



(19)



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA

(11) Número de publicación: **2 276 427**

(51) Int. Cl.:

C12N 15/52 (2006.01)

C12N 9/00 (2006.01)

C07K 14/36 (2006.01)

C12Q 1/68 (2006.01)

C12N 15/63 (2006.01)

C12N 1/21 (2006.01)

C12P 19/00 (2006.01)

(12)

TRADUCCIÓN DE PATENTE EUROPEA

T3

(86) Número de solicitud europea: **97927087 .3**

(86) Fecha de presentación : **30.05.1997**

(87) Número de publicación de la solicitud: **0915981**

(87) Fecha de publicación de la solicitud: **19.05.1999**

(54) Título: **Aislamiento de los genes de biosíntesis de pseudo-oligosacáridos de *Streptomyces glaucescens* GLA.O y su uso.**

(30) Prioridad: **07.06.1996 DE 196 22 783**

(45) Fecha de publicación de la mención BOPI:
16.06.2007

(45) Fecha de la publicación del folleto de la patente:
16.06.2007

(73) Titular/es: **Sanofi-Aventis Deutschland GmbH**
Brüningstrasse 50
65929 Frankfurt am Main, DE

(72) Inventor/es: **Decker, Heinrich**

(74) Agente: **Elzaburu Márquez, Alberto**

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

Aislamiento de los genes de biosíntesis de pseudo-oligosacáridos de *Streptomyces glaucescens* GLA.O y su uso.

La presente invención se refiere al aislamiento de genes que codifican enzimas para la biosíntesis de inhibidores de la α -amilasa, los denominados pseudo-oligosacáridos. En particular, se trata de genes procedentes de la cepa de estreptomicetos, *Streptomyces glaucescens* GLA.O (DSM 40716). Además, describe el uso de estos genes para producir acarbosa y sustancias homólogas con ayuda de *Streptomyces glaucescens* (DSM 40716), la expresión heteróloga de estos genes en otras cepas que producen pseudo-oligosacáridos (p. ej., *Actinoplanes* sp. SE50/100) con el fin de incrementar y estabilizar la producción y también su expresión heteróloga en otros microorganismos, tales como *E. coli*, *Bacillus subtilis*, *Actinomycetales*, como p. ej., las cepas de *Streptomyces*, *Actinoplanes*, *Ampullariella*, *Streptoporangium*, *Streptomyces hygroscopicus* var. *limoneus*, *Streptomyces glaucescens*, así como los hongos relevantes biotecnológicamente (p. ej., *Aspergillus niger*, *Penicillium chrysogenum*) y levaduras (p. ej., *Saccharomyces cerevisiae*). La invención también se refiere a genes homólogos en otros microorganismos y a métodos para aislarlos.

Streptomyces glaucescens (DSM 40716) que en la bibliografía también se denomina *Streptomyces glaucescens* GLA.O, produce ambos antibióticos, hidroxiestreptomicina (Hütter (1967) Systematik der Streptomyceten. Basel Karger Verlag) y tetradenomicina (Weber y col. (1979) Arch. Microbiol. 121:111-116). Es conocido que los estreptomicetos son capaces de sintetizar productos naturales estructuralmente distintos. Sin embargo, las condiciones bajo las cuales se producen estos compuestos, son frecuentemente desconocidas o las sustancias sólo se producen en cantidades muy pequeñas y no se detectan.

El inhibidor de la α -amilasa, la acarbosa, ha sido aislado a partir de una variedad de cepas de *Actinoplanes* (SE50, SE82 y SE18) (Schmidt y col. (1977) Naturwissenschaften 64:535-536). La sustancia activa se descubrió en asociación con un escrutinio de los inhibidores de la α -amilasa, procedentes de organismos de los géneros *Actinoplanes*, *Ampullariella* y *Streptoporangium*. En el caso de la acarbosa, se trata de un pseudo-tetrasacárido que está compuesto por una unidad inusual de ciclitol insaturado a la que se une un amino azúcar 4,6 dideoxi-4-amino-D-glucopiranos. Unidades de D-glucopiranos adicionales, unidas de forma α -1,4-glicosídica se pueden unir al amino azúcar. De este modo, la acarbosa contiene, p. ej., dos moléculas adicionales de D-glucosa. La cepa productora sintetiza una mezcla de productos de pseudo-oligosacáridos que poseen cadenas laterales de azúcar de longitudes diferentes (Schmidt y col. (1977) Naturwissenschaften 64:535-536). El radical ciclitol de la acarbosa es idéntico al compuesto valienamina, que es un componente del antibiótico validamicina A (Iwasa y col. (1979) J. Antibiot. 32: 595-602) procedente de *Streptomyces hygroscopicus* var. *limoneus*.

La acarbosa se puede producir por fermentación empleando una cepa de *Actinoplanes* y ha conseguido una gran importancia económica como agente terapéutico para los diabéticos. Aunque *Actinoplanes* sintetiza una mezcla de los productos inhibidores de la α -amilasa, sólo el compuesto que tiene el peso molecular relativo de 645,5 (acarviosina con dos unidades de glucosa (Truscheit (1984) VIIIth International Symposium on Medicinal Chemistry, Proc. Vol. I. Swedisch Academy of Pharmaceutical Sciences, Estocolmo, Suecia), se emplea con el nombre genérico de acarbosa. Las condiciones de fermentación se seleccionan para asegurar que la acarbosa es el producto principal de la fermentación. Alternativamente, se emplean agentes de selección y cepas determinadas en las que se forma acarbosa como producto principal o se emplean en procesos de purificación para el aislamiento selectivo (Truscheit (1984) VIIIth International Symposium on Medicinal Chemistry, Proc. Vol. I. Swedisch Academy of Pharmaceutical Sciences, Estocolmo, Suecia). También es posible transformar químicamente la mezcla del producto para obtener finalmente el producto deseado de acarbosa.

En el documento de solicitud de patente alemán, con número de publicación 2 209 834, se describe un procedimiento para la preparación de inhibidores de la sacarosa a partir de la cepa *Actinoplanes* SE 50. Stockmann y Piepersberg describían sondas génicas para la detección del metabolismo de la 6-desoxihexosa en estreptomicetos que producen metabolitos secundarios (FEMS Microbiology Letters 90, (1992), 185-190).

Al contrario que el género *Streptomyces*, el género *Actinoplanes* no se ha investigado hasta la fecha genéticamente a fondo. Los métodos que se establecieron para el género *Streptomyces*, no son transferibles o no son siempre transferibles al género *Actinoplanes*. Para mejorar la producción de acarbosa de forma deliberada, con métodos de biología molecular, se tienen que aislar y caracterizar los genes de la biosíntesis de acarbosa. En este contexto, se sugiere el intentar establecer un sistema de hospedador-vector para *Actinoplanes* sp. Sin embargo, esto es muy tedioso y complicado debido a las escasas investigaciones sobre *Actinoplanes*.

La invención descrita en la presente solicitud de patente consigue el objetivo de clonar los genes de la biosíntesis de acarbosa y de pseudo-oligosacáridos homólogos, clonándose estos genes a partir de *Streptomyces glaucescens* GLA.O, que es un estreptomiceto que está bien investigado genéticamente (Crameri y col. (1983) J. Gen. Microbiol. 129:519-527; Hintermann y col. (1984) Mol. Gen. Genet. 196:513-520; Motamedi y Hutchinson (1987) PNAS USA 84:4445-4449; Geistlich y col. (1989) Mol. Microbiol. 3:1061-1069) y que, sorprendentemente, es un productor de acarbosa. En un medio que contiene almidón, *Streptomyces glaucescens* GLA.O produce pseudo-oligosacáridos que tienen los pesos moleculares 645, 807 y 970.

Un objeto de la invención también es el aislamiento de los genes correspondientes para la biosíntesis, procedentes de *Streptomyces glaucescens* GLA.O y su empleo para aislar las regiones colindantes de ADN para completar la agrupación génica de dichos genes de la biosíntesis.

El aislamiento de los genes para la biosíntesis de los pseudo-oligosacáridos y la caracterización de estos genes, tienen gran importancia para conseguir una mejor comprensión de la biosíntesis y de su regulación. Este conocimiento se puede utilizar para incrementar la productividad de la cepa de *Streptomyces glaucescens* GLA.O, en relación con la producción de acarbosa, mediante métodos establecidos clásicos y de biología molecular. Adicionalmente, la agrupación génica completa que codifica la síntesis de los pseudo-oligosacáridos o genes individuales, también se puede expresar en otros microorganismos de importancia biotecnológica, para conseguir un mayor incremento en la producción o una simplificación de la preparación de los pseudo-oligosacáridos, como p. ej., la acarbosa. La modificación específica de los genes de la biosíntesis también se puede utilizar para preparar una cepa que produzca exclusivamente acarbosa con un peso molecular de 645. Puesto que los genes para biosintetizar antibióticos están siempre presentes en agrupaciones y frecuentemente están muy conservados (Stockmann y Piepersberg (1992) FEMS Microbiol. Letters 90:185-190; Malpartida y col. (1987) Nature 314:642-644), los genes de *Streptomyces glaucescens* GLA.O también se pueden emplear como sonda para aislar los genes que codifican la acarbosa procedentes de, p. ej., *Actinoplanes* sp. La expresión de los genes reguladores o de los genes que codifican etapas limitantes de la biosíntesis, puede dar como resultado un aumento de la productividad en *Streptomyces glaucescens* GLA.O, en *Actinoplanes* sp. o en las cepas productoras correspondientes. Asimismo, un incremento de la productividad también se puede conseguir desconectando ("knock out" o mutagénesis) los genes de la biosíntesis de acarbosa que tienen un efecto inhibitor en la biosíntesis.

Una estrategia posible para clonar los genes de la biosíntesis de antibióticos que no se han aislado previamente, es la de emplear sondas específicas del gen (Stockmann y Piepersberg (1992) FEMS Microbiol. Letters 90:185-190; Malpartida y col. (1987) Nature 314:642-644). Estas sondas pueden ser fragmentos de ADN que están marcados con P³² o que están marcadas de otro modo o los genes adecuados se pueden amplificar directamente a partir de las cepas que se van a investigar, empleando cebadores degenerados para PCR y ADN cromosómico aislado como el molde.

El procedimiento mencionado en último lugar, se ha empleado en el presente estudio. Los pseudo-oligosacáridos como la acarbosa contienen una unidad de 4,6-desoxiglucosa como elemento estructural. Se conoce que la enzima 4,6-deshidratasa de dTDP-glucosa está implicada en la biosíntesis de la 4,6-desoxiglucosa (Stockmann y Piepersberg (1992) FEMS Microbiol. Letters 90:185-190). Puesto que los desoxiazúcares son un componente frecuente de los productos naturales y antibióticos, esta enzima puede ser posiblemente un medio para aislar los genes correspondientes para la biosíntesis de antibióticos. Puesto que estos genes están siempre presentes en forma de agrupaciones, inicialmente es suficiente aislar un gen; el aislamiento y la caracterización de las regiones adyacentes del ADN se puede realizar posteriormente.

Por ejemplo, una 4,6-deshidratasa de dTDP-glucosa cataliza una etapa de la biosíntesis de la hidroxiestreptomina en *Streptomyces glaucescens* GLA.O (Retzlaff y col. (1993) Industrial Microorganisms. Basic and applied molecular genetics ASM, Washington DC, EE.UU). Otras 4,6-deshidratasas de dTDP-glucosa se han aislado a partir de otros microorganismos, por ejemplo, a partir de *Streptomyces griseus* (Pissowotzki y col. (1991) Mol. Gen. Genet. 231:113-123), *Streptomyces fradiae* (Merson-Davies y Cundcliffe (1994) Mol. Microbiol. 13:349-355) y *Streptomyces violaceoruber* (Bechthold, y col. (1995) Mol. Gen. Genet. 248:610-620).

Por ello fue posible deducir las secuencias de los cebadores de la PCR, para amplificar una 4,6-deshidratasa de dTDP-glucosa a partir de las secuencias de aminoácidos de los genes de biosíntesis que ya se conocían. Para ello, se seleccionaron las regiones conservadas en las secuencias proteicas de estas enzimas y se tradujo la secuencia de aminoácidos en una secuencia de ácidos nucleicos, de acuerdo con el código genético. Las secuencias proteicas se obtuvieron de las bases de datos del EMBL y de Genbank. Se emplearon las siguientes secuencias: *Streptomyces griseus*; número de orden: X62567: gen strE (de fecha 30-10-1993); *Streptomyces violaceoruber*; número de orden: L37334: gen graE (de fecha 10-04-1995); *Saccharopolyspora etythraea*; número de orden: L37354: gen gdh (de fecha 09-11-1994). Se obtiene un gran número de posibles secuencias de cebadores como resultado de la degeneración del código genético. Ya que los estreptomicetos contienen generalmente una G o una C en la tercera posición de un codón (Wright y Bibb (1992) Gene 113:55-65), se reduce el número de cebadores que se tienen que sintetizar. Estas mezclas de cebadores se pueden emplear a continuación para llevar a cabo una amplificación con la PCR, del ADN procedente de las cepas que se van a investigar, lo que conduce de forma ideal a un fragmento de ADN amplificado. En el caso de proteínas altamente conservadas, este fragmento muestra una longitud pronosticable que es el resultado de la distancia entre los cebadores en la secuencia de ácido nucleico del gen correspondiente. Sin embargo, una mezcla experimental de esta naturaleza no da obligatoriamente como resultado un fragmento amplificado. Los cebadores pueden ser demasiado inespecíficos y amplificar un gran número de fragmentos o bien no obtenerse ningún producto de la PCR si no hay ningún sitio de unión complementaria para los cebadores de la PCR en el cromosoma.

La investigación de las cepas de estreptomicetos *Streptomyces glaucescens* GLA.O aportó un fragmento de ADN amplificado (acbD*) con la longitud esperada de 550 pb. Una investigación posterior mostraba que, además de contener un gen de la 4,6-deshidratasa de dTDP-glucosa para biosintetizar hidroxiestreptomina, esta cepa contiene sorprendentemente un segundo gen de la 4,6-deshidratasa de dTDP-glucosa para la biosíntesis de pseudo-oligosacáridos, tales como la acarbosa. Aunque ambos genes muestran un alto grado de homología, sólo son idénticos en un 65% a nivel de aminoácidos.

La sonda acbD* (véase el Ejemplo 2 y la Tabla 2A) se empleó para aislar, un fragmento de ADN con PstI de 6,8 kb de longitud procedente de *Streptomyces glaucescens* GLA.O, que codifica una variedad de genes (acbA, acbB, acbC, acbD, acbE, acbF), que participan en la biosíntesis de los pseudo-oligosacáridos.

5 Mediante la delección de los genes acbBCD (aminotransferasa acbB, dTDP-glucosa-sintetasa acbC, 4,6-deshidratación de dTDP-glucosa acbD, véase el Ejemplo 6) se produjo un mutante de *Streptomyces glaucescens* GLA.O, que ya no producía ningún pseudo-oligosacárido más en el medio de producción. La participación de los genes acbBCD en la síntesis de pseudo-oligosacáridos se verificó por tanto deleccionando los loci correspondientes.

10 Los dos genes, es decir, la dTDP-glucosa-sintetasa y la 4,6-deshidratación de dTDP-glucosa podrían participar en la biosíntesis del desoxiazúcar de los pseudo-oligosacáridos, lo que se puede concluir por la función de las enzimas homólogas estudiadas a fondo (véase más arriba). La aminotransferasa (codificada por el gen acbB) es responsable probablemente de transferir el grupo amino del radical azúcar o del radical ciclitol. Analizando la secuencia de proteínas de acbB, se encontró un motivo de aminoácido que está implicado en la unión del piridoxal-fosfato. Este motivo es típico
15 de las aminotransferasas de la clase III (EC 2.6.1.11; EC 2.6.1.13; EC 2.6.1.18; EC 2.6.1.19; EC 2.6.1.62; EC 2.6.1.64; EC 5.4.3.8). La función enzimática exacta de acbB sólo se puede aclarar con la investigación adicional de la biosíntesis de los pseudo-oligosacáridos. acbE codifica una proteína reguladora de la transcripción que muestra gran similitud con las proteínas que se unen al ADN que poseen un motivo de hélice-giro-hélice (p. ej., *Bacillus subtilis* DegA, P37947: Base de datos SwissProt). De este modo, el activador de la transcripción CcpA procedente de *Bacillus subtilis* inhibe
20 la formación de α -amilasa, en presencia, p. ej., de glucosa (Henkin y col. (1991) Mol. Microbiol. 5:575-584). Otros representantes de este grupo son proteínas que reconocen unidades de azúcar y son capaces de mostrar un efecto positivo o negativo sobre la biosíntesis de las rutas metabólicas. La biosíntesis de los pseudo-oligosacáridos también está regulada en *Streptomyces glaucescens* GLA.O. La síntesis de los pseudo-oligosacáridos sólo se podía demostrar hasta la fecha, en medios que contenían almidón. Por tanto, acbE podría ser responsable de la regulación de la síntesis de
25 pseudo-oligosacáridos, pero el mecanismo preciso no se conoce todavía. Pero el gen se puede modificar específicamente para obtener una tasa incrementada de biosíntesis de pseudo-oligosacáridos, por métodos de la biología molecular. Además, el sitio del ADN al que se une acbE se puede identificar mediante los ensayos denominados de retardo en gel (del inglés, "gel shift") (Miwa y col. (1994) Microbiology 140:2567-2575). Después de la identificación y la posterior modificación de los promotores y de otras regiones reguladoras del ADN que son responsables de la transcripción
30 de los genes de los pseudo-oligosacáridos, se puede conseguir un incremento de la tasa de biosíntesis de la acarbosa.

Hasta la fecha, la función de acbF no está bien definida. El producto génico correspondiente muestra homologías con proteínas de unión a azúcares, tales como la proteína que se une al azúcar de *Streptococcus mutans* (MsmE; Q00749: base de datos de SwissProt), siendo probable que esté implicado en la biosíntesis de los pseudo-oligosacáridos.
35 El producto génico del gen acbA muestra homologías con proteínas bacterianas conocidas para la unión a ATP (p. ej., de *Streptomyces peucetius* DrrA, P32010: base de datos de SwissProt). La proteína AcbA posee el típico motivo de unión a ATP/GTP, el denominado bucle P. Estas proteínas constituyen un componente importante de los denominados transportadores ABC que están implicados en el transporte activo de metabolitos en membranas biológicas (Higgins (1995) Cell 82:693-696). Por tanto, AcbA podría ser responsable de la exportación de pseudo-oligosacáridos fuera de
40 la célula o estar implicada en la importación de unidades de azúcar para biosintetizar inhibidores de la α -amilasa, tales como p. ej., la maltosa.

Todos los genes analizados hasta la fecha de la biosíntesis de metabolitos secundarios, están dispuestos en una agrupación. Por ello, se asume que los genes de la biosíntesis de acarbosa, de acuerdo con la solicitud, también
45 están dispuestos en una agrupación génica tal. Los genes restantes que tienen importancia en la biosíntesis de pseudo-oligosacáridos, se pueden aislar de este modo aislando las regiones de ADN colindantes con el fragmento de ADN PstI de 6,8 kb, de acuerdo con la invención. Tal y como se ha mencionado anteriormente, se pueden aislar sin dificultad agrupaciones de genes homólogos, a partir de microorganismos distintos de *Streptomyces glaucescens* (DSM 40716).

50 Por ello, es un objeto de la invención una molécula de ADN recombinante que contiene los genes para la biosíntesis de acarbosa, la cual

(a) contiene una secuencia de ADN según la Tabla 4;

55 (b) contiene una secuencia de ADN que es capaz de hibridarse con la molécula de ADN, de acuerdo con (a)

(a) bajo condiciones rigurosas; o

60 (c) contiene una secuencia de ADN que debido a la degeneración del código genético, difiere de las moléculas de ADN según (a) y (b), pero permite la expresión de las proteínas que se pueden expresar correspondientes a las moléculas de ADN según (a) y (b).

La invención se refiere adicionalmente a una molécula de ADN recombinante que comprende el gen acbA, en particular que se caracteriza porque comprende la secuencia de ADN de los nucleótidos 1 a 720 según la Tabla 4 o partes de la misma; a una molécula de ADN recombinante que comprende el gen acbB, en particular que está
65 caracterizada porque comprende la secuencia de ADN de los nucleótidos 720 a 2006 según la Tabla 4, o partes de la misma; a una molécula de ADN recombinante que comprende el gen acbC, en particular que está caracterizada porque comprende la secuencia de ADN de los nucleótidos 2268 a 3332 según la Tabla 4, o partes de la misma; a una

molécula de ADN recombinante que comprende el gen *acbD*, en particular que está caracterizada porque comprende la secuencia de ADN de los nucleótidos 3332 a 4306 según la Tabla 4, o partes de la misma; a una molécula de ADN recombinante que comprende el gen *acbE*, en particular que está caracterizada porque comprende la secuencia de ADN de los nucleótidos 4380 a 5414 según la Tabla 4, o partes de la misma; a una molécula de ADN recombinante que comprende el gen *acbF*, en particular que está caracterizada porque comprende la secuencia de ADN de los nucleótidos 5676 a 6854 según la Tabla 4, o partes de la misma;

La invención se refiere adicionalmente a un cebador de oligonucleótidos para la amplificación con PCR de una molécula de ADN recombinante que es tal y como se ha descrito anteriormente y que comprende genes para biosintetizar acarbosa, teniendo el cebador en particular la secuencia según la Tabla 1.

La invención se refiere adicionalmente a un vector que comprende una molécula de ADN recombinante que comprende una molécula de ADN tal y como se ha descrito en el penúltimo y antepenúltimo párrafos, en particular la que está caracterizada porque el vector es un vector de expresión y dicha molécula de ADN está ligada funcionalmente a una secuencia de promotor, siendo adecuado el vector para la expresión en organismos hospedadores que se seleccionan a partir del grupo consistente en *E. coli*, *Bacillus subtilis*, *Actinomycetales*, como, p. ej., las cepas de *Streptomyces*, *Actinoplanes*, *Ampullariella*, *Streptosporangium*, *Streptomyces hygroscopicus* var. *limoneus*, *Streptomyces glaucescens*, así como hongos con importancia biotecnológica (p. ej., *Aspergillus niger*, *Penicillium chrysogenum*) y levaduras (p. ej., *Saccharomyces cerevisiae*), estando particularmente preferidos *Streptomyces glaucescens* (DSM 40716) o *Actinoplanes* sp. Dado que esta molécula de ADN está ligada funcionalmente a secuencias de promotor del vector, es sólo una realización preferida de la invención, también es posible una expresión empleando secuencias de promotor que son endógenas de la molécula de ADN, p. ej., los promotores que son en cada caso los naturales o los promotores naturales que se han mutado para mejorar el rendimiento de acarbosa. Tales promotores naturales son parte de la molécula de ADN de acuerdo con la invención.

Otro objeto de la invención es un vector que contiene una molécula de ADN de acuerdo con la invención, para uso en un procedimiento para inactivar o modificar los genes naturales de la biosíntesis de acarbosa, en un microorganismo productor de acarbosa. Un vector tal se selecciona preferentemente a partir del grupo consistente en pGM160 y de vectores tales como los descritos en los documentos de patente europea EP 0 334 282 y EP 0 158 872.

Otro objeto de la invención es una célula hospedadora que se transforma con una de las moléculas de ADN o los vectores descritos anteriormente, caracterizada en particular porque dicha célula hospedadora se selecciona entre el grupo consistente en *E. coli*, *Bacillus subtilis*, *Actinomycetales*, tales como p. ej., las cepas de *Streptomyces*, *Actinoplanes*, *Ampullariella*, *Streptosporangium*, *Streptomyces hygroscopicus* var. *limoneus*, *Streptomyces glaucescens*, así como hongos de importancia biotecnológica (p. ej., *Aspergillus niger*, *Penicillium chrysogenum*) y levaduras (p. ej., *Saccharomyces cerevisiae*); se prefiere particularmente que se escoja a partir del grupo compuesto por *Streptomyces glaucescens* (DSM 40716) y *Actinoplanes* sp.

Otro objeto de la invención son proteínas aisladas que se pueden obtener expresando los genes que son codificados por la molécula de ADN descrita en el párrafo anterior.

Las siguientes declaraciones se aplican a todos los genes individuales, identificados en el contexto de la presente invención y únicamente se reúnen por razones de claridad: otro objeto de la invención es una proteína, codificada por una molécula de ADN recombinante, tal y como se ha descrito en el párrafo anterior, caracterizada en particular porque comprende la secuencia de ADN de los nucleótidos 1 a 720 o 720 a 2006 o 2268 a 3332 o 3332 a 4306 o 4380 a 5414 o 5676 a 6854, según la Tabla 4 o partes de la misma; se prefiere muy especialmente una proteína que está codificada por el gen *acbA* o el gen *acbB* o el gen *acbC* o el gen *acbD* o el gen *acbE* o el gen *acbF*, que contiene la secuencia de aminoácidos según la Tabla 4 o partes de la misma.

Otro objeto de la invención es un procedimiento para obtener las proteínas que se han descrito anteriormente como parte del objeto de la invención, caracterizado porque

(a) las proteínas se expresan en una célula hospedadora adecuada, en particular las que se caracterizan porque dicha célula hospedadora se selecciona entre el grupo consistente en *E. coli*, *Bacillus subtilis*, *Actinomycetales*, tales como las cepas de *Streptomyces*, *Actinoplanes*, *Ampullariella*, *Streptosporangium*, *Streptomyces hygroscopicus* var. *limoneus*, *Streptomyces glaucescens*, así como hongos de importancia biotecnológica (p. ej., *Aspergillus niger*, *Penicillium chrysogenum*) y levaduras (p. ej., *Saccharomyces cerevisiae*); se prefiere especialmente que se seleccione entre el grupo consistente en *Streptomyces glaucescens* (DSM 40716) y *Actinoplanes* sp. y

(b) se aíslan.

Otro objeto de la invención es un procedimiento para la preparación de acarbosa, caracterizado porque

(a) uno o varios genes de la molécula de ADN recombinante que comprende una secuencia de ADN de acuerdo con la Tabla 4 o partes de la misma o que comprende una secuencia de ADN que es capaz de hibridarse, bajo condiciones rigurosas, con la molécula de ADN de acuerdo con la Tabla 4 o partes de la misma, o que comprende una secuencia de ADN que debido a la degeneración del código genético, difiere de las

moléculas de ADN que se acaban de describir, pero que permite la expresión de las proteínas que se pueden expresar que se corresponden con estas moléculas de ADN o con partes de las mismas, se emplean para la expresión en una célula hospedadora adecuada, que se selecciona en particular, a partir del mismo grupo que el descrito en el párrafo anterior, y

(b) la acarbosa se aísla a partir del material sobrenadante del cultivo de dicha célula hospedadora.

Otro objeto de la invención es un procedimiento para la preparación de acarbosa, caracterizado porque

(a) uno o varios genes de la molécula de ADN recombinante que comprende una secuencia de ADN de acuerdo con la Tabla 4 o con partes de la misma o que comprende una secuencia de ADN que es capaz de hibridarse, bajo condiciones rigurosas, con la molécula de ADN de acuerdo con la Tabla 4 o con partes de la misma, o que comprende una secuencia de ADN que, debido a la degeneración del código genético, difiere de las moléculas de ADN que se acaban de describir, pero que permite la expresión de las proteínas que se pueden expresar de forma correspondiente empleando las moléculas de ADN o partes de las mismas, se inactivan en una célula hospedadora productora de acarbosa, en particular de *Streptomyces glaucescens* (DSM 40716) y de *Actinoplanes* sp. y

(b) la acarbosa se aísla a partir de dicha célula hospedadora.

En este contexto la inactivación de uno o varios genes se puede efectuar mediante métodos convencionales de la biología molecular, por ejemplo, empleando los vectores descritos anteriormente (pGM160 y otros). Un gen que se va a inactivar podría ser, p. ej., el gen *acbE* que tiene probablemente una función reguladora. Del mismo modo se podrían inactivar genes con el fin de obtener acarbosa pura como el único producto de la fermentación y no obtener más una mezcla de de pseudo-oligosacáridos homólogos (véase, más arriba). La inactivación de esos genes se consigue preferentemente empleando los vectores que se han descrito anteriormente con este fin.

Otro objeto de la invención es un procedimiento para la preparación de acarbosa, caracterizado porque los procesos para preparar la acarbosa que se han descrito en los dos párrafos anteriores, se combinan entre sí, de modo que uno o varios de dichos genes se expresa artificialmente y uno o varios de dichos genes se inactiva.

Otro objeto de la invención es un procedimiento para modificar la expresión génica de los genes endógenos de la biosíntesis de acarbosa, mutando el gen promotor respectivo para obtener rendimientos mejorados de acarbosa. En este contexto, se pueden emplear métodos conocidos de recombinación homóloga, para introducir las mutaciones en la cepa de producción que se va a mejorar. Tales mutaciones pueden ser transiciones, deleciones y/o adiciones. Una “adición” puede indicar, por ejemplo, la adición de un solo nucleótido o de varios nucleótidos o de una o varias secuencias de ADN que tienen un efecto regulador positivo y que producen una mejora de la expresión de un gen endógeno para la biosíntesis de acarbosa. También el caso inverso, es decir, la adición de una secuencia de ADN que tiene un efecto regulador negativo sobre la represión de un gen endógeno para la biosíntesis de acarbosa, es una forma preferida de adición. Las “transiciones” pueden ser, p. ej., intercambios de nucleótidos que reducen o aumentan el efecto de los elementos reguladores que actúan negativa o positivamente. Con las “deleciones” se pueden eliminar elementos reguladores positivos o negativos. Los genes endógenos de este procedimiento están presentes preferentemente en los *Actinomycetales*, como p. ej., las cepas de *Streptomyces*, *Actinoplanes*, *Ampullariella*, *Streptosporangium*, *Streptomyces hygroscopicus* var. *limoneus*, *Streptomyces glaucescens*; muy especialmente están presentes en *Streptomyces glaucescens* (DSM 40716) y *Actinoplanes* sp.

Otro objeto de la invención es el uso de *Streptomyces* (DSM 40716) para la obtención de acarbosa.

Otro objeto de la invención es el uso de *Streptomyces* (DSM 40716) para la preparación de mutantes de esta cepa por la “vía clásica”, con la que se consigue una producción de acarbosa más abundante. Los procedimientos para la preparación de productores mejorados de productos naturales de este tipo se conocen desde hace tiempo y se emplean frecuentemente etapas clásicas de la mutagénesis y de la selección.

Otro objeto de la invención es un procedimiento para completar la agrupación génica para biosintetizar acarbosa y polisacáridos homólogos, de acuerdo con la Tabla 4, caracterizado porque

(a) se preparan las sondas para la hibridación que se obtienen a partir de la molécula de ADN, de acuerdo con la Tabla 4,

(b) estas sondas para la hibridación se emplean en el escrutinio genómico de las genotecas de ADN obtenidas a partir de *Streptomyces glaucescens* (DSM 40716), bajo condiciones rigurosas y

(c) los clones encontrados se aíslan y se caracterizan.

Otro objeto de la invención es un procedimiento para completar la agrupación génica para biosintetizar acarbosa y pseudo-oligosacáridos homólogos de acuerdo con la Tabla 4, caracterizado porque partiendo de la molécula de ADN recombinante, de acuerdo con la Tabla 4,

- (a) se preparan cebadores para la PCR,
- (b) estos cebadores para la PCR se emplean para acumular fragmentos de ADN del ADN genómico procedente de *Streptomyces glaucescens* (DSM 40716), combinándose estos cebadores con los cebadores que se hibridan con las secuencias del sistema del vector empleado,
- (c) los fragmentos acumulados se aíslan y se caracterizan.

Otro objeto de la invención es un procedimiento para aislar una agrupación génica para la biosíntesis de acar-bosa y pseudo-oligosacáridos homólogos a partir de microorganismos productores de acar-bosa que son diferentes de *Streptomyces glaucescens* (DSM 40716), caracterizado porque partiendo de la molécula de ADN recombinante según la reivindicación 4

- (a) se preparan las sondas para la hibridación,
- (b) estas sondas para la hibridación se emplean para el escrutinio del ADN genómico o del ADNc de las genotecas de ADN que se han obtenido a partir del microorganismo correspondiente y
- (c) los clones que se han encontrado se aíslan y se caracterizan.

Otro objeto de la invención es un procedimiento para aislar una agrupación génica para la biosíntesis de acar-bosa y pseudo-oligosacáridos homólogos, a partir de microorganismos productores de acar-bosa que sean diferentes de *Streptomyces glaucescens* (DSM 40716), caracterizado porque partiendo de la molécula de ADN recombinante según la reivindicación 4

- (a) se preparan cebadores para la PCR,
- (b) estos cebadores para la PCR se emplean para acumular fragmentos de ADN de ADN genómico o de ADNc de un microorganismo correspondiente,
- (c) los fragmentos acumulados se aíslan y se caracterizan y
- (d) eventualmente, se emplean en un procedimiento tal y como se ha descrito en el párrafo anterior.

Los procedimientos descritos para aislar una agrupación génica para la biosíntesis de acar-bosa y pseudo-oligosacáridos homólogos a partir de microorganismos productores de acar-bosa que sean distintos de *Streptomyces glaucescens* (DSM 40716), se caracterizan porque los microorganismos se seleccionan a partir del grupo consistente en *Actinomycetales*, tales como, p. ej., las cepas de *Streptomyces*, *Actinoplanes*, *Ampullariella*, *Streptosporangium*, *Streptomyces hygroscopicus* var. *limoneus*, *Streptomyces glaucescens*, preferentemente a partir del grupo consistente en *Streptomyces glaucescens* (DSM 40716) y *Actinoplanes* sp.

Otro objeto de la invención es el uso de *Streptomyces glaucescens* (DSM 40716) para aislar acar-bosa.

A continuación, se explica la invención con mayor detalle con ayuda de los ejemplos, las tablas y las figuras sin estar restringida a los mismos.

Todos los aislamientos de plásmidos se realizaron empleando un equipo de reactivos para el aislamiento (Nucleobond®) de la firma Macherey and Nagel (Düren, Alemania) según las instrucciones del fabricante. Las etapas de trabajo de biología molecular se realizaron según protocolos convencionales (Sambrock y col. (1989) Molecular cloning: A Laboratory Manual, 2ª ed. Cold Spring Harbor Laboratory, EE.UU.) o según indicaciones de los respectivos fabricantes. Las secuencias de ADN y proteicas se examinaron empleando el programa informático de Genetics Computer Group, versión 8 (programas: FastA, TFASTA, BlastX, Motifs, GAP, CODONPREFERENCE) y las bases de datos Swiss-Prot (versión 32), EMBL (versión 46) y Prosite (versión 12.2). La manipulación con técnicas de biología molecular de *Streptomyces glaucescens* y de *Actinoplanes* (aislamiento del ADN, transformaciones del ADN) se realizaron tal y como describen Hopwood y col.: Genetic Manipulation of Streptomycetes: A Laboratory Manual. The John Innes Foundation, Norwich, GB, 1985 y Motamedi y Hutchinson: Cloning and heterologous expression of a gene cluster for the biosynthesis of tetracenomycin C, the anthracycline antitumor antibiotic of *Streptomyces glaucescens*. Proc. Natl. Acad. Sci. USA 84:4445-4449 (1987).

En general, las hibridaciones se realizaron empleando un equipo de reactivos para marcar el ADN de forma no radiactiva de Boehringer Mannheim (nº de cat. 1175033). La visualización tenía lugar con el equipo de reactivos para detectar luminiscencia de Boehringer Mannheim (nº de cat. 1363514). En todos los ejemplos aportados en esta solicitud de patente, la hibridación se realizó bajo condiciones rigurosas: 68°C, 16 h, 5 x SSC, 0,1% de N-laurilsarcosina, 0,02% de SDS, 1% de reactivo de bloqueo (Boehringer Mannheim). SSC significa NaCl 0,15 M/citrato sódico 0,015 M. La definición de "condiciones rigurosas" empleada en esta memoria, se aplica a todos los aspectos de la presente invención en los que se hace referencia a "condiciones rigurosas". En este contexto, la forma de conseguir este rigor, es decir, las condiciones de hibridación citadas, no deben tener un efecto limitante, ya que el experto en la técnica

ES 2 276 427 T3

puede seleccionar también otras condiciones para obtener las mismas condiciones rigurosas, por ejemplo, mediante el uso de otras soluciones para la hibridación, junto con otras temperaturas.

Ejemplo 1

Síntesis y secuencias de los cebadores de la PCR y amplificación de los fragmentos procedentes de S. glaucescens (DSM 40716)

La PCR se realizó bajo condiciones convencionales empleando en cada caso 100 pmol de cebador 1 y 2 en 100 μ l de mezcla de reacción

Tampón para PCR ¹	10 μ l
Cebador para PCR	en cada caso 2,5 μ l
dNTPs	en cada caso 0,2 mM
BSA (10 mg/ml)	1 μ l
ADN molde	1 μ g (1 μ l)
Polimerasa Taq ² (5 unidades/ml)	1,5 μ l
H ₂ O	rellenar hasta 100 μ l
¹ : Promega	
² : Boehringer Mannheim	

Las muestras se recubren con 75 μ l de aceite mineral y la amplificación se realiza empleando un ciclador térmico de ADN TC1 de la firma Perkin Elmer.

Parámetros

Ciclos	Temperatura	Duración
1	96°C	5 min
	74°C	5 min
30	95°C	1,5 min
	74°C	1,5 min
1	74°C	5 min

En la Tabla 1 se presentan-introducen las secuencias de los cebadores degenerados que se deberían utilizar para amplificar las deshidratasa de dTDP-glucosa procedentes de diferentes estreptomicetos.

TABLA 1

Secuencias de cebadores para amplificar las 4,6-deshidratasa de dTDP-glucosa

Cebador 1:	CSGGSGSSGCSGGSTTCATSGG	(SEQ ID NO.: 1)
Cebador 2:	GGGWVCTGGYVSGGSCCGTAGTTG	(SEQ ID NO.: 2)

En esta tabla, S= G o C, W= A o T, V= A o G, Y= T o C.

Ejemplo 2

Secuencias de ADN de los fragmentos de la PCR aislados a partir de Streptomyces glaucescens (DSM 40716)

La secuenciación tuvo lugar por el método de terminación de cadena didesoxi según Sanger y col. (PNAS USA, 74:5463-5467 (1977)). Las reacciones se realizaron empleando el equipo de reactivos para secuenciación Auto Read Sequencing Kit[®] de la firma Pharmacia Biotech (Freiburg, Alemania) según las instrucciones del fabricante. Para la separación y la detección se empleó un secuenciador de ADN, ALF DNA Sequencer[®] de la firma Pharmacia Biotech (Freiburg, Alemania).

La posterior clonación de los fragmentos de la PCR (Sure Clone Kit[®], Pharmacia Biotech, Freiburg) en el vector de *E. coli* pUC18 y la secuenciación del fragmento, apoyaban la sospecha de que el fragmento codificaba una 4,6-deshidratasa de dTDP-glucosa. Sin embargo, se aislaron 2 genes diferentes, mostrando ambos un alto grado de homología con la 4,6-deshidratasa de dTDP-glucosa, pero que no eran idénticos. Los fragmentos de la PCR se designan de aquí en adelante acbD* y HstrE*.

ES 2 276 427 T3

Las secuencias de los fragmentos aislados de la PCR se muestran en la Tabla 2A y 2B y la comparación de la homología de las secuencias de aminoácidos deducidas de HstrE* y acbD*, se muestra en la Tabla 2C. Las dos proteínas muestran una identidad de sólo el 65%.

TABLA 2A

Secuencia de ADN de acbD (los sitios de ligación al cebador están subrayados, SEQ ID NO.: 3)*

Cebador 1

```

1  CCCGGGCGGG GCGGGGTTCA TCGGCTCCGC CTACGTCCGC CGGCTCCTGT
51  CGCCCGGGGC CCCCGGCGGC GTCGCGGTGA CCGTCCTCGA CAAACTCACC
101 TACGCCGGCA GCCTCGCCCG CCTGCACGCG GTGCGTGACC ATCCCGGCCT
151 CACCTTCGTC CAGGGCGACG TGTGCGACAC CGCGCTCGTC GACACGCTGG
201 CCGCGCGGCA CGACGACATC GTGCACTTCG CGGCCGAGTC GCACGTCGAC
251 CGCTCCATCA CCGACAGCGG TGCCTTCACC CGCACCAACG TGCTGGGCAC
301 CCAGGTCCTG CTCGACGCCG CGCTCCGCCA CGGTGTGCGC ACCCTCGTGC
351 ACGTCTCCAC CGACGAGGTG TACGGCTCCC TCCCGCACGG GGCCGCCGCG
401 GAGAGCGACC CCCTGCTCCC GACCTCGCCG TACGCGGCGT CGAAGGCGGC
451 CTCGGACCTC ATGGCGCTCG CCCACCACCG CACCCACGGC CTGGACGTCC
501 GGGTGACCCG CTGTTCTGAAC AACTACGGCC CGCACCAGTT CCCGGG

```

Cebador 2

TABLA 2B

Secuencia de ADN de HstrE (los sitios de ligación al cebador están subrayados, SEQ ID NO.: 4)*

Cebador 2

```

1  CCCCGGGTGC TGGTAGGGGC CGTAGTTGTT GGAGCAGCGG GTGATGCGCA
51  CGTCCAGGCC GTGGCTGACG TGCATGGCCA GCGCGAGCAG GTCGCCCGAC
45  101 GCCTTGGAGG TGGCATAGGG GCTGTTGGGG CGCAGCGGCT CGTCCTCCGT
151 CCACGACCCC GTCTCCAGCG AGCCGTAGAC CTCGTGCGTG GACACCTGCA
50  201 CGAAGGGGGC CACGCCGTGC CGCAGGGCCG CGTCGAGGAG TGTCTGCGTG
251 CCGCCGGCGT TGGTCCGCAC GAACGCGGCG GCATCGAGCA GCGAGCGGTC
55  301 CACGTGCGAC TCGGCGGCGA GGTGCACGAC CTGGTCCTGG CCGGCCATGA
351 CCCGGTTCGAC CAGGTCCGCG TCGCAGATGT CGCCGTGGAC GAAGCGCAGC
401 CGGGGGTGGT CGCGGACCGG GTCGAGGTTG GCGAGGTTGC CGGCGTAGCT
60  451 CAGGGCGTCG AGCACGGTGA CGACGGCGTC GGGCGGCCCG TCCGGACCGA
501 GGAGGGTGCG GACGTAGTGC GAGCCCATGA ACCCCGCCGC C

```

Cebador 1

Ejemplo 4

Clonación del fragmento *PstI* de 6,8 kb de longitud

Entre otros, mediante el fragmento de PCR *acbD**, se marca un fragmento de ADN *PstI* de 6,8 kb de longitud (Fig. 2). Este fragmento de ADN se aisló del modo siguiente. La región del gel se escindió con una cuchilla de afeitar y el ADN se aisló del gel empleando un equipo de reactivos para aislamiento de la firma Pharmacia Biotech y se clonó en el plásmido pUC19 (plásmido *pacb1*) cortado con la enzima de restricción *PstI* y se transformó en la cepa DH5 α de *E. coli*. Los clones individuales se subcultivaron a partir de estas placas y se realizó el aislamiento del ADN plasmídico empleando estos clones. Se realizó una amplificación con PCR, con los cebadores 1 y 2, descritos anteriormente (Tab. 1), empleando el ADN procedente de estos clones. De este modo, se aisló el clon adecuado de *E. coli*, que contenía el fragmento *PstI* con una longitud de 6,8 kb.

Ejemplo 5

Secuenciación del fragmento de ADN *PstI* aislado de 6,8 kb de longitud

El ADN se digirió con diversas enzimas de restricción y los fragmentos de ADN aislados se clonaron en pUC19. A continuación, se determinó la secuencia de ADN del fragmento completo, que se muestra en la Tab. 4 (SEQ ID NO.: 7). La secuencia de ADN del fragmento *PstI* de 6,8 kb de longitud se confirmó sólo parcialmente, mediante secuenciación suplementaria de la cadena opuesta. Se encontraron varios marcos de lectura abiertos que codificaban diversas proteínas (programas: CODONPREFERENCE y BlastX). Se encontró un total de 6 regiones codificadoras. Un gen que tenía un alto grado de homología con la proteína que se une a ATP, *acbA*, una aminotransferasa, *acbB*, una dTDP-glucosa-sintetasa, *acbC*, una dTDP-glucosa-deshidratasa, *acbD*, un gen regulador con homologías con la familia de proteínas LacI, *acbE* y una proteína con similitudes con las proteínas que se unen a azúcar, *acbF*. La secuencia de los genes *acbA* y *acbF* sólo se determinó parcialmente. Las homologías con otras proteínas procedentes de bases de datos y las propiedades de las proteínas putativas, se sumarizan en la Tab. 3. La Fig. 3 muestra, de forma resumida, un mapa de restricción del fragmento que contiene los sitios de corte de restricción más importantes, mencionados en el texto, y la disposición de los marcos de lectura abiertos identificados.

TABLA 3

Análisis de los marcos de lectura abiertos identificados sobre el fragmento *PstI* de 6,8 kb de longitud procedente de *Streptomyces glaucescens* (DSM 40716)

ORF	Aminoácidos	PM	FastA [§]	% de identidad	Número de orden [§]
<i>acbA</i>	239	*	MalK, <i>E. coli</i>	29 %	P02914
<i>acbB</i>	429	45618	DgdA, <i>Burkholderia cepacia</i>	32 %	P16932
<i>acbC</i>	355	37552	StrD, <i>Streptomyces griseus</i>	60 %	P08075
<i>acbD</i>	325	35341	StrE, <i>Streptomyces griseus</i>	62 %	P29782
<i>acbE</i>	345	36549	DegA, <i>Bacillus subtilis</i>	31 %	P37947
<i>acbF</i>	396	*	MalE, <i>E. coli</i>	22 %	P02928
* marco de lectura abierto incompleto; [§] Base de datos Swiss-Prot (versión 32)					

Ejemplo 6

Deleción de los genes *acbBCD* para la biosíntesis de pseudo-oligosacáridos procedentes de *Streptomyces glaucescens* (DSM 40716)

Una prueba de que el fragmento de ADN identificado codificaba los genes para la biosíntesis de pseudo-oligosacáridos, se realizó del modo siguiente. Una región génica de 3,4 kb (fragmento b *EcoRI*-*SstI*, Fig. 3) se sustituyó con el gen de resistencia a la eritromicina (1,6 kb) y se clonó con las regiones de ADN flanqueantes del fragmento

PstI de 6,8 kb (pacb1) en el plásmido pGM160 sensible a la temperatura. El plásmido se construyó tal y como se describe a continuación: el fragmento EcoRI-HindIII de 2,2 kb de longitud (c, Fig. 3) procedente del plásmido pacb1, se clonó en pGEM7zf (Promega, Madison, WI, EE.UU.; plásmido pacb2), así como el fragmento SstI de 1 kb de longitud procedente de pacb1 (a, Fig. 3) se clonó en pUC19 (plásmido pacb3). A continuación se realizó una ligación empleando los siguientes fragmentos. El plásmido pGM160 (Muth y col. (1989) Mol. Gen Genet. 219:341-348) se cortó con BamHI/HindIII, el plásmido pacb2 con XbaI/BamHI (c, Fig. 3), el plásmido pacb3 con EcoRI/HindIII (a, Fig. 3) y el plásmido pIJ4026 (Bibb y col. (1985) Gene 38:215-226) se cortó con EcoRI/XbaI para aislar el gen de resistencia ermE de 1,6 kb de longitud. Los fragmentos se ligaron en una mezcla y se transformaron en *E. coli* DH5 α y se seleccionaron sobre ampicilina. El plásmido pacb4 resultante se aisló a partir de *E. coli* DH5 α , se verificó si era correcto mediante digestión con restricción y a continuación se transfirió mediante transformación con protoplastos en *S. glaucescens* (DSM 40716). Los transformantes se seleccionaron con tioestreptona a 27°C, sobre agar R2YENG. A continuación, los transformantes se incubaron a la temperatura no permisiva de 39°C y de este modo, se dirigió la integración del plásmido en el genoma mediante recombinación homóloga (selección con tioestreptona (25 μ g/ml) y eritromicina (50 μ g/ml)). Bajo esas condiciones, sólo podían crecer los clones en los que se había integrado el plásmido en el genoma. Los clones correspondientes se aislaron, se llevaron a esporulación (medio 1, véase a continuación) y se extendieron en placas sobre agar que contenía eritromicina (medio 1). Los clones individuales se aislaron de nuevo desde esas placas y se extendieron en medio que contenía tioestreptona y en medio que contenía eritromicina. Los clones que eran resistentes a la eritromicina pero ya no eran resistentes a la tioestreptona, se analizaron. En esos clones, los genes acbBCD se reemplazaron por ermE. Se examinaron diversos clones y la cepa *S. glaucescens* (DSM 40716) Δ acb se seleccionó finalmente como cepa de referencia (resistente a la eritromicina, sensible a la tioestreptona) para una investigación posterior.

Medio 1

25	Extracto levadura	de 4 g/L
	Extracto malta	de 10 g/L
	Glucosa	4 g/L
	Agar	15 g/L
30	pH	7,2

En otro experimento se examinó si la cepa correspondiente todavía producía acarbosa. Algunos clones se dejaron crecer y se investigó la formación del inhibidor de la α -amilasa, pero no se encontró actividad. A continuación, los mutantes se caracterizaron adicionalmente mediante hibridación de tipo Southern. La integración del gen ermE había tenido lugar en el sitio pronosticado. La Fig. 4 muestra una hibridación de tipo Southern que se había realizado con el tipo silvestre y con el mutante por delección Δ acb de *Streptomyces glaucescens* (DSM 40716). Como sonda se empleó el fragmento SstI de pacb3. El ADN cromosómico se aisló a partir del tipo silvestre y de los mutantes y se digirió con las enzimas PstI y PstI/HindIII. El patrón del fragmento obtenido para el mutante por delección se corresponde con el acontecimiento de recombinación pronosticado. El tipo silvestre muestra un fragmento PstI invariable de 6,8 kb de longitud, mientras que el mutante muestra un fragmento que está reducido en 1,8 kb (comparar la pista 1 y la 3, Fig. 4). Con la integración del gen de resistencia ermE, se introducía adicionalmente un sitio de corte interno para HindIII en el fragmento PstI (comparar la pista 2 y la 4, Fig. 4).

Ejemplo 7

Inhibición de la α -amilasa mediante acarbosa

Empleando un ensayo enzimático para detectar almidón (TC-Starch, Boehringer-Mannheim, n° de cat. 297748) era posible mostrar que el compuesto aislado a partir de *Streptomyces glaucescens* (DSM 40716) inhibe la α -amilasa. Principio del ensayo: el almidón se escinde en D-glucosa mediante la amiloglucosidasa. A continuación, la glucosa se convierte con hexoquinasa en glucosa-6-fosfato y ésta se convierte con glucosa-6-fosfato-deshidrogenasa en D-gluconato-6-fosfato. Esta reacción produce NADPH, cuya formación se puede determinar fotométricamente. La acarbosa inhibe la α -amilasa e impide de este modo la formación de D-glucosa y finalmente también la formación de NADPH.

Ejemplo 8

*Medio de crecimiento para *S. glaucescens* (DSM 40716) y producción de acarbosa*

La fermentación se realizó a 27°C en un agitador por rotación a 120 rpm, en frascos Erlenmeyer de 500 ml con una punción lateral y 100 ml de medio 2. La fermentación se interrumpió al cabo de 2 o 3 días. Los pseudo-oligosacáridos se detectaron en un ensayo de difusión en placa, tal y como se describe en el Ejemplo 9. No se produjeron inhibidores de la α -amilasa cuando se empleó el medio 3. Esto significa que la glucosa inhibe la producción de los pseudo-oligosacáridos. También otros azúcares, tales como maltosa o sacarosa, o fuentes complejas de azúcar (extracto de malta) se pueden tomar en consideración para producir pseudo-oligosacáridos con *S. glaucescens* (DSM 40716).

ES 2 276 427 T3

Medio 2

5	Harina de soja	20 g/L
	Almidón	20 g/L
	pH	7,2

Medio 3

10	Harina de soja	20 g/L
	Glucosa	20 g/L
15	pH	7,2

Ejemplo 9

Bioensayo que emplea Mucor miehei

20 Una suspensión de esporas de la cepa *Mucor miehei* se vertió en agar (medio 5) (10^5 esporas/ml) y 10 ml de esta mezcla se vertieron en cada caso en placas de Petri. Se cargaron discos de papel para ensayo (diámetro 6 mm) con 10 μ l de acarbose (1 mg/ml) o con una muestra procedente del cultivo de *S. glaucescens* y se depositaron sobre las placas del ensayo. La incubación de las placas tuvo lugar a 37°C. Se observaron halos de inhibición sobre el medio 5
25 que contenía almidón. Como testigo se empleó una placa que se había preparado con glucosa (medio 4) en lugar de almidón. Sobre este medio no se formaba ningún halo de inhibición alrededor de los discos de filtro cargados con el compuesto activo.

Medios 4 y 5

30	KH ₂ PO ₄ x 3 H ₂ O	0,5 g
	MgSO ₄ x 7 H ₂ O	0,2 g
	NaCl	0,1 g
35	Sulfato de amonio	5 g
	Base de levadura nitrogenada	1,7 g
	Glucosa (4) o almidón (5)	5 g
	Agar	15 g

Ejemplo 10

Transformación de S. glaucescens (DSM 40716)

45 Los protoplastos de la cepa de *Streptomyces glaucescens* se aislaron tal y como describen Motamedi y Hutchinson ((1987) PNAS USA 84:4445-4449) y el ADN plasmídico aislado se transfirió a las células mediante transformación PEG, tal y como describen Hopwood y col. ((1985) Genetic manipulations of Streptomyces: A Laboratory Manual. The John Innes Foundation, Norwich GB). Los protoplastos se regeneraron sobre medio R2YENG a 30°C (Motamedi y Hutchinson (1987) PNAS USA 84:4445-4449). Las placas de agar se recubrieron después de 18 h con una solución
50 que contenía tioestreptona y se incubaron a 30°C (concentración final de tioestreptona: 20 μ g/ml).

Ejemplo 11

55 *Aislamiento de los pseudo-oligosacáridos procedentes de Streptomyces glaucescens (DSM 40716), análisis HPLC y espectroscopía de masas*

Aislamiento

60 El caldo de cultivo se separó del micelio por filtración. El material filtrado del cultivo obtenido de este modo, se cargó a continuación en una columna XAD16, después de lo cual, la columna se lavó con agua y los componentes activos se eluyeron con metanol al 30%. El material eluido se concentró hasta la fase acuosa y se extrajo con acetato de etilo para eliminar las impurezas lipofílicas. La fase acuosa se concentró a continuación y los componentes activos se purificaron adicionalmente en metanol al 5% empleando una columna P2 de biogel. Las fracciones individuales se recogieron en un recolector de fracciones. Las fracciones individuales se analizaron con bioensayo con *Mucor miehei*.
65 El material eluido activo se cromatografió de nuevo para purificar adicionalmente, en metanol al 5% sobre biogel P2. El material aislado de este modo se investigó adicionalmente con HPLC y MS.

HPLC

Columna: Nucleosil® 100 C-18

Eluyente ácido fosfórico al 0,1% = A /acetonitrilo = B

Gradiente: de 0-100% de B en 15 min

Detección: 215 nm

Flujo 2 ml/min

Volumen de inyección: 10-20 µl

Empleando HPLC no era posible distinguir la preparación de pseudo-oligosacárido de *S. glaucescens* (DSM 40716) de la acarbosa auténtica. El tiempo de retención y el espectro de absorción de rayos UV de los dos componentes eran idénticos en este sistema de eluyente. La mezcla de pseudo-oligosacáridos no se fraccionó bajo estas condiciones.

Análisis mediante espectroscopía de masas (MS)

Los pesos moleculares y el patrón de fragmentación de la acarbosa auténtica y de los pseudo-oligosacáridos aislados a partir de *Streptomyces glaucescens* (DSM 40716) se determinaron mediante MS con electrovaporización. El análisis de la acarbosa que se puede obtener comercialmente en Bayer (Glucobay), proporcionó un pico de masa a 645,5 (acarbosa). Las muestras purificadas procedentes de *S. glaucescens* (DSM 40716) contienen una mezcla de diferentes pseudo-oligosacáridos cuyas cadenas laterales de azúcar tienen longitudes diferentes: 969 (acarbosa + 2 unidades de glucosa), 807 (acarbosa + 1 unidad de glucosa), 645 (se corresponde con la acarbosa auténtica). Cuando la acarbosa y el compuesto que se aísla a partir de *S. glaucescens* (DSM 40716) y que tiene un peso molecular de 645 se fragmentan, se forman los mismos fragmentos de moléculas: 145 (4-amino-4,6-desoxiglucosa) 303 (acarviosina) y 465 (303 con una unidad de glucosa).

Actinoplanes sp. SE50 también produce una mezcla de moléculas de acarbosa que tienen cadenas laterales de azúcar con longitudes diferentes (Truscheit (1984) VIIIth International Symposium on Medicinal Chemistry, Proc. Vol. I. Swedish Academy of Pharmaceutical Sciences, Estocolmo, Suecia). La longitud de las cadenas laterales de azúcar se pueden influir con la elección de los parámetros de la fermentación y del sustrato en la solución del nutriente.

Ejemplo 12

Hibridación Southern con Actinoplanes sp. SE50/110 (ATCC31044)

El ADN cromosómico se aisló a partir de la cepa *Actinoplanes* sp. SE50/100 y se digirió con las enzimas de restricción (PstI y BamHI). A continuación, se realizó una hibridación Southern empleando una sonda que incluye la región codificadora acbD de la 4,6-deshidratasa de dTDP-glucosa procedente de *Streptomyces glaucescens* (DSM 40716) (fragmento d, Fig. 3). La sonda se hibrida con distintas bandas de *Actinoplanes* sp. SE50/110 (Fig. 5, pistas 1 y 2). Esto ofrece la posibilidad de aislar los fragmentos correspondientes de *Actinoplanes* sp. SE50/100 y de otras líneas celulares. El que estas regiones de ADN estén de hecho implicadas en la biosíntesis de acarbosa se tiene que mostrar todavía en investigaciones adicionales. Alternativamente, los cebadores 1 y 2 de la PCR (Tab. 1) también podrían utilizarse para amplificar la 4,6-deshidratasa de dTDP-glucosa procedente de *Actinoplanes* sp.

Leyendas

Fig. 1: Hibridación Southern con *S. glaucescens* (DSM 40716). Pista 1: PstI, pista 2: BamHI, pista 3: BglII. Como sonda se empleó el fragmento de PCR HstrE* marcado. Marcación de los fragmentos de ADN que están implicados en la biosíntesis de hidroxistreptomina.

Fig. 2: Hibridación Southern con *S. glaucescens* (DSM 40716). Pista 1: PstI, pista 2: BamHI, pista 3: BglII. Como sonda se empleó el fragmento de PCR acbD* marcado. Marcación de fragmentos de ADN que participan en la biosíntesis de los acbD* pseudo-oligosacáridos.

Fig. 3: Mapa de restricción del fragmento PstI de 6,8 kb de longitud procedente de *Streptomyces glaucescens* (DSM 40716). Los marcos de lectura abiertos y la dirección en la que se transcribe cada uno, se indican con flechas. Los fragmentos a, b, c y d identifican las regiones de ADN que se explican con más detalle en el texto.

Fig. 4: Hibridación Southern que emplea Δacb de *Streptomyces glaucescens*: pista 1: PstI, pista 2: PstI/HindIII y *Streptomyces glaucescens* (DSM 40716) pista 3: PstI, pista 4: PstI/HindIII. Como sonda se empleó el fragmento a SstI marcado de 1,0 kb de longitud (Fig. 3).

ES 2 276 427 T3

Fig. 5: Hibridación Southern empleando *Actinoplanes* sp. SE50/100: pista 1: PstI, pista 2: BamHI y *Streptomyces glaucescens* (DSM 40716) pista 3: PstI. Como sonda se empleó el fragmento d marcado SmaI/EcoRI de 1,0 kb de longitud (4,6-deshidratasa de dTDP-glucosa, Fig. 3). Las flechas indican los fragmentos de ADN marcados (BamHI: 2,1 y 0,7 kb, PstI: ~11-12 kb)

Tab. 4: Secuencia de ADN del fragmento PstI de 6,8 kb de longitud procedente de *Streptomyces glaucescens* (DSM 40716) (SEQ ID NO.: 7). Las secuencias de aminoácidos deducidas (SEQ ID NO.: 8-13) de los marcos de lectura abiertos identificados

se indican debajo de la secuencia de ADN. Los codones de inicio, de detención y los sitios potenciales de unión a ribosomas, están subrayados.

acbA: SEQ ID NO.: 8

acbB: SEQ ID NO.: 9

acbC: SEQ ID NO.: 10

acbD: SEQ ID NO.: 11

acbE: SEQ ID NO.: 12

acbF: SEQ ID NO.: 13

(Tabla pasa a página siguiente)

ES 2 276 427 T3

TABLA 4

(SEQ ID NO.: 7, 8, 9, 10, 11, 12, 13)

5
P
s
t
I

10
CTGCAGGGTTCCCTGGTGCACGACCCGCCCTGGTGCAGCACCAGGGCGCTGTCGCAGAT
-----+-----+-----+-----+-----+ 60
GACGTCCCAAGGGACCACGTGCTGGGCGGGGACCAGCTGCTGGTCCCGCGACAGCGTCTA
Q L T G Q H V V R G Q D V V L A S D C I -

15
CGCGGCGATGTCGGCGATGTCGTGGCTGGTGAGCACCACGGTGGTGCCAGTTCCCGGTG
-----+-----+-----+-----+-----+ 120
GCGCCGCTACAGCCGCTACAGCACCAGCCACTCGTGGTGCCACCACGGGTCAAGGGCCAC
A A I D A I D H S T L V V T T G L E R H -

20
GGCGCGGTTGACCAGCCGGCGCACCGCGTCCTTCAGCACCATGTCGAGGCCGATCGTGGG
-----+-----+-----+-----+-----+ 180
CCGCGCCAACCTGGTCCGCGCGTGGCGCAGGAAGTCGTGGTACAGCTCCGGCTAGCACCC
A R N V L R R V A D K L V M D L G I T P -

25
CTCGTCCAGAACAGCAGCGGCCGGGTGCTGCAGCAGGCTCGCCGCGATCTCGGCGCGCAT
-----+-----+-----+-----+-----+ 240
GAGCAGGGTCTTGTGCTGCGCGGCCAGCACGTGCTCCGAGCGCGCTAGAGCCGCGCGTA
E D W F L V A P D H L L S A A I E A R M -

30
S
P
h
I
CGCGTGTCCGAGGCTGAGCTGCCGCACGGGGGTGGACCCAGCGCGTGCATGTCGAGGAG
-----+-----+-----+-----+-----+ 300
CGCGACAGGCTCCGACTCGACGGCGTGCCCCACCTGGGGTCCGCGAGCTACAGCTCCTC
R Q G L S L Q R V P T S G L A D I D L L -

35
GTCCCGGAACAGGGCGAGGTTGCCCGGTTAGACCGGTCCGGGGATGTCGTAGATGCCGGC
-----+-----+-----+-----+-----+ 360
CAGGGCCTTGTCCCGCTCCAACGCGGCCATCTGGCCAGGCCCTACAGCATCTACGCGC
D R F L A L N R R Y V P G P I D Y I R R -

40
K
P
n
I
CAGGATGCGGAAGGAGTCCGGGTACCGACAGGTCCCACCAGAGCTGGCTGCGCTGCCCGAA
-----+-----+-----+-----+-----+ 420
GTCCTACGCCTTCCTCAGCCCATGGCTGTCCAGGGTGGTCTCGACCGACGCGACCGGCTT
L I R F S D P V S L D W W L Q S R Q G F -

45
GACGACGCGGATCGTGCGGGCGTTGCGCTGCCGGTGCCGGTAGGGCTCCAGCCCGGCGAC
-----+-----+-----+-----+-----+ 480
CTGCTGCGGCTAGCACGCCCGCAACGCGACGGCCACGGCCATCCCGAGGTCCGGGCGCTG
V V G I T R A N R Q R H R Y P E L G A V -

50
CGTGACGCGGCCGGAGGTGGGGGTCATGATGCCGGTCAGCATCTTGATCGTGGTGCAGTT
-----+-----+-----+-----+-----+ 540
GCACGTGCGCGGCCTCCACCCCACTACTACGGCCAGTCGTAGAACTAGCACCAGCTGAA
T C R G S T P T M I G T L M K I T T S K -

55
GCCCGCTCCGTTGGCGCCGATGTAGGCGGTCTTCGTGCCGCGCGGTATCTCGAAGGAGAC
-----+-----+-----+-----+-----+ 600
CGGCCGAGGCAACCGCGCTACATCCGCCAGAAGCACGGCCGGCCATAGAGCTTCCTCTG
G A G N A G I Y A T K T G A P I E F S V -

ES 2 276 427 T3

K
P
n
I

5 GTCTGACGGCGCGCACGACGGGTACCGGCGGGTCAGGAGGGTGGAGAGGCTGCCGAG
-----+-----+-----+-----+-----+ 560
CAGCAGCTGCCGCGCGTGTGCGCCATGGCGGCCAGTCTCTCCACCTCTCCGACGGCTC
D D V A R V V R Y R R T L L T S L S G L -

10 CAGGCGGGGCTCGCGTTCGGCCAGCCGGAACCTCTTGACGAGGTGTTCCGCCACGATCAC
-----+-----+-----+-----+-----+ 720
GTCCGGCCCCGAGCGCAAGCCGGTCCGCCTTGAGGAACTGCTCCACAAGCCGGTGTAGTG
*
L G P E R E A L R P E K V L H E A V I V -
-----+-----+-----+-----+-----+ acbA
15 GCGATCACCCTCGACGGCCGTCTCCAGCAGGCGCAGGCCCTCGTCGAGCAGCGCCTCG
-----+-----+-----+-----+-----+ 780
CGCTAGTGGCGAGCTGCCGGCAGAGGTGCTCCGCGTCCGGGAGCAGCTCGTCGCGGAGC
A I V R E V A T E L L R L G E D L L A E -

20 TCGAGGGTGAACGGCGGTGCCAGCCGAGGATGTGGCCGCCAGGGAGGTGCGCAGCCCC
-----+-----+-----+-----+-----+ 840
AGCTCCCACTTGCCGCCACGGTCCGCGTCTACACGGCGGGTCCCTCCACGCGTCGGGG
D L T F P P A L R L I H G G L S T R L G -

S
m
a
I

25 AGGTCGAGGGCGGTGGTGTAGACGGCCCGGGCGGTCTCGGGGGCGGGTGCCCGGCCGACG
-----+-----+-----+-----+-----+ 900
TCCAGCTCCCGCCACCACATCTGCCGGGCCCGCCAGAGCCCCCGCCACGGGCCGGTGC
L D L A T T Y V A R A T E P A P A R G V -

30 GCGTCGGTGACGAACCTCCAGGCCCCACAGCAGTCCGAGGCCGCGTACCTGGCCGAGCTGG
-----+-----+-----+-----+-----+ 960
CGCAGCCACTGCTTGAGGTCCGGGGTGTCTGTCAGGCTCCGGCGCATGGACCGGCTCGACC
A D T V F E L G W L L G L G R V Q G L Q -

S
s
t
I

35 GGAAGCGGGACTCCAGGGCGCGCAGCCGCTCCTGGATGAGCTCGCCGAGGACGCGCACG
-----+-----+-----+-----+-----+ 1020
CCCTTCGCCCTGAGGTCCCGCGCGTCCGGCAGGACCTACTCGAGCGGCTCCTGCGCGTGC
P F R S E L A R L R E Q I L E G L V R V -

40 CCGTCGATCAGCCGGTCCGCGCTCGACGACCTCCAGCGTGGCGCGGGCGGGCGATCCCC
-----+-----+-----+-----+-----+ 1080
GCCAGCTAGTCCGGCCAGCGCGAGTGTGAGGTCGCACCGCGCCCGCCCGCTAGGGG
R D I L R D R E V V E L T A R A A A I G -

S
m
a
I

50 AGTGGGTTGCTCGCGTACGTCGAGGCGTACGCCCCGGGGTGGCCGCTCCGGCCTGCGCA
-----+-----+-----+-----+-----+ 1140
TCACCAACGAGCGCATGCAGCTCCGCATGCGGGGCCCCACCGCGGAGGCCGACGCGT
L P N S A Y T S A Y A G P H G G G A Q A -

ES 2 276 427 T3

GCTTCCGCGCGTCCGGCCAGCACGGCGAAGGGGAATCCGCTCGCGGTGCCCTTGGACAGC
 -----+-----+-----+-----+-----+ 1200
 CGAAGGCGCGCAGGCCGGTCTGTGCCGCTTCCCCTTAGGCGAGCGCCACGGGAACCTGTCTG
 5 A E A R G A L V A F P F G S A T G K S L -

 ATCGCCAGGTCCGGCTCGATGCCGAACAGTTTCGTGGCGAGGAAGGCGCCGGTGCGCCCG
 -----+-----+-----+-----+-----+ 1260
 TAGCGGTCCAGGCCGAGCTACGGCTTGTCAAGCGACCGCTCCTTCCGCGGCCACGCGGGC
 10 M A L D P E I G F L E S A L F A G T R G -

 CCGCCGGTGAGGACCTCGTCCGGCGACGAGCAGCACGCCGCCGTCCCGGCAGGCGCCGGCG
 -----+-----+-----+-----+-----+ 1320
 GCGCGCCACTCCTGGAGCAGCCGCTGCTCGTCTGTGCGGCGGCAGGGCCGTCCGCGGCCGC
 15 G G T L V E D A V L L V G G D R C A G A -

 ATCCGCTCCAGTAGCCGGGGGGCGGCACGATGACGCCTGCCGCGCCGAGGACGGGTTCG
 -----+-----+-----+-----+-----+ 1380
 TAGGCGAGGGTCATCGGCCCCCGCCGTGCTACTGCGGACGGCGCGGCTCCTGCCCAAGC
 20 I R E W Y G P P P V I V G A A G L V P E -

 AAGACCAGGGCCGAGACGTTGGGCTTCTCCGCGATGTGCCGGCGCACGAGGGTCGCGCAC
 -----+-----+-----+-----+-----+ 1440
 TTCTGGTCCCGGCTCTGCAACCCGAAGAGGCGCTACACGGCCGCGTGTCTCCAGCGCGTG
 25 F V L A S V N P K E A I H R R V L T A C -

 CGCACGTGCGACGAGGGGTACTCCAGGCCAGGGGACAGCGGTAGCCAGTAGGGGCTGTA
 -----+-----+-----+-----+-----+ 1500
 GCGTGCGAGCGTGTCTCCCATGAGGTCCGGGTCCCCTGTGCGCATCGGTCTATCCCGACAT
 30 R V D C S P Y E L G L P C R Y G T P A T -

 GCCAGCACGCTGTTGCCGCTGAAGGCCTGGTGGCCGATGTCCAGTGGACCAGCATCCGG
 -----+-----+-----+-----+-----+ 1560
 CGGTCTGTGCGACAACGGCGACTTCCGGACCACCGGTACAGGGTCACCTGGTCTAGGCC
 35 A L V S N G S F A Q H G I D W H V L M R -

 GCGCCCATGGTCTTGCCGCTGGAAGCCGTGGCGCAGGGCGCAGATCCGGTTGCGGCCCGGC
 -----+-----+-----+-----+-----+ 1620
 CGCGGGTACCAGAACGGCACCTTCGGCACCGCGTCCCGCGTCTAGGCCAACGCCGGGCCG
 40 A G M T K G H F G H R L A C I R N R G P -

 GCGGCGGTGCGCTGGACGACCCGCGAGGGCGGCCTCGACCACCTCCGCGCCGGTGGAGAAG
 -----+-----+-----+-----+-----+ 1680
 CGCCGCCAGCGGACCTGCTGGGCGTCCCGCCGGAGCTGGTGGAGGCGCGGCCACCTCTTC
 45 A A T A Q V V R L A A E V V E A G T S F -

 AAGGCGTAGGTGTGAGCTGTTTCGGGCAGCAGCCTGGCGAGCAGTTCAGCAGGCCGSGC
 -----+-----+-----+-----+-----+ 1740
 TTCCGCATCCACAGCTCGACAAGCCCGTCTCGGACCGCTCGTCAAGGTCTCCGGCCGC
 50 F A Y T D L Q E P L L R A L L E L L G A -

 CGGTCCGGCGTGGCGCTGTCTGTGGACGTTCCACAGGCGGCGGGCCTGGGTGGTGAGTGCC
 -----+-----+-----+-----+-----+ 1800
 GCCAGGCCGACCGCGACAGCACCTGCAAGGTGTCCGCCGCCCGGACCCACCACTCACGG
 55 R D P T A S D H V N W L R R A Q T T L A -

 TCGACGACCTCCGGGTGCCCGTGGCCAGTGACTGGGTGAGGGTCCCGGCCGGAAGTCG
 -----+-----+-----+-----+-----+ 1860
 AGCTGCTGGAGGCCACGGGCACCGGTCACTGACCACTCCAGGGCCGGCGCTTCAGC
 60 E V V E P H G H G L S Q T L T G A A F D -

 65

ES 2 276 427 T3

AGGTACTGGTTGCCGTCCAGGTCCGGTCAGAACGGGACCGCGTCCCTCGGCGAAGACCCGG
 -----+-----+-----+-----+-----+ 1920
 TCCATGACCAACGGCAGGTCCAGCCAGTCTTGCCCTGGCGCAGGGAGCCGCTTCTGGGCC
 5 L Y Q N G D L D T L V P G R G E A F V R -

 CGTCCGTGGACGGCTTCCTCGGAGGCGCCCGGCCAGGTGGCGGGCCTCCCGTGCCAGG
 -----+-----+-----+-----+-----+ 1980
 GCAGGCACCTGCCGAAGGAGCCTCCGCGGGCCGCGGTCCACCGCCCGGAGGGCACGGTCC
 10 R G H V A E E S A G P A L H R A E R A L -

 TGCTGTGTCTGCCGTAAGCCTGTCTATCGCTGCCTCTGCTCGTCCGACCGGCTGACGCGAT
 -----+-----+-----+-----+-----+ 2040
 ACGACACAGACGGCATTCGGACAGTAGCGACGGAGACGAGCAGCCTGGCCGACTGCGCTA
 15 H Q T Q R L G T M
 -----+-----+-----+-----+-----+
 CGCCGGCGAACTGCGTTGTGGCGCACCACGGTTGGGGCGGCTCGGCGCTGAGTCAAACAC
 -----+-----+-----+-----+-----+ 2100
 GCGGCCGCTTGACGCAACACCGCGTGGTGCCAAACCCCGCGAGCCGCGACTCAGTTTGTG

 TTGAACACACACCGCTGCAAGAGTTTGGCGGTTGTTTCAGAAAGTTGTTGCGAGCGGCC
 -----+-----+-----+-----+-----+ 2160
 AACTTGTGTGTGGCGACGTTCTCAAACGCCCAACAAAGTCTTTCAACAACGCTCGCCGGG

 CGGCACTCTGGTTGAGTCGACGTGCTTACGGCGCCACCACGCCTCACGTTGAGGAGGGA
 -----+-----+-----+-----+-----+ 2220
 GCCGTGAGACCAACTCAGCTGCACGAATGCCGCGGTGGTGCGGAGTGCAAGCTCCTCCCT

 CCTGTGAGAACAAGCCCGCAGACCGACCCGCTCCCGCGGAGGCGGAGGTGAAGGCCCTGG
 -----+-----+-----+-----+-----+ 2280
 GGACACTCTTGTTCGGGCGTCTGGCTGGGCGAGGGCGCCTCCGGCTCCACTTCCGGGACC
 30 V K A L V -
 -----+-----+-----+-----+-----+
 acbC -----

 P
 v
 u
 35 I
 I

 TCCTGGCAGGTGGAACCGGCAGCAGACTGAGGCCGTTACCCACACCGCCGCCAAGCAGC
 -----+-----+-----+-----+-----+ 2340
 AGGACCGTCCACCTTGGCCGTCGTCTGACTCCGGCAAGTGGGTGTGGCGGCGGTTGCTCG
 40 L A G G T G S R L R P F T H T A A K Q L -

 TGCTCCCCATCGCCAACAAGCCCGTGCTCTTCTACGCGCTGGAGTCCCTCGCCGCGGCGG
 -----+-----+-----+-----+-----+ 2400
 ACGAGGGGTAGCGGTTGTTTCGGGCACGAGAAGATGCCGACCTCAGGGAGCGGCGCCGCC
 45 L P I A N K P V L F Y A L E S L A A A G -

 GTGTCCGGGAGGCCGGCGTCTGTCGTGGGCGCGTACGGCCGGGAGATCCGCGAACTCACCG
 -----+-----+-----+-----+-----+ 2460
 CACAGGCCCTCCGGCCCGCAGCAGCACCCGCGCATGCCGGCCCTCTAGGCGCTTGTAGTGGC
 V R E A G V V V G A Y G R E I R E L T G -

 GCGACGGCACCGCGTTCCGGTTACGCATCACCTACCTCCACCAGCCCCGCCCGCTCGGTC
 -----+-----+-----+-----+-----+ 2520
 CGCTGCCGTGGCGCAAGCCCAATGCGTAGTGGATGGAGGTGGTTCGGGGCGGGCGAGCCAG
 D G T A F G L R I T Y L H Q P R P L G L -

 TCGCGCACGCGGTGCGCATCGCCCGCGGCTTCCTGGGCGACGACGACTTCCTGCTGTACC
 -----+-----+-----+-----+-----+ 2580
 AGCGCGTGCGCCACGCGTAGCGGGCGCCGAAGGACCCGCTGCTGCTGAAGGACGACATGG
 A H A V R I A R G F L G D D D F L L Y L -

ES 2 276 427 T3

TGGGGGACAACTACCTGCCCCAGGGCGTCACCGACTTCGCCCCCAATCGGGCCCGGATC
 -----+-----+-----+-----+-----+ 2640
 ACCCCCTGTTGATGGACGGGGTCCCGCAGTGGCTGAAGCGGGCGGTTAGCCGGCGGCTAG
 G D N Y L P Q G V T D F A R Q S A A D P -
 5
 CCGCGGCGGGCCCGGCTGCTGCTCAGCCCGGTGCGGGACCGTCCGCCTTCGGCGTCCGGG
 -----+-----+-----+-----+-----+ 2700
 GGCGCCCGGGCCGACGACGAGTGGGGCCAGCGCCTGGGCAGGCGGAAGCCGCAGCGCC
 A A A R L L L T P V A D P S A F G V A E -
 10
 AGGTGACGCGGACGGGAACGTGCTGCGCTTGGAGGAGAAACCGACGTCCCGCGCAGCT
 -----+-----+-----+-----+-----+ 2760
 TCCAGCTGCGCCTGCCCTTGCACGACGGAACCTCCTCTTTGGGCTGCAGGGCGGCTCGA
 V D A D G N V L R L E E K P D V P R S S -
 15
 CGCTCGCGCTCATCGGCGTGTACGCCTTCAGCCCGGCGCTCCACGAGGCGGTACGGGCCA
 -----+-----+-----+-----+-----+ 2820
 GCGAGCGGAGTAGCCGCACATGCGGAAGTGGGGCCGGCAGGTGCTCCGCCATGCCCGGT
 L A L I G V Y A F S P A V H E A V R A I -
 20
 TCACCCCTCGCCCCGCGGAGCTGGAGATACCCACGCCGTGCAGTGGATGATCGACC
 -----+-----+-----+-----+-----+ 2880
 AGTGGGGGAGCGGGCGCGCTCGACCTCTAGTGGGTGCGGCACGTACCTACTAGCTGG
 T P S A R G E L E I T H A V Q W M I D R -
 25
 GGGGCCTGCGCGTACGGGCGGAGACCACCACCGGCCCTGGCGCGACACCGGCAGCGCGG
 -----+-----+-----+-----+-----+ 2940
 CCGCGACGCGCATGCCCGGCTCTGGTGGTGGGCGGGACCGCGCTGTGGCCGTGCGGCC
 G L R V R A E T T T R P W R D T G S A E -
 30
 AGGACATGCTGGAGGTCAACCGTCACGTCTTGACGGACTGGAGGGCCGCATCGAGGGGA
 -----+-----+-----+-----+-----+ 3000
 TCCTGTACGACCTCCAGTTGGCAGTGCAGGACCTGCCTGACCTCCCGGCTAGCTCCCT
 D M L E V N R H V L D G L E G R I E G K -
 35
 AGGTGACGCGCACAGCACGCTGGTGGGCCGGGTGGTGTATCGGCGCGGGTCCGTCTCAGCAACTCCA
 -----+-----+-----+-----+-----+ 3060
 TCCAGCTGCGCGTGTGCTGCGACCGCGGCCAGGCCACCGGCTTCGCGCTAGCAGG
 V D A H S T L V G R V R V A E G A I V R -
 40
 GGGGGTCACACGTGGTGGGCCCGGTGGTGTATCGGCGCGGGTCCGTCTCAGCAACTCCA
 -----+-----+-----+-----+-----+ 3120
 CCGCGAGTGTGACACCGCGGCCACCACTAGCCGCGCCACGGCAGCAGTCTGTTAGGT
 G S H V V G P V V I G A G A V V S N S S -
 45
 GTGTCGGCCCGTACACCTCCATCGGGGAGGACTGCCGGTTCGAGGACAGCGCCATCGACT
 -----+-----+-----+-----+-----+ 3180
 CACAGCCGGGCATGTGGAGGTAGCCCTCCTGACGGCCAGCTCCTGTGCGGGTAGCTCA
 V G P Y T S I G E D C R V E D S A I E Y -
 50
 ACTCCGTCTGTGTCGCGCGGCCAGGTTCGAGGGGGCGTCCCGCATCGAGGCGTCCCTCA
 -----+-----+-----+-----+-----+ 3240
 TGAGGCAGGACGACGCGCGCGGTCCAGCTCCCCGCGGGCGTAGCTCCGCAGGGAGT
 S V L L R G A Q V E G A S R I E A S L I -
 55
 TCGGCCGCGGCGCGCTCGTGGGCCGGGCCCGCTCTCCCGCAGGCTACCGACTGGTGA
 -----+-----+-----+-----+-----+ 3300
 AGCCGCGCGCGCGGACGAGCCGGGGCGGGGGCAGAGGGCGTCCGAGTGGCTGACCACT
 G R G A V V G P A P R L P Q A H R L V I -
 60
 65

ES 2 276 427 T3

TCGGCGACCACAGCAAGGTGTATCTCACCCCATGACCACGACCATCCTCGTCACCGGCGG
 -----+-----+-----+-----+-----+ 3360
 AGCCGCTGGTGTCTGTTCCACATAGAGTGGGGTACTGGTGTGGTAGGAGCAGTGGCCCGCC

G D H S K V Y L T P * M T T T I L V T G G -
 acbD -----

AGCGGGCTTCATTGCTCCGCCTACGTCCGCCGGCTCCTGTGCGCCGGGGCCCCCGGCGG
 -----+-----+-----+-----+-----+ 3420
 TCGCCCGAAGTAAGCGAGGCGGATGCAGGCGGCCGAGGACAGCGGGCCCCGGGGCCGCC
 A G F I R S A Y V R R L L S P G A P G G -
 CGTCGCGGTGACCGTCTCGACAACTCACCTACGCCGGCAGCCTCGCCCGCCTGCACGC
 -----+-----+-----+-----+-----+ 3480
 GCAGCGCCACTGGCAGGAGCTGTTTGAAGTGGATGCGGCCGTGCGAGCGGGCGGACGTGCG
 V A V T V L D K L T Y A G S L A R L H A -
 GGTGCGTGACCATCCCGGCCTCACCTTCGTCCAGGGCGACGTGTGCGACACCGCGCTCGT
 -----+-----+-----+-----+-----+ 3540
 CCACGCACTGGTAGGGCCGGAGTGAAGCAGGTCCCGCTGCACACGCTGTGGCGCGAGCA
 V R D H P G L T F V Q G D V C D T A L V -
 CGACACGCTGGCCGCGCGGCACGACGACATCGTGCACTTCGCGGGCCGAGTCGCACGTCGA
 -----+-----+-----+-----+-----+ 3600
 GCTGTGCGACCGGCGCGCCGTGCTGCTGTAGCACGTGAAGCGCCGGCTCAGCGTGCAGCT
 D T L A A R H D D I V H F A A E S H V D -
 CCGCTCCATCACCGACAGCGGTGCCTTCACCCGCACCAACGTGCTGGGCACCCAGGTCCT
 -----+-----+-----+-----+-----+ 3660
 GGCGAGGTAGTGGCTGTGCGCCACGGAAGTGGGCGTGGTTGCACGACCCGTGGGTCCAGGA
 R S I T D S G A F T R T N V L G T Q V L -
 GCTCGACGCGCGCTCCGCCACGGTGTGCGCACCTTCGTGCACGTCTCCACCGACGAGGT
 -----+-----+-----+-----+-----+ 3720
 CGAGCTGCGGCGCGAGGCGGTGCCACACGCGTGAAGCACGTGCAGAGGTGGCTGCTCCA
 L D A A L R H G V R T F V H V S T D E V -
 GTACGGCTCCCTCCCGCACGGGGCCGCGCGGAGAGCGACCCCTGCTTCCGACCTCGCC
 -----+-----+-----+-----+-----+ 3780
 CATGCCGAGGGAGGGCGTCCCCGGCGGCGCCTCTCGCTGGGGGACGAAGGCTGGAGCGG
 Y G S L P H G A A A E S D P L L P T S P -
 GTACGCGGCGTCAAGGCGGCCTCGGACCTCATGGCGCTCGCCCACCACCGCACCCACGG
 -----+-----+-----+-----+-----+ 3840
 CATGCGCCGACGCTTCCGCCGAGCCTGGAGTACCGCGAGCGGGTGGTGGCGTGGGTGCC
 Y A A S K A A S D L M A L A H H R T H G -
 CCTGGACGTCCGGGTGACCCGCTGTTTGAACAACTTCGGCCCCCACCAGCATCCCGAGAA
 -----+-----+-----+-----+-----+ 3900
 GGACCTGCAGGCCCCACTGGGCGACAAGCTTGTGAAGCCGGGGGTGGTCTAGGGCTCTT
 L D V R V T R C S N N F G P H Q H P E K -
 GCTCATACCGCGCTTCCCTGACCAGCCTCCTGTCCGGCGGCACCGTTCCCTCTACGGCGA
 -----+-----+-----+-----+-----+ 3960
 CGAGTATGCGCGAAGGACTGGTCCGAGGACAGGCCCGCTGGCAAGGGGAGATGCCGCT
 L I P R F L T S L L S G G T V P L Y G D -

ES 2 276 427 T3

CGGGCGGCACGTGCGCGACTGGCTGCACGTCGACGACCCACGTGAGGGCCGTCGAACTCGT
-----+-----+-----+-----+-----+ 4020
GCGCGCGGTGCACGCGTGCACGCGTGCAGCTGCTGGTGCAGTCCCGGCAGCTTGAGCA
G R H V R D W L H V D D H V R A V E L V -

B
9
1
I
I

CCGCGTGTGCGGGCCGGCCGGGAGAGATCTACAACATCGGGGGCGGCACCTCGCTGCCAA
-----+-----+-----+-----+ 4080
GGCGCACAGCCCGCGCGGCCCTCTCTAGATGTTGTAGCCCCGCGCGTGGAGCGACGGGTT
R V S G R P G E I Y N I G G G T S L P N -

S
s
t
I

CCTGGAGCTCACGCACCGGTTGCTCGCACTGTGCGGCGCGGGCCCGAGCGCATCGTCCA
-----+-----+-----+-----+ 4140
GGACCTCGAGTTCGTGGCCAACGAGCGTGACACGCGCGCCCGGGCCCTCGCGTAGCAGGT
L E L T H R L L A L C G A G P E R I V H -

CGTCGAGAACCAGGAGGGGACGACCGGCGTACGCGGTGACCCAGCAAGATCACCGC
-----+-----+-----+-----+ 4200
GCAGCTCTTGGCGTCCCCGTGCTGGCCGCGATGCGCCAGCTGGTGTCTAGTGGCG
V E N R K G H D R R Y A V D H S K I T A -

N
F
u
I

GGAACCTCGGTTACCGGCCGCGCACCGACTTCGCGACCGCGCTGGCCGACACCGCGAAGTG
-----+-----+-----+-----+ 4260
CCTTGAGCCAATGGCCGGCGCGTGGCTGAAGCGCTGGCGCGACCGGCTGTGGCGCTTCAC
E L G Y R P R T D F A T A L A D T A K W -

GTACGAGCGGCACGAGGACTGGTGGCGTCCCTGCTCGCCGCGACATGACGTGCGGCCGG
-----+-----+-----+-----+ 4320
CATGCTCGCCGTGCTCCTGACCACCGCAGGGGACGAGCGGCGCTGTACTGCAGCCCGGCC
Y E R H E D W W R P L L A A T *

ACCGCAACCACCGGCCCCGGCCGGCACACCGCCGCCCGCGGCGGTGGCCGGCCGGTCAG
-----+-----+-----+-----+ 4380
TGGCGTTGGTGGCCGGGGCCGGCCGTGTGGCGGCGGGCGCCGCCACCGGCCGGCCAGTC
* -

CGTCCGTGAGCCGGGCGCCGGCCGCCCGCGGGCCGGCGGCGGTGGACCCCCGGACCACC
-----+-----+-----+-----+ 4440
GCAGGCACTCGGCCCGCGGCCGGCGGGGCGCCCGGCCGCCACCTGGGGGCCCTGGTGG
R G H A P R R G G R P G A A T S G R V V -

E
c
o
r
I

AGTTCGGGCATGAAGACGAATTCGGTGCAGCGCGCGGCGGCGTTCGCTCATCTCCTCCAGC
-----+-----+-----+-----+ 4500
TCAAGGCCGTACTTCTGCTTAAGCCACGCGCCCGCCCGCAAGGCGAGTAGAGGAGGTGC
L E P M F V F E T R P P P T G S M E E L -

ES 2 276 427 T3

AGTGCCTCCACGGCGACCTGCCCCATCGCCTTGACGGGCTGTCTGATGGTGGTCAGGGGA
 -----+-----+-----+-----+-----+ 4560
 TCACGCAGGTGCCGCTGGACGGGTAGCGGAAGTCCCCGACAGACTACCACCAAGTCCCCT
 5 L A D V A V Q G M A K V P Q R I T T L P -

 GGGTCGGTGAAGGCCATGAGCGGCGAGTCGTGGAAGCCGACCACCGAGATGTCACCGGGA
 -----+-----+-----+-----+-----+ 4620
 CCCAGCCACTTCCGGTACTCGCCGCTCAGCAGCTTCGGCTGGTGGCTCTACAGTGGCCCT
 10 P D T F A M L P S D D F G V V S I D G P -

 ACCGTGAGACCCCGCCGGCGCGCGCCCGCACGGCGCCGAGGGCCATCATGTCGCTGGCG
 -----+-----+-----+-----+-----+ 4680
 TGGCACTCTGGGGCGGCCGCGCGCGGGCGTGCCTCGGCTCCCGGTAGTACAGCGACCGC
 15 V T L G R R R A A R V A G L A M M D S A -

 CACATGACGGCGGTGCAGCCAGGTGATCAGCGCGGACGCGGCGGCTGGCCCCCTCC
 -----+-----+-----+-----+-----+ 4740
 GTGTAAGTCCCGCCAGCTCGGGTCCAGCTAGTCGCGCCTGCGCCGCGGACCGGGGGGAGG
 20 C M V A T C G L D I L A S A A A Q G G E -

 S
 s
 t
 I
 AGGGAGAACAGCGAGTGTGTCAGAGCTCCTCGGACTCCCGCGCCGACACTCCCAGGTGC
 -----+-----+-----+-----+-----+ 4800
 TCCCTCTTGTGCTCAGGACGTGCTCGAGGAGCCTGAGGGCGCGGCTGTGAGGGTCCACG
 25 L S F L S H Q V L E E S E R A S V G L H -

 TCCCGCACGCCGCGCCCGGAACCCCTCGATCTTCCGCTGCACCGGCACGAAGCGGGCGGGC
 -----+-----+-----+-----+-----+ 4860
 AGGGCGTGCGGCGGGCCCTTGGGGAGCTAGAAGGCGACGTGGCCGTGCTTCGCCCGCCCG
 30 E R V G A R F G E I K R Q V P V F R A P -

 CCGACGGCGAGGCCGACGCGCTCGTGCCCCAGCTCCGCCAGGTGCGCCACGGCCAGGCGC
 -----+-----+-----+-----+-----+ 4920
 GGCTGCCGCTCCGGCTGCGGAGCAGGGGTGAGGCGGTCCACGCGGTGCGCGTCCGCG
 35 G V A L G V R E H G L E A L H A V A L R -

 ATCGCGGCGCGGTGCTCCGGGGAGACGAAGGGTGCCTCGATCCGGGGCGAGAACCCGTTT
 -----+-----+-----+-----+-----+ 4980
 TAGCGCCGGGCCAGCAGGCCCCCTGCTTCCACGGAGCTAGGCCCGCTCTTGGGCAAG
 40 M A A R D D P S V F P A E I R P S F G N -

 ACGAGGACGAAGGGCACCTGCCGCTCGTGACCGGCGCGTACCGTCCGGTCTCGGCGGTG
 -----+-----+-----+-----+-----+ 5040
 TGCTCCTGCTTCCCGTGGACGGCGAGCACGTGCGCCGGCATGGCAGGCCAGAGCCGCCAC
 45 V L V F P V Q R E H L R G Y R G T E A T -

 GTGTCCGCGTGCAGTCCGGAGACGAAGATGATGCCGGACACCCCGCGGTCCACGAGCATC
 -----+-----+-----+-----+-----+ 5100
 CACAGGCGCACGTGAGGCTCTGCTTCTACTACGGCCTGTGGGGCGCCAGGTGCTCGTAG
 50 T D A H L G S V F I I G S V G R D V L M -

 S
 m
 a
 I
 TCCGTGAGTTCGTCTCGGTGAGCCGCCCGGGGTCTGCGTGGCGAGCACGGGCGTGTAG
 -----+-----+-----+-----+-----+ 5160
 AGGCACTCAAGCAGGAGCCAGCTCGGCGGGCCCCAGACGCACCGCTCGTGCCCGCACATC
 55 E T L E D E T S G G P T Q T A L V P T Y -

 60

 65

ES 2 276 427 T3

CCCTGACGCGTGAGCGCCTGCCCCATCACCTGGGCCAGTGCGGGGAAGAAGGGGTTGTCC
 -----+-----+-----+-----+-----+-----+-----+ 5220
 GGGACTGCGCACTCGCGGACGGGGTAGTGGACCCGGTCACGCCCTTCTTCCCAACAGG
 G Q R T L A Q G M V Q A L A P F F P N D -
 AGTTCGGGGGTGACCAGTCCGACCAGCTCGGCGCGGCGCTGTCGCGCCGGCTGCTCGTAG
 -----+-----+-----+-----+-----+-----+-----+ 5280
 TCAAGCCCCACTGGTCAGGCTGGTCAGCCGCGCGCCGACAGCGCGCCGACGAGCATC
 L E P T V L G V L E A R R Q R A P Q E Y -
 CCCAGCGCGTCCAGTGCGGTGAGCACCAGTTCGCGGGTGCCGGTGCCACACCGCGCGCA
 -----+-----+-----+-----+-----+-----+-----+ 5340
 GGGTCGCGCAGGTACGCCAGTCGTGGCTCAGCGCCACGGCCACCGGTGTGGCGCGCT
 G L A D L A T L V S D R T G T A V G R A -
 S
 m
 a
 I
 CCGTTCAGCACCCGGCTGACCGTGGCCTTGCTGACGCCCGCCGGGCTGCGATGTGCGCG
 -----+-----+-----+-----+-----+-----+-----+ 5400
 GCGAAGTCGTGGGCGGACTGGCACCGAACGACTGCGGGCGGGCCGACGCTACAGCCCG
 G N L V R S V T A K S V G A R A A I D A -
 AGCCGCATGGTCATGGCAACGCACTCTACCTGTGCGGGCGTCAAGGGCGTGCCACCGCGC
 -----+-----+-----+-----+-----+-----+-----+ 5460
 TCGGCGTACGAGTACCGTTGCGTGAGATGACAGCCCCGAGTCCCGCACGGGTGGCGCG
 L R M T M
 acbE
 GCGGAACCGCGGACTGCGGGGCACGGCCCGTCCGCCGCCACGGACCACGCGCCGAAA
 -----+-----+-----+-----+-----+-----+-----+ 5520
 CGCCTTGCCCGCCTGACGCCCGTGCCGGGCAGGCGGCGGGTGCTTGTGCGCGGGCTTT
 CGATGGCTGAAAATGCTTGACGCAAAATGCCGCAACGTCTTTTCGGCGGCTTTTCGATCCT
 -----+-----+-----+-----+-----+-----+-----+ 5580
 GCTACCGACTTTTACGAACGTGTTTTAACGGCGTTGACAGAAAGCCGCCGAAAAGCTAGGA
 GTTACGTTCTTGCAACCCCGGCGCGCGCAGAAAGCGGTTGGCGTGAGGCGTCCAGACCT
 -----+-----+-----+-----+-----+-----+-----+ 5640
 CAATGCAAGGACCGTTGGGGCCGCGGCGCTTTCGCCAACCGCACTCCGCAGGTCTGGA
 CCGCCCGATTCCGGGATCACTCAGGGGAGTTACAATGCGGCGTGGCATTGCGGCCACCG
 -----+-----+-----+-----+-----+-----+-----+ 5700
 GCGGGGCTAAGGCCCTAGTGAGTCCCCTCAAGTGTTACGCCGACCGTAACGCCGGTGGC
 M R R G I A A T A -
 acbF
 CGCTGTTTCGCGGCTGTGGCCATGACGGCATCGGCGTGTGGCGGGGGCGACAACGGCGGAA
 -----+-----+-----+-----+-----+-----+-----+ 5760
 GCGACAAGCGCCGACACCGGTACTGCCGTAGCCGACACCGCCCCGCTGTTGCCGCCTT
 L F A A V A M T A S A C G G G D N G G S -
 K
 P
 n
 I
 GCGGTACCGACGCGGGCGGCACGGAGCTGTGCGGGACCGTCACCTTCTGGGACACGTCCA
 -----+-----+-----+-----+-----+-----+-----+ 5820
 CGCCATGGCTGCGCCCGCCGTCCTCGACAGCCCTGGCAGTGGAAGACCCTGTGCAGGT
 G T D A G G T E L S G T V T F W D T S N -

ES 2 276 427 T3

5 ACGAAGCCGAGAAGGCGACGTACCAGGCCCTCGCGGAGGGCTTCGAGAAGGAGCACCCGA
 -----+-----+-----+-----+-----+ 5880
 TGCTTCGGCTCTTCCGCTGCATGGTCCGGGAGCGCCTCCCGAAGCTCTTCCTCGTGGGCT
 E A E K A T Y Q A L A E G F E K E H P K -
 10 AGGTGACGTCAAGTACGTCAACGTCCCCTTCGGCGAGGCGAACGCCAAGTTCAAGAACG
 -----+-----+-----+-----+-----+ 5940
 TCCAGCTGCAGTTCATGCAGTTGCAGGGCAAGCCGCTCCGCTTGCAGTTCAAGTTCTTGC
 V D V K Y V N V P F G E A N A K F K N A -
 15 CCGCGGGCGGCAACTCCGGTGCCCCGGACGTGATGCGTACGGAGGTGCGCTGGGTGCGGG
 -----+-----+-----+-----+-----+ 6000
 GCGCGCCCGCGTTGAGGCCACGGGGCCTGCACTACGCATGCCTCCAGCGGACCCAGCGCC
 A G G N S G A P D V M R T E V A W V A D -
 20 ACTTCGCCAGCATCGGCTACCTCGCCCCGCTCGACGGCACGCCCGCCCTCGACGACGGGT
 -----+-----+-----+-----+-----+ 6060
 TGAAGCGGTCTAGCCGATGGAGCGGGCGAGCTGCCGTGCGGGCGGGAGCTGCTGCCCA
 F A S I G Y L A P L D G T P A L D D G S -
 25 CGGACCACCTTCCCCAGGGCGGCGAGCACCAGGTACGAGGGGAAGACCTACGCGGTCCCGC
 -----+-----+-----+-----+-----+ 6120
 GCCTGGTGGAAAGGGTCCCGCCGTCTGGTCCATGCTCCCCTTCTGGATGCGCCAGGGCG
 D H L P Q G G S T R Y E G K T Y A V P Q -
 30 AGGTGATCGACACCCTGGCGCTCTTCTACAACAAGGAAGTCTGACGAAGGCCGGTGTCTG
 -----+-----+-----+-----+-----+ 6180
 TCCACTAGCTGTGGGACCGCGAGAAGATGTTGTTCTTGACGACTGCTTCCGGCCACAGC
 V I D T L A L F Y N K E L L T K A G V E -
 35 AGGTGCCGGGCTCCCTCGCCGAGCTGAAGACGGCCCGCCGAGATCACCGAGAAGACCG
 -----+-----+-----+-----+-----+ 6240
 TCCACGGCCCGAGGGAGCGGCTCGACTTCTGCGGGCGGCGGCTCTAGTGGCTCTTCTGGC
 V P G S L A E L K T A A A E I T E K T G -
 40 GCGCGAGCGGCTCTACTGCGGGGCGACGACCCGTAAGTCTGTTCTGCGCTACCTCTACG
 -----+-----+-----+-----+-----+ 6300
 CGCGCTCGCCGAGATGACGCCCGCTGCTGGGCATGAACCAAGGACGGGATGGAGATGC
 A S G L Y C G A T T R T W F L P Y L Y G -
 45 GGGAGGGCGGCGACCTGGTTCGACGAGAAGAACAAGACCGTCACGGTCGACGACGAAGCCG
 -----+-----+-----+-----+-----+ 6360
 CCCTCCCGCCGCTGGACCAGCTGCTCTTCTGTTCTGGCAGTGGCAGCTGCTGCTTCGGC
 E G G D L V D E K N K T V T V D D E A G -
 50 GTGTGCGGCCTACCGCGTCATCAAGACCTCGTGGACAGCAAGGCGGCCATCACCGACG
 -----+-----+-----+-----+-----+ 6420
 CACACGCGCGGATGGCGCAGTAGTTCTTGGAGCACCTGTCTGTTCCGCCGGTAGTGGCTGC
 V R A Y R V I K D L V D S K A A I T D A -
 55 CGTCCGACGGCTGGAACAACATGCAGAAGCCCTCAAGTCGGGCAAGGTGCGCATGATGG
 -----+-----+-----+-----+-----+ 6480
 GCAGGCTGCCGACCTTGTGTACGTCTTGCGGAAGTTCAAGCCGTTCCAGCGGTACTACC
 S D G W N N M Q N A F K S G K V A M M V -
 60 TCAACGGCCCCCTGGGCCATCGAGGACGTCAAGCGGGAGCCCGCTCAAGGACGCCGGCA
 -----+-----+-----+-----+-----+ 6540
 AGTTGCCGGGGACCCGGTAGCTCCTGCAGTTCCGCCCTCGGGCGAAGTTCCTGCGGCCGT
 N G P W A I E D V K A G A R F K D A G N -
 65

ES 2 276 427 T3

ACCTGGGGGTGCGCCCCGTCCCGGCCGGCAGTGCCGGACAGGGCTCTCCCCAGGGCGGGT
 -----+-----+-----+-----+-----+ 6600
 TGGACCCCCAGCGGGGGCAGGGCCGGCCGTACGGCCCTGTCCCGAGAGGGGTCCCGCCCA
 5 L G V A P V P A G S A G Q G S P Q G G W -

 GGAACCTCTCGGTGTACGCGGGCTCGAAGAACCTCGACGCCTCCTACGCCTTCGTGAAGT
 -----+-----+-----+-----+-----+ 6660
 CCTTGGAGAGCCACATGCGCCCGAGCTTCTTGGAGCTGCGGAGGATGCGGAAGCACTTCA
 10 N L S V Y A G S K N L D A S Y A F V K Y -

 S
 S
 t
 I
 15 ACATGAGCTCCGCCAAGGTGCAGCAGCAGACCACCGAGAAGCTGAGCCTGCTGCCACCC
 -----+-----+-----+-----+-----+ 6720
 TGTACTCGAGGCGGTTCCACGTCGTCGTCTGGTGGCTCTTCGACTCGGACGACGGGTGGG
 M S S A K V Q Q Q T T E K L S L L P T R -

 GCACGTCCGTCTACGAGGTCCCGTCCGTGCGGACAACGAGATGGTGAAGTTCTTCAAGC
 -----+-----+-----+-----+-----+ 6780
 CGTGCAGGCAGATGCTCCAGGGCAGGCAGCGCCTGTTGCTCTACCACTTCAAGAAGTTTCG
 T S V Y E V P S V A D N E M V K F F K P -

 CGGCCGTGACAAGGCCGTGGAACGCCGTGGATCGCCGAGGGCAATGCCCTCTTCGAGC
 -----+-----+-----+-----+-----+ 6840
 GCCGGCAGCTGTTCCGGCAGCTTGCCGGCACCTAGCGGCTCCCGTTACGGGAGAAGCTCG
 A V D K A V E R P W I A E G N A L F E P -

 P
 S
 t
 I
 30 CGATCCGGCTGCAG
 -----+----- 6854
 35 GCTAGGCCGACGTC
 I R L Q -

 40

 45

 50

 55

 60

 65

REIVINDICACIONES

1. Molécula de ADN que comprende los genes para la biosíntesis de acarbosa, **caracterizada** porque

(a) comprende una secuencia de ADN según la Tabla 4; o

(b) comprende una secuencia de ADN que es capaz de hibridarse, bajo condiciones rigurosas, con la molécula de ADN según (a), o

(c) contiene una secuencia de ADN que, debido a la degeneración del código genético, difiere de las moléculas de ADN según (a) y (b), pero permite la expresión de las proteínas que se pueden expresar que se corresponden con las moléculas de ADN según (a) y (b).

2. Molécula de ADN según la reivindicación 1, **caracterizada** porque comprende como secuencia mencionada en (a), la secuencia de ADN de los nucleótidos 1 a 720 (gen acbA) de acuerdo con la Tabla 4.

3. Molécula de ADN según la reivindicación 1, **caracterizada** porque comprende como secuencia mencionada en (a), la secuencia de ADN de los nucleótidos 720 a 2006 (gen acbB) de acuerdo con la Tabla 4.

4. Molécula de ADN según la reivindicación 1, **caracterizada** porque comprende como secuencia mencionada en (a), la secuencia de ADN de los nucleótidos 2268 a 3332 (gen acbC) de acuerdo con la Tabla 4.

5. Molécula de ADN según la reivindicación 1, **caracterizada** porque comprende como secuencia mencionada en (a), la secuencia de ADN de los nucleótidos 3332 a 4306 (gen acbD) de acuerdo con la Tabla 4.

6. Molécula de ADN según la reivindicación 1, **caracterizada** porque comprende como secuencia mencionada en (a), la secuencia de ADN de los nucleótidos 4380 a 5414 (gen acbE) de acuerdo con la Tabla 4.

7. Molécula de ADN según la reivindicación 1, **caracterizada** porque comprende como secuencia mencionado en (a), la secuencia de ADN de los nucleótidos 5676 a 6854 (gen acbF) de acuerdo con la Tabla 4.

8. Cebador oligonucleótido para la amplificación con PCR de la molécula de ADN según la reivindicación 1.

9. Vector que contiene una molécula de ADN según una o varias de las reivindicaciones 1 a 8.

10. Vector según la reivindicación 9, para uso en un procedimiento de inactivación o de modificación de los genes naturales para la biosíntesis de acarbosa en un microorganismo productor de acarbosa.

11. Vector según la reivindicación 10, **caracterizado** porque se selecciona a partir del grupo consistente en pGM160 o vectores relacionados.

12. Vector según la reivindicación 9, **caracterizado** porque es un vector de expresión y dicha molécula de ADN está ligada funcionalmente con una secuencia promotora.

13. Vector según la reivindicación 12, que es adecuado para la expresión en organismos hospedadores que se seleccionan entre el grupo consistente en *E. coli*, *Bacillus subtilis*, cepas de *Streptomyces*, *Actinoplanes*, *Ampullariella*, *Streptosporangium*, *Streptomyces hygroscopicus* var. *limoneus*, *Streptomyces glaucescens*, así como *Aspergillus niger*, *Penicillium chrysogenum* y *Saccharomyces cerevisiae*.

14. Vector según la reivindicación 12, que es adecuado para la expresión en *Streptomyces glaucescens* (DSM 40716) o en *Actinoplanes* sp.

15. Célula hospedadora que se transforma con un vector según una de las reivindicaciones 9 a 14.

16. Célula hospedadora según la reivindicación 15, **caracterizada** porque se selecciona a partir del grupo consistente en *E. coli*, *Bacillus subtilis*, las cepas de *Streptomyces*, *Actinoplanes*, *Ampullariella*, *Streptosporangium*, *Streptomyces hygroscopicus* var. *limoneus*, *Streptomyces glaucescens*, así como *Aspergillus niger*, *Penicillium chrysogenum* y *Saccharomyces cerevisiae*.

17. Célula hospedadora según la reivindicación 16, **caracterizada** porque se selecciona a partir del grupo consistente en *Streptomyces glaucescens* (DSM 40716) y *Actinoplanes* sp.

18. Proteína (producto del gen acbA), codificada por un ADN según la reivindicación 2.

19. Proteína (producto del gen acbB), codificada por un ADN según la reivindicación 3.

20. Proteína (producto del gen acbC), codificada por un ADN según la reivindicación 4.

21. Proteína (producto del gen *acbD*), codificada por un ADN según la reivindicación 5.

22. Proteína (producto del gen *acbE*), codificada por un ADN según la reivindicación 6.

23. Proteína (producto del gen *acbF*), codificada por un ADN según la reivindicación 7.

24. Procedimiento para obtener las proteínas según una de las reivindicaciones 18 bis 23, **caracterizado** porque

(a) las proteínas se expresan en una célula hospedadora adecuada y

(b) se aíslan.

25. Procedimiento según la reivindicación 24, **caracterizado** porque la célula hospedadora se selecciona a partir del grupo consistente en *E. coli*, *Bacillus subtilis*, las cepas de *Streptomyces*, *Actinoplanes*, *Ampullariella*, *Streptosporangium*, *Streptomyces hygroscopicus* var. *limoneus*, *Streptomyces glaucescens*, así como *Aspergillus niger*, *Penicillium chrysogenum* y *Saccharomyces cerevisiae*.

26. Procedimiento según la reivindicación 24, **caracterizado** porque la célula hospedadora se selecciona a partir del grupo consistente en *Streptomyces glaucescens* (DSM 40716) y *Actinoplanes* sp.

27. Procedimiento para la preparación de acarbosa, **caracterizado** porque

(a) uno o varios genes según una o varias de las reivindicaciones 2 a 7, se emplean para la expresión en una célula hospedadora adecuada y

(b) la acarbosa se aísla a partir del material sobrenadante del cultivo de dicha célula hospedadora.

28. Procedimiento para la preparación de acarbosa según la reivindicación 27, **caracterizado** porque las células hospedadoras se seleccionan según una de las reivindicaciones 16 o 17.

29. Procedimiento para la preparación de acarbosa, **caracterizado** porque

(a) uno o varios genes según una o varias de las reivindicaciones 2 a 7, se inactivan en una célula hospedadora productora natural de acarbosa y

(b) se aísla la acarbosa a partir de dicha célula hospedadora.

30. Procedimiento para la preparación de acarbosa según la reivindicación 29, **caracterizado** porque se seleccionan las células hospedadoras según la reivindicación 17.

31. Procedimiento para la preparación de acarbosa, **caracterizado** porque un procedimiento según una de las reivindicaciones 27 a 28, se combina con un procedimiento según una de las reivindicaciones 29 a 30.

32. Procedimiento para completar la agrupación génica para biosintetizar acarbosa según la reivindicación 1, **caracterizado** porque

(a) las regiones colindantes de ADN genómico se aíslan por métodos de hibridación empleando sondas de hibridación que se hibridan, bajo condiciones rigurosas, con la molécula de ADN según la reivindicación 1 y

(b) se secuencian.

33. Procedimiento para completar la agrupación génica para la biosíntesis de acarbosa según la reivindicación 1, **caracterizado** porque

a) se aíslan las regiones de ADN genómico adyacentes mediante PCR, empleando cebadores de PCR que se hibridan, bajo condiciones rigurosas, con las secuencias de ADN procedentes de la molécula de ADN según la reivindicación 1 y los cebadores que poseen una secuencia que permite la hibridación con secuencias del sistema del vector empleado y

b) se secuencian.

34. Procedimiento para aislar una agrupación génica para la biosíntesis de acarbosa y de pseudo-oligosacáridos homólogos procedentes de microorganismos productores de acarbosa distintos de *Streptomyces glaucescens* (DSM 40716), **caracterizado** porque

a) se preparan las sondas de la hibridación que se hibridan, bajo condiciones rigurosas, con la molécula de ADN recombinante según la reivindicación 1,

b) estas sondas de hibridación se emplean para el escrutinio del ADN genómico o del ADNc que se ha obtenido a partir del microorganismo correspondiente, y

c) los clones encontrados se aíslan y se caracterizan.

35. Procedimiento para aislar una agrupación génica para la biosíntesis de acarbosa y de pseudo-oligosacáridos homólogos procedentes de microorganismos productores de acarbosa, distintos de *Streptomyces glaucescens* (DSM 40716), **caracterizado** porque

a) se preparan los cebadores de la PCR que se hibridan, bajo condiciones rigurosas, con la molécula de ADN recombinante según la reivindicación 1,

b) estos cebadores para la PCR se emplean para acumular fragmentos de ADN genómico o del ADNc de un microorganismo correspondiente,

c) los fragmentos acumulados se aíslan y se caracterizan y

d) eventualmente se emplean para un procedimiento según la reivindicación 34.

36. Procedimiento según las reivindicaciones 34 o 35, **caracterizado** porque los microorganismos se seleccionan a partir del grupo consistente en *Actinomycetales*, las cepas de *Streptomyces*, *Actinoplanes*, *Ampullariella*, *Streptosporangium*, *Streptomyces hygrosopicus* var. *limoneus*, *Streptomyces glaucescens*.

37. Procedimiento según la reivindicación 36, **caracterizado** porque los microorganismos se seleccionan, en particular, a partir del grupo consistente en *Streptomyces glaucescens* (DSM 40716) y *Actinoplanes* sp.

38. Uso de *Streptomyces glaucescens* (DSM 40716) para obtener acarbosa.

39. Procedimiento para modificar la expresión génica de los genes para la biosíntesis endógena de acarbosa, para obtener un rendimiento mejorado de acarbosa, en donde

a) se introducen mutaciones en uno o en varios de los promotores génicos respectivos y

b) el rendimiento en acarbosa de la cepa productora así obtenida se compara con el de la cepa de partida.

40. Procedimiento según la reivindicación 39, **caracterizado** porque las mutaciones son

a) transiciones

b) deleciones y/o

c) adiciones.

Fig. 1: Hibridación Southern con *Streptomyces glaucescens* GLA.0

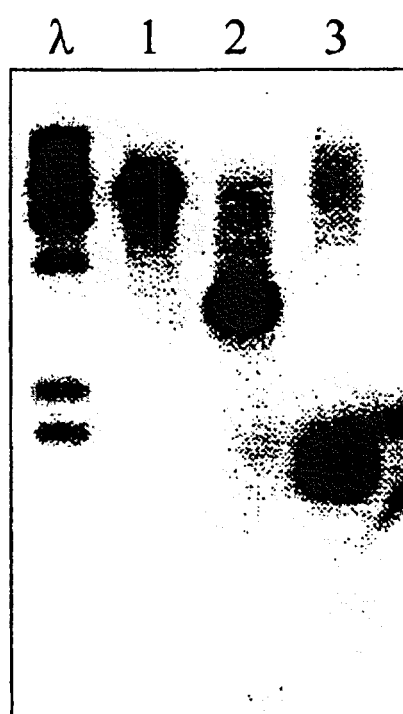


Fig. 2: Hibridación Southern con
Streptomyces glaucescens GLA.0

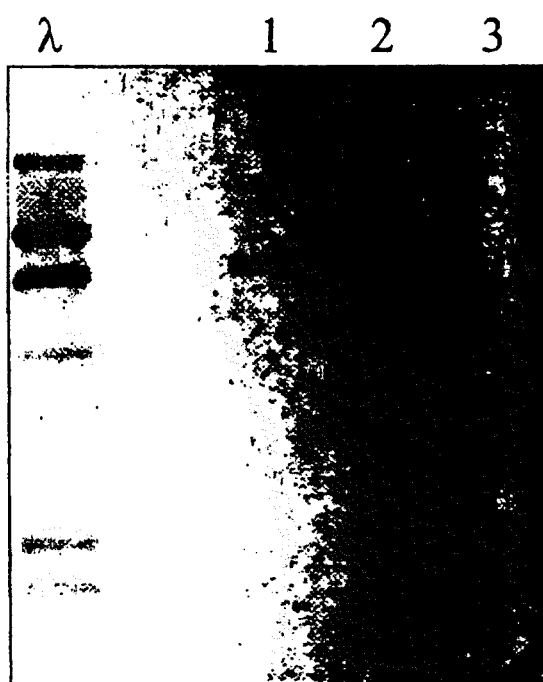


Fig. 3: Mapa de restricción del fragmento PstI de 6,8 kb de longitud de *Streptomyces glaucescens* GLA.0

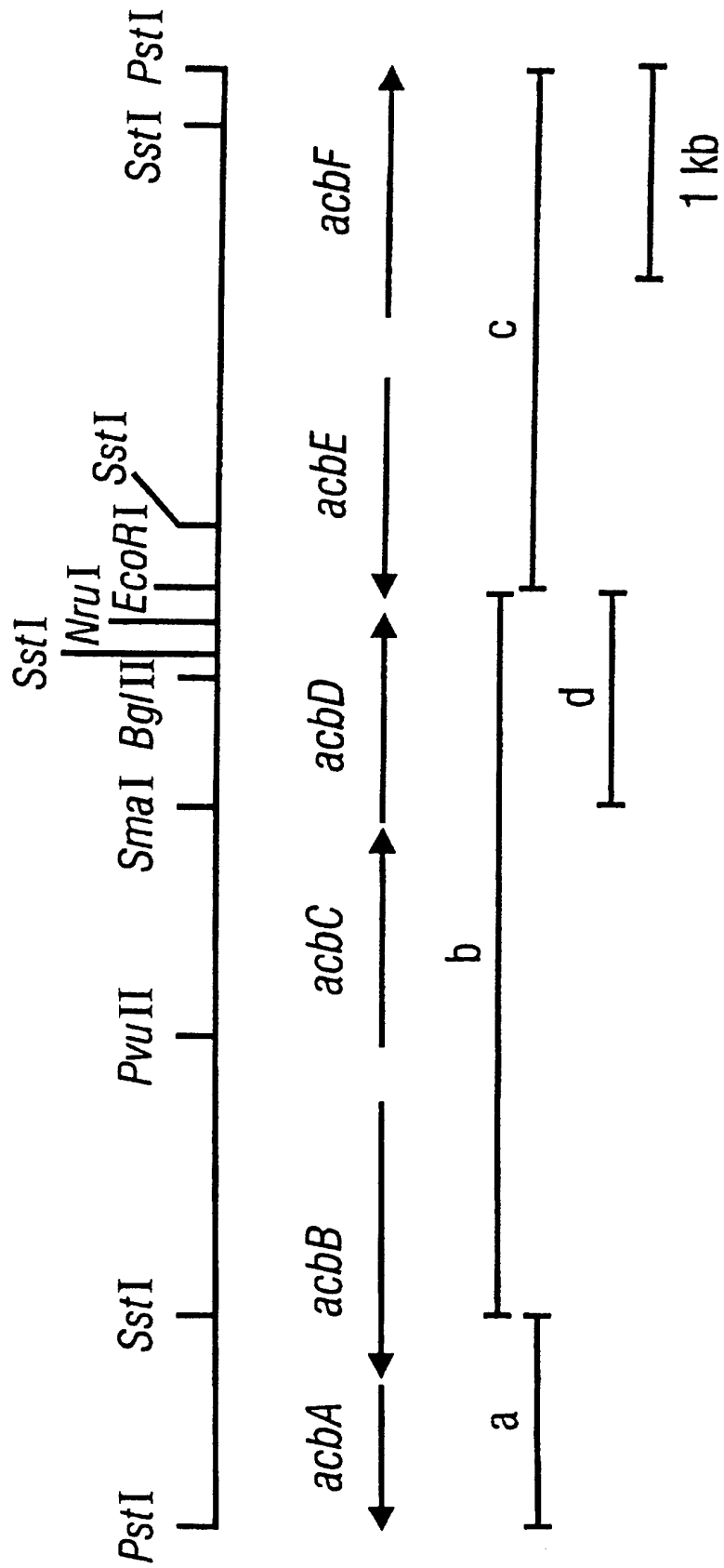


Fig. 4: Hibridación Southern con
Streptomyces glaucescens Δ acb

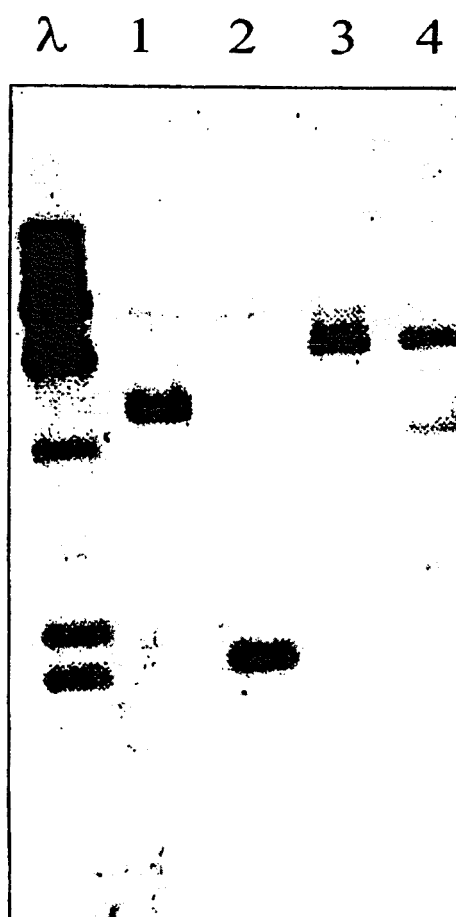
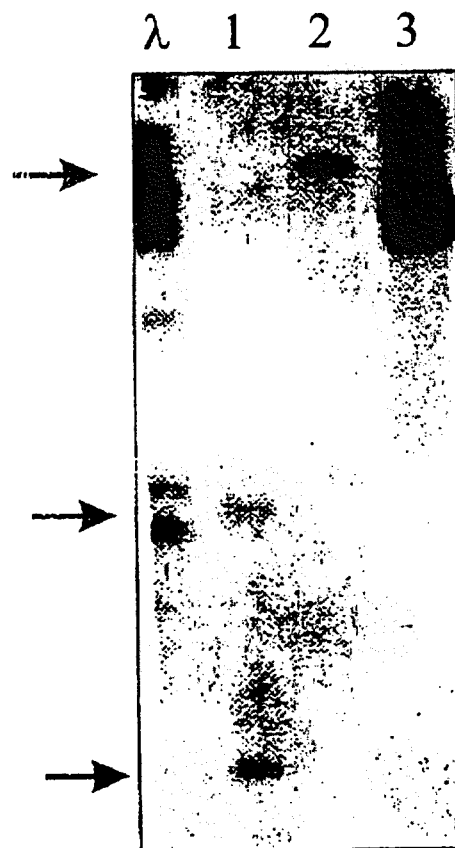


Fig. 5: Hibridación Southern con
Actinoplanes sp. SE50/100



LISTA DE SECUENCIAS

(1) INFORMACIÓN GENERAL:

- 5 (i) SOLICITANTE:
- (A) NOMBRE: Hoechst Aktiengesellschaft
- (B) CALLE: -
- 10 (C) LUGAR: Frankfurt
- (D) ESTADO FEDERADO: -
- (E) PAÍS: Alemania
- (F) CÓDIGO POSTAL: 65926
- 15 (G) TELÉFONO: 069-305-3005
- (H) TELEFAX: 069-35-7175
- (I) TELEX: -
- 20 (ii) TÍTULO DE LA SOLICITUD: Aislamiento de genes para la biosíntesis de pseudo-oligosacáridos procedentes de *Streptomyces glaucescens* GLA.O y su uso
- (iii) NÚMERO DE SECUENCIAS: 13
- 25 (iv) FORMA LEGIBLE POR ORDENADOR:
- (A) SOPORTE DE DATOS: Disquete
- (B) ORDENADOR: Compatible con PC IBM
- (C) SISTEMA OPERATIVO: PC-DOS/MS-DOS
- 30 (D) PROGRAMA: Patent In Release nº 1.0, versión nº 1.25 (EPA)

(2) INFORMACIÓN DE LA SEQ ID NO: 1:

- 35 (i) CARACTERÍSTICAS DE LA SECUENCIA:
- (A) LONGITUD: 22 pares de bases
- (B) TIPO: ácido nucleico
- 40 (C) FORMA DE LA CADENA: única
- (D) TOPOLOGÍA: lineal
- (ii) TIPO DE MOLÉCULA: ADN
- 45 (ix) CARACTERÍSTICAS:
- (A) NOMBRE/CLAVE: exón
- (B) UBICACIÓN: 1..22
- 50 (xi) DESCRIPCIÓN DE LA SECUENCIA: SEQ ID NO: 1:
- CSGGSGSSGC SGGSTTCATS GG

22

(2) INFORMACIÓN DE LA SEQ ID NO: 2:

- 55 (i) CARACTERÍSTICAS DE LA SECUENCIA:
- (A) LONGITUD: 24 pares de bases
- (B) TIPO: ácido nucleico
- 60 (C) FORMA DE CADENA: única
- (D) TOPOLOGÍA: lineal
- (ii) TIPO DE MOLÉCULA: ADN
- 65 (ix) CARACTERÍSTICAS:
- (A) NOMBRE/CLAVE: exón

ES 2 276 427 T3

(B) UBICACIÓN: 1..24

(xi) DESCRIPCIÓN DE LA SECUENCIA: SEQ ID NO: 2

5 GGGWVCTGGY VSGGSCCGTA GTTG

24

(2) INFORMACIÓN DE LA SEQ ID NO: 3:

10 (i) CARACTERÍSTICAS DE LA SECUENCIA:

(A) LONGITUD: 546 pares de bases

(B) TIPO: ácido nucleico

(C) FORMA DE CADENA: única

15 (D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: ADN (genómico)

20 (ix) CARACTERÍSTICAS:

(A) NOMBRE/CLAVE: exón

(B) UBICACIÓN: 1..546

25 (xi) DESCRIPCIÓN DE LA SECUENCIA: SEQ ID NO: 3:

CCCCGGCGGG	GCGGGGTTCA	TCGGCTCCGC	CTACGTCCGC	CGGCTCCTGT	CGCCCGGGGC	60
CCCCGGCGGC	GTGCGGTGA	CCGTCTCGA	CAAACTCACC	TACGCCGGCA	GCCTCGCCCG	120
CCTGCACGCG	GTGCGTGACC	ATCCCGGCCT	CACCTTCGTC	CAGGGCGACG	TGTGCGACAC	180
CGCGCTCGTC	GACACGCTGG	CCGCGCGGCA	CGACGACATC	GTGCACTTCG	CGGCCGAGTC	240
GCACGTCGAC	CGCTCCATCA	CCGACAGCGG	TGCCTTCACC	CGCACCACG	TGCTGGGCAC	300
CCAGGTCTCG	CTCGACGCCG	CGCTCCGCCA	CGGTGTGCGC	ACCCTCGTGC	ACGTCTCCAC	360
CGACGAGGTG	TACGGCTCCC	TCCCGCACGG	GGCCGCCGCG	GAGAGCGACC	CCCTGCTCCC	420
GACCTCGCCG	TACGCGGCGT	CGAAGGCGGC	CTCGGACCTC	ATGGCGCTCG	CCCACCACCG	480
CACCCACGGC	CTGGACGTCC	GGGTGACCCG	CTGTTTGAAC	AACTACGGCC	CGCACCAGTT	540
CCCCGG						546

45 (2) INFORMACIÓN DE LA SEQ ID NO: 4:

50 (i) CARACTERÍSTICAS DE LA SECUENCIA:

(A) LONGITUD: 541 pares de bases

(B) TIPO: ácido nucleico

(C) FORMA DE CADENA: única

55 (D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: ADN (genómico)

(ix) CARACTERÍSTICAS:

(A) NOMBRE/CLAVE: exón

60 (B) UBICACIÓN: 1..541

(xi) DESCRIPCIÓN DE LA SECUENCIA: SEQ ID NO: 4:

ES 2 276 427 T3

	CCCCGGGTGC TGGTAGGGGC CGTAGTTGTT GGAGCAGCGG GTGATGCGCA CGTCCAGGCC	60
	GTGGCTGACG TGCATGGCCA GCGCGAGCAG GTCGCCCCGAC GCCTTGGAGG TGGCATAGGG	120
5	GCTGTTGGGG CGCAGCGGCT CGTCCTCCGT CCACGACCCC GTCTCCAGCG AGCCGTAGAC	180
	CTCGTCGGTG GACACCTGCA CGAAGGGGGC CACGCCGTGC CGCAGGGCCG CGTCGAGGAG	240
	CACGTGCGAC TCGGCGGCGA GGTGCACGAC CTGGTCCTGG CCGGCCATGA CCCGGTCGAC	360
10	CAGGTCCGCG TCGCAGATGT CGCCGTGGAC GAAGCGCAGC CGGGGGTGGT CGCGGACCGG	420
	GTCGAGGTTG GCGAGGTTGC CGGCGTAGCT CAGGGCGTCG AGCACGGTGA CGACGGCGTC	480
15	GGGCGGCCCCG TCCGGACCGA GGAGGGTGCG GACGTAGTGC GAGCCCATGA ACCCGCCCGC	540
	C	541

(2) INFORMACIÓN DE LA SEQ ID NO: 5:

(i) CARACTERÍSTICAS DE LA SECUENCIA:

- (A) LONGITUD: 180 aminoácidos
- (B) TIPO: aminoácido
- (C) FORMA DE CADENA: única
- (D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: proteína

(ix) CARACTERÍSTICAS:

- (A) NOMBRE/CLAVE: PCRstrE.Pep
- (B) UBICACIÓN: 1..180

(xi) DESCRIPCIÓN DE LA SECUENCIA: SEQ ID NO: 5:

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

ES 2 276 427 T3

(2) INFORMACIÓN DE LA SEQ ID NO: 6:

(i) CARACTERÍSTICAS DE LA SECUENCIA:

- (A) LONGITUD: 181 aminoácidos
- (B) TIPO: aminoácido
- (C) FORMA DE CADENA: única
- (D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: proteína

(ix) CARACTERÍSTICAS:

- (A) NOMBRE/CLAVE: PCR acbD.Pep
- (B) UBICACIÓN: 1..181

(xi) DESCRIPCIÓN DE LA SECUENCIA: SEQ ID NO: 6:

```

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

Pro His Gln Phe Pro
180

```

(2) INFORMACIÓN DE LA SEQ ID NO: 7:

(i) CARACTERÍSTICAS DE LA SECUENCIA:

- (A) LONGITUD: 6854 pares de bases
- (B) TIPO: ácido nucleico
- (C) FORMA DE CADENA: doble
- (D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: ADN (genómico)

(ix) CARACTERÍSTICAS:

- (A) NOMBRE/CLAVE: Agrupación de genes de la biosíntesis “acarbosa”
- (B) UBICACIÓN: 1..6854

ES 2 276 427 T3

(xi) DESCRIPCIÓN DE LA SECUENCIA: SEQ ID NO: 7:

5	CTGCAGGGTT CCCTGGTGCA CGACCCGCCC CTGGTCGACG ACCAGGGCGC TGTCGCAGAT	60
	CGCGGCGATG TCGGCGATGT CGTGGCTGGT GAGCACCACG GTGGTGCCCA GTTCCCGGTG	120
	GGCGCGGTTG ACCAGCCGGC GCACCGCGTC CTTCAGCACC ATGTCGAGGC CGATCGTGGG	180
10	CTCGTCCCAG AACAGCACGG CCGGGTCGTG CAGCAGGCTC GCCGCGATCT CGGCGCGCAT	240
	GCGCTGTCCG AGGCTGAGCT GCCGCACGGG GGTGGACCCC AGCGCGTCCA TGTCGAGGAG	300
	GTCCCAGAAC AGGGCGAGGT TCGCGCGGTA GACCGGTCCG GGGATGTCGT AGATGCGGCG	360
15	CAGGATGCGG AAGGAGTCGG GTACCGACAG GTCCCACCAG AGCTGGCTGC GCTGGCCGAA	420
	GACGACCCG ATCGTGCGGG CGTTGCGCTG CCGGTGCCCG TAGGGCTCCA GCCCGGCGAC	480
	CGTGCAGCGG CCGGAGGTGG GGGTCATGAT GCCGGTCAGC ATCTTGATCG TGGTGCACTT	540
20	GCCGGTCCG TTGGCGCCGA TGTAGGCGGT CTTCTGTCCG GCCGGTATCT CGAAGGAGAC	600
	GTCGTCGACG GCGCGCACGA CGCGGTACCG GCGGGTCAGG AGGGTGGAGA GGCTGCCGAG	660
	CAGGCCGGGC TCGCGTTCGG CCAGCCGGAA CTCCTTGACG AGGTGTTCCG CCACGATCAC	720
25	GCGATACCC GCTCGACGGC CGTCTCCAGC AGGCGCAGGC CCTCGTCGAG CAGCGCCTCG	780
	TCGAGGGTGA ACGGCGGTGC CAGCCGCAGG ATGTGGCCGC CCAGGGAGGT GCGCAGCCCC	840
	AGGTCGAGGG CGGTGGTGTA GACGGCCCCG GCGGTCTCGG GGGCGGGTGC CCGGCCGACG	900
30	GCGTCGGTGA CGAACTCCAG GCCCCACAGC AGTCCGAGGC CGCGTACCTG GCCGAGCTGG	960
	GGGAAGCGGG ACTCCAGGGC GCGCAGCCGC TCCTGGATGA GCTCGCCGAG GACGCGCACG	1020
	CGGTGATCA GCCGGTCGCG CTCGACGACC TCCAGCGTGG CGCGGGCGGC GGCGATCCCC	1080
35	AGTGGGTTGC TCGCGTACGT CGAGGCGTAC GCCCCGGGGT GGCCGCCCTCC GGCCTGCGCA	1140
	GCTTCCGCGC GTCCGGCCAG CACGGCGAAG GGAATCCGC TCGCGGTGCC CTTGGACAGC	1200
	ATCGCCAGGT CCGGCTCGAT GCCGAACAGT TCGCTGGCGA GGAAGGCGCC GGTGCGCCCC	1260
40	CCGCCGGTGA GGACCTCGTC GGCGACGAGC AGCACGCCGC CGTCCCGGCA GGCGCCGGCG	1320
	ATCCGCTCCC AGTAGCCGGG GGGCGGCACG ATGACGCCTG CCGCGCCGAG GACGGGTTCC	1380
	AAGACCAGGG CCGAGACGTT GGGCTTCTCC GCGATGTGCC GGCGCACGAG GGTGCGGCAC	1440
45	CGCACGTGCG ACGAGGGGTA CTCCAGGCCC AGGGGACAGC GGTAGCCAGT AGGGGCTGTA	1500
	GCCAGCACGC TGTTGCCGCT GAAGGCCTGG TGGCCGATGT CCCAGTGGAC CAGCATCCGG	1560
	GCGCCCATGG TCTTGCCGTG GAAGCCGTGG CGCAGGGCGC AGATCCGGTT CGGGCCCGGC	1620
50	GCGGCGGTG CCTGGACGAC CCGCAGGGCG GCCTCGACCA CCTCCGCGCC GGTGGAGAAG	1680
	AAGGCGTAGG TGTCGAGCTG TTCGGGCAGC AGCCTGGCGA GCAGTTCCAG CAGGCCGGCG	1740
	CGGTCCGGCG TGGCGCTGTC GTGGACGTTT CACAGGCGGC GGGCCTGGGT GGTGAGTGCC	1800

ES 2 276 427 T3

	TCGACGACCT	CCGGGTGCCC	GTGGCCCAGT	GACTGGGTGA	GGGTCCCAGC	CGCGAAGTCG	1860
	AGGTACTGGT	TGCCGTCCAG	GTCCGTCCAG	ACGGGACCGC	GTCCCTCCGC	GAAGACCCGG	1920
5	CGTCCGTGGA	CGGCTTCCTC	GGAGGCGCCC	GGCGCCAGGT	GGCGGGCCTC	CCGTGCCAGG	1980
	TGCTGTGTCT	GCCGTAAGCC	TGTCATCGCT	GCCTCTGCTC	GTCGGACCGG	CTGACGCGAT	2040
	CGCCGGCGAA	CTGCGTTGTG	GCGCACCACG	GTTGGGGCGG	CTCGGCGCTG	AGTCAAACAC	2100
10	TTGAACACAC	ACCGCTGCAA	GAGTTTGCGG	GTTGTTTCAG	AAAGTTGTTG	CGAGCGGCCC	2160
	CGGCACTCTG	GTTGAGTCGA	CGTGCTTACG	GCGCCACCAC	GCCTCACGTT	CGAGGAGGGA	2220
	CCTGTGAGAA	CAAGCCCGCA	GACCGACCCG	CTCCCGCGGA	GGCCGAGGTG	AAGGCCCTGG	2280
15	TCCTGGCAGG	TGGAACCGGC	AGCAGACTGA	GGCCGTTCAC	CCACACCGCC	GCCAAGCAGC	2340
	TGCTCCCCAT	CGCCAACAAG	CCCCTGCTCT	TCTACGCGCT	GGAGTCCCTC	GCCGCGGCGG	2400
	GTGTCCGGGA	GGCCGGCGTC	GTGTTGGGCG	CGTACGGCCG	GGAGATCCGC	GAACCTCACCG	2460
20	GCGACGGCAC	CGCGTTCGGG	TTACGCATCA	CCTACCTCCA	CCAGCCCCGC	CCGCTCGGTC	2520
	TCGCGCACGC	GGTGCGCATC	GCCCCGGGCT	TCCTGGGCGA	CGACGACTTC	CTGCTGTACC	2580
	TGGGGGACAA	CTACCTGCCC	CAGGGCGTCA	CCGACTTCGC	CCGCCAATCG	GCCGCCGATC	2640
25	CCGCGGCGGC	CCGGCTGCTG	CTCACCCTCG	TCGCGGACCC	GTCCGCCTTC	GGCGTCGCGG	2700
	AGGTCGACGC	GGACGGGAAC	GTGCTGCGCT	TGGAGGAGAA	ACCCGACGTC	CCGCGCAGCT	2760
	CGCTCGCGCT	CATCGGCGTG	TACGCCTTCA	GCCCCGGCGT	CCACGAGGCG	GTACGGGCCA	2820
30	TCACCCCTC	CGCCCGCGGC	GAGCTGGAGA	TCACCCACGC	CGTGCAGTGG	ATGATCGACC	2880
	GGGGCTGCG	CGTACGGGCC	GAGACCACCA	CCCGGCCCTG	GCGCGACACC	GGCAGCGCGG	2940
35	AGGACATGCT	GGAGGTCAAC	CGTCACGTCC	TGGACGGACT	GGAGGGCCGC	ATCGAGGGGA	3000
	AGGTCGACGC	GCACAGCACG	CTGGTCGGCC	GGGTCCGGGT	GGCCGAAGGC	GCGATCGTGC	3060
	GGGGGTACA	CGTGGTGGGC	CCGGTGGTGA	TCGGCGCGGG	TGCCGTGCTC	AGCAACTCCA	3120
40	GTGTCGGCCC	GTACACCTCC	ATCGGGGAGG	ACTGCCGGGT	CGAGGACAGC	GCCATCGAGT	3180
	ACTCCGTCTC	GCTGCGCGGC	GCCCAGGTGC	AGGGGGCGTC	CCGCATCGAG	GCGTCCCTCA	3240
	TCGGCCGCGG	CGCCGTGCTC	GGCCCGGCCC	CCCCTCTCCC	GCAGGCTCAC	CGACTGGTGA	3300
45	TCGGCGACCA	CAGCAAGGTG	TATCTCACCC	CATGACCACG	ACCATCCTCG	TCACCGGCGG	3360
	AGCGGGCTTC	ATTGCTCCG	CCTACGTCCG	CCGGCTCCTG	TCGCCCCGGG	CCCCCGGCGG	3420
	CGTCGCGGTG	ACCGTCCTCG	ACAACTCAC	CTACGCCGGC	AGCCTCGCCC	GCCTGCACGC	3480
50	GGTGCGTGAC	CATCCCGGCC	TCACCTTCGT	CCAGGGCGAC	GTGTGCGACA	CCGCGCTCGT	3540
	CGACACGCTG	GCCGCGCGGC	ACGACGACAT	CGTGCACTTC	GCGGCCGAGT	CGCACGTGCA	3600
	CCGCTCCATC	ACCGACAGCG	GTGCCCTTCA	CCGCACCAAC	GTGCTGGGCA	CCCAGGTCCT	3660
55	GCTCGACGCC	GCGCTCCGCC	ACGGTGTGCG	CACCTTCGTG	CACGTCTCCA	CCGACGAGGT	3720
	GTACGGCTCC	CTCCCGCACG	GGGCCGCGCG	GGAGAGCGAC	CCCCTGCTTC	CGACCTCGCC	3780
60	GTACGCGGCG	TCGAAGGCGG	CCTCGGACCT	CATGGCGCTC	GCCCACCACC	GCACCCACGG	3840

65

ES 2 276 427 T3

	CCTGGACGTC	CGGGTGACCC	GCTGTTTCGAA	CAACTTCGGC	CCCCACCAGC	ATCCCGAGAA	3900
	GCTCATACCG	CGCTTCCTGA	CCAGCCTCCT	GTCCGGCGGC	ACCGTTCCCC	TCTACGGCGA	3960
5	CGGGCGGCAC	GTGCGCGACT	GGCTGCACGT	CGACGACCAC	GTCAGGGCCG	TCGAACTCGT	4020
	CCGCGTGTCT	GGCCGGCCCG	GAGAGATCTA	CAACATCGGG	GGCGGCACCT	CGCTGCCCAA	4080
	CCTGGAGCTC	ACGCACCGGT	TGCTCGCACT	GTGCGGCGCG	GGCCCGGAGC	GCATCGTCCA	4140
10	CGTCGAGAAC	CGCAAGGGGC	ACGACCGGCG	CTACGCGGTC	GACCACAGCA	AGATCACC GC	4200
	GGAACTCGGT	TACCGGCCCG	GCACCGACTT	CGCGACCGCG	CTGGCCGACA	CCGCGAAGTG	4260
	GTACGAGCGG	CACGAGGACT	GGTGGCGTCC	CCTGCTCGCC	GCGACATGAC	GTCGGGCGCG	4320
15	ACCGCAACCA	CCGGCCCCGG	CCGGCACACC	GCCGCCCGCG	GCCGGTGGCC	GGCCGGTCAG	4380
	CGTCCGTGAG	CCGGGCGCCG	GCCGCCCGCG	GGGCGGCGCG	CGGTGGACCC	CCGGACCACC	4440
	AGTTCGCGCA	TGAAGACGAA	TTCGTTGCGC	GGCGGCGCGC	TTCCGCTCAT	CTCCTCCAGC	4500
20	AGTGCCTCCA	CGGCGACCTG	CCCCATCGCC	TTGACGGGCT	GTCTGATGGT	GGTCAGGGGA	4560
	GGGTCCGTGA	AGGCCATGAG	CGGCGAGTCG	TGCAAGCCGA	CCACCGAGAT	GTCACCGGGA	4620
	ACCGTGAGAC	CCCGCCGCGC	CGCGGCCCGC	ACGGCGCCGA	GGGCCATCAT	GTCGCTGGCG	4680
25	CACATGACGG	CGGTGCAGCC	CAGGTCGATC	AGCGCGGACG	CGGCGGCCTG	GCCCCCTCC	4740
	AGGGAGAACA	GCGAGTGCTG	CACGAGCTCC	TCGGACTCCC	GCGCCGACAC	TCCCAGGTGC	4800
	TCCCGCACGC	CGGCCCGGAA	CCCCTCGATC	TTCCGCTGCA	CCGGCACGAA	GCGGGCGGGC	4860
30	CCGACGCGCA	GGCCGACGCG	CTCGTGCCCC	AGCTCCGCCA	GGTGCGCCAC	GGCCAGGCGC	4920
	ATCGCGGCCC	GGTCGTCCGG	GGAGACGAAG	GGTGCCTCGA	TCCGGGGCGA	GAACCCGTTC	4980
35	ACGAGGACGA	AGGGCACCTG	CCGCTCGTGC	AGCCGGCCGT	ACCGTCCGGT	CTCGGCGGTG	5040
	GTGTCCGCGT	GCAGTCCGGA	GACGAAGATG	ATGCCGACAC	CCCCGCGGTC	CACGAGCATC	5100
	TCCGTGAGTT	CGTCCTCGGT	CGAGCCGCCC	GGGGTCTGCG	TGGCGAGCAC	GGGCGTGTAG	5160
40	CCCTGACCGG	TGAGCGCCTG	CCCCATCACC	TGGGCCAGTG	CGGGGAAGAA	GGGGTTGTCC	5220
	AGTTCGGGGG	TGACCAGTCC	GACCAGCTCG	GCGCGGCGCT	GTCGCGCCGG	CTGCTCGTAG	5280
	CCCAGCGCGT	CCAGTGCGGT	CAGCACCGAG	TCGCGGGTGC	CGGTGGCCAC	ACCGCGCGCA	5340
45	CCGTTCAGCA	CCCGGCTGAC	CGTGGCCTTG	CTGACGCCCC	CCCGGGCTGC	GATGTCGGCG	5400
	AGCCGCATGG	TCATGGCAAC	GCACTCTACC	TGTCGGGGCG	TCAGGGCGTG	CCCACCGCGC	5460
	GCGGAACCGG	CGGACTGCGG	GGCACGGCCC	GTCCGCGGCC	CACGGACCAC	GCGCCCGAAA	5520
50	CGATGGCTGA	AAATGCTTGC	AGCAAATTGC	CGCAACGTCT	TTCCGGCGGCT	TTTCGATCCT	5580
	GTTACGTTCC	TGGCAACCCC	GGCGCCGCGC	AGAAGCGGTT	GGCGTGAGGC	GTCCAGACCT	5640
	CCGCCCGATT	CCGGGATCAC	TCAGGGGAGT	TCACAATGCG	GCGTGGCATT	GCGGCCACCG	5700
55	CGCTGTTTCG	GGCTGTGGCC	ATGACGGCAT	CGGCGTGTGG	CGGGGGCGAC	AACGGCGGAA	5760
	GCGGTACCGA	CGCGGGCGGC	ACGGAGCTGT	CGGGGACCGT	CACCTTCTGG	GACACGTCCA	5820
	ACGAAGCCGA	GAAGGCGACG	TACCAGGCCC	TCGCGGAGGG	CTTCGAGAAG	GAGCACCCGA	5880

ES 2 276 427 T3

	AGGTCGACGT CAAGTACGTC AACGTCCCGT TCGGCGAGGC GAACGCCAAG TTCAAGAACG	5940
	CCGCGGGCGG CAACTCCGGT GCGCCGGACG TGATGCGTAC GGAGGTCGCC TGGGTCGCGG	6000
5	ACTTCGCCAG CATCGGCTAC CTCGCCCCGC TCGACGGCAC GCGCGCCCTC GACGACGGGT	6060
	CGGACCACCT TCCCCAGGGC GGCAGACCA GGTACGAGGG GAAGACCTAC GCGGTCCCCG	6120
10	AGGTGATCGA CACCCTGGCG CTCTTCTACA ACAAGGAAC TCTGACGAAG GCCGGTGTCG	6180
	AGGTGCCGGG CTCCCTCGCC GAGCTGAAGA CGGCCGCCGC CGAGATCACC GAGAAGACCG	6240
	GCGCGAGCGG CCTCTACTGC GGGGCGACGA CCCGTACTTG GTTCCTGCCC TACCTCTACG	6300
15	GGGAGGGCGG CGACCTGGTC GACGAGAAGA ACAAGACCGT CACGGTCGAC GACGAAGCCG	6360
	GTGTGCGCGC CTACCGCGTC ATCAAGGACC TCGTGGACAG CAAGGCGGCC ATCACCGACG	6420
	CGTCCGACGG CTGGAACAAC ATGCAGAACG CTTTCAAGTC GGGCAAGGTC GCCATGATGG	6480
20	TCAACGGCCC CTGGGCCATC GAGGACGTCA AGGCGGGAGC CCGCTTCAAG GACGCCGGCA	6540
	ACCTGGGGGT CGCCCCCGTC CCGGCCGGCA GTGCCGGACA GGGCTCTCCC CAGGGCGGGT	6600
	GGAACCTCTC GGTGTACGCG GGCTCGAAGA ACCTCGACGC CTCCTACGCC TTCGTGAAGT	6660
25	ACATGAGCTC CGCCAAGGTG CAGCAGCAGA CCACCGAGAA GCTGAGCCTG CTGCCCACCC	6720
	GCACGTCCGT CTACGAGGTC CCGTCCGTCG CGGACAACGA GATGGTGAAG TTCTTCAAGC	6780
	CGGCCGTCGA CAAGGCCGTC GAACGGCCGT GGATCGCCGA GGGCAATGCC CTCTTCGAGC	6840
30	CGATCCGGCT GCAG	6854

(2) INFORMACIÓN DE LA SEQ ID NO: 8:

- 35 (i) CARACTERÍSTICAS DE LA SECUENCIA:
- (A) LONGITUD: 240 aminoácidos
 - (B) TIPO: aminoácido
 - 40 (C) FORMA DE CADENA: única
 - (D) TOPOLOGÍA: lineal
- (ii) TIPO DE MOLÉCULA: proteína
- 45 (ix) CARACTERÍSTICAS:
- (A) NOMBRE/CLAVE: acbA
 - (B) UBICACIÓN: 1..240
- 50 (xi) DESCRIPCIÓN DE LA SECUENCIA: SEQ ID NO: 8:

55

60

65

ES 2 276 427 T3

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

(2) INFORMACIÓN DE LA SEQ ID NO: 9:

(i) CARACTERÍSTICAS DE LA SECUENCIA:

- (A) LONGITUD: 429 aminoácidos
- (B) TIPO: aminoácido
- (C) FORMA DE CADENA: única
- (D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: proteína

(ix) CARACTERÍSTICAS:

- (A) NOMBRE/CLAVE: acbB
- (B) UBICACIÓN: 1..429

(xi) DESCRIPCIÓN DE LA SECUENCIA: SEQ ID NO: 9:

ES 2 276 427 T3

Met Thr Gly Leu Arg Gln Thr Gln His Leu Ala Arg Glu Ala Arg His
 1 5 10 15
 Leu Ala Pro Gly Ala Ser Glu Glu Ala Val His Gly Arg Arg Val Phe
 20 25 30
 Ala Glu Gly Arg Gly Pro Val Leu Thr Asp Leu Asp Gly Asn Gln Tyr
 35 40 45
 Leu Asp Phe Ala Ala Gly Thr Leu Thr Gln Ser Leu Gly His Gly His
 50 55 60
 Pro Glu Val Val Glu Ala Leu Thr Thr Gln Ala Arg Arg Leu Trp Asn
 65 70 75 80
 Val His Asp Ser Ala Thr Pro Asp Arg Ala Gly Leu Leu Glu Leu Leu
 85 90 95
 Ala Arg Leu Leu Pro Glu Gln Leu Asp Thr Tyr Ala Phe Phe Ser Thr
 100 105 110
 Gly Ala Glu Val Val Glu Ala Ala Leu Arg Val Val Gln Ala Thr Ala
 115 120 125
 Ala Pro Gly Arg Asn Arg Ile Cys Ala Leu Arg His Gly Phe His Gly
 130 135 140
 Lys Thr Met Gly Ala Arg Met Leu Val His Trp Asp Ile Gly His Gln
 145 150 155 160
 Ala Phe Ser Gly Asn Ser Val Leu Ala Thr Ala Pro Thr Gly Tyr Arg
 165 170 175
 Cys Pro Leu Gly Leu Glu Tyr Pro Ser Cys Asp Val Arg Cys Ala Thr
 180 185 190
 Leu Val Arg Arg His Ile Ala Glu Lys Pro Asn Val Ser Ala Leu Val
 195 200 205
 Phe Glu Pro Val Leu Gly Ala Ala Gly Val Ile Val Pro Pro Pro Gly
 210 215 220
 Tyr Trp Glu Arg Ile Ala Gly Ala Cys Arg Asp Gly Gly Val Leu Leu
 225 230 235 240
 Val Ala Asp Glu Val Leu Thr Gly Gly Gly Arg Thr Gly Ala Phe Leu
 245 250 255
 Ala Ser Glu Leu Phe Gly Ile Glu Pro Asp Leu Ala Met Leu Ser Lys
 260 265 270
 Gly Thr Ala Ser Gly Phe Pro Phe Ala Val Leu Ala Gly Arg Ala Glu
 275 280 285
 Ala Ala Gln Ala Gly Gly Gly His Pro Gly Ala Tyr Ala Ser Thr Tyr
 290 295 300
 Ala Ser Asn Pro Leu Gly Ile Ala Ala Ala Arg Ala Thr Leu Glu Val
 305 310 315 320
 Val Glu Arg Asp Arg Leu Ile Asp Arg Val Arg Val Leu Gly Glu Leu
 325 330 335
 Ile Gln Glu Arg Leu Arg Ala Leu Glu Ser Arg Phe Pro Gln Leu Gly
 340 345 350
 Gln Val Arg Gly Leu Gly Leu Leu Trp Gly Leu Glu Phe Val Thr Asp
 355 360 365
 Ala Val Gly Arg Ala Pro Ala Pro Glu Thr Ala Arg Ala Val Tyr Thr
 370 375 380
 Thr Ala Leu Asp Leu Gly Leu Arg Thr Ser Leu Gly Gly His Ile Leu
 385 390 395 400
 Arg Leu Ala Pro Pro Phe Thr Leu Asp Glu Ala Leu Leu Asp Glu Gly
 405 410 415
 Leu Arg Leu Leu Glu Thr Ala Val Glu Arg Val Ile Ala
 420 425

ES 2 276 427 T3

(2) INFORMACIÓN DE LA SEQ ID NO: 10:

(i) CARACTERÍSTICAS DE LA SECUENCIA:

(A) LONGITUD: 355 aminoácidos

(B) TIPO: aminoácido

(C) FORMA DE CADENA: única

(D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: proteína

(ix) CARACTERÍSTICAS:

(A) NOMBRE/CLAVE: acbC

(B) UBICACIÓN: 1..355

(xi) DESCRIPCIÓN DE LA SECUENCIA: SEQ ID NO: 10:

```

Val Lys Ala Leu Val Leu Ala Gly Gly Thr Gly Ser Arg Leu Arg Pro
1      5      10
Phe Thr His Thr Ala Ala Lys Gln Leu Leu Pro Ile Ala Asn Lys Pro
20      25      30
Val Leu Phe Tyr Ala Leu Glu Ser Leu Ala Ala Ala Gly Val Arg Glu
35      40      45
Ala Gly Val Val Val Gly Ala Tyr Gly Arg Glu Ile Arg Glu Leu Thr
50      55      60
Gly Asp Gly Thr Ala Phe Gly Leu Arg Ile Thr Tyr Leu His Gln Pro
65      70      75      80
Arg Pro Leu Gly Leu Ala His Ala Val Arg Ile Ala Arg Gly Phe Leu
85      90      95
Gly Asp Asp Asp Phe Leu Leu Tyr Leu Gly Asp Asn Tyr Leu Pro Gln
100      105      110
Gly Val Thr Asp Phe Ala Arg Gln Ser Ala Ala Asp Pro Ala Ala Ala
115      120      125
Arg Leu Leu Leu Thr Pro Val Ala Asp Pro Ser Ala Phe Gly Val Ala
130      135      140
Glu Val Asp Ala Asp Gly Asn Val Leu Arg Leu Glu Glu Lys Pro Asp
145      150      155      160
Val Pro Arg Ser Ser Leu Ala Leu Ile Gly Val Tyr Ala Phe Ser Pro
165      170      175
Ala Val His Glu Ala Val Arg Ala Ile Thr Pro Ser Ala Arg Gly Glu
180      185      190
Leu Glu Ile Thr His Ala Val Gln Trp Met Ile Asp Arg Gly Leu Arg
195      200      205
Val Arg Ala Glu Thr Thr Thr Arg Pro Trp Arg Asp Thr Gly Ser Ala
210      215      220
Glu Asp Met Leu Glu Val Asn Arg His Val Leu Asp Gly Leu Glu Gly
225      230      235      240
Arg Ile Glu Gly Lys Val Asp Ala His Ser Thr Leu Val Gly Arg Val
245      250      255

```

ES 2 276 427 T3

		Arg	Val	Ala	Glu	Gly	Ala	Ile	Val	Arg	Gly	Ser	His	Val	Val	Gly	Pro
					260					265					270		
5		Val	Val	Ile	Gly	Ala	Gly	Ala	Val	Val	Ser	Asn	Ser	Ser	Val	Gly	Pro
				275					280					285			
		Tyr	Thr	Ser	Ile	Gly	Glu	Asp	Cys	Arg	Val	Glu	Asp	Ser	Ala	Ile	Glu
		290						295					300				
10		Tyr	Ser	Val	Leu	Leu	Arg	Gly	Ala	Gln	Val	Glu	Gly	Ala	Ser	Arg	Ile
		305					310					315					320
		Glu	Ala	Ser	Leu	Ile	Gly	Arg	Gly	Ala	Val	Val	Gly	Pro	Ala	Pro	Arg
					325						330				335		
15		Leu	Pro	Gln	Ala	His	Arg	Leu	Val	Ile	Gly	Asp	His	Ser	Lys	Val	Tyr
				340						345					350		
		Leu	Thr	Pro													
20				355													

(2) INFORMACIÓN DE LA SEQ ID NO: 11:

(i) CARACTERÍSTICAS DE LA SECUENCIA:

- (A) LONGITUD: 325 aminoácidos
- (B) TIPO: aminoácido
- (C) FORMA DE CADENA: única
- (D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: proteína

(ix) CARACTERÍSTICAS:

- (A) NOMBRE/CLAVE: acbD
- (B) UBICACIÓN: 1..325

(xi) DESCRIPCIÓN DE LA SECUENCIA: SEQ ID NO: 11:

ES 2 276 427 T3

1 Met Thr Thr Thr Ile Leu Val Thr Gly Gly Ala Gly Phe Ile Arg Ser
 5 5 10 15
 20 Ala Tyr Val Arg Arg Leu Leu Ser Pro Gly Ala Pro Gly Gly Val Ala
 25 30
 35 Val Thr Val Leu Asp Lys Leu Thr Tyr Ala Gly Ser Leu Ala Arg Leu
 40 45
 50 His Ala Val Arg Asp His Pro Gly Leu Thr Phe Val Gln Gly Asp Val
 55 60
 65 Cys Asp Thr Ala Leu Val Asp Thr Leu Ala Ala Arg His Asp Asp Ile
 70 75 80
 85 Val His Phe Ala Ala Glu Ser His Val Asp Arg Ser Ile Thr Asp Ser
 90 95
 100 Gly Ala Phe Thr Arg Thr Asn Val Leu Gly Thr Gln Val Leu Leu Asp
 105 110
 115 Ala Ala Leu Arg His Gly Val Arg Thr Phe Val His Val Ser Thr Asp
 120 125
 130 Glu Val Tyr Gly Ser Leu Pro His Gly Ala Ala Ala Glu Ser Asp Pro
 135 140
 145 Leu Leu Pro Thr Ser Pro Tyr Ala Ala Ser Lys Ala Ala Ser Asp Leu
 150 155 160
 165 Met Ala Leu Ala His His Arg Thr His Gly Leu Asp Val Arg Val Thr
 170 175
 180 Arg Cys Ser Asn Asn Phe Gly Pro His Gln His Pro Glu Lys Leu Ile
 185 190
 195 Pro Arg Phe Leu Thr Ser Leu Leu Ser Gly Gly Thr Val Pro Leu Tyr
 200 205
 210 Gly Asp Gly Arg His Val Arg Asp Trp Leu His Val Asp Asp His Val
 215 220
 225 Arg Ala Val Glu Leu Val Arg Val Ser Gly Arg Pro Gly Glu Ile Tyr
 230 235 240
 245 Asn Ile Gly Gly Gly Thr Ser Leu Pro Asn Leu Glu Leu Thr His Arg
 250 255
 260 Leu Leu Ala Leu Cys Gly Ala Gly Pro Glu Arg Ile Val His Val Glu
 265 270
 275 Asn Arg Lys Gly His Asp Arg Arg Tyr Ala Val Asp His Ser Lys Ile
 280 285
 290 Thr Ala Glu Leu Gly Tyr Arg Pro Arg Thr Asp Phe Ala Thr Ala Leu
 295 300
 305 Ala Asp Thr Ala Lys Trp Tyr Glu Arg His Glu Asp Trp Trp Arg Pro
 310 315 320
 325 Leu Leu Ala Ala Thr

60 (2) INFORMACIÓN DE LA SEQ ID NO: 12:

(i) CARACTERÍSTICAS DE LA SECUENCIA:

- (A) LONGITUD: 345 aminoácidos
- (B) TIPO: aminoácido
- (C) FORMA DE CADENA: única

ES 2 276 427 T3

(D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: proteína

(ix) CARACTERÍSTICAS:

(A) NOMBRE/CLAVE: acbE

(B) UBICACIÓN: 1..345

(xi) DESCRIPCIÓN DE LA SECUENCIA: SEQ ID NO: 12:

```

Met Thr Met Arg Leu Ala Asp Ile Ala Ala Arg Ala Gly Val Ser Lys
 1          5          10          15
Ala Thr Val Ser Arg Val Leu Asn Gly Ala Arg Gly Val Ala Thr Gly
          20          25          30
Thr Arg Asp Ser Val Leu Thr Ala Leu Asp Ala Leu Gly Tyr Glu Gln
          35          40          45
Pro Ala Arg Gln Arg Arg Ala Glu Leu Val Gly Leu Val Thr Pro Glu
 20          50          55          60
Leu Asp Asn Pro Phe Phe Pro Ala Leu Ala Gln Val Met Gly Gln Ala
 65          70          75          80
Leu Thr Arg Gln Gly Tyr Thr Pro Val Leu Ala Thr Gln Thr Pro Gly
 25          85          90          95
Gly Ser Thr Glu Asp Glu Leu Thr Glu Met Leu Val Asp Arg Gly Val
          100          105          110
Ser Gly Ile Ile Phe Val Ser Gly Leu His Ala Asp Thr Thr Ala Glu
 30          115          120          125
Thr Gly Arg Tyr Gly Arg Leu His Glu Arg Gln Val Pro Phe Val Leu
          130          135          140
Val Asn Gly Phe Ser Pro Arg Ile Glu Ala Pro Phe Val Ser Pro Asp
 35          145          150          155          160
Asp Arg Ala Ala Met Arg Leu Ala Val Ala His Leu Ala Glu Leu Gly
          165          170          175
His Glu Arg Val Gly Leu Ala Val Gly Pro Ala Arg Phe Val Pro Val
 40          180          185          190
Gln Arg Lys Ile Glu Gly Phe Arg Ala Gly Val Arg Glu His Leu Gly
          195          200          205
Val Ser Ala Arg Glu Ser Glu Glu Leu Val Gln His Ser Leu Phe Ser
 45          210          215          220
Leu Glu Gly Gly Gln Ala Ala Ala Ser Ala Leu Ile Asp Leu Gly Cys
 225          230          235          240
Thr Ala Val Met Cys Ala Ser Asp Met Met Ala Leu Gly Ala Val Arg
 50          245          250          255
Ala Ala Arg Arg Arg Gly Leu Thr Val Pro Gly Asp Ile Ser Val Val
          260          265          270
Gly Phe Asp Asp Ser Pro Leu Met Ala Phe Thr Asp Pro Pro Leu Thr
 55          275          280          285
Thr Ile Arg Gln Pro Val Lys Ala Met Gly Gln Val Ala Val Asp Ala
          290          295          300
Leu Leu Glu Glu Met Ser Gly Thr Pro Pro Pro Arg Thr Glu Phe Val
 60          305          310          315          320
Phe Met Pro Glu Leu Val Val Arg Gly Ser Thr Ala Ala Gly Pro Arg
          325          330          335
Gly Gly Arg Arg Pro Ala His Gly Arg
 65          340          345

```

(2) INFORMACIÓN DE LA SEQ ID NO: 13:

(i) CARACTERÍSTICAS DE LA SECUENCIA:

(A) LONGITUD: 393 aminoácidos

(B) TIPO: aminoácido

(C) FORMA DE CADENA: única

(D) TOPOLOGÍA: lineal

(ii) TIPO DE MOLÉCULA: proteína

(ix) CARACTERÍSTICAS:

(A) NOMBRE/CLAVE: acbF

(B) UBICACIÓN: 1..393

(xi) DESCRIPCIÓN DE LA SECUENCIA: SEQ ID NO: 13:

```

Met Arg Arg Gly Ile Ala Ala Thr Ala Leu Phe Ala Ala Val Ala Met
1      5      10      15

Thr Ala Ser Ala Cys Gly Gly Gly Asp Asn Gly Gly Ser Gly Thr Asp
20      25      30

Ala Gly Gly Thr Glu Leu Ser Gly Thr Val Thr Phe Trp Asp Thr Ser
35      40      45

Asn Glu Ala Glu Lys Ala Thr Tyr Gln Ala Leu Ala Glu Gly Phe Glu
50      55      60

Lys Glu His Pro Lys Val Asp Val Lys Tyr Val Asn Val Pro Phe Gly
65      70      75      80

Glu Ala Asn Ala Lys Phe Lys Asn Ala Ala Gly Gly Asn Ser Gly Ala
85      90      95

Pro Asp Val Met Arg Thr Glu Val Ala Trp Val Ala Asp Phe Ala Ser
100     105     110

Ile Gly Tyr Leu Ala Pro Leu Asp Gly Thr Pro Ala Leu Asp Asp Gly
115     120     125

Ser Asp His Leu Pro Gln Gly Gly Ser Thr Arg Tyr Glu Gly Lys Thr
130     135     140

Tyr Ala Val Pro Gln Val Ile Asp Thr Leu Ala Leu Phe Tyr Asn Lys
145     150     155     160

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

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

Leu Tyr Cys Gly Ala Thr Thr Arg Thr Trp Phe Leu Pro Tyr Leu Tyr
195     200     205

Gly Glu Gly Gly Asp Leu Val Asp Glu Lys Asn Lys Thr Val Thr Val
210     215     220

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

Asp Ser Lys Ala Ala Ile Thr Asp Ala Ser Asp Gly Trp Asn Asn Met
245     250     255

Gln Asn Ala Phe Lys Ser Gly Lys Val Ala Met Met Val Asn Gly Pro
260     265     270

Trp Ala Ile Glu Asp Val Lys Ala Gly Ala Arg Phe Lys Asp Ala Gly
275     280     285

Asn Leu Gly Val Ala Pro Val Pro Ala Gly Ser Ala Gly Gln Gly Ser
290     295     300

Pro Gln Gly Gly Trp Asn Leu Ser Val Tyr Ala Gly Ser Lys Asn Leu
305     310     315     320

```

ES 2 276 427 T3

	Asp	Ala	Ser	Tyr	Ala	Phe	Val	Lys	Tyr	Met	Ser	Ser	Ala	Lys	Val	Gln	
					325					330					335		
5	Gln	Gln	Thr	Thr	Glu	Lys	Leu	Ser	Leu	Leu	Pro	Thr	Arg	Thr	Ser	Val	
					340					345				350			
	Tyr	Glu	Val	Pro	Ser	Val	Ala	Asp	Asn	Glu	Met	Val	Lys	Phe	Phe	Lys	
			355					360					365				
10	Pro	Ala	Val	Asp	Lys	Ala	Val	Glu	Arg	Pro	Trp	Ile	Ala	Glu	Gly	Asn	
		370					375					380					
	Ala	Leu	Phe	Glu	Pro	Ile	Arg	Leu	Gln								
	385					390											
15																	
20																	
25																	
30																	
35																	
40																	
45																	
50																	
55																	
60																	
65																	