaagcacac gaaacttttt ccttccttca ttcacgcaca      840
ctactctcta atgagcaacg gtatacggcc ttccttccag ttacttgaat ttgaaataaa      900
aaaaagtttg ctgtcttgct atcaagtata aatagacctg caattattaa tcttttgttt      960
cctcgtcatt gttctcgttc cctttcttcc ttgtttcttt ttctgcacaa tatttcaagc     1020
tataccaagc atacaatcaa ctatctcata tacaatgaag gaagttgtta ttgcttctgc     1080
```

10

ES 2 607 894 T3

tgtcagaact gccattgggt cttacggtaa gtctttgaag gacgtcccag ctgtcgactt	1140
gggtgctacc gccatcaagg aagctgtcaa gaaggctggt atcaagccag aagatgtaa	1200
cgaagttatc ttaggtaacg ttttgcaagc tggtttaggt caaaaccag ctcgtaagc	1260
ttctttcaag gctggttgc cagttgaaat tccagccatg accatcaaca aggtttgtgg	1320
ttctggtttg agaactgttt ctttgctgc tcaaatcatc aaggctggtg acgctgatgt	1380
catcattgct ggtggtatgg aaaacatgtc cagagctcca tacttggtca acaatgctag	1440
atggggttac agaatgggta acgccaagtt cgtcgatgaa atgatcactg acggtttatg	1500
ggacgctttc aacgactacc acatgggtat cactgctgaa aacattgctg aaagatggaa	1560
catctccaga gaagaacaag atgaatttgc tttggcttct caaaagaag ctgaagaagc	1620
catcaaactt ggtcaattca aggacgaaat tgtcccagtt gtcatcaagg gtagaaagg	1680
tgaaccggtt gtcgacaccg atgaacacc aagattcgggt tccaccattg aaggtttggc	1740
caagttgaaa ccagctttca agaaggatgg taccgctcact gctggtaacg cttccggttt	1800
gaacgactgt gctgctgttt tggttatcat gtctgctgaa aaggccaagg aattgggtgt	1860
caagccattg gccaagattg tctcctacgg ttctgctggt gttgaccag ccatcatggg	1920
ttacggctct ttctacgcta ccaaggctgc tatcgaaaag gctggttggga ccggtgacga	1980
attggatttg attgaatcca acgaagcttt cgctgctcaa tctttggctg ttgccaagga	2040
cttgaaattc gacatgaaca aggtcaacgt taacgggtgt gccattgctt tgggtcacc	2100
aattgggtgt tccggtgcca gaatcttgggt tacttttagtc cacgctatgc aaaagcgtga	2160
tgccaagaag ggtttggcta ctctatgtat cgggtggtgt caaggactg ccatcttatt	2220
ggaaaagtgt taggcccggt cataaaagca tcttgatgag gataatgatt ttttttgaa	2280
tatacataaa tactaccggt tttctgctag attttgtaa gacgtaaata agtacatatt	2340
actttttaag ccaagacaag attaagcatt aactttacc tttctcttc taagtttcaa	2400
tactagtat cactgtttaa aagttatggc gagaacgtcg gcggttaaaa tatattacc	2460
tgaacgtggt gaattgaagt tctaggatgg tttaaagatt tttcctttt gggaaataag	2520
taaacaatat attgctgctt ttgcaaacg cacataacca caatatgtga ctattggcaa	2580
agaacgcatt atcctttgaa gaggtggata ctgatactaa gagagtctct attccggctc	2640
cacttttagt ccagagatta cttgtcttct tacgtatcag aacaagaaag catttccaaa	2700
gtaattgcat ttgcccttga gcagtatata tatactaaga agtttaaca tttaaacgtg	2760
tgtgtgcatt atatatatta aaaattaaga attagactaa ataaagtgtt tctaaaaaaa	2820

ES 2 607 894 T3

tattaaagtt gaaatgtgcg tgttgtgaat tgtgctctat tagaataatt atgacttgtg 2880
tgcgtttcat attttaaaat aggaaataac caagaaagaa aaagtacat ccagagaaac 2940
caattataatc aaatcaaata aaacaaccag cttcgggtgtg tgtgtgtgtg tgaagctaag 3000
agttgatgcc atttaatcta aaaattttaa ggtgtgtgtg tggataaaat attagaatga 3060
caattcgaga tgaaatttta agcaaactct agtaggaaat aagcggctta ttcttgttgg 3120
ctcctaattc tttttagtgt atcagttccc attgataaaa aaattaaaat taaaattaga 3180
aaaattaaac cagaaaaatc aagttgatta aaatgtgaca aaaattatga ttaaatgcta 3240
cttcaacagg agcccgggcc tatttggagt agtcgtagaa acccttacca gactttctac 3300
ctaaccaacc agctctaacy tacttcttca ataaagtgtg aggtctgtac ttagagtcac 3360
cggtttcaga gtataagaca tccatgatgg ccaaacagat atccaaaccg atgaagtcac 3420
ctaattccaa tggaccatt ggggtggttag caccaatth catggccttg tcgatatctt 3480
caacagaagc aataccttca gccaaaatac cgacagcttc gttgatcatt ggaatcaaga 3540
ttctgttgac aacgaaacct ggagctttag caacttcaac tgggtctta ccaatggcaa 3600
tggagtttc cttgacagca tcgaaagttt cttgagaggt ggcaataacct ctgatgactt 3660
cgaccaactt catgactgga gctgggttga agaagtgcac accgataacc ttgtctggtc 3720
tcttgtagc agaagcaact tcagtgatgg acaagaaga agtgttgaa gccaaaatgg 3780
tttctggctt acagatgttg tccaaatcag caaagattg cttcttgatg tccattcttt 3840
caacggcagc tcaatgacc aaatcacagt cagcagccat gttcaagtca acagtaccgg 3900
agattctggg caagatttcg accttggttag cttcttcaat cttacccttc ttgaccaact 3960
tggacaagtt cttgttgatg aaatccaaac cacgggtcaac gaattcgtcc ttgatatctc 4020
tcaaaacaac ttcgaaacct ttggcagcga aagcttgagc aataccagaa cccatggtac 4080
cggcaccaat gacacaaacc ttcttcattt tgatttagtg tttgtgtgtt gataagcagt 4140
tgcttggtt tttatgaaaa atagctagaa ggaataaggg attacaagag agatgttaca 4200
agaagaagt aaaataaatt tgattaatat tgccattatc aaaagctatt tatatgttga 4260
aatcgtggag atcatgtgtg ccagaaaagg ccacagttc cggggagagg cataccttga 4320
ggtggetagg aatcacggag acctcttgac ttgcagggtg ggctagctag aattaagtga 4380
ggtgacaagg tttccataca gttttgacct tgagacgttg ctacttacga tttgcagtat 4440
gcaagtctca tgctgcaaac aaaagaggac cgctcaggta atcgetcaat tagtggacgt 4500

ES 2 607 894 T3

tatcaggggc gggagaggcg aaagtggttt ttggtggtgt aagtaaaggt cgtccaaata	4560
tgcaggtggt tgggtgctat cctagtgga gctcggatca gtagataacc cgcctagaag	4620
cggatTTTT cTTTTTTTT cttccttctt ttctgctatt atttcaaagc cTTTTgcgtc	4680
aagtaatgaa tatctggcgg ttccgcggta atgcgacaat ttgtgatatg cactcttaaa	4740
accccgccac gatgatcgca cgtgccggca tttatagacg acttttctgg ttgtcccgt	4800
tcacggcaca tgcattgcatc aatgaccgaa ttcaggttgc tactaacctat tgtgttgtgt	4860
tattgctgtg catgaggtgc tcaagtgcc gcggcatctg actagtggta actctagacg	4920
gcttcgatgc agagagttcc tcaaaatttt tcttttcaat tgtttgcctg gtttccgcgg	4980
cgtatatcag tttttggcga tatggtaacg cgataactta cggcaccttc acggtatag	5040
tcttttttaa aagtgactgt taattccagg attgaaagga agtgtcgaat agtatagtat	5100
gctttctagg ccggccggtt aaatgggcc gcggccggtt taaacggccg gcccttcct	5160
tttacagtgc ttcggaaaag cacagcgttg tccaaggga caatttttct tcaagttaat	5220
gcataagaaa tatctTTTT tatgttttagc taagtaaaag cagcttgag taaaaaaaa	5280
aatgagtaaa tttctogatg gattagtctc tcacaggtaa cataacaaaa accaagaaaa	5340
gcccgcttct gaaaactaca gttgacttgt atgctaaagg gccagactaa tgggaggaga	5400
aaaagaaacg aatgtatatg ctcatctaca ctctatatca ccatatggag gataagttgg	5460
gctgagcttc tgatccaatt tattctatcc attagtgtct gatatgtccc accagccaac	5520
acttgatagt atctactcgc cattcacttc cagcagcgc agtaggggtg ttgagcttag	5580
taaaaatgtg cgcaccacaa gcctacatga ctccacgtca catgaaacca caccgtggg	5640
ccttgttgcg ctaggaatag gatatgagc gaagacgctt ctgcttagta accacaccac	5700
atTTTcaggg ggtcagatctg cttgcttctt ttactgtcac gagcggcca taatcgcgct	5760
TTTTTTTTaa aaggcgcgag acagcaaaca ggaagctcgg gtttcaacct tcggagtgg	5820
cgcagatctg gagactggat ctttacaata cagtaaggca agccaccatc tgcttcttag	5880
gtgcatgca cggtatccac gtgcagaaca acatagtctg aagaaggggg ggaggagcat	5940
gttcattctc tgtagcagta agagcttggg gataatgacc aaaactggag tctcgaatc	6000
atataaatag acaatatatt ttcacacaat gagatttcta gtacagttct attctctctc	6060
ttgcataaat aagaaattca tcaagaactt ggtttgatat ttcaccaaca cacacaaaa	6120
acagtacttc actaaattta cacacaaaac aaaatggaat tgaacaacgt tatcttgaa	6180
aaggaagta aggttgcctg tgtcaccatc aacagaccaa aggctttgaa tgctttgaac	6240

ES 2 607 894 T3

tctgacactt tgaaggaaat ggactacgtc attggtgaaa ttgaaaacga ttctgaagtt	6300
ttggctgtca tcttgaccgg tgccggtgaa aagtctttcg ttgctgggtgc tgatatctct	6360
gaaatgaagg aaatgaacac cattgaaggt agaaagtctg gtatcttagg taacaaggtt	6420
ttcagaagat tggaaattgt ggaaaagcca gtcattgctg ctgtcaacgg ttctcgctttg	6480
ggtggtggtt gtgaaattgc catgtcctgt gacatcagaa ttgcttcttc taacgctcgt	6540
ttcgggtcaac cagaagtcgg tctaggtatc actccaggtt tcggtgggtac tcaaagatta	6600
tccagattgg ttggtatggg tatggccaag caattgatct tcaccgctca aaacatcaag	6660
gctgacgaag ctttgagaat tggtttagtc aacaaggttg ttgaaccatc tgaattgatg	6720
aacactgcca aggaaattgc taacaagatc gtctccaacg ctccagttgc tgtcaaattg	6780
tccaagcaag ccatcaacag aggtatgcaa tgtgatatcg acaccgcttt ggcctttgaa	6840
tctgaagctt tcggtgaatg tttctccact gaagaccaa aggatgctat gaccgctttc	6900
atcgaaaaga gaaagattga aggtttcaag aacaggtgat gagccgggc gcgaatttct	6960
tatgatttat gatttttatt attaaataag ttataaaaaa aataagtgtg taaaaatttt	7020
aaagtgactc ttaggtttta aaacgaaaat tcttattctt gagtaactct ttctgtagg	7080
tcaggttget ttctcaggta tagcatgagg tcgctcttat tgaccacacc tctaccgca	7140
tgccgagcaa atcgctgcaa atcgctccc atttcacca attgtagata tgctaactcc	7200
agcaatgagt tgatgaatct cgggtgtgat tttatgtcct cagaggacia cacctgttgt	7260
aatcgttctt ccacacggat ccacagccta gccttcagtt gggctctatc ttcacgtca	7320
ttcattgcat ctactagccc cttacctgag cttcaagacg ttatatcgct tttatgtatc	7380
atgatcttat cttgagatat gaatacataa atatatttac tcaagtgtat acgtgcatgc	7440
tttttttacg gtttaaacat ttaaattggc cgctctagag gatccccggg taccgagctc	7500
gggccagcgc ctactagtcc cggtaatttg aaaacaaacc cggctctcga gcggagatcc	7560
ggcgataatt accgcagaaa taaaccata cacgagacgt agaaccagcc gcacatggcc	7620
ggagaaactc ctgcgagaat ttcgtaaact cgcgcgcatt gcactctgat ttcctaatgc	7680
ggcacttcca ggcctcgaga cctctgacat gcttttgaca ggaatagaca ttttcagaat	7740
gttatecata tgcccttcgg gttttttcc ttcctttcc atcatgaaaa atctctcgag	7800
accgtttacc cattgctttt ttggtgtctt tttccctcgt tcacagaaag tctgaagaag	7860
ctatagtaga actatgagct ttttttgtt ctggtttcct ttttttttt tttacctctg	7920

ES 2 607 894 T3

tggaaattgt tactctcaca ctcttttagtt cgtttgtttg ttttgtttat tccaattatg	7980
accggtgacg aaacgtggtc gatggtgggt accgcttatg ctcccctcca ttagtttoga	8040
ttatataaaa aggccaaata ttgtattatt ttcaaagtgc ctatcattat cgtctaacat	8100
ctaatttctc ttaaattttt tctctttctt tcctataaca ccaatagtga aaatcttttt	8160
ttcttctata tctacaaaaa cttttttttt ctatcaacct cgttgataaa ttttttcttt	8220
aaaaatcgtt aataattaat taattggaaa ataaccattt tttctctctt ttatacacac	8280
attcaaaaaga aagaaaaaaa atatacccca gctagttaaa gaaaatcatt gaaaagaata	8340
agaagataag aaagatttaa ttatcaaaca atatcaatat gcctcaatcc tgggaagaac	8400
tggccgctga taagcgcgcc cgcctcgcaa aaaccatccc tgatgaatgg aaagtcaga	8460
cgctgcctgc ggaagacagc gttattgatt tcccaaagaa atcggggatc ctttcagagg	8520
ccgaactgaa gatcacagag gcctccgctg cagatcttgt gtccaagctg gcggccggag	8580
agttgacctc ggtggaagtt acgctagcat tctgtaaacg ggcagcaatc gcccagcagt	8640
taacaaactg cgcccacgag ttcttccctg acgccgctct cgcgcaggca agggaaactc	8700
atgaatacta cgcaaagcac aagagaccgc ttgggtccact ccatggcctc cccatctctc	8760
tcaaagacca gcttcgagtc aagggtctacg aaacatcaat gggctacatc tcatggctaa	8820
acaagtacga cgaaggggac tcggttctga caaccatgct ccgcaaagcc ggtgccgtct	8880
tctacgtcaa gacctctgtc ccgcagacc tgatggtctg cgagacagtc aacaacatca	8940
tcgggcgcac cgtcaaccca cgcaacaaga actggtcgtg cggcggcagt tctggtggtg	9000
agggtgcgat cgttgggatt cgtggtggcg tcatcgggtg aggaacggat atcggtggt	9060
cgattcgagt gccggccgcg ttcaacttcc tgtacggtct aaggccgagt catgggcggc	9120
tgccgatgac aaagatggcg aacagcatgg agggtcagga gacggtgcac agcgttgctg	9180
ggccgattac gactctggt gaggacctcc gcctcttcac caaatccgtc ctcggtcagg	9240
agccatgaa atacgactcc aaggatcatcc ccatgccctg gcgccagtcc gagtcggaca	9300
ttattgcctc caagatcaag aacggcgggc tcaatatcgg ctactacaac ttcgacggca	9360
atgtccttcc acaccctcct atcctgcgcy gcgtggaac caccgtcgcc gcaactcgca	9420
aagccggtca caccgtgacc ccgtggacgc catacaagca cgatttcggc cacgatctca	9480
tctcccatat ctacgcggt gacggcagcy ccgacgtaat gcgcgatatc agtgcacccg	9540
gcgagccggc gattccaaat atcaaagacc tactgaacct gaacatcaa gctgttaaca	9600
tgaacgagct ctgggacacg catctccaga agtgaatta ccagatggag taccttgaga	9660

ES 2 607 894 T3

```

aatggcggga ggctgaagaa aaggccggga aggaactgga cgccatcacc ggcgcgatta 9720
cgctaccgc tgcggtacgg catgaccagt tccggtacta tgggtatgcc tctgtgatca 9780
acctgctgga tttcacgagc gtggttggtc cggttacctt tgcggataag aacatcgata 9840
agaagaatga gagtttcaag gcggttagtg agcttgatgc cctcgtgcag gaagagtatg 9900
atccggaggc gtaccatggg gcaccggttg cagtgcaggt tatcggacgg agactcagtg 9960
aagagaggac gttggcgatt gcagaggaag tggggaagtt gctgggaaat gtggtgactc 10020
cataggtcga gaatttatac ttagataagt atgtacttac aggtatattt ctatgagata 10080
ctgatgtata catgcatgat aatattttaa cggttattag tgccgattgt cttgtgcat 10140
aatgacgttc ctatcaaagc aatacactta ccacctatta catgggcaa gaaaatattt 10200
tcgaacttgt ttagaatatt agcacagagt atatgatgat atccgtaga ttatgcatga 10260
ttcattccta caacttttcc gtagcataag gattaattac ttggatgcca ataaaaaaaa 10320
aaaacatcga gaaaatttca gcatgctcag aaacaattgc agtgtatcaa agtaaaaaaa 10380
agattttcgc tacatgttcc ttttgaagaa agaaaatcat ggaacattag atttcaaaaa 10440
atttaaccac cgctgattaa cgattagacc gttaagcga caacaggtta ttagtacaga 10500
gaaagcattc tgtggtggtg ccccgactt tcttttgcga cataggtaa tcgaatacca 10560
tcatactatc tttccaatg actccctaaa gaaagactct tcttcgatgt tgtatacgtt 10620
ggagcatagg gcaagaattg tggcttgaga tgaattcact ggcgctcgtt ttacaacgtc 10680
gtgactgga aaaccctggc gttaccacac ttaatcgcct tgcagacat cccctttcg 10740
ccagctggcg taatagcgaa gaggcccgca ccgatcgccc ttccaacag ttgcgagcc 10800
tgaatggcga atgggcctg atgcggtatt ttctccttac gcactgtgc ggtatttcac 10860
accgcatatg gtgactctc agtacaatct gctctgatgc cgcatagta agccagcccc 10920
gacaccgcc aacaccgct gacgcgcct gacgggcttg tctgctccc gcatccgctt 10980
acagacaagc tgtgaccgtc tccgggagct gcatgtgtca gaggttttca ccgtcatcac 11040
cgaaacgcgc gagacgaaag ggcctcgtga tacgcctatt tttatagggt aatgtcatga 11100
taataatggg ttcttagacg tcaggtggca cttttcgggg aatgtgccc ggaacccta 11160
tttgtttatt tttctaaata cattcaaata tgtatecgtc catgagaaa taaccctgat 11220
aaatgcttca ataatttga aaaaggaaga gtatgagat tcaacatttc cgtgtcggcc 11280
ttattccctt tttgoggca tttgccttc ctgttttgc tcaccagaa acgctggtga 11340

```

ES 2 607 894 T3

aagtaaaaga tgctgaagat cagttgggtg cacgagtggg ttacatcgaa ctggatctca 11400
 acagcggtaa gatccttgag agttttcgcc ccgaagaacg tttccaatg atgagcactt 11460
 ttaaagttct gctatgtggc gcggtattat cccgtattga cgccgggcaa gagcaactcg 11520
 gtgcgccat aactattct cagaatgact tggttgagta ctcaccagtc acagaaaagc 11580
 atcttacgga tggcatgaca gtaagagaat tatgcagtgc tgccataacc atgagtgata 11640
 aactgcggc caacttactt ctgacaacga tcggaggacc gaaggagcta accgcttttt 11700
 tgcacaacat gggggatcat gtaactcgcc ttgatcgttg ggaaccggag ctgaatgaag 11760
 ccataccaaa cgacgagcgt gacaccacga tgcctgtagc aatggcaaca acgttgcgca 11820
 aactattaac tggcgaacta cttactctag cttcccggca acaattaata gactggatgg 11880
 agcgggataa agttgcagga ccacttctgc gctcggccct tccggctggc tggtttattg 11940
 ctgataaatc tggagccggt gagcgtgggt ctcgcggtat cattgcagca ctggggccag 12000
 atggtaagcc ctcccgtatc gtagttatct acacgacggg gagtcaggca actatggatg 12060
 aacgaaatag acagatcgct gagatagtg cctcactgat taagcattgg taactgtcag 12120
 accaagttta ctcatatata ctttagattg atttaaaact tcatttttaa tttaaaagga 12180
 tctaggtgaa gatccttttt gataatctca tgacccaaaat cccttaacgt gagttttcgt 12240
 tccactgagc gtcagacccc gtagaaaaga tcaaaggatc ttcttgagat cttttttttc 12300
 tgcgcgtaat ctgctgcttg caaacaaaaa aaccaccgct accagcggtg gtttgtttgc 12360
 cggatcaaga gctaccaact ctttttccga aggtaactgg cttcagcaga gcgcagatac 12420
 caaataactgt ctttctagt tagccgtagt taggccacca cttcaagaac tctgtagcac 12480
 cgcctacata cctcgctctg ctaatcctgt taccagtggc tgctgccagt ggcgataagt 12540
 cgtgtcttac cgggttgac tcaagacgat agttaccgga taaggcgcag cggtcgggct 12600
 gaacgggggg ttcgtgcaca cagcccagct tggagcgaac gacctacacc gaactgagat 12660
 acctacagcg tgagctatga gaaagcgcga cgcttcccga agggagaaag gcggacaggt 12720
 atccggtaag cggcagggtc ggaacaggag agcgcacgag ggagcttcca ggggaaacg 12780
 cctggtatct ttatagtcct gtcgggtttc gccacctctg acttgagcgt cgatttttgt 12840
 gatgctcgtc aggggggagg agcctatgga aaaacgccag caacgcggcc tttttacggt 12900
 tcctggcctt ttgctggcct tttgctcaca tgttctttcc tgcgttatcc cctgattctg 12960
 tgataaccg tattaccgcc tttgagttag ctgataccgc tcgccgcagc cgaacgaccg 13020
 agcgcagcga gtcagtgagc gaggaagcgg aagagcggcc aatacgcaaa ccgcctctcc 13080
 ccgcgcgttg gccgattcat taatgcagct ggcacgacag gtttcccagc tggaaagcgg 13140
 gcagtgagcg caacgcaatt aatgtgagtt agctcactca ttaggcaccc caggctttac 13200
 actttatgct tccgctcgt atgttggtg gaattgtgag cggataaaaa tttcacacag 13260
 gaaacagcta tgaccatgat tacgcc 13286

<210> 42

5

<211> 16359

ES 2 607 894 T3

<212> ADN

<213> Artificial

<220>

5

<223> pBOL36

<400> 42

