(19) OFICINA ESPAÑOLA DE PATENTES Y MARCAS

ESPAÑA

(11) Número de publicación: **2 681 692**

(51) Int. CI.:

C07K 14/195 (2006.01)
A23C 9/12 (2006.01)

(12) TRADUCCIÓN DE PATENTE EUROPEA T3

(86) Fecha de presentación y número de la solicitud internacional: **29.03.2011 PCT/EP2011/054865**

(87) Fecha y número de publicación internacional: **06.10.2011 WO11120993**

(96) Fecha de presentación y número de la solicitud europea: **29.03.2011 E 11713215 (9)**

(97) Fecha y número de publicación de la concesión europea: **02.05.2018 EP 2552945**

(54) Título: **Polipéptidos que tienen actividad transgalactosilante**

(30) Prioridad:

29.03.2010 US 318580 P
29.03.2010 EP 10158172

(45) Fecha de publicación y mención en BOPI de la traducción de la patente:
14.09.2018

(73) Titular/es:

DUPONT NUTRITION BIOSCIENCES APS (100.0%)
Langebrogade 1
1411 Copenhagen K, DK

(72) Inventor/es:

LARSEN, MORTEN KROG y
POULSEN, CHARLOTTE, HORSMANS

(74) Agente/Representante:

ELZABURU, S.L.P

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

Polipéptidos que tienen actividad transgalactosilante

Campo de la invención

La presente invención se refiere a polipéptidos, específicamente polipéptidos que tienen actividad transgalactosilante y a los ácidos nucleicos que los codifican, y sus usos en, p. ej. productos lácteos.

Antecedentes de la invención

Los galactooligosacáridos (GOS) son hidratos de carbono que no son digeribles en seres humanos y animales que comprenden dos o más moléculas de galactosa, típicamente hasta nueve, unidas mediante enlaces glucosídicos. Los GOS también pueden incluir una o más moléculas de glucosa. Uno de los efectos beneficiosos de GOS es su capacidad de actuar como compuestos prebióticos estimulando selectivamente la proliferación de microorganismos colónicos beneficiosos, tales como bacterias, para proporcionar beneficios fisiológicos al consumidor. Los efectos sobre la salud contrastados han dado lugar a un creciente interés en los GOS como ingredientes alimentarios para diversos tipos de alimentos.

La enzima β-galactosidasa (EC 3.2.1.23) generalmente hidroliza la lactosa a los monosacáridos D-glucosa y D-galactosa. En la reacción enzimática normal de las β-galactosidasas, la enzima hidroliza lactosa y une transitoriamente el monosacárido galactosa en un complejo galactosa-enzima que transfiere galactosa al grupo hidroxilo del agua, que da como resultado la liberación de D-galactosa y D-glucosa. Sin embargo, a altas concentraciones de lactosa, algunas β-galactosidasas son capaces de transferir galactosa a los grupos hidroxilo de D-galactosa o D-glucosa en un proceso llamado transgalactosilación mediante el cual se producen galactooligosacáridos.

El género Bifidobacterium es uno de los tipos de cultivos de bacterias más comúnmente utilizados en la industria láctea para la fermentación de una variedad de productos lácteos. La ingestión de productos que contienen Bifidobacterium tiene además un efecto promotor de la salud. Este efecto no solo se logra por un pH reducido de los contenidos intestinales sino también por la capacidad de Bifidobacterium para repoblar la flora intestinal en individuos a los que se les ha alterado la flora intestinal por ejemplo por la ingesta de antibióticos. Además, Bifidobacterium tiene el potencial de desplazar los microorganismos intestinales potencialmente nocivos.

Los galactooligosacáridos son conocidos por potenciar el crecimiento de Bifidobacterium. Este efecto probablemente se logra a través de la capacidad única de Bifidobacterium para aprovechar los galactooligosacáridos como fuente de carbono. También se cree que el suplemento dietético de galactooligosacáridos tiene una serie de efectos de protección de enfermedades a largo plazo. Por ejemplo, la ingesta de galactooligosacáridos ha demostrado ser altamente protectora contra el desarrollo de cáncer colorrectal en ratas. Por lo tanto, existe un gran interés en desarrollar métodos económicos y eficientes para producir galactooligosacáridos para uso en la industria para mejorar los suplementos dietéticos y los productos lácteos.

Se conoce un polipéptido beta-galactosidasa de Ruminococcus/Blautia hansenii que tiene 1807 aminoácidos (con SEQ ID NO: 12) de la base de datos UniProt, 24 de noviembre de 2009, "Subname: Full=Beta-galactosidase" XP002591904 obtenido del no. de acceso EBI UNIPROT: C9LAL1.

Se conoce una glicosidasa que tiene 1768 aminoácidos (con SEQ ID NO: 13) de la base de datos UniProt, 14 de octubre de 2008, "Subname: Full=Putative no caracterised protein" XP002610554 obtenido del no. de acceso EBI UN-IPROT: B5CQV4.

Se ha descrito una lactasa extracelular de *Bifidobacterium bifidum* DSM20215 truncada con aproximadamente 580 aminoácidos (BIF3) como una enzima transgalactosilante en una solución que contiene lactosa solubilizada en agua (Jørgensen et al. (2001), Appl. Microbiol. Biotechnol., 57: 647 - 652). En el documento WO 2009/071539 se describe un fragmento truncado diferente comparado con BIF3 que da como resultado una hidrólisis eficiente y una producción muy baja de GOS cuando se ensaya en leche.

El documento WO 01/90317 describe un polipéptido que tiene actividad transgalactosilante de *Bifidobacterium bifidum* (OLGA5) y su uso para producir productos lácteos.

Las enzimas lactasa de *Bifidobacterium bifidum* descritas anteriormente tienen la desventaja de que requieren altas concentraciones de lactosa para exhibir actividad transgalactosilasa o que tienen predominantemente actividad beta-galactosilasa (hidrolasa).

Todavía existe la necesidad de desarrollar enzimas que sean eficientes en la producción de GOS y que, además, puedan funcionar a bajos niveles de sustrato lactosa, tal como en leche.

Objeto de la invención

Es un objeto de las realizaciones de la invención proporcionar polipéptidos que tengan actividad transgalactosilante. Polipéptidos que tienen una relación útil de actividad de transgalactosilación a hidrólisis y, por lo tanto, son productores eficientes de GOS cuando se incuban con lactosa incluso a bajos niveles de lactosa tal como en un producto a base de leche. Describimos un método para la producción de galactooligosacáridos (GOS) in situ en productos lácteos. También describimos un método para desarrollar un método más barato y más eficiente para la producción de galactooligosacáridos (GOS) para uso en la industria.

Compendio de la invención

La presente invención describe un polipéptido aislado que tiene actividad transgalactosilante que comprende una secuencia de aminoácidos que tiene al menos 80% de identidad de secuencia con la secuencia de aminoácidos del polipéptido maduro de SEQ ID NO 1, cuyo polipéptido tiene una longitud entre 900 y 1180 aminoácidos. Los polipéptidos descritos en la presente memoria son sorprendentemente capaces de producir galactooligosacáridos in situ cuando se incuban con lactosa tal como leche. Por lo tanto, cuando el polipéptido, como se describe en la presente memoria, o una célula anfitriona que expresa el polipéptido se incuba con lactosa bajo condiciones apropiadas, se producen galactooligosacáridos con una alta eficiencia y, por lo tanto, se reduce la lactosa. La presencia de galactooligosacáridos en productos lácteos u otros productos comestibles tiene la ventaja de potenciar el crecimiento de *Bifdobacterium sp.* promotora de la salud en el producto o en la flora intestinal del consumidor después de la ingesta del producto o de ambos.

También describimos un polipéptido aislado que tiene actividad transgalactosilante seleccionada del grupo que consiste en:

a. un polipéptido que comprende una secuencia de aminoácidos que tiene al menos 66% de identidad de secuencia con la secuencia de aminoácidos del polipéptido maduro de SEQ ID NO: 1,

b. un polipéptido que comprende una secuencia de aminoácidos que tiene al menos 66% de identidad de secuencia con la secuencia de aminoácidos del polipéptido maduro de SEQ ID NO: 2,

c. un polipéptido codificado por un polinucleótido que se hibrida bajo al menos condiciones de baja rigurosidad con i) la secuencia de ácido nucleico comprendida en la SEQ ID NO: 10 que codifica el polipéptido maduro de SEQ ID NO: 1; ii) la secuencia de ADNc de i) o iii) la cadena complementaria de i) o ii);

d. un polipéptido codificado por un polinucleótido que se hibrida bajo al menos condiciones de baja rigurosidad con i) la secuencia de ácido nucleico comprendida en la SEQ ID NO: 11 que codifica el polipéptido maduro de SEQ ID NO: 2; ii) la secuencia de ADNc de i) o iii) la cadena complementaria de i) o ii);

e. un polipéptido que comprende una sustitución conservadora, deleción y/o inserción de uno o más aminoácidos de SEQ ID NO: 1, y

f. un polipéptido que comprende una sustitución conservadora, deleción y/o inserción de uno o más aminoácidos de SEQ ID NO: 2,

siempre que el polipéptido de los puntos anteriores a, c, y e tenga como máximo una longitud de 1806 aminoácidos y siempre que el polipéptido de los puntos anteriores b, d, y f tenga como máximo una longitud de 1767 aminoácidos.

En un aspecto, se describe en la presente memoria un método para expresar un polipéptido, comprendiendo el método obtener una célula como se describe en la presente memoria y expresar el polipéptido en la célula, y opcionalmente purificar el polipéptido. En un aspecto adicional, se describe en la presente memoria una composición que comprende un polipéptido como se describe en la presente memoria, preferiblemente una composición alimentaria, más preferiblemente un producto lácteo. En un aspecto adicional, se describe en la presente memoria un método para producir un producto alimentario tratando un sustrato que comprende lactosa con un polipéptido como se describe en la presente memoria tal como producir un producto lácteo tratando un sustrato a base de leche que comprende lactosa con un polipéptido como se describe en la presente memoria. En un aspecto adicional, los polipéptidos se utilizan para tratar un sustrato con una beta-galactosidasa hidrolizante. En un aspecto adicional, se describe en la presente memoria un producto alimentario, preferiblemente un producto lácteo, que comprende una enzima transgalactosilante obtenida a partir de *Ruminococcus hansenii* o *Ruminococcus lactaris*, preferiblemente como se define en los puntos a-f anteriores, y más preferiblemente un polipéptido como se define adicionalmente en la presente memoria. En un aspecto adicional, se describe en la presente memoria un galactooligosacárido o composición del mismo obtenido tratando un sustrato que comprende lactosa con un polipéptido como se describe en la presente memoria.

En un aspecto, se proporciona un polipéptido que tiene actividad transgalactosilante que comprende una secuencia de aminoácidos que tiene

a. al menos 66% de identidad de secuencia con la secuencia de aminoácidos de SEQ ID NO: 1, y/o

b. al menos 66% de identidad de secuencia con la secuencia de aminoácidos de SEQ ID NO: 2.

En otro aspecto, se proporciona un polipéptido que tiene una relación entre actividad transgalactosilante y actividad β-galactosidasa de al menos 1 medida a una concentración de 6 LAU/ml en un ensayo basado en leche a 37ºC y 5% p/p de lactosa después de 30 minutos de reacción que comprende una secuencia de aminoácidos que tiene

a. al menos 66% de identidad de secuencia con la secuencia de aminoácidos de SEQ ID NO: 1, y/o

b. al menos 66% de identidad de secuencia con la secuencia de aminoácidos de SEQ ID NO:2.

En un aspecto adicional, se proporciona un polipéptido que comprende una secuencia de aminoácidos que tiene al menos 60% de identidad de secuencia con la secuencia de aminoácidos de SEQ ID NO: 5. En un aspecto adicional, se proporciona un polipéptido que comprende una secuencia de aminoácidos que tiene al menos 94% de identidad de secuencia con la secuencia de aminoácidos de SEQ ID NO: 8. En un aspecto adicional, se proporciona el uso de un polipéptido que tiene actividad transgalactosilante que comprende una secuencia de aminoácidos que tiene al menos 60% de identidad de secuencia con la secuencia de aminoácidos de SEQ ID NO: 5 para producir galactooligosacáridos. En un aspecto adicional, se proporciona el uso de un polipéptido que tiene actividad transgalactosilante que comprende una secuencia de aminoácidos que tiene al menos 94% de identidad de secuencia con la secuencia de aminoácidos de SEQ ID NO: 8 para producir galactooligosacáridos.

En un aspecto adicional, se proporciona el uso de un polipéptido que tiene actividad transgalactosilante que comprende una secuencia de aminoácidos que tiene

a. al menos 60% de identidad de secuencia con la secuencia de aminoácidos de SEQ ID NO: 3,

b. al menos 60% de identidad de secuencia con la secuencia de aminoácidos de SEQ ID NO: 4,

c. al menos 60% de identidad de secuencia con la secuencia de aminoácidos de SEQ ID NO: 6, o

d. al menos 60% de identidad de secuencia con la secuencia de aminoácidos de SEQ ID NO: 7,

para producir galactooligosacáridos.

En un aspecto adicional, se proporciona un polipéptido que tiene una relación entre actividad transgalactosilante y actividad β-galactosidasa de al menos 1 medida a una concentración de 6 LAU/ml en un ensayo basado en leche a 37ºC y 5% p/p de lactosa después de 30 minutos de reacción. En un aspecto adicional, se proporciona un ácido nucleico capaz de codificar un polipéptido como se describe en la presente memoria. En un aspecto adicional, se proporciona un plásmido que comprende un ácido nucleico como se describe en la presente memoria. En un aspecto adicional, se proporciona un vector de expresión que comprende un ácido nucleico como se describe en la presente memoria, o capaz de expresar un polipéptido como se describe en la presente memoria. En un aspecto adicional, se proporciona una célula anfitriona que comprende, y preferiblemente transformada con, un plásmido como se describe en la presente memoria, o un vector de expresión como se describe en la presente memoria. En un aspecto adicional, se proporciona una célula capaz de expresar un polipéptido como se describe en la presente memoria. En un aspecto adicional, se proporciona un método para expresar un polipéptido, comprendiendo el método obtener una célula anfitriona o una célula como se describe en la presente memoria y la expresión del polipéptido a partir de la célula o célula anfitriona, y opcionalmente la purificación del polipéptido. En un aspecto adicional, se proporciona una composición que comprende un polipéptido como se describe en la presente memoria y un estabilizador. En un aspecto adicional, se proporciona una composición que comprende un polipéptido como se describe en la presente memoria y un sustrato de hidrato de carbono. En un aspecto adicional, se proporciona un método para producir un producto lácteo mediante tratamiento de un sustrato a base de leche que comprende lactosa con un polipéptido que tiene una relación entre actividad transgalactosilante y actividad β-galactosidasa de al menos 1, al menos 2,5, al menos 3, al menos 4, al menos 5, al menos 6, al menos 7, al menos 8, al menos 9, al menos 10, al menos 11 o al menos 12, medido a una concentración de 6 LAU/ml en un ensayo basado en leche a 37ºC y 5% p/p de lactosa después de 30 minutos de reacción. En un aspecto adicional, se proporciona un método para producir un producto lácteo mediante tratamiento de un sustrato a base de leche que comprende lactosa con un polipéptido como se describe en la presente memoria. En un aspecto adicional, se proporciona un uso de una célula como se describe en la presente memoria para producir un producto seleccionado del grupo que consiste en yogur, queso, producto de leche fermentada, suplemento dietético y producto comestible probiótico. En un aspecto adicional, se proporciona un producto lácteo que comprende una célula como se describe en la presente memoria. En un aspecto adicional, se proporciona un producto lácteo que comprende un polipéptido como se describe en la presente memoria. En un aspecto adicional, se proporciona un producto lácteo que comprende un polipéptido como se describe en la presente memoria en una concentración de 0,01-1000 ppm. En un aspecto adicional, se proporciona un producto lácteo que comprende un polipéptido inactivado como se describe en la presente memoria. En un aspecto adicional, se proporciona un producto lácteo que comprende GOS formado in situ mediante un polipéptido como se describe en la presente memoria. En un aspecto adicional, se proporciona un uso de un polipéptido transgalactosilante como se describe en la presente memoria o una célula como se describe en la

presente memoria para producir galactooligosacáridos. En un aspecto adicional, se proporciona un uso de un polipéptido transgalactosilante descrito en la presente memoria o una célula como se describe en la presente memoria, para producir galactooligosacáridos para ser parte de un producto seleccionado del grupo que consiste en yogur, queso, productos lácteos fermentados, suplementos dietéticos y productos comestibles probióticos. En un aspecto adicional, se proporciona un uso de un polipéptido transgalactosilante como se describe en la presente memoria o una célula como se describe en la presente memoria, para producir galactooligosacáridos para potenciar el crecimiento de Bifidobacterium. En un aspecto adicional, se proporciona un uso de un polipéptido transgalactosilante como se describe en la presente memoria o una célula como se describe en la presente memoria, para producir galactooligosacáridos para potenciar el crecimiento de Bifidobacterium en una fermentación con cultivo mixto. En un aspecto adicional, se proporciona un método para producir un polipéptido transgalactosilante como se describe en la presente memoria, que comprende cultivar una célula como se describe en la presente memoria en un medio de cultivo adecuado bajo condiciones que permiten la expresión de dicho polipéptido, y recuperar el polipéptido resultante del cultivo. En un aspecto adicional, un método para producir galactooligosacáridos, que comprende el contacto de un polipéptido como se describe en la presente memoria o una célula como se describe en la presente memoria con una solución a base de leche que comprende lactosa.

Lista de secuencias

La SEQ ID NO: 1 es un fragmento truncado de 1125 aminoácidos de la SEQ ID NO: 12.

La SEQ ID NO: 2 es un fragmento truncado de 1150 aminoácidos de la SEQ ID NO: 13.

La SEQ ID NO: 3 es los restos de aminoácidos 559-649 de la SEQ ID NO: 1.

La SEQ ID NO: 4 es los restos de aminoácidos 579-649 de la SEQ ID NO: 1.

La SEQ ID NO: 5 es los restos de aminoácidos 579-636 de la SEQ ID NO: 1.

La SEQ ID NO: 6 es los restos de aminoácidos 575-665 de la SEQ ID NO: 2.

La SEQ ID NO: 7 es los restos de aminoácidos 594-665 de la SEQ ID NO: 2.

La SEQ ID NO: 8 es los restos de aminoácidos 594-652 de la SEQ ID NO: 2.

La SEQ ID NO: 9 es un péptido señal del vector de expresión de Bacillus subtilis pBN.

La SEQ ID NO: 10 es la secuencia de nucleótidos que codifica la SEQ ID NO: 1 que incluye la secuencia que codifica el péptido señal.

La SEQ ID NO: 11 es la secuencia de nucleótidos que codifica la SEQ ID NO: 2 que incluye la secuencia que codifica el péptido señal.

La SEQ ID NO: 12 es una beta-galactosidasa de Ruminococcus/Blautia hansenii DSM 20583.

La SEQ ID NO: 13 es una glucosidasa de Ruminococcus lactaris ATCC 29176.

SEQ ID NO: 14 es la secuencia de nucleótidos que codifica la SEQ ID NO: 12 sin la secuencia señal.

SEQ ID NO: 15 es la secuencia de nucleótidos que codifica la SEQ ID NO: 13 sin la secuencia señal.

La SEQ ID NO: 16 es la secuencia de nucleótidos que codifica la SEQ ID NO: 1.

La SEQ ID NO: 17 es la secuencia de nucleótidos que codifica la SEQ ID NO: 2.

Leyendas de las figuras

La Fig. 1 muestra un mapa de plásmido de la construcción de expresión de *Ruminococcus hansenii*. La secuencia codificante de rhBIF3d3 se fusionó en fase con la secuencia señal aprE utilizando BssHII y PacI como sitios de restricción.

La Fig. 2 muestra la acumulación de galactosa y glucosa durante el tratamiento enzimático de una solución de lactosa al 5% p/p en tampón T con Lactozym® como control, *Ruminococcus hansenii* (SEQ ID NO: 1), *Ruminococcus lactaris* (SEQ ID NO: 2) y *Bifidobacterium bifidum* BIF3d3 (truncado) (como se describe por Jørgensen et al. (2001), Appl. Microbiol. Biotechnol., 57: 647 - 652 y patente EP 1.283.876).

La Fig. 3 muestra el resultado de la cromatografía en capa fina de los polipéptidos en 9% p/p de leche reconstituida que proporciona una concentración final de lactosa de 5% p/p. Los polipéptidos se dosificaron en base a la actividad LAU determinada como se describe en el ejemplo 1 a una concentración final de 6 LAU/ml.

La Fig. 4 muestra los resultados de la cromatografía de intercambio aniónico de variantes de la Ruminococcus hansenii (SEQ ID NO: 1). El gel es un Nu-PAGE 4 - gel de acrilamida al 12% teñido con tinción azul brillante de coomassie.

Descripción detallada de la invención

En la presente memoria se describe un polipéptido aislado que tiene actividad transgalactosilante que comprende una secuencia de aminoácidos que tiene al menos 80% de identidad de secuencia con la secuencia de aminoácidos del polipéptido maduro de SEQ ID NO: 1, cuyo polipéptido tiene una longitud entre 900 y 1180 aminoácidos.

En la presente memoria se describe un polipéptido aislado que tiene actividad transgalactosilante seleccionada del grupo que consiste en:

a. un polipéptido que comprende una secuencia de aminoácidos que tiene al menos 66% de identidad de secuencia con la secuencia de aminoácidos del polipéptido maduro de SEQ ID NO: 1,

b. un polipéptido que comprende una secuencia de aminoácidos que tiene al menos 66% de identidad de secuencia con la secuencia de aminoácidos del polipéptido maduro de SEQ ID NO: 2,

c. un polipéptido codificado por un polinucleótido que se hibrida bajo al menos condiciones de baja rigurosidad con i) la secuencia de ácido nucleico comprendida en la SEQ ID NO: 10 que codifica el polipéptido maduro de SEQ ID NO: 1; ii) la secuencia de ADNc de i) o iii) la cadena complementaria de i) o ii);

d. un polipéptido codificado por un polinucleótido que se hibrida bajo al menos condiciones de baja rigurosidad con i) la secuencia de ácido nucleico comprendida en la SEQ ID NO: 11 que codifica el polipéptido maduro de SEQ ID NO: 2; ii) la secuencia de ADNc de i) o iii) la cadena complementaria de i) o ii);

e. un polipéptido que comprende una sustitución conservadora, deleción y/o inserción de uno o más aminoácidos de SEQ ID NO: 1, y

f. un polipéptido que comprende una sustitución conservadora, deleción y/o inserción de uno o más aminoácidos de SEQ ID NO: 2,

siempre que el polipéptido de los puntos anteriores a, c, y e tenga como máximo una longitud de 1806 aminoácidos y siempre que el polipéptido de los puntos b, d, y f anteriores tenga como máximo una longitud de 1767 aminoácidos.

De acuerdo con esta descripción detallada, se aplican las siguientes abreviaturas y definiciones. Se debe observar que tal como se utiliza en la presente memoria, las formas singulares "un", "una" y "el" incluyen referencias plurales a menos que el contexto indique claramente lo contrario. Por lo tanto, por ejemplo, la referencia a "un polipéptido" incluye una pluralidad de tales polipéptidos, y la referencia a "la formulación" incluye la referencia a una o más formulaciones y equivalentes de los mismos conocidos por los expertos en la técnica, y así sucesivamente.

A menos que se defina lo contrario, todos los términos técnicos y científicos utilizados en la presente memoria tienen el mismo significado como se entiende comúnmente por un experto en la técnica. Los siguientes términos se proporcionan a continuación.

"Transgalactosilasa" significa una enzima que, entre otras cosas, es capaz de transferir galactosa a los grupos hidroxilo de D-galactosa o D-glucosa por lo que se producen galactooligosacáridos. En un aspecto, se identifica una transgalactosilasa mediante reacción de la enzima sobre lactosa en la que la cantidad de galactosa generada es menor que la cantidad de glucosa generada en cualquier momento dado.

En el presente contexto, el término "actividad transgalactosilante" significa la transferencia de una parte de galactosa a una molécula distinta de agua y se mide como [glucosa] - [galactosa] generada en cualquier momento dado durante la reacción.

En el presente contexto, el término "actividad β-galactosidasa" significa la capacidad de una enzima para hidrolizar β-galactósidos tal como, por ejemplo, lactosa en monosacáridos, glucosa y galactosa.

En el presente contexto, el término "actividad de transgalactosilación relativa" significa ([glucosa] - [galactosa]*100)/[glucosa]) medida en un punto de tiempo después de 15 minutos de tiempo de reacción.

En el presente contexto, el término [glucosa] significa la concentración de glucosa en % en peso medida mediante HPLC.

En el presente contexto, el término [galactosa] significa la concentración de galactosa en % en peso medida mediante HPLC.

En el presente contexto, el término "después de 15 min de reacción" significa la cantidad de tiempo que debe pasar antes de la medición de la actividad después de la incubación con el polipéptido descrito en la presente memoria en un ensayo.

En un aspecto, la actividad se mide después de 15 min de reacción, 30 min de reacción, 60 min de reacción, 90 min de reacción, 120 min de reacción o 180 min de reacción. Por lo tanto, en un aspecto, como ejemplo la actividad transgalactosilante relativa es medida 15 minutos después de la adición de la enzima, tal como 30 minutos después de la adición de la enzima, tal como 60 minutos después de la adición de la enzima, tal como 90 minutos después de la adición de la enzima, tal como 120 minutos después de la adición de la enzima o 180 minutos después de la adición de la enzima.

En el presente contexto, el término "relación de actividad transgalactosilante:actividad β-galactosidasa" significa ([glucosa]-[galactosa]/[galactosa]).

En el presente contexto, el término "lactosa ha sido transgalactosilada" significa que una molécula de galactosa se ha unido covalentemente a la molécula de lactosa tal como, por ejemplo, unida covalentemente a cualquiera de los grupos hidroxilo libres en la molécula de lactosa o generada por transgalatosilación interna por ejemplo formando alolactosa.

En el presente contexto, el término "ensayo basado en leche" significa un ensayo realizado en leche, leche reconstituida o soluciones que contienen constituyentes de leche principales tales como, por ejemplo, lactosa. Se puede realizar un ensayo basado en leche preparando muestras en leche reconstituida al 9% p/p a partir de leche en polvo desnatada (tal como, p. ej., Humana Milk Union, DE NW508 EG) dando una concentración final de lactosa de 5% p/p. Las enzimas se dosifican basándose en la actividad de LAU determinada como se describe a continuación, dando la concentración final deseada en LAU/ml.

Se toma una muestra antes de la adición de la enzima y se toman muestras adicionales en puntos de tiempo indicados y las enzimas se inactivan inmediatamente incubando a 95°C durante 10 minutos. Las muestras se diluyen 1:10 y se aplican 2 µl sobre placas de gel de sílice 60 HPTLC activadas (161ºC durante 10 min) (Merck Cat # 1.05641.0001) con un CAMAG Automatic TLC Sampler 4. Las placas de TLC se eluyen con un eluyente que contiene (80) acetonitrilo: (20) acetato de etilo: (50) 1-propanol: (40) agua. Las muestras se visualizan mediante calentamiento (161ºC durante 10 min) y se dejan enfriar antes de sumergirlas en H2SO4 al 5% p/p en etanol al 99,9% p/p. Las placas se desarrollan con calentamiento a 161ºC durante 3 min.

En un aspecto, tal ensayo es como se describe en el ejemplo 3.

En el contexto de la presente solicitud, 1 unidad de lactasa (1 LAU) es la cantidad de enzima que libera 1 micromol de glucosa por minuto en tampón M a pH 6,5 y 37ºC con una concentración de lactosa de 4,75% p/v. El tampón M se prepara disolviendo 3,98 g de $C_6H_5Na_3O_7$-5 $2H_2O$, 8,31 g de ácido cítrico, 0,9 g de K_2SO_4, 2,6 g de K_2HPO_4, 7,35 g de KH_2PO_4, 5,45 g de KOH, 4,15 g, $MgCl_2$ $6H_2O$, 3,75 g de $CaCl_2$ $2H_2O$ y 1,4 g de $NaHCO_3$ en 4 litros de agua, añadiendo 12,5 ml de NaOH 4N, ajustando a pH 6,5 utilizando HCl, y agregando agua hasta un volumen total de 5 litros.

La actividad en LAU de una lactasa específica se puede determinar mediante medición directa de la glucosa liberada a partir de lactosa bajo las condiciones descritas anteriormente. El experto sabrá cómo determinar tal actividad. Alternativamente, la actividad se puede determinar utilizando el ensayo de actividad de lactasa descrito en el Ejemplo 1 de la presente solicitud. Aquí, la actividad se obtiene comparando con una curva estándar con una lactasa de actividad conocida, y la actividad de la muestra desconocida se calcula a partir de esto. La lactasa de actividad conocida puede, p. ej., ser Lactozym obtenida de Novozymes A/S, Dinamarca.

En el presente contexto, el término "cuyo polipéptido está liofilizado" significa que el polipéptido se ha obtenido liofilizando un líquido del polipéptido a una presión apropiada y durante un período apropiado retirando el agua.

En el presente contexto, el término "cuyo polipéptido está en solución" se refiere a un polipéptido que es soluble en un disolvente sin precipitación de la solución. Un disolvente para este fin incluye cualquier medio en el que pueda aparecer el polipéptido, tal como un tampón acuoso o una solución salina, un caldo de fermentación o el citoplasma de un anfitrión de expresión.

En el presente contexto, el término "estabilizador" significa cualquier estabilizador para estabilizar el polipéptido, p. ej., un poliol tal como, p. ej., glicerol o propilenglicol, un azúcar o un alcohol de azúcar, ácido láctico, ácido bórico o un derivado de ácido bórico (p. ej., un éster de borato aromático). En un aspecto, el estabilizador es glicerol.

En el presente contexto, el término "sustrato de hidrato de carbono" significa un compuesto orgánico con la fórmula general $C_m(H_2O)_n$, es decir, que consiste únicamente en carbono, hidrógeno y oxígeno, los dos últimos en la proporción de átomos 2:1, tal como un disacárido.

En el presente contexto, el término "disacárido" es dos unidades monosacáridas unidas por un enlace covalente conocido como un enlace glucosídico formado mediante una reacción de deshidratación, que da como resultado la

pérdida de un átomo de hidrógeno de un monosacárido y un grupo hidroxilo del otro. La fórmula de disacáridos no modificados es $C_{12}H_{22}O_{11}$. En un aspecto, el disacárido es lactulosa, trehalosa, ramnosa, maltosa, sacarosa, lactosa o celobiosa. En un aspecto adicional, el disacárido es lactosa.

El término "aislado" significa que la secuencia está al menos sustancialmente libre de al menos otro componente con el que la secuencia está asociado de forma natural en la naturaleza y tal como se encuentra en la naturaleza. En un aspecto, "polipéptido aislado" como se utiliza en la presente memoria se refiere a un polipéptido que es al menos 30% puro, al menos 40% puro, al menos 60% puro, al menos 80% puro, al menos 90% puro, y al menos 95% puro, según se determinó mediante SDS-PAGE.

El término "polipéptido sustancialmente puro" significa en la presente memoria una preparación de polipéptido que contiene como máximo 10%, preferiblemente como máximo 8%, más preferiblemente como máximo 6%, más preferiblemente como máximo 5%, más preferiblemente como máximo 4%, como máximo 3% , incluso más preferiblemente como máximo 2%, más preferiblemente como máximo 1%, e incluso más preferiblemente como máximo 0,5% en peso de otro material polipeptídico con el que está asociado de forma natural. Por lo tanto, se prefiere que el polipéptido sustancialmente puro sea al menos 92% puro, preferiblemente al menos 94% puro, más preferiblemente al menos 95% puro, más preferiblemente al menos 96% puro, más preferiblemente al menos 96% puro, más preferiblemente al menos 97% puro, más preferiblemente al menos 98% puro, incluso más preferiblemente al menos 99%, lo más preferiblemente al menos 99,5% puro, e incluso lo más preferiblemente 100% puro en peso del material polipeptídico total presente en la preparación. Los polipéptidos descritos en la presente memoria están preferiblemente en una forma sustancialmente pura. En particular, se prefiere que los polipéptidos estén en "forma esencialmente pura", es decir, que la preparación del polipéptido esté esencialmente libre de otro material polipeptídico con el que esté asociado de forma natural. Esto se puede lograr, por ejemplo, preparando el polipéptido por medio de métodos recombinantes bien conocidos o mediante métodos de purificación clásicos. En la presente memoria, el término "polipéptido sustancialmente puro" es sinónimo de los términos "polipéptido aislado" y "polipéptido en forma aislada".

El término "purificado" o "puro" significa que un componente dado está presente en un estado de alto nivel p. ej. al menos aproximadamente 51% puro, o al menos aproximadamente 75%, o al menos aproximadamente 80%, o al menos aproximadamente 90% puro, o al menos aproximadamente 95% puro o al menos aproximadamente 98% puro. El componente es deseablemente el componente activo predominante presente en una composición.

El término "microorganismo" en el presente contexto incluye cualquier microorganismo que podría comprender una secuencia de nucleótidos que codifica un polipéptido que tiene las propiedades específicas como se definen en la presente memoria y/o productos obtenidos del mismo.

En el presente contexto, "microorganismo" puede incluir cualquier bacteria u hongo que sea capaz de fermentar un sustrato de leche.

El término "célula anfitriona" - en el presente contexto incluye cualquier célula que comprende una secuencia de nucleótidos que codifica un polipéptido que tiene las propiedades específicas como se define en la presente memoria o un vector de expresión como se describió anteriormente y que se utiliza en la producción de un polipéptido que tiene las propiedades específicas como se define en la presente memoria. En un aspecto, la producción es producción recombinante.

El término "leche", en el contexto de la presente invención, se debe entender como la secreción láctea obtenida a partir de cualquier mamífero, tal como vacas, ovejas, cabras, búfalos o camellos.