```

aagcttgcat gectgcaggt cgacggcgcg ccggggcccg ttaaacaatg gcaaactgag      60
cacaacaata ccagtcgga tcaactggca ccatctctcc cgtagtctca tctaattttt      120
cttcggatg aggttcagata tataaccgca cacctttatt atggtttccc tgagggaata      180
atagaatgtc ccattcgaaa tcaccaattc taaacctggg cgaattgtat ttcgggtttg      240
ttaactcgtt ccagtcagga atgttccacg tgaagctatc ttccagcaaa gtctccactt      300
cttcatcaaa ttgtgggaga atactcccaa tgctcttata tatgggactt ccgggaaaca      360
cagtaccgat acttcccaat tcgtcttcag agctcattgt ttgtttgaag agactaatca      420
aagaatcgtt ttctcaaaaa aattaatata ttaactgata gtttgatcaa aggggcaaaa      480
cgtaggggca aacaaacgga aaaatcgttt ctcaaatfff ctgatgcaa gaactetaac      540
cagtcttata taaaaattgc cttatgatcc gtctctccgg ttacagcctg tgtaactgat      600
taatcctgcc tttctaatac ccattctaata gttttaatta agggattttg tcttcattaa      660
cggctttcgc tcataaaaaat gttatgacgt tttgcccgca ggcgggaaac catccacttc      720
acgagactga tctcctctgc cggaacaccg ggcactctca acttataagt tggagaaata      780
agagaatttc agattgagag aatgaaaaaa aaaaaaaaaa aaaaggcaga ggagagcata      840
gaaatggggt tcactttttg gtaaagctat agcatgccta tcacatataa atagagtgcc      900
agtagcgact tttttcacac tcgaaatact cttactactg ctctcttggt gttttatca      960
cttcttggtt cttcttggtt aatagaatat caagctaca aaagcataca atcaactatc     1020
aactattaac tatatcgtaa tacacaggcc ggccaaaatg aaggccaaat caaggcggga     1080
agggacaacc aggaacgtaa gggtagcctc cccataacat aaactcaata aaatatatag     1140

```

ES 2 607 894 T3

tcttcaactt gaaaaaggaa caagctcatg caaagaggtg gtaccgcac gccgaaatgc 1200
atgcaagtaa cctattcaaa gtaatatctc atacatgttt catgagggta acaacatgcy 1260
actgggtgag catatgttcc gctgatgtga tgtgcaagat aaacaagcaa gacagaaact 1320
aacttcttct tcatgtaata aacacacccc gcgtttatth acctatctth aaacttcaac 1380
accttatatc ataactaata tttcttgaga taagcacact gcaccatac cttccttaaa 1440
aacgtagctt ccagtttttg gtggttctgg cttocttccc gattccgccc gctaaaacga 1500
taatthttgtt gcctggtggc atthgcaaaa tgcataacct atgcatttaa aagattatgt 1560
atgctcttct gactthttcgt gtgatgaggc tegtggaaaa aatgaataat ttatgaatth 1620
gagaacaatt ttgtgttgtt acggtattht actatggaat aatcaatcaa ttgaggatth 1680
tatgcaaata tegtthtgaat atthttccga cctthtgagt actthttctc ataattgcat 1740
aatattgtcc gctgcccgth tttctgttag acggtgtctt gatctacttg ctatcgthca 1800
acaccacctt atthttctaac taththththt ttagctcatt tgaatcagct tatggtgatg 1860
gcacatththt gcataaacct agctgtctc gttgaacata ggaaaaaaa atatataaac 1920
aaggctctth cactctcctt ggaatcagat ttgggtthtgt tccctthatt ttcatathtc 1980
ttgtcatatt cthttctcaa ttattatctt ctactcataa cctcacgcaa aataacacag 2040
tcaaatcaat caaaatggac ttcaacttga ccagagaaca agaattggtc agacaaatgg 2100
ttagagaatt tgctgaaaac gaagttaagc caattgctgc tgaatcgat gaaactgaaa 2160
gattccaat ggaaaacgct aagaagatgg gtcaatacgg tatgatgggt attccattct 2220
ctaaggaata cgggtggtgct ggtggtgacg tcttgtctta catcattgct gtcgaagaat 2280
tgtccaaggt ttgtggtacc actggtgtca tcttatctgc tcacacttct ctatgtgect 2340
ccttgatcaa cgaacacggt actgaagaac aaaagcaaaa gtacttggtt ccattggcca 2400
aggtgaaaa gattggtgcc tacggtthtga ctgaaccaa cgctggtact gactctggtg 2460
ctcaacaaac tgttgccgth ttggaaggth accactacgt catcaacggt tccaagatct 2520
tcatcaccaa cgggtggtgth gctgacacct ttgtcatctt cgctatgacc gatcgtaacca 2580
aggtaccaa ggttatctct gctthcatta ttgaaaaggg thtcaagggt thctecatcy 2640
gtaaggctga acaaaagtht ggtatcagag cttoctctac cactgaattg gthttcgaag 2700
acatgattgt tccagthtga aacatgatcy gtaaggaagg taagggtthc ccaattgcca 2760
tgaagactth agatggtggt agaattggtt ttgctgctca agctthgggt attgctgaag 2820
gtgccttcaa cgaagctaga gcttacatga aggaaagaaa gcaattcggt agatctthgg 2880

ES 2 607 894 T3

acaaattcca aggtttggct tggatgatgg ctgacatgga cgttgccatc gaatctgctc	2940
gttacttggt ctacaaggct gcttacttga agcaagctgg tttgccatac accgtegatg	3000
ctgccagagc taagttgcac gctgccaacg ttgccatgga tgtcaccacc aaggctgtcc	3060
aattattcgg tggttacggt tacaccaagg actaccagc tgaagaatg atgagagatg	3120
ctaagatcac tgaaatctac gaaggtactt ctgaagtcca aaagttggtt atctccggt	3180
agatcttcag ataggcccg gcataaagca atcttgatga ggataatgat ttttttttga	3240
atatacataa atactaccgt ttttctgcta gattttgtga agacgtaa atagtacatat	3300
tactttttaa gccaaagaca gattaagcat taactttacc cttttctctt ctaagtttca	3360
atactagtta tcaactgtta aaagttatgg cgagaacgct gccgggttaa atatattacc	3420
ctgaacgtgg tgaattgaag ttctagatg gtttaagat ttttctttt tgggaaataa	3480
gtaacaata tattgctgcc tttgcaaac gcacataccc acaatatgtg actattggca	3540
aagaacgcat taccctttga agaggtgat actgatacta agagagtctc tattccggct	3600
ccacttttag tccagagatt acttgtcttc ttacgtatca gaacaagaaa gcatttccaa	3660
agtaattgca tttgcccttg agcagtatat atatactaag aagtttaaac atttaaacgg	3720
ccggcctaga aagcatacta tactattcga cacttccttt caatcctgga attaacagtc	3780
acttttaaaa aagacatcta ccgtgaaggt gccgtagagt atcgcgttac catatcgcca	3840
aaaactgata tacgccgagg aaaccaggca acaattgaa aagaaaaatt ttgaggaact	3900
ctctgcatcg aagccgtcta gagttaccac tagtcagatg ccgcccggcac ttgagcaact	3960
catgcacagc aataacacaa cacaatggtt agtagcaacc tgaattcggc cattgatgca	4020
tgcattgtcc gtgaagcggg acaaccagaa aagtcgtcta taaatgccg cactgctgat	4080
catcgtggcg gggttttaag agtgcatatc acaaatgtc gcattaccgc ggaaccgcca	4140
gatattcatt acttgacgca aaagcgtttg aaataatgac gaaaaagaag gaagaaaaa	4200
aaagaaaaat accgcttcta ggccgggttat ctactgatcc gagcttccac taggatagca	4260
cccaaacacc tgcataattg gacgacctt acttacacca ccaaaaacca ctttcgctc	4320
tcccgccctt gataacgtcc actaatgag cgattacctg agcgtctctc ttttgtttgc	4380
agcatgagac ttgcatactg caaatcgtaa gtagcaacgt ctcaaggcca aaactgtatg	4440
gaaaccttgt cacctcactt aattctagct agcctaccct gcaagtcaag aggtctccgt	4500
gattcctagc cacctcaagg tatgcctctc cccggaaact gtggcctttt ctggcacaca	4560

ES 2 607 894 T3

tgatctccac gatttcaaca tataaatagc ttttgataat ggcaatatta atcaaattta	4620
ttttacttct ttcttgtaac atctctcttg taatccctta ttccttctag ctatTTTTca	4680
taaaaaacca agcaactgct tatcaacaca caaacactaa atcaaatgg tCGatttCGa	4740
atactctatc ccaaccagaa tcttcttcgg taaggacaag atcaacgttt tgggtagaga	4800
attgaagaaa tacggttcca aggttttgat tgtctacggt ggtggttcca tcaagagaaa	4860
cggtatctac gacaaggctg tctccatTTT ggaaaagaac tctatcaaT tctacgaatt	4920
ggctgggtgtt gaaccaaacc caagagttac cacCGtCGaa aagggtgtca agatctgtcg	4980
tGaaaacggt gttgaagttg ttttggccat cggTggTggt tctgccattg actgtGCCaa	5040
ggtcattgct gctgcctgtg aatacgatgg taaccatgg gacattgtct tggatggTtc	5100
taagatcaag cgtgtcttac caattgcttc catcttgact atCGtGctA ctggttctga	5160
aatggacacc tgggctgTta tcaacaacat ggacactaac gaaaagttga ttgctgctca	5220
cccagatatg gccccaaagt tctctatTTT ggacccaacc tacacttaca ctgTtccaac	5280
caaccaaact gctgctggta ctgctgatat catgtctcac atctttgaag tttacttctc	5340
caacaccaag accgcttact tgcaagacag aatggctgaa gctctattaa gaacctgtat	5400
caagtacggt ggtattgctt tggaaaagcc agatgactac gaagccagag ctaacttgat	5460
gtgggcttcc tctttggcta tcaacggttT attgacttac ggtAaggaca ccaactggtc	5520
cgttcatttg atggaacacg aattgtctgc ttactacgat atcactcacg gtgtcggttT	5580
ggccatcttg actccaaact ggatggaata cttttgaaC aacgacactg tctacaagtt	5640
cgtcgaatac ggtgttaacg tctggggtat tgacaaggaa aagaaccact acgacattgc	5700
tcaccaagcc atccaaaaga ccagagacta tttcgtcaac gttttgggtt taccatccag	5760
attaagagat gttggtattg aagaagaaaa attggatata atggctaagg aatctgtcaa	5820
attgactggt ggtaccattg gtaacttgag acctgttaac gcttctgaag ttttgcaaat	5880
cttcaagaaa tctgttttag cccgggctcc tgttgaagta gcatttaacT ataatttttg	5940
tcacatttta atcaacttga tttttctggt ttaatttttc taattttaat ttttaatttt	6000
ttatcaatgg gaactgatac actaaaaaga attaggagcc aacaagaata agccgcttat	6060
ttcctactag agtttgctta aaatttcatc tcgaattgtc attctaatat tttatccaca	6120
cacacacctt aaaattttta gattaaatgg catcaactct tagcttcaca cacacacaca	6180
caccgaagct ggttgTTTT tttgatttga tataattggT ttctctggat ggtactTTTT	6240
ctttcttggT tatttctat tttaaaatat gaaacgcaca caagtcataa ttattctaT	6300

ES 2 607 894 T3

```

agagcacaat tcacaacacg cacatttcaa ctttaatat ttttagaaa cactttattt 6360
agtctaattc ttaattttta atatatataa tgcacacaca cgtttaaag ggcccgcggc 6420
ccgtttaaac ggccggccct tcccttttac agtgcttcgg aaaagcacag cgttgtccaa 6480
gggaacaatt tttcttcaag ttaatgcata agaaatatct tttttatgt ttagctaagt 6540
aaaagcagct tggagtaaaa aaaaaatga gtaaatttct cgatggatta gtttctcaca 6600
ggtaacatag caaaaaccaa gaaaagcccg cttctgaaaa ctacagttga cttgtatgct 6660
aaagggccag actaatggga ggagaaaaag aaacgaatgt atatgctcat ttacactcta 6720
tatcaccata tggaggataa gttgggctga gcttctgac caatttattc tatccattag 6780
ttgctgatat gtcccaccag ccaacacttg atagtatcta ctgccattc acttccagca 6840
gcccagtag ggttgttgag cttagtataa atgtgcgcac cacaagccta catgaactca 6900
cgtcacatga aaccacaccg tggggccttg ttgcgctagg aataggatat gcgacgaaga 6960
cgcttctgct tagtaaccac accacatttt cagggggtcg atctgcttgc ttcctttact 7020
gtcacgagcg gcccataatc gcgctttttt tttaaaaggc gcgagacagc aaacaggaag 7080
ctcgggtttc aaccttcgga gtggctgcag atctggagac tggatcttta caatacagta 7140
aggcaagcca ccatctgctt cttaggtgca tgcgacggta tccacgtgca gaacaacata 7200
gtctgaagaa gggggggagg agcatgttca ttctctgtag cagtaagagc ttgggtataa 7260
tgacaaaaac tggagtctcg aaatcatata aatagacaat atattttcac acaatgagat 7320
ttgtagtaca gttctattct ctctcttgca taaataagaa attcatcaag aacttggttt 7380
gatatttcac caacacacac aaaaaacagt acttactaa atttacacac aaaacaaaat 7440
gaaggttacc aacaaaagg aattgaagca aaagttgaac gaattgagag aagctcaaaa 7500
gaagtctgct acctacactc aagaacaagt tgacaagatc ttcaagcaat gtgccattgc 7560
tgctgccaag gaacgtatca acttggccaa gttggctgtc gaagaaaccg gtattggttt 7620
ggttgaagac aagatcatca agaaccactt cgctgctgaa tacatctaca acaagtacaa 7680
gaacgaaaag acctgtggta tcatcgacca cgatgactct ttgggtatca ccaaggttgc 7740
tgaaccaatc ggtattgtcg ccgccattgt cccaaccact aacccaactt ccaactgccat 7800
cttcaaactc ttgatctect tgaagaccag aaacgetatc ttcttctccc cacacceaaag 7860
agccaagaag tccaccattg ctgctgccaa attaactctg gatgctgctg ttaagctgg 7920
tgcccaaaag aacattattg gttgatoga tgaaccttcc attgaattgt ctcaagactt 7980

```


ES 2 607 894 T3

gatgtctgaa gctgatatca tcttggtctac cgggtggtcca tccatggtca aggccgctta 8040
ctcttctggt aagccagcta ttggtgttgg tgctggtaac actccagcta tcatcgatga 8100
atctgctgac attgacatgg ctgtctcctc cattatcttg tccaagactt atgacaacgg 8160
tgtcatctgt gcctctgaac aatccatctt ggttatgaac tctatctacg aaaaggtaa 8220
ggaagaattt gttaagagag gttcctacat cttaaaccaa aatgaaattg ccaagatcaa 8280
ggaaacatg ttcaagaacg gtgccatcaa cgctgacatt gtcggtaaat ctgcttacat 8340
cattgccaaag atggctggta ttgaagttcc acaaaccact aagatcttga tccggtgaagt 8400
tcaatctgtc gaaaagtctg aattattctc tcacgaaaag ttgtctccag tcttggtctat 8460
gtacaaggtc aaggatctcg acgaagcttt gaagaaggct caaagattaa ttgaattagg 8520
tggttctggt cacacctctt ctctatacat tgactctcaa aacaacaagg acaaggtaa 8580
ggaattcggc ctagctatga agacttccag aactttcatc aacatgccat cttctcaagg 8640
tgcttctggt gatttgtaca actttgccat tgctccatct ttcactttag gttgtggtac 8700
ctggggtggt aactctgttt ctcaaaacgt tgaaccaaag catttgctaa acatcaagtc 8760
cgttgctgaa agaagagaaa acatgttggg gttcaagggt ccacaaaaga tctacttcaa 8820
atacggttgt ttgagatttg ctttgaagga attgaaagat atgaacaaga agcgtgcttt 8880
catcgttact gacaaggatt tgttcaaatt gggttacggt aacaagatca ctaaggtttt 8940
ggatgaaatt gatatcaagt actccatctt cactgatatc aaatctgacc caaccattga 9000
ctccgtcaag aagggtgcta aggaaatggt gaacttcgaa ccagatacca ttatctccat 9060
tgggtggtgt tctccaatgg atgctgcaa ggttatgcat ttgtgttacg aataaccaga 9120
agctgaaatc gaaaacttgg ccatcaactt catggacatc agaaagagaa tctgtaactt 9180
cccaaagtg ggtaccaagg ccatttctgt tgccattcca accaccgctg gtaccggttc 9240
tgaagctact ccatttgctg tcatcaccaa cgacgaaacc ggtatgaagt acccattgac 9300
ctcttacgaa ttgactcaa acatggccat cattgacact gaattgatgt tgaacatgcc 9360
aagaaagtg actgctgcta ccggtattga cgctttagtc cacgctatcg aagcttacgt 9420
ctccgttatg gccactgact acactgacga attggctttg agagctatca agatgatctt 9480
caagtacttg ccaagagctt acaagaacgg tactaacgat atcgaagctc gtgaaaagat 9540
ggctcacgct tccaacattg ctggtatggc ttctgctaac gctttcttgg gtgtttgtca 9600
ctccatggc cacaagttgg gtgctatgca ccagttcct cacggtattg cttgtgctgt 9660
tttgattgaa gaagtcacatc agtacaacgc tactgactgt ccaaccaagc aaactgcttt 9720

ES 2 607 894 T3

```

cccacaatac aagtctccaa acgccaaagag aaagtacgct gaaattgctg aataacttgaa 9780
cttgaaaggt acttctgaca ctgaaaaggt cactgcttta atcgaagcta tctccaagtt 9840
gaagattgac ttatctattc ctcaaaacat ctctgctgct ggtattaaca agaaggactt 9900
ctacaacact ttagacaaga tgtccgaatt ggctttcgat gaccaatgta ccaccgctaa 9960
ccaagatac ccattgatct ctgaattgaa ggatatctac atcaagtcct tttaagcccg 10020
ggcggggatc tcttatgtct ttacgattta tagttttcat tatcaagtat gcctatatta 10080
gtatatagea tctttagatg acagtgttcg aagtttcacg aataaaagat aatattctac 10140
tttttgctcc caccgcgttt gctagcacga gtgaacacca tccctcgcct gtgagttgta 10200
cccattcctc taaactgtag acatggtagc ttcagcagtg ttcgttatgt acggcatcct 10260
ccaacaaaca gtcggttata gtttgcctg ctctctgaa tcgtctcct cgatatttct 10320
cattttcctt cgcatgccag cattgaaatg atcgaagttc aatgatgaaa cggtaattct 10380
tctgtcattt actcatctca tctcatcaag ttatataatt ctatacggat gtaatttttc 10440
acttttcgtc ttgacgtcca ccctataatt tcaattattg aaccctcaca aatgatgcac 10500
tgcaatgtac acaccctcat atagtttaa catttaaagtg gcccgctcta gaggatcccc 10560
gggtaccgag ctccggccca gcgctactag ttcgggtaat ttgaaaacaa acccggtctc 10620
gaagcggaga tccggcgata attaccgag aaataaaccc atacacgaga cgtagaacca 10680
gccgcacatg gccggagaaa ctctgcgag aatttcgtaa actcgcgcgc attgcatctg 10740
tattttcctaa tgcggcaact ccaggcctcg agacctctga catgcttttg acaggaatag 10800
acattttcag aatgttatcc atatgccttt cgggtttttt tccttcttt tccatcatga 10860
aaaatctctc gagaccgttt atccattgct tttttgttgt ctttttccct cgttcacaga 10920
aagtctgaag aagctatagt agaactatga gctttttttg tttctgtttt cctttttttt 10980
ttttttacct ctgtggaaat tgttactctc aactcttta gttcgtttgt ttgttttggt 11040
tattccaatt atgaccggtg acgaaacgtg gtogatggtg ggtaccgctt atgctcccct 11100
ccattagttt cgattatata aaaagccaa atattgtatt attttcaaat gtcctatcat 11160
tatcgtctaa catctaattt ctcttaaatt ttttctcttt ctttctata acaccaatag 11220
tgaaaatctt tttttcttct atatctacaa aaactttttt tttctatcaa cctcgttgat 11280
aaattttttc tttacaatc gtaataaatt aattaattgg aaaataacca ttttttctct 11340
cttttataca cacattcaaa agaaagaaaa aaaatatacc ccagctagtt aaagaaaaac 11400

```

ES 2 607 894 T3

```

attgaaaaga ataagaagat aagaaagatt taattatcaa acaatatcaa tatgcctcaa 11460
tcttggaag aactggccgc tgataagcgc gccgcctcg caaaaacat ccctgatgaa 11520
tgaaagtcc agacgctgcc tgcggaagac agcgttattg atttccaaa gaaatcgggg 11580
atcctttcag aggccgaact gaagatcaca gaggcctccg ctgcagatct tgtgtccaag 11640
ctggcgccg gagagttgac ctcggtgaa gttacgctag cattctgtaa acgggcagca 11700
atcgcccagc agttaacaaa ctgcgcccac gagttcttcc ctgacgccc tctcgcgcag 11760
gcaagggaac tcgatgaata ctacgcaaag cacaagagac ccgttggtcc actccatggc 11820
ctccccatct ctctcaaaga ccagcttoga gtcaagggct acgaaacatc aatgggctac 11880
atctcatggc taaacaagta cgacgaagg gactcggttc tgacaacat gctccgcaaa 11940
gccggtgccg tcttctacgt caagacctct gtcccgcaga ccctgatggt ctgcgagaca 12000
gtcaacaaca tcatcgggcg caccgtcaac ccacgcaaca agaactggtc gtgcgcgccg 12060
agttctggtg gtgaggggtgc gatcgttggg attcgtggtg gcgtcatcgg tgtaggaacg 12120
gatatcggtg gctcgattcg agtgccggcc gcgttcaact tctgtacgg tctaagccg 12180
agtcatgggc ggctgcccga tgcaaagatg gcgaacagca tggaggggtca ggagacggtg 12240
cacagcgttg tcgggccgat tacgactct gttgaggacc tccgcctctt caccaaacc 12300
gtcctcgtc aggagccatg gaaatacgac tccaaggtca tccccatgcc ctgggccag 12360
tccgagtcgg acattattgc ctccaagatc aagaacggcg ggctcaatat cggctactac 12420
aacttcgacg gcaatgtcct tccacacct cctatcctgc gcggcgtgga aaccaccgtc 12480
gccgactcg ccaaagccgg tcacaccgtg acccgtgga cgcatacaa gcacgatctc 12540
ggccacgatc tcatctccca tatctacgcg gctgacggca gcgccgacgt aatgccccgat 12600
atcagtgcat ccggcgagcc ggcgattcca aatatcaaag acctactgaa cccgaacatc 12660
aaagctgta acatgaacga gctctgggac acgcatctcc agaagtggaa ttaccagatg 12720
gagtacctg agaaatggcg ggaggctgaa gaaaaggccg ggaaggaact ggacgccatc 12780
atcgcgccga ttacgcctac cgctgcggtg cggcatgacc agttccggta ctatgggtat 12840
gcctctgtga tcaacctgct ggatttcacg agcgtggtt ttccggttac ctttgcggat 12900
aagaacatcg ataagaagaa tgagagtttc aaggcgggta gtgagcttga tgccctcgtg 12960
caggaagagt atgatccgga ggcgtaccat ggggcaccgg ttgcagtgca ggttatcggg 13020
cggagactca gtgaagagag gacggtggcg attgcagagg aagtggggaa gttgctggga 13080
aatgtggtga ctccataggt cgagaattta tacttagata agtatgtact tacaggtata 13140

```

ES 2 607 894 T3

tttctatgag atactgatgt atacatgcat gataatattt aaacggttat tagtgccgat 13200
 tgtcttgtgc gataatgacg ttcctatcaa agcaatacac ttaccaccta ttacatgggc 13260
 caagaaaata ttttcgaact tgtttagaat attagcacag agtatatgat gatatccggt 13320
 agattatgca tgattcattc ctacaacttt ttcgtagcat aaggattaat tacttggatg 13380
 ccaataaaaa aaaaaaacat cgagaaaatt tcagcatgct cagaaacaat tgcagtgat 13440
 caagtaaaa aaaagatttt cgctacatgt tccttttgaa gaaagaaaat catggaacat 13500
 tagatttaca aaaatttaac caccgctgat taacgattag accgttaagc gcacaacagg 13560
 ttattagtac agagaaagca ttctgtggtg ttgccccgga ctttcttttg cgacataggt 13620
 aaatcgaata ccatcatact atcttttcca atgactccct aaagaaagac tcttcttcca 13680
 tgttgatatac gttggagcat agggcaagaa ttgtggcttg agatgaattc actggccgtc 13740
 gttttacaac gtcgtgactg ggaaaacctt ggcgttacct aacttaatcg ccttgcagca 13800
 catccccctt tcgccagctg gcgtaatagc gaagaggccc gcaccgatcg cccttcccaa 13860
 cagttgcgca gctgaatgg cgaatggcgc ctgatgcggt attttctctc tacgcatctg 13920
 tgcggtattt cacaccgcat atggtgcact ctcagtacaa tctgctctga tgccgcatag 13980
 ttaagccaagc cccgacaccc gccaacaccc gctgacgcgc cctgacgggc ttgtctgctc 14040
 ccggcatccg cttacagaca agctgtgacc gtctccggga gctgcatgtg tcagaggttt 14100
 tcaccgcat caccgaaacg cgcgagacga aagggcctcg tgatacgcct atttttatag 14160
 gttaatgtca tgataataat ggtttcttag acgtcaggtg gcacttttctg gggaaatgtg 14220
 cgcggaaccc ctatttgttt atttttctaa atacattcaa atatgtatcc gctcatgaga 14280
 caataaccct gataaatgct tcaataatat tgaaaaagga agagtatgag tattcaacat 14340
 ttcogtgcg cccttattcc cttttttgcg gcattttgcc ttctgtttt tgctcacca 14400
 gaaacgctgg tgaaagtaaa agatgctgaa gatcagttgg gtgcacgagt gggttacatc 14460
 gaactggatc tcaacagcgg taagatcctt gagagtttct gccccgaaga acgttttcca 14520
 atgatgagca cttttaaagt tctgctatgt ggcgcggtat tatcccgat tgacgccggg 14580
 caagagcaac tcggtcgccg catacactat tctcagaatg acttggttga gtactacca 14640
 gtcacagaaa agcatcttac ggatggcatg acagtaagag aattatgcag tgctgccata 14700
 accatgagtg ataacactgc ggccaactta cttctgacaa cgatcggagg accgaaggag 14760
 ctaaccgctt ttttgcacaa catgggggat catgtaactc gccttgatcg ttgggaaccg 14820

ES 2 607 894 T3

```

gagctgaatg aagccatacc aaacgacgag cgtgacacca cgatgcctgt agcaatggca 14880
acaacgttgc gcaaactatt aactggcgaa ctacttactc tagcttcccc gcaacaatta 14940
atagactgga tggaggcgga taaagttgca ggaccacttc tgcgctcggc ccttccggct 15000
ggctggttta ttgctgataa atctggagcc ggtgagcgtg ggtctcggc tatcattgca 15060
gcaactatgg atgaacgaaa tagacagatc gctgagatag gtgcctcact gattaagcat 15180
tggtaaactgt cagaccaagt ttactcatat atactttaga ttgatttaaa acttcatttt 15240
taatttaaaa ggatctaggt gaagatcctt tttgataatc tcatgaccaa aatcccttaa 15300
cgtgagtttt cgttccactg agcgtcagac cccgtagaaa agatcaaagg atcttcttga 15360
gatccttttt ttctgocgt aatctgctgc ttgcaaacia aaaaaccacc gctaccagcg 15420
gtggtttggt tgccggatca agagctacca actctttttc cgaaggtaac tggcttcagc 15480
agagcgcaga taccaaatac tgtccttcta gtgtagccgt agttaggcca ccacttcaag 15540
aactctgtag caccgcctac atacctcgtc ctgctaatec tgttaccagt ggctgctgcc 15600
agtggcgata agtcgtgtct taccgggttg gactcaagac gatagttacc ggataaggcg 15660
cagcggtcgg gctgaacggg gggttcgtgc acacagccca gcttggagcg aacgacctac 15720
accgaactga gatacctaca gcgtgagcta tgagaaagcg ccacgcttcc cgaagggaga 15780
aaggcggaca ggtatccggt aagcggcagg gtcggaacag gagagcgac gagggagctt 15840
ccaggggaa acgcttgga tctttatagt cctgtcgggt ttcgccacct ctgactgag 15900
cgtcgatfff tgtgatgctc gtcagggggg cggagcctat ggaaaaacgc cagcaacgcg 15960
gcctttttac ggttctcggc cttttgctgg ctttttgcct acatgttctt tctcgcgta 16020
tcccctgatt ctgtggataa ccgtattacc gcctttgagt gagctgatac cgctcggcgc 16080
agccgaacga ccgagcgcag cgagtcagtg agcggaggaag cggaagagcg cccaatacgc 16140
aaaccgcctc tcccgcgcgc ttggccgatt cattaatgca gctggcacga caggtttccc 16200
gactggaaaag cgggcagtga gcgcaacgca attaatgtga gttagctcac tcattaggca 16260
cccagcgtt tacactttat gcttccggct cgtatgttgt gtggaattgt gagcggataa 16320
caatttcaca caggaaacag ctatgaccat gattacgcc 16359

```

<210> 43

<211> 8684

5

<212> ADN

<213> Artificial

<220>

<223> pBOL113

10

<400> 43

ES 2 607 894 T3

tcgcgcgttt cggatgatgac ggtgaaaacc tctgacacat gcagctcccg gagacggtca	60
cagcttgtct gtaagcggat gccgggagca gacaagcccg tcagggcgcg tcagcgggtg	120
ttggcgggtg tcggggctgg cttaactatg cggcatcaga gcagattgta ctgagagtgc	180
accataccac agcttttcaa ttcaattcat cttttttttt ttattctttt ttttgatttc	240
ggtttctttg aaattttttt gattcggtaa tctccgaaca gaaggaagaa cgaaggaagg	300
agcacagact tagattggta tatatacgca tatgtagtgt tgaagaaaca tgaaattgcc	360
cagtattctt aacccaactg cacagaacaa aaacctgcag gaaacgaaga taaatcatgt	420
cgaaagctac atataaggaa cgtgctgcta ctcatcctag tcctgttgct gccaaagctat	480
ttaatatcat gcacgaaaag caaacaact tgtgtgcttc attggatggt cgtaccacca	540
aggaattact ggagttagtt gaagcattag gtccaaaat ttgtttacta aaaacacatg	600
tggatatctt gactgatttt tccatggagg gcacagttaa gccgctaaag gcattatccg	660
ccaagtacaa ttttttactc ttcgaagaca gaaaatttgc tgacattggt aatacagtca	720
aattgcagta ctctcgggt gtatacagaa tagcagaatg ggcagacatt acgaatgcac	780
acgggtggtt gggcccagg attgtagcg gtttgaagca ggcgcagaa gaagtaacaa	840
aggaacctag aggccttttg atgtagcag aattgtcatg caagggctcc ctatctactg	900
gagaatatac taagggtact gttgacattg cgaagagcga caaagatttt gttatcggct	960
ttattgctca aagagacatg ggtggaagag atgaaggta cgattggtt attatgacac	1020
ccggtgtggg ttttagatgac aagggagacg cattgggtca acagtataga accgtggatg	1080
atgtggtctc tacaggatct gacattatta ttggtggaag aggactattt gcaaagggaa	1140
gggatgctaa ggtagagggt gaacgttaca gaaaagcagg ctgggaagca tatttgagaa	1200
gatgcggcca gcaaaaactaa aaaactgtat tataagtaaa tgcattgata ctaaaactcac	1260
aaattagagc ttcaatttaa ttatatcagt tattacccta tgcggtgtga aataccgcac	1320
agatgcgtaa ggagaaaata ccgcatcagg aaattgtaaa cgtaaatatt ttgttaaaat	1380
tcgcgtaaa tttttgttaa atcagctcat tttttaacca ataggccgaa atcggcaaaa	1440
tcccttataa atcaaaaagaa tagaccgaga tagggttgag tgttgttcca gtttgaaca	1500
agagtccact attaagaac gtggactcca acgtcaaagg gcgaaaaacc gtctatcagg	1560

ES 2 607 894 T3

gcgatggccc actacgtgaa ccatcacct aatcaagttt tttggggtcg aggtgcccga 1620
 aagcactaaa tcggaaccct aaagggagcc cccgatttag agcttgacgg ggaagccgg 1680
 cgaacgtggc gagaaaggaa gggaagaaag cgaaaggagc gggcgctagc gcgctggcaa 1740
 gtgtagcggc cacgctgcgc gtaaccacca cccccccgc gcttaatgcg ccgctacagg 1800
 gcgctgcgc ccattcgcca ttcaggctgc gcaactgttg ggaagggcga tcgggtcggg 1860
 cctcttcgct attacgccag ctggcgaaag ggggatgtgc tgcaaggcga ttaagttggg 1920
 taagccagg gttttccag tcacgacgtt gtaaacgac ggccagttag cgcgcgtaat 1980
 acgactcact atagggcgaa ttgggtaccg ggccccct cgaggtcgac ggtatcgata 2040
 agcttgatat cgaattcctg cagcccggg gatccactag ttctagagcg gccatttaa 2100
 acggccggcc ctagatcaga gggtggtaaa tgaagtgtaa tagtattcat ttttctata 2160
 aatcatcct tccgtgattt atacaaaaga agaggagaat atgctgaata cttggtatat 2220
 tactctacat tatactctta tcttgacggg tattctgagc atcttactca gtttcaagat 2280
 cttttaatgt ccaaaaacat ttgagccgat ctaaatactt ctgtgttttc attaatttat 2340
 aaattgtact cttttaagac atgaaagta ccaacatcgg ttgaaacagt ttttcattta 2400
 cttatggttt attggttttt ccagtgaatg attatttgc gttacccttt cgtaaaagtt 2460
 caaacacgtt ttaagtatt gtttagttgc tctttcgaca tatatgatta tccctgcgcg 2520
 gctaaagtta aggatgcaaa aaacataaga caactgaagt taatttacgt caattaagtt 2580
 ttccagggtat atgatgtttt gggcttccac taattcaata agtatgtcat gaaatacgtt 2640
 gtgaagagca tccagaaata atgaaaagaa acaacgaaac tgggtcggcc tgttgtttct 2700
 tttctttacc acgtgatctg cggcatttac aggaagtcgc gcgttttgcg cagttgttgc 2760
 aacgcagcta cggctaacia agcctagtgg aactcgactg atgtgttagg gcctaaaact 2820
 ggtggtgaca gctgaagtga actattcaat ccaatcatgt catggctgtc acaaagacct 2880
 tgcggaccgc acgtacgaac acatacgtat gctaatatgt gttttgatag taccagtgta 2940
 tcgcagacct gcaatttttt tgtaggtttg gaagaatata taaaggttgc actcattcaa 3000
 gatagttttt ttcttgtgtg tctattcatt ttattattgt ttgtttaaat gttaaaaaaa 3060
 ccaagaactt agtttcaaat taaattcatc acacaaacia acaaaacia atgaacattg 3120
 ttgtttgttt gaagcaagtt ccagacactg ctgaagtcag aattgacca gtcaagggtta 3180
 ctttaatcag agaaggtgtt ccatctatca tcaaccaga cgacaagaac gctttggaag 3240
 aagctttggt tttgaaggac aactacggtg ctcacgttac cgtcatttcc atgggtccac 3300

ES 2 607 894 T3

ctcaagccaa gaacgctttg gttgaagctt tggccatggg tgctgatgaa gctgtcttat 3360
 tgactgacag agctttcggg ggtgctgata ctttagctac ctctcacacc attgctgctg 3420
 gtatcaagaa attgaaatac gatatcgtct ttgccggtcg tcaagccatc gatggtgata 3480
 ccgctcaagt cggccagaa attgctgaac atttgggtat tccacaagtc acctacgttg 3540
 aaaaggttga agttgacggg gacactttga agatcagaaa ggcttgggaa gacggttacg 3600
 aagttgttga agtcaagact ccagttctat tgactgccat caaggaattg aacgttccaa 3660
 gatacatgtc cgttgaaaag atcttcggtg ctttcgacaa ggaagtcaag atgtggactg 3720
 ctgatgatat cgatgtcgac aaggccaact tgggtttgaa aggttctcca accaaggtca 3780
 agaaatcttc taccaaggaa gtcaagggtc aagggtgaagt cattgacaaa ccagtcaagg 3840
 aagctgccgc ttacgttggt tccaagttga aggaagaaca ctacatctaa agcccgggcg 3900
 gagattgata agacttttct agttgcatat cttttatatt taaatcttat ctattagtta 3960
 attttttgta atttatcctt atatatagtc tggttattct aaaatatcat ttcagtatct 4020
 aaaaattccc ctcttttttc agttatatct taacaggcga cagtccaaat gttgatttat 4080
 cccagtcgga ttcacaggg ttgtgaagca ttttgtcaat ggtcgaaatc acatcagtaa 4140
 tagtgcctct tacttgcctc atagaatttc tttctcttaa cgtcacctgt tggcttttta 4200
 tagtttcgaa atctatggg ataccaaatg gtgttcccaa ttcacgtta cgggctatt 4260
 ttttaccat tgaagtattg gaatcgtcaa ttttaaagta tatctctctt ttacgtaaag 4320
 cctgcgagat cctcttaagt atagcgggga agccatcgtt attcgatatt gtcgtaacaa 4380
 atactttgat cggcgctatg tttaaatggt taaacatgga cagatatgcg atgaaaacgc 4440
 taagtgatac tccaaatggt gaaaggtacg atgcttgaa acaatacttg gaaatcaccg 4500
 gaaacaccat atgcggcgaa aagccaatta gtgtgatact aagtgcttta tcgaaaatcc 4560
 gtgatgccgg tccttcaggc atcaaatttc agtggcctaa ttattcacag agttctcatg 4620
 tgacaagtat tgatgatagt agtgtcagtt atgcttcagg ttatgttact ataggataat 4680
 gatcacggct aaaacggtcg aatgtaagca tatacttttc gattgtataa ttgttcccaa 4740
 atactacagc atctcaagga aaaaaaaaaa aaaacttcca aaaaaatcga atccctgagg 4800
 aatctttaat acattttcaa tctatttaag ttttataaac gtgtatatga gatgtcatga 4860
 gcatgaatta ttaataataa aaactaaatc attaaagtaa cttaaggagt taaagcccgg 4920
 gctttaattg ttagcagcct tgacttgagc aatcaattct ggaacaacct tgttgacatc 4980

ES 2 607 894 T3

accgacaatg gccaaatcag caaccttcat gattggagct tcgacatcct tgttgatggc 5040
 aatgatgtag tcagagtctt gcataaccagc caagtgttgg atggcaccag agataccaca 5100
 agcaatgtac aaagtggtc tgacggctct accggttga ccgacttgca agtccttgtc 5160
 aaccattcc ttttcaatgg cagctctgga agcagcaatg gtaccacca acaaagaagc 5220
 taattcttcc aatttttcga agttttcctt ggaaccaaca ccacgaccac cagcaacca 5280
 aaccttggct tcaccgatat cagcaatgtc cttggccaat ttgacaacct tggaaacctt 5340
 ggttctgata tcagaagcag tcaatttgat ggcaaccttt tcgatcttgt catcagaaac 5400
 gttagcatcg ttaactggca atttttcaaa gacacctggt ctgacggtag ccatttgagg 5460
 tctgtggca gaacagacaa tggtagcaat caagttacca ccgaaagctg gtctggtagc 5520
 caacaagtca cggttttcga catcgatata caaagagta cagtcagcag tcaaaccagt 5580
 agacaatctg gcagcaattc ttggacccaa gtctctaccg atgaaagtag caccgatgaa 5640
 taagatttct ggttttctt cgttgaccaa gtcacagata accttggcgt aaccgtcagt 5700
 ggagaaatga gctaataatt cgttgcagc agccaaaacc ttgtcagcac cgtgggacaa 5760
 caagtccttg gacatctttt cagtgttgag acccaataag acagcagtca attcaacacc 5820
 caatttttca gccatttctt tacccttacc tagcaattcc aaagaacct tttgtaattc 5880
 accatctctt tgttcagcga aaaccagac accctttag tagccttgt tcatgttag 5940
 ttaattatag ttcgttgacc gtatattcta aaaacaagta ctcttaaaa aaaaaccttg 6000
 aagggaataa acaagtagaa tagatagaga gaaaaataga aatgcaaga gaatttatat 6060
 attagaaaga gagaagaaa aatggaaaa aaaaaatagg aaaagccaga aatagcacta 6120
 gaaggagcga caccagaaaa gaaggtgatg gaaccaattt agctatata agttaactac 6180
 cggctcgatc atctctgcct ccagcatagt cgaagaagaa tttttttttt cttgaggctt 6240
 ctgtcagcaa ctogtatttt ttctttcttt tttggtgagc ctaaaaagtt cccacgttct 6300
 cttgtacgac gccgtcacia acaaccttat gggttaattg tcgcggtctg ggtgtataaa 6360
 tgtgtgggtg caggccggcc gtttaaacgg gccgccaccg cgggtggagct ccagcttttg 6420
 ttccctttag tgagggttaa ttgcgcgctt ggcgtaatca tggcatagc tgtttcctgt 6480
 gtgaaattgt taccgctca caattccaca caacatagga gccggaagca taaagtgtaa 6540
 agcctggggt gcctaagtag tgaggttaact cacattaatt gcgttgcgct cactgcccgc 6600
 tttccagtcg gaaaacctgt cgtgccagct gcattaatga atcgccaac gcgcggggag 6660
 agggggttg cgtattgggc gctcttccgc ttctctgctc actgactcgc tgcgctcggc 6720

ES 2 607 894 T3

cgttcggctg cggcagcgcg tadcagctca ctcaaaggcg gtaatacggg tatccacaga 6780
 atcaggggat aacgcaggaa agaacatgtg agcaaaaggc cagcaaaagg ccaggaaccg 6840
 taaaaaggcc gcgttgctgg cgtttttcca taggctccgc cccctgacg agcatcacia 6900
 aaatcgacgc tcaagtcaga ggtggcgaaa cccgacagga ctataaagat accagcgtt 6960
 tccccctgga agctcctcg tgcgctctcc tgttccgacc ctgccgctta ccggatacct 7020
 gtccgccttt ctcccttcgg gaagcgtggc gctttctcat agctcacgct gtaggtatct 7080
 cagttcggtg taggtcgttc gctccaagct gggctgtgtg cacgaacccc ccgttcagcc 7140
 cgaccgctgc gccttatccg gtaactatcg tcttgagtcc aaccggtaa gacacgactt 7200
 atcgccactg gcagcagcca ctggtaacag gattagcaga gcgaggtatg taggcggtgc 7260
 tacagagttc ttgaagtggg ggcctaacta cggctacact agaaggacag tatttggtat 7320
 ctgcgctctg ctgaagccag ttacctcgg aaaaagagtt ggtagctctt gatccgcaa 7380
 acaaaccacc gctggtagcg gtggtttttt tgtttgcaag cagcagatta cgcgcagaaa 7440
 aaaaggatct caagaagatc ctttgatctt ttctacgggg tctgacgctc agtggaaacga 7500
 aaactcacgt taagggattt tggatcatgag attatcaaaa aggatcttca cctagatcct 7560
 tttaaattaa aaatgaagtt ttaaatcaat ctaaagtata tatgagtaaa cttggtctga 7620
 cagttaccaa tgcttaataca gtgaggcacc tatctcagcg atctgtctat ttcgttcatc 7680
 catagttgcc tgactccccg tcgtgtagat aactacgata cgggagggct taccatctgg 7740
 cccagtgct gcaatgatac cgcgagacc acgctcaccg gctccagatt tatcagcaat 7800
 aaaccagcca gccggaaggg ccgagcgcag aagtggctct gcaactttat ccgcctccat 7860
 ccagtctatt aattgttgc gggaaagctag agtaagtagt tcgccagtta atagtttgcg 7920
 caacgttgtt gccattgcta caggcatcgt ggtgtcacgc tcgtcgtttg gtatggcttc 7980
 attcagctcc ggttcccaac gatcaaggcg agttacatga tccccatgt tgtgcaaaaa 8040
 agcggttagc tccttcggtc ctccgatcgt tgtcagaagt aagtggccg cagtgtatc 8100
 actcatggtt atggcagcac tgcataattc tcttactgtc atgccatccg taagatgctt 8160
 ttctgtgact ggtgagtact caaccaagtc attctgagaa tagtgtatgc ggcgaccgag 8220
 ttgctcttgc ccggcgtcaa tacgggataa taccgcgcca catagcagaa ctttaaaagt 8280
 gctcatcatt ggaaaacggt ctccggggcg aaaactctca aggatcttac cgctgttgag 8340
 atccagttcg atgtaaccca ctcggtcacc caactgatct tcagcatctt ttactttcac 8400
 cagcgtttct gggtgagcaa aaacaggaag gcaaaatgcc gcaaaaaagg gaataagggc 8460
 gacacgaaa tgttgaatac tcatactctt cctttttcaa tattattgaa gcatttatca 8520
 gggttattgt ctcagagcg gatacatatt tgaatgtatt tagaaaaata aacaaatagg 8580
 ggttccgcgc acatttcccc gaaaagtgcc acctgacgtc taagaaacca ttattatcat 8640
 gacattaacc tataaaaata ggcgtatcac gaggccttt cgtc 8684

<210> 44

5

<211> 12314

ES 2 607 894 T3

<212> ADN

<213> Artificial

<220>

5

<223> pBOL115

<400> 44

tcgcgcgttt cggatgatgac ggtgaaaacc tctgacacat gcagctcccg gagacgggtca	60
cagcttgtct gtaagcggat gccgggagca gacaagcccg tcagggcgcg tcagcgggtg	120
ttggcgggtg tcggggctgg cttaactatg cggcatcaga gcagattgta ctgagagtgc	180
accataccac agcttttcaa ttcaatcoat cttttttttt ttattctttt ttttgatttc	240
ggtttctttg aaattttttt gattcggtaa tctccgaaca gaaggaagaa cgaaggaagg	300
agcacagact tagattggta tatatacgca tatgtagtgt tgaagaaaca tgaaattgcc	360
cagtattctt aacccaactg cacagaacaa aaacctgcag gaaacgaaga taaatcatgt	420
cгааagctac atataaggaa cgtgctgcta ctcatcctag tcctgttgct gccaaactat	480
ttaatatcat gcacgaaaag caaacaaact tgtgtgcttc attggatggt cgtaccacca	540
aggaattact ggagttagtt gaagcattag gtcccaaaat ttgtttacta aaaacacatg	600
tggatatctt gactgatttt tccatggagg gcacagttaa gccgctaaag gcattatccg	660
ccaagtacaa ttttttactc ttcgaagaca gaaaatttgc tgacattggt aatacagtca	720
aattgcagta ctctgcgggt gtatacagaa tagcagaatg ggcagacatt acgaatgcac	780
acggtgtggt gggcccagggt attgttagcg gtttgaagca ggcggcagaa gaagtaacaa	840
aggaacctag aggccttttg atgttagcag aattgtcatg caagggtcc ctatctactg	900
gagaatatac taagggtact gttgacattg cgaagagcga caaagatttt gttatcggct	960
ttattgctca aagagacatg ggtggaagag atgaaggta cgattggttg attatgacac	1020
ccggtgtggg tttagatgac aaggagagcg cattgggtca acagtataga accgtggatg	1080
atgtggtctc tacaggatct gacattatta ttgttgaag aggactatth gcaaagggaa	1140

ES 2 607 894 T3

```

gggatgctaa ggtagagggg gaacgttaca gaaaagcagg ctgggaagca tatttgagaa 1200
gatgcggcca gcaaaactaa aaaactgtat tataagtaaa tgcattgata ctaaactcac 1260
aaattagagc ttcaatttaa ttatatcagt tattacccta tgcggtgtga aataccgcac 1320
agatgcgtaa ggagaaaata ccgcatcagg aaattgtaaa cgtaaatatt ttgttaaaat 1380
tcgcggttaa tttttgttaa atcagctcat tttttaacca ataggccgaa atcggcaaaa 1440
tcccttataa atcaaaagaa tagaccgaga tagggttgag tgttgttcca gtttggaaca 1500
agagtccact attaagaac gtggactcca acgtcaaagg gcgaaaaacc gtctatcagg 1560
gcatggtccc actacgtgaa ccatcacct aatcaagttt tttgggtcgc aggtgctgta 1620
aagcactaaa tcggaaccct aaagggagcc cccgatttag agcttgacgc ggaaagccgc 1680
cgaacgtggc gagaaaggaa ggaagaaag cgaaaggagc gggcgctagg gcgctgcaa 1740
gtgtagcggc cacgctgccc gtaaccacca caccgcccgc gcttaatgag ccgctacagg 1800
gcgctgccc ccattcgcca ttcagctgac gcaactgttg ggaagggcga tcggtgcccg 1860
cctcttcgct attacgccag ctggcgaaag ggggatgtgc tgcaaggcga ttaagttggg 1920
taacgccagg gtttccag tcacgacgtt gtaaaacgac ggccagtgag cgcgctaat 1980
acgactcact ataggcgcaa ttgggtaccg ggccccct cgaggtcgac ggtatcgata 2040
agcttgatat cgaattcctg cagcccggg gatccactag ttctagagcg gccatttaa 2100
acggccggcc ctagatcaga ggggtgtaa tgaagtgtaa tagtattcat tttcttata 2160
aatcatcctt tccgtgattt atacaaaaga agaggagaat atgctgaata cttggtatat 2220
tactctacat tatactctta tcttgacggg tattctgagc atcttactca gtttcaagat 2280
cttttaatgt ccaaaaacat ttgagccgat ctaaatactt ctgtgttttc attaatat 2340
aaattgtact ctttaagac atggaaagta ccaacatcgc ttgaaacagt tttcattta 2400
cttatggttt attggttttt ccagtgaatg attatgtgc gttacccttt cgtaaaagtt 2460
caaacacgtt ttaagtatt gtttagttgc tcttccgaca tatatgatta tccctgccc 2520
gctaaagtta aggatgcaaa aaacataaga caactgaagt taatttacgt caattaagtt 2580
ttccagggtg atgatgtttt gggcttccac taattcaata agtatgtcat gaaatcagtt 2640
gtgaagagca tccagaaata atgaaaagaa acaacgaaac tgggtcggcc tgttgtttct 2700
tttcttacc acgtgatctg cggcatttac aggaagtcgc gcgttttgcg cagttgttgc 2760
aacgcagcta cggtaacaa agcctagtgg aactcgactg atgtgttagg gcctaaaact 2820

```

ES 2 607 894 T3

```

ggtggtgaca gctgaagtga actattcaat ccaatcatgt catggctgtc acaaagacct 2880
tgcggaccgc acgtacgaac acatacgtat gctaatatgt gttttgatag taccagtgga 2940
tcgcagacct gcaatTTTTT tgtaggTTTg gaagaatata taaaggTTgC actcattcaa 3000
gatagTTTTT ttcttTgtgTg tctattcatt ttattattgt ttgtttaaat gttaaaaaaa 3060
ccaagaactt agtttcaaTt taaattcatc acacaaacaa acaaaacaaa atgaacattg 3120
ttgTTTgTtTt gaagcaagTt ccagacactg ctgaagtcag aattgaccca gtcaagggta 3180
ctttaatcag agaaggtgTt ccatctatca tcaaccaga cgacaagaac gctttggaag 3240
aagctTTTgTt tttgaaggac aactacggTg ctcacgttac cgtcatttcc atgggtccac 3300
ctcaagccaa gaacgctTTg gttgaagctt tggccatggg tgctgatgaa gctgtcttat 3360
tgactgacag agctttcggt ggtgctgata ctttagctac ctctcacacc attgctgctg 3420
gtatcaagaa attgaaatac gatatcgtct ttgccggtcg tcaagccatc gatggtgata 3480
ccgctcaagt cggTccagaa attgctgaac atttgggtat tccacaagtc acctacgttg 3540
aaaaggTtga agttgacggt gacacttTga agatcagaaa ggcttgggaa gacggTtacg 3600
aagTtGtTga agtcaagact ccagTtctat tgactgccat caaggaattg aacgttccaa 3660
gatacatgtc cgttgaaaag atcttcggTg ctttcgacaa ggaagtcaag atgtggactg 3720
ctgatgatat cgatgtcgac aaggccaact tgggtttgaa aggttctcca accaaggtca 3780
agaaatcttc taccaaggaa gtcaagggtc aaggTgaagt cattgacaaa ccagtcaagg 3840
aagctgccgc ttacgtTgtt tccaagTtga aggaagaaca ctacatctaa agccccggcg 3900
gagattgata agactTTTTt agttgcatat cttttatatt taaatcttat ctattagtta 3960
atTTTTtGta atttatcctt atatatagtc tggTtattct aaaatatcat ttCagtatct 4020
aaaaattccc ctctTTTTtC agttatatct taacaggcga cagtccaaat gttgatttat 4080
cccagtccga ttcacTcaggg ttgtgaagca tttTgtcaat ggtcgaaatc acatcagtaa 4140
tagTgcctct tactTgcctc atagaatttc tttctcttaa cgtcaccgTt tggTctTTta 4200
tagTttcgaa atctatggTg ataccaaTg gtgttcccaa ttcacTgTta cgggCgtatt 4260
TTTTaccaat tgaagTattg gaatcgtcaa ttttaaagta tatctctctt ttacgTaaag 4320
cctgcgagat cctctTaaTg atagcgggga agccatcgtt attcgatatt gtcgTaaCa 4380
atactttgat cggcgtatg tttaaTgTt taaacatgga cagatatgCg atgaaaacgc 4440
taagtgatac tccaaTggt gaaaggtacg atgcttgGaa acaatactTg gaaatcaccg 4500
gaaacaccat atgcggcgaa aagccaatta gtgtgatact aagTgctTta tcgaaaatcc 4560

```

ES 2 607 894 T3

gtgatgccgg tccttcaggc atcaaatttc agtggcctaa ttattcacag agttctcatg 4620
 tgacaagtat tgatgatagt agtgtcagtt atgcttcagg ttatgttact ataggataat 4680
 gatcacggct aaaacggtcg aatgtaagca tatatctttc gattgtataa ttgttcccaa 4740
 atactacagc atctcaagga aaaaaaaca aaaacttcca aaaaaatcga atccctgagg 4800
 aatctttaat acattttcaa tctatttaag ttttataaac gtgtatatga gatgtcatga 4860
 gcatgaatta ttaataataa aaactaaatc attaaagtaa cttaaggagt taaagcccgg 4920
 gctttaattg ttagcagcct tgacttgagc aatcaattct ggaacaacct tgttgacatc 4980
 accgacaatg gccaaatcag caaccttcat gattggagct tcgacatctt tgttgatggc 5040
 aatgatgtag tcagagtctt gcataccagc caagtgttgg atggcaccag agataccaca 5100
 agcaatgtac aaagttggtc tgacggtctt accggtttga ccgacttgca agtccttgtc 5160
 aaccattcc ttttcaatgg cagctctgga agcagcaatg gtaccacca acaagaagc 5220
 taattcttcc aatttttcga agttttcctt ggaaccaaca ccacgaccac cagcaacca 5280
 aaccttggct tcaccgatat cagcaatgct cttggccaat ttgacaacct tggaaacctt 5340
 ggttctgata tcagaagcag tcaatttgat ggcaaccttt tcgatcttgt catcagaaac 5400
 gttagcatcg ttaactggca atttttcaaa gacacctggt ctgacggtgg ccatttgagg 5460
 tctgtggtca gaacagacia tggtagcaat caagttacca ccgaaagctg gtctggtagc 5520
 caacaagtca cggttttcga catcgatatc caaagaggtc cagtcagcag tcaaaccagt 5580
 agacaatctg gcagcaatc ttggaccaa gtctctaccg atgaaagtag caccgatgaa 5640
 taagatttct ggctttctt cgttgaccaa gtcacagata accttggcgt aaccgtcagt 5700
 ggagaaatga gctaataatt cgttgtcagc agccaaaacc ttgtcagcac cgtgggacia 5760
 caagtccttg gacatctttt cagtgttgtg acccaataag acagcagtca attcaacacc 5820
 caatttttca gccatttctt tacccttacc tagcaattcc aaagaacct tttgtaattc 5880
 accatctctt tgttcagcga aaaccagac acccttgtag tcagccttgt tcatgtttag 5940
 ttaattatag ttcggtgacc gtatattcta aaaacaagta ctccttaaaa aaaaaccttg 6000
 aagggaataa acaagtagaa tagatagaga gaaaaataga aaatgcaaga gaatttatat 6060
 attagaaaga gagaagaaa aatggaaaaa aaaaaatagg aaaagccaga aatagcacta 6120
 gaaggagcga caccagaaaa gaaggtgatg gaaccaatth agctatatat agttaactac 6180
 cgctcgcac atctctgcct ccagcatagt cgaagaagaa tttttttttt cttgaggctt 6240

ES 2 607 894 T3

ctgtcagcaa ctcgatTTTT ttctttcttt tttggtgagc ctaaaaagtt cccacgttct	6300
cttgtacgac gccgtcacia acaaccttat gggtaatTTg tcgCGgtctg ggtgtataaa	6360
tgtgtgggtg caggccggcc gtttaaaccg gccgccaccg cggTggagcc tgtgtggaag	6420
aacgattaca acaggtgTTg tcctctgagg acataaaata cacaccgaga ttcatcaact	6480
cattgctgga gttagcatat ctacaattgg gtgaaatggg gagcgatttg caggcatttg	6540
ctcgcatgc cggtagaggt gtggTcaata agagcgacct catgctatac ctgagaaagc	6600
aacctgacct acaggaaaga gttactcaag aataagaatt ttcgTTTTaa aacctaagag	6660
tcactTTaaa atttgtatac acttattTTTT ttataactt atttaataat aaaaatcata	6720
aatcataaga aattcgctcg agtcgactgc agtttactgc ttgtagtcgt aagaagTTTg	6780
gatgatatcc ttgatttcag aaatcaaagc ttctTTTggg ttggcagtgg tacattggtc	6840
ttcgaagcc aattcagcca ttctgtcaat ggattcgttc aattcttctt cagagacacc	6900
ttgagatttc aagttcattt caataccaac agattgacct aattcgtaga cagcctTggc	6960
caaagattca accaaagcct cagtggTgtt acctttcaaa cctaagaact tggcgatac	7020
agcgtaatcg gtgtcagctc tgaagaatc gtactTTggg aacaaagcgt gctTTTgagg	7080
gtcctTggcg ttgtatctga tgatgtgagg caacaagatg gcgTtagctc taccatgtgg	7140
aataccgtat tcaccaccaa ttttTgagc aatggagtga gcaatacca agaaagcgtt	7200
agcaaagcc ataccggcca aagtagaagc gttgtgcatc ttttctctgg aaacctTgtc	7260
acctTTTTca acggaagatt tcaagtattc aaaggtcaat ttgatagctt gtagggacia	7320
acctctggtg taatcggagg ccatgacaga aacgtaagat tccatagcgt gagtcaaaac	7380
gtccataccg gtatcagcag tgacagattt tgggacggac atgacaaatt gagggTcaat	7440
gatggcgaca tctggagtca aagcgaatc agccaatggg tatttgacgt tggTTTcaga	7500
gtcagtgata acagcaaTg gagtgacttc agaaccagta ccagaagtgg ttggaataca	7560
gatgaaagTg gcgTTTTctg gcataccaat cttgtaggtt ctcttaccaa tgtccaagaa	7620
TTTTTgctTg gcaccgaaga aagaagTTTc tggatgtTcg aagaacatcc acatagcctt	7680
ggcagcatcc atggcagaac caccaccaa agcaatgatg gtgtctggTt ggaaatcgac	7740
catcattTcc aaacctTgt agacagtgtt ggtggatgga tttggtTcaa cttcagagaa	7800
gatcttgatt tgaggTTgtt cagttctTTg acgtaagacg ttttcgacgg tttTggtgta	7860
accaaattca accatacctg ggtcaciaac gatcatgacc ttttcaatct tgtccatggt	7920
ggTcaagGac atgatagcgt tttctTcGaa atagattTga gctggaacct tgaagattTg	7980

ES 2 607 894 T3

```

agtgttgttt cttctcttgg caatggtcct gatgttcaac aaatcggtag cagagacggt 8040
gtgggagatg gagtttctac cgtaagaacc acaacccaaa gtcaaggatg gaatcaattc 8100
gttgtacatg tcaccgatac caccaacagc agatggagtg ttgacccaaa cagacaagc 8160
cttcattctt agaccgaaat ctttttcaa agtttcgtct tcagtgtgga taacagcagt 8220
gtgacctaaa ccaccgaagt gcaaagtgtc ttcacagatt tggaaagctt gcttggtaga 8280
ttgagccttg actaaagcca aaactggaga caacttttct ctggacaatg ggtagttaga 8340
accgacaccg gagatttcag caatgatcaa tttggtgttt tctggaactg ggataccggc 8400
caattcagca atttcaacag cagacttacc aacgatatct ggcttgatac cagtcttttg 8460
ttcgttcatg atggcgtttt ctaatctttg taattcgtcc ttcttgacga agtaagcttg 8520
gtgagccttg aattcattgg tgacatcctt gtagatttcc ttgtcaatga caacaacttg 8580
ttcagaagca cagatcatac cattatcgaa agtcttgaa cccgatgatc cgttgacagc 8640
acgcttgata tgagcagtct tttcagatga agatggaacg ttacctggac caacacccaa 8700
agctggttta ccagtggagt aagcagactt aaccatacca gaaccaccgg tagccaagac 8760
taaagcaata cccttgtggt tcatcaattg cttggtagct tcaatggatg gaacttcaat 8820
ccattggatg atatcctttg gagcaccagc cttcatggca gcttccaaaa caacttcagc 8880
agctctcttg gaagattcct gagcagatgg gtggaaagcg aaaataattg ggttaccagt 8940
cttgatggca atcatagcct tgaagatggt ggtagaagtt gggttggtgg ttggagtgac 9000
accacagatg acaccaattg gttcagcaac gtaggtcaaa cccttttctt tgtcttcacc 9060
aatgatacca acagtcttgt tgtccttgat ggagttccag atgtattcag aggcgtataa 9120
gttcttgata gccttgtcct cgtagatacc tctaccggtt tcttcatgag ccaacttggc 9180
caaaacatg tgttgggtcaa cagcagccaa ggacatttgg tggacaatgt ggtcaatttc 9240
ttcttgagac ttcttggaca aagcttccaa agccttctta cctttgtcag ctaaagcatc 9300
aatcatgatg gcaacttctt gttccttggg acctctgttt tccttttctg gaatggtcaa 9360
cattttttac tagttctaga atccgtcgaa actaagttct ggtgttttaa aactaaaaaa 9420
aagactaact ataaaagtag aatttaagaa gtttaagaaa tagatttaca gaattacaat 9480
caatacctac cgtctttata tacttattag tcaagtaggg gaataatttc agggaactgg 9540
tttcaacctt ttttttcagc tttttccaaa tcagagagag cagaaggtaa tagaaggtgt 9600
aagaaaatga gatagataca tgcgtgggtc aattgccttg tgcatcatt tactccaggc 9660

```


ES 2 607 894 T3

```

aggttgcatc actccattga ggttgtgcc gttttttgcc tgtttgtdcc cctgttctct 9720
gtagttgdcg taagagaatg gacctatgaa ctgatggttg gtgaagaaaa caatattttg 9780
gtgctgggat tctttttttt tctggatgcc agcttaaaaa gcgggctcca ttatatttag 9840
tggatgccag gaataaactg ttcaccaga cacctacgat gttatatatt ctgtgtaacc 9900
cgccccctat tttgggatg tacgggttac agcagaatta aaaggctaata tttttgacta 9960
aataaagtta ggaaaatcac tactattaat tatttacgta ttctttgaaa tggcgagtat 10020
tgataatgat aaactgagct ccagcttttg ttccctttag tgagggttaa ttgdcgctt 10080
ggcgtaatca tggcatagc tgtttctgt gtgaaattgt tatccgctca caattccaca 10140
caacatagga gccggaagca taaagtgtaa agcctgggt gcctaagtag tgaggtaact 10200
cacattaatt gcgttgctc cactgccgc tttccagtcg gaaacctgt cgtgccagct 10260
gcattaatga atcgccaac gcgcgggag agcggtttg cgtattggg gctctccgc 10320
ttctctctc actgactcgc tgcctcggg cgttcggctg cggcgagcgg tatcagctca 10380
ctcaaaggcg gtaatacggg tatccacaga atcaggggat aacgcaggaa agaactgtg 10440
agcaaaaggc cagcaaaagg ccaggaaccg taaaaggcc gcgttgctgg cgtttttcca 10500
taggctccgc cccctgacg agcatcacia aaatcgacgc tcaagtcaga ggtggcgaia 10560
cccgacagga ctataagat accagcgtt tccccctgga agctccctcg tgcgctctcc 10620
tgttccgacc ctgccctta ccggatacct gtcgccttt ctccctcgg gaagcgtggc 10680
gctttctcat agctcacgct gtaggtatct cagttcggg taggtcgtt gctccaagct 10740
gggctgtgtg cacgaacccc ccgttcagcc cgaccgctgc gccttatccg gtaactatcg 10800
tcttgagtc aaccggtaa gacacgactt atcgccactg gcagcagcca ctggtaacag 10860
gattagcaga gcgaggtat taggcggtgc tacagagttc ttgaagtggg ggcctaacta 10920
cggctacact agaaggacag tatttggtat ctgcgctctg ctgaagccag ttacctcgg 10980
aaaaagagtt gtagctctt gatccggcaa acaaacacc gctggtagcg gtggtttttt 11040
tgtttgcaag cagcagatta cgcgcagaaa aaaaggatct caagaagatc ctttgatctt 11100
ttctacggg tctgacgctc agtggaacga aaactcacgt taagggattt tggteatgag 11160
attatcaaaa aggatcttca cctagatcct tttaaattaa aaatgaagtt ttaaatcaat 11220
ctaaagtata tatgagtaaa cttggtctga cagttaccaa tgcttaatca gtgaggcacc 11280
tatctcagcg atctgtctat ttcgttcac catagttgcc tgactcccc tcgtgtagat 11340
aactacgata cgggagggt taccatctgg cccagtgct gcaatgatac cgcgagacc 11400

```

ES 2 607 894 T3

acgctcaccg gctccagatt tatcagcaat aaaccagcca gccggaaggg ccgagcgcag 11460
aagtggctct gcaactttat ccgectccat ccagtetatt aattggtgcc gggaagctag 11520
agtaagtagt tcgccagtta atagtttgcg caacgttggt gccattgcta caggcatcgt 11580
ggtgtcacgc tcgtcgtttg gtatggcttc attcagctcc ggttcccaac gatcaaggcg 11640
agttacatga tccccatgt tgtgcaaaaa agcggttagc tccttcggtc ctccgatcgt 11700
tgtcagaagt aagttggccg cagtgttata actcatggtt atggcagcac tgcataattc 11760
tcttactgtc atgcatccg taagatgctt ttctgtgact ggtgagtaact caaccaagtc 11820
attctgagaa tagtgtatgc ggcgaccgag ttgctcttgc ccggcgtcaa tacgggataa 11880
taccgcgcca catagcagaa ctttaaaagt gctcatcatt ggaaaacggt ctccggggcg 11940
aaaactctca aggatcttac cgctgttgag atccagttcg atgtaacca ctctgtcacc 12000
caactgatct tcagcatctt ttactttcac cagcgtttct gggtgagcaa aaacaggaag 12060
gcaaaatgcc gcaaaaaagg gaataagggc gacacggaaa tgttgaatac tcatactctt 12120
cctttttcaa tattattgaa gcatttatca gggttattgt ctcatgagcg gatacatatt 12180
tgaatgtatt tagaaaaata aacaaatagg ggttccgcgc acatttcccc gaaaagtgcc 12240
acctgacgtc taagaaacca ttattatcat gacattaacc tataaaaata ggcgtatcac 12300
gaggcccttt cgtc 12314

<210> 45

<211> 11180

5

<212> ADN

<213> Artificial

<220>

<223> pBOL116

10

<400> 45

tcgcgcgttt cggatgatgac ggtgaaaacc tctgacacat gcagctcccg gagacggtca 60
cagcttgtct gtaagcggat gccgggagca gacaagcccg tcagggcgcg tcagcgggtg 120
ttggcgggtg tcggggctgg cttaactatg cggcatcaga gcagattgta ctgagagtgc 180
accataccac agcttttcaa ttcaattcat cttttttttt ttattctttt ttttgatttc 240
ggtttctttg aaattttttt gattcggtaa tctccgaaca gaaggaagaa cgaaggaagg 300
agcacagact tagattggta tatatacgca tatgtagtgt tgaagaaca tgaaattgcc 360
cagtattctt aacccaactg cacagaacaa aaacctgcag gaaacgaaga taaatcatgt 420

ES 2 607 894 T3

cgaaagctac atataaggaa cgtgctgcta ctcatcctag tctgttgct gccaaagctat	480
ttaatatcat gcacgaaaag caaacaact tgtgtgcttc attggatggt cgtaccacca	540
aggaattact ggagttagtt gaagcattag gtcccaaat ttgtttacta aaaacacatg	600
tgatatctt gactgatttt tccatggagg gcacagttaa gccgctaaag gcattatccg	660
ccaagtacaa ttttttactc ttcgaagaca gaaaatttgc tgacattggt aatacagtca	720
aattgcagta ctctgcgggt gtatacagaa tagcagaatg ggcagacatt acgaatgcac	780
acgggtggtt gggcccagggt attgttagcg gtttgaagca ggcggcagaa gaagtaacaa	840
aggaacctag aggccctttg atgttagcag aattgtcatg caagggtcc ctatctactg	900
gagaatatac taagggtact gttgacattg cgaagagcga caaagatttt gttatcggct	960
ttattgctca aagagacatg ggtggaagag atgaaggta cgattgggtt attatgacac	1020
ccggtgtggg ttttagatgac aaggagagcg cattgggtca acagtataga accgtggatg	1080
atgtggtctc tacaggatct gacattatta ttgttgaag aggactattt gcaaaggaa	1140
gggatgctaa ggtagagggt gaacgttaca gaaaagcagg ctgggaagca tatttgagaa	1200
gatgcggcca gcaaaactaa aaaactgtat tataagtaaa tgcattgata ctaaactcac	1260
aaattagagc ttcaatttaa ttatatcagt tattacccta tgcggtgtga aataccgcac	1320
agatgcgtaa ggagaaaata ccgcatcagg aaattgtaaa cgtaaatatt ttgttaaaat	1380
tcggtttaa tttttgttaa atcagctcat tttttaacca ataggccgaa atcgcaaaa	1440
tcccttataa atcaaaagaa tagaccgaga tagggttgag tgttgttcca gtttgaaca	1500
agagtccact attaagaac gtggactcca acgtcaaagg gcgaaaacc gtctatcagg	1560
gcgatggccc actacgtgaa ccatcacctc aatcaagttt tttgggtcag aggtgcccga	1620
aagcactaaa tcggaaccct aaaggagacc cccgatttag agcttgacgg ggaaagccgg	1680
cgaacgtggc gagaaaggaa ggaagaaag cgaagagac gggcgctagg gcgctgcaa	1740
gtgtagcggc cacgctgccc gtaaccacca caccgcccgc gcttaatgag ccgctacagg	1800
gcgctgccc ccattgccca ttcaggctgc gcaactgtt ggaagggcga tcggtgccc	1860
ctctctcgtc attacgccag ctggcgaaag ggggatgtgc tgcaaggcga ttaagtggg	1920
taacgccagg gtttccag tcacgacgtt gtaaaacgac ggccagttag cgcgctaata	1980
acgactcact ataggcgcaa ttgggtaccg ggccccctc cgaggtcgac ggtatcgata	2040
agcttgatat cgaattcctg cagcccggg gatccactag ttctagagcg gccatttaa	2100
acggccggcc ctagatcaga ggggtgtaaa tgaagtgtaa tagtattcat ttttctata	2160

ES 2 607 894 T3

aatcatccct tccgtgattt atacaaaaga agaggagaat atgctgaata cttggtatat	2220
tactctacat tatactctta tcttgacggg tattctgagc atcttactca gtttcaagat	2280
cttttaaatgt ccaaaaacat ttgagccgat ctaaatactt ctgtgttttc attaatattat	2340
aaattgtact cttttaagac atggaaagta ccaacatcgg ttgaacacgt ttttcattta	2400
cttatggttt attggttttt ccagtgaatg attatattgtc gttacccttt cgtaaaagtt	2460
caaacacgtt tttaagtatt gtttagttgc tctttcgaca tatatgatta tccctgcgcg	2520
gctaaagtta aggatgcaaa aaacataaga caactgaagt taatttacgt caattaagtt	2580
ttccagggta atgatgtttt gggcttccac taattcaata agtatgtcat gaaatacgtt	2640
gtgaagagca tccagaaata atgaaaagaa acaacgaaac tgggtcggcc tgttgtttct	2700
tttctttacc acgtgatctg cggcatttac aggaagtcgc gcgttttgcg cagttgttgc	2760
aacgcagcta cggctaacaa agcctagtgg aactcgactg atgtgttagg gcctaaaact	2820
ggtggtgaca gctgaagtga actattcaat ccaatcatgt catggctgtc acaaagacct	2880
tgcggaccgc acgtacgaac acatacgtat gctaatatgt gttttgatag taccagtgga	2940
tgcgagacct gcaatttttt tgtaggtttg gaagaatata taaaggttgc actcattcaa	3000
gatagttttt ttcttgtgtg tctattcatt ttattattgt ttgtttaaat gttaaaaaaa	3060
ccaagaactt agtttcaaat taaattcatc acacaaacaa acaaaacaaa atgaacattg	3120
ttgtttgitt gaagcaagtt ccagacactg ctgaagtcag aattgacca gtcaagggta	3180
ctttaatcag agaaggtggt ccatctatca tcaaccaga cgacaagaac gctttggaag	3240
aagctttggt tttgaaggac aactacggtg ctcacgttac cgtcatttcc atgggtccac	3300
ctcaagccaa gaacgctttg gttgaagcct tggccatggg tgcgatgaa gctgtcttat	3360
tgactgacag agctttcggg ggtgctgata ctttagctac ctctcacacc attgctgctg	3420
gtatcaagaa attgaaatac gatatcgtct ttgccggtcg tcaagccatc gatggtgata	3480
cgcctcaagt cggccagaa attgctgaac atttgggtat tccacaagtc acctacgttg	3540
aaaaggttga agttgacggt gacactttga agatcagaaa ggcttgggaa gacggttacg	3600
aagttgttga agtcaagact ccagttctat tgactgccat caaggaattg aacggtccaa	3660
gatacatgtc cgttgaaaag atcttcggtg ctttcgacaa ggaagtcaag atgtggactg	3720
ctgatgatat cgatgtcgac aaggccaact tgggtttgaa aggttctcca accaaggtca	3780
agaaatcttc taccaaggaa gtcaagggtc aaggtgaagt cattgacaaa ccagtcaagg	3840

ES 2 607 894 T3

aagctgccgc ttacgttggt tccaagttga aggaagaaca ctacatctaa agccccggcg 3900
gagattgata agacttttct agttgcatat cttttatatt taaatcttat ctattagtta 3960
atTTTTtgta atttatcctt atatatagtc tggttattct aaaatatcat ttcagtatct 4020
aaaaattccc ctcttttttc agttatatct taacaggcga cagtccaaat gttgatttat 4080
cccagtcgga ttcacacaggg ttgtgaagca ttttgtcaat ggtcgaatc acatcagtaa 4140
tagtgectct tacttgectc atagaatttc tttctcttaa cgtcacggtt tggctcttta 4200
tagtttcgaa atctatgggt ataccaaatg gtgttcccaa ttcacgtta cgggcgtatt 4260
ttttaccaat tgaagtattg gaatcgtcaa ttttaaagta tatctctctt ttacgtaaag 4320
cctgcgagat cctcttaagt atagcgggga agccatcgtt attcgatatt gtcgtaacaa 4380
atactttgat cggcgctatg tttaaagtgt taaacatgga cagatatgcg atgaaaacgc 4440
taagtgatac tccaaatggt gaaagtgacg atgcttgaa acaatacttg gaaatcaccg 4500
gaaacaccat atgcggcgaa aagccaatta gtgtgatact aagtgcctta tcgaaaatcc 4560
gtgatgccgg tccttcaggc atcaaatttc agtggcctaa ttattcacag agttctcatg 4620
tgacaagtat tgatgatagt agtgtcagtt atgcttcagg ttatgttact ataggataat 4680
gatcacggct aaaacggtcg aatgtaagca tatatctttc gattgtataa ttgttcccaa 4740
atactacagc atctcaagga aaaaaaaca aaaacttcca aaaaaatcga atccctgagg 4800
aatctttaat acattttcaa tctatttaag ttttataaac gtgtatatga gatgtcatga 4860
gcatgaatta ttaataataa aaactaaatc attaaagtaa cttaaggagt taaagcccgg 4920
gctttaattg ttagcagcct tgacttgagc aatcaattct ggaacaacct tgttgacatc 4980
accgacaatg gccaaatcag caaccttcat gattggagct tcgacatctt tgttgatggc 5040
aatgatgtag tcagagtctt gcataccagc caagtgttg atggcaccag agataccaca 5100
agcaatgtac aaagttggtc tgacggtctt accggtttga ccgacttgca agtccttgtc 5160
aaccattcc ttttcaatgg cagctctgga agcagcaatg gtaccacca acaaagaagc 5220
taattcttcc aatttttcga agttttcctt ggaaccaaca ccacgaccac cagcaaccaa 5280
aaccttgget tcaccgatat cagcaatgtc cttggccaat ttgacaacct tggaaacctt 5340
ggttctgata tcagaagcag tcaatttgat ggcaacctt tcgatcttgt catcagaaac 5400
gttagcatcg ttaactggca atttttcaaa gacacctggt ctgacggtgg ccatttgagg 5460
tctgtggtca gaacagacia tggtagcaat caagttacca ccgaaagctg gtctggtagc 5520
caacaagtca cggttttcga catcgatatc caaagaggt aagtcagcag tcaaaccagt 5580

ES 2 607 894 T3

```

agacaatctg gcagcaattc ttggacceca gtctctaccg atgaaagtag caccgatgaa 5640
taagatttct ggctttcttt cgttgacca gtcacagata accttggcgt aaccgtcagt 5700
ggagaaatga gctaataatt cgttgtcagc agccaaaacc ttgtcagcac cgtgggacaa 5760
caagtccttg gacatctttt cagtgttggtg acccaataag acagcagtca attcaacacc 5820
caatthttca gccatttctt tacccttacc tagcaattcc aaagaaacct tttgtaattc 5880
accatctctt tgttcagcga aaaccagac accctttag tagccttgt tcatgtttag 5940
ttaattatag ttcgttgacc gtatattcta aaaacaagta ctccttaaaa aaaaaccttg 6000
aagggaataa acaagtagaa tagatagaga gaaaaataga aaatgcaaga gaatttatat 6060
attagaaaga gagaagaaa aatggaaaaa aaaaaatagg aaaagccaga aatagcacta 6120
gaaggagcga caccagaaaa gaaggtgatg gaaccaatth agctatatat agttaactac 6180
cggctcgatc atctctgcct ccagcatagt cgaagaagaa thttthtttt cttgagcctt 6240
ctgtcagcaa ctcgtattht ttctthcttt tttggtgagc ctaaaaagtt cccacgttct 6300
ctgttacgac gccgtcacia acaaccttat gggtaattht tgcggtctg ggtgtataaa 6360
tgtgtgggtg caggccggcc gthtaaacgg gccgccaccg cgtgggagcc tgtgtggaag 6420
aacgattaca acaggtgtht tctctgaggt acataaaaata cacaccgaga thcatcaact 6480
cattgctgga gthtagatat ctacaattht gtgaaatggg gagcgttht caggcattth 6540
ctcggcatgc cgttagaggt gtggtcaata agagcgacct catgctatac ctgagaaagc 6600
aacctgacct acaggaaaga gthtactcaag aataagaatt thcgtthtaa aacctaaagag 6660
tcactthaaa atthgtatac actththttt thtataactt atthtaaat aaaaatcata 6720
aatcataaga aatthcgtctg agthgactgc agthttacaag thcaattht ccatgatggc 6780
cttaacaatg gcttgaacat ctthcgttht thctggtthc gctggagcag aagaagaagc 6840
agcaccgaca ccgaaagctt ctctgattth thctgacgtg gtggtaccgt aagcaacctt 6900
thctgatgtht aacaagthtt ctggaccaac gthgtcagag gtggcagaac caccaacagc 6960
accacaacct aaagtcaaa gthggactaa gthggtagca ccaccgatac caccaaaaga 7020
acctggagag thaaccaaaa thctggaaac aggtthctth aaagcaaat ctctaatgat 7080
thctthcgtth thgaggtgga thgatcaaggt thgaccagaa cthtggtht gcaataaagc 7140
caaagactth thcaagctt catgcccagth thctgacgtg thgaaagcca agactggagc 7200
caatthttcc thagcgtatg gaththtagg agaaacatc gthtthctcg atagtaagat 7260

```

ES 2 607 894 T3

aacagcatca gatggaatgg aaataccagc caatthggcc aaagcttggg catccttacc 7320
 aacgatggct gggthtgag taccgttggc acgtaataga atcttaccaa cctthtcaga 7380
 ttcttcagca ttcaagaagt aaccctthtg tctcttgaat tcttcgatga thtcagcctt 7440
 cttgacggtt tcagcaatga tggattgttc agaagcacag atgacaccgt tgtcgaagt 7500
 gtcagaaccg ataacctthc taacagcagt tggaatgtca gcagthctth cgatgaaaca 7560
 tggaccgtta cctggaccga caccgatggc tggagtacca gaggagtaag cagctctaac 7620
 cataccttca ccaccggtag ccaagatcaa agcgggtgtcc thgttcttca tcaattcagc 7680
 agtaccttca acggtcaaaa tggacataca thggatcaaa ccatctggag caccagcttc 7740
 aacagcagcc thttgcatga tcttaacggc thcagtgatg gaaccggacag cagthgggtg 7800
 tggagagaag acaatggcgt taccagcctt caaagcaatc aagacctga aatggcagt 7860
 ggaagthggg thggtagatg gaatcaaac agcaatgaca cctaatggga cagcaatgtc 7920
 aatcaatthc thttccttgt ctctcttcaa gataccaacg gtcttcaaat ccttgatgta 7980
 gthgtagaca acaatggagg agaathgtt cthgatgacc thgtcttccc atthaccgta 8040
 accggtgtct tcgtaagcca atthggccaa thtgacagct tcaacttcag tagccttggc 8100
 gatctthtcg atgaccttgt taacagcttc thgggaaaag thcttgaatt cagcttgagc 8160
 ctthctggcc thggcaatca aagthctaac thcttgatg gattgcaaat ccttgtccat 8220
 gattthcatt thttactagt tctagaatcc gtcgaaacta agthctgggt thttaaaact 8280
 aaaaaaaga ctaactataa aagtagaatt taagaagtht aagaataga thtacagaat 8340
 tacaatcaat acctaccgtc thtatatact tathagtcaa gtaggggaat aatthcaggg 8400
 aactgththc aacctththt thcagcttht tccaaatcag agagagcaga aggtaataga 8460
 aggtgtaaga aatgagata gatacatgcg tgggtcaatt gccttgtgtc atcatttact 8520
 ccaggcaggt tgcactcctc cattgaggtt gtgcccttht thtgctgtt tgtgccctg 8580
 thctctgtag thgcgctaag agaatggacc tatgaactga tggthgtgta agaaaacaat 8640
 atthtggtgc tgggattctt thththctg gatgccagct taaaagcgg gctccattat 8700
 atthagtgga tgcagggaat aaactgttca cccagacacc tacgatgtha tatattctgt 8760
 gtaaccgcc cctatththg ggcattgtac ggttacagca gaatthaaag gctaattht 8820
 tgactaaata aagthaggaa aatcactact athaattatt tacgtattct thgaaatggc 8880
 gagtattgat aatgataaac tgagctccag cththgttcc cthtagtgag ggttaattgc 8940
 gcgcttggcg taatcattgt catagctgtt tctgtgtga aatgttatc cgctcacaat 9000

ES 2 607 894 T3

tccacacaac ataggagccg gaagcataaa gtgtaaagcc tggggtgcct aatgagtgag	9060
gtaactcaca ttaattgcgt tgcgctcact gcccgctttc cagtcgggaa acctgtcgtg	9120
ccagctgcat taatgaatcg gccaacgcgc ggggagaggc ggtttgcgta ttgggcgctc	9180
ttccgcttcc tcgctcactg actcgctgcg ctccggtcgtt cggctgcggc gagcggatc	9240
agctcactca aaggcggtaa tacggttata cacagaatca ggggataacg caggaaagaa	9300
catgtgagca aaaggccagc aaaaggccag gaaccgtaaa aaggccgctg tgctggcgtt	9360
tttccatagg ctccgcccc ctgacgagca tcacaaaaat cgacgctcaa gtcagaggty	9420
gcgaaacccg acaggactat aaagatacca ggcggttccc cctggaagct ccctcgtgcg	9480
ctctcctgtt ccgaccctgc cgcttaccgg atacctgtcc gcctttctcc ctccggaag	9540
cgtggcgctt tctcatagct cagctgtag gtatctcagt tcggtgtagg tcgttcgctc	9600
caagctgggc tgtgtgcacg aacccccgt tcagcccgac cgctgcgcct tatccgtaa	9660
ctatcgtctt gagtccaacc cggtaagaca cgacttatcg cactggcag cagccactgg	9720
taacaggatt agcagagcga ggtatgtagg cgggtctaca gaggttctga agtggtgcc	9780
taactacggc tacactagaa ggacagtatt tggatatctgc gctctgctga agccagttac	9840
cttcggaaaa agagttggtg gctcttgatc cggcaaaaa accaccgctg gtagcggtyg	9900
tttttttgtt tgcaagcagc agattacgcg cagaaaaaaaa ggatctcaag aagatccttt	9960
gatcttttct acggggtctg acgctcagtg gaacgaaaa tcacgttaag ggattttggt	10020
catgagatta tcaaaaagga tcttcaccta gatcctttta aattaaaat gaagttttaa	10080
atcaatctaa agtatatatg agtaaacttg gtctgacagt taccaatgct taatcagtga	10140
ggcacctatc tcagcgatct gtctatctcg ttcatecata gttgcctgac tcccgcgtg	10200
gtagataact acgatacggg agggcttacc atctggcccc agtgctgcaa tgataccgcy	10260
agaccacgc tcaccggctc cagatttata agcaataaac cagccagccg gaagggccga	10320
gcgcagaagt ggtcctgcaa ctttatccgc ctccatccag tctattaatt gttgccggga	10380
agctagagta agtagttcgc cagttaatag tttgcgcaac gttggtgcca ttgctacagg	10440
catcgtggtg tcacgctcgt cgtttggtat ggcttcattc agctccggtt cccaacgatc	10500
aaggcgagtt acatgatecc ccatgttgty caaaaaagcg gttagctcct tcggctcctc	10560
gatcgttgty agaagtaagt tggccgcagt gttatcactc atggttatgg cagcactgca	10620
taattctctt actgtcatgc catccgtaag atgcttttct gtgactggtg agtactcaac	10680

ES 2 607 894 T3

caagtcattc tgagaatagt gtatgcggcg accgagttgc tcttgcccgg cgtcaatacg 10740
 ggataatacc ggcacacata gcagaacttt aaaagtgtc atcattggaa aacgttcttc 10800
 ggggcgaaaa ctctcaagga tcttaccgct gttgagatcc agttcgatgt aaccactcgc 10860
 tgcaccaaac tgatcttcag catcttttac tttcaccagc gtttctgggt gagcaaaaac 10920
 aggaaggcaa aatgccgcaa aaaaggaat aagggcgaca cggaaatggt gaatactcat 10980
 actcttcctt tttcaatatt attgaagcat ttatcagggg tattgtctca tgagcggata 11040
 catatttgaa tgtatttaga aaaataaaca aataggggtt cgcgcacat ttccccgaaa 11100
 agtgccacct gacgtctaag aaaccattat tatcatgaca ttaacctata aaaataggcg 11160
 taccacgagg ccctttcgtc 11180

<210> 46

<211> 11108

5

<212> ADN

<213> Artificial

<220>

<223> pBOL118

10

<400> 46

tcgcgcgttt cggatgatgac ggtgaaaacc tctgacacat gcagctcccg gagacggta 60
 cagcttgtct gtaagcggat gccgggagca gacaagcccg tcagggcgcg tcagcgggtg 120
 ttggcgggtg tcggggctgg cttaactatg cggcatcaga gcagattgta ctgagagtgc 180
 accataccac agcttttcaa ttcaattcat catTTTTTTT ttattctttt ttttgatttc 240
 ggtttctttg aaatTTTTTt gattcggtaa tctccgaaca gaaggaagaa cgaaggaagg 300
 agcacagact tagattggtat tatatacgca tatgtagtgt tgaagaaaca tgaattgcc 360
 cagtattctt aaccactcgc cacagaacaa aaacctgcag gaaacgaaga taaatcatgt 420
 cgaaagctac atataaggaa cgtgctgcta ctcatcctag tcctggtgct gccaaagctat 480
 ttaatatcat gcacgaaaag caaacaact tgtgtgcttc attggatggt cgtaccacca 540
 aggaattact ggagttagtt gaagcattag gtcccaaat ttgtttacta aaaacacatg 600
 tggatattctt gactgatttt tccatggagg gcacagttaa gccgctaaag gcattatccg 660
 ccaagtacaa ttttttactc ttcgaagaca gaaaatttgc tgacattggt aatacagtca 720
 aattgcagta ctctgcgggt gtatacagaa tagcagaatg ggcagacatt acgaatgcac 780
 acggtgtggt gggcccagggt attgtagcgt gtttgaagca ggcggcagaa gaagtaacaa 840
 aggaacctag aggccttttg atgtagcag aattgtcatg caagggtccc ctatctactg 900

ES 2 607 894 T3

```

gagaatatac taagggtact gttgacattg cgaagagcga caaagatfff gttatcggct    960
ttattgctca aagagacatg ggtggaagag atgaaggtta cgattgggtg attatgacac   1020
ccggtgtggg tttagatgac aaggagacg cattgggtca acagtataga accgtggatg   1080
atgtggtctc tacaggatct gacattatta ttggtggaag aggactatff gcaaagggaa   1140
gggatgctaa ggtagagggt gaacgttaca gaaaagcagg ctggaagca tatttgagaa   1200
gatgcgccca gcaaaactaa aaaactgtat tataagtaaa tgcatgtata ctaaactcac   1260
aaattagagc ttcaatftaa ttatatcagt tattacceta tgcggtgtga aataaccgac   1320
agatgcgtaa ggagaaaata ccgcatcagg aaattgtaaa cgftaatatt ttgttaaaat   1380
tcgcgftaaa tttttgttaa atcagctcat tttttaacca ataggccgaa atcggcaaaa   1440
tcccttataa atcaaaagaa tagaccgaga tagggttgag tgttgttcca gtttgaaca   1500
agagtccact attaaagaac gtggactcca acgtcaaagg gcgaaaaacc gtctatcagg   1560
gcatggtccc actacgtgaa ccatcacct aatcaagttt tttgggtcg aggtgccgta   1620
aagcactaaa tcggaacct aaagggagcc cccgatttag agcttgacgg ggaaagccgg   1680
cgaacgtggc gagaaaggaa gggaagaaag cgaaaggagc gggcgctagg gcgctggcaa   1740
gtgtagcggc cacgctgcgc gtaaccacca caccgcccgc gcttaatgcg ccgctacagg   1800
gcgctgcgc ccattcgcca ttcagctgc gcaactgtg ggaagggcga tcggtgcggg   1860
cctcttcgct attacgccag ctggcgaaag ggggatgtgc tgcaaggcga ttaagttggg   1920
taacgccagg gttttccag tcacgacgtt gtaaacgac ggccagtgag cgcgcgtaat   1980
acgactcact ataggcgaa ttgggtaccg ggccccct cgaggtcgac ggtatcgata   2040
agcttgatat cgaattcctg cageccgggg gatccactag ttctagagcg gccattftaa   2100
acggccggcc ctagatcaga gggtggtaaa tgaagtgtaa tagtattcat ttttcttata   2160
aatcatccct tccgtgattt atacaaaaga agaggagaat atgctgaata cttggtatat   2220
tactctacat tatactctta tcttgacggg tattctgagc atcttactca gtttcaagat   2280
cttttaatgt ccaaaaacat ttgagccgat ctaaatactt ctgtgttttc attaatftat   2340
aaattgtact cttttaagac atggaaagta ccaacatcgg ttgaaacagt ttttcattta   2400
cttatggttt attggttttt ccagtgaatg attatftgtc gttaccctft cgtaaaagtt   2460
caaacacgtt tftaagtatt gtttagttgc tctttcgaca tatatgatta tccctgcgcg   2520
gctaaagtta aggatgcaaa aaacataaga caactgaagt taatftacgt caattaagtt   2580

```

ES 2 607 894 T3

```

ttccagggta atgatgtttt gggcttccac taattcaata agtatgtcat gaaatacgtt 2640
gtgaagagca tccagaaata atgaaaagaa acaacgaaac tgggtcggcc tggtgtttct 2700
tttctttacc acgtgatctg cggcatttac aggaagtcgc gcgttttgcg cagttgttgc 2760
aacgcagcta cggctaacaa agcctagtgg aactcgactg atgtgttagg gcctaaaact 2820
ggtggtgaca gctgaagtga actattcaat ccaatcatgt catggctgtc acaaagacct 2880
tgcggaccgc acgtacgaac acatacgtat gctaatatgt gttttgatag taccacgtga 2940
tcgcagacct gcaatTTTTT tgtaggtttg gaagaatata taaaggttgc actcattcaa 3000
gatagttttt ttcttgtgtg tctattcatt ttattattgt ttgtttaaat gttaaaaaaa 3060
ccaagaactt agtttcaaat taaattcatc acacaaacaa acaaaacaaa atgaacattg 3120
ttgtttgitt gaagcaagtt ccagacactg ctgaagtcag aattgacca gtcaagggta 3180
ctttaatcag agaaggtgtt ccatctatca tcaaccaga cgacaagaac gctttggaag 3240
aagctttggt tttgaaggac aactacggtg ctcacgttac cgtcatttcc atgggtccac 3300
ctcaagccaa gaacgctttg gttgaagctt tggccatggg tgcctgatga gctgtcttat 3360
tgactgacag agctttcggg ggtgctgata ctttagctac ctctcacacc attgctgctg 3420
gtatcaagaa attgaaatac gatatcgtct ttgccggtcg tcaagccatc gatggtgata 3480
ccgctcaagt cggtcagaa attgctgaac atttgggtat tccacaagtc acctacgttg 3540
aaaagggtga agttgacggt gacacttga agatcagaaa ggcttgggaa gacggttacg 3600
aagttgttga agtcaagact ccagttctat tgactgccat caaggaattg aacgttccaa 3660
gatacatgtc cgttgaaaag atcttcggtg ctttcgacaa ggaagtcaag atgtggactg 3720
ctgatgatat cgatgtcgac aaggccaact tgggtttgaa aggttctcca accaaggtca 3780
agaaatcttc taccaaggaa gtcaagggtc aagggtgaagt cattgacaaa ccagtcaagg 3840
aagctgccgc ttacgttggt tccaagttga aggaagaaca ctacatctaa agcccgggcg 3900
gagattgata agacttttct agttgcatat cttttatatt taaatcttat ctattagtta 3960
atTTTTTgta atttatcctt atatatagtc tggttattct aaaatatcat ttcagtatct 4020
aaaaattccc ctctTTTTTc agttatatct taacaggcga cagtccaaat gttgatttat 4080
cccagtcoga ttcacaggg ttgtgaagca ttttgtcaat ggtcgaatc acatcagtaa 4140
tagtgcctct tacttgcctc atagaatttc tttctcttaa cgtcaccggt tggcttttta 4200
tagtttcgaa atctatggtg ataccaaatg gtgttcccaa ttcacgtta cgggcgtatt 4260
ttttaccaat tgaagtattg gaatcgtcaa ttttaaagta tatctctctt ttacgtaaag 4320

```

ES 2 607 894 T3

```

cctgcgagat cctcttaagt atagcgggga agccatcggt attcgatatt gtcgtaacaa 4380
atactttgat cggcgctatg tttaaatggt taaacatgga cagatatgcg atgaaaacgc 4440
taagtgatac tccaaatggt gaaaggtagc atgcttgga acaatacttg gaaatcaccg 4500
gaaacaccat atgcggcgaa aagccaatta gtgtgatact aagtgcctta tcgaaaatcc 4560
gtgatgccgg tccttcaggc atcaaatttc agtggcctaa ttattcacag agttctcatg 4620
tgacaagtat tgatgatagt agtgtcagtt atgcttcagg ttatgttact ataggataat 4680
gatcacggct aaaacggctc aatgtaagca tatacttttc gattgtataa ttgttcccaa 4740
atactacagc atctcaagga aaaaaaaaaa aaaacttcca aaaaaatcga atccctgagg 4800
aatctttaat acatcttcaa tctatttaag ttttataaac gtgtatatga gatgtcatga 4860
gcatgaatta ttaataataa aaactaaatc attaaagtaa cttaggagat taaagcccgg 4920
gctttaattg ttagcagcct tgacttgagc aatcaattct ggaacaacct tgttgacatc 4980
accgacaatg gccaaatcag caaccttcat gattggagct tcgacatctt tgttgatggc 5040
aatgatgtag tcagagtctt gcataccagc caagtgttgg atggcaccag agataccaca 5100
agcaatgtac aaagtggtc tgacggctt accggttga cgcacttga agtccttgtc 5160
aaccattcc ttttcaatgg cagctctgga agcagcaatg gtaccacca acaaagaagc 5220
taattcttcc aatttttcga agttttcctt ggaaccaaca ccacgaccac cagcaaccaa 5280
aaccttggct tcaccgatat cagcaatgtc cttggccaat ttgacaacct tggaaacctt 5340
ggttctgata tcagaagcag tcaatttgat ggcaaccttt tcgatcttgt catcagaaac 5400
gttagcatcg ttaactggca atttttcaa gacacctggc ctgacggtgg ccatttgagg 5460
tctgtggtca gaacagacaa tggtagcaat caagttacca ccgaaagctg gtctggtagc 5520
caacaagtca cggttttcga catcgatata caaagaggta cagtcagcag tcaaaccagt 5580
agacaatctg gcagcaattc ttggaccaa gtctctaccg atgaaagtag caccgatgaa 5640
taagatttct ggctttctt cgttgaccaa gtcacagata accttggcgt aaccgtcagt 5700
ggagaaatga gctaataatt cgttgtcagc agccaaaacc ttgtcagcac cgtgggacaa 5760
caagtcttg gacatctttt cagtgttgag acccaataag acagcagtca attcaacacc 5820
caatcttca gccatttctt tacccttacc tagcaattcc aaagaaacct tttgtaattc 5880
accatctctt tgttcagcga aaaccagac acctttagtag tcagccttgt tcatgtttag 5940
ttaattatag ttogttgacc gtatattcta aaaacaagta ctccttaaaa aaaaaccttg 6000

```

ES 2 607 894 T3

aaggaataa acaagtagaa tagatagaga gaaaaataga aaatgcaaga gaatttatat 6060
attagaaaaga gagaagaaaa aatggaaaaa aaaaaatagg aaaagccaga aatagcacta 6120
gaaggagcga caccagaaaa gaaggtgatg gaaccaatth agctatatat agttaactac 6180
cggctcgatc atctctgcct ccagcatagt cgaagaagaa tttttttttt cttgaggctt 6240
ctgtcagcaa ctcgtattht ttctttcttt tttggtgagc ctaaaaagtt cccacgttct 6300
cttgtagcag gccgtcacia acaaccttat gggtaatthg tcgctgctct ggtgtataaa 6360
tgtgtgggtg caggccggcc gtttaaacgg gccgccaccg cggaggagcc tgtgtggaag 6420
aacgattaca acaggtgthg tcctctgagg acataaaaata cacaccgaga ttcacaaact 6480
cattgctgga gttagcatat ctacaatthg gtgaaatggg gagcagatthg caggcattthg 6540
ctcggcatgc cggtagaggt gtggtcaata agagcgacct catgctatac ctgagaaagc 6600
aacctgacct acaggaaaga gttactcaag aataagaatt ttcgthttta aacctaagag 6660
tcactthtaa atthgtatac acttatthtt tttataactt atthtaaat aaaaatcata 6720
aatcataaga aatcgcctcg agtcgactgc agthtaaaaca atthctgaaag catcgaccaa 6780
aacacaacga cgtaatctga cgaaagthct ggcagaagtg acacctcac cagthggggt 6840
ggtgatggtc atggtggtcc aaccttcacc acccaaacc aaaccagcga tacatggacc 6900
gthcttgaca aagatggaag tgtcaatggc gthtagccatt tggthcatgt thctgatgt 6960
thctggagtgc atggcagcag tgtggtgaca accaccttc aatthgacag ccaaagcaat 7020
agcgtcagca acgthtagca cacggacaac tggtaagact ggcacatca atthcagtagc 7080
agcaaatggg thttcagcgg tggthtcgac gaataaat ctgthttctt tggaacctt 7140
caaaccgatg gcagcagcaa thttaccagc atctctacca acccagthctc tggagacggt 7200
accttacct thttcatcga thttcttcaa caaaactggt tgcaatthgt gagctthgtc 7260
agcagthcaac thgacggcat gthgacctc catcaatctc atcaatthct cagcaacgga 7320
gtcaacaaca atcaaacct thctgthcagc acagatgatg thgtthctga aagaagcacc 7380
cttgacaatg gathgagcag ctctggccaa atcagcggth thctcgacaa caacaggagg 7440
gthaccagea ccagcagcaa thaatctct gthggtgthc thctggcag thtcaacaac 7500
agctthacca ccagthcaga thaatagacc gatacctggg aactthgaaat atctthgagc 7560
agththcgata thctgggthg caacagthc caacaagtht thctggaccac cagcagcaac 7620
aatggctthg thcaatagag thgatgctct thgagaaacc thctthggcag thgggtgthg 7680
agcgaagata acggagthc cagcagcaat caaagagatg gcgthgthga thcagtagc 7740

ES 2 607 894 T3

agctgggttg gtagatgggg tgacggaagc aacaacaccc catggagcat tttcaatcaa	7800
agtcaaacca ttatcacccg tcaagacttg tggagacaaa cattcgacac ctggagtacc	7860
tctagcttga gcaacgttct tagcgaatth gtcttcaact ctaccatac cggtttcgga	7920
gacagccaat tcagccaagt ctctggcatg cttttcacca gcttctctga tggcagcaat	7980
ggccaattgt ctcatggcaa cagatttcaa accttggtga gcaaccttgg cagcagcaac	8040
agcgtcgtcc aaagaagcga aaacacccat ttcgtggaca gcagcagatg gagtgtcaga	8100
agattgcatt ttcaacaaga cagccttgac aacttggtcg atatcttgtt ggttcatttt	8160
ttactagttc tagaatccgt cgaaactaag ttctgggtgt ttaaaactaa aaaaaagact	8220
aactataaaa gtagaattta agaagtttaa gaaatagatt tacagaatta caatcaatac	8280
ctaccgtctt tatatactta ttagtcaagt aggggaataa tttcaggaa ctggtttcaa	8340
cctttttttt cagctttttc caaatcagag agagcagaag gtaatagaag gtgtaagaaa	8400
atgagataga tacatgctg ggtcaattgc cttgtgtcat catttactcc aggcaggttg	8460
catcactcca ttgaggttgt gcccgttttt tgccctgtttg tgcccctggt ctctgtagtt	8520
gcgctaagag aatggaccta tgaactgatg gttggtgaag aaaacaatat tttggtgctg	8580
ggattctttt tttttctgga tgccagctta aaaagcgggc tccattatat ttagtggatg	8640
ccaggaataa actgttcacc cagacaccta cgatgttata tattctgtgt aaccgccccc	8700
ctattttggg catgtacggg ttacagcaga attaaaaggc taattttttg actaaataaa	8760
gttaggaaaa tcaactactat taattattta cgtattcttt gaaatggcga gtattgataa	8820
tgataaactg agctccagct tttgttccct ttagtgaggg ttaattgctg ccttggcgta	8880
atcatggtca tagctgtttc ctgtgtgaaa ttgttatccg ctcaacaattc cacacaacat	8940
aggagccgga agcataaagt gtaaagcctg ggggtgcctaa tgagtgaggt aactcacatt	9000
aattgcgttg cgctcactgc ccgctttcca gtcgggaaac ctgctgtgcc agctgcatta	9060
atgaatcggc caacgcgcgg ggagagcgg tttgcgtatt gggcgcctct ccgcttctc	9120
gctcactgac tcgctgcgct cggtcgttcg gctgcggcga gcggtatcag ctcaactcaa	9180
ggcggtaata cggttatcca cagaatcagg ggataacgca ggaaagaaca tgtgagcaaa	9240
aggccagcaa aaggccagga accgtaaaaa ggccgcgttg ctggcgtttt tccataggct	9300
ccgccccct gacgagcatc acaaaaaatcg acgctcaagt cagaggtggc gaaaccgcac	9360
aggactataa agataccagg cgtttccccc tggaaagctcc ctctgtgcgct ctctgttcc	9420

ES 2 607 894 T3

```

gaccctgccg cttaccggat acctgtccgc ctttctccct tcggaagcg tggcgctttc 9480
tcatagctca cgctgtaggt atctcagttc ggtgtaggtc gttcgctcca agctgggctg 9540
tgtgcacgaa cccccgttc agcccgaccg ctgcgctta tccggtaact atcgtcttga 9600
gtccaaccg gtaagacacg acttatcgcc actggcagca gccactggta acaggattag 9660
cagagcgagg tatgtaggag gtgctacaga gttcttgaag tggggccta actacggeta 9720
cactagaagg acagtatttg gtatctgcgc tctgctgaag ccagttacct tcggaaaaag 9780
agttggtagc tcttgatccg gcaaacaaac caccgctggt agcggtggtt tttttgtttg 9840
caagcagcag attacgcgca gaaaaaaagg atctcaagaa gatcctttga tcttttctac 9900
ggggtctgac gctcagtgga acgaaaactc acgtaaggg attttggta tgagattatc 9960
aaaaggatc ttcacctaga tccttttaa ttaaaaatga agttttaat caatctaaag 10020
tatatatgag taaacttggc ctgacagtta ccaatgctta atcagtgagg cacctatctc 10080
agcgatctgt ctatttcgtt catccatagt tgcctgactc cccgtcgtgt agataactac 10140
gatacgggag ggcttaccat ctggccccag tgctgcaatg ataccgagag acccacgctc 10200
accggctcca gatttatcag caataaacca gccagccgga agggccgagc gcagaagtgg 10260
tcctgcaact ttatccgct ccatccagtc tattaattgt tgccgggaag ctagagtaag 10320
tagttcgcca gtaaatagtt tgcgcaacgt tgttgccatt gctacaggca tcgtggtgtc 10380
acgctcgtcg tttggtatgg cttcatcag ctccggttcc caacgatcaa ggcgagttac 10440
atgatcccc atgttgtgca aaaaagcggg tagctcctc ggtcctccga tcgttgtcag 10500
aagtaagttg gccgcagtgt tatcactcat ggttatggca gcaactgcata attctcttac 10560
tgatcatgca tccgtaagat gcttttctgt gactggtgag tactcaacca agtcattctg 10620
agaatagtgt atgcggcgac cgagttgctc ttgcccggcg tcaatacggg ataataccgc 10680
gccacatagc agaactttaa aagtgtcat cattggaaaa cgttcttcg ggcgaaaact 10740
ctcaaggatc ttaccgctgt tgagatccag ttcgatgtaa cccactcgtg caccctactg 10800
atcttcagca tcttttactt tcaccagcgt ttctgggtga gcaaaaacag gaaggcaaaa 10860
tgccgcaaaa aagggaataa gggcgacacg gaaatgttga atactcatac tcttctttt 10920
tcaatattat tgaagcattt atcagggtta ttgtctcatg agcggataca tatttgaatg 10980
tatttagaaa aataaaciaa taggggttcc gcgcacattt ccccgaaaag tgccacctga 11040
cgtctaagaa accattatta tcatgacatt aacctataaa aataggcgta tcacgaggcc 11100
ctttcgtc 11108

```

<210> 47

<211> 11114

5

<212> ADN

<213> Artificial

<220>

ES 2 607 894 T3

<223> pBOL120

<400> 47

tcgcgcgttt	cggtgatgac	ggtgaaaacc	tctgacacat	gcagctcccc	gagacggtca	60
cagcttgtct	gtaagcggat	gccgggagca	gacaagcccc	tcagggcgcg	tcagcgggtg	120
ttggcgggtg	tcggggctgg	cttaactatg	cggcatcaga	gcagattgta	ctgagagtgc	180
accataccac	agcttttcaa	ttcaattcat	catttttttt	ttattctttt	ttttgatttc	240
ggtttctttg	aaattttttt	gattcggtaa	tctccgaaca	gaaggaagaa	cgaaggaagg	300
agcacagact	tagattggta	tatatacgca	tatgtagtgt	tgaagaaaca	tgaaattgcc	360
cagtattctt	aaccaactg	cacagaacaa	aaacctgcag	gaaacgaaga	taaatcatgt	420
cgaagctac	atataaggaa	cgtgctgcta	ctcatcctag	tcctggtgct	gccaaagctat	480
ttaatatcat	gcacgaaaag	caaacaaact	tgtgtgcttc	attggatggt	cgtaccacca	540
aggaattact	ggagttagtt	gaagcattag	gtcccaaaat	ttgtttacta	aaaacacatg	600
tggatatctt	gactgattht	tccatggagg	gcacagttaa	gccgctaaag	gcattatccg	660
ccaagtacaa	ttttttactc	ttcgaagaca	gaaaatttgc	tgacattggt	aatacagtca	720
aattgcagta	ctctgcgggt	gtatacagaa	tagcagaatg	ggcagacatt	acgaatgcac	780
acggtgtggt	gggccaggt	attgttagcg	gtttgaagca	ggcggcagaa	gaagtaacaa	840
aggaacctag	aggccttttg	atgttagcag	aattgtcatg	caagggctcc	ctatctactg	900
gagaatatac	taagggtact	gttgacattg	cgaagagcga	caaagatttt	gttatcggct	960
ttattgctca	aagagacatg	ggtggaagag	atgaaggtta	cgattgggtg	attatgacac	1020
ccggtgtggg	tttagatgac	aagggagacg	cattgggtca	acagtataga	accgtggatg	1080
atgtggtctc	tacaggatct	gacattatta	ttgttggaag	aggactatth	gcaaagggaa	1140
gggatgctaa	ggtagagggt	gaacgttaca	gaaaagcagg	ctggaagca	tatttgagaa	1200
gatgcggcca	gcaaaactaa	aaaactgtat	tataagtaaa	tgcattgata	ctaaactcac	1260
aaattagagc	ttcaatttaa	ttatatcagt	tattacccta	tgcggtgtga	aataccgcac	1320
agatgcgtaa	ggagaaaata	cgcacacagg	aaattgtaaa	cgttaatatt	ttgttaaaat	1380

5

ES 2 607 894 T3

tcgcgttaaa tttttgtaa atcagctcat tttttaacca ataggccgaa atcggcaaaa	1440
tcccttataa atcaaaagaa tagaccgaga tagggttgag tgttgttcca gtttgaaca	1500
agagtccact attaaagaac gtggactcca acgtcaaagg gcgaaaaacc gtctatcagg	1560
gcgatggccc actacgtgaa ccatcacct aatcaagttt tttggggtcg aggtgccgta	1620
aagcactaaa tcggaaccct aaaggagacc cccgatttag agcttgacgg ggaagccgg	1680
cgaacgtggc gagaaaggaa gggaaagaaag cgaaaggagc gggcgctagg gcgctggcaa	1740
gtgtagcggc cacgctgcgc gtaaccacca caccgcgcgc gcttaatgcg ccgctacagg	1800
gcgctgcgc ccattcgcca ttcaggctgc gcaactgttg ggaagggcga tcggtgcggg	1860
cctcttcgct attacgccag ctggcgaaag ggggatgtgc tgcaaggcga ttaagtggg	1920
taacgccagg gttttccag tcacgacgtt gtaaacgac ggccagtgag cgcgcgtaat	1980
acgactcact ataggcgaa ttgggtaccg ggccccct cgaggtcgac ggtatcgata	2040
agcttgatat cgaattcctg cagccccggg gatccactag ttctagagcg gccatttaa	2100
acggccggcc ctagatcaga gggtggtaaa tgaagtgtaa tagtattcat tttcttata	2160
aatcatcct tccgtgattt atacaaaaga agaggagaat atgctgaata cttggtatat	2220
tactctacat tatactctta tcttgacggg tattctgagc atcttactca gtttcaagat	2280
cttttaatgt ccaaaaacat ttgagccgat ctaaatactt ctgtgttttc attaatttat	2340
aaattgtact cttttaagac atggaaagta ccaacatcg ttgaaacagt ttttcattta	2400
cttatggttt attggttttt ccagtgaatg attatgtgc gttacccttt cgtaaaagt	2460
caaacacgtt ttttaagtatt gtttagttgc tctttcgaca tatatgatta tccctgcgcg	2520
gctaaagtta aggatgcaaa aaacataaga caactgaagt taatttacgt caattaagtt	2580
ttccagggtat atgatgtttt gggcttccac taattcaata agtatgcat gaaatcagtt	2640
gtgaagagca tccagaaata atgaaaagaa acaacgaaac tgggtcggcc tgttgtttct	2700
tttctttacc acgtgatctg cggcatttac aggaagtcgc gcgttttgcg cagttgttgc	2760
aacgcagcta cggctaacaa agcctagtgg aactcgactg atgtgttagg gcctaaaact	2820
ggtggtgaca gctgaagtga actattcaat ccaatcatgt catggctgtc acaaagacct	2880
tgcggaccgc acgtacgaac acatacgtat gtaatatgt gttttgatag taccagtg	2940
tcgcagacct gcaatttttt tgtaggtttg gaagaatata taaaggttgc actcattcaa	3000
gatagttttt ttcttgtgtg tctattcatt ttattattgt ttgtttaaat gttaaaaaaa	3060
ccaagaactt agtttcaaat taaattcatc acacaaacaa acaaaacaaa atgaacattg	3120

ES 2 607 894 T3

ttgtttgttt gaagcaagtt ccagacactg ctgaagtcag aattgaccca gtcaagggta	3180
ctttaatcag agaaggtggt ccatctatca tcaacccaga cgacaagaac gctttggaag	3240
aagctttggt tttgaaggac aactacggtg ctcacgttac cgtcatttcc atgggtccac	3300
ctcaagccaa gaacgctttg gttgaagcct tggccatggg tgctgatgaa gctgtcttat	3360
tgactgacag agctttcggg ggtgctgata ctttagctac ctctcacacc attgctgctg	3420
gtatcaagaa attgaaatac gatatcgtct ttgccggctg tcaagccatc gatggtgata	3480
ccgctcaagt cggtcagaa attgctgaac atttgggtat tccacaagtc acctacgttg	3540
aaaaggttga agttgacggt gacactttga agatcagaaa ggcttgggaa gacggttacg	3600
aagttgttga agtcaagact ccagttctat tgactgccat caaggaattg aacgttccaa	3660
gatacatgtc cgttgaaaag atcttcggtg ctttcgacaa ggaagtcaag atgtggactg	3720
ctgatgatat cgatgtcgac aaggccaact tgggtttgaa aggttctcca accaaggtca	3780
agaaatcttc taccaaggaa gtcaagggtc aagggtgaagt cattgacaaa ccagtcaagg	3840
aagctgccgc ttacgttgtt tccaagttga aggaagaaca ctacatctaa agccccggcg	3900
gagattgata agacttttct agttgcatat cttttatatt taaatcttat ctattagtta	3960
atTTTTTgta atttatcctt atatatagtc tggttattct aaaatatcat ttcagtatct	4020
aaaaattccc ctcttttttc agttatatct taacaggcga cagtccaaat gttgatattat	4080
cccagtcgga ttcacacagg ttgtgaagca ttttgtcaat ggtcgaaatc acatcagtaa	4140
tagtgectct tacttgectc atagaatttc tttctcttaa cgtcacctgt tggctcttta	4200
tagtttcgaa atctatggtg ataccaaag gtgttcccaa ttcacogtta cgggcgtatt	4260
ttttaccaat tgaagtattg gaatcgtcaa ttttaaagta tatctctctt ttacgtaaag	4320
cctgcgagat cctcttaagt atagcgggga agccatcgtt attcgatatt gtcgtaacaa	4380
atactttgat cggcgctatg tttaaagtgt taaacatgga cagatatgcg atgaaaacgc	4440
taagtgatac tccaaatggt gaaaggtacg atgcttgaa acaatacttg gaaatcaccg	4500
gaaacaccat atgcggcgaa aagccaatta gtgtgatact aagtgcctta tcgaaaatcc	4560
gtgatgccgg tccttcaggc atcaaatttc agtggcctaa ttattcacag agttctcatg	4620
tgacaagtat tgatgatagt agtgtcagtt atgcttcagg ttatgttact ataggataat	4680
gatcacggct aaaacggtcg aatgtaagca tatatctttc gattgtataa ttgttcccaa	4740
atactacagc atctcaagga aaaaaaaca aaaacttcca aaaaaatcga atccctgagg	4800

ES 2 607 894 T3

aatctttaat acattttcaa tctatttaag ttttataaac gtgtatatga gatgtcatga 4860
gcatgaatta ttaataataa aaactaaatc attaaagtaa ctttaaggagt taaagcccgg 4920
gctttaattg ttagcagcct tgacttgagc aatcaattct ggaacaacct tgttgacatc 4980
accgacaatg gccaaatcag caaccttcat gattggagct tgcacatctt tgttgatggc 5040
aatgatgtag tcagagtctt gcataaccagc caagtgttgg atggcaccag agataccaca 5100
agcaatgtac aaagtggtc tgacggtctt accggtttga cegacttgca agtccttgtc 5160
aaccattcc ttttcaatgg cagctctgga agcagcaatg gtaccacca acaagaagc 5220
taattcttcc aatttttcga agttttcctt ggaaccaaca ccacgaccac cagcaacca 5280
aaccttggtc tcaccgatat cagcaatgtc cttggccaat ttgacaacct tggaaacctt 5340
ggttctgata tcagaagcag tcaatttgat ggcaaccttt tcgatcttgt catcagaaac 5400
gttagcatcg ttaactggca atttttcaa gacacctggc ctgacggtgg ccatttgagg 5460
tctgtggtca gaacagacaa tggtagcaat caagttacca ccgaaagctg gtctggtagc 5520
caacaagtca cgtttttcga catcgatata caaagaggta cagtcagcag tcaaaccagt 5580
agacaatctg gcagcaattc ttggaccaa gtctctaccg atgaaagtag caccgatgaa 5640
taagatttct ggctttcttt cgttgaccaa gtcacagata accttggcgt aaccgtcagt 5700
ggagaaatga gctaataatt cgttgtcagc agccaaaacc ttgtcagcac cgtgggacaa 5760
caagtccttg gacatctttt cagtgttgtg acccaataag acagcagtca attcaacacc 5820
caatttttca gccatttctt tacccttacc tagcaattcc aaagaacct tttgtaattc 5880
accatctctt tgttcagcga aaaccagac accctttagc tcagccttgt tcatgtttag 5940
ttaattatag ttcgttgacc gtatattcta aaaacaagta ctcttaaaa aaaaaccttg 6000
aaggaataa acaagtagaa tagatagaga gaaaaataga aaatgcaaga gaatttatat 6060
attagaaaga gagaagaaa aatggaaaa aaaaaatagg aaaagccaga aatagcacta 6120
gaaggagcga caccagaaaa gaaggtgatg gaaccaattt agctatata agttaactac 6180
cggctcgatc atctctgcct ccagcatagt cgaagaagaa tttttttttt cttgaggctt 6240
ctgtcagcaa ctcttatttt ttctttcttt tttggtgagc ctaaaaagtt cccacgttct 6300
cttgtacgac gccgtcacia acaaccttat gggttaattg tcgcggtctg ggtgtataaa 6360
tgtgtgggtg caggccggcc gtttaaacgg gccgccaccg cggaggagcc tgtgtggaag 6420
aacgattaca acagtggttgc tcctctgagg acataaaata cacaccgaga ttcatcaact 6480
cattgctgga gttagcatat ctacaattgg gtgaaatggg gagcgatttg caggcatttg 6540

ES 2 607 894 T3

```

ctcggcatgc cggtagaggt gtggtcaata agagcgacct catgctatac ctgagaaagc 6600
aacctgacct acaggaaaga gttactcaag aataagaatt ttcgttttaa aacctaagag 6660
tcactttaa atttgtatac acttattttt ttataactt atttaataat aaaaatcata 6720
aatcataaga aattcgctcg agtcgactgc agtttatcta atggagaaac catcagtcaa 6780
gacacatctt cttcttctag cgaagtgacg ggcagtggta gtaccttcac cagttggagt 6840
agcaatggtg aaagtggtag aaccttcacc tctgaaacct aaaccagcga aagatggacc 6900
gttcttgaca aagatggagg tttgcatgtc acgggcagcc ttgttcaatc tggagatggt 6960
ttgagagtgc atggtagcag tgtgatgtag accttgttcc aattcaatgg caacttccaa 7020
agcttcatcg aagtctgga ctctgacaac tggacaatt ggcatcaaca attcaacagt 7080
agcgaatggg tgggactttt cagtttcgac aatgatcaat cttgggggta aatcacaagc 7140
aataccagct tctttcaaga tttcagtggc agacttacca accaatttct tgttggtgac 7200
acccttgta gtgacggcaa ctttttccaa tttttggata tcagatgggt tagtgacgtg 7260
caaagcaccg tcttttcca tttggaacaa caagaagtca gcaatggagt caacggcaac 7320
aacagacttt tcagcgatac acaagatatt atggtcaaag gaagcacctg cgacaatgtc 7380
agcagcagcc ttttcaatgt tagcggtttc gtcaacgatg gatggagggt taccagcacc 7440
agcaccgata accttcttac cagattgcat agcttgcaag acaacacctg gaccaccagt 7500
gatgaccaac aatggaacct ttgggtgggt catcatttct tgagcagctt ggatagatgg 7560
cttggaacg gtgacaatca agttgtcaat accacaagaa tctctgacga tagtgttaa 7620
cttttcaatc aaccataaag agatgttctt ggcacctggg tgaggagagt agaaaacggc 7680
gttaccagca gccaacatac cgatggagtt acagatcaaa gtttcagttg ggttggtaga 7740
tgagcaaca gcaccgatga caccgatgg agataattcg tataaagtca taccgttgtc 7800
accggtagca acttcagtgt acaagtctc aacacctgga gtcttttcca tagctaaagt 7860
gttcttcaag attttatcgg tgacattacc cataccggtt tcagcaacag ctctggtagc 7920
aatggtttcg atttctgggt ataaagctc tctgatggcc ttgacaacgt ttcttcttcc 7980
ttccaaagat ttttccttgt aacagttttg agcaatgacg gcagcttggg cagcttcac 8040
gacggtatcg aaaacaccgg acttggcacc ttgggtgggt gtcttgggtg gaacttcctt 8100
ttgttcagcc aatttttcca acaaaacctt cttgactaat tgttccaatt ccaaagattc 8160
cattttttac tagttctaga atccgtcgaa actaagttct ggtgttttaa aactaaaaa 8220

```

ES 2 607 894 T3

aagactaact ataaaagtag aatttaagaa gttaagaaa tagatttaca gaattacaat 8280
caatacctac cgtctttata tacttattag tcaagtaggg gaataatttc agggaaactgg 8340
tttcaacctt ttttttcagc tttttccaaa tcagagagag cagaaggtaa tagaagggtg 8400
aagaaaatga gatagataca tgcgtgggtc aattgccttg tgcatcatt tactccaggc 8460
aggttgcatc actccattga ggttggtgcc gttttttgcc tgtttggtcc cctgttctct 8520
gtagttgctc taagagaatg gacctatgaa ctgatggtg gtgaagaaaa caatattttg 8580
gtgctgggat tctttttttt tctggatgcc agcttaaaaa gcgggctcca ttatatttag 8640
tggatgccag gaataaactg ttcaccacaga cacctacgat gttatatatt ctgtgtaacc 8700
cgccccctat tttggcatg tacgggttac agcagaatta aaaggctaata tttttgacta 8760
aataaagtta ggaaaatcac tactattaat tatttacgta ttctttgaaa tggcgagtat 8820
tgataatgat aaactgagct ccagcttttg ttccctttag tgagggttaa ttgctgctt 8880
ggcgtaatca tggctatagc tgtttcctgt gtgaaattgt tatccgctca caattccaca 8940
caacatagga gccggaagca taaagtgtaa agcctggggg gcctaataag tgaggtaact 9000
cacattaatt gcgttgctc cactgccgcg tttccagctg gaaaacctgt cgtgccagct 9060
gcattaatga atcgccaac gcgcggggag agcgggtttg cgtattgggc gctcttccgc 9120
ttcctcgtc actgactcgc tgcgctcggg cgttcggctg cgcgagcgg tatcagctca 9180
ctcaaaggcg gtaatacggg tatccacaga atcaggggat aacgcaggaa agaacatgtg 9240
agcaaaaggc cagcaaaagg ccaggaaccg taaaaggcc gcgttgctgg cgtttttcca 9300
taggctccgc cccctgacg agcatcacia aaatcgacgc tcaagtcaga ggtggcgaaa 9360
cccgacagga ctataaagat accagcggtt tccccctgga agctccctcg tgcgctctcc 9420
tgttccgacc ctgctgctta ccgatacct gtccgccttt ctcccttcgg gaagcgtggc 9480
gctttctcat agctcacgct gtaggtatct cagttcgggt taggtcgttc gctccaagct 9540
gggctgtgtg cacgaacccc ccgttcagcc cgaccgctgc gccttatccg gtaactatcg 9600
tcttgagtcc aaccgggtaa gacacgactt atcgccactg gcagcagcca ctggtaacag 9660
gattagcaga gcgaggtatg taggcggtgc tacagagttc ttgaagtggg ggcctaacta 9720
cggctacact agaaggacag tatttggtat ctgcgctctg ctgaagccag ttacctcgg 9780
aaaaagagtt gtagctcct gatccggcaa acaaaccacc gctggtagcg gtggtttttt 9840
tgtttgcaag cagcagatta cgcgcagaaa aaaaggatct caagaagatc ctttgatctt 9900
ttctacgggg tctgacgctc agtgaacga aaactcacgt taagggattt tggctatgag 9960

ES 2 607 894 T3

attatcaaaa aggatcttca cctagatcct tttaaattaa aatgaagtt ttaaatcaat 10020
 ctaaagtata tatgagtaaa cttggtctga cagttaccaa tgcttaatca gtgaggcacc 10080
 tatctcagcg atctgtctat ttcgttcatc catagttgcc tgactccccg tcgtgtagat 10140
 aactacgata cgggagggct taccatctgg cccagtgct gcaatgatac cgcgagacc 10200
 acgctcaccg gctccagatt tatcagcaat aaaccagcca gccggaaggg ccgagcgcag 10260
 aagtggtcct gcaactttat cgcctccat ccagtctatt aattggtgcc gggaagctag 10320
 agtaagtagt tggccagtta atagtttgcg caacgttggt gccattgcta caggcatcgt 10380
 ggtgtcacgc tcgtcgtttg gtatggcttc attcagctcc ggttcccaac gatcaaggcg 10440
 agttacatga tccccatgt tgtgcaaaaa agcggttagc tccttcggtc ctccgatcgt 10500
 tgtcagaagt aagtggccg cagtgtatc actcatggtt atggcagcac tgcataattc 10560
 tcttactgtc atgccatccg taagatgctt ttctgtgact ggtgagtact caaccaagtc 10620
 attctgagaa tagtgtatgc ggcgaccgag ttgctcttgc ccggcgtcaa tacgggataa 10680
 taccgcgcca catagcagaa ctttaaaagt gctcatcatt ggaaaacggt cttcggggcg 10740
 aaaactctca aggatcttac cgctgttgag atccagttcg atgtaacca ctctgtcacc 10800
 caactgatct tcagcatcct ttactttcac cagcgtttct gggtgagcaa aaacaggaag 10860
 gcaaatgcc gcaaaaaagg gaataagggc gacacgaaa tgttgaatac tcatactctt 10920
 cctttttcaa tattattgaa gcatttatca gggttattgt ctcatgagcg gatacatatt 10980
 tgaatgtatt tagaaaaata aacaaatagg ggttccgcgc acatttcccc gaaaagtgcc 11040
 acctgacgtc taagaaacca ttattatcat gacattaacc tataaaaaata ggcgtatcac 11100
 gaggccttt cgtc 11114

<210> 48

<211> 2613

5

<212> ADN

<213> Entamoeba histolytica

<400> 48

atgtcaacac aacaaactat gactgtagat gaacatatta atcaacttgt tgctaaagca 60
 caagttgcac ttaaagaata tcttaacca gaatatacac aagaaaaaat agattatatt 120
 gtaaagaaag catcagttgc agcacttgat caacatttg cacttgcagc agctgcagtt 180
 gaagaaacag gaagaggtat ttttgaagat aaagctacta aaaatatatt tgcatgtgaa 240
 catgttacac atgaaatgag acatgctaaa acagttggta ttattaatgt agatccactt 300

10

ES 2 607 894 T3

tatggaatta cagaaattgc agaaccagtt ggagttgttt gtggagttac accagttact	360
aatccaacat caacagctat tttcaagtca cttatttcaa ttaaaacaag aatccaatt	420
gtattttcat tccatccatc agcacttaaa tgttctatta tggcagctaa aattgttaga	480
gatgcagcta ttgcagcagg agcaccagaa aattgtattc aatggattga atttgagga	540
attgaagcat caaataaatt aatgaatcat ccaggagttg ctactattct tgctacagga	600
ggaaatgcta tggttaaagc agcataattca tcaggaaaac cagcacttgg agtaggagca	660
ggaaatgtac caacatatat tgaaaaaaca tgtaatatta aacaagcagc aatgatgta	720
gttatgtcaa aatcatttga taatggtagt atttgtgcat cagaacaagc agcaattatt	780
gataaagaaa tttatgatca agtagttgaa gaaatgaaaa cacttggagc atatttcatt	840
aatgaagaag aaaaagctaa attagaaaag tttatgtttg gagttaatgc atattcagca	900
gatgttaata atgcaagact taatccaaaa tgtccaggtg tgtcaccaca atggtttgct	960
gaacaagttg gaattaaagt tccagaagat tgtaatatta tttgtgcagt ttgtaaagaa	1020
gttggacca atgaaccatt aacaagagaa aaattatcac cagttcttgc tattcttaaa	1080
gcagaaaata cacaagatgg tattgataaa gctgaagcta tggttgaatt taatggtaga	1140
ggacattcag cagctattca ttcgaatgat aaagcagtag ttgaaaagta tgcacttaca	1200
atgaaagcat gcagaatttt acataataca ccatcatcac aaggaggaat tggatcaatt	1260
tataactata tttggccatc atttacactt ggatgtggat catatggagg aaattcggta	1320
tcagctaatg ttacatatca taatttatta aatattaaaa gacttgcaga tagaagaaac	1380
aaccttcaat ggttcagagt tccaccaaag attttctttg aaccacattc tattagatat	1440
cttgctgaac ttaaggaact tagtaaaata ttcattgttt cagatagaat gatgtataaa	1500
ttaggatatg tagatagagt tatggatgta ttgaaaagaa gaagtaatga agtagaaatt	1560
gaaattttca ttgatgtaga accagatcca tctattcaaa ccgttcaaaa aggacttgct	1620
gttatgaata catttggacc agataatatt attgctattg gaggaggatc agctatggat	1680
gcagctaaga ttatgtggtt actttatgaa catccagaag ccgatttctt tgcaatgaaa	1740
caaaaattca ttgatcttag aaagagagca tttaaattcc caacaatggg taagaaagct	1800
agattaatth gtattccaac aacatcagga actggatcag aagttacacc atttgcagtt	1860
atttcagatc atgaaacag taagaaatat ccacttgctg attattcact tacaccatca	1920
gttgctattg ttgatccaat gtttactatg tcaacttcaa agagagctat tgctgatact	1980

ES 2 607 894 T3

```

ggacttgatg tattggttca tgcaacagaa gcatatgttt cagttatggc taatgaatat 2040
actgatggac ttgctagaga agcagttaaa ttagtctttg aaaatcttct taaatcatat 2100
aatggagatt tagaagcaag agaaaagatg cacaatgctg caacaattgc aggtatggca 2160
tttgcacatcag cattccttgg tatggacat tccatggcac ataagttgg agcagcattc 2220
catcttccac atggtagatg tgtagcagta ttattaccac atgtcattag atataatgga 2280
caaaaaccaa gaaagcttgc aatgtggcca aatataatt tetataagge agaccaaga 2340
tatatggaac ttgcacaaat ggttggactt aaatgtaata caccagctga aggagttaa 2400
gcatttgcta aagcatgtga agaattaatg aaagccacag agactattac tggattcaag 2460
caagcaata ttgatgaagc agcatggatg agtaaagtac cagaaatggc acttcttgca 2520
tttgaagatc aatgttcacc agctaacca agagtccca tggtaagga tatggaaaag 2580
attctcaaag ctgcatatta tccaattgct tga 2613

```

<210> 49

<211> 2610

5

<212> ADN

<213> Secuencia artificial

<220>

<223> adh2 E. histolytica optimizada en el par de codones

10

<400> 49

```

atgtccactc aacaaacat gaccgttgat gaacacatta accaattggt cagaaaaggct 60
caagttgctt tgaaggaata cttgaaacca gaatacactc aagaaaagat cgattacatt 120
gtcaagaagg cttctgttgc tgctctagac caacactgtg ctttggctgc tgctgctgtc 180
gaagaaactg gtcgtggtat ctttgaagac aaagctacca agaacatttt cgcttgtgaa 240
cacgtcactc acgaaatgag acacgccaag accgttggtg tcatcaacgt tgatccatta 300
tacggtatca ctgaaattgc tgaaccagtc ggtgttgtct gtggtgtcac cccagttacc 360
aaccacaactt ctactgcat tttcaaatct ttgatttcca tcaagaccag aaaccgaatt 420
gttttctcct tccaccatc tgctttgaaa tgttccatca tggctgcaa gatcgtcaga 480
gatgctgcea ttgctgctgg tgctccagaa aactgtatcc aatggatcga atttggtggt 540
attgaagctt ccaacaaatt gatgaacat cctggtgttg ctaccatctt agctactggt 600
ggtaacgcta tggtaaggc tgcttactct tctggttaagc cagctttggg tgtcgggtgct 660
ggtaacgtcc caacttcat cgaaaagacc tgtaatatca agcaagctgc taacgatgtt 720
gtcatgtcca agtctttcga caacggtatg atctgtgcct ccgaacaagc tgccatcatc 780

```


ES 2 607 894 T3

```

gacaaagaaa tctacgacca agttgttgaa gaaatgaaga ctttgggtgc ttacttcac 840
aacgaagaag aaaaggccaa attggaaaaa ttcattgttcg gtgttaatgc ttactctgct 900
gatgtcaaca acgccagatt gaacccaaag tgtccaggta tgtctccaca atggttcgct 960
gaacaagtcg gtatcaaggt tccagaagac tgtaacatca tctgtgccgt ttgtaaggaa 1020
gttgggtccaa acgaaccatt gaccagagaa aagttgtctc cagttttggc ctttttgaag 1080
gctgaaaaca ctcaagatgg tattgacaag gctgaagcta tggtcgaatt caacggtcgt 1140
ggtcactctg ctgccattca ctccaatgac aaggctgttg ttgaaaaata cgctttgacc 1200
atgaaggctt gtcgtatctt gcacaacact ccatcttctc aagggtggtat cggttccatt 1260
tacaactaca tctggccatc tttcacttta gtttgtgggt cttacgggtg taactccggt 1320
tctgccaatg ttacctacca caacttgttg aacatcaaga gattggctga cagaagaaac 1380
aacttacaat gggtcagagt cccaccaaag atcttcttcg aacctcactc cattagatac 1440
ttggctgaat tgaaggaatt gtccaagatt ttcattgtct ctgacagaat gatgtacaaa 1500
ttgggttacg ttgacagagt tatggatgtc ttgaagagaa gatccaacga agttgaaatt 1560
gaaatcttca tcgatgttga accagacca tccattcaaa cctccaaaaa gggtttggct 1620
gtcatgaaca ctttcgggtc agacaacatc attgccattg gtgggtggtc tgccatggat 1680
gctgccaaaga tcattgtggtt attatacгаа catccagaag ctgatttctt cgctatgaag 1740
caaaaattca tcgatttaag aaagagagct ttcaagttcc caaccatggg taagaaggcc 1800
agattaatct gtatcccaac cacttctggt accggttctg aagtcacccc attcgtctgc 1860
atctctgacc acgaaactgg taagaagtat ccattggctg actactctt gaccccatcc 1920
gttgccattg ttgacccaat gtttaccatg tccttgctta agagagccat tgctgacact 1980
ggtttggatg tcttagtcca cgctactgaa gcttacgttt ctgttatggc taacgaatac 2040
actgacgggt tggccagaga agctgtcaaa ttggttttcg aaaacttgtt gaaatcttac 2100
aacgggtgact tggaaagctc tgaaaagatg cacaacgctg ctaccattgc tggatggcc 2160
tttgcttctg ctttcttggg tatggacat tccatggctc acaaggctcg tgctgctttc 2220
catttgccac acggtagatg tgttgccgtt ttgttgctc acgttatcag atacaacggt 2280
caaaagccaa gaaagttggc catgtggcca aagtacaact tctacaaggc tgatcaaaga 2340
tacatggaat tggctcaaat ggtcggtttg aagtgaaca cccagctga aggtgtcgaa 2400
gcctttgcca aggttgtga agaattgatg aaggctactg aaaccatcac tggtttcaag 2460
aaggccaaca ttgatgaagc tgcttgatg tccaaggttc cagaaatggc tctattggct 2520
ttcgaagacc aatgttctcc agctaacca agagtccaa tggtaagga catggaagag 2580
atthtgaagg ctgcttacta cccaatcgct 2610

```

<210> 50

5 <211> 870

<212> PRT

<213> Entamoeba histolytica

ES 2 607 894 T3

<400> 50

Met Ser Thr Gln Gln Thr Met Thr Val Asp Glu His Ile Asn Gln Leu
 1 5 10 15

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

Thr Gln Glu Lys Ile Asp Tyr Ile Val Lys Lys Ala Ser Val Ala Ala
 35 40 45

Leu Asp Gln His Cys Ala Leu Ala Ala Ala Val Glu Glu Thr Gly
 50 55 60

Arg Gly Ile Phe Glu Asp Lys Ala Thr Lys Asn Ile Phe Ala Cys Glu
 65 70 75 80

His Val Thr His Glu Met Arg His Ala Lys Thr Val Gly Ile Ile Asn
 85 90 95

Val Asp Pro Leu Tyr Gly Ile Thr Glu Ile Ala Glu Pro Val Gly Val
 100 105 110

Val Cys Gly Val Thr Pro Val Thr Asn Pro Thr Ser Thr Ala Ile Phe
 115 120 125

Lys Ser Leu Ile Ser Ile Lys Thr Arg Asn Pro Ile Val Phe Ser Phe
 130 135 140

His Pro Ser Ala Leu Lys Cys Ser Ile Met Ala Ala Lys Ile Val Arg
 145 150 155 160

Asp Ala Ala Ile Ala Ala Gly Ala Pro Glu Asn Cys Ile Gln Trp Ile
 165 170 175

ES 2 607 894 T3

Glu Phe Gly Gly Ile Glu Ala Ser Asn Lys Leu Met Asn His Pro Gly
 180 185 190

Val Ala Thr Ile Leu Ala Thr Gly Gly Asn Ala Met Val Lys Ala Ala
 195 200 205

Tyr Ser Ser Gly Lys Pro Ala Leu Gly Val Gly Ala Gly Asn Val Pro
 210 215 220

Thr Tyr Ile Glu Lys Thr Cys Asn Ile Lys Gln Ala Ala Asn Asp Val
 225 230 235 240

Val Met Ser Lys Ser Phe Asp Asn Gly Met Ile Cys Ala Ser Glu Gln
 245 250 255

Ala Ala Ile Ile Asp Lys Glu Ile Tyr Asp Gln Val Val Glu Glu Met
 260 265 270

Lys Thr Leu Gly Ala Tyr Phe Ile Asn Glu Glu Glu Lys Ala Lys Leu
 275 280 285

Glu Lys Phe Met Phe Gly Val Asn Ala Tyr Ser Ala Asp Val Asn Asn
 290 295 300

Ala Arg Leu Asn Pro Lys Cys Pro Gly Met Ser Pro Gln Trp Phe Ala
 305 310 315 320

Glu Gln Val Gly Ile Lys Val Pro Glu Asp Cys Asn Ile Ile Cys Ala
 325 330 335

Val Cys Lys Glu Val Gly Pro Asn Glu Pro Leu Thr Arg Glu Lys Leu
 340 345 350

Ser Pro Val Leu Ala Ile Leu Lys Ala Glu Asn Thr Gln Asp Gly Ile
 355 360 365

Asp Lys Ala Glu Ala Met Val Glu Phe Asn Gly Arg Gly His Ser Ala
 370 375 380

Ala Ile His Ser Asn Asp Lys Ala Val Val Glu Lys Tyr Ala Leu Thr
 385 390 395 400

ES 2 607 894 T3

Met Lys Ala Cys Arg Ile Leu His Asn Thr Pro Ser Ser Gln Gly Gly
 405 410 415

Ile Gly Ser Ile Tyr Asn Tyr Ile Trp Pro Ser Phe Thr Leu Gly Cys
 420 425 430

Gly Ser Tyr Gly Gly Asn Ser Val Ser Ala Asn Val Thr Tyr His Asn
 435 440 445

Leu Leu Asn Ile Lys Arg Leu Ala Asp Arg Arg Asn Asn Leu Gln Trp
 450 455 460

Phe Arg Val Pro Pro Lys Ile Phe Phe Glu Pro His Ser Ile Arg Tyr
 465 470 475 480

Leu Ala Glu Leu Lys Glu Leu Ser Lys Ile Phe Ile Val Ser Asp Arg
 485 490 495

Met Met Tyr Lys Leu Gly Tyr Val Asp Arg Val Met Asp Val Leu Lys
 500 505 510

Arg Arg Ser Asn Glu Val Glu Ile Glu Ile Phe Ile Asp Val Glu Pro
 515 520 525

Asp Pro Ser Ile Gln Thr Val Gln Lys Gly Leu Ala Val Met Asn Thr
 530 535 540

Phe Gly Pro Asp Asn Ile Ile Ala Ile Gly Gly Gly Ser Ala Met Asp
 545 550 555 560

Ala Ala Lys Ile Met Trp Leu Leu Tyr Glu His Pro Glu Ala Asp Phe
 565 570 575

Phe Ala Met Lys Gln Lys Phe Ile Asp Leu Arg Lys Arg Ala Phe Lys
 580 585 590

Phe Pro Thr Met Gly Lys Lys Ala Arg Leu Ile Cys Ile Pro Thr Thr
 595 600 605

Ser Gly Thr Gly Ser Glu Val Thr Pro Phe Ala Val Ile Ser Asp His
 610 615 620

Glu Thr Gly Lys Lys Tyr Pro Leu Ala Asp Tyr Ser Leu Thr Pro Ser

ES 2 607 894 T3

625		630		635		640
Val Ala Ile	Val Asp	Pro Met Phe Thr	Met Ser Leu Pro	Lys Arg Ala		
	645		650		655	
Ile Ala Asp	Thr Gly	Leu Asp Val	Leu Val His Ala	Thr Glu Ala Tyr		
	660		665		670	
Val Ser Val	Met Ala Asn	Glu Tyr Thr	Asp Gly Leu Ala	Arg Glu Ala		
	675		680		685	
Val Lys Leu	Val Phe Glu	Asn Leu Leu	Lys Ser Tyr Asn	Gly Asp Leu		
	690		695		700	
Glu Ala Arg	Glu Lys Met	His Asn Ala Ala	Thr Ile Ala Gly	Met Ala		
	705		710		715	720
Phe Ala Ser	Ala Phe Leu	Gly Met Asp	His Ser Met Ala	His Lys Val		
		725		730		735
Gly Ala Ala	Phe His Leu	Pro His Gly	Arg Cys Val Ala	Val Leu Leu		
		740		745		750
Pro His Val	Ile Arg Tyr	Asn Gly Gln	Lys Pro Arg Lys	Leu Ala Met		
	755		760		765	
Trp Pro Lys	Tyr Asn Phe	Tyr Lys Ala	Asp Gln Arg Tyr	Met Glu Leu		
	770		775		780	
Ala Gln Met	Val Gly Leu	Lys Cys Asn	Thr Pro Ala Glu	Gly Val Glu		
		790		795		800
Ala Phe Ala	Lys Ala Cys	Glu Glu Leu	Met Lys Ala Thr	Glu Thr Ile		
		805		810		815
Thr Gly Phe	Lys Lys Ala	Asn Ile Asp	Glu Ala Ala Trp	Met Ser Lys		
		820		825		830
Val Pro Glu	Met Ala Leu	Leu Ala Phe	Glu Asp Gln Cys	Ser Pro Ala		
	835		840		845	
Asn Pro Arg	Val Pro Met	Val Lys Asp	Met Glu Lys Ile	Leu Lys Ala		
	850		855		860	
			Ala Tyr Tyr	Pro Ile Ala		
			865		870	

<210> 51

5 <211> 2658

<212> ADN

<213> E2 de Piromyces sp.

ES 2 607 894 T3

<400> 51

atgtccggat taaaaatggt ccaaaacctt tctctttacg gtagtctcgc cgaaatcgat	60
actagcgaaa agcttaacga agctatggac aaattaactg ctgcccaaga acaattcaga	120
gaatacaacc aagaacaagt tgacaaaatc ttcaaggctg ttgctttagc tgctttctcaa	180
aaccgtgttg ctttcgctaa gtacgcacac gaagaaaccc aaaaggggtg tttcgaagat	240
aaggttatca agaacgaatt cgctgctgat tacatctacc acaagtactg caatgacaag	300
accgccggta tcattgaata tgatgaagcc aatggctcta tggaaattgc tgaaccagtt	360
ggtccagttg ttggtattgc tccagttact aacccaactt ctactatcat ctacaagtct	420
ttaattgcct taaagaccgg taactgtatt atcttctcac cacatccagg agctcacaag	480
gcctctgttt tcggtgtaa ggtcttacac caagctgctg ttaaggctgg tgccccagaa	540
aactgtattc aatcatctt cccaaagatg gatttaacta ctgaattatt acaccaccaa	600
aagactcgtt tcatttgggc tactgggtgt ccaggtttag ttcacgcctc ttacaattct	660
ggtaagccag ctcttgggtg tgggccagggt aatgctccag ctcttattga tgaaacttgt	720
gatatgaacg aagctgttgg ttctatcgtt gtttctaaga ctttcgattg tggatgatc	780
tgtgccactg aaaaacgctg tgctcgttgc gaatctgtct acgaaaactt cgttgctacc	840
atgaagaagc gtggtgccta ctcatgact ccagaagaaa ccaagaaggc ttctaactt	900
cttttcggag aaggtatgag attaaatgct aaggctgttg gtcaaactgc caagacttta	960
gctgaaatgg ccggtttcga agtcccagaa aacaccgttg ttctctgtgg tgaagcttct	1020
gaagttaa at tcgaagaacc aatggctcac gaaaagttaa ctactatcct cggatctac	1080
aaggctaagg actttgacga tgggtgcaga ttatgtaagg aattagttac tttcgggtgt	1140
aagggtcaca ctgctgttct ctacaccaac caaaacaaca aggaccgtat tgaaaagtac	1200
caaaacgaag ttccagcctt ccacatctta gttgacatgc catcttcctt cggttgtatt	1260
ggtgatatgt acaacttccg tcttgctcca gctcttacca ttacttgtgg tactatgggt	1320
ggtggttcct cctctgataa cattggtcca aagcacttac ttaacatcaa gcgtgttgg	1380

ES 2 607 894 T3

atgagacgcg aaaacatgct ttggttcaag attccaaagt ctgtctactt caagcgtgct 1440
atcctttctg aagctttatc tgacttacgt gacaccaca agcgtgctat cattattacc 1500
gatagaacta tgactatggt aggtcaaact gacaagatca ttaaggcttg tgaaggatcat 1560
ggtatggtct gcaactgtcta cgataagggt gtcccagatc caactatcaa gtgtattatg 1620
gaagggtgta atgaaatgaa cgtcttcaag ccagatttag ctattgctct tgggtggtggt 1680
tctgctatgg atgccgctaa gatgatgcgt ttattctacg aataaccaga ccaagactta 1740
caagatattg ctactcgttt cgtcgatata cgtaagcgtg ttgttggttg tccaaagctt 1800
ggtagactta ttaagactct tgtctgtatc ccaactacct ctggtactgg tgccgaagtt 1860
actccattcg ctgctgttac ctctgaagaa ggtcgtaagt acccattagt cgactacgaa 1920
cttactccag atatggctat tgttgatcca gaattcgctg ttggtatgcc aaagcgttta 1980
acttcttga ctggtattga tgctcttacc cacgccattg aatcttacct ttctattatg 2040
gctactgact tcaactagacc atactctctc cgtgctgttg gtcttatctt cgaatccctt 2100
tcccttgctt acaacaacgg taaggatatt gaagctcgtg aaaagatgca caatgcttct 2160
gctattgctg gtatggcctt tgccaacgct ttccttggtt gttgtcactc tgttgetcac 2220
caacttggtt cgtctacca cattccacac ggtcttgcca acgctttaat gctttctcac 2280
atcattaagt acaacgctac tgactctcca gtaagatgg gtaccttccc acaatacaag 2340
taccacaag ctatgctca ctacgctgaa attgctgaac tcttattacc accaactcaa 2400
gttgtaaga tgactgatgt tgataagggt caatacttaa ttgaccgtgt tgaacaatta 2460
aaggctgacg ttggtattcc aaagtctatt aaggaaactg gaatggttac tgaagaagac 2520
ttcttcaaca aggttgacca agttgctatc atggccttcg atgaccaatg tactggtgct 2580
aaccacgctt acccattagt ttctgaatta aaacaattaa tgattgatgc ctggaacggt 2640
gttgteccaa agctctaa 2658

<210> 52

<211> 885

5 <212> PRT

<213> E2 de *Piromyces* sp.

<400> 52

10 Met Ser Gly Leu Gln Met Phe Gln Asn Leu Ser Leu Tyr Gly Ser Leu
1 5 10 15

ES 2 607 894 T3

Ala Glu Ile Asp Thr Ser Glu Lys Leu Asn Glu Ala Met Asp Lys Leu
 20 25 30

Thr Ala Ala Gln Glu Gln Phe Arg Glu Tyr Asn Gln Glu Gln Val Asp
 35 40 45

Lys Ile Phe Lys Ala Val Ala Leu Ala Ala Ser Gln Asn Arg Val Ala
 50 55 60

Phe Ala Lys Tyr Ala His Glu Glu Thr Gln Lys Gly Val Phe Glu Asp
 65 70 75 80

Lys Val Ile Lys Asn Glu Phe Ala Ala Asp Tyr Ile Tyr His Lys Tyr
 85 90 95

Cys Asn Asp Lys Thr Ala Gly Ile Ile Glu Tyr Asp Glu Ala Asn Gly
 100 105 110

Leu Met Glu Ile Ala Glu Pro Val Gly Pro Val Val Gly Ile Ala Pro
 115 120 125

Val Thr Asn Pro Thr Ser Thr Ile Ile Tyr Lys Ser Leu Ile Ala Leu
 130 135 140

Lys Thr Arg Asn Cys Ile Ile Phe Ser Pro His Pro Gly Ala His Lys
 145 150 155 160

Ala Ser Val Phe Val Val Lys Val Leu His Gln Ala Ala Val Lys Ala
 165 170 175

Gly Ala Pro Glu Asn Cys Ile Gln Ile Ile Phe Pro Lys Met Asp Leu
 180 185 190

Thr Thr Glu Leu Leu His His Gln Lys Thr Arg Phe Ile Trp Ala Thr
 195 200 205

Gly Gly Pro Gly Leu Val His Ala Ser Tyr Thr Ser Gly Lys Pro Ala
 210 215 220

Leu Gly Gly Gly Pro Gly Asn Ala Pro Ala Leu Ile Asp Glu Thr Cys
 225 230 235 240

Asp Met Asn Glu Ala Val Gly Ser Ile Val Val Ser Lys Thr Phe Asp

ES 2 607 894 T3

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

ES 2 607 894 T3

Ile Leu Ser Glu Ala Leu Ser Asp Leu Arg Asp Thr His Lys Arg Ala
 485 490 495

Ile Ile Ile Thr Asp Arg Thr Met Thr Met Leu Gly Gln Thr Asp Lys
 500 505 510

Ile Ile Lys Ala Cys Glu Gly His Gly Met Val Cys Thr Val Tyr Asp
 515 520 525

Lys Val Val Pro Asp Pro Thr Ile Lys Cys Ile Met Glu Gly Val Asn
 530 535 540

Glu Met Asn Val Phe Lys Pro Asp Leu Ala Ile Ala Leu Gly Gly Gly
 545 550 555 560

Ser Ala Met Asp Ala Ala Lys Met Met Arg Leu Phe Tyr Glu Tyr Pro
 565 570 575

Asp Gln Asp Leu Gln Asp Ile Ala Thr Arg Phe Val Asp Ile Arg Lys
 580 585 590

Arg Val Val Gly Cys Pro Lys Leu Gly Arg Leu Ile Lys Thr Leu Val
 595 600 605

Cys Ile Pro Thr Thr Ser Gly Thr Gly Ala Glu Val Thr Pro Phe Ala
 610 615 620

Val Val Thr Ser Glu Glu Gly Arg Lys Tyr Pro Leu Val Asp Tyr Glu
 625 630 635 640

Leu Thr Pro Asp Met Ala Ile Val Asp Pro Glu Phe Ala Val Gly Met
 645 650 655

Pro Lys Arg Leu Thr Ser Trp Thr Gly Ile Asp Ala Leu Thr His Ala
 660 665 670

Ile Glu Ser Tyr Val Ser Ile Met Ala Thr Asp Phe Thr Arg Pro Tyr
 675 680 685

Ser Leu Arg Ala Val Gly Leu Ile Phe Glu Ser Leu Ser Leu Ala Tyr
 690 695 700

ES 2 607 894 T3

Asn 705	Asn	Gly	Lys	Asp	Ile 710	Glu	Ala	Arg	Glu	Lys 715	Met	His	Asn	Ala	Ser 720
Ala	Ile	Ala	Gly	Met 725	Ala	Phe	Ala	Asn	Ala	Phe 730	Leu	Gly	Cys	Cys	His 735
Ser	Val	Ala	His 740	Gln	Leu	Gly	Ser	Val 745	Tyr	His	Ile	Pro	His	Gly	Leu 750
Ala	Asn	Ala	Leu	Met	Leu	Ser	His 760	Ile	Ile	Lys	Tyr	Asn	Ala	Thr	Asp 765
Ser	Pro	Val	Lys	Met	Gly	Thr 775	Phe	Pro	Gln	Tyr	Lys	Tyr	Pro	Gln	Ala 780
Met	Arg	His	Tyr	Ala	Glu 790	Ile	Ala	Glu	Leu	Leu	Leu	Pro	Pro	Thr	Gln 800
Val	Val	Lys	Met	Thr	Asp 805	Val	Asp	Lys	Val	Gln	Tyr	Leu	Ile	Asp	Arg 815
Val	Glu	Gln	Leu	Lys	Ala	Asp	Val	Gly 825	Ile	Pro	Lys	Ser	Ile	Lys	Glu 830
Thr	Gly	Met 835	Val	Thr	Glu	Glu	Asp 840	Phe	Phe	Asn	Lys	Val	Asp	Gln	Val 845
Ala	Ile	Met	Ala	Phe	Asp	Asp	Gln	Cys	Thr	Gly	Ala	Asn	Pro	Arg	Tyr 850
Pro	Leu	Val	Ser	Glu	Leu	Lys	Gln	Leu	Met	Ile	Asp	Ala	Trp	Asn	Gly 865
Val	Val	Pro	Lys	Leu 885											

REIVINDICACIONES

1. Un método de identificación de un polipéptido heterólogo que tiene actividad enzimática para convertir piruvato o acetaldehído en acetil-CoA en una única etapa de conversión en (el citosol de) una célula de levadura que comprende:
- 5 - proporcionar una célula de levadura mutada, en la que dicha mutación comprende una inactivación de al menos un gen de la desviación de piruvato deshidrogenasa (PDH), seleccionado de los genes que codifican las enzimas piruvato descarboxilasa (PDC; E.C. 4.1.1.1), acetaldehído deshidrogenasa (ALD; E.C. 1.2.1.3. E.C. 1.2.1.4 o E.C. 1.2.1.5) y acetil-CoA sintetasa (ACS; E.C. 6.2.1.1);
- 10 - transformar dicha célula de levadura mutada con un vector de expresión que comprende al menos una secuencia de nucleótidos heteróloga operativamente unida a un promotor funcional en levadura y teniendo dicha secuencia de nucleótidos heteróloga que codifica un polipéptido candidato posible actividad enzimática para convertir piruvato o acetaldehído en acetil-CoA;
- probar dicha célula de levadura mutada recombinante para su capacidad para crecer sobre medio mínimo que contiene glucosa como única fuente de carbono, e
- 15 - identificar dicho polipéptido candidato como un polipéptido heterólogo que tiene actividad enzimática para convertir piruvato o acetaldehído en acetil-CoA en una única etapa de conversión en (el citosol de) dicha célula de levadura cuando se observa el crecimiento de dicha célula.
2. Método según la reivindicación 1, en el que dicha célula de levadura es una célula de *Saccharomyces cerevisiae* y en el que dicha secuencia de nucleótidos heteróloga está optimizada en el par de codones para la expresión en *Saccharomyces cerevisiae*.
- 20 3. Método según la reivindicación 2, en el que dicha mutación comprende una inactivación del gen para acetil-CoA sintetasa isoforma 2 (*acs2*).
4. Método según una cualquiera de las reivindicaciones 1 a 3, en el que dicho polipéptido candidato que tiene actividad enzimática para convertir acetaldehído en acetil-CoA en una única etapa de conversión es una (supuesta) acetaldehído deshidrogenasa acetilante (*acdh*; EC 1.2.1.10).
- 25 5. Una célula de levadura recombinante que comprende una secuencia de nucleótidos heteróloga que codifica un polipéptido que tiene actividad enzimática para la conversión directa de acetaldehído en acetil-CoA en (el citosol de) dicha célula de levadura, en la que el polipéptido es una acetaldehído deshidrogenasa (EC 1.2.1.10) acetilante, en la que dicho polipéptido tiene más del 70 %, 80 %, 90 % o 95 % de identidad de secuencia con la secuencia de aminoácidos seleccionada de SEQ ID NO: 19, 22, 25, 28 y 52.
- 30 6. Una célula de levadura recombinante según la reivindicación 5 que comprende un vector para la expresión de los polipéptidos heterólogos en levadura, comprendiendo dicho vector la secuencia de nucleótidos heteróloga operativamente unida a un promotor funcional en levadura.
- 35 7. Una levadura recombinante según una cualquiera de las reivindicaciones 5 o 6, en la que la levadura es *Saccharomyces cerevisiae*, y en la que dicha secuencia de nucleótidos heteróloga está optimizada en el par de codones para la expresión en *Saccharomyces cerevisiae*.
8. Una levadura recombinante según la reivindicación 7, en la que dicha secuencia de nucleótidos heteróloga está seleccionada de SEQ ID NO: 20, 23, 26, 29 y 51.
- 40 9. Célula de levadura según una cualquiera de las reivindicaciones 5 a 8, que comprende además una inactivación de al menos un gen de la desviación de PDH, seleccionado de los genes que codifican las enzimas piruvato descarboxilasa (PDC; E.C. 4.1.1.1), acetaldehído deshidrogenasa (ALD; E.C. 1.2.1.3. E.C. 1.2.1.4 o E.C. 1.2.1.5) y acetil-CoA sintetasa (ACS; E.C. 6.2.1.1).
10. Célula de levadura según una cualquiera de las reivindicaciones 5 a 9, en la que la célula de levadura comprende una inactivación de un gen que codifica una acetil-CoA sintasa.
- 45 11. Célula de levadura según una cualquiera de las reivindicaciones 5 a 10, en la que dicha célula muestra crecimiento sobre medio mínimo que contiene glucosa como única fuente de carbono.
12. Célula de levadura según una cualquiera de las reivindicaciones 5 a 11, que comprende además una inactivación de un gen que codifica una enzima que cataliza la conversión de acetaldehído en etanol, preferentemente una alcohol deshidrogenasa.
- 50 13. Célula de levadura según una cualquiera de las reivindicaciones 5 a 12, que comprende además uno o más genes introducidos que codifican una vía recombinante para la formación de 1-butanol a partir de acetil-CoA.

14. Célula de levadura según la reivindicación 13, en la que dicho uno o más genes introducidos codifican enzimas que producen acetoacetyl-CoA, 3-hidroxi-butiril-CoA, crotonil-CoA, butiril-CoA, butilaldehído y/o 1-butanol.
15. Célula de levadura según una cualquiera de las reivindicaciones 5 a 14, en la que dicha levadura es *Saccharomyces cerevisiae*
- 5 16. Un método de producción de un producto de fermentación, que comprende las etapas de fermentar un sustrato de carbono adecuado con una célula de levadura según una cualquiera de las reivindicaciones 5 a 15 y recuperar el producto de fermentación producido durante dicha fermentación.
17. Método según la reivindicación 16, en el que el producto de fermentación es butanol.

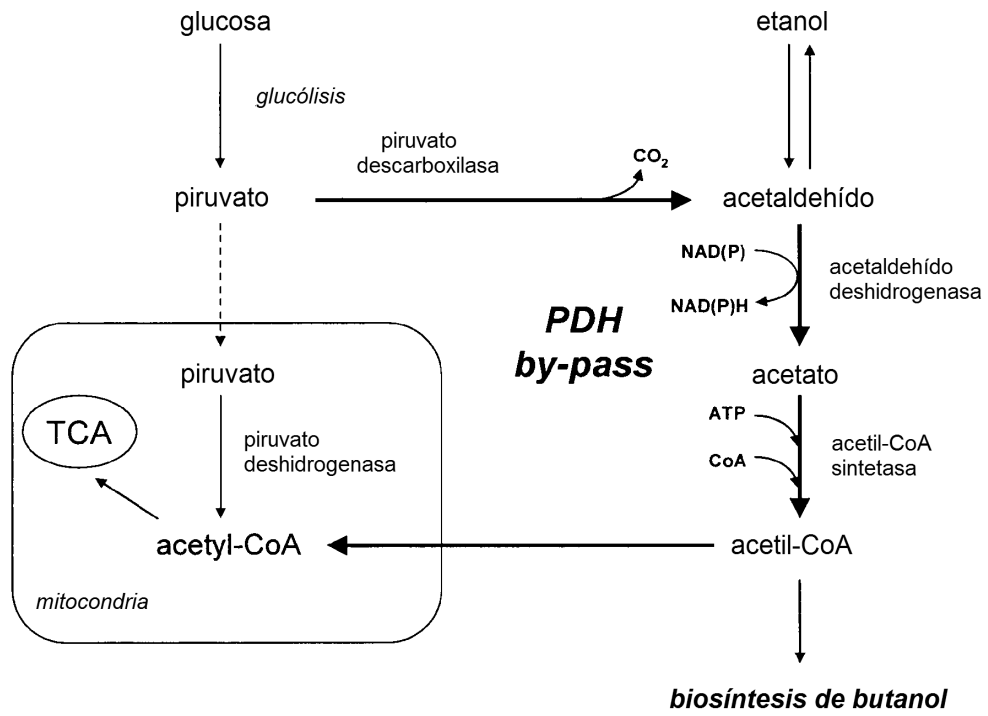


Figura 1

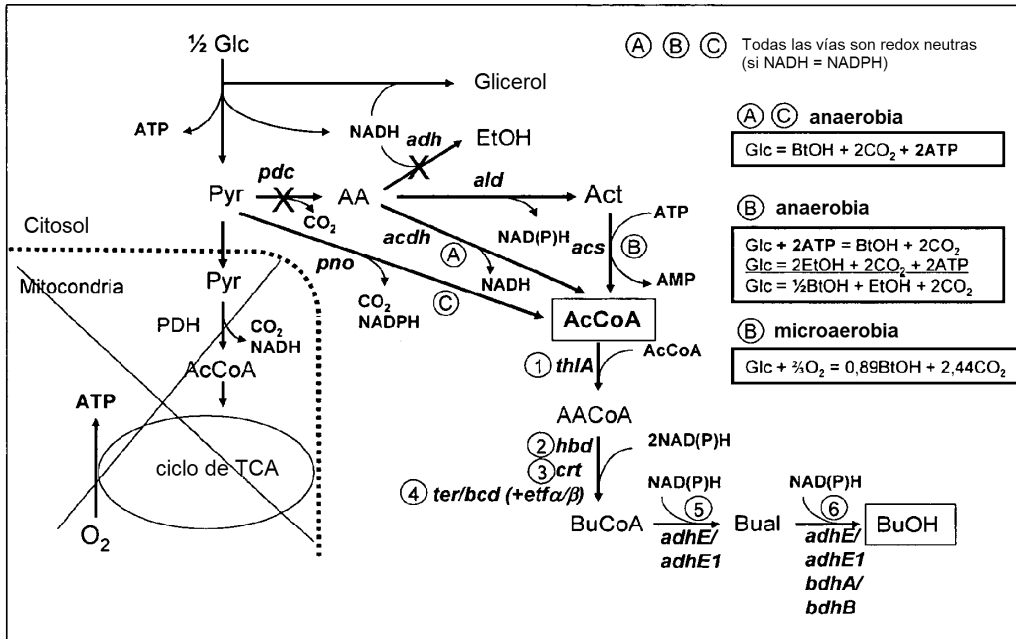


Figura 2

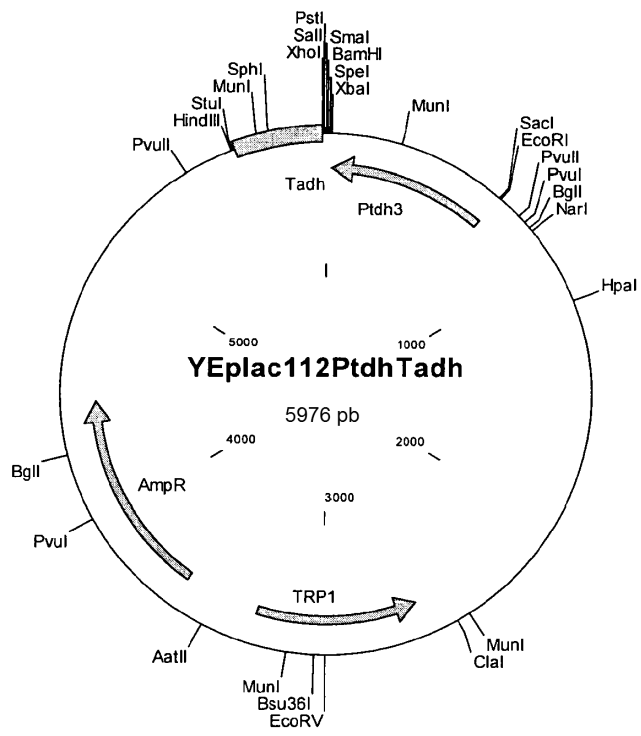


Figura 3

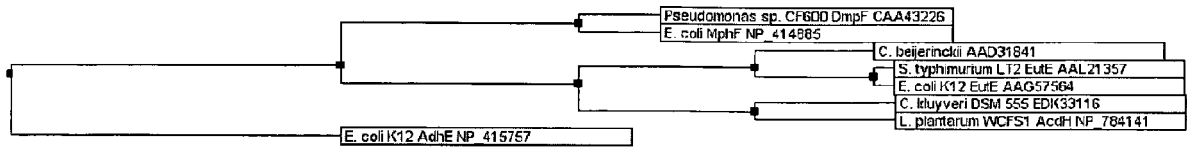


Figura 4