En el presente contexto, el término "sustrato a base de leche" significa cualquier material de leche cruda y/o procesada o un material derivado de constituyentes de leche. Los sustratos útiles a base de leche incluyen pero no se limitan a soluciones/suspensiones de cualquier leche o productos lácteos que contengan lactosa, tal como leche entera o baja en grasa, leche desnatada, suero de leche, leche en polvo reconstituida, leche condensada, soluciones de leche en polvo, leche UHT, suero, filtrado de suero, suero ácido o nata. Preferiblemente, el sustrato a base de leche es leche o una solución acuosa de leche desnatada en polvo. El sustrato a base de leche puede estar más concentrado que la leche cruda. El sustrato a base de leche puede tener una proporción de proteína a lactosa de al menos 0,2, preferiblemente al menos 0,3, al menos 0,4, al menos 0,5, al menos 0,6 o, lo más preferiblemente, al menos 0,7. El sustrato a base de leche se puede homogeneizar y/o pasteurizar según métodos conocidos en la técnica.

"Homogenizar" como se utiliza en la presente memoria significa un mezclado intensivo para obtener una suspensión o emulsión soluble. Se puede realizar para dividir la grasa de la leche en tamaños más pequeños para que ya no se separe de la leche. Esto se puede lograr forzando la leche a alta presión a través de pequeños orificios.

"Pasteurizar" como se utiliza en la presente memoria significa reducir o eliminar la presencia de organismos vivos, tales como microorganismos, en el sustrato a base de leche. Preferiblemente, la pasteurización se logra manteniendo una temperatura especificada durante un período de tiempo específico. La temperatura especificada generalmente se alcanza calentando. La temperatura y duración se pueden seleccionar para matar o inactivar

ciertas bacterias, tales como bacterias dañinas, y/o para inactivar enzimas en la leche. Una etapa de enfriamiento rápido puede seguir.

Un "producto alimentario" o "composición alimentaria" en el presente contexto puede ser cualquier alimento comestible o producto alimentario adecuado para el consumo por un animal o ser humano.

Un "producto lácteo" en el presente contexto puede ser cualquier producto alimentario en donde uno de los principales constituyentes esté basado en leche. Preferible, el componente principal es a base de leche. Más preferiblemente, el constituyente principal es un sustrato a base de leche que se ha tratado con una enzima que tiene actividad transgalactosilante.

En el presente contexto, "uno de los principales constituyentes" significa un constituyente que tiene una materia seca que constituye más del 20%, preferiblemente más del 30% o más del 40% de la materia seca total del producto lácteo, mientras que "el principal constituyente" significa un constituyente que tiene una materia seca que constituye más del 50%, preferiblemente más del 60% o más del 70% de la materia seca total del producto lácteo.

Un "producto lácteo fermentado" en el presente contexto se debe entender como cualquier producto lácteo en donde cualquier tipo de fermentación forma parte del proceso de producción. Ejemplos de productos lácteos fermentados son productos como yogur, suero de leche, nata fresca, requesón y queso fresco. Un producto lácteo fermentado se puede producir mediante cualquier método conocido en la técnica.

El término "fermentación" significa la conversión de hidratos de carbono en alcoholes o ácidos a través de la acción de un microorganismo tal como un cultivo iniciador. En un aspecto, la fermentación comprende la conversión de lactosa en ácido láctico.

En el presente contexto, el término "dominios Pfam" significa regiones dentro de una secuencia proteica que se identifican como Pfam-A o Pfam-B basadas en alineamientos de secuencias múltiples y la presencia de Hidden Markov Motifs ("The Pfam protein families database": RD Finn, J. Mistry, J. Tate, P. Coggill, A. Heger, JE Pollington, OL Gavin, P. Gunesekaran, G. Ceric, K. Forslund, L. Holm, EL Sonnhammer, SR Eddy, A. Bateman Nucleic Acids Research (2010) Database Issue 38: D211-222.). Como ejemplos de dominios Pfam, se pueden mencionar Glyco_hydro2N (PF02837), Glyco_hydro (PF00703), Glyco_hydro 2C (PF02836) y el dominio similar a Ig bacteriana (grupo 4) (PF07532).

Como se utiliza en la presente memoria "una posición que corresponde a la posición" significa que se realiza un alineamiento como se describe en la presente memoria entre un polipéptido pregunta particular y el polipéptido de referencia. La posición que corresponde a una posición específica en el polipéptido de referencia se identifica luego como el aminoácido correspondiente en el alineamiento con la identidad de secuencia más alta.

En un aspecto, se proporciona un polipéptido que tiene actividad transgalactosilante que comprende una secuencia de aminoácido que tiene

a. al menos 66% de identidad de secuencia con la secuencia de aminoácido de SEQ ID NO: 1, y/o

b. al menos 66% de identidad de secuencia con la secuencia de aminoácido de SEQ ID NO: 2.

En un aspecto, se proporciona un polipéptido, en donde la secuencia de aminoácido comprende al menos uno o más restos de aminoácidos seleccionados de los siguientes grupos:

a. un resto de aminoácido seleccionado del grupo que consiste en D/E/N/Q en una posición que corresponde a la posición 576 en la SEQ ID NO: 1,

b. un resto de aminoácido seleccionado del grupo que consiste en D/E/N/Q en una posición que corresponde a la posición 588 en la SEQ ID NO: 1,

c. un resto de aminoácido seleccionado del grupo que consiste en E/D/Q/N en una posición que corresponde a la posición 592 en la SEQ ID NO: 1 y/o

d. un resto de aminoácido seleccionado del grupo que consiste en D/E/Q/N en una posición que corresponde a la posición 625 en la SEQ ID NO: 1.

En un aspecto, se proporciona un polipéptido, en donde la secuencia de aminoácido comprende al menos uno o más restos de aminoácidos seleccionados de los siguientes grupos:

a. un resto de aminoácido seleccionado del grupo que consiste en D/E/N/Q en una posición que corresponde a la posición 592 en la SEQ ID NO: 2,

b. un resto de aminoácido seleccionado del grupo que consiste en D/E/N/Q en una posición que corresponde a la posición 604 en la SEQ ID NO: 2,

c. un resto de aminoácido seleccionado del grupo que consiste en E/D/Q/N en una posición que corresponde a la posición 608 en la SEQ ID NO: 2 y/o

d. un resto de aminoácido seleccionado del grupo que consiste en D/E/Q/N en una posición que corresponde a la posición 641 en la SEQ ID NO: 2.

Se ha encontrado que el aminoácido en una posición que corresponde a la posición 576, 588, 592 y 625 en la SEQ ID NO: 1 y los respectivos aminoácidos en una posición que corresponde a la posición 592, 604, 608 y 641 en la SEQ ID NO: 2 tienen un efecto sobre la actividad de los polipéptidos descritos en la presente memoria.

En un aspecto, se describe en la presente memoria un polipéptido, en donde la secuencia de aminoácidos comprende al menos uno o más restos de aminoácidos ácidos tales como D o E, en una posición que corresponde a la posición 576, 588, 592 y 625 en la SEQ ID NO: 1 o en una posición que corresponde a la posición 592, 604, 608 y 641 en la SEQ ID NO: 2.

En otro aspecto, describimos un polipéptido que tiene una relación entre actividad transgalactosilante y actividad β-galactosidasa de al menos 1 medida a una concentración de 6 LAU/ml en un ensayo basado en leche a 37°C y 5% p/p de lactosa después de 30 minutos de reacción que comprende una secuencia de aminoácidos que tiene

a. al menos 66% de identidad de secuencia con la secuencia de aminoácidos de SEQ ID NO: 1, y/o

b. al menos 66% de identidad de secuencia con la secuencia de aminoácidos de SEQ ID NO: 2.

En un aspecto adicional, se proporciona un polipéptido que comprende una secuencia de aminoácidos que tiene al menos 60% de identidad de secuencia con la secuencia de aminoácidos de SEQ ID NO: 5. En un aspecto adicional, se proporciona un polipéptido que comprende una secuencia de aminoácidos que tiene al menos 94% de identidad de secuencia con la secuencia de aminoácidos de SEQ ID NO: 8. En un aspecto adicional, se proporciona el uso de un polipéptido que tiene actividad transgalactosilante que comprende una secuencia de aminoácidos que tiene al menos 60% de identidad de secuencia con la secuencia de aminoácidos de SEQ ID NO: 5 para producir galactooligosacáridos. En un aspecto adicional, se proporciona el uso de un polipéptido que tiene actividad transgalactosilante que comprende una secuencia de aminoácidos que tiene al menos 94% de identidad de secuencia con la secuencia de aminoácidos de SEQ ID NO: 8 para producir galactooligosacáridos.

En un aspecto adicional, se proporciona el uso de un polipéptido que tiene actividad transgalactosilante que comprende una secuencia de aminoácidos que tiene

a. al menos 60% de identidad de secuencia con la secuencia de aminoácidos de SEQ ID NO: 3,

b. al menos 60% de identidad de secuencia con la secuencia de aminoácidos de SEQ ID NO: 4,

c. al menos 60% de identidad de secuencia con la secuencia de aminoácidos de SEQ ID NO: 6, o

d. al menos 60% de identidad de secuencia con la secuencia de aminoácidos de SEQ ID NO: 7,

para producir galactoligosacáridos.

En un aspecto adicional, se proporciona un polipéptido que comprende una secuencia de aminoácidos que tiene al menos 66% de identidad de secuencia con la secuencia de aminoácidos de SEQ ID NO: 1, y al menos 60% de identidad de secuencia con la secuencia de aminoácidos de SEQ ID NO: 5.

En un aspecto adicional, se proporciona un polipéptido que comprende una secuencia de aminoácidos que tiene al menos 66% de identidad de secuencia con la secuencia de aminoácidos de SEQ ID NO: 2, y al menos 94% de identidad de secuencia con la secuencia de aminoácidos de SEQ ID NO: 8.

En un aspecto adicional, se proporciona un polipéptido que contiene uno o más dominios Pfam seleccionados de: Glyco_hydro2N (PF02837), Glyco_hydro (PF00703), Glyco_hydro 2C (PF02836) y dominio similar a Ig bacteriana (grupo 4) (PF07532). En un aspecto adicional, se proporciona un polipéptido que contiene los dominios Pfam Glyco_hydro2N (PF02837), Glyco_hydro (PF00703), Glyco_hydro 2C (PF02836) y el dominio similar a Ig bacteriano (grupo 4) (PF07532). En un aspecto adicional, se proporciona un polipéptido que contiene los dominios Glyco_hydro2N (PF02837), Glyco_hydro (PF00703) y Glyco_hydro 2C (PF02836) cuyos dominios constituyen el dominio catalítico del polipéptido.

En un aspecto adicional, se proporciona un polipéptido que comprende una secuencia de aminoácidos y que tiene una relación entre actividad transgalactosilante y actividad β-galactosidasa de al menos 1, al menos 2,5, al menos 3, al menos 4, al menos 5, al menos 6, en al menos 7, al menos 8, al menos 9, al menos 10, al menos 11 o al menos 12 medidos a una concentración de 6 LAU/ml en un ensayo basado en leche a 37ºC y 5% p/p de lactosa después de 15 o 30 minutos de reacción. En un aspecto adicional, el polipéptido se obtiene a partir de *Ruminococcus hansenii* o *Ruminococcus lactaris*.

En un aspecto, el o los polipéptidos descritos en la presente memoria tienen una actividad transgalactosilante tal que más del 20%, más del 30%, más del 40%, hasta el 50% de la lactosa inicial está transgalactosilada como se mide a una concentración de 6 LAU/ml en un ensayo basado en leche a 37ºC y 5% p/p de lactosa después de 30 minutos de reacción.

En un aspecto adicional, el o los polipéptidos descritos en la presente memoria tienen una actividad β-galactosidasa tal que menos del 80%, menos del 70%, menos del 60%, menos del 50%, menos del 40%, menos del 30%, menos del 20% de la lactosa se ha hidrolizado, como se mide a una concentración de 6 LAU/ml en un ensayo basado en leche a 37ºC y 5% p/p de lactosa después de 30 minutos de reacción.

En un aspecto, la actividad β-galactosidasa y/o la actividad transgalactosilante se miden a una concentración de 6 LAU/ml, 3 LAU/ml o 1 LAU/ml.

En un aspecto adicional, la presente memoria describe polipéptido o polipéptidos que tienen una o más de las siguientes características:

a) una relación entre actividad transgalactosilante y actividad β-galactosidasa de al menos 1, al menos 2,5, al menos 3, al menos 4, al menos 5, al menos 6, al menos 7, al menos 8, al menos 9, al menos 10, al menos 11, o al menos 12 como se mide a una concentración de 6 LAU/ml en un ensayo basado en leche a 37ºC y 5% p/p de lactosa después de 30 minutos de reacción, y/o

b) tiene una actividad transgalactosilante tal que más del 20%, más del 30%, más del 40% y hasta el 50% de la lactosa inicial se ha transgalactosilado como se mide a una concentración de 6 LAU/ml en un ensayo basado en leche en 37ºC y 5% p/p de lactosa después de 30 minutos de reacción.

En un aspecto adicional, se proporciona un polipéptido que comprende una secuencia de aminoácidos que tiene al menos 60% de identidad de secuencia con la secuencia de aminoácidos de SEQ ID NO: 5. En un aspecto adicional, se proporciona un polipéptido que comprende una secuencia de aminoácidos que tiene al menos 94% de identidad de secuencia con la secuencia de aminoácidos de SEQ ID NO: 8. En un aspecto adicional, se proporciona un polipéptido que comprende una secuencia de aminoácidos que tiene al menos 60% de identidad de secuencia con la secuencia de aminoácidos de SEQ ID NO: 3. En otro aspecto adicional, se proporciona un polipéptido que comprende una secuencia de aminoácidos que tiene al menos 60% de identidad de secuencia con la secuencia de aminoácidos de SEQ ID NO: 4. En un aspecto adicional, se proporciona un polipéptido que comprende una secuencia de aminoácidos que tiene al menos 94% de identidad de secuencia con la secuencia de aminoácidos de SEQ ID NO: 6. En un aspecto adicional, se proporciona un polipéptido que comprende una secuencia de aminoácidos que tiene al menos 94% de identidad de secuencia con la secuencia de aminoácidos de SEQ ID NO: 7.

Las proteínas generalmente están compuestas por una o más regiones funcionales, comúnmente denominadas dominios. La presencia de diferentes dominios en diversas combinaciones en diferentes proteínas da lugar al repertorio diverso de proteínas que se encuentran en la naturaleza. Una forma de describir los dominios es con la ayuda de la base de datos Pfam que es una gran colección de familias de dominios de proteínas como se describe en "The Pfam protein families database": RD Finn, J. Mistry, J. Tate, P. Coggill, A Heger, JE Pollington, OL Gavin, P. Gunesekaran, G. Ceric, K. Forslund, L. Holm, E.L. Sonnhammer, S.R. Eddy, A. Bateman Nucleic Acids Research (2010) Database Issue 38: D211-222. Cada familia está representada por alineamientos de secuencias múltiples y modelos de Markov ocultos (HMM). En un aspecto adicional, los presentes autores han descubierto que el o los polipéptidos proporcionados en la presente memoria contienen uno o más de los dominios Pfam Glyco_hydro2N (PF02837), Glyco_hydro (PF00703), Glyco_hydro 2C (PF02836) y dominio similar a Ig bacteriana (grupo 4) (PF07532). En un aspecto, el o los polipéptidos proporcionados en la presente memoria contienen Glyco_hydro2N (PF02837), Glyco_hydro (PF00703), Glyco_hydro 2C (PF02836) y dominio similar a Ig bacteriana (grupo 4) (PF07532).

En un aspecto, el o los polipéptidos descritos en la presente memoria comprenden una secuencia de aminoácidos que tienen un resto de aminoácido seleccionado del grupo que consiste en D, E, N y Q en una posición que corresponde a la posición 576 en la SEQ ID NO: 1. En un aspecto, el o los polipéptidos descritos en la presente memoria comprenden una secuencia de aminoácidos que tiene un resto de aminoácido seleccionado del grupo que consiste en D, E y N en una posición que corresponde a la posición 576 en la SEQ ID NO: 1. En un aspecto, el o los polipéptidos descritos en la presente memoria comprenden una secuencia de aminoácidos que tiene un resto de aminoácido seleccionado del grupo que consiste en D y E en una posición que corresponde a la posición 576 en la SEQ ID NO: 1. En un aspecto, el o los polipéptidos descritos en la presente memoria comprenden una secuencia de aminoácidos cuyo resto de aminoácido D en una posición corresponde a la posición 576 en la SEQ ID NO: 1.

En un aspecto, el o los polipéptidos descritos en la presente memoria comprenden una secuencia de aminoácidos que tiene un resto de aminoácido seleccionado del grupo que consiste en D, E, N y Q en una posición que corresponde a la posición 588 en la SEQ ID NO: 1. En un aspecto, el o los polipéptidos descritos en la presente memoria comprenden una secuencia de aminoácidos que tiene un resto de aminoácido seleccionado del grupo que consiste en D, E y N en una posición que corresponde a la posición 588 en la SEQ ID NO: 1. En un aspecto, el o los polipéptidos descritos en la presente memoria comprenden una secuencia de aminoácidos que tiene un resto de

aminoácido seleccionado del grupo que consiste en D y E en una posición que corresponde a la posición 588 en la SEQ ID NO: 1. En un aspecto, el o los polipéptidos descritos en la presente memoria comprenden una secuencia de aminoácidos cuyo resto de aminoácido D en una posición corresponde a la posición 588 en la SEQ ID NO: 1.

En un aspecto, el o los polipéptidos descritos en la presente memoria comprenden una secuencia de aminoácidos que tiene un resto de aminoácido seleccionado del grupo que consiste en D, E, N y Q en una posición que corresponde a la posición 592 en la SEQ ID NO: 1. En un aspecto, el o los polipéptidos descritos en la presente memoria comprenden una secuencia de aminoácidos que tiene un resto de aminoácido seleccionado del grupo que consiste en D, E y Q en una posición que corresponde a la posición 592 en la SEQ ID NO: 1. En un aspecto, el o los polipéptidos descritos en la presente memoria comprenden una secuencia de aminoácidos que tiene un resto de aminoácido seleccionado del grupo que consiste en D y E en una posición que corresponde a la posición 592 en la SEQ ID NO: 1. En un aspecto, el o los polipéptidos descritos en la presente memoria comprenden una secuencia de aminoácidos cuyo resto de aminoácido E en una posición corresponde a la posición 592 en la SEQ ID NO: 1.

En un aspecto, el o los polipéptidos descritos en la presente memoria comprenden una secuencia de aminoácidos que tiene un resto de aminoácido seleccionado del grupo que consiste en D, E, N y Q en una posición que corresponde a la posición 625 en la SEQ ID NO: 1. En un aspecto, el o los polipéptidos descritos en la presente memoria comprenden una secuencia de aminoácidos que tiene un resto de aminoácido seleccionado del grupo que consiste en D, E y N en una posición que corresponde a la posición 625 en la SEQ ID NO: 1. En un aspecto, el o los polipéptidos descritos en la presente memoria comprenden una secuencia de aminoácidos que tiene un resto de aminoácido seleccionado del grupo que consiste en D y E en una posición que corresponde a la posición 625 en la SEQ ID NO: 1. En un aspecto, el o los polipéptidos descritos en la presente memoria comprenden una secuencia de aminoácidos cuyo resto de aminoácido D en una posición corresponde a la posición 625 en la SEQ ID NO: 1.

En un aspecto, los polipéptidos tienen actividad transgalactosilante útil en un intervalo de pH de 4-9, tal como 5-8, tal como 5,5-7,5.

Describimos polipéptidos que tienen un cierto grado de identidad de secuencia u homología de secuencia con la o las secuencias de aminoácidos definidas en la presente memoria o con un polipéptido que tiene las propiedades específicas definidas en la presente memoria. La presente invención abarca, péptidos que tienen un grado de identidad de secuencia de al menos 80% con la secuencia de aminoácidos del polipéptido maduro de SEQ ID NO: 1, cuyo polipéptido tiene una longitud entre 900 y 1180 aminoácidos.

En un aspecto, la secuencia de aminoácidos y/o la secuencia de nucleótidos homóloga debería proporcionar y/o codificar un polipéptido que retiene la actividad transgalactosilante funcional y/o potencia la actividad transgalactosilante en comparación con un polipéptido de SEQ ID NO: 1 o 2.

En el presente contexto, se toma una secuencia homóloga para incluir una secuencia de aminoácidos que puede ser al menos 66%, 70%, 75%, 78%, al menos 80%, al menos 85%, al menos 90%, al menos 95%, al menos 96%, al menos 97%, al menos 98% o al menos 99%, idéntica a la secuencia encontrada. Típicamente, los homólogos comprenderán los mismos sitios activos, etc., que la secuencia de aminoácidos encontrada. Aunque la homología también se puede considerar en términos de similitud (es decir, restos de aminoácidos que tienen propiedades/funciones químicas similares), en el presente contexto se prefiere expresar la homología en términos de identidad de secuencia.

Las comparaciones de identidad de secuencia se pueden realizar a simple vista, o más habitualmente, con la ayuda de programas de comparación de secuencias fácilmente disponibles. Estos programas informáticos disponibles en el mercado utilizan algoritmos de comparación complejos para alinear dos o más secuencias que reflejan mejor los eventos evolutivos que podrían haber conducido a la o las diferencias entre las dos o más secuencias. Por lo tanto, estos algoritmos operan con un sistema de puntuación que recompensa el alineamiento de aminoácidos idénticos o similares y penaliza la inserción de huecos, extensiones de huecos y el alineamiento de aminoácidos no similares. El sistema de puntuación de los algoritmos de comparación incluye:

i) asignación de una puntuación de penalización cada vez que se inserta un hueco (puntuación de penalización por huecos),

ii) asignación de una puntuación de penalización cada vez que se extiende un hueco existente con una posición adicional (puntuación de penalización por extensión),

iii) asignación de puntuaciones altas al alineamiento de aminoácidos idénticos, y

iv) asignación de puntuaciones variables al alineamiento de aminoácidos no idénticos.

La mayoría de los programas de alineamiento permiten modificar las penalizaciones por huecos. Sin embargo, se prefiere utilizar los valores predeterminados cuando se utiliza tal software para comparaciones de secuencias.

Las puntuaciones dadas para alineamiento de aminoácidos no idénticos se asignan según una matriz de puntuación también llamada matriz de sustitución. Las puntuaciones proporcionadas en tales matrices de sustitución están

reflejando el hecho de que la probabilidad de que un aminoácido se sustituya por otro durante la evolución varía y depende de la naturaleza física o química del aminoácido que se va a sustituir. Por ejemplo, la probabilidad de que un aminoácido polar se sustituya por otro aminoácido polar es mayor en comparación con que se sustituya por un aminoácido hidrófobo. Por lo tanto, la matriz de puntuación asignará la puntuación más alta para aminoácidos idénticos, una puntuación más baja para aminoácidos no idénticos pero similares e incluso una puntuación más baja para aminoácidos no idénticos no similares. Las matrices de puntuación utilizadas más frecuentemente son las matrices PAM (Dayhoff et al. (1978), Jones et al. (1992)), las matrices BLOSUM (Henikoff and Henikoff (1992)) y la matriz Gonnet (Gonnet et al. (1992)).

Los programas informáticos adecuados para llevar a cabo dicho alineamiento incluyen, pero no se limitan a, Vector NTI (Invitrogen Corp.) y los programas ClustalV, ClustalW y ClustalW2 (Higgins DG y Sharp PM (1988), Higgins et al. (1992), Thompson et al. (1994), Larkin et al. (2007). Se encuentra disponible una selección de diferentes herramientas de alineamiento del servidor ExPASy Proteomics en www.expasy.org. Otro ejemplo de software que puede realizar el alineamiento de secuencias es BLAST (herramienta básica de búsqueda de alineamiento local, por sus siglas en inglés), que está disponible en la página web del National Center for Biotechnology Information que se puede encontrar actualmente en http://www.ncbi.nlm.nih.gov/ y que fue descrito en primer lugar en Altschul et al. (1990) J. Mol. Biol. 215; 403-410.

Una vez que el software ha producido un alineamiento, es posible calcular el % de similitud y el % de identidad de secuencia. El software generalmente hace esto como parte de la comparación de secuencias y genera un resultado numérico.

El software ClustalW se puede utilizar para realizar alineamientos de secuencia. Preferiblemente, el alineamiento con ClustalW se realiza con los siguientes parámetros para el alineamiento por pares:

Matriz de sustitución	Gonnet 250
Penalización por abrir hueco	20
Penalización por extender hueco	0,2
Penalización por terminar hueco	Ninguna

Por ejemplo, ClustalW2 está disponible en Internet por el European Bioinformatics Institute en la página web EMBL-EBI www.ebi.ac.uk under tools - sequence analysis - ClustalW2. Actualmente, la dirección exacta de la herramienta ClustalW2 es www.ebi.ac.uk/Tools/clustalw2.

Alternativamente, el programa Align X en Vector NTI (Invitrogen) se puede utilizar para realizar alineamientos de secuencia. Exp10 se puede utilizar con la configuración predeterminada:

Penalización por abrir hueco: 10

Penalización por extender de hueco: 0,05

Intervalo de penalización por separación de huecos: 8

El porcentaje de identidad de una secuencia de aminoácidos con, o para, otra secuencia de aminoácidos se puede determinar mediante el uso de la matriz de puntuación: blosum62mt2 y la configuración de alineamiento por pares Vector NTI.

Configuración	K-tupla	1
	Número de mejores diagonales	5
	Tamaño de ventana	5
	Penalización de hueco	3
	Penalización por abrir hueco	10
	Penalización por extender hueco	0,1

Por lo tanto, describimos variantes, homólogos y derivados de cualquier secuencia de aminoácidos de una proteína o polipéptido como se define en la presente memoria, particularmente las de SEQ ID NO: 1 o aquellas de SEQ ID NO: 2, 3, 4, 5, 6, 7 o 8 definidas a continuación.

Las secuencias, particularmente las de variantes, homólogos y derivados de SEQ ID NO: 1, 2, 3, 4, 5, 6, 7 u 8 definidas a continuación, también pueden tener deleciones, inserciones o sustituciones de restos de aminoácidos que producen restos de aminoácidos que producen un cambio silencioso y dan como resultado una sustancia funcionalmente equivalente. Las sustituciones de aminoácidos deliberadas se pueden realizar sobre la base de la similitud en polaridad, carga, solubilidad, hidrofobicidad, hidrofilicidad y/o la naturaleza anfipática de los restos, siempre que se retenga la actividad de unión secundaria de la sustancia. Por ejemplo, los aminoácidos cargados negativamente incluyen ácido aspártico y ácido glutámico; los aminoácidos cargados positivamente incluyen lisina y arginina; y los aminoácidos con grupos de cabeza polar no cargados que tienen valores de hidrofilicidad similares incluyen leucina, isoleucina, valina, glicina, alanina, asparagina, glutamina, serina, treonina, fenilalanina y tirosina.

La sustitución conservadora (tanto sustitución como reemplazo se utilizan en la presente memoria para referirse al intercambio de un resto de aminoácido existente, con un resto alternativo) que se puede producir, es decir, sustitución homogénea tal como básico por básico, ácido por ácido, polar por polar, etc. La sustitución no conservadora también se puede producir, es decir, de una clase de resto a otra o alternativamente implica la inclusión de aminoácidos no naturales tales como ornitina (en lo sucesivo denominada Z), ácido diaminobutírico ornitina (en lo sucesivo denominada B), norleucina ornitina (en lo sucesivo denominado O), pirilalanina, tienilalanina, naftilalanina y fenilglicina.

Las sustituciones conservadoras que se pueden realizar son, por ejemplo, dentro de los grupos de aminoácidos básicos (arginina, lisina e histidina), aminoácidos ácidos (ácido glutámico y ácido aspártico), aminoácidos alifáticos (alanina, valina, leucina, isoleucina), aminoácidos polares (glutamina, asparagina, serina, treonina), aminoácidos aromáticos (fenilalanina, triptófano y tirosina), hidroxilaminoácidos (serina, treonina), aminoácidos grandes (fenilalanina y triptófano) y aminoácidos pequeños (glicina, alanina).

El polipéptido puede ser un polipéptido que tiene la secuencia mostrada en la SEQ ID NO: 1 o una variante polipeptídica que tiene al menos al menos 66%, al menos 70%, al menos 75%, al menos 78%, al menos 80%, al menos 85%, al menos 90%, al menos 95%, al menos 96%, al menos 97%, al menos 98% o al menos 99% de identidad de secuencia de aminoácidos con la misma. El polipéptido puede ser un polipéptido que tiene la secuencia mostrada en la SEQ ID NO: 1 o una variante polipeptídica que tiene al menos al menos 70% de identidad de secuencia de aminoácidos con la misma. El polipéptido puede ser un polipéptido que tiene la secuencia que se muestra en la SEQ ID NO: 1 o una variante polipeptídica que tiene al menos al menos 75% de identidad de secuencia de aminoácidos con la misma. El polipéptido puede ser un polipéptido que tiene la secuencia mostrada en la SEQ ID NO: 1 o una variante polipeptídica que tiene al menos al menos 80% de identidad de secuencia de aminoácidos con la misma.

El polipéptido puede ser un polipéptido que tiene la secuencia mostrada en la SEQ ID NO: 3 o una variante polipeptídica que tiene al menos 60%, al menos 65%, al menos 70%, al menos 75%, al menos 78%, al menos 80%, al menos 85%, al menos 90%, al menos 95%, al menos 96%, al menos 97%, al menos 98% o al menos 99% de identidad de secuencia de aminoácidos con la misma.

El polipéptido puede ser un polipéptido que tiene la secuencia mostrada en la SEQ ID NO: 4 o una variante polipeptídica que tiene al menos 60%, al menos 65%, al menos 70%, al menos 75%, al menos 78%, al menos 80% , al menos el 85%, al menos el 90%, al menos el 95%, al menos el 96%, al menos el 97%, al menos el 98% o al menos el 99% de identidad de secuencia de aminoácidos con la misma.

El polipéptido puede ser un polipéptido que tiene la secuencia mostrada en la SEQ ID NO: 5 o una variante polipeptídica que tiene al menos 60%, al menos 65%, al menos 70%, al menos 75%, al menos 78%, al menos 80% , al menos el 85%, al menos el 90%, al menos el 95%, al menos el 96%, al menos el 97%, al menos el 98% o al menos el 99% de identidad de secuencia de aminoácidos con la misma.

El polipéptido puede ser un polipéptido que tiene la secuencia mostrada en la SEQ ID NO: 2 o una variante polipeptídica que tiene al menos 60%, al menos 65%, al menos 75%, al menos 78%, al menos 80%, al menos 85% , al menos el 90%, al menos el 95%, al menos el 96%, al menos el 97%, al menos el 98% o al menos el 99% de identidad de secuencia de aminoácidos con la misma. El polipéptido puede ser un polipéptido que tiene la secuencia que se muestra en la SEQ ID NO: 2 o una variante polipeptídica que tiene al menos al menos 70% de identidad de secuencia de aminoácidos con la misma. El polipéptido puede ser un polipéptido que tiene la secuencia mostrada en la SEQ ID NO: 2 o una variante polipeptídica que tiene al menos al menos un 75% de identidad de secuencia de aminoácidos con la misma. El polipéptido puede ser un polipéptido que tiene la secuencia mostrada en la SEQ ID NO: 2 o una variante polipeptídica que tiene al menos al menos 80% de identidad de secuencia de aminoácidos con la misma.

El polipéptido puede ser un polipéptido que tiene la secuencia mostrada en la SEQ ID NO: 6 o una variante polipeptídica que tiene al menos al menos 65%, al menos 70%, al menos 75%, al menos 78%, al menos 80%, al menos 85%, al menos 90%, al menos 94%, al menos 95%, al menos 96%, al menos 97%, al menos 98% o al menos 99% de identidad de secuencia de aminoácidos con la misma.

El polipéptido puede ser una variante polipeptídica que tiene la secuencia mostrada en la SEQ ID NO: 7 o una variante polipeptídica que tiene al menos al menos 65%, al menos 70%, al menos 75%, al menos 78%, al menos 80%, a al menos el 85%, al menos el 90%, al menos el 94%, al menos el 95%, al menos el 96%, al menos el 97%, al menos el 98% o al menos el 99% de identidad de secuencia de aminoácidos con la misma.

El polipéptido puede ser una variante polipeptídica que tiene la secuencia mostrada en la SEQ ID NO: 8 o una variante polipeptídica que tiene al menos al menos 65%, al menos 70%, al menos 75%, al menos 78%, al menos 80%, a al menos el 85%, al menos el 90%, al menos el 94%, al menos el 95%, al menos el 96%, al menos el 97%, al menos el 98% o al menos el 99% de identidad de secuencia de aminoácidos con la misma.

En un aspecto, la secuencia polipeptídica descrita en la presente memoria puede estar en una forma purificada.

En un aspecto, el polipéptido descrito o proteína descrita en la presente memoria puede estar en una forma aislada.

Una "variante" o una "variantes" se refiere a polipéptidos o ácidos nucleicos. El término "variante" se puede utilizar indistintamente con el término "mutante". Las variantes incluyen inserciones, sustituciones, transversiones, truncamientos y/o inversiones en una o más ubicaciones en la secuencia de aminoácidos o nucleótidos, respectivamente. Las frases "polipéptido variante", "variante polipeptídica", "polipéptido", "variante" y "enzima variante" significan un polipéptido/proteína que tiene o comprende la secuencia de aminoácidos de, o está modificada comparada con la secuencia de aminoácidos de SEQ ID NO: 1, 2, 3, 4, 5, 6, 7 u 8. Los polipéptidos variantes incluyen un polipéptido que tiene un cierto porcentaje, por ejemplo, 60%, 65%, 66%, 68%, 70%, 72%, 74%, 76%, 78%, 80%, 85%, 90%, 91 %, 92%, 93%, 94%, 95%, 96%, 97%, 98% o 99% de identidad de secuencia con las SEQ ID NO: 1, 2, 3, 4, 5, 6, 7 ó 8. Como se utiliza en la presente memoria, "enzimas parentales", "secuencia parental", "polipéptido parental" significan enzimas y polipéptidos a partir de los cuales se basan cualquiera de los polipéptidos variantes, p. ej., SEQ ID NO: 1 o 2. Un "ácido nucleico parental" significa una secuencia de ácido nucleico que codifica el polipéptido parental. La secuencia señal de una "variante" puede ser la misma o puede diferir de la secuencia señal del Ruminococcus lactaris de tipo salvaje o Blautia/Ruminococcus hansenii o un péptido señal de Bacillus o cualquier secuencia señal que secretará el polipéptido. Una variante se puede expresar como una proteína de fusión que contiene un polipéptido heterólogo. Por ejemplo, la variante puede comprender un péptido señal de otra proteína o una secuencia diseñada para ayudar a la identificación o purificación de la proteína de fusión expresada, tal como una secuencia His-Tag.

Para describir las diversas variantes que se contemplan para abarcar la presente descripción, se adoptará la siguiente nomenclatura para facilitar la referencia. Donde la sustitución incluye un número y una letra, p. ej., 592P, entonces esto se refiere a {posición según el sistema de numeración/aminoácido sustituido}. Por consiguiente, por ejemplo, la sustitución de un aminoácido por prolina en la posición 592 se designa como 592P. Cuando la sustitución incluye una letra, un número y una letra, p. ej., D592P, esto se refiere a {aminoácido original/posición según el sistema de numeración/aminoácido sustituido}. Por consiguiente, por ejemplo, la sustitución de alanina con prolina en la posición 592 se designa como A592P.

Cuando dos o más sustituciones son posibles en una posición particular, esto se designará con letras contiguas, que pueden estar separadas opcionalmente por barras "/", p. ej., G303ED o G303E/D.

La o las posiciones y las sustituciones se enumeran con referencia a SEQ ID NO: 1 o SEQ ID NO: 2. Se pueden encontrar posiciones equivalentes en otra secuencia alineando esta secuencia con SEQ ID NO: 1 o SEQ ID NO: 2 para encontrar un alineamiento con el mayor porcentaje de identidad y luego determinar qué aminoácido se alinea para corresponderse con un aminoácido de una posición específica de SEQ ID NO: 1 o SEQ ID NO: 2. Tal alineamiento y uso de una secuencia como primera referencia es simplemente una cuestión de rutina para un experto en la técnica.

Los "ácidos nucleicos variantes" pueden incluir secuencias que son complementarias a secuencias que son capaces de hibridarse con las secuencias de nucleótidos presentadas en la presente memoria, en particular con SEQ ID NO: 10-11. Por ejemplo, una secuencia variante es complementaria de secuencias capaces de hibridarse bajo condiciones rigurosas, p. ej., 50ºC y 0,2X SSC (1X SSC = 0,15 M NaCl, 0,015 M citrato de sodio, pH 7,0), con las secuencias de nucleótidos presentadas en la presente memoria, en particular con SEQ ID NO: 10-11. Más particularmente, el término variante abarca secuencias que son complementarias a secuencias que son capaces de hibridarse bajo condiciones altamente restrictivas, p. ej., 65ºC y 0,1X SSC, con las secuencias de nucleótidos presentadas en la presente memoria, en particular con SEQ ID NO: 10-11. El punto de fusión (Tf) de un ácido nucleico variante puede ser aproximadamente 1, 2 o 3ºC menor que el Tf del ácido nucleico de tipo salvaje.

En un aspecto, describimos polipéptidos aislados que tienen actividad transgalactosilante como se indicó anteriormente que están codificados por polinucleótidos que se hibridan bajo condiciones de rigurosidad muy baja, preferiblemente condiciones de rigurosidad baja, más preferiblemente condiciones de rigurosidad media, más preferiblemente condiciones de rigurosidad media-alta, incluso más preferiblemente condiciones de rigurosidad alta, y lo más preferiblemente condiciones de rigurosidad muy alta con i) la secuencia de ácido nucleico comprendida en la SEQ ID NO: 10 que codifica el polipéptido maduro de SEQ ID NO: 1; ii) la secuencia de ADNc de i) o iii) la cadena complementaria de i) o ii) o con i) la secuencia de ácido nucleico comprendida en la SEQ ID NO: 11 que codifica el

polipéptido maduro de SEQ ID NO: 2; ii) la secuencia de ADNc de i) o iii) la cadena complementaria de i) o ii); (J. Sambrook, E. F. Fritsch, y T. Maniatis, 1989, Molecular Cloning, A Laboratory Manual, 2d edition, Cold Spring Harbor, Nueva York). Una subsecuencia de SEQ ID NO: 10 u 11 contiene al menos 100 nucleótidos contiguos o preferiblemente al menos 200 nucleótidos contiguos. Además, la subsecuencia puede codificar un fragmento polipeptídico que tiene actividad lactasa.

La secuencia de nucleótidos de SEQ ID NO: 10 u 11 o una subsecuencia de la misma, así como la secuencia de aminoácidos de SEQ ID NO: 1 o 2 o un fragmento de la misma, se puede utilizar para diseñar una sonda de ácido nucleico para identificar y clonar ADN que codifica polipéptidos que tienen actividad transgalactosilasa a partir de cepas de diferentes géneros o especies según métodos bien conocidos en la técnica. En particular, tales sondas se pueden utilizar para la hibridación con el ADN genómico o ADNc del género o especie de interés, siguiendo métodos de transferencia Southern convencionales, con el fin de identificar y aislar el gen correspondiente en el mismo. Tales sondas pueden ser considerablemente más cortas que la secuencia completa, pero deberían ser al menos 14, preferiblemente al menos 25, más preferiblemente al menos 35, y más preferiblemente al menos 70 nucleótidos de longitud. Sin embargo, se prefiere que la sonda de ácido nucleico tenga al menos 100 nucleótidos de longitud. Por ejemplo, la sonda de ácido nucleico puede tener al menos 200 nucleótidos, preferiblemente al menos 300 nucleótidos, más preferiblemente al menos 400 nucleótidos, o más preferiblemente al menos 500 nucleótidos de longitud. Se pueden utilizar sondas incluso más largas, p. ej., sondas de ácido nucleico que son al menos 600 nucleótidos, al menos preferiblemente al menos 700 nucleótidos, más preferiblemente al menos 800 nucleótidos, o más preferiblemente al menos 900 nucleótidos de longitud. Se pueden utilizar sondas de ADN y ARN. Las sondas están típicamente marcadas para detectar el gen correspondiente (por ejemplo, con ^{32}P, ^{3}H, ^{35}S, biotina o avidina).

Una biblioteca de ADN genómico preparada a partir de cualquier otro organismo puede, por lo tanto, cribarse para ADN que se hibrida con las sondas descritas anteriormente y que codifica un polipéptido que tiene actividad lactasa. El ADN genómico u otro a partir de cualquier otro organismo se puede separar mediante electroforesis en gel de agarosa o poliacrilamida u otras técnicas de separación. El ADN de las bibliotecas o el ADN separado se puede transferir e inmovilizar sobre nitrocelulosa u otro material de soporte adecuado. Con el fin de identificar un clon o ADN que sea homólogo con la SEQ ID NO: 10 u 11 o una subsecuencia de la misma, el material de soporte se utiliza en una transferencia Southern.

La hibridación indica que la secuencia de nucleótidos se hibrida con una sonda de ácido nucleico marcada que corresponde con la secuencia de nucleótidos mostrada en la SEQ ID NO: 10 u 11, su cadena complementaria, o una subsecuencia de la misma, bajo condiciones de rigurosidad muy baja a muy alta. Las moléculas a las que la sonda de ácido nucleico se hibrida bajo estas condiciones se pueden detectar utilizando una película de rayos X.

En un aspecto preferido, la sonda de ácido nucleico es los nucleótidos 175 a 2011 o los nucleótidos 198 a 2040 de la SEQ ID NO: 10 o SEQ ID NO: 11 respectivamente. En otro aspecto preferido, la sonda de ácido nucleico es una secuencia de polinucleótidos que codifica el polipéptido de SEQ ID NO: 1 o SEQ ID NO: 2, o una subsecuencia de la misma. En otro aspecto preferido, la sonda de ácido nucleico es la SEQ ID NO: 10 o SEQ ID NO: 11. En otro aspecto preferido, la sonda de ácido nucleico es la región que codifica el polipéptido maduro de SEQ ID NO: 10 o SEQ ID NO: 11.

Para sondas largas de al menos 100 nucleótidos de longitud, las condiciones de rigurosidad muy baja a muy alta se definen como prehibridación e hibridación a 42ºC en 5X SSPE, SDS al 0,3%, 200 g/ml de ADN de esperma de salmón fragmentado y desnaturalizado, y formamida al 25% para rigurosidades muy bajas y bajas, formamida al 35% para rigurosidades medias y medias-altas, o formamida al 50% para rigurosidades altas y muy altas, siguiendo procedimientos estándar de transferencia Southern durante 12 a 24 horas de manera óptima.

Para sondas largas de al menos 100 nucleótidos de longitud, el material de soporte se lava finalmente tres veces cada uno durante 15 minutos utilizando 2X SSC, SDS al 0,2% preferiblemente al menos a una temperatura de 45ºC (rigurosidad muy baja), más preferiblemente al menos a 50ºC (rigurosidad baja), más preferiblemente al menos a 55ºC (rigurosidad media), más preferiblemente al menos a 60ºC (rigurosidad media-alta), incluso más preferiblemente al menos a 65ºC (rigurosidad alta), y lo más preferiblemente al menos a 70ºC (rigurosidad muy alta).

El lavado se puede realizar utilizando 0,2X SSC, SDS al 0,2% preferiblemente al menos a una temperatura de 45ºC (rigurosidad muy baja), más preferiblemente al menos a 50ºC (rigurosidad baja), más preferiblemente al menos a 55ºC (rigurosidad media), más preferiblemente al menos a 60ºC (rigurosidad media-alta), incluso más preferiblemente al menos a 65ºC (rigurosidad alta), y lo más preferiblemente al menos a 70ºC (rigurosidad muy alta). El lavado se puede realizar utilizando 0,1X SSC, SDS al 0,2% preferiblemente al menos a una temperatura de 45ºC (rigurosidad muy baja), más preferiblemente al menos a 50ºC (rigurosidad baja), más preferiblemente al menos a 55ºC (rigurosidad media), más preferiblemente al menos a 60ºC (rigurosidad media-alta),

incluso más preferiblemente al menos a 65ºC (rigurosidad alta), y lo más preferiblemente al menos a 70ºC (rigurosidad muy alta).

Para sondas cortas que tienen aproximadamente 15 nucleótidos a aproximadamente 70 nucleótidos de longitud, las condiciones de rigurosidad se definen como prehibridación, hibridación y lavado posterior a la hibridación a una

temperatura de aproximadamente 5ºC a aproximadamente 10ºC por debajo del T_f calculado utilizando el cálculo según Bolton y McCarthy (1962, Proceedings of the National Academy of Sciences USA 48:1390) en NaCl 0,9 M, Tris-HCl 0,09 M, pH 7,6, EDTA 6 mM, NP-40 al 0,5%, 1X solución de Denhardt, pirofosfato de sodio 1 mM, fosfato monobásico de sodio 1 mM, ATP 0,1 mM y 0,2 mg de ARN de levadura por ml siguiendo métodos de transferencia Southern estándar.

Para sondas cortas que tienen aproximadamente 15 nucleótidos a aproximadamente 70 nucleótidos de longitud, el material de soporte se lava una vez en 6X SCC más SDS al 0,1% durante 15 minutos y dos veces cada uno durante 15 minutos utilizando 6X SSC a una temperatura de 5ºC a 10ºC por debajo del T_f calculado.

En condiciones de hibridación que contienen sal, el T_f efectivo es lo que controla el grado de identidad requerido entre la sonda y el ADN unido al filtro para una hibridación exitosa. El T_f efectivo se puede determinar utilizando la fórmula siguiente para determinar el grado de identidad requerido para que dos ADN se hibriden en diversas condiciones de rigurosidad.

$$T_f \text{ efectivo} = 81,5 + 16,6(\log M[Na^+]) + 0,41(\%G+C) – 0,72(\% \text{ de formamida})$$

(Véase www.ndsu.nodak.edu/instruct/mcclean/plsc731/dna/dna6.htm). El contenido de G+C de SEQ ID NO: 10 es 42% y el contenido de G+C de SEQ ID NO: 11 es 44%. Para rigurosidad media, la formamida es 35% y la concentración de Na^+ para 5X SSPE es 0,75 M.

Otra relación relevante es que un desajuste del 1% de dos ADN reduce el T_f en 1,4ºC. Para determinar el grado de identidad requerido para que dos ADN se hibriden bajo condiciones de rigurosidad media a una temperatura de 42ºC, se utiliza la siguiente fórmula:

$$\% \text{ de homología} = 100 - [(Tf \text{ efectivo} – \text{Temperatura})/1,4]$$

(Véase www.ndsu.nodak.edu/instruct/mcclean/plsc731/dna/dna6.htm)

Los ácidos nucleicos variantes incluyen un polinucleótido que tiene un cierto porcentaje, p. ej., 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95% o 99% de identidad de secuencia con la codificación de ácido nucleico SEQ ID NO: 1 o 2. En un aspecto, se proporciona un ácido nucleico capaz de codificar un polipéptido como se describe en la presente memoria. En un aspecto adicional, el ácido nucleico descrito en la presente memoria tiene una secuencia de ácido nucleico que es al menos 60%, tal como al menos 65%, tal como al menos 70%, tal como al menos 75%, tal como al menos 80%, tal como al menos el 85%, tal como al menos el 90%, tal como al menos el 95%, tal como al menos el 99% idéntico a la SEQ ID NO: 10 u 11.

En un aspecto, los polipéptidos descritos en la presente memoria comprenden una secuencia de aminoácidos que tiene al menos 66% de identidad de secuencia con la secuencia de aminoácidos del polipéptido maduro codificado por la secuencia de nucleótidos que codifica la transgalatosilasa contenida en el nº de acceso DSM: 20583. En un aspecto, los polipéptidos descritos en la presente memoria comprenden una secuencia de aminoácidos que tiene al menos 66% de identidad de secuencia con la secuencia de aminoácidos del polipéptido maduro codificado por la secuencia de nucleótidos que codifica la transgalatosilasa contenida en el nº de acceso ATCC: 29176. Todas las consideraciones y limitaciones relacionadas con las identidades de secuencia y funcionalidad discutidas en términos de la SEQ ID NO: 1 o 2 se aplican según corresponda a las identidades de secuencia y a la funcionalidad de estos polipéptidos y nucleótidos.

Como se utiliza en la presente memoria, el término "expresión" se refiere al método mediante el cual se produce un polipéptido en base a la secuencia de ácido nucleico de un gen. El método incluye transcripción y traducción.

Como se utiliza en la presente memoria, "polipéptido" se utiliza de forma intercambiable con los términos "secuencia de aminoácidos", "enzima", "péptido" y/o "proteína". Como se utiliza en la presente memoria, "secuencia de nucleótidos" o "secuencia de ácidos nucleicos" se refiere a una secuencia de oligonucleótidos o secuencia de polinucleótidos y variantes, homólogos, fragmentos y derivados de los mismos. La secuencia de nucleótidos puede ser de naturaleza genómica, sintética o recombinante y puede ser de doble cadena o monocatenaria, ya sea representando la cadena sentido o antisentido. Como se utiliza en la presente memoria, el término "secuencia de nucleótidos" incluye ADN genómico, ADNc, ADN sintético y ARN.

"Homólogo" significa una entidad que tiene un cierto grado de identidad u "homología" con las secuencias de aminoácidos y las secuencias de nucleótidos encontradas. En un aspecto, la secuencia de aminoácidos encontrada es SEQ ID NO: 1-8, y la secuencia de aminoácidos encontrada es preferiblemente SEQ ID NO: 10-11.

Una "secuencia homóloga" incluye un polinucleótido o un polipéptido que tiene un cierto porcentaje, por ejemplo, 80%, 85%, 90%, 95% o 99%, de identidad de secuencia con otra secuencia. El porcentaje de identidad significa que, cuando se alinea, ese porcentaje de bases o restos de aminoácidos es el mismo cuando se comparan las dos secuencias. Las secuencias de aminoácidos son no idénticas, donde un aminoácido está sustituido, eliminado o añadido en comparación con la secuencia en cuestión. El porcentaje de identidad de secuencia típicamente se mide con respecto a la secuencia madura de la proteína encontrada, es decir, después de la eliminación de una

secuencia señal, por ejemplo. Típicamente, los homólogos comprenderán los mismos restos del sitio activo que la secuencia de aminoácidos encontrada. Los homólogos también retienen actividad enzimática, aunque el homólogo puede tener diferentes propiedades enzimáticas que el tipo salvaje.

Como se utiliza en la presente memoria, "hibridación" incluye el proceso mediante el cual una cadena de ácido nucleico se une con una cadena complementaria a través del emparejamiento de bases, así como el proceso de amplificación como se lleva a cabo en tecnologías de reacción en cadena de la polimerasa (PCR). El ácido nucleico variante puede existir como ADN o ARN monocatenario o bicatenario, un ARN/ADN heterodúplex o un ARN/ADN copolímero. Como se utiliza en la presente memoria, "copolímero" se refiere a una única cadena de ácido nucleico que comprende tanto ribonucleótidos como desoxirribonucleótidos. El ácido nucleico variante puede estar optimizado con codones para aumentar adicionalmente la expresión.

Como se utiliza en la presente memoria, un compuesto "sintético" se produce mediante síntesis química o enzimática *in vitro*. Incluye, pero no se limita a, ácidos nucleicos variantes preparados con el uso óptimo de codones para organismos anfitriones, tales como una célula anfitriona de levadura u otros anfitriones de expresión de elección.

Como se utiliza en la presente memoria, "célula transformada" incluye células, que incluyen tanto células bacterianas como fúngicas, que se han transformado mediante el uso de técnicas de ADN recombinante. La transformación típicamente se produce mediante la inserción de una o más secuencias de nucleótidos en una célula. La secuencia de nucleótidos insertada puede ser una secuencia de nucleótidos heteróloga, es decir, es una secuencia que no es natural para la célula que se va a transformar, tal como una proteína de fusión.

Como se utiliza en la presente memoria, "unidos operablemente" significa que los componentes descritos están en una relación que les permite funcionar de la manera prevista. Por ejemplo, una secuencia reguladora unida operablemente a una secuencia codificante está ligada de tal forma que la expresión de la secuencia codificante se consigue bajo condiciones compatibles con las secuencias control.

Como se utiliza en la presente memoria, el término "fragmento" se define como un polipéptido que tiene uno o más (varios) aminoácidos eliminados del extremo amino y/o carboxilo, por ejemplo, del polipéptido de SEQ ID NO: 12 ó 13; en donde el fragmento tiene actividad transgalactosilante.

En un aspecto, el término "fragmento polipeptídico" se define en la presente memoria como un polipéptido que tiene uno o más (varios) aminoácidos eliminados del extremo amino y/o carboxilo del polipéptido de SEQ ID NO: 1 ó 2; en donde el fragmento tiene actividad transgalactosilante.

En un aspecto, un fragmento contiene al menos 500, 700, 900 o 1000 restos de aminoácidos. En un aspecto, un fragmento contiene como máximo 1250, 1200, 1180, 1170, 1150 u 1125 restos de aminoácidos.

En un aspecto adicional, la longitud del polipéptido descrito en la presente memoria es de 500 a 1250 aminoácidos. En un aspecto adicional, la longitud de la variante polipeptídica es de 500 a 1200 aminoácidos. En un aspecto adicional, la longitud de la variante polipeptídica es de 700 a 1170 aminoácidos. En un aspecto adicional, la longitud de la variante polipeptídica es de 900 a 1180 aminoácidos. En un aspecto adicional, la longitud de la variante polipeptídica es de 900 a 1150 aminoácidos. En un aspecto adicional, la longitud de la variante polipeptídica es de 1000 a 1125 aminoácidos.

En un aspecto, se proporciona un plásmido que comprende un ácido nucleico como se describe en la presente memoria.

En un aspecto, se proporciona un vector de expresión que comprende un ácido nucleico como se describe en la presente memoria, o capaz de expresar un polipéptido como se describe en la presente memoria.

En un aspecto adicional, se proporciona una célula anfitriona que comprende, preferiblemente transformada con, un plásmido como se describe en la presente memoria o un vector de expresión como se describe en la presente memoria.

En un aspecto adicional, se proporciona una célula capaz de expresar un polipéptido como se describe en la presente memoria.

En un aspecto, la célula anfitriona como se describe en la presente memoria, o la célula como se describe en la presente memoria es una célula bacteriana, fúngica o de levadura.

En un aspecto adicional, la célula anfitriona se selecciona del grupo que consiste en *Ruminococcus*, *Bifidobacterium*, *Lactococcus*, *Lactobacillus*, *Streptococcus*, *Leuconostoc*, *Escherichia*, *Bacillus*, *Streptomyces*, *Saccharomyces*, *Kluyveromyces*, *Candida*, *Torula*, *Torulopsis* y *Aspergillus*.

En un aspecto adicional, la célula de célula anfitriona se selecciona del grupo que consiste en *Ruminococcus hansenii*, *Bifidobacterium breve*, *Bifidobacterium longum*, *Bifidobacterium infantis*, *Bifidobacterium bifidum* y *Lactococcus lactis*.

En un aspecto adicional, un método de expresión de un polipéptido como se describe en la presente memoria comprende obtener una célula anfitriona o una célula como se describe en la presente memoria y expresar el polipéptido en la célula o célula anfitriona, y opcionalmente purificar el polipéptido.

Variantes polipeptídicas de SEQ ID NO: 1 o SEQ ID NO: 2

En un aspecto, se proporciona una variante de SEQ ID NO: 1 ó 2 que tiene una sustitución en una o más posiciones que afecta a una propiedad alterada tal como transgalactosilación mejorada, con respecto a SEQ ID NO: 1 ó 2. Dichos polipéptidos variantes también se mencionan en este documento por conveniencia como "polipéptido variante", "variante polipeptídica" o "variante". En un aspecto, los polipéptidos como se definen en la presente memoria tienen una actividad transgalactosilante mejorada en comparación con el polipéptido de SEQ ID NO: 1 ó 2. En otro aspecto, los polipéptidos como se definen en la presente memoria tienen una velocidad de reacción mejorada en comparación con el polipéptido de SEQ ID NO: 1 ó 2.

En un aspecto, los polipéptidos y variantes como se definen en la presente memoria exhiben actividad enzimática. En un aspecto, los polipéptidos y los polipéptidos variantes descritos en la presente memoria comprenden actividad transgalactosilante.

En un aspecto, la relación de actividad transgalactosilante:actividad de β-galactosidasa es al menos 2,5, tal como al menos 3, tal como al menos 4, tal como al menos 5, tal como al menos 6, tal como al menos 7, tal como al menos 8, tal como al menos 9, tal como al menos 10, tal como al menos 11, o tal como al menos 12 después de 30 min de reacción.

En un aspecto, los polipéptidos y las variantes como se definen en la presente memoria se pueden obtener a partir de fuentes microbianas, en particular de un hongo filamentoso o levadura, o de una bacteria. La enzima se puede, p. ej., obtener a partir de una cepa de Agaricus, p. ej. A. bisporus; Ascovaginospora; Aspergillus, p. ej. A. niger, A. awamori, A. foetidus, A. japonicus, A. oryzae; Candida; Chaetomium; Chaetotomastia; Dictyostelium, p. ej. D. discoideum; Kluveromyces, p. ej. K. fragilis, K. lactis; Mucor, p. ej. M. javanicus, M. mucedo, M. subtilissimus; Neurospora, p. ej. N. crassa; Rhizomucor, p. ej. R. pusillus; Rhizopus, p. ej. R. arrhizus, R. japonicus, R. stolonifer; Sclerotinia, p. ej. S. libertiana; Torula; Torulopsis; Trichophyton, p. ej. T. rubrum; Whetzelinia, p. ej. W. sclerotiorum; Bacillus, p. ej. B. coagulans, B. circulans, B. megaterium, B. novalis, B. subtilis, B. pumilus, B. stearothermophilus, B. thuringiensis; Bifidobacterium, p. ej. B. longum, B. bifidum, B. animalis; Chryseobacterium; Citrobacter, p. ej. C. freundii; Clostridium, p. ej. C. perfringens; Diplodia, p. ej. D. gossypina; Enterobacter, p. ej. E. aerogenes, E. cloacae Edwardsiella, E. tarda; Erwinia, p. ej. E. herbicola; Escherichia, p. ej. E. coli; Klebsiella, p. ej. K. pneumoniae; Miriococcum; Mirotesio; Mucor; Neurospora, p. ej. N. crassa; Proteus, p. ej. P. vulgaris; Providencia, p. ej. P. stuartii; Pycnoporus, p. ej. Pycnoporus cinnabarinus, Pycnoporus sanguineus; Ruminococcus, p. ej. R. torques; Salmonella, p. ej. S. typhimurium; Serratia, p. ej. S. liquefasciens, S. marcescens; Shigella, p. ej. S. flexneri; Streptomyces, p. ej. S. antibioticus, S. castaneoglobisporus, S. violeceoruber; Trametes; Trichoderma, p. ej. T. reesei, T. viride; Yersinia, p. ej. Y. enterocolitica.

Se proporciona un polipéptido aislado y/o purificado que comprende un polipéptido o un polipéptido variante como se define en la presente memoria. El polipéptido variante puede ser una forma madura del polipéptido (SEQ ID NO: 1 ó 2). En un aspecto, las variantes incluyen un dominio C-terminal.

En un aspecto, un polipéptido variante como se define en la presente memoria incluye variantes en donde se han añadido o eliminado entre uno y aproximadamente 25 restos de aminoácidos con respecto a la SEQ ID NO: 1 o la SEQ ID NO: 2. En un aspecto, la variante tiene la secuencia de aminoácidos de SEQ ID NO: 1 o SEQ ID NO: 2, en donde se ha sustituido cualquier número entre uno y aproximadamente 25 aminoácidos. En un aspecto adicional, la variante tiene la secuencia de aminoácidos de SEQ ID NO: 1 o SEQ ID NO: 2, en donde se ha sustituido cualquier número entre tres y doce aminoácidos. En un aspecto adicional, la variante tiene la secuencia de aminoácidos de SEQ ID NO: 1 o SEQ ID NO: 2, en donde se ha sustituido cualquier número entre cinco y nueve aminoácidos.

En un aspecto, al menos dos, en otro aspecto al menos tres, y en otro aspecto, se han sustituido al menos cinco aminoácidos de la SEQ ID NO: 1 o SEQ ID NO: 2.

En un aspecto, el o los polipéptidos descritos en la presente memoria tienen la secuencia de SEQ ID NO: 1 ó 2.

En un aspecto, el o los polipéptidos descritos en la presente memoria tienen la secuencia de SEQ ID NO: 1 ó 2, en donde el aminoácido 10, tal como 9, tal como 8, tal como 7, tal como 6, tal 5, tal como 4, tal como 3, tal como 2, tal como 1 en el extremo N-terminal se sustituye y/o se elimina.

En un aspecto adicional, la longitud de la variante polipeptídica es de 500 a 1250 aminoácidos. En un aspecto adicional, la longitud de la variante polipeptídica es de 500 a 1200 aminoácidos. En un aspecto adicional, la longitud de la variante polipeptídica es de 700 a 1170 aminoácidos. En un aspecto adicional, la longitud de la variante polipeptídica es de 900 a 1180 aminoácidos. En un aspecto adicional, la longitud de la variante polipeptídica es de 900 a 1150 aminoácidos. En un aspecto adicional, la longitud de la variante polipeptídica es de 1000 a 1125 aminoácidos.

Caracterización de polipéptidos

Las enzimas y variantes enzimáticas de las mismas se pueden caracterizar por su ácido nucleico y secuencias polipeptídicas primarias, por modelado estructural tridimensional y/o por su actividad específica. Las características adicionales de los polipéptidos o variantes polipeptídicas como se define en la presente memoria incluyen estabilidad, rango de pH, estabilidad a la oxidación y termoestabilidad, por ejemplo. Los niveles de expresión y la actividad de la enzima se pueden evaluar utilizando ensayos estándar conocidos por los expertos en la técnica. En otro aspecto, las variantes demuestran características de rendimiento mejoradas con respecto al polipéptido con SEQ ID NO: 1 ó 2, tales como estabilidad mejorada a altas temperaturas, p. ej., 65-85ºC.

Una característica de expresión significa un nivel alterado de expresión de la variante, cuando la variante se produce en una célula anfitriona particular. La expresión generalmente se refiere a la cantidad de variante activa que es recuperable a partir de un caldo de fermentación utilizando técnicas estándar conocidas en esta técnica durante un período de tiempo dado. La expresión también se puede relacionar con la cantidad o relación de variante producida dentro de la célula anfitriona o secretada por la célula anfitriona. La expresión también se puede relacionar con la velocidad de traducción del ARNm que codifica el polipéptido variante.

Se proporciona un ácido nucleico complementario a un ácido nucleico que codifica cualquiera de las variantes polipeptídicas expuestas como se define en la presente memoria. Adicionalmente, se proporciona un ácido nucleico capaz de hibridarse con el complemento. La secuencia para uso en los métodos y composiciones descritos en la presente memoria puede ser una secuencia sintética. Incluye, pero no se limita a, secuencias hechas con el uso de codones óptimos para la expresión en organismos anfitriones, tales como levadura.

Las variantes polipeptídicas como se proporcionan en la presente memoria se pueden producir sintéticamente o mediante expresión recombinante en una célula anfitriona, según procedimientos bien conocidos en la técnica. En un aspecto, el o los polipéptidos descritos en la presente memoria son polipéptidos recombinantes. La variante polipeptídica expresada como se define en la presente memoria opcionalmente se aísla antes del uso.

La variante polipeptídica como se define en la presente memoria se puede purificar después de la expresión. Los métodos de modificación genética y producción recombinante de variantes polipeptídicas se describen, por ejemplo, en las Patentes de EE.UU. núms. de publicación 7.371.552, 7.166.453; 6.890.572; y 6.667.065; y los Núms. de solicitudes publicadas de EE.UU. 2007/0141693; 2007/0072270; 2007/0020731; 2007/0020727; 2006/0073583; 2006/0019347; 2006/0018997; 2006/0008890; 2006/0008888; y 2005/0137111. Las enseñanzas relevantes de estas descripciones, que incluyen secuencias polinucleotídicas que codifican polipéptidos, cebadores, vectores, métodos de selección, células anfitrionas, purificación y reconstitución de variantes polipeptídicas expresadas, y caracterización de variantes polipeptídicas como se define en la presente memoria, que incluyen tampones útiles, intervalos de pH, concentraciones de Ca^{2+}, concentraciones de sustrato y concentraciones de enzimas para ensayos enzimáticos.

Las células anfitrionas adecuadas incluyen una bacteria Gram positiva seleccionada del grupo que consiste en *Bacillus subtilis, B. licheniformis, B. lentus, B. brevis, B. stearothermophilus, B. alkalophilus, B. amyloliquefaciens, B. coagulans, B. circulans, B. lautus, B. thuringiensis, Streptomyces lividans* o *S. murinus*; o una bacteria Gram negativa, en donde dicha bacteria Gram negativa es *Escherichia coli* o una especie de *Pseudomonas*. En un aspecto, la célula anfitriona es *B. subtilus* o *B. licheniformis*. La célula anfitriona puede ser *B. subtilis*, y la proteína expresada se modifica por ingeniería genética para que comprenda una secuencia señal de *B. subtilis*, como se expone con más detalle a continuación. En un aspecto, la célula anfitriona expresa el polinucleótido como se establece en las reivindicaciones.

Una célula anfitriona se puede modificar genéticamente para expresar una variante polipeptídica como se define en la presente memoria con una secuencia de aminoácidos que tiene al menos aproximadamente 66%, 68%, 70%, 72%, 74%, 78%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 100% de identidad con el polipéptido de SEQ ID NO: 1 ó 2. El polinucleótido que codifica una variante polipeptídica como se define en la presente memoria puede tener una secuencia de ácido nucleico que codifica la proteína de SEQ ID NO: 1 o una secuencia de ácido nucleico que tiene al menos aproximadamente 66%, 68%, 70%, 72%, 74%, 78% , 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% o 100% de identidad de secuencia con un ácido nucleico que codifica la proteína de SEQ ID NO: 1 ó 2. La secuencia de ácido nucleico puede tener al menos aproximadamente 60%, 66%, 68%, 70%, 72%, 74%, 78%, 80%, 85%, 90%, 95%, 96%, 97%, 98% o 99% de identidad de secuencia con el ácido nucleico de SEQ ID NO: 10-11.

Vectores

En un aspecto, un vector puede comprender un polinucleótido. En un aspecto, una célula bacteriana comprende el vector. Una construcción de ADN puede comprender un ácido nucleico que codifica una variante que se transfiere a una célula hospedadora en un vector de expresión que comprende secuencias reguladoras unidas operablemente a una secuencia codificante. El vector puede ser cualquier vector que se pueda integrar en un genoma de célula anfitriona fúngica y replicarse cuando se introduzca en la célula anfitriona. El FGSC Catalogue of Strains, University of Missouri, enumera vectores adecuados. Se proporcionan ejemplos adicionales de vectores de expresión y/o

integración adecuados en Sambrook et al., MOLECULAR CLONING: A LABORATORY MANUAL, 3rd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York (2001); Bennett et al., MORE GENE MANIPULATIONS IN FUNGI, Academic Press, San Diego (1991), pp. 396-428; y Num de patente de EE.UU. 5.874.276. Los vectores ilustrativos incluyen pFB6, pBR322, PUC18, pUC100 y pENTR/D, pDON™201, pDONR™221, pENTR™, pGEM®3Z y pGEM®4Z. Ejemplos de uso en células bacterianas incluyen pBR322 y pUC19, que permiten la replicación en *E. coli*, y pE194, por ejemplo, que permite la replicación en *Bacillus*.

Un ácido nucleico que codifica una variante se puede unir operablemente a un promotor adecuado, que permite la transcripción en la célula anfitriona. El promotor se puede obtener a partir de genes que codifican proteínas homólogas o heterólogas a la célula anfitriona. Los ejemplos no limitantes adecuados de promotores incluyen promotores cbh1, cbh2, egl1 y egl2. El promotor puede ser uno que sea nativo de la célula anfitriona. Por ejemplo, cuando *P. saccharophila* es el anfitrión, el promotor es un promotor nativo de *P. saccharophila*. Un "promotor inducible" es un promotor que está activo bajo regulación ambiental o de desarrollo. El promotor puede ser uno que sea heterólogo para la célula anfitriona.

La secuencia codificante puede estar unida operablemente a una secuencia de ADN que codifica una secuencia señal. Un péptido señal representativo es la SEQ ID NO: 9 que es la secuencia señal nativa del precursor de aprE de Bacillus subtilis. El ADN que codifica la secuencia señal se puede reemplazar con una secuencia de nucleótidos que codifica una secuencia señal de otros precursores de Bacillus subtilis extracelulares. El polinucleótido que codifica la secuencia señal puede estar inmediatamente secuencia arriba y dentro del marco del polinucleótido que codifica el polipéptido. La secuencia señal se puede seleccionar de la misma especie que la célula anfitriona.

Una secuencia señal y una secuencia promotora que comprende una construcción de ADN o vector para ser introducida en una célula anfitriona fúngica se pueden obtener a partir de la misma fuente. El vector de expresión también puede incluir una secuencia de terminación. La secuencia de terminación y la secuencia promotora se pueden obtener a partir de la misma fuente. La secuencia de terminación puede ser homóloga a la célula anfitriona.

Un vector de expresión puede incluir un marcador seleccionable. Los ejemplos de marcadores seleccionables adecuados incluyen aquellos que confieren resistencia a agentes antimicrobianos, p.ej., higromicina o fleomicina. Los marcadores selectivos nutricionales también son adecuados e incluyen *amdS*, *argB* y *pyr4*. El marcador selectivo puede ser el gen *amdS*, que codifica la enzima acetamidasa; permite a las células transformadas crecer sobre acetamida como fuente de nitrógeno. El uso de un gen *amdS* de *A. nidulans* como marcador selectivo se describe en Kelley et al., EMBO J. 4: 475-479 (1985) y Penttila et al., Gene 61: 155-164 (1987).

Un vector de expresión adecuado que comprende una construcción de ADN con un polinucleótido que codifica una variante puede ser cualquier vector que sea capaz de replicarse de forma autónoma en un organismo hospedador dado o integrarse en el ADN del hospedador. El vector de expresión puede ser un plásmido. Se contemplan dos tipos de vectores de expresión para obtener la expresión de genes. El primer vector de expresión comprende secuencias de ADN en las que el promotor, la región codificante y el terminador se originan todos a partir del gen que se va a expresar. El truncamiento génico se puede obtener eliminando secuencias de ADN no deseadas para dejar que el dominio se exprese bajo el control de sus propias secuencias reguladoras transcripcionales y traduccionales. El segundo tipo de vector de expresión está preensamblado y contiene las secuencias requeridas para la transcripción de alto nivel y un marcador seleccionable. La región codificante de un gen o parte del mismo se puede insertar en este vector de expresión de uso general, de manera que está bajo el control transcripcional de las secuencias promotora y terminadora de la construcción de expresión. Los genes o parte de los mismos se pueden insertar secuencia abajo del promotor *cbh1* fuerte.

Transformación, expresión y cultivo de células anfitrionas

La introducción de una construcción o vector de ADN en una célula anfitriona incluye técnicas tales como transformación; electroporación; microinyección nuclear; transducción; transfección, p. ej., mediada por lipofección y mediada por DEAE-Dextrina; incubación ADN precipitado con fosfato cálcico; bombardeo a alta velocidad con microproyectiles recubiertos de ADN; y fusión de protoplastos. Las técnicas de transformación generales son conocidas en la técnica. Véase, p. ej., Ausubel *et al.* (1987), *supra*, chapter 9; Sambrook *et al.* (2001), *supra*; y Campbell *et al.*, Curr. Gineta. 16: 53-56 (1989). La expresión de proteínas heterólogas en *Trichoderma* se describe, por ejemplo, en la Patente de EE.UU. Nº de publicación 6.022.725; Patente de EE.UU. Nº de publicación 6.268.328; Harkki et al., Enzyme Microb. Technol. 13: 227-233 (1991); Harkki et.al., BioTechnol. 7: 596-603 (1989); EP 244,234; y EP 215,594. Los transformantes genéticamente estables se pueden construir con sistemas de vectores en los que el ácido nucleico que codifica una variante se integra de forma estable en un cromosoma de la célula anfitriona. Los transformantes se purifican a continuación mediante técnicas conocidas.

En un ejemplo no limitante, los transformantes estables que incluyen un marcador *amdS* se distinguen de los transformantes inestables por su velocidad de crecimiento más rápida y la formación de colonias circulares con un contorno liso, en lugar de irregular sobre un medio de cultivo sólido que contiene acetamida. Además, en algunos casos se lleva a cabo una prueba adicional de estabilidad cultivando los transformantes sobre medio sólido no selectivo, p. ej., un medio que carece de acetamida, recolectando esporas de este medio de cultivo y determinando

el porcentaje de estas esporas que posteriormente germinan y crecen sobre medio selectivo que contiene acetamida. Otros métodos conocidos en la técnica se pueden utilizar para seleccionar transformantes.

Identificación de actividad

Para evaluar la expresión de una variante en una célula anfitriona, los ensayos pueden medir la proteína expresada, el ARNm correspondiente o la actividad β-galactosidasa. Por ejemplo, los ensayos adecuados incluyen transferencia Northern y Southern, RT-PCR (reacción en cadena de la polimerasa con transcriptasa inversa) e hibridación *in situ*, utilizando una sonda de hibridación marcada apropiadamente. Los ensayos adecuados también incluyen medir la actividad en una muestra. Los ensayos adecuados de la actividad de la variante incluyen, pero no se limitan a ensayos basados en ONPG o determinación de glucosa en mezclas de reacción tales como las descritas en los ejemplos de la presente memoria.

Métodos para purificar polipéptidos descritos en la presente memoria

En general, una variante producida en cultivo celular se secreta en el medio y se puede purificar o aislar, p. ej., retirando componentes no deseados del medio de cultivo celular. En algunos casos, una variante se puede recuperar de un lisado celular. En tales casos, la enzima se purifica a partir de las células en las que se produjo utilizando técnicas empleadas rutinariamente por los expertos en la técnica. Los ejemplos incluyen, pero no se limitan a, cromatografía de afinidad, métodos cromatográficos de intercambio iónico, que incluyen intercambio de iones de alta resolución, cromatografía de interacción hidrofóbica, división en dos fases, precipitación con etanol, HPLC de fase inversa, cromatografía sobre sílice o sobre resina de intercambio catiónico, tal como DEAE, cromatoenfoque, SDS-PAGE, precipitación con sulfato de amonio y filtración en gel utilizando Sephadex G-75, por ejemplo. Dependiendo del uso pretendido, el o los polipéptidos descritos en la presente memoria se pueden, por ejemplo, liofilizar o preparar en una solución. En un aspecto, el o los polipéptidos descritos en la presente memoria están en forma liofilizada. En otro aspecto, el o los polipéptidos descritos en la presente memoria están en solución.

Métodos para inmovilizar y formular los polipéptidos descritos en la presente memoria

Las composiciones polipeptídicas se pueden preparar de acuerdo con métodos conocidos en la técnica y pueden estar en forma de una composición líquida o seca. Por ejemplo, la composición polipeptídica puede estar en forma de un granulado o de un microgranulado. El polipéptido que se incluirá en la composición se puede estabilizar de acuerdo con métodos conocidos en la técnica.

La preparación enzimática, tal como en forma de un ingrediente alimentario puede estar en forma de una solución o como un sólido dependiendo del uso y/o el modo de aplicación y/o el modo de administración. La forma sólida puede ser como un polvo enzimático desecado o como una enzima granulada.

Los ejemplos de formulaciones enzimáticas secas incluyen productos secados por atomización, productos de granulación por mezclador, productos estratificados tales como gránulos de lecho fluido, gránulos extruidos o peletizados, productos prensados o productos liofilizados.

La preparación enzimática, tal como en forma de un ingrediente alimentario puede estar en forma de una solución o como un sólido dependiendo del uso y/o el modo de aplicación y o el modo de administración. La forma sólida puede ser como un polvo enzimático desecado o como una enzima granulada.

En un aspecto, describimos una preparación de complejo enzimático que comprende un complejo enzimático, un portador de enzimas y opcionalmente un estabilizador y/o un conservante.

En otro aspecto adicional, el portador de enzima se puede seleccionar del grupo que consiste en glicerol o agua.

En un aspecto adicional, la preparación comprende un estabilizador. En un aspecto, el estabilizador se selecciona del grupo que consiste en sales inorgánicas, polioles, azúcares y combinaciones de los mismos. En un aspecto, el estabilizador es una sal inorgánica tal como cloruro de potasio. En otro aspecto, el poliol es glicerol, propilenglicol o sorbitol. En otro aspecto más, el azúcar es un hidrato de carbono de molécula pequeña, en particular cualquiera de los hidratos de carbono de sabor dulce, tales como glucosa, galactosa, fructosa y sacarosa.

Y en un aspecto adicional, la preparación comprende un conservante. En un aspecto, el conservante es metilparabeno, propilparabeno, benzoato, sorbato u otros conservantes autorizados para uso alimentario o una mezcla de los mismos.

Un método para producir un producto alimentario se puede ensayar con enzimas inmovilizadas, p. ej. una lactasa inmovilizada u otro galactooligosacárido productor de enzimas. La enzima se puede inmovilizar sobre cualquier soporte orgánico o inorgánico. Los soportes inorgánicos ilustrativos incluyen alúmina, celita, Dowex-1-cloruro, perlas de vidrio y gel de sílice. Los soportes orgánicos ilustrativos incluyen DEAE-celulosa, hidrogeles de alginato o perlas de alginato o equivalentes. En varios aspectos, la inmovilización de la lactasa se puede optimizar mediante adsorción física sobre el soporte inorgánico. Las enzimas se pueden inmovilizar en diferentes medios, incluyendo

agua, tampón Tris-HCl y solución tamponada con fosfato. La enzima se puede inmovilizar con cualquier tipo de sustrato, p. ej. filtros, fibras, columnas, perlas, coloides, geles, hidrogeles, mallas y similares.

Uso de los polipéptidos descritos en la presente memoria

En un aspecto, se proporciona un método para producir un producto lácteo tratando un sustrato a base de leche que comprende lactosa con un polipéptido como se describe en la presente memoria. En un aspecto adicional, se proporciona un método para producir un producto lácteo tratando un sustrato a base de leche que comprende lactosa con un polipéptido que tiene una actividad transgalactosilante relativa por encima de 60%, tal como por encima de 70%, tal como por encima de 75% después de 15 min de reacción. En un aspecto, la actividad transgalactosilante relativa está por encima de 3 después de 30 min de reacción. En un aspecto adicional, la actividad transgalactosilante relativa está por encima de 6 después de 30 min de reacción. Y en otro aspecto, la actividad transgalactosilante relativa está por encima de 12 después de 30 min de reacción. En un aspecto, se proporciona un método, en donde el tratamiento con un polipéptido como se describe en la presente memoria tiene lugar a una temperatura óptima para la actividad de la enzima. En un aspecto adicional, el polipéptido se añade al sustrato a base de leche a una concentración de 0,01-1000 ppm. Y en un aspecto adicional, el polipéptido se añade al sustrato a base de leche a una concentración de 0,1-100 ppm. En un aspecto adicional, el polipéptido se añade al sustrato a base de leche a una concentración de 1-10 ppm. En un aspecto, se proporciona un método que comprende además fermentar un sustrato tal como un producto lácteo con un microorganismo. En un aspecto adicional, el producto lácteo es yogurt. En un aspecto adicional, el tratamiento con el polipéptido y el microorganismo se realiza esencialmente al mismo tiempo. En un aspecto, el polipéptido y el microorganismo se añaden al sustrato a base de leche esencialmente al mismo tiempo.

En un aspecto, se proporciona una composición preferiblemente una composición alimentaria, más preferiblemente un producto lácteo que comprende una célula o un polipéptido como se describe en la presente memoria.

En un aspecto, se proporciona un producto lácteo que comprende una célula o un polipéptido como se describe en la presente memoria. En un aspecto, el polipéptido como se define en la presente memoria se añade en una concentración de 0,01-1000 ppm. En un aspecto, se proporciona un producto lácteo que comprende un polipéptido inactivado como se define en la presente memoria. En un aspecto, se proporciona un producto lácteo que comprende un polipéptido inactivado como se define en la presente memoria en una concentración de 0,01-1000 ppm. En un aspecto, se proporciona un producto lácteo que comprende GOS formado in situ por un polipéptido como se define en la presente memoria. En un aspecto, se proporciona un producto lácteo que comprende una célula como se define en la presente memoria.

Un producto lácteo como se describe en la presente memoria puede ser, p. ej., leche desnatada, leche baja en grasa, leche entera, nata, leche UHT, leche que tiene una vida útil prolongada, un producto lácteo fermentado, queso, yogur, mantequilla, leche, suero de leche, bebida de leche acidificada, crema agria, bebida a base de suero de leche, helado, leche condensada, dulce de leche o una bebida láctea con sabor. Un producto lácteo se puede fabricar mediante cualquier método conocido en la técnica.

Un producto lácteo puede comprender adicionalmente componentes no lácteos, p. ej. componentes vegetales tales como, p. ej., aceite vegetal, proteína vegetal y/o hidratos de carbono vegetales. Los productos lácteos también pueden comprender otros aditivos tales como, p. ej., enzimas, aromatizantes, cultivos microbianos tales como cultivos probióticos, sales, edulcorantes, azúcares, ácidos, frutas, zumos de frutas o cualquier otro componente conocido en la técnica como componente de, o aditivo a, un producto lácteo.

Uno o más componentes de leche y/o fracciones de leche representan al menos el 50% (peso/peso), tal como al menos 70%, p. ej. al menos 80%, preferiblemente al menos 90%, del producto lácteo.

Uno o más sustratos a base de leche que se han tratado con una enzima como se define en la presente memoria que tiene actividad transgalactosilante pueden representar al menos 50% (peso/peso), tal como al menos 70%, p. ej. al menos 80%, preferiblemente al menos 90%, del producto lácteo.

El producto lácteo puede ser un producto lácteo que no se enriquece mediante la adición de galactooligosacáridos preproducidos.

El sustrato a base de leche tratado con polipéptido puede no secarse antes de utilizarse como un ingrediente en el producto lácteo.

El producto lácteo puede ser helado. En el presente contexto, el helado puede ser cualquier tipo de helado, como el helado con toda la crema, el helado bajo en grasa o el helado a base de yogur u otros productos lácteos fermentados. El helado se puede fabricar mediante cualquier método conocido en la técnica.

El producto lácteo puede ser leche o leche condensada.

El producto lácteo puede ser leche UHT. La leche UHT en el presente contexto es leche que ha sido sometida a un proceso de esterilización que está destinado a matar todos los microorganismos, incluidas las esporas bacterianas.

El tratamiento UHT (temperatura ultra alta, por sus siglas en inglés) puede ser, p. ej., tratamiento térmico durante 30 segundos a 130ºC, o tratamiento térmico durante un segundo a 145ºC.

El producto lácteo puede ser leche ESL. La leche ESL en el presente contexto es leche que tiene una vida útil prolongada debido a la microfiltración y/o tratamiento térmico y que se puede mantener fresca durante al menos 15 días, preferiblemente durante al menos 20 días, en el estante de la tienda a 2-5ºC.

El producto lácteo puede ser un producto lácteo fermentado, p. ej., yogur.

Los microorganismos utilizados para la mayoría de los productos lácteos fermentados se seleccionan del grupo de bacterias generalmente denominadas bacterias ácido lácticas. Como se utiliza en la presente memoria, el término "bacteria ácido láctica" designa una bacteria gram-positiva, microaerófila o anaeróbica, que fermenta azúcares con la producción de ácidos que incluye ácido láctico como ácido producido predominantemente, ácido acético y ácido propiónico. Las bacterias de ácido láctico más útiles industrialmente se encuentran dentro del orden "Lactobacillales" que incluye *Lactococcus spp.*, *Streptococcus spp.*, *Lactobacillus spp.*, *Leuconostoc spp.*, *Pseudoleuconostoc spp.*, *Pediococcus spp.*, *Brevibacterium spp.*, *Enterococcus spp.* y *Propionibacterium spp.* Adicionalmente, las bacterias productoras de ácido láctico que pertenecen al grupo de bacterias anaerobias, bifidobacterias, es decir, *Bifidobacterium spp.*, que se utilizan frecuentemente como cultivos alimenticios solos o en combinación con bacterias ácido lácticas, se incluyen generalmente en el grupo de bacterias de ácido lácticas.

Las bacterias ácido lácticas normalmente se suministran a la industria láctea como cultivos congelados o liofilizados para la propagación de iniciador a granel o los llamados cultivos "Direct Vat Set" (DVS), destinados a la inoculación directa en un recipiente de fermentación o cuba para la producción de un producto lácteo fermentado. Tales cultivos se conocen en general como "cultivos iniciadores" o "iniciadores".

Las cepas de cultivo iniciador comúnmente utilizadas de bacterias ácido lácticas generalmente se dividen en organismos mesófilos que tienen temperaturas de crecimiento óptimas a aproximadamente 30ºC y organismos termófilos que tienen temperaturas de crecimiento óptimas en el intervalo de aproximadamente 40 a aproximadamente 45ºC. Los organismos típicos que pertenecen al grupo mesófilo incluyen *Lactococcus lactis, Lactococcus lactis subsp. cremoris, Leuconostoc mesenteroides subsp. cremoris, Pseudoleuconostoc mesenteroides subsp. cremoris, Pediococcus pentosaceus, Lactococcus lactis subsp. lactis biovar. diacetylactis, Lactobacillus casei subsp. casei y Lactobacillus paracasei subsp. paracasei*. Las especies bacterianas ácido lácticas termófilas incluyen como ejemplos *Streptococcus thermophilus, Enterococcus faecium, Lactobacillus delbrueckii subsp. lactis, Lactobacillus helveticus, Lactobacillus delbrueckii subsp. bulgaricus* y *Lactobacillus acidophilus*.

También las bacterias anaeróbicas que pertenecen al género Bifidobacterium que incluyen *Bifidobacterium bifidum*, *Bifidobacterium animalis* y *Bifidobacterium longum* se utilizan comúnmente como cultivos lácteos iniciadores y generalmente se incluyen en el grupo de bacterias ácido lácticas. Además, las especies de Propionibacteria se utilizan como cultivos iniciadores lácteos, en particular en la fabricación de queso. Además, los organismos que pertenecen al género Brevibacterium se utilizan comúnmente como cultivos iniciadores alimentarios.

Otro grupo de cultivos microbianos iniciadores son cultivos fúngicos, que incluyen cultivos de levadura y cultivos de hongos filamentosos, que se utilizan particularmente en la fabricación de ciertos tipos de queso y bebidas. Los ejemplos de hongos incluyen *Penicillium roqueforti, Penicillium candidum, Geotrichum candidum, Torula kefir, Saccharomyces kefir y Saccharomyces cerevisiae*.

El microorganismo utilizado para la fermentación del sustrato a base de leche puede ser *Lactobacillus casei* o una mezcla de *Streptococcus thermophilus* y *Lactobacillus delbrueckii subsp. bulgaricus*.

Los métodos de fermentación son bien conocidos y la persona experta en la técnica sabrá cómo seleccionar condiciones del método adecuadas, tales como temperatura, oxígeno, cantidad y características de microorganismo o microorganismos, aditivos tales como p. ej. hidratos de carbono, aromas, minerales, enzimas y tiempo de proceso. Obviamente, las condiciones de fermentación se seleccionan para apoyar el logro deseado.

Como resultado de la fermentación, se reducirá el pH del sustrato a base de leche. El pH de un producto lácteo fermentado puede estar, p. ej., en el rango 3,5-6, tal como en el rango 3,5-5, preferiblemente en el rango 3,8-4,8.

En un aspecto, se proporciona un método para utilizar los polipéptidos o utilizar uno o más de los tipos de células mencionados anteriormente para producir oligosacáridos. Los oligosacáridos comprenden, pero no se limitan a fructooligosacáridos, galactooligosacáridos, isomaltooligosacáridos, maltooligosacáridos, lactosacarosa y xilooligosacáridos.

Los oligosacáridos se pueden producir incubando la célula que expresa el polipéptido en un medio que comprende un sustrato de disacárido tal como, por ejemplo, lactulosa, trehalosa, ramnosa, maltosa, sacarosa, lactosa o celobiosa. La incubación se lleva a cabo en condiciones en las que se producen oligosacáridos. Las células pueden ser parte de un producto seleccionado del grupo que consiste en yogur, queso, productos lácteos fermentados, suplementos dietéticos y productos comestibles probióticos. Alternativamente, los oligosacáridos se pueden recuperar y posteriormente añadirse al producto de interés antes o después de su preparación.

En un aspecto, se proporciona el uso de una célula descrita en la presente memoria para producir un producto seleccionado del grupo que consiste en yogur, queso, producto de leche fermentada, suplemento dietético y producto comestible probiótico.

En un aspecto, los polipéptidos descritos en la presente memoria se pueden utilizar para preparar productos de queso y en métodos para preparar los productos de queso. Los productos de queso pueden, p. ej. ser seleccionados del grupo que consiste en queso crema, requesón y queso procesado. Añadiendo polipéptidos, los quesos pueden contener niveles significativamente aumentados de galactooligosacáridos y niveles reducidos de lactosa. En un aspecto, los niveles de lactosa en el producto de queso final se pueden reducir en al menos aproximadamente 25 por ciento, preferiblemente al menos aproximadamente 50 por ciento, y más preferiblemente al menos aproximadamente 75 por ciento. Los polipéptidos se pueden utilizar para reducir la lactosa en los productos de queso a menos de aproximadamente 1 gramo por porción, una cantidad que puede ser tolerada por la mayoría de los individuos intolerantes a la lactosa.

Los productos de queso proporcionados en la presente memoria son productos de queso enriquecidos nutricionalmente que tienen un contenido de fibra soluble aumentada, contenido calórico reducido, excelentes propiedades organolépticas, textura y sabor mejorados. Además, los polipéptidos descritos en la presente memoria pueden reducir el índice glucémico de los productos de queso porque los GOS se absorben más lentamente que la lactosa o sus productos de hidrólisis. Finalmente, los polipéptidos pueden reducir el coste de producción de productos de queso, particularmente productos de queso crema, porque los GOS sorprendentemente proporcionan una textura mejorada al producto de queso crema, permitiendo de este modo el uso reducido de estabilizadores, o permitiendo un contenido de humedad incrementado sin sinéresis.

En un aspecto adicional, se proporciona una composición que comprende un polipéptido como se describe en la presente memoria y un sustrato de hidrato de carbono. En un aspecto adicional, el sustrato de hidrato de carbono es un disacárido. En un aspecto adicional, el disacárido es, por ejemplo, lactulosa, trehalosa, ramnosa, maltosa, sacarosa, lactosa o celobiosa. En otro aspecto más, el sustrato de hidrato de carbono es lactosa. La composición se prepara de forma tal que se producen oligosacáridos. El polipéptido como se describe en la presente memoria puede ser parte de un producto seleccionado del grupo que consiste en yogur, queso, productos lácteos fermentados, suplementos dietéticos y productos comestibles probióticos. En un aspecto, se proporciona una composición que comprende un polipéptido como se describe en la presente memoria y un estabilizante. Ejemplos de estabilizantes son, p. ej., un poliol tal como, p. ej., glicerol o propilenglicol, un azúcar o un alcohol de azúcar, ácido láctico, ácido bórico o un derivado de ácido bórico (p. ej., un éster de borato aromático).

En un aspecto, se proporciona el uso de un polipéptido transgalactosilante como se describe en la presente memoria o una célula como se describe en la presente memoria, para producir galactooligosacáridos. En un aspecto, se proporciona el uso de un polipéptido transgalactosilante descrito en la presente memoria o una célula como se describe en la presente memoria, para producir galactooligosacáridos para ser parte de un producto seleccionado del grupo que consiste en yogur, queso, productos lácteos fermentados, suplementos dietéticos y productos comestibles probióticos. En un aspecto, el producto es yogur, queso o productos lácteos fermentados. En un aspecto, se proporciona el uso de un polipéptido transgalactosilante como se describe en la presente memoria o una célula como se describe en la presente memoria, para producir galactooligosacáridos para potenciar el crecimiento de Bifidobacterium. En un aspecto, se proporciona el uso de un polipéptido transgalactosilante como se describe en la presente memoria o una célula como se describe en la presente memoria, para producir galactooligosacáridos para potenciar el crecimiento de Bifidobacterium en una fermentación de cultivo mixto.

En un aspecto, se proporciona un método para producir un polipéptido transgalactosilante como se describe en la presente memoria, que comprende cultivar una célula como se describe en la presente memoria en un medio de cultivo adecuado bajo condiciones que permiten la expresión de dicho polipéptido y recuperar el polipéptido resultante del cultivo. Se proporciona un método para producir galactooligosacáridos, que comprende poner en contacto un polipéptido como se describe en la presente memoria o una célula como se describe en la presente memoria con una solución a base de leche que comprende lactosa.

La adición de oligosacáridos puede potenciar el crecimiento de Bifidobacterium solo o de Bifidobacterium en un cultivo mixto.

El tratamiento de productos lácteos con enzimas que convierten la lactosa en monosacáridos o GOS tiene varias ventajas. En primer lugar, los productos pueden ser consumidos por personas con intolerancia a la lactosa que, de lo contrario, presentarían síntomas como la flatulencia y la diarrea. En segundo lugar, los productos lácteos tratados con lactasa tendrán un dulzor más alto que los productos no tratados similares debido al mayor dulzor percibido de la glucosa y la galactosa en comparación con la lactosa. Este efecto es particularmente interesante para aplicaciones tales como yogur y helado, donde se desea un alto dulzor del producto final y esto permite una reducción neta de hidratos de carbono en el producto consumido. En tercer lugar, en la producción de helados se observa a menudo un fenómeno denominado arenosidad, donde las moléculas de lactosa cristalizan debido a la baja solubilidad relativa de la lactosa. Cuando la lactosa se convierte en monosacáridos o GOS, la sensación en la boca del helado mejora mucho con respecto a los productos no tratados. La presencia de una sensación arenosa debido a la cristalización de lactosa se puede eliminar y los costos de la materia prima se pueden reducir

reemplazando la leche desnatada en polvo por suero en polvo. Los principales efectos del tratamiento enzimático fueron dulzor aumentado.

En un aspecto, el o los polipéptidos transgalactosilantes como se describe en la presente memoria se pueden utilizar junto con otras enzimas tales como proteasas tales como quimosina o renina, lipasas tales como fosfolipasas, amilasas, transferasas y lactasas. En un aspecto, el o los polipéptidos transgalactosilantes como se describe en la presente memoria se pueden utilizar junto con lactasa. Esto puede ser especialmente útil cuando existe el deseo de reducir la lactosa residual después del tratamiento con el o los polipéptidos transgalactosilantes como se describe en la presente memoria especialmente a bajos niveles de lactosa. Una lactasa en la presente memoria es cualquier glucósido hidrolasa que tiene la capacidad de hidrolizar el disacárido lactosa en galactosa y en monómeros de glucosa constituyentes. El grupo de lactasas comprende, pero no se limita a, enzimas asignadas a la subclase EC 3.2.1.108. Las enzimas asignadas a otras subclases, tales como, p. ej., EC 3.2.1.23, también pueden ser lactasas en el presente contexto. Una lactasa en el presente contexto puede tener otras actividades además de la actividad hidrolítica de la lactosa, tal como, por ejemplo, una actividad transgalactosilante. En el presente contexto, la actividad hidrolítica de lactosa de la lactasa se puede denominar como su actividad lactasa o su actividad beta-galactosidasa. Las enzimas que tienen actividad lactasa pueden ser de origen animal, vegetal o microbiano. Las enzimas preferidas se obtienen a partir de fuentes microbianas, en particular a partir de un hongo filamentoso o levadura, o de una bacteria. La enzima se puede, p. ej., obtener a partir de una cepa de Agaricus, p. ej. A. bisporus; Ascovaginospora; Aspergillus, p. ej. A. niger, A. awamori, A. foetidus, A. japonicus, A. oryzae; Candida; Chaetomium; Chaetotomastia; Dictyostelium, p. ej. D. discoideum; Klu- veromyces, p. ej. K. fragilis, K. lactis; Mucor, p. ej. M. javanicus, M. mucedo, M. subtilissimus; Neurospora, p. ej. N. crassa; Rhizomucor, p. ej. R. pusillus; Rhizopus, p. ej. R. arrhizus, R, japonicus, R. stolonifer; Sclerotinia, p. ej. S. libertiana; Torula; Torulopsis; Trichophyton, p. ej. T. rubrum; Whetzelinia, p. ej. W. sclerotiorum; Bacillus, p. ej. B. coagulans, B. circulans, B. megaterium, B. novalis, B. subtilis, B. pumilus, B. stearothermophilus, B. thuringiensis; Bifidobacterium, p. ej. B. longum, B. bifidum, B. animalis; Chryseobacterium; Citrobacter, p. ej. C. freundii; Clostridium, p. ej. C. perfringens; Diplodia, p. ej. D. gossypina; Enterobacter, p. ej. E. aerogenes, E. cloacae Edwardsiella, E. tarda; Erwinia, p. ej. E. herbicola; Escherichia, p. ej. E. coli; Klebsiella, p. ej. K. pneumoniae; Miriococcum; Myrothesium; Mucor; Neurospora, p. ej. N. crassa; Proteus, p. ej. P. vulgaris; Providencia, p. ej. P. stuartii; Pycnoporus, p. ej. Pycnoporus cinnabarinus, Pycnoporus sanguineus; Ruminococcus, p. ej. R. torques; Salmonella, p. ej. S. typhimurium; Serratia, p. ej. S. liquefasciens, S. marcescens; Shigella, p. ej. S. flexneri; Streptomyces, p. ej. S. antibioticus, S. castaneoglobisporus, S. violeceoruber; Trametes; Trichoderma, p. ej. T. reesei, T. viride; Yersinia, p. ej. Y. enterocolitica. La lactasa puede ser un componente intracelular de microorganismos como Kluyveromyces y Bacillus. Kluyveromyces, especialmente K. fragilis y K. lactis, y otros hongos tales como los de los géneros Candida, Torula y Torulopsis, son una fuente común de lactasas fúngicas, mientras que B. coagulans y B circulans son fuentes bien conocidas de lactasas bacterianas. Varias preparaciones de lactasa comerciales derivadas de estos organismos están disponibles tales como Lactozym.RTM. (disponible de Novozymes, Dinamarca), HA-Lactase (disponible de Chr. Hansen, Dinamarca) y Maxilact.RTM. (disponible de DSM, Países Bajos), todos a partir de K. lactis. Todas estas lactasas se denominan lactasas neutras que tienen un pH óptimo entre pH 6 y pH 8. Cuando tales lactasas se utilizan en la producción de, p. ej., yogur bajo en lactosa, el tratamiento enzimático se tendrá que hacer en un paso separado antes de la fermentación o se deben utilizar dosificaciones enzimáticas altas, porque su actividad disminuye a medida que disminuye el pH durante la fermentación. Además, estas lactasas no son adecuadas para la hidrólisis de lactosa en leche realizada a alta temperatura, lo que en algunos casos sería beneficioso para mantener el recuento microbiano bajo y garantizar así una buena calidad de la leche.

La enzima puede ser una lactasa de una bacteria, p. ej. de la familia Bifidobacteriaceae, tal como del género Bifidobacterium tal como la lactasa descrita en el documento WO 2009/071539.

Ejemplo 1

Producción de polipéptidos

Se adquirió un gen sintético de *Ruminococcus hansenii* con codones optimizados para la expresión en *Bacillus subtilus* de GeneART (Regensburg, Alemania). El gen sintético se clonó en el vector de expresión pBN *Bacillus subtilus* (figura 1) y se transformó en la cepa BG6006 de *Bacillus subtilus*. Los transformantes se volvieron a sembrar dos veces en placas LB que contenían 10 mg/ml de neomicina como selección.

Se preparó un precultivo en medio LB que contenía 10 mg/ml de neomicina y se cultivó durante 7 horas a 37ºC y a 180 rpm de agitación. Se utilizaron 500 ml de este precultivo para inocular 50 ml de medio modificado de Grant que contenía 10 mg/ml de neomicina al que se dejó crecer durante 48 horas a 33ºC y a 180 rpm de agitación.

Los cultivos se recogieron por centrifugación a 10.000 x g y se filtraron en condiciones estériles. Los caldos de fermentación se concentraron por arriba utilizando Sartorius Vivaspin20 MWCO 10,000 Dalton (Código de producto VS2002) a 4000 rpm en una centrífuga de mesa. El concentrado se estabilizó con un 20% p/p de glicerol.

El medio modificado de Grant se prepararó según las siguientes instrucciones:

PARTE I (Autoclave)	
Soytone	10g
Llevar a	500 ml por litro
PARTE II	
K_2NPO_4 1M	3 ml
Glucosa	75 g
Urea	3,6 g
10X MOPS de Grant	100 ml
Llevar a 400 ml por litro	

La PARTE I (2% p/p de Soytone) se preparó, y se trató en un autoclave durante 20-25 minutos. La PARTE II se preparó, y se mezcló con la PARTE 1 y el pH se ajustó a pH a 7,3 con HCl/NaOH.

El volumen se llevó a volumen completo y se esterilizó a través de un filtro de PES de 0,22 µm.

El tampón 10X MOPS se preparó según las siguientes instrucciones:

83,72 g	Ticina
7,17 g	Gránulos de KOH
12 g	NaCl
29,22 g	K_2SO_4 0,276M
10 ml	$MgCl_2$ 0,528M
10 ml	Micronutrientes de Grant 100X
Llevar a 100 ml.	

Los Micronutrientes 100X se preparan según las siguientes instrucciones:

Citrato de sodio.2H2O	1,47 g
CaCl2. 2H2O	1,47 g
FeSO4.7H2O	0,4 g
MnSO4.H2O	0,1 g
ZnSO4.H2O	0,1 g
CuCl2.2H2O	0,05 g
CoCl2.6H2O	0,1 g
Na2MoO4.2H2O	0,1 g

El volumen se alcanzó con agua milliQ.

La esterilización se realizó a través de un filtro de PES de 0,2 µm.

La protección de la luz fue envolviendo en papel de aluminio.

El almacenamiento fue a 4ºC.

Determinación de la actividad de hidrólisis de las preparaciones enzimáticas

La actividad enzimática de *Ruminococcus hansenii* (SEQ ID NO: 1), *Ruminococcus lactaris* (SEQ ID NO:2) y *Bifidobacterium bifidum BIF3d3* (truncado) (como se describe por Jørgensen et al. (2001), Appl. Microbiol.

Biotechnol., 57: 647 - 652 y patente EP 1.283.876) se midió usando el sustrato comercialmente disponible 2-nitrofenil-β-D-galactopiranósido (ONPG) (Sigma N1127).

Composición de tampón 1 x ONPG:

50	mM Na-Citrato
100	mM NAPO4
2	mM CaCl2
1	mM MgCl2
20	mM ONPG

Las series de dilución de las enzimas anteriores y Lactozym® (de Novozymes) como control estándar se realizaron en placas de microtitulación de 96 pocillos. Se transfirieron 75 ml de las diluciones a una nueva placa de microtitulación y se mezclaron con 75 ml de tampón 2 x ONPG concentrado. Las mediciones de absorbancia se registraron a 450 nm en un dispositivo molecular SpectraMax controlado por el paquete de software Softmax. La cámara se equilibró a 37ºC y se realizaron grabaciones cada 15 segundos durante 10 minutos en total. Se midió la generación de ONP y se determinó la Vmax de la reacción. El Vmax para cada preparación enzimática se comparó con concentraciones conocidas de (3000 LAU/ml) Lactozym® y la actividad en LAU/ml se calculó a partir del estándar Lactozym® (véase la Tabla 1 a continuación).

Tabla: 1

Enzima	LAU/ml
Lactozym®	3000
Bifidobacterium bifidum BIF3s3 (truncada)*	105
Ruminococcus lactaris (SEQ ID NO:2)	45
Ruminococcus hansenii (SEQ ID NO:1)	42

Ejemplo 2

Definición de unidad de enzima productora de GOS

En la presente solicitud, la actividad transgalactosilante relativa se define como la diferencia entre la cantidad de glucosa liberada restada de la cantidad de galactosa liberada dividida por la cantidad de galactosa generada en tampón T a 37ºC.

Ecuación 1:

$$\text{Actividad relativa transgalactosilante} = \frac{[\text{Glucosa}] - [\text{Galactosa}]}{[\text{Glucosa}]}$$

El tampón T se preparó de la siguiente manera:

50	mM No-citrato
100	mM Na – PO4
2	mM CaCl2
1	mM MgCl2
5	p/p % Lactosa
pH 6,0	

Medición de galactosa y glucosa por cromatografía HPLC

La galactosa y la glucosa se analizaron utilizando un sistema Dionex ICS3000 que consta del automuestreador ICS-3000 AS, el detector ED ICS-3000, el módulo de cromatografía DC ICS-3000 y una bomba de gradiente DP (Dionex Corp, Sunnyvale, California, EE.UU.).

La galactosa y la glucosa se separaron utilizando una columna CarboPac PA1 de 4 mm con una columna de protección CarboPac PA1 de 4 mm (Dionex Corp, Sunnyvale, California, EE.UU.). El flujo fue de 1 ml/min. El gradiente se realizó según la tabla 2, y la cuantificación se realizó con el uso de estándares externos.

Tabla 2: Programa de gradiente (p/p %) utilizado para el análisis de monosacáridos en muestras

Tiempo (min)	Agua Milli Q	NaOH 150 mM
0-12	90%-85%	10%-15%
12-25	85%-0%	15%-100%
25-30	0%	100%
30-32	0%-90%	100%-10%
32-34	90%	10%

Los eluyentes utilizados fueron agua y NaOH 150 mM. Se preparó NaOH 150 mM (eluyente) desgasificando 2 l de agua Milli Q durante 10 minutos y añadiendo 16 ml de NaOH al 50% p/p y desgasificando durante otros 5 minutos.

Cálculo de la actividad transgalactosilante

La actividad relativa transgalactosilante se calculó según la ecuación 1 y las concentraciones de glucosa y galactosa se midieron mediante HPLC.

Tabla 3: Concentración de Galactosa en %:

Tiempo/min	0	15	30	60	120	180
Lactozym®	0	1,5	1,9	2,1	2,3	2,1
Bifidobacterium, bifidum BIF3d3 (truncado)	0	0,5	1	1	1	0,9
Ruminococcus hansenii (SEQ ID NO:1)	0	0,1	0,1	0,2	0,2	0,2
Ruminococcus lactaris (SEQ ID NO:2)	0	0,1	0,1	0,1	0,2	0,3

Tabla 4: Concentración de Glucosa en %:

Tiempo/min	0	15	30	60	120	180
Lactozym®	0	1,9	2,3	2,3	2,4	2,2
Bifidobacterium, bifidum BIF3d3 (truncado)	0	1,5	1,8	1,8	1,6	1,8
Ruminococcus hansenii (SEQ ID NO:1)	0	0,7	1,3	1,2	1,3	1,3
Ruminococcus lactaris (SEQ ID NO:2)	0	0,4	0,5	0,8	0,7	0,8

Tabla 5: Relación de actividad transgalactosilante:Actividad β-galactosidasa

Tiempo/min	0	15	30	60	120	180
Lactozym®	nd	0,27	0,21	0,10	0,04	0,05
Bifidobacterium, bifidum BIF3d3 (truncado)	nd	2,00	0,80	0,80	0,60	1,00
Ruminococcus hansenii (SEQ ID NO:1)	nd	6,00	12,00	5,00	5,50	5,50
Ruminococcus lactaris (SEQ ID NO:2)	nd	3,00	4,00	7,00	2,50	1,67
nd: No determinado para este punto de tiempo.						

Tabla 6: Actividad transgalactosilante relativa en %:

Tiempo/min	0	15	30	60	120	180
Lactozym®	nd	21,05	17,39	8,70	4,17	4,55
Bifidobacterium, bifidum BIF3d3 (truncado)	nd	66,67	44,44	44,44	37,50	50,00
Ruminococcus hansenii (SEQ ID NO:1)	nd	85,71	92,31	83,33	84,62	84,62
Ruminococcus lactaris (SEQ ID NO:2)	nd	75,00	80,00	87,50	71,43	62,50
nd: No determinado para este punto de tiempo.						

La Fig. 2 muestra la acumulación de glucosa y galactosa en el tiempo extra. Como es claramente evidente a partir de la Fig. 2 y las tablas anteriores, las enzimas de *Ruminococcus hansenii* (SEQ ID NO: 1) y *Ruminococcus lactaris* (SEQ ID NO: 2) generan solo entre 10 - 20% de galactosa en relación con la enzima de *Bifidobacterium, bifidum* BIF3d3 (truncada). Estos hallazgos sugieren que las enzimas de *Ruminococcus hansenii* (SEQ ID NO: 1) y *Ruminococcus lactaris* (SEQ ID NO: 2) son capaces de excluir agua del sitio activo de manera más eficiente que Lactozym® y la enzima de *Bifidobacterium*, *bifidum* BIF3d3 (truncada).

Ejemplo 3

Ensayo en leche

Se prepararon muestras en 9% p/p de leche reconstituida a partir de leche desnatada en polvo (Humana Milk Union, DE NW508 EG) dando una concentración final de lactosa de 5% p/p. Las enzimas se dosificaron en base a la actividad de LAU determinada como se describió anteriormente a una concentración final de 6 LAU/ml. Se tomó una muestra antes de la adición de la enzima y se tomaron muestras adicionales en puntos de tiempo indicados y las enzimas se inactivaron inmediatamente incubando a 95ºC durante 10 minutos. Las muestras se diluyeron 1:10 y se aplicaron 2 ml sobre placas de gel de sílice 60 HPTLC activadas (161ºC durante 10 minutos) (Merck Cat NO 1.05641.0001) con un CAMAG Automatic TLC Sampler 4. Las placas de TLC se eluyeron con un eluyente que contiene (80) acetonitrilo: (20) acetato de etilo: (50) 1-propanol: (40) agua. Las muestras se visualizaron calentando (161ºC durante 10 minutos) y se dejaron enfriar antes de sumergirlas en 5% p/p de H2SO4 en etanol al 99,9%. Las placas se desarrollaron con calentamiento a 161ºC durante 3 min.

Tabla 7: Composición de estándares:

	Conc Std A (p/p %)	Conc Std B (p/p %)	Conc Std C (p/p %)
Glucosa	0,5	0,4	0,1
Lactosa	0,3	0,2	0,5
Galactosa	0,1	0,05	0,3

El número de muestra en la fig. 3 es como se muestra en la tabla a continuación:

Número de muestra:

1	Std A	8	Lactozym® 120 min	15	*B.bifidum* 180 min	22	*R. lactaris* 0 min
2	Std B	9	Lactozym® 180 min	16	*R. hansenii* 0 min	23	*R. lactaris* 15 min
3	Std C	10	*B.bifidum* 0 min	17	*R. hansenii* 15 min	24	*R. lactaris* 30 min
4	Lactozym® 0 min	11	*B.bifidum* 15 min	18	*R. hansenii* 30 min	25	*R. lactaris* 60 min
5	Lactozym® 15 min	12	*B.bifidum* 30 min	19	*R. hansenii* 60 min	26	*R. lactaris* 120 min
6	Lactozym® 30 min	13	*B.bifidum* 60 min	20	*R. hansenii* 120 min	27	*R. lactaris* 180 min
7	Lactozym® 60 min	14	*B.bifidum* 120 min	21	*R. hansenii* 180 min		

La Fig. 3 muestra la composición de azúcar de la leche en diversos puntos de tiempo durante la incubación. Mientras que Lactozym® genera cantidades aproximadamente iguales de glucosa y galactosa, las enzimas de *Bifidobacterium, bifidum* BIF3d3 (truncadas), *Ruminococcus hansenii* (SEQ ID NO: 1) y *Ruminococcus lactaris* (SEQ ID NO: 2) generan todas más glucosa que galactosa. Estos resultados son indicativos de que todas estas enzimas son capaces de realizar transgalactosilación en leche reconstituida con una concentración de lactosa inicial de 5% p/p.

Ejemplo 4

Actividad de mutantes de núcleo catalítico de *Ruminococcus hansenii*

Purificación de la enzima a partir de muestras crudas

Se obtuvieron muestras de enzimas crudas como se describe en el ejemplo 1.

Método de purificación

Cromatografía de intercambio iónico, Q HiTrap HP FF 5 ml

La columna se preparó como se describe por el fabricante y se equilibró en tampón Tris/HCl 20 mM, pH 8,0 (tampón A).

La muestra (5 ml) se desaló en Tampón A y se aplicó a la columna a un caudal de 4 ml/min. La columna se lavó con tampón A y las proteínas unidas se eluyeron con un gradiente lineal de NaCl 0-0,6 M en tampón A. Durante todo el recorrido se recogieron fracciones de 4 ml.

Ensayo de actividad

Se mezclaron 90 ml de tampón de reacción con 30 ml de la muestra de enzima diluida indicada (tabla 8) en una placa de PCR Eppendorf twin tech de 96 pocillos (Cat. 951020401) y se incubaron durante 30 minutos a 42ºC en un equipo PCR Eppendorf Mastercycler Gradient. La reacción se detuvo transfiriendo la mezcla a una placa Costar 9017 de 96 pocillos que contenía 120 ml de Na-carbonato 10% (solución de parada). Las reacciones se midieron a 420 nm en un lector de placas Spectra Max 190 de Molecular Devices.

Tabla 8:

	Proteína, µg/ml	Diluída	OD420	Tiempo/min	Actividad/min, %
E592Q Frac. 13	62	10X	0,011	25	3
D625N Frac. 14	11	2X	0,168	25	50
D588N Frac. 14	6	1X	0,019	25	6
D576N Frac. 17	29	5X	0,270	5	81
WT Frac. 17	33	5,5X	0,335	5	100

Las concentraciones de proteína se ajustaron a la de la muestra D588N diluyendo con tampón Na-P 50 mM (pH 7,0) y se midió la actividad como se describe anteriormente.

La Tabla 8 muestra la concentración de proteína en las fracciones indicadas, el factor de dilución para alcanzar la concentración de D588N, la medición de OD420, el tiempo de reacción en minutos y la actividad relativa por minuto de la enzima de tipo salvaje (WT) de *Ruminococcus hansenii*.

La Fig. 4 muestra los resultados de la cromatografía de intercambio aniónico de las variantes anteriores de *Ruminococcus hansenii* (SEQ ID NO: 1). El gel es un gel de acrilamida Nu-PAGE 4-12% teñido con tinción azul brillante coomassie.

Calle	Muestra
1	E592Q cruda
2	Frac. 13 eluida
3	Frac. 14 eluida
4	D625N cruda
5	Frac. 14 eluida
6	Frac. 15 eluida
7	D588N cruda
8	Frac.14 eluida
9	Frac.17 eluida

Calle	Muestra
13	D576N cruda
14	Frac. 17 eluida
15	Frac. 18 eluida
16	WT cruda
17	Frac. 17 eluida
18	Frac. 18 eluida

Lista de secuencias

La SEQ ID NO: 1 es un fragmento truncado de 1125 aminoácidos de la SEQ ID NO: 12.

```
KADSQTQMSS EPEQVAVKDY GSNSARTQNF DSDWKFNLGD VSNAQTPTFD DSKWRTLSLP    60
HDYSIEQEYS QSLEAESGYL PGGVGWYRKN FTLGEEAKGK RIRIDFDGVY MNATVYVNGK   120
EVGTHPYGYT PFSFDITDYI SYDKENTIAV KVDHQTPSSR WYSGSGIYRS VNLTTTNDVH   180
VDLNGIKVES NNLEKEAGKT VNTDVKTTVV NGSKEAKNIT ITHTVFKKGE KPDKAIGTFT   240
TEAQEIGAGK KTEISATVPV KNPELWSVEN PALYTIRTEV KAGDKLLDSY DTEYGFHYLN   300
FDTETGFQLN GKNVKLKGVC MHHDQGALGA VANRRAIERQ VEILQEMGCN SIRVTHNPAS   360
KDLIEVCNEK GILVIEEVFD GWHRAKNGNS NDYSVWFEKA IEEDNAILGK EADMTWAEYD   420
LKAIMKRDQN APSIIEWSLG NEIQEGAGGS GYAERADKLI KWAKEADATK TLTIGSNAVK   480
RGDWEQVSIG DKLTKAGGTS GTNYSDGASY DKIHKEHPDW KLYGSETASS VNSRGIYSVT   540
GNQEATSDQQ LTAYDNSRVN WGALASQAWY DVIQRDFVAG EYVWTGFDYI GEPTPWNGTD   600
PGAKGTWPSP KNSYFGIIDT AGFPKDSYYF YQSQWNEEVN TLHVLPAWNE DVVKKNSDGT   660
VPVVVYSDAK EVELFFTPAN GGEKKSLGKK TFKTETTKAG YSYQVLENGK KKHKDLYMEW   720
QVPYEAGTLE AVAKDAKGNV IKDTEGRSVV KTTGEEAKLS AKTDRNSIQA DGKDLSYITV   780
DVTDKDGNIV PDAANRVTFD VQGAGKLVGV DNGSSPDHDS YKADNRKAFS GKVLAIVQST   840
EKAGEITVTA KADGLESSTV KITTTPVKEE PSERYVESYK YSKSYYVKTG TKPQLPKKIE   900
AQYSDRTKED VAVKWDEISD EQISKTGSFT VEGTVGKRDI TVNINMIDDV AALLNYSGAT   960
QKGVKPQLPD VRPAVLPDGT VLAASFPVQW DEKDADTFQK PDEIVTVNGS ADIFGKTIPV  1020
TASIRVQKED IKIGSSVTNV AKLSQNIQGS DTLEAIKDGK TEMSLNNDGG PNESAWSNWD  1080
ASQKGTKEAE LTFTFDTQQR IGEIVIHFAK DNNSIRFPDA GTTEI                  1125
```

La SEQ ID NO: 2 es un fragmento truncado de 1150 aminoácidos de la SEQ ID NO: 13.

```
AGVSVPALAQ QAVRTESQTQ MSSDPELVYV NNYSSTAQRS QNFNSNWKFY FGDAGNAQGA    60
TFDDSKWEQV SLPHDYSISQ EYSKSMEAES GYLGGGTGWY RKNFTLSSDT QGKRVRIDFD   120
GVYMNATVWV NGHEVGTHPY GYTSFSFDIT DYVKYDGENT IAVKVVNNTP SSRWYSGSGI   180
YRDVDLTITD DVHVDLNGTK VTTPNLETEK GSTVNTDVTA TVANDSDAAK SVAVRHTVFP   240
KDGSADQSIG TVTTNAQSIA AGATAEIQAT VPVSNPELWS VENPSLYTVR TEVLVDGQVT   300
DTYDTEYGFR YFNFDSNTGF SLNGENMKLK GVCMHHDQGS LGAAAYDSAI DRQVKILKEM   360
GCNSIRVTHN PAAQDLIDAC NEQGILVVEE AFDTWTRPKN GNSNDYSVWF NQTVASDNEI   420
LGATNGETWA QFDLESMISR DYNAPSVIMW SLGNEVMEGI SGGTDAEYEA TATKLINWAY   480
DADNTRPMTI GDNKLKANWQ ISKTFARLLT EKGGTVGFNY ADGRVLDSYH SSNSNWLLYG   540
SETASAINSR GIYYRTTGGG QTSDKQLTSY DNSNVGWGAT ASNAWYTVLT RDFAAGEYVW   600
TGFDYLGEPT PWNGTGSGAV GSWPSPKNSY FGIIDTAGFA KDSYYFYQSQ WNDDVTTLHV   660
LPAWNNNVVS KDSSGNVPVV VYSDAASVEL FFQAKGSDTK TSLGKKTFTQ KTTDAGYTYQ   720
IYEGSDKNST TDKNLYLTWN VPYADGTVSA VAYNSNGQKI TDTVGQSSVT TTGKASKLKA   780
SADHKKIAAD GESLSYITVD VTDANGNIVP DAENRVKFTV EGDGELVGVD NGSSPDHDSY   840
QADNRKAFSG KVLAIVKSTK EAGTITVTAS ADGLDSASVK ITTTAVDNGS TEKQIDSFKM   900
SRTYYVKVGS TPELPEKIVT RYTDGTSEEL PVTWDAITED QIAAAGSFQV KGTVKGGYSV   960
AVNVNMIDEV GGLLNYSTNT AVGVAPVLPT SRPAVLQDGT VMDVTFPVTW EDKAASAYDK  1020
AGTVTVNGTA NVLGKEIAVT ASVRVQEETI TIGDSVSADA LNLTQSVPAD KQSDTLNAIK  1080
DGSTTISSNT SGGANPTVWS NYDYSQDGNT TADIIFEYAT EQRLGQIVTH FARDSWSMRY  1140
PDAGATEIYV                                                         1150
```

La SEQ ID NO: 3 es los restos de aminoácidos 559-649 de la SEQ ID NO: 1.

```
VNWGALASQA WYDVIQRDFV AGEYVWTGFD YIGEPTPWNG TDPGAKGTWP SPKNSYFGII    60
DTAGFPKDSY YFYQSQWNEE VNTLHVLPAW N                                   91
```

La SEQ ID NO: 4 es los restos de aminoácidos 579-649 de la SEQ ID NO: 1.

```
AGEYVWTGFD YIGEPTPWNG TDPGAKGTWP SPKNSYFGII DTAGFPKDSY YFYQSQWNEE      60
VNTLHVLPAW N                                                           71
```

La SEQ ID NO: 5 es los restos de aminoácidos 579-636 de la SEQ ID NO: 1.

AGEYVWTGFD YIGEPTPWNG TDPGAKGTWP SPKNSYFGII DTAGFPKDSY YFYQSQWN 58

La SEQ ID NO: 6 es los restos de aminoácidos 575-665 de la SEQ ID NO: 2.

```
VGWGATASNA WYTVLTRDFA AGEYVWTGFD YLGEPTPWNG TGSGAVGSWP SPKNSYFGII      60
DTAGFAKDSY YFYQSQWNDD VTTLHVLPAW N                                     91
```

La SEQ ID NO: 7 es los restos de aminoácidos 594-665 de la SEQ ID NO: 2.

```
AGEYVWTGFD YLGEPTPWNG TGSGAVGSWP SPKNSYFGII DTAGFAKDSY YFYQSQWNDD      60
VTTLHVLPAW N                                                           71
```

La SEQ ID NO: 8 es los restos de aminoácidos 594-652 de la SEQ ID NO: 2.

AGEYVWTGFD YLGEPTPWNG TGSGAVGSWP SPKNSYFGII DTAGFAKDSY YFYQSQWN 58

La SEQ ID NO: 9 es un péptido señal del vector de expresión de Bacillus subtilis pBN.

VRSKKLWISLLFALALIFTMAFGSTSSAQA

La SEQ ID NO: 10 es la secuencia de nucleótidos que codifica la SEQ ID NO: 1 que incluye la secuencia que codifica el péptido señal.

```
gtgagaagcaaaaaattgtggatcagtttgctgtttgctttagcgttaatctttacgatggcgttcggcagcacatccagcgcgcaggc
ggcagggaaaaaagcagatagccaaacacaaatgtcatcagaaccggaacaagttgcggttaaagattatggctcaaatagcgc
acgcacacagaattttgatagcgattggaaatttaacctgggagatgttagcaatgcacagacaccgacatttgatgattcaaaatg
```

gcgcacactgtcactgccgcatgattatagcatcgaacaggaatattcacaatcactggaagcagaatcaggctatcttccgggagg
cgttggctggtatcgcaaaaattttacactgggcgaagaagcgaaaggcaaacgcattcgcattgattttgatggcgtctatatgaat
gcaacagtctatgtgaatggcaaagaagttggcacacatccgtatggctatacaccgtttagctttgatatcacagattatatcagcta
tgataaagaaaacacaattgcggtcaaagtcgatcatcaaacaccgtcatcaagatggtattcaggcagcggcatttatagatcagt
caacctgacaacaacaaatgatgtccatgtcgatctgaatggcattaaagtcgaaagcaacaacctggaaaaagaagcaggcaaa
acagtcaacacagatgtgaaaacaacagttgtgaacggctcaaaagaagcgaaaaacatcacaattacacatacagtctttaaaa
aaggcgaaaaaccggataaagcgatcggcacatttacaacagaagcgcaagaaattggcgcaggcaaaaaaacagaaatcagc
gcaacagtcccggttaaaaatccggaactgtggtcagttgaaaatccggcactgtatacaattcgcacagaagttaaagcaggcga
taaactgctggatagctatgatacagaatatggctttcattatctgaactttgatacagaaacaggctttcagctgaatggcaaaaac
gttaaactgaaaggcgtttgcatgcatcatgatcaaggcgcacttggcgcagttgcaaatagaagagcaattgaacgccaagtcga
aattctgcaagaaatgggctgcaatagcattagagtcacacataatccggcaagcaaagatctgattgaagtctgcaacgaaaaag
gcattctggtcattgaagaagtttttgacggctggcatagagcaaaaaatggcaacagcaacgattatagcgtctggtttgaaaaag
cgatcgaagaagataacgcgattctgggaaaagaagcggatatgacttgggcagaatatgatctgaaagcgattatgaaacgcga
tcaaaatgcaccgagcattattgaatggtcactgggcaatgaaattcaagaaggcgcaggcggatcaggctatgcagaaagagcg
gataaactgatcaaatgggcgaaagaagcagacgcaacaaaaacactgacaattggcagcaatgcagttaaagaggcgattgg
gaacaagttagcatcggcgataaacttacaaaagcaggcggaacatcaggcacaaattattcagatggcgcatcatatgataaaat
tcataaagaacatccggattggaaactgtatggctcagaaacagcatcatcagttaatagccgtggcatttattcagttacaggcaat
caagaagcaacaagcgatcaacaactgacagcgtatgataatagcagagttaattggggagcactggcatcacaagcatggtatg
atgttatccagagagattttgtcgcaggcgaatatgtttggacaggctttgattatatcggcgaaccgacaccgtggaatggcacaga
tccgggagcaaaaggcacatggccgtcaccgaaaaacagctactttggcattatcgatacagcaggctttccgaaagattcatatta
ttttttatcagagccagtggaatgaagaagtcaatacactgcacgttcttccggcatggaatgaagatgtcgtcaaaaaaaactcaga
tggcacagttccggttgttgtttattcagatgcgaaagaagtcgaactgtttttacaccggcaaatggcggagaaaaaaaaagcctg
ggaaaaaaaacatttaaaacagaaacaacaaaagctggctatagctatcaagttctggaaaacggcaaaaaaaaacataaagat
ctgtatatggaatggcaagttccgtatgaagcaggcacacttgaagcagttgcgaaagatgcaaaaggcaacgtcattaaagatac
agaaggcagaagcgtcgttaaaacaacaggcgaagaagcaaaactgtcagcaaaaacggatcgcaatagcattcaagcagatg
gcaaagatctgtcatatattacagtcgatgtcacagataaagatggcaatattgttccggatgcagcaaatagagtcacatttgatgt
ccaaggcgcaggaaaactggttggcgttgataatggctcatcaccggatcatgatagctataaagcggataaccgcaaagcatttc
aggcaaagttctggcaattgttcagtcaacagaaaaagcaggcgaaattacagttacagcaaaagcagatggcctggaatcaagc
acagtcaaaatcacaacaacaccggttaaagaagaaccgagcgaaagatatgtcgaaagctataaatacagcaaaagctattatg
tgaaaacaggcacaaaaccgcaactgccgaaaaaaattgaagcgcagtatagcgatcgcacaaaagaggatgttgcggtcaaat
gggatgaaatctcagatgaacaaattagcaaaacaggcagctttacagttgaaggcacagttggcaaaagagatatcacagtcaa
cattaacatgatcgatgatgttgcagcactgctgaattattcaggcgcaacacaaaaaggcgttaaaccgcaacttccggatgttaga
ccggcagttctgcctgatggcacagtcctggcagcatcatttccggttcagtgggatgaaaaagatgcggatacatttcagaaaccg
gatgaaattgttacagttaacggcagcgcagatatctttggcaaaacaattccggttacagcaagcattagagtgcagaaagaaga
tatcaaaattggcagcagcgttacaaatgttgcaaaactgagccaaaatattcaaggcagcgatacactggaagcaatcaaagatg
gcaaaacagaaatgagcctgaataatgatggcggaccgaatgaatcagcatggtcaaattgggatgcatcacagaaaggcacaa
aagaagccgaactgacatttacatttgatacacagcaacgcattggcgaaattgtcattcattttgcgaaagataacaactcaatcag
atttccggatgctggcacaacagaaatctaa

La SEQ ID NO: 11 es la secuencia de nucleótidos que codifica la SEQ ID NO: 2 que incluye la secuencia que codifica el péptido señal.

gtggatcagtttgctgtttgctttagcgttaatctttacgatggcgttcggcagcacatccagcgcgcaggcggcagggaagcaggcg
tttcagttccggcactggcacaacaagcagttagaacagaaagccaaacacaaatgtcatcagatccggaactggtctatgtgaata
actatagcagcacagcacaaagaagccagaactttaacagcaactggaaattctacttcggagatgcgggaaatgcacaaggcgc
aacatttgatgatagcaaatgggaacaagtttcactgccgcatgattattcaatcagccaagaatatagcaaatcaatggaagcaga
atcaggctatcttggcggaggcacaggctggtatcgcaaaaattttacactgagcagcgatacacaaggcaaaagagtccgcattg
attttgatggcgtctatatgaatgcaacagtttgggttaatggccatgaagttggcacacatccgtatggctatacaagctttagctttg
atatcacagattatgtgaaatatgatggcgaaaacacaattgcagtcaaagtcgtcaataatacaccgtcaagcagatggtattcag
gctcaggcatttatagagatgtcgatctgacaatcacagatgatgttcatgttgatctgaacggcacaaaagttacaacaccgaacct
ggaaacagaaaaaggcagcacagtcaatacagatgttacagcaacagttgcgaatgattcagatgcagcaaaatcagttgcagtt
cgccatacagtttttccgaaagatggcagcgcagatcaatcaattggcacagtcacaacaaatgcacaatcaattgcagcaggcgc
aacagcagaaattcaagcaacggttccggtttcaaatcctgaactgtggtcagttgaaaatccgtcactgtatacagtcagaacaga
agttctggtcgacggccaagtcacagatacatatgatacagaatatggctttcgctattttaactttgatagcaacacaggcttttcact
gaatggcgaaaatatgaaactgaaaggcgtctgcatgcatcatgatcaaggctcacttggcgcagcagcatacgactcagcaattg
atcgccaggtcaaaatcctgaaagaaatgggctgcaatagcattagagtcacacataatccggcagcacaagatctgattgatgcg
tgcaatgaacaaggcattctggttgttgaagaagcgtttgatacttggacaagaccgaaaaatggcaacagcaacgattatagcgtc
tggtttaatcagacagttgcgagcgataatgaaattctgggagcgacaaatggcgaaacatgggcacaatttgatctggaaagcat
gatctcacgcgattataatgcaccgtcagtcattatgtggtcactgggcaatgaagttatggaaggcattagcggaggcacagatgc

agaatatgaagcgacagcgacgaaactgattaactgggcgtatgatgcggataatacacgtccgatgacaattggcgataacaaa
ctgaaagcgaactggcagatctcaaaaacatttgcgagactgctgacagaaaaaggcggaacagtgggctttaattatgcagatg
gcagagttctggattcatatcatagcagcaatagcaattggctgctgtatggctcagaaacagcatcagcgattaatagccgtggcat
ctattatagaacaacaggcggaggccaaacatcagataaacagctgacaagctatgataattcaaatgttggctggggagcaaca
gcatcaaatgcatggtatacagttctgacaagagattttgcggcaggcgaatatgtttggacaggctttgattatctgggcgaaccga
caccgtggaatggcacaggctcaggcgcagttggctcatggccgtcaccgaaaaattcttattttggcattatcgatacagcaggctt
cgcaaaagatagctattatttttatcagagccagtggaatgatgatgttacaacactgcatgttcttccggcatggaataataatgtcg
tcagcaaagattcatcaggcaatgttccggttgttgtttattcagatgcggcatcagtcgaactgttttttcaagcaaaaggcagcgat
acaaaaacaagcctgggcaaaaaaacatttacacagaaaacaacagacgcaggctatacatatcagatctatgaaggctcagata
aaaacagcacaacagacaaaaacctgtatctgacatggaatgttccgtatgcagatggaacagtttcagcagttgcgtataatagca
acggccagaaaattacagatacagttggccagtcctcagttacaacaacaggcaaagcgtcaaaactgaaagcatcagcggatca
taaaaaaattgcagcggatggcgaatcactgtcatatatcacagtcgatgtcacagatgcgaatggcaatattgttccggatgcaga
aaatcgcgtcaaatttacagttgaaggcgatggcgaactggttggcgttgataatggctcatcaccggatcatgattcatatcaagcg
gataaccgcaaagcattttcaggcaaagttctggcaattgtgaaaagcacaaaagaagctggcacaattacagttacagcatcagc
agatggcctggattcagcatcagtcaaaatcacaacaacagcagtcgataatggcagcacagaaaaacaaatcgatagctttaaa
atgagccgcacatattatgttaaagttggcagcacaccggaactgccggaaaaaattgtcacacgctatacagatggcacatcaga
agaactgcctgttacttgggatgcaattacagaagatcaaattgcagcagcaggctcatttcaagttaaaggcacagtcaaaggcg
gatattcagttgcagtcaacgtcaacatgattgatgaagttggcggactgctgaattattcaacaaatacagcagttggcgttgcacc
ggttctgccgacatcaagaccggcagttctgcaagatggcacagttatggatgttacatttccggtcacatgggaagataaagcagc
aagcgcatatgataaagcaggcacagtgacagtcaatggcacagcaaatgttctgggcaaagaaattgcagttacagcgagcgtt
agagttcaggaagaaacaatcacaattggagattcagtttcagcggatgcactgaatctgacacaaagcgttccggcagataaaca
aagcgatacactgaacgcaattaaagatggctcaacaacaattagctcaaatacaagcggaggcgcaaatccgacagtttggagc
aactatgactatagccaggatggcaatacgacagcggatatcatttttgaatatgcgacagaacaaagactgggccaaatcgttaca
cattttgcgagagatagctggtcaatgagatatcctgatgcaggcgctacagaaatttatgtctaa

La SEQ ID NO: 12 es una beta-galactosidasa de Ruminococcus/Blautia hansenii DSM 20583.

myffgrsaimmltvktrkeifmrkqrlarigaatlaavltvqgmgfsstvyakeepvrvkadsqtqmssepeqvavkdygsnsar
tqnfdsdwkfnlgdvsnaqtptfddskwrtislphdysieqeysqsleaesgylpggvgwyrknftlgeeakgkriridfdgvymna
tvyvngkevgthpygytpfsfditdyisydkentiavkvdhqtpssrwysgsgiyrsvnlttndvhvdlngikvesnnlekeagktv
ntdvkttvvngskeaknitithtvfkkgekpdkaigtfttеaqeigagkkteisatvpvknpelwsvenpalytirtevkagdklldsy
dteygfhylnfdtetgfqlngknvklkgvcmhhdqgalgavanrraierqveilqemgcnsirvthnpaskdlievcnekgilvieev
fdgwhraknqnsndysvwfekaieednailgkeadmtwaeydlkaimkrdqnapsiiewslgneiqegaggsgyaeradklik
wakeadatktltigsnavkrgdweqvsigdkltkaggtsgtnysdgasydkihkehpdwklygsetassvnsrgiysvtgnqeats
dqqltaydnsrvnwgalasqawydviqrdfvageyvwtgfdyigeptpwngtdpgakgtwpspknsyfgiidtagfpkdsyyfy
qsqwneevntlhvlpawnedvvkknsdgtvpvvvysdakevelfftpanggekkslgkktfktettkagysyqvlengkkkhkdl
ymewqvpyeagtleavakdakgnvikdtegrsvvkttgeeaklsaktdrnsiqadgkdlsyitvdvtdkdgnivpdaanrvtfdv
qgagklvgvdngsspdhdsykadnrkafsgkvlaivqstekageitvtakadglesstvkitttpvkeepseryvesykysksyyv
ktgtkpqlpkkieaqysdrtkedvavkwdeisdeqisktgsftvegtvgkrditvninmiddvaallnysgatqkgvkpqlpdvrpa
vlpdgtvlaasfpvqwdekdadtfqkpdeivtvngsadifgktipvtasirvqkedikigssvtnvaklsqniqgsdtleaikdgkte
mslnndggpnesawsnwdasqkgtkeaeltftfdtqqrigeivihfakdnnsirfpdagtteifvsetgkdgtwekvevkehigee
kdrvkayryeiapvtatyvkvkvvnanatdtgnrkpctaitevelkkaegsfkvnetaeleevkvgervlpnaayaldsysvpetd
aavtaktkdnasltilpkhenvvrmilesedhkatknfavrmgeeetvlpdddsrdypvekitatagseykpgtanegpvkyvldg
kaethwhtnwsvsgegskpehrtvtlqlgndeeeapmidalrymprsngangrvteyeiqysldgdkwqtaatgeidkkqtgw
milgfeepvqakyvrfigthttsdqgndkhmavselrarvateapapsekytitanvndktmgavtldsetgeyekgtkatltavp
kegfafvnwtidgqevskenpyihtvetdatitanferievenegwvqtengweyyengqkvvgwkevsgkwyyfeenglmqt
gwvfvnnhwyymdqwgamcigwvavdghwyymdqwgamctgwvsvnghwyhmdqwgamqtgwalvdsnwyyln
tdgsmaigwvavnghwyymdqwgamqtgwalvdsnwyylntdgsmaigwvavnghwyymdqwgamqtgwvlvgsd
wyylntdgsmassqwidgyyvdasgkmk

La SEQ ID NO: 13 es una glucosidasa de Ruminococcus lactaris ATCC 29176.

mkkkkrctrvgagalaavlavtaagvsvpalaqqavrtesqtqmssdpelvyvnnysstaqrsqnfnsnwkfyfgdagnaqgat
fddskweqvslphdyslsqeysksmeaesgylgggtgwyrknftlssdtqgkrvridfdgvymnatvwvnghevgthpygytsf
sfditdyvkydgentiavkvvnntpssrwysgsgiyrdvdltitddvhvdlngtkvttpnletekgstvntdvtatvandsdaaksva
vrhtvfpkdgsadqsigtvttnaqslaagataelqatvpvsnpelwsvenpslytvrtevlvdgqvtdtydteygfryfnfdsntgfsl
ngenmklkgvcmhhdqgslgaaaydsaidrqvkllkemgcnsirvthnpaaqdlidacneqgilvveeafdtwtrpkngnsndy
svwfnqtvasdneilgatngetwaqfdlesmisrdynapsvimwslgnevmegisggtdaeyeatatklinwaydadntrpmtl
gdnklkanwqisktfarlltekggtvgfnyadgrvldsyhssnsnwllygsetasalnsrgiyyrttgggqtsdkqltsydnsnvgwg

atasnawytvltrdfaageyvwtgfdylgeptpwngtgsgavgswpspknsyfgiidtagfakdsyyfyqsqwnddvttlhvlpa
wnnnvvskdssgnvpvvvysdaasvelffqakgsdtktslgkktftqkttdagytyqiyegsdknsttdknlyltwnvpyadgtvs
avaynsngqkitdtvgqssvtttgkasklkasadhkkiaadgeslsyitvdvtdangnivpdaenrvkftvegdgelvgvdngssp
dhdsyqadnrkafsgkvlaivkstkeagtitvtasadgldsasvkitttavdngstekqidsfkmsrtyyvkvgstpelpekivtrytd
gtseelpvtwdaitedqiaaagsfqvkgtvkggysvavnvnmidevggllnystntavgvapvlptsrpavlqdgtvmdvtfpvt
wedkaasaydkagtvtvngtanvlgkeiavtasvrvqeetitigdsvsadalnltqsvpadkqsdtlnaikdgsttissntsgganpt
vwsnydysqdgnttadiifeyateqrlgqivthfardswsmrypdagateiyvspdgtnwakldttetigtesgnvkpytydfapv
gatfvkfhltnstqatgttakactgiteielkvatgsrttnttaelqtltvngkevpqtaldskvyttpailaeieatakdnasvtvlpaynd
viriivesedhqtrntyevrlneaeqttpdsdsrdypvskltasagseqsttgvegpasnakdgdestlwhtrwsapaatsdqlwfty
eleeetvldalrylprqgtadgqnngrvneyrvevstdgstwttvstgnwedsqdwklaeftepvaakyvrltgvhtygssaanvd
kymsaaeirlrmaesktdiadaangvtvtapdslevakadaenpvmfdlsdivvkagdttlrygvdyvisyenntdfgtaklvikgi
dgytgtleheftitqkakvmtgitwntkpekviytegetldvtglvinvvyddd steavayseanadeftfspaldtklaatdktvtvt
ykgasliyditvnpkkvdptdpdqpdkpdtpdngndngndnngngnnngtddgkkdpgqsgvtdnknqgnnsnngtaagn
kanaaaktgdtanmllpmiaamlagtavvgtisirrrrr

SEQ ID NO: 14 es la secuencia de nucleótidos que codifica la SEQ ID NO: 12 sin la secuencia señal.

aaagcagatagccaaacacaaatgtcatcagaaccggaacaagttgcggttaaagattatggctcaaatagcgcacgcacacaga
attttgatagcgattggaaatttaacctgggagatgttagcaatgcacagacaccgacatttgatgattcaaaatggcgcacactgtc
actgccgcatgattatagcatcgaacaggaatattcacaatcactggaagcagaatcaggctatcttccgggaggcgttggctggta
tcgcaaaaattttacactgggcgaagaagcgaaaggcaaacgcattcgcattgattttgatggcgtctatatgaatgcaacagtctat
gtgaatggcaaagaagttggcacacatccgtatggctatacaccgtttagctttgatatcacagattatatcagctatgataaagaaa
acacaattgcggtcaaagtcgatcatcaaacaccgtcatcaagatggtattcaggcagcggcatttatagatcagtcaacctgacaa
caacaaatgatgtccatgtcgatctgaatggcattaaagtcgaaagcaacaacctggaaaaagaagcaggcaaaacagtcaacac
agatgtgaaaacaacagttgtgaacggctcaaaagaagcgaaaaacatcacaattacacatacagtctttaaaaaaggcgaaaaa
ccggataaagcgatcggcacatttacaacagaagcgcaagaaattggcgcaggcaaaaaaacagaaatcagcgcaacagtcccg
gttaaaaatccggaactgtggtcagttgaaaatccggcactgtatacaattcgcacagaagttaaagcaggcgataaactgctggat
agctatgatacagaatatggctttcattatctgaactttgatacagaaacaggctttcagctgaatggcaaaaacgttaaactgaaag
gcgtttgcatgcatcatgatcaaggcgcacttggcgcagttgcaaatagaagagcaattgaacgccaagtcgaaattctgcaagaa
atgggctgcaatagcattagagtcacacataatccggcaagcaaagatctgattgaagtctgcaacgaaaaaggcattctggtcatt
gaagaagtttttgacggctggcatagagcaaaaaatggcaacagcaacgattatagcgtctggtttgaaaaagcgatcgaagaag
ataacgcgattctgggaaaagaagcggatatgacttgggcagaatatgatctgaaagcgattatgaaacgcgatcaaaatgcacc
gagcattattgaatggtcactgggcaatgaaattcaagaaggcgcaggcggatcaggctatgcagaaagagcggataaactgatc
aaatgggcgaaagaagcagacgcaacaaaaacactgacaattggcagcaatgcagttaaaagaggcgattgggaacaagttag
catcggcgataaacttacaaaagcaggcggaacatcaggcacaaattattcagatggcgcatcatatgataaaattcataaagaac
atccggattggaaactgtatggctcagaaacagcatcatcagttaatagccgtggcatttattcagttacaggcaatcaagaagcaa
caagcgatcaacaactgacagcgtatgataatagcagagttaattggggagcactggcatcacaagcatggtatgatgttatccag
agagattttgtcgcaggcgaatatgtttggacaggctttgattatatcggcgaaccgacaccgtggaatggcacagatccgggagca
aaaggcacatggccgtcaccgaaaaacagctactttggcattatcgatacagcaggctttccgaaagattcatattatttttatcagag
ccagtggaatgaagaagtcaatacactgcacgttcttccggcatggaatgaagatgtcgtcaaaaaaaactcagatggcacagttc
cggttgttgtttattcagatgcgaaagaagtcgaactgtttttttacaccggcaaatggcggagaaaaaaaaagcctgggaaaaaaa
acatttaaaacagaaacaacaaaagctggctatagctatcaagttctggaaaacggcaaaaaaaaacataaagatctgtatatgga
atggcaagttccgtatgaagcaggcacacttgaagcagttgcgaaagatgcaaaaggcaacgtcattaaagatacagaaggcag
aagcgtcgttaaaacaacaggcgaagaagcaaaactgtcagcaaaaacggatcgcaatagcattcaagcagatggcaaagatct

gtcatatattacagtcgatgtcacagataaagatggcaatattgttccggatgcagcaaatagagtcacatttgatgtccaaggcgca
ggaaaactggttggcgttgataatggctcatcaccggatcatgatagctataaagcggataaccgcaaagcattttcaggcaaagtt
ctggcaattgttcagtcaacagaaaaagcaggcgaaattacagttacagcaaaagcagatggcctggaatcaagcacagtcaaaa
tcacaacaacaccggttaaagaagaaccgagcgaaagatatgtcgaaagctataaatacagcaaaagctattatgtgaaaacagg
cacaaaaccgcaactgccgaaaaaaattgaagcgcagtatagcgatcgcacaaaagaggatgttgcggtcaaatgggatgaaat
ctcagatgaacaaattagcaaaacaggcagctttacagttgaaggcacagttggcaaaagagatatcacagtcaacattaacatga
tcgatgatgttgcagcactgctgaattattcaggcgcaacacaaaaaggcgttaaaccgcaacttccggatgttagaccggcagttct
gcctgatggcacagtcctggcagcatcatttccggttcagtgggatgaaaaagatgcggatacatttcagaaaccggatgaaattgt
tacagttaacggcagcgcagatatctttggcaaaacaattccggttacagcaagcattagagtgcagaaagaagatatcaaaattg
gcagcagcgttacaaatgttgcaaaactgagccaaaatattcaaggcagcgatacactggaagcaatcaaagatggcaaaacaga
aatgagcctgaataatgatggcggaccgaatgaatcagcatggtcaaattgggatgcatcacagaaaggcacaaaagaagccga
actgacatttacatttgatacacagcaacgcattggcgaaattgtcattcattttgcgaaagataacaactcaatcagatttccggatg
ctggcacaacagaaatctttgtttcagaaacaggcaaagatggcacatgggaaaaagttgaagtcaaagagcatattggcgaaga
aaaagatcgcgtcaaagcatatcgctatgaaattgcaccggttacagcgacatatgttaaagttaaagtcgtcaatgcgaacgcgac
agatacaggcaatagaaaaccgtgcacagcaattacagaagtcgaactgaaaaaagcagaaggcagctttaaagtcaacgaaac
agcagaactggaagaagttaaagttggcgaacgtgttctgccgaatgcagcatatgcactggattcatattcagttccggaaacgga
tgcagcagttacagcaaaaacaaaagataatgcgagcctgacaatcctgccgaaacatgaaaatgtcgtcagaatgattctggaaa
gcgaagaccataaagcgacgaaaaactttgcagttagaatgggcgaagaagaaacagttctgccggatgatgattcaagagatta
tccggtcgaaaaaatcacagcaacagcaggctcagaatataaaccgggaacagcaaatgaaggaccggttaaatatgttctggat
ggcaaagcagaaacacattggcatacaaattggtcagtttcaggcgaaggctcaaaaccggaacatagaacagttacactgcaac
tgggcaatgatgaagaagaagcaccgatgattgacgcactgagatatatgccgagatcaaatggcgcaaatggcagagttacgga
atatgaaattcagtatagcctggatggcgataaatggcaaacagcagcaacaggcgaaatcgataaaaaacaaacaggctggat
gatcctgggctttgaagaaccggttcaagcaaaatatgtccgctttattggcacacatacaacatcagatcagggcaatgataaacat
atggcagtttcagaactgagagcaagagttgcaacagaagcaccggcaccgtcagagaagtatacaattacagcgaacgtcaacg
ataaaacaatgggagcagttacacttgatagcgaaacaggcgaatatgaaaaaggcacgaaagcaacactgacagcagttccga
aagaaggctttgcatttgtcaactggacaattgatggccaagaagtctcaaaagaaaacccgtatatccatacagttgaaacggatg
cgacaatcacagcgaattttgaacgcattgaagtcgaaaatgaaggctgggttcaaacagaaaatggctgggaatattatgagaat
ggccaaaaagttgtcggctggaaagaagtttcaggcaaatggtactactttgaagaaaatggcctgatgcaaacaggatgggtcttt
gttaacaaccattggtattatatggatcagtgggggcaatgtgcattggctgggttgcagttgatggccattggtactacatggacc
aatggggtgctatgtgtacaggctgggttagcgtcaatggacattggtatcatatggaccaatggggagccatgcaaacaggctgg
gcactggttgattcaaattggtattacctgaatacggatggctcaatggcaattggatgggtcgcagtgaacggccactggtattaca
tggatcaatggggagctatgcagacgggatgggctcttgttgatagcaactggtattatcttaacacagatggcagcatggcaatcg
gctgggtggcggttaatggacactggtactatatggatcaatggggtgcaatgcagacaggctgggttctggtcggcagcgattggt
actatttaaacacggatggatctatggcatcaagccaatggattgatggctattatgttgatgcaagcggcaagatgaag

SEQ ID NO: 15 es la secuencia de nucleótidos que codifica la SEQ ID NO: 13 sin la secuencia señal.

gcaggcgtttcagttccggcactggcacaacaagcagttagaacagaaagccaaacacaaatgtcatcagatccggaactggtcta
tgtgaataactatagcagcacagcacaaagaagccagaactttaacagcaactggaaattctacttcggagatgcgggaaatgcac

aaggcgcaacatttgatgatagcaaatgggaacaagtttcactgccgcatgattattcaatcagccaagaatatagcaaatcaatgg
aagcagaatcaggctatcttggcggaggcacaggctggtatcgcaaaaattttacactgagcagcgatacacaaggcaaaagagt
ccgcattgattttgatggcgtctatatgaatgcaacagtttgggttaatggccatgaagttggcacacatccgtatggctatacaagctt
tagctttgatatcacagattatgtgaaatatgatggcgaaaacacaattgcagtcaaagtcgtcaataatacaccgtcaagcagatg
gtattcaggctcaggcatttatagagatgtcgatctgacaatcacagatgatgttcatgttgatctgaacggcacaaaagttacaaca
ccgaacctggaaacagaaaaaggcagcacagtcaatacagatgttacagcaacagttgcgaatgattcagatgcagcaaaatcag
ttgcagttcgccatacagtttttccgaaagatggcagcgcagatcaatcaattggcacagtcacaacaaatgcacaatcaattgcagc
aggcgcaacagcagaaattcaagcaacggttccggtttcaaatcctgaactgtggtcagttgaaaatccgtcactgtatacagtcag
aacagaagttctggtcgacggccaagtcacagatacatatgatacagaatatggctttcgctattttaactttgatagcaacacaggc
ttttcactgaatggcgaaaatatgaaactgaaaggcgtctgcatgcatcatgatcaaggctcacttggcgcagcagcatacgactca
gcaattgatcgccaggtcaaaatcctgaaagaaatgggctgcaatagcattagagtcacacataatccggcagcacaagatctgat
tgatgcgtgcaatgaacaaggcattctggttgttgaagaagcgtttgatacttggacaagaccgaaaaatggcaacagcaacgatt
atagcgtctggtttaatcagacagttgcgagcgataatgaaattctgggagcgacaaatggcgaaacatgggcacaatttgatctgg
aaagcatgatctcacgcgattataatgcaccgtcagtcattatgtggtcactgggcaatgaagttatggaaggcattagcggaggca
cagatgcagaatatgaagcgacagcgacgaaactgattaactgggcgtatgatgcggataatacacgtccgatgacaattggcgat
aacaaactgaaagcgaactggcagatctcaaaaacatttgcgagactgctgacagaaaaaggcggaacagtgggctttaattatg
cagatggcagagttctggattcatatcatagcagcaatagcaattggctgctgtatggctcagaaacagcatcagcgattaatagcc
gtggcatctattatagaacaacaggcggaggccaaacatcagataaacagctgacaagctatgataattcaaatgttggctgggga
gcaacagcatcaaatgcatggtatacagttctgacaagagattttgcggcaggcgaatatgtttggacaggctttgattatctgggcg
aaccgacaccgtggaatggcacaggctcaggcgcagttggctcatggccgtcaccgaaaaattcttattttggcattatcgatacagc
aggcttcgcaaaagatagctattatttttatcagagccagtggaatgatgatgttacaacactgcatgttcttccggcatggaataata
atgtcgtcagcaaagattcatcaggcaatgttccggttgttgtttattcagatgcggcatcagtcgaactgttttttcaagcaaaaggca
gcgatacaaaaacaagcctgggcaaaaaaacatttacacagaaaacaacagacgcaggctatacatatcagatctatgaaggctc
agataaaaacagcacaacagacaaaaacctgtatctgacatggaatgttccgtatgcagatggaacagtttcagcagttgcgtata
atagcaacggccagaaaattacagatacagttggccagtcctcagttacaacaacaggcaaagcgtcaaaactgaaagcatcagc
ggatcataaaaaaattgcagcggatggcgaatcactgtcatatatcacagtcgatgtcacagatgcgaatggcaatattgttccgga
tgcagaaaatcgcgtcaaatttacagttgaaggcgatggcgaactggttggcgttgataatggctcatcaccggatcatgattcatat
caagcggataaccgcaaagcattttcaggcaaagttctggcaattgtgaaaagcacaaaagaagctggcacaattacagttacagc
atcagcagatggcctggattcagcatcagtcaaaatcacaacaacagcagtcgataatggcagcacagaaaaacaaatcgatagc
tttaaaatgagccgcacatattatgttaaagttggcagcacaccggaactgccggaaaaaattgtcacacgctatacagatggcaca
tcagaagaactgcctgttacttgggatgcaattacagaagatcaaattgcagcagcaggctcatttcaagttaaaggcacagtcaaa
ggcggatattcagttgcagtcaacgtcaacatgattgatgaagttggcggactgctgaattattcaacaaatacagcagttggcgttg
caccggttctgccgacatcaagaccggcagttctgcaagatggcacagttatggatgttacatttccggtcacatgggaagataaag
cagcaagcgcatatgataaagcaggcacagtgacagtcaatggcacagcaaatgttctgggcaaagaaattgcagttacagcgag
cgttagagttcaggaagaaacaatcacaattggagattcagtttcagcggatgcactgaatctgacacaaagcgttccggcagata
aacaaagcgatacactgaacgcaattaaagatggctcaacaacaattagctcaaatacaagcggaggcgcaaatccgacagtttg
gagcaactatgactatagccaggatggcaatacgacagcggatatcatttttgaatatgcgacagaacaaagactgggccaaatcg
ttacacattttgcgagagatagctggtcaatgagatatcctgatgcaggcgctacagaaatttatgtctcaccggatggcacaaattg
ggcaaaactggatacaacagaaacaattggcacagaaagcggcaatgttaaaccgtatacatatgattttgcaccggttggcgcaa
catttgttaaatttcatctgacaaacagcacacaagcaacaggcacaacagcaaaagcatgcacaggcattacagaaattgaactg

aaagttgcaacaggctcacgcacaacaaatacaacagcagaactgcaaacactgacagttaatggcaaagaagttccgcaaaca
gcactggatagcaaagtttatacaacaccggcaattctggcagaaattgaagcaacagcgaaagataatgcaagcgttacagttctt
ccggcatataatgatgtcattcgcattattgtcgaaagcgaagatcatcaaacacgcaatacatatgaagtcagactgaatgaagcg
gaacaaacaacaccggattcagattcaagagattatccggttagcaaactgacagcatcagcaggctcagaacaatcaacaacag
gcgttgaaggaccggcatcaaatgcaaaagacggtgatgaatcaacactgtggcatacaagatggtcagcaccggcagcaacatc
agatcaactgtggtttacatatgaactggaagaagaaacggtactggacgcactgagatatctgccgagacaaggcacagcagat
ggccaaaataatggcagagttaatgaatatcgcgtcgaagttagcacagatggcagcacatggacaacagtttcaacaggcaattg
ggaagatagccaagattggaaactggcagaatttacagaaccggttgcagcaaaatatgtcagactgacaggcgttcatacatatg
gctcatcagcagcaaacgtcgataaatacatgagcgcagcagaaattagactgagaatggcagaaagcaaaacggatattgcag
atgcagcaaatggcgttacagttacagcaccggattcaattgaagttgcaaaagcagatgcagaaaacccggttatgtttgatctga
gcgatattgttgtcaaagcaggcgatacaacactgagatatggcgttgattatgtcattagctatgaaaacaacacagattttggcac
agcgaaactggtcattaaaggcattgatggctatacaggcacactggaacatgaattcacaatcacgcagaaagccaaagtcatga
caggcatcacatggaatacaaaaccggaaaaagtcatttatacggaaggtgaaacgctggatgttacaggcctggttattaatgtcg
tctatgatgatgatagcacagaagcagttgcatatagcgaagcaaatgcggatgaatttacattttcaccggcactggatacaaaac
tggcagcgacagataaaacagtcacagttacatataaaggcgcaagcctgatttatgatattacagtcaacccgaaaaaagtcgat
ccgacagatccggatcagcctgataaaccggatacaccggataatggcaatgataacggcaacgataataatggcaacggcaata
acaacggcacagatgatggcaaaaaagatccgggacaatcaggcgttacagataacaaaaatcagggcaataacagcaataatg
gaacagcagcaggcaataaagcaaatgcagcagcaaaaacaggcgatacagcaaatatgctgctgccgatgattgcagcaatgc
tggcaggcacagcagttgttggcacaatttcaattcgcagacgcagacgc

La SEQ ID NO: 16 es la secuencia de nucleótidos que codifica la SEQ ID NO: 1.

aaagcagatagccaaacacaaatgtcatcagaaccggaacaagttgcggttaaagattatggctcaaatagcgcacgcacacagaattttgatagcgattggaaat
ttaacctgggagatgttagcaatgcacagacaccgacatttgatgattcaaaatggcgcacactgtcactgccgcatgattatagcatcgaacaggaatattcacaat
cactggaagcagaatcaggctatcttccgggaggcgttggctggtatcgcaaaaattttacactgggcgaagaagcgaaaggcaaacgcattcgcattgattttgat
ggcgtctatatgaatgcaacagtctatgtgaatggcaaagaagttggcacacatccgtatggctatacaccgtttagctttgatatcacagattatatcagctatgataa
agaaaacacaattgcggtcaaagtcgatcatcaaacaccgtcatcaagatggtattcaggcagcggcatttatagatcagtcaacctgacaacaacaaatgatgtcc
atgtcgatctgaatggcattaaagtcgaaagcaacaacctggaaaaagaagcaggcaaaacagtcaacacagatgtgaaaacaacagttgtgaacggctcaaaa
gaagcgaaaaacatcacaattacacatacagtctttaaaaaaggcgaaaaaccggataaagcgatcggcacatttacaacagaagcgcaagaaattggcgcagg
caaaaaaacagaaatcagcgcaacagtcccggttaaaaatccggaactgtggtcagttgaaaatccggcactgtatacaattcgcacagaagttaaagcaggcga
taaactgctggatagctatgatacagaatatggctttcattatctgaactttgatacagaaacaggctttcagctgaatggcaaaaacgttaaactgaaaggcgtttgca
tgcatcatgatcaaggcgcacttggcgcagttgcaaatagaagagcaattgaacgccaagtcgaaattctgcaagaaatgggctgcaatagcattagagtcacaca
taatccggcaagcaaagatctgattgaagtctgcaacgaaaaaggcattctggtcattgaagaagtttttgacggctggcatagagcaaaaaatggcaacagcaac
gattatagcgtctggtttgaaaaagcgatcgaagaagataacgcgattctgggaaaagaagcggatatgacttgggcagaatatgatctgaaagcgattatgaaac
gcgatcaaaatgcaccgagcattattgaatggtcactgggcaatgaaattcaagaaggcgcaggcggatcaggctatgcagaaagagcggataaactgatcaaa
tgggcgaaagaagcagacgcaacaaaaacactgacaattggcagcaatgcagttaaaagaggcgattgggaacaagttagcatcggcgataaacttacaaaag
caggcggaacatcaggcacaaattattcagatggcgcatcatatgataaaattcataaagaacatccggattggaaactgtatggctcagaaacagcatcatcagtt
aatagccgtggcatttattcagttacaggcaatcaagaagcaacaagcgatcaacaactgacagcgtatgataatagcagagttaattggggagcactggcatcac
aagcatggtatgatgttatccagagagattttgtcgcaggcgaatatgtttggacaggctttgattatatcggcgaaccgacaccgtggaatggcacagatccggga

gcaaaaggcacatggccgtcaccgaaaaacagctactttggcattatcgatacagcaggctttccgaaagattcatattatttttatcagagccagtggaatgaagaa
gtcaatacactgcacgttcttccggcatggaatgaagatgtcgtcaaaaaaaactcagatggcacagttccggttgttgtttattcagatgcgaaagaagtcgaactgt
ttttacaccggcaaatggcggagaaaaaaaaagcctgggaaaaaaaacatttaaaacagaaacaacaaaagctggctatagctatcaagttctggaaaacggca
aaaaaaaacataaagatctgtatatggaatggcaagttccgtatgaagcaggcacacttgaagcagttgcgaaagatgcaaaaggcaacgtcattaaagatacag
aaggcagaagcgtcgttaaaacaacaggcgaagaagcaaaactgtcagcaaaaacggatcgcaatagcattcaagcagatggcaaagatctgtcatatattaca
gtcgatgtcacagataaagatggcaatattgttccggatgcagcaaatagagtcacatttgatgtccaaggcgcaggaaaactggttggcgttgataatggctcatc
accggatcatgatagctataaagcggataaccgcaaagcatttttcaggcaaagttctggcaattgttcagtcaacagaaaaagcaggcgaaattacagttacagca
aaagcagatggcctggaatcaagcacagtcaaaatcacaacaacaccggttaaagaagaaccgagcgaaagatatgtcgaaagctataaatacagcaaaagcta
ttatgtgaaaacaggcacaaaaccgcaactgccgaaaaaaattgaagcgcagtatagcgatcgcacaaaagaggatgttgcggtcaaatgggatgaaatctcag
atgaacaaattagcaaaacaggcagctttacagttgaaggcacagttggcaaaagagatatcacagtcaacattaacatgatcgatgatgttgcagcactgctgaat
tattcaggcgcaacacaaaaaggcgttaaaccgcaacttccggatgttagaccggcagttctgcctgatggcacagtcctggcagcatcatttccggttcagtggg
atgaaaaagatgcggatacatttcagaaaccggatgaaattgttacagttaacggcagcgcagatatctttggcaaaacaattccggttacagcaagcattagagtg
cagaaagaagatatcaaaattggcagcagcgttacaaatgttgcaaaactgagccaaaatattcaaggcagcgatacactggaagcaatcaaagatggcaaaaca
gaaatgagcctgaataatgatggcggaccgaatgaatcagcatggtcaaattgggatgcatcacagaaaggcacaaaagaagccgaactgacatttacatttgata
cacagcaacgcattggcgaaattgtcattcattttgcgaaagataacaactcaatcagatttccggatgctggcacaacagaaatc

La SEQ ID NO: 17 es la secuencia de nucleótidos que codifica la SEQ ID NO: 2.

gcaggcgtttcagttccggcactggcacaacaagcagttagaacagaaagccaaacacaaatgtcatcagatccggaactggtctatgtgaataactatagcagc
acagcacaaagaagccagaactttaacagcaactggaaattctacttcggagatgcgggaaatgcacaaggcgcaacatttgatgatagcaaatgggaacaagtt
tcactgccgcatgattattcaatcagccaagaatatagcaaatcaatggaagcagaatcaggctatcttggcggaggcacaggctggtatcgcaaaaatttttacact
gagcagcgatacacaaggcaaaagagtccgcattgattttgatggcgtctatatgaatgcaacagtttgggttaatggccatgaagttggcacacatccgtatggct
atacaagctttagctttgatatcacagattatgtgaaatatgatggcgaaaacacaattgcagtcaaagtcgtcaataatacaccgtcaagcagatggtattcaggctc
aggcatttatagagatgtcgatctgacaatcacagatgatgttcatgttgatctgaacggcacaaaagttacaacaccgaacctggaaacagaaaaaggcagcaca
gtcaatacagatgttacagcaacagttgcgaatgattcagatgcagcaaaatcagttgcagttcgccatacagtttttccgaaagatggcagcgcagatcaatcaatt
ggcacagtcacaacaaatgcacaatcaattgcagcaggcgcaacagcagaaattcaagcaacggttccggtttcaaatcctgaactgtggtcagttgaaaatccgt
cactgtatacagtcagaacagaagttctggtcgacggccaagtcacagatacatatgatacagaatatggctttcgctattttaactttgatagcaacacaggcttttca
ctgaatggcgaaaatatgaaactgaaaggcgtctgcatgcatcatgatcaaggctcacttggcgcagcagcatacgactcagcaattgatcgccaggtcaaaatcc
tgaaagaaatgggctgcaatagcattagagtcacacataatccggcagcacaagatctgattgatgcgtgcaatgaacaaggcattctggttgttgaagaagcgttt
gatacttggacaagaccgaaaaatggcaacagcaacgattatagcgtctggtttaatcagacagttgcgagcgataatgaaattctgggagcgacaaatggcgaa
acatgggcacaatttgatctggaaagcatgatctcacgcgattataatgcaccgtcagtcattatgtggtcactgggcaatgaagttatggaaggcattagcggagg
cacagatgcagaatatgaagcgacagcgacgaaactgattaactgggcgtatgatgcggataatacacgtccgatgacaattggcgataacaaactgaaagcga
actggcagatctcaaaaacatttgcgagactgctgacagaaaaaggcggaacagtgggctttaattatgcagatggcagagttctggattcatatcatagcagcaat
agcaattggctgctgtatggctcagaaacagcatcagcgattaatagccgtggcatctattatagaacaacaggcggaggccaaacatcagataaacagctgaca
agctatgataattcaaatgttggctggggagcaacagcatcaaatgcatggtatacagttctgacaagagattttgcggcaggcgaatatgtttggacaggctttgatt
atctgggcgaaccgacaccgtggaatggcacaggctcaggcgcagttggctcatggccgtcaccgaaaaattcttattttggcattatcgatacagcaggcttcgc
aaaagatagctattatttttatcagagccagtggaatgatgatgttacaacactgcatgttcttccggcatggaataataatgtcgtcagcaaagattcatcaggcaatg
ttccggttgttgtttattcagatgcggcatcagtcgaactgtttttttcaagcaaaaggcagcgatacaaaacaagcctgggcaaaaaaacatttacacagaaaacaa
cagacgcaggctatacatatcagatctatgaaggctcagataaaaacagcacaacagacaaaaacctgtatctgacatggaatgttccgtatgcagatggaacagt
ttcagcagttgcgtataatagcaacggccagaaaattacagatacagttggccagtcctcagttacaacaacaggcaaagcgtcaaaactgaaagcatcagcggat
cataaaaaaaattgcagcggatggcgaatcactgtcatatatcacagtcgatgtcacagatgcgaatggcaatattgttccggatgcagaaaatcgcgtcaaatttaca
gttgaaggcgatggcgaactggttggcgttgataatggctcatcaccggatcatgattcatatcaagcggataaccgcaaagcattttcaggcaaagttctggcaatt
gtgaaaagcacaaaagaagctggcacaattacagttacagcatcagcagatggcctggattcagcatcagtcaaaatcacaacaacagcagtcgataatggcag
cacagaaaaacaaatcgatagctttaaaatgagccgcacatattatgttaaagttggcagcacaccggaactgccggaaaaaattgtcacacgctatacagatggc
acatcagaagaactgcctgttacttgggatgcaattacagaagatcaaattgcagcagcaggctcatttcaagttaaaggcacagtcaaaggcggatattcagttgc
agtcaacgtcaacatgattgatgaagttggcggactgctgaattattcaacaaatacagcagttggcgttgcaccggttctgccgacatcaagaccggcagttctgc
aagatggcacagttatggatgttacatttccggtcacatgggaagataaagcagcaagcgcatatgataaagcaggcacagtgacagtcaatggcacagcaaatg
ttctgggcaaagaaattgcagttacagcgagcgttagagttcaggaagaaacaatcacaattggagattcagtttcagcggatgcactgaatctgacacaaagcgtt
ccggcagataaacaaagcgatacactgaacgcaattaaagatggctcaacaacaattagctcaaatacaagcggaggcgcaaatccgacagtttggagcaactat
gactatagccaggatggcaatacgacagcggatatcatttttgaatatgcgacagaacaaagactgggccaaatcgttacacattttgcgagagatagctggtcaat
gagatatcctgatgcaggcgctacagaaatttatgtc

Listado de secuencias

<110> Danisco A/S

<120> Polipéptidos que tienen actividad transgalactosilante

<130> 18028PCT00

<160> 17

<170> PatentIn versión 3.5

<210> 1
<211> 1125
<212> PRT
<213> Ruminococcus hansenii

<400> 1

```
Lys Ala Asp Ser Gln Thr Gln Met Ser Ser Glu Pro Glu Gln Val Ala
1               5                  10                  15

Val Lys Asp Tyr Gly Ser Asn Ser Ala Arg Thr Gln Asn Phe Asp Ser
            20                  25                  30

Asp Trp Lys Phe Asn Leu Gly Asp Val Ser Asn Ala Gln Thr Pro Thr
        35                  40                  45

Phe Asp Asp Ser Lys Trp Arg Thr Leu Ser Leu Pro His Asp Tyr Ser
    50                  55                  60

Ile Glu Gln Glu Tyr Ser Gln Ser Leu Glu Ala Glu Ser Gly Tyr Leu
65                  70                  75                  80

Pro Gly Gly Val Gly Trp Tyr Arg Lys Asn Phe Thr Leu Gly Glu Glu
                85                  90                  95

Ala Lys Gly Lys Arg Ile Arg Ile Asp Phe Asp Gly Val Tyr Met Asn
            100                 105                 110

Ala Thr Val Tyr Val Asn Gly Lys Glu Val Gly Thr His Pro Tyr Gly
        115                 120                 125

Tyr Thr Pro Phe Ser Phe Asp Ile Thr Asp Tyr Ile Ser Tyr Asp Lys
    130                 135                 140

Glu Asn Thr Ile Ala Val Lys Val Asp His Gln Thr Pro Ser Ser Arg
145                 150                 155                 160

Trp Tyr Ser Gly Ser Gly Ile Tyr Arg Ser Val Asn Leu Thr Thr Thr
```

```
                165                 170                 175

Asn Asp Val His Val Asp Leu Asn Gly Ile Lys Val Glu Ser Asn Asn
            180                 185                 190

Leu Glu Lys Glu Ala Gly Lys Thr Val Asn Thr Asp Val Lys Thr Thr
        195                 200                 205

Val Val Asn Gly Ser Lys Glu Ala Lys Asn Ile Thr Ile Thr His Thr
    210                 215                 220

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

Thr Glu Ala Gln Glu Ile Gly Ala Gly Lys Lys Thr Glu Ile Ser Ala
                245                 250                 255

Thr Val Pro Val Lys Asn Pro Glu Leu Trp Ser Val Glu Asn Pro Ala
            260                 265                 270

Leu Tyr Thr Ile Arg Thr Glu Val Lys Ala Gly Asp Lys Leu Leu Asp
        275                 280                 285

Ser Tyr Asp Thr Glu Tyr Gly Phe His Tyr Leu Asn Phe Asp Thr Glu
    290                 295                 300

Thr Gly Phe Gln Leu Asn Gly Lys Asn Val Lys Leu Lys Gly Val Cys
305                 310                 315                 320

Met His His Asp Gln Gly Ala Leu Gly Ala Val Ala Asn Arg Arg Ala
                325                 330                 335

Ile Glu Arg Gln Val Glu Ile Leu Gln Glu Met Gly Cys Asn Ser Ile
            340                 345                 350

Arg Val Thr His Asn Pro Ala Ser Lys Asp Leu Ile Glu Val Cys Asn
        355                 360                 365

Glu Lys Gly Ile Leu Val Ile Glu Glu Val Phe Asp Gly Trp His Arg
    370                 375                 380

Ala Lys Asn Gly Asn Ser Asn Asp Tyr Ser Val Trp Phe Glu Lys Ala
385                 390                 395                 400

Ile Glu Glu Asp Asn Ala Ile Leu Gly Lys Glu Ala Asp Met Thr Trp
```

405 410 415

Ala Glu Tyr Asp Leu Lys Ala Ile Met Lys Arg Asp Gln Asn Ala Pro
420 425 430

Ser Ile Ile Glu Trp Ser Leu Gly Asn Glu Ile Gln Glu Gly Ala Gly
435 440 445

Gly Ser Gly Tyr Ala Glu Arg Ala Asp Lys Leu Ile Lys Trp Ala Lys
450 455 460

Glu Ala Asp Ala Thr Lys Thr Leu Thr Ile Gly Ser Asn Ala Val Lys
465 470 475 480

Arg Gly Asp Trp Glu Gln Val Ser Ile Gly Asp Lys Leu Thr Lys Ala
485 490 495

Gly Gly Thr Ser Gly Thr Asn Tyr Ser Asp Gly Ala Ser Tyr Asp Lys
500 505 510

Ile His Lys Glu His Pro Asp Trp Lys Leu Tyr Gly Ser Glu Thr Ala
515 520 525

Ser Ser Val Asn Ser Arg Gly Ile Tyr Ser Val Thr Gly Asn Gln Glu
530 535 540

Ala Thr Ser Asp Gln Gln Leu Thr Ala Tyr Asp Asn Ser Arg Val Asn
545 550 555 560

Trp Gly Ala Leu Ala Ser Gln Ala Trp Tyr Asp Val Ile Gln Arg Asp
565 570 575

Phe Val Ala Gly Glu Tyr Val Trp Thr Gly Phe Asp Tyr Ile Gly Glu
580 585 590

Pro Thr Pro Trp Asn Gly Thr Asp Pro Gly Ala Lys Gly Thr Trp Pro
595 600 605

Ser Pro Lys Asn Ser Tyr Phe Gly Ile Ile Asp Thr Ala Gly Phe Pro
610 615 620

Lys Asp Ser Tyr Tyr Phe Tyr Gln Ser Gln Trp Asn Glu Glu Val Asn
625 630 635 640

Thr Leu His Val Leu Pro Ala Trp Asn Glu Asp Val Val Lys Lys Asn

645 650 655

Ser Asp Gly Thr Val Pro Val Val Val Tyr Ser Asp Ala Lys Glu Val
660 665 670

Glu Leu Phe Phe Thr Pro Ala Asn Gly Gly Glu Lys Lys Ser Leu Gly
675 680 685

Lys Lys Thr Phe Lys Thr Glu Thr Thr Lys Ala Gly Tyr Ser Tyr Gln
690 695 700

Val Leu Glu Asn Gly Lys Lys Lys His Lys Asp Leu Tyr Met Glu Trp
705 710 715 720

Gln Val Pro Tyr Glu Ala Gly Thr Leu Glu Ala Val Ala Lys Asp Ala
725 730 735

Lys Gly Asn Val Ile Lys Asp Thr Glu Gly Arg Ser Val Val Lys Thr
740 745 750

Thr Gly Glu Glu Ala Lys Leu Ser Ala Lys Thr Asp Arg Asn Ser Ile
755 760 765

Gln Ala Asp Gly Lys Asp Leu Ser Tyr Ile Thr Val Asp Val Thr Asp
770 775 780

Lys Asp Gly Asn Ile Val Pro Asp Ala Ala Asn Arg Val Thr Phe Asp
785 790 795 800

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

Asp His Asp Ser Tyr Lys Ala Asp Asn Arg Lys Ala Phe Ser Gly Lys
820 825 830

Val Leu Ala Ile Val Gln Ser Thr Glu Lys Ala Gly Glu Ile Thr Val
835 840 845

Thr Ala Lys Ala Asp Gly Leu Glu Ser Ser Thr Val Lys Ile Thr Thr
850 855 860

Thr Pro Val Lys Glu Glu Pro Ser Glu Arg Tyr Val Glu Ser Tyr Lys
865 870 875 880

Tyr Ser Lys Ser Tyr Tyr Val Lys Thr Gly Thr Lys Pro Gln Leu Pro

885 890 895

Lys Lys Ile Glu Ala Gln Tyr Ser Asp Arg Thr Lys Glu Asp Val Ala
900 905 910

Val Lys Trp Asp Glu Ile Ser Asp Glu Gln Ile Ser Lys Thr Gly Ser
915 920 925

Phe Thr Val Glu Gly Thr Val Gly Lys Arg Asp Ile Thr Val Asn Ile
930 935 940

Asn Met Ile Asp Asp Val Ala Ala Leu Leu Asn Tyr Ser Gly Ala Thr
945 950 955 960

Gln Lys Gly Val Lys Pro Gln Leu Pro Asp Val Arg Pro Ala Val Leu
965 970 975

Pro Asp Gly Thr Val Leu Ala Ala Ser Phe Pro Val Gln Trp Asp Glu
980 985 990

Lys Asp Ala Asp Thr Phe Gln Lys Pro Asp Glu Ile Val Thr Val Asn
995 1000 1005

Gly Ser Ala Asp Ile Phe Gly Lys Thr Ile Pro Val Thr Ala Ser
1010 1015 1020

Ile Arg Val Gln Lys Glu Asp Ile Lys Ile Gly Ser Ser Val Thr
1025 1030 1035

Asn Val Ala Lys Leu Ser Gln Asn Ile Gln Gly Ser Asp Thr Leu
1040 1045 1050

Glu Ala Ile Lys Asp Gly Lys Thr Glu Met Ser Leu Asn Asn Asp
1055 1060 1065

Gly Gly Pro Asn Glu Ser Ala Trp Ser Asn Trp Asp Ala Ser Gln
1070 1075 1080

Lys Gly Thr Lys Glu Ala Glu Leu Thr Phe Thr Phe Asp Thr Gln
1085 1090 1095

Gln Arg Ile Gly Glu Ile Val Ile His Phe Ala Lys Asp Asn Asn
1100 1105 1110

Ser Ile Arg Phe Pro Asp Ala Gly Thr Thr Glu Ile

1115 1120 1125

<210> 2
<211> 1150
<212> PRT
<213> Ruminococcus lactaris

<400> 2

```
Ala Gly Val Ser Val Pro Ala Leu Ala Gln Gln Ala Val Arg Thr Glu
1               5                  10                  15

Ser Gln Thr Gln Met Ser Ser Asp Pro Glu Leu Val Tyr Val Asn Asn
            20                  25                  30

Tyr Ser Ser Thr Ala Gln Arg Ser Gln Asn Phe Asn Ser Asn Trp Lys
        35                  40                  45

Phe Tyr Phe Gly Asp Ala Gly Asn Ala Gln Gly Ala Thr Phe Asp Asp
    50                  55                  60

Ser Lys Trp Glu Gln Val Ser Leu Pro His Asp Tyr Ser Ile Ser Gln
65                  70                  75                  80

Glu Tyr Ser Lys Ser Met Glu Ala Glu Ser Gly Tyr Leu Gly Gly Gly
                85                  90                  95

Thr Gly Trp Tyr Arg Lys Asn Phe Thr Leu Ser Ser Asp Thr Gln Gly
            100                 105                 110

Lys Arg Val Arg Ile Asp Phe Asp Gly Val Tyr Met Asn Ala Thr Val
        115                 120                 125

Trp Val Asn Gly His Glu Val Gly Thr His Pro Tyr Gly Tyr Thr Ser
    130                 135                 140

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

Ile Ala Val Lys Val Val Asn Asn Thr Pro Ser Ser Arg Trp Tyr Ser
                165                 170                 175

Gly Ser Gly Ile Tyr Arg Asp Val Asp Leu Thr Ile Thr Asp Asp Val
            180                 185                 190

His Val Asp Leu Asn Gly Thr Lys Val Thr Thr Pro Asn Leu Glu Thr
        195                 200                 205
```

```
Glu Lys Gly Ser Thr Val Asn Thr Asp Val Thr Ala Thr Val Ala Asn
    210                 215                 220

Asp Ser Asp Ala Ala Lys Ser Val Ala Val Arg His Thr Val Phe Pro
225                 230                 235                 240

Lys Asp Gly Ser Ala Asp Gln Ser Ile Gly Thr Val Thr Thr Asn Ala
                245                 250                 255

Gln Ser Ile Ala Ala Gly Ala Thr Ala Glu Ile Gln Ala Thr Val Pro
            260                 265                 270

Val Ser Asn Pro Glu Leu Trp Ser Val Glu Asn Pro Ser Leu Tyr Thr
        275                 280                 285

Val Arg Thr Glu Val Leu Val Asp Gly Gln Val Thr Asp Thr Tyr Asp
    290                 295                 300

Thr Glu Tyr Gly Phe Arg Tyr Phe Asn Phe Asp Ser Asn Thr Gly Phe
305                 310                 315                 320

Ser Leu Asn Gly Glu Asn Met Lys Leu Lys Gly Val Cys Met His His
                325                 330                 335

Asp Gln Gly Ser Leu Gly Ala Ala Ala Tyr Asp Ser Ala Ile Asp Arg
            340                 345                 350

Gln Val Lys Ile Leu Lys Glu Met Gly Cys Asn Ser Ile Arg Val Thr
        355                 360                 365

His Asn Pro Ala Ala Gln Asp Leu Ile Asp Ala Cys Asn Glu Gln Gly
    370                 375                 380

Ile Leu Val Val Glu Glu Ala Phe Asp Thr Trp Thr Arg Pro Lys Asn
385                 390                 395                 400

Gly Asn Ser Asn Asp Tyr Ser Val Trp Phe Asn Gln Thr Val Ala Ser
                405                 410                 415

Asp Asn Glu Ile Leu Gly Ala Thr Asn Gly Glu Thr Trp Ala Gln Phe
            420                 425                 430

Asp Leu Glu Ser Met Ile Ser Arg Asp Tyr Asn Ala Pro Ser Val Ile
        435                 440                 445
```

```
Met Trp Ser Leu Gly Asn Glu Val Met Glu Gly Ile Ser Gly Gly Thr
    450                 455                 460

Asp Ala Glu Tyr Glu Ala Thr Ala Thr Lys Leu Ile Asn Trp Ala Tyr
465                 470                 475                 480

Asp Ala Asp Asn Thr Arg Pro Met Thr Ile Gly Asp Asn Lys Leu Lys
                485                 490                 495

Ala Asn Trp Gln Ile Ser Lys Thr Phe Ala Arg Leu Leu Thr Glu Lys
            500                 505                 510

Gly Gly Thr Val Gly Phe Asn Tyr Ala Asp Gly Arg Val Leu Asp Ser
        515                 520                 525

Tyr His Ser Ser Asn Ser Asn Trp Leu Leu Tyr Gly Ser Glu Thr Ala
    530                 535                 540

Ser Ala Ile Asn Ser Arg Gly Ile Tyr Tyr Arg Thr Thr Gly Gly Gly
545                 550                 555                 560

Gln Thr Ser Asp Lys Gln Leu Thr Ser Tyr Asp Asn Ser Asn Val Gly
                565                 570                 575

Trp Gly Ala Thr Ala Ser Asn Ala Trp Tyr Thr Val Leu Thr Arg Asp
            580                 585                 590

Phe Ala Ala Gly Glu Tyr Val Trp Thr Gly Phe Asp Tyr Leu Gly Glu
        595                 600                 605

Pro Thr Pro Trp Asn Gly Thr Gly Ser Gly Ala Val Gly Ser Trp Pro
    610                 615                 620

Ser Pro Lys Asn Ser Tyr Phe Gly Ile Ile Asp Thr Ala Gly Phe Ala
625                 630                 635                 640

Lys Asp Ser Tyr Tyr Phe Tyr Gln Ser Gln Trp Asn Asp Asp Val Thr
                645                 650                 655

Thr Leu His Val Leu Pro Ala Trp Asn Asn Asn Val Val Ser Lys Asp
            660                 665                 670

Ser Ser Gly Asn Val Pro Val Val Val Tyr Ser Asp Ala Ala Ser Val
        675                 680                 685
```

```
Glu Leu Phe Phe Gln Ala Lys Gly Ser Asp Thr Lys Thr Ser Leu Gly
    690                 695                 700

Lys Lys Thr Phe Thr Gln Lys Thr Thr Asp Ala Gly Tyr Thr Tyr Gln
705                 710                 715                 720

Ile Tyr Glu Gly Ser Asp Lys Asn Ser Thr Thr Asp Lys Asn Leu Tyr
                725                 730                 735

Leu Thr Trp Asn Val Pro Tyr Ala Asp Gly Thr Val Ser Ala Val Ala
            740                 745                 750

Tyr Asn Ser Asn Gly Gln Lys Ile Thr Asp Thr Val Gly Gln Ser Ser
        755                 760                 765

Val Thr Thr Thr Gly Lys Ala Ser Lys Leu Lys Ala Ser Ala Asp His
    770                 775                 780

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

Val Thr Asp Ala Asn Gly Asn Ile Val Pro Asp Ala Glu Asn Arg Val
                805                 810                 815

Lys Phe Thr Val Glu Gly Asp Gly Glu Leu Val Gly Val Asp Asn Gly
            820                 825                 830

Ser Ser Pro Asp His Asp Ser Tyr Gln Ala Asp Asn Arg Lys Ala Phe
        835                 840                 845

Ser Gly Lys Val Leu Ala Ile Val Lys Ser Thr Lys Glu Ala Gly Thr
    850                 855                 860

Ile Thr Val Thr Ala Ser Ala Asp Gly Leu Asp Ser Ala Ser Val Lys
865                 870                 875                 880

Ile Thr Thr Thr Ala Val Asp Asn Gly Ser Thr Glu Lys Gln Ile Asp
                885                 890                 895

Ser Phe Lys Met Ser Arg Thr Tyr Tyr Val Lys Val Gly Ser Thr Pro
            900                 905                 910

Glu Leu Pro Glu Lys Ile Val Thr Arg Tyr Thr Asp Gly Thr Ser Glu
        915                 920                 925
```

```
Glu Leu Pro Val Thr Trp Asp Ala Ile Thr Glu Asp Gln Ile Ala Ala
    930                 935                 940

Ala Gly Ser Phe Gln Val Lys Gly Thr Val Lys Gly Gly Tyr Ser Val
945                 950                 955                 960

Ala Val Asn Val Asn Met Ile Asp Glu Val Gly Gly Leu Leu Asn Tyr
                965                 970                 975

Ser Thr Asn Thr Ala Val Gly Val Ala Pro Val Leu Pro Thr Ser Arg
            980                 985                 990

Pro Ala Val Leu Gln Asp Gly Thr  Val Met Asp Val Thr  Phe Pro Val
        995                 1000                1005

Thr Trp  Glu Asp Lys Ala Ala  Ser Ala Tyr Asp Lys  Ala Gly Thr
    1010                1015                1020

Val Thr  Val Asn Gly Thr Ala  Asn Val Leu Gly Lys  Glu Ile Ala
    1025                1030                1035

Val Thr  Ala Ser Val Arg Val  Gln Glu Glu Thr Ile  Thr Ile Gly
    1040                1045                1050

Asp Ser  Val Ser Ala Asp Ala  Leu Asn Leu Thr Gln  Ser Val Pro
    1055                1060                1065

Ala Asp  Lys Gln Ser Asp Thr  Leu Asn Ala Ile Lys  Asp Gly Ser
    1070                1075                1080

Thr Thr  Ile Ser Ser Asn Thr  Ser Gly Gly Ala Asn  Pro Thr Val
    1085                1090                1095

Trp Ser  Asn Tyr Asp Tyr Ser  Gln Asp Gly Asn Thr  Thr Ala Asp
    1100                1105                1110

Ile Ile  Phe Glu Tyr Ala Thr  Glu Gln Arg Leu Gly  Gln Ile Val
    1115                1120                1125

Thr His  Phe Ala Arg Asp Ser  Trp Ser Met Arg Tyr  Pro Asp Ala
    1130                1135                1140

Gly Ala  Thr Glu Ile Tyr Val
    1145                1150
```

<210> 3
<211> 91

<212> PRT
<213> Ruminococcus hansenii

<400> 3

```
Val Asn Trp Gly Ala Leu Ala Ser Gln Ala Trp Tyr Asp Val Ile Gln
1               5                   10                  15

Arg Asp Phe Val Ala Gly Glu Tyr Val Trp Thr Gly Phe Asp Tyr Ile
            20                  25                  30

Gly Glu Pro Thr Pro Trp Asn Gly Thr Asp Pro Gly Ala Lys Gly Thr
        35                  40                  45

Trp Pro Ser Pro Lys Asn Ser Tyr Phe Gly Ile Ile Asp Thr Ala Gly
    50                  55                  60

Phe Pro Lys Asp Ser Tyr Tyr Phe Tyr Gln Ser Gln Trp Asn Glu Glu
65                  70                  75                  80

Val Asn Thr Leu His Val Leu Pro Ala Trp Asn
                85                  90
```

<210> 4
<211> 71
<212> PRT
<213> Ruminococcus hansenii

<400> 4

```
Ala Gly Glu Tyr Val Trp Thr Gly Phe Asp Tyr Ile Gly Glu Pro Thr
1               5                   10                  15

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

Lys Asn Ser Tyr Phe Gly Ile Ile Asp Thr Ala Gly Phe Pro Lys Asp
        35                  40                  45

Ser Tyr Tyr Phe Tyr Gln Ser Gln Trp Asn Glu Glu Val Asn Thr Leu
    50                  55                  60

His Val Leu Pro Ala Trp Asn
65                  70
```

<210> 5
<211> 58
<212> PRT
<213> Ruminococcus hansenii

<400> 5

```
Ala Gly Glu Tyr Val Trp Thr Gly Phe Asp Tyr Ile Gly Glu Pro Thr
1               5                   10                  15

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

Lys Asn Ser Tyr Phe Gly Ile Ile Asp Thr Ala Gly Phe Pro Lys Asp
        35                  40                  45

Ser Tyr Tyr Phe Tyr Gln Ser Gln Trp Asn
    50                  55
```

<210> 6
<211> 91
<212> PRT
<213> Ruminococcus lactaris

<400> 6

```
Val Gly Trp Gly Ala Thr Ala Ser Asn Ala Trp Tyr Thr Val Leu Thr
1               5                   10                  15

Arg Asp Phe Ala Ala Gly Glu Tyr Val Trp Thr Gly Phe Asp Tyr Leu
            20                  25                  30

Gly Glu Pro Thr Pro Trp Asn Gly Thr Gly Ser Gly Ala Val Gly Ser
        35                  40                  45

Trp Pro Ser Pro Lys Asn Ser Tyr Phe Gly Ile Ile Asp Thr Ala Gly
    50                  55                  60

Phe Ala Lys Asp Ser Tyr Tyr Phe Tyr Gln Ser Gln Trp Asn Asp Asp
65                  70                  75                  80

Val Thr Thr Leu His Val Leu Pro Ala Trp Asn
                85                  90
```

<210> 7
<211> 71
<212> PRT
<213> Ruminococcus lactaris

<400> 7

```
Ala Gly Glu Tyr Val Trp Thr Gly Phe Asp Tyr Leu Gly Glu Pro Thr
```

```
1               5                   10                  15

Pro Trp Asn Gly Thr Gly Ser Gly Ala Val Gly Ser Trp Pro Ser Pro
            20                  25                  30

Lys Asn Ser Tyr Phe Gly Ile Ile Asp Thr Ala Gly Phe Ala Lys Asp
        35                  40                  45

Ser Tyr Tyr Phe Tyr Gln Ser Gln Trp Asn Asp Asp Val Thr Thr Leu
    50                  55                  60

His Val Leu Pro Ala Trp Asn
65                  70
```

<210> 8
<211> 58
<212> PRT
<213> Ruminococcus lactaris

<400> 8

```
Ala Gly Glu Tyr Val Trp Thr Gly Phe Asp Tyr Leu Gly Glu Pro Thr
1               5                   10                  15

Pro Trp Asn Gly Thr Gly Ser Gly Ala Val Gly Ser Trp Pro Ser Pro
            20                  25                  30

Lys Asn Ser Tyr Phe Gly Ile Ile Asp Thr Ala Gly Phe Ala Lys Asp
        35                  40                  45

Ser Tyr Tyr Phe Tyr Gln Ser Gln Trp Asn
    50                  55
```

<210> 9
<211> 30
<212> PRT
<213> Bacillus subtilis

<400> 9

```
Val Arg Ser Lys Lys Leu Trp Ile Ser Leu Leu Phe Ala Leu Ala Leu
1               5                   10                  15

Ile Phe Thr Met Ala Phe Gly Ser Thr Ser Ser Ala Gln Ala
            20                  25                  30
```

<210> 10
<211> 3477
<212> ADN
<213> Ruminococcus hansenii

<400> 10

gtgagaagca aaaaattgtg gatcagtttg ctgtttgctt tagcgttaat ctttacgatg 60

gcgttcggca gcacatccag cgcgcaggcg gcagggaaaa aagcagatag ccaaacacaa 120

atgtcatcag aaccggaaca agttgcggtt aaagattatg gctcaaatag cgcacgcaca 180

cagaattttg atagcgattg gaaatttaac ctgggagatg ttagcaatgc acagacaccg 240

acatttgatg attcaaaatg gcgcacactg tcactgccgc atgattatag catcgaacag 300

gaatattcac aatcactgga agcagaatca ggctatcttc cgggaggcgt tggctggtat 360

cgcaaaaatt ttacactggg cgaagaagcg aaaggcaaac gcattcgcat tgattttgat 420

ggcgtctata tgaatgcaac agtctatgtg aatggcaaag aagttggcac acatccgtat 480

ggctatacac cgtttagctt tgatatcaca gattatatca gctatgataa agaaaacaca 540

attgcggtca aagtcgatca tcaaacaccg tcatcaagat ggtattcagg cagcggcatt 600

tatagatcag tcaacctgac aacaacaaat gatgtccatg tcgatctgaa tggcattaaa 660

gtcgaaagca acaacctgga aaaagaagca ggcaaaacag tcaacacaga tgtgaaaaca 720

acagttgtga acggctcaaa agaagcgaaa aacatcacaa ttacacatac agtctttaaa 780

aaaggcgaaa aaccggataa agcgatcggc acatttacaa cagaagcgca agaaattggc 840

gcaggcaaaa aaacagaaat cagcgcaaca gtcccggtta aaaatccgga actgtggtca 900

gttgaaaatc cggcactgta tacaattcgc acagaagtta aagcaggcga taaactgctg 960

gatagctatg atacagaata tggctttcat tatctgaact ttgatacaga aacaggcttt 1020

cagctgaatg gcaaaaacgt taaactgaaa ggcgtttgca tgcatcatga tcaaggcgca 1080

cttggcgcag ttgcaaatag aagagcaatt gaacgccaag tcgaaattct gcaagaaatg 1140

ggctgcaata gcattagagt cacacataat ccggcaagca aagatctgat tgaagtctgc 1200

aacgaaaaag gcattctggt cattgaagaa gtttttgacg gctggcatag agcaaaaaat 1260

ggcaacagca acgattatag cgtctggttt gaaaaagcga tcgaagaaga taacgcgatt 1320

ctgggaaaag aagcggatat gacttgggca gaatatgatc tgaaagcgat tatgaaacgc 1380

gatcaaaatg caccgagcat tattgaatgg tcactgggca atgaaattca agaaggcgca 1440

ggcggatcag gctatgcaga aagagcggat aaactgatca aatgggcgaa agaagcagac 1500

gcaacaaaaa cactgacaat tggcagcaat gcagttaaaa gaggcgattg ggaacaagtt 1560

agcatcggcg ataaacttac aaaagcaggc ggaacatcag gcacaaatta ttcagatggc 1620

gcatcatatg ataaaattca taaagaacat ccggattgga aactgtatgg ctcagaaaca 1680

gcatcatcag ttaatagccg tggcatttat tcagttacag gcaatcaaga agcaacaagc 1740

gatcaacaac tgacagcgta tgataatagc agagttaatt ggggagcact ggcatcacaa 1800

gcatggtatg atgttatcca gagagatttt gtcgcaggcg aatatgtttg gacaggcttt 1860

gattatatcg gcgaaccgac accgtggaat ggcacagatc cgggagcaaa aggcacatgg 1920

ccgtcaccga aaaacagcta ctttggcatt atcgatacag caggctttcc gaaagattca 1980

tattattttt atcagagcca gtggaatgaa gaagtcaata cactgcacgt tcttccggca 2040

tggaatgaag atgtcgtcaa aaaaaactca gatggcacag ttccggttgt tgtttattca 2100

gatgcgaaag aagtcgaact gtttttttaca ccggcaaatg gcggagaaaa aaaaagcctg 2160

ggaaaaaaaa catttaaaac agaaacaaca aaagctggct atagctatca agttctggaa 2220

aacggcaaaa aaaaacataa agatctgtat atggaatggc aagttccgta tgaagcaggc 2280

acacttgaag cagttgcgaa agatgcaaaa ggcaacgtca ttaaagatac agaaggcaga 2340

agcgtcgtta aaacaacagg cgaagaagca aaactgtcag caaaaacgga tcgcaatagc 2400

attcaagcag atggcaaaga tctgtcatat attacagtcg atgtcacaga taaagatggc 2460

aatattgttc cggatgcagc aaatagagtc acatttgatg tccaaggcgc aggaaaactg 2520

gttggcgttg ataatggctc atcaccggat catgatagct ataaagcgga taaccgcaaa 2580

gcattttcag gcaaagttct ggcaattgtt cagtcaacag aaaaagcagg cgaaattaca 2640

gttacagcaa aagcagatgg cctggaatca agcacagtca aaatcacaac aacaccggtt 2700

aaagaagaac cgagcgaaag atatgtcgaa agctataaat acagcaaaag ctattatgtg 2760

aaaacaggca caaaaccgca actgccgaaa aaaattgaag cgcagtatag cgatcgcaca 2820

aaagaggatg ttgcggtcaa atgggatgaa atctcagatg aacaaattag caaaacaggc 2880

agctttacag ttgaaggcac agttggcaaa agagatatca cagtcaacat taacatgatc 2940

gatgatgttg cagcactgct gaattattca ggcgcaacac aaaaaggcgt taaaccgcaa 3000

cttccggatg ttagaccggc agttctgcct gatggcacag tcctggcagc atcatttccg 3060

gttcagtggg atgaaaaaga tgcggataca tttcagaaac cggatgaaat tgttacagtt 3120

aacggcagcg cagatatctt tggcaaaaca attccggtta cagcaagcat tagagtgcag 3180

aaagaagata tcaaaattgg cagcagcgtt acaaatgttg caaaactgag ccaaaatatt 3240

caaggcagcg atacactgga agcaatcaaa gatggcaaaa cagaaatgag cctgaataat 3300

gatggcggac cgaatgaatc agcatggtca aattgggatg catcacagaa aggcacaaaa 3360

gaagccgaac tgacatttac atttgataca cagcaacgca ttggcgaaat tgtcattcat 3420

tttgcgaaag ataacaactc aatcagattt ccggatgctg gcacaacaga aatctaa 3477

<210> 11
<211> 3534

<212> ADN
<213> Ruminococcus lactaris

<400> 11

gtggatcagt ttgctgtttg ctttagcgtt aatctttacg atggcgttcg gcagcacatc 60

cagcgcgcag gcggcaggga agcaggcgtt tcagttccgg cactggcaca acaagcagtt 120

agaacagaaa gccaaacaca aatgtcatca gatccggaac tggtctatgt gaataactat 180

agcagcacag cacaaagaag ccagaacttt aacagcaact ggaaattcta cttcggagat 240

gcgggaaatg cacaaggcgc aacatttgat gatagcaaat gggaacaagt ttcactgccg 300

catgattatt caatcagcca agaatatagc aaatcaatgg aagcagaatc aggctatctt 360

ggcggaggca caggctggta tcgcaaaaat tttacactga gcagcgatac acaaggcaaa 420

agagtccgca ttgattttga tggcgtctat atgaatgcaa cagtttgggt taatggccat 480

gaagttggca cacatccgta tggctataca agctttagct ttgatatcac agattatgtg 540

aaatatgatg gcgaaaacac aattgcagtc aaagtcgtca ataatacacc gtcaagcaga 600

tggtattcag gctcaggcat ttatagagat gtcgatctga caatcacaga tgatgttcat 660

gttgatctga acggcacaaa agttacaaca ccgaacctgg aaacagaaaa aggcagcaca 720

gtcaatacag atgttacagc aacagttgcg aatgattcag atgcagcaaa atcagttgca 780

gttcgccata cagttttcc gaaagatggc agcgcagatc aatcaattgg cacagtcaca 840

acaaatgcac aatcaattgc agcaggcgca acagcagaaa ttcaagcaac ggttccggtt 900

tcaaatcctg aactgtggtc agttgaaaat ccgtcactgt atacagtcag aacagaagtt 960

ctggtcgacg gccaagtcac agatacatat gatacagaat atggctttcg ctattttaac 1020

tttgatagca acacaggctt ttcactgaat ggcgaaaata tgaaactgaa aggcgtctgc 1080

atgcatcatg atcaaggctc acttggcgca gcagcatacg actcagcaat tgatcgccag 1140

gtcaaaatcc tgaaagaaat gggctgcaat agcattagag tcacacataa tccggcagca 1200

caagatctga ttgatgcgtg caatgaacaa ggcattctgg ttgttgaaga agcgtttgat 1260

acttggacaa gaccgaaaaa tggcaacagc aacgattata gcgtctggtt taatcagaca 1320

gttgcgagcg ataatgaaat tctgggagcg acaaatggcg aaacatgggc acaatttgat 1380

ctggaaagca tgatctcacg cgattataat gcaccgtcag tcattatgtg gtcactgggc 1440

aatgaagtta tggaaggcat tagcggaggc acagatgcag aatatgaagc gacagcgacg 1500

aaactgatta actgggcgta tgatgcggat aatacacgtc cgatgacaat tggcgataac 1560

aaactgaaag cgaactggca gatctcaaaa acatttgcga gactgctgac agaaaaaggc 1620

ggaacagtgg gctttaatta tgcagatggc agagttctgg attcatatca tagcagcaat 1680

agcaattggc tgctgtatgg ctcagaaaca gcatcagcga ttaatagccg tggcatctat 1740

tatagaacaa caggcggagg ccaaacatca gataaacagc tgacaagcta tgataattca 1800

aatgttggct ggggagcaac agcatcaaat gcatggtata cagttctgac aagagatttt 1860

gcggcaggcg aatatgtttg gacaggcttt gattatctgg gcgaaccgac accgtggaat 1920

ggcacaggct caggcgcagt tggctcatgg ccgtcaccga aaaattctta ttttggcatt 1980

atcgatacag caggcttcgc aaaagatagc tattattttt atcagagcca gtggaatgat 2040

gatgttacaa cactgcatgt tcttccggca tggaataata atgtcgtcag caaagattca 2100

tcaggcaatg ttccggttgt tgtttattca gatgcggcat cagtcgaact gttttttcaa 2160

gcaaaaggca gcgatacaaa aacaagcctg ggcaaaaaaa catttacaca gaaaacaaca 2220

gacgcaggct atacatatca gatctatgaa ggctcagata aaaacagcac aacagacaaa 2280

aacctgtatc tgacatggaa tgttccgtat gcagatggaa cagtttcagc agttgcgtat 2340

aatagcaacg gccagaaaat tacagataca gttggccagt cctcagttac aacaacaggc 2400

aaagcgtcaa aactgaaagc atcagcggat cataaaaaaa ttgcagcgga tggcgaatca 2460

ctgtcatata tcacagtcga tgtcacagat gcgaatggca atattgttcc ggatgcagaa 2520

aatcgcgtca aatttacagt tgaaggcgat ggcgaactgg ttggcgttga taatggctca 2580

tcaccggatc atgattcata tcaagcggat aaccgcaaag cattttcagg caaagttctg 2640

gcaattgtga aaagcacaaa agaagctggc acaattacag ttacagcatc agcagatggc 2700

ctggattcag catcagtcaa aatcacaaca acagcagtcg ataatggcag cacagaaaaa 2760

caaatcgata gctttaaaat gagccgcaca tattatgtta aagttggcag cacaccggaa 2820

ctgccggaaa aaattgtcac acgctataca gatggcacat cagaagaact gcctgttact 2880

tgggatgcaa ttacagaaga tcaaattgca gcagcaggct catttcaagt taaaggcaca 2940

gtcaaaggcg gatattcagt tgcagtcaac gtcaacatga ttgatgaagt tggcggactg 3000

ctgaattatt caacaaatac agcagttggc gttgcaccgg ttctgccgac atcaagaccg 3060

gcagttctgc aagatggcac agttatggat gttacatttc cggtcacatg ggaagataaa 3120

gcagcaagcg catatgataa agcaggcaca gtgacagtca atggcacagc aaatgttctg 3180

ggcaaagaaa ttgcagttac agcgagcgtt agagttcagg aagaaacaat cacaattgga 3240

gattcagttt cagcggatgc actgaatctg acacaaagcg ttccggcaga taaacaaagc 3300

gatacactga acgcaattaa agatggctca acaacaatta gctcaaatac aagcggaggc 3360

gcaaatccga cagtttggag caactatgac tatagccagg atggcaatac gacagcggat 3420

atcatttttg aatatgcgac agaacaaaga ctgggccaaa tcgttacaca ttttgcgaga 3480
gatagctggt caatgagata tcctgatgca ggcgctacag aaatttatgt ctaa 3534

<210> 12
<211> 1807
<212> PRT
<213> Ruminococcus hansenii

<400> 12

```
Met Tyr Phe Phe Gly Arg Ser Ala Ile Met Met Leu Thr Val Lys Thr
1               5                  10                  15

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

Ala Thr Leu Ala Ala Val Leu Thr Val Gln Gly Met Gly Phe Ser Ser
        35                  40                  45

Thr Val Tyr Ala Lys Glu Glu Pro Val Arg Val Lys Ala Asp Ser Gln
    50                  55                  60

Thr Gln Met Ser Ser Glu Pro Glu Gln Val Ala Val Lys Asp Tyr Gly
65                  70                  75                  80

Ser Asn Ser Ala Arg Thr Gln Asn Phe Asp Ser Asp Trp Lys Phe Asn
                85                  90                  95

Leu Gly Asp Val Ser Asn Ala Gln Thr Pro Thr Phe Asp Asp Ser Lys
            100                 105                 110

Trp Arg Thr Leu Ser Leu Pro His Asp Tyr Ser Ile Glu Gln Glu Tyr
        115                 120                 125

Ser Gln Ser Leu Glu Ala Glu Ser Gly Tyr Leu Pro Gly Gly Val Gly
    130                 135                 140

Trp Tyr Arg Lys Asn Phe Thr Leu Gly Glu Glu Ala Lys Gly Lys Arg
145                 150                 155                 160

Ile Arg Ile Asp Phe Asp Gly Val Tyr Met Asn Ala Thr Val Tyr Val
                165                 170                 175

Asn Gly Lys Glu Val Gly Thr His Pro Tyr Gly Tyr Thr Pro Phe Ser
            180                 185                 190

Phe Asp Ile Thr Asp Tyr Ile Ser Tyr Asp Lys Glu Asn Thr Ile Ala
```

```
        195                 200                 205

Val Lys Val Asp His Gln Thr Pro Ser Ser Arg Trp Tyr Ser Gly Ser
    210                 215                 220

Gly Ile Tyr Arg Ser Val Asn Leu Thr Thr Thr Asn Asp Val His Val
225                 230                 235                 240

Asp Leu Asn Gly Ile Lys Val Glu Ser Asn Asn Leu Glu Lys Glu Ala
                245                 250                 255

Gly Lys Thr Val Asn Thr Asp Val Lys Thr Thr Val Val Asn Gly Ser
            260                 265                 270

Lys Glu Ala Lys Asn Ile Thr Ile Thr His Thr Val Phe Lys Lys Gly
        275                 280                 285

Glu Lys Pro Asp Lys Ala Ile Gly Thr Phe Thr Thr Glu Ala Gln Glu
    290                 295                 300

Ile Gly Ala Gly Lys Lys Thr Glu Ile Ser Ala Thr Val Pro Val Lys
305                 310                 315                 320

Asn Pro Glu Leu Trp Ser Val Glu Asn Pro Ala Leu Tyr Thr Ile Arg
                325                 330                 335

Thr Glu Val Lys Ala Gly Asp Lys Leu Leu Asp Ser Tyr Asp Thr Glu
            340                 345                 350

Tyr Gly Phe His Tyr Leu Asn Phe Asp Thr Glu Thr Gly Phe Gln Leu
        355                 360                 365

Asn Gly Lys Asn Val Lys Leu Lys Gly Val Cys Met His His Asp Gln
    370                 375                 380

Gly Ala Leu Gly Ala Val Ala Asn Arg Arg Ala Ile Glu Arg Gln Val
385                 390                 395                 400

Glu Ile Leu Gln Glu Met Gly Cys Asn Ser Ile Arg Val Thr His Asn
                405                 410                 415

Pro Ala Ser Lys Asp Leu Ile Glu Val Cys Asn Glu Lys Gly Ile Leu
            420                 425                 430

Val Ile Glu Glu Val Phe Asp Gly Trp His Arg Ala Lys Asn Gly Asn
```

```
        435                 440                 445

Ser Asn Asp Tyr Ser Val Trp Phe Glu Lys Ala Ile Glu Glu Asp Asn
    450                 455                 460

Ala Ile Leu Gly Lys Glu Ala Asp Met Thr Trp Ala Glu Tyr Asp Leu
465                 470                 475                 480

Lys Ala Ile Met Lys Arg Asp Gln Asn Ala Pro Ser Ile Ile Glu Trp
                485                 490                 495

Ser Leu Gly Asn Glu Ile Gln Glu Gly Ala Gly Gly Ser Gly Tyr Ala
            500                 505                 510

Glu Arg Ala Asp Lys Leu Ile Lys Trp Ala Lys Glu Ala Asp Ala Thr
        515                 520                 525

Lys Thr Leu Thr Ile Gly Ser Asn Ala Val Lys Arg Gly Asp Trp Glu
    530                 535                 540

Gln Val Ser Ile Gly Asp Lys Leu Thr Lys Ala Gly Gly Thr Ser Gly
545                 550                 555                 560

Thr Asn Tyr Ser Asp Gly Ala Ser Tyr Asp Lys Ile His Lys Glu His
                565                 570                 575

Pro Asp Trp Lys Leu Tyr Gly Ser Glu Thr Ala Ser Ser Val Asn Ser
            580                 585                 590

Arg Gly Ile Tyr Ser Val Thr Gly Asn Gln Glu Ala Thr Ser Asp Gln
        595                 600                 605

Gln Leu Thr Ala Tyr Asp Asn Ser Arg Val Asn Trp Gly Ala Leu Ala
    610                 615                 620

Ser Gln Ala Trp Tyr Asp Val Ile Gln Arg Asp Phe Val Ala Gly Glu
625                 630                 635                 640

Tyr Val Trp Thr Gly Phe Asp Tyr Ile Gly Glu Pro Thr Pro Trp Asn
                645                 650                 655

Gly Thr Asp Pro Gly Ala Lys Gly Thr Trp Pro Ser Pro Lys Asn Ser
            660                 665                 670

Tyr Phe Gly Ile Ile Asp Thr Ala Gly Phe Pro Lys Asp Ser Tyr Tyr
```

675 680 685

Phe Tyr Gln Ser Gln Trp Asn Glu Glu Val Asn Thr Leu His Val Leu
690 695 700

Pro Ala Trp Asn Glu Asp Val Val Lys Lys Asn Ser Asp Gly Thr Val
705 710 715 720

Pro Val Val Val Tyr Ser Asp Ala Lys Glu Val Glu Leu Phe Phe Thr
725 730 735

Pro Ala Asn Gly Gly Glu Lys Lys Ser Leu Gly Lys Lys Thr Phe Lys
740 745 750

Thr Glu Thr Thr Lys Ala Gly Tyr Ser Tyr Gln Val Leu Glu Asn Gly
755 760 765

Lys Lys Lys His Lys Asp Leu Tyr Met Glu Trp Gln Val Pro Tyr Glu
770 775 780

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

Lys Asp Thr Glu Gly Arg Ser Val Val Lys Thr Thr Gly Glu Glu Ala
805 810 815

Lys Leu Ser Ala Lys Thr Asp Arg Asn Ser Ile Gln Ala Asp Gly Lys
820 825 830

Asp Leu Ser Tyr Ile Thr Val Asp Val Thr Asp Lys Asp Gly Asn Ile
835 840 845

Val Pro Asp Ala Ala Asn Arg Val Thr Phe Asp Val Gln Gly Ala Gly
850 855 860

Lys Leu Val Gly Val Asp Asn Gly Ser Ser Pro Asp His Asp Ser Tyr
865 870 875 880

Lys Ala Asp Asn Arg Lys Ala Phe Ser Gly Lys Val Leu Ala Ile Val
885 890 895

Gln Ser Thr Glu Lys Ala Gly Glu Ile Thr Val Thr Ala Lys Ala Asp
900 905 910

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

915 920 925

Glu Pro Ser Glu Arg Tyr Val Glu Ser Tyr Lys Tyr Ser Lys Ser Tyr
930 935 940

Tyr Val Lys Thr Gly Thr Lys Pro Gln Leu Pro Lys Lys Ile Glu Ala
945 950 955 960

Gln Tyr Ser Asp Arg Thr Lys Glu Asp Val Ala Val Lys Trp Asp Glu
965 970 975

Ile Ser Asp Glu Gln Ile Ser Lys Thr Gly Ser Phe Thr Val Glu Gly
980 985 990

Thr Val Gly Lys Arg Asp Ile Thr Val Asn Ile Asn Met Ile Asp Asp
995 1000 1005

Val Ala Ala Leu Leu Asn Tyr Ser Gly Ala Thr Gln Lys Gly Val
1010 1015 1020

Lys Pro Gln Leu Pro Asp Val Arg Pro Ala Val Leu Pro Asp Gly
1025 1030 1035

Thr Val Leu Ala Ala Ser Phe Pro Val Gln Trp Asp Glu Lys Asp
1040 1045 1050

Ala Asp Thr Phe Gln Lys Pro Asp Glu Ile Val Thr Val Asn Gly
1055 1060 1065

Ser Ala Asp Ile Phe Gly Lys Thr Ile Pro Val Thr Ala Ser Ile
1070 1075 1080

Arg Val Gln Lys Glu Asp Ile Lys Ile Gly Ser Ser Val Thr Asn
1085 1090 1095

Val Ala Lys Leu Ser Gln Asn Ile Gln Gly Ser Asp Thr Leu Glu
1100 1105 1110

Ala Ile Lys Asp Gly Lys Thr Glu Met Ser Leu Asn Asn Asp Gly
1115 1120 1125

Gly Pro Asn Glu Ser Ala Trp Ser Asn Trp Asp Ala Ser Gln Lys
1130 1135 1140

Gly Thr Lys Glu Ala Glu Leu Thr Phe Thr Phe Asp Thr Gln Gln

```
    1145                1150                1155

Arg Ile  Gly Glu Ile Val Ile  His Phe Ala Lys Asp  Asn Asn Ser
    1160                1165                1170

Ile Arg  Phe Pro Asp Ala Gly  Thr Thr Glu Ile Phe  Val Ser Glu
    1175                1180                1185

Thr Gly  Lys Asp Gly Thr Trp  Glu Lys Val Glu Val  Lys Glu His
    1190                1195                1200

Ile Gly  Glu Glu Lys Asp Arg  Val Lys Ala Tyr Arg  Tyr Glu Ile
    1205                1210                1215

Ala Pro  Val Thr Ala Thr Tyr  Val Lys Val Lys Val  Val Asn Ala
    1220                1225                1230

Asn Ala  Thr Asp Thr Gly Asn  Arg Lys Pro Cys Thr  Ala Ile Thr
    1235                1240                1245

Glu Val  Glu Leu Lys Lys Ala  Glu Gly Ser Phe Lys  Val Asn Glu
    1250                1255                1260

Thr Ala  Glu Leu Glu Glu Val  Lys Val Gly Glu Arg  Val Leu Pro
    1265                1270                1275

Asn Ala  Ala Tyr Ala Leu Asp  Ser Tyr Ser Val Pro  Glu Thr Asp
    1280                1285                1290

Ala Ala  Val Thr Ala Lys Thr  Lys Asp Asn Ala Ser  Leu Thr Ile
    1295                1300                1305

Leu Pro  Lys His Glu Asn Val  Val Arg Met Ile Leu  Glu Ser Glu
    1310                1315                1320

Asp His  Lys Ala Thr Lys Asn  Phe Ala Val Arg Met  Gly Glu Glu
    1325                1330                1335

Glu Thr  Val Leu Pro Asp Asp  Asp Ser Arg Asp Tyr  Pro Val Glu
    1340                1345                1350

Lys Ile  Thr Ala Thr Ala Gly  Ser Glu Tyr Lys Pro  Gly Thr Ala
    1355                1360                1365

Asn Glu  Gly Pro Val Lys Tyr  Val Leu Asp Gly Lys  Ala Glu Thr
```

```
    1370                1375                1380

His Trp  His Thr Asn Trp Ser  Val Ser Gly Glu Gly  Ser Lys Pro
    1385                1390                1395

Glu His  Arg Thr Val Thr Leu  Gln Leu Gly Asn Asp  Glu Glu Glu
    1400                1405                1410

Ala Pro  Met Ile Asp Ala Leu  Arg Tyr Met Pro Arg  Ser Asn Gly
    1415                1420                1425

Ala Asn  Gly Arg Val Thr Glu  Tyr Glu Ile Gln Tyr  Ser Leu Asp
    1430                1435                1440

Gly Asp  Lys Trp Gln Thr Ala  Ala Thr Gly Glu Ile  Asp Lys Lys
    1445                1450                1455

Gln Thr  Gly Trp Met Ile Leu  Gly Phe Glu Glu Pro  Val Gln Ala
    1460                1465                1470

Lys Tyr  Val Arg Phe Ile Gly  Thr His Thr Thr Ser  Asp Gln Gly
    1475                1480                1485

Asn Asp  Lys His Met Ala Val  Ser Glu Leu Arg Ala  Arg Val Ala
    1490                1495                1500

Thr Glu  Ala Pro Ala Pro Ser  Glu Lys Tyr Thr Ile  Thr Ala Asn
    1505                1510                1515

Val Asn  Asp Lys Thr Met Gly  Ala Val Thr Leu Asp  Ser Glu Thr
    1520                1525                1530

Gly Glu  Tyr Glu Lys Gly Thr  Lys Ala Thr Leu Thr  Ala Val Pro
    1535                1540                1545

Lys Glu  Gly Phe Ala Phe Val  Asn Trp Thr Ile Asp  Gly Gln Glu
    1550                1555                1560

Val Ser  Lys Glu Asn Pro Tyr  Ile His Thr Val Glu  Thr Asp Ala
    1565                1570                1575

Thr Ile  Thr Ala Asn Phe Glu  Arg Ile Glu Val Glu  Asn Glu Gly
    1580                1585                1590

Trp Val  Gln Thr Glu Asn Gly  Trp Glu Tyr Tyr Glu  Asn Gly Gln
```

```
    1595                1600                1605

Lys Val  Val Gly Trp Lys Glu  Val Ser Gly Lys Trp  Tyr Tyr Phe
    1610                1615                1620

Glu Glu  Asn Gly Leu Met Gln  Thr Gly Trp Val Phe  Val Asn Asn
    1625                1630                1635

His Trp  Tyr Tyr Met Asp Gln  Trp Gly Ala Met Cys  Ile Gly Trp
    1640                1645                1650

Val Ala  Val Asp Gly His Trp  Tyr Tyr Met Asp Gln  Trp Gly Ala
    1655                1660                1665

Met Cys  Thr Gly Trp Val Ser  Val Asn Gly His Trp  Tyr His Met
    1670                1675                1680

Asp Gln  Trp Gly Ala Met Gln  Thr Gly Trp Ala Leu  Val Asp Ser
    1685                1690                1695

Asn Trp  Tyr Tyr Leu Asn Thr  Asp Gly Ser Met Ala  Ile Gly Trp
    1700                1705                1710

Val Ala  Val Asn Gly His Trp  Tyr Tyr Met Asp Gln  Trp Gly Ala
    1715                1720                1725

Met Gln  Thr Gly Trp Ala Leu  Val Asp Ser Asn Trp  Tyr Tyr Leu
    1730                1735                1740

Asn Thr  Asp Gly Ser Met Ala  Ile Gly Trp Val Ala  Val Asn Gly
    1745                1750                1755

His Trp  Tyr Tyr Met Asp Gln  Trp Gly Ala Met Gln  Thr Gly Trp
    1760                1765                1770

Val Leu  Val Gly Ser Asp Trp  Tyr Tyr Leu Asn Thr  Asp Gly Ser
    1775                1780                1785

Met Ala  Ser Ser Gln Trp Ile  Asp Gly Tyr Tyr Val  Asp Ala Ser
    1790                1795                1800

Gly Lys  Met Lys
    1805
```

<210> 13
<211> 1768
<212> PRT
<213> Ruminococcus lactaris

<400> 13

```
Met Lys Lys Lys Lys Arg Cys Thr Arg Val Gly Ala Gly Ala Leu Ala
1               5                   10                  15

Ala Val Leu Ala Val Thr Ala Ala Gly Val Ser Val Pro Ala Leu Ala
            20                  25                  30

Gln Gln Ala Val Arg Thr Glu Ser Gln Thr Gln Met Ser Ser Asp Pro
        35                  40                  45

Glu Leu Val Tyr Val Asn Asn Tyr Ser Ser Thr Ala Gln Arg Ser Gln
    50                  55                  60

Asn Phe Asn Ser Asn Trp Lys Phe Tyr Phe Gly Asp Ala Gly Asn Ala
65                  70                  75                  80

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

His Asp Tyr Ser Ile Ser Gln Glu Tyr Ser Lys Ser Met Glu Ala Glu
            100                 105                 110

Ser Gly Tyr Leu Gly Gly Gly Thr Gly Trp Tyr Arg Lys Asn Phe Thr
        115                 120                 125

Leu Ser Ser Asp Thr Gln Gly Lys Arg Val Arg Ile Asp Phe Asp Gly
    130                 135                 140

Val Tyr Met Asn Ala Thr Val Trp Val Asn Gly His Glu Val Gly Thr
145                 150                 155                 160

His Pro Tyr Gly Tyr Thr Ser Phe Ser Phe Asp Ile Thr Asp Tyr Val
                165                 170                 175

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

Pro Ser Ser Arg Trp Tyr Ser Gly Ser Gly Ile Tyr Arg Asp Val Asp
        195                 200                 205

Leu Thr Ile Thr Asp Asp Val His Val Asp Leu Asn Gly Thr Lys Val
    210                 215                 220
```

```
Thr Thr Pro Asn Leu Glu Thr Glu Lys Gly Ser Thr Val Asn Thr Asp
225                 230                 235                 240

Val Thr Ala Thr Val Ala Asn Asp Ser Asp Ala Ala Lys Ser Val Ala
                245                 250                 255

Val Arg His Thr Val Phe Pro Lys Asp Gly Ser Ala Asp Gln Ser Ile
            260                 265                 270

Gly Thr Val Thr Thr Asn Ala Gln Ser Ile Ala Ala Gly Ala Thr Ala
        275                 280                 285

Glu Ile Gln Ala Thr Val Pro Val Ser Asn Pro Glu Leu Trp Ser Val
    290                 295                 300

Glu Asn Pro Ser Leu Tyr Thr Val Arg Thr Glu Val Leu Val Asp Gly
305                 310                 315                 320

Gln Val Thr Asp Thr Tyr Asp Thr Glu Tyr Gly Phe Arg Tyr Phe Asn
                325                 330                 335

Phe Asp Ser Asn Thr Gly Phe Ser Leu Asn Gly Glu Asn Met Lys Leu
            340                 345                 350

Lys Gly Val Cys Met His His Asp Gln Gly Ser Leu Gly Ala Ala Ala
        355                 360                 365

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

Cys Asn Ser Ile Arg Val Thr His Asn Pro Ala Ala Gln Asp Leu Ile
385                 390                 395                 400

Asp Ala Cys Asn Glu Gln Gly Ile Leu Val Val Glu Glu Ala Phe Asp
                405                 410                 415

Thr Trp Thr Arg Pro Lys Asn Gly Asn Ser Asn Asp Tyr Ser Val Trp
            420                 425                 430

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

Gly Glu Thr Trp Ala Gln Phe Asp Leu Glu Ser Met Ile Ser Arg Asp
    450                 455                 460
```

```
Tyr Asn Ala Pro Ser Val Ile Met Trp Ser Leu Gly Asn Glu Val Met
465                 470                 475                 480

Glu Gly Ile Ser Gly Gly Thr Asp Ala Glu Tyr Glu Ala Thr Ala Thr
                485                 490                 495

Lys Leu Ile Asn Trp Ala Tyr Asp Ala Asp Asn Thr Arg Pro Met Thr
            500                 505                 510

Ile Gly Asp Asn Lys Leu Lys Ala Asn Trp Gln Ile Ser Lys Thr Phe
        515                 520                 525

Ala Arg Leu Leu Thr Glu Lys Gly Gly Thr Val Gly Phe Asn Tyr Ala
    530                 535                 540

Asp Gly Arg Val Leu Asp Ser Tyr His Ser Ser Asn Ser Asn Trp Leu
545                 550                 555                 560

Leu Tyr Gly Ser Glu Thr Ala Ser Ala Ile Asn Ser Arg Gly Ile Tyr
                565                 570                 575

Tyr Arg Thr Thr Gly Gly Gly Gln Thr Ser Asp Lys Gln Leu Thr Ser
            580                 585                 590

Tyr Asp Asn Ser Asn Val Gly Trp Gly Ala Thr Ala Ser Asn Ala Trp
        595                 600                 605

Tyr Thr Val Leu Thr Arg Asp Phe Ala Ala Gly Glu Tyr Val Trp Thr
    610                 615                 620

Gly Phe Asp Tyr Leu Gly Glu Pro Thr Pro Trp Asn Gly Thr Gly Ser
625                 630                 635                 640

Gly Ala Val Gly Ser Trp Pro Ser Pro Lys Asn Ser Tyr Phe Gly Ile
                645                 650                 655

Ile Asp Thr Ala Gly Phe Ala Lys Asp Ser Tyr Tyr Phe Tyr Gln Ser
            660                 665                 670

Gln Trp Asn Asp Asp Val Thr Thr Leu His Val Leu Pro Ala Trp Asn
        675                 680                 685

Asn Asn Val Val Ser Lys Asp Ser Ser Gly Asn Val Pro Val Val Val
    690                 695                 700
```

```
Tyr Ser Asp Ala Ala Ser Val Glu Leu Phe Phe Gln Ala Lys Gly Ser
705                 710                 715                 720

Asp Thr Lys Thr Ser Leu Gly Lys Lys Thr Phe Thr Gln Lys Thr Thr
                725                 730                 735

Asp Ala Gly Tyr Thr Tyr Gln Ile Tyr Glu Gly Ser Asp Lys Asn Ser
            740                 745                 750

Thr Thr Asp Lys Asn Leu Tyr Leu Thr Trp Asn Val Pro Tyr Ala Asp
        755                 760                 765

Gly Thr Val Ser Ala Val Ala Tyr Asn Ser Asn Gly Gln Lys Ile Thr
    770                 775                 780

Asp Thr Val Gly Gln Ser Ser Val Thr Thr Thr Gly Lys Ala Ser Lys
785                 790                 795                 800

Leu Lys Ala Ser Ala Asp His Lys Lys Ile Ala Ala Asp Gly Glu Ser
                805                 810                 815

Leu Ser Tyr Ile Thr Val Asp Val Thr Asp Ala Asn Gly Asn Ile Val
            820                 825                 830

Pro Asp Ala Glu Asn Arg Val Lys Phe Thr Val Glu Gly Asp Gly Glu
        835                 840                 845

Leu Val Gly Val Asp Asn Gly Ser Ser Pro Asp His Asp Ser Tyr Gln
    850                 855                 860

Ala Asp Asn Arg Lys Ala Phe Ser Gly Lys Val Leu Ala Ile Val Lys
865                 870                 875                 880

Ser Thr Lys Glu Ala Gly Thr Ile Thr Val Thr Ala Ser Ala Asp Gly
                885                 890                 895

Leu Asp Ser Ala Ser Val Lys Ile Thr Thr Thr Ala Val Asp Asn Gly
            900                 905                 910

Ser Thr Glu Lys Gln Ile Asp Ser Phe Lys Met Ser Arg Thr Tyr Tyr
        915                 920                 925

Val Lys Val Gly Ser Thr Pro Glu Leu Pro Glu Lys Ile Val Thr Arg
    930                 935                 940
```

```
Tyr Thr Asp Gly Thr Ser Glu Glu Leu Pro Val Thr Trp Asp Ala Ile
945                 950                 955                 960

Thr Glu Asp Gln Ile Ala Ala Ala Gly Ser Phe Gln Val Lys Gly Thr
                965                 970                 975

Val Lys Gly Gly Tyr Ser Val Ala Val Asn Val Asn Met Ile Asp Glu
            980                 985                 990

Val Gly Gly Leu Leu Asn Tyr Ser  Thr Asn Thr Ala Val  Gly Val Ala
        995                 1000                1005

Pro Val  Leu Pro Thr Ser Arg  Pro Ala Val Leu Gln  Asp Gly Thr
    1010                1015                1020

Val Met  Asp Val Thr Phe Pro  Val Thr Trp Glu Asp  Lys Ala Ala
    1025                1030                1035

Ser Ala  Tyr Asp Lys Ala Gly  Thr Val Thr Val Asn  Gly Thr Ala
    1040                1045                1050

Asn Val  Leu Gly Lys Glu Ile  Ala Val Thr Ala Ser  Val Arg Val
    1055                1060                1065

Gln Glu  Glu Thr Ile Thr Ile  Gly Asp Ser Val Ser  Ala Asp Ala
    1070                1075                1080

Leu Asn  Leu Thr Gln Ser Val  Pro Ala Asp Lys Gln  Ser Asp Thr
    1085                1090                1095

Leu Asn  Ala Ile Lys Asp Gly  Ser Thr Thr Ile Ser  Ser Asn Thr
    1100                1105                1110

Ser Gly  Gly Ala Asn Pro Thr  Val Trp Ser Asn Tyr  Asp Tyr Ser
    1115                1120                1125

Gln Asp  Gly Asn Thr Thr Ala  Asp Ile Ile Phe Glu  Tyr Ala Thr
    1130                1135                1140

Glu Gln  Arg Leu Gly Gln Ile  Val Thr His Phe Ala  Arg Asp Ser
    1145                1150                1155

Trp Ser  Met Arg Tyr Pro Asp  Ala Gly Ala Thr Glu  Ile Tyr Val
    1160                1165                1170
```

```
Ser Pro  Asp Gly Thr Asn Trp  Ala Lys Leu Asp Thr  Thr Glu Thr
    1175                1180                1185

Ile Gly  Thr Glu Ser Gly Asn  Val Lys Pro Tyr Thr  Tyr Asp Phe
    1190                1195                1200

Ala Pro  Val Gly Ala Thr Phe  Val Lys Phe His Leu  Thr Asn Ser
    1205                1210                1215

Thr Gln  Ala Thr Gly Thr Thr  Ala Lys Ala Cys Thr  Gly Ile Thr
    1220                1225                1230

Glu Ile  Glu Leu Lys Val Ala  Thr Gly Ser Arg Thr  Thr Asn Thr
    1235                1240                1245

Thr Ala  Glu Leu Gln Thr Leu  Thr Val Asn Gly Lys  Glu Val Pro
    1250                1255                1260

Gln Thr  Ala Leu Asp Ser Lys  Val Tyr Thr Thr Pro  Ala Ile Leu
    1265                1270                1275

Ala Glu  Ile Glu Ala Thr Ala  Lys Asp Asn Ala Ser  Val Thr Val
    1280                1285                1290

Leu Pro  Ala Tyr Asn Asp Val  Ile Arg Ile Ile Val  Glu Ser Glu
    1295                1300                1305

Asp His  Gln Thr Arg Asn Thr  Tyr Glu Val Arg Leu  Asn Glu Ala
    1310                1315                1320

Glu Gln  Thr Thr Pro Asp Ser  Asp Ser Arg Asp Tyr  Pro Val Ser
    1325                1330                1335

Lys Leu  Thr Ala Ser Ala Gly  Ser Glu Gln Ser Thr  Thr Gly Val
    1340                1345                1350

Glu Gly  Pro Ala Ser Asn Ala  Lys Asp Gly Asp Glu  Ser Thr Leu
    1355                1360                1365

Trp His  Thr Arg Trp Ser Ala  Pro Ala Ala Thr Ser  Asp Gln Leu
    1370                1375                1380

Trp Phe  Thr Tyr Glu Leu Glu  Glu Glu Thr Val Leu  Asp Ala Leu
    1385                1390                1395
```

```
Arg Tyr  Leu Pro Arg Gln Gly  Thr Ala Asp Gly Gln  Asn Asn Gly
    1400                 1405                 1410

Arg Val  Asn Glu Tyr Arg Val  Glu Val Ser Thr Asp  Gly Ser Thr
    1415                 1420                 1425

Trp Thr  Thr Val Ser Thr Gly  Asn Trp Glu Asp Ser  Gln Asp Trp
    1430                 1435                 1440

Lys Leu  Ala Glu Phe Thr Glu  Pro Val Ala Ala Lys  Tyr Val Arg
    1445                 1450                 1455

Leu Thr  Gly Val His Thr Tyr  Gly Ser Ser Ala Ala  Asn Val Asp
    1460                 1465                 1470

Lys Tyr  Met Ser Ala Ala Glu  Ile Arg Leu Arg Met  Ala Glu Ser
    1475                 1480                 1485

Lys Thr  Asp Ile Ala Asp Ala  Ala Asn Gly Val Thr  Val Thr Ala
    1490                 1495                 1500

Pro Asp  Ser Ile Glu Val Ala  Lys Ala Asp Ala Glu  Asn Pro Val
    1505                 1510                 1515

Met Phe  Asp Leu Ser Asp Ile  Val Val Lys Ala Gly  Asp Thr Thr
    1520                 1525                 1530

Leu Arg  Tyr Gly Val Asp Tyr  Val Ile Ser Tyr Glu  Asn Asn Thr
    1535                 1540                 1545

Asp Phe  Gly Thr Ala Lys Leu  Val Ile Lys Gly Ile  Asp Gly Tyr
    1550                 1555                 1560

Thr Gly  Thr Leu Glu His Glu  Phe Thr Ile Thr Gln  Lys Ala Lys
    1565                 1570                 1575

Val Met  Thr Gly Ile Thr Trp  Asn Thr Lys Pro Glu  Lys Val Ile
    1580                 1585                 1590

Tyr Thr  Glu Gly Glu Thr Leu  Asp Val Thr Gly Leu  Val Ile Asn
    1595                 1600                 1605

Val Val  Tyr Asp Asp Asp Ser  Thr Glu Ala Val Ala  Tyr Ser Glu
    1610                 1615                 1620
```

```
Ala Asn  Ala Asp Glu Phe Thr  Phe Ser Pro Ala Leu  Asp Thr Lys
    1625                 1630                 1635

Leu Ala  Ala Thr Asp Lys Thr  Val Thr Val Thr Tyr  Lys Gly Ala
    1640                 1645                 1650

Ser Leu  Ile Tyr Asp Ile Thr  Val Asn Pro Lys Lys  Val Asp Pro
    1655                 1660                 1665

Thr Asp  Pro Asp Gln Pro Asp  Lys Pro Asp Thr Pro  Asp Asn Gly
    1670                 1675                 1680

Asn Asp  Asn Gly Asn Asp Asn  Asn Gly Asn Gly Asn  Asn Asn Gly
    1685                 1690                 1695

Thr Asp  Asp Gly Lys Lys Asp  Pro Gly Gln Ser Gly  Val Thr Asp
    1700                 1705                 1710

Asn Lys  Asn Gln Gly Asn Asn  Ser Asn Asn Gly Thr  Ala Ala Gly
    1715                 1720                 1725

Asn Lys  Ala Asn Ala Ala Ala  Lys Thr Gly Asp Thr  Ala Asn Met
    1730                 1735                 1740

Leu Leu  Pro Met Ile Ala Ala  Met Leu Ala Gly Thr  Ala Val Val
    1745                 1750                 1755

Gly Thr  Ile Ser Ile Arg Arg  Arg Arg Arg
    1760                 1765
```

<210> 14
<211> 5244
<212> ADN
<213> Ruminococcus hansenii

<400> 14

```
aaagcagata gccaaacaca aatgtcatca gaaccggaac aagttgcggt taaagattat       60

ggctcaaata gcgcacgcac acagaatttt gatagcgatt ggaaatttaa cctgggagat      120

gttagcaatg cacagacacc gacatttgat gattcaaaat ggcgcacact gtcactgccg      180

catgattata gcatcgaaca ggaatattca caatcactgg aagcagaatc aggctatctt      240

ccgggaggcg ttggctggta tcgcaaaaat tttacactgg gcgaagaagc gaaaggcaaa      300

cgcattcgca ttgattttga tggcgtctat atgaatgcaa cagtctatgt gaatggcaaa      360

gaagttggca cacatccgta tggctataca ccgtttagct ttgatatcac agattatatc      420
```

agctatgata aagaaaacac aattgcggtc aaagtcgatc atcaaacacc gtcatcaaga 480

tggtattcag gcagcggcat ttatagatca gtcaacctga caacaacaaa tgatgtccat 540

gtcgatctga atggcattaa agtcgaaagc aacaacctgg aaaaagaagc aggcaaaaca 600

gtcaacacag atgtgaaaac aacagttgtg aacggctcaa aagaagcgaa aaacatcaca 660

attacacata cagtctttaa aaaaggcgaa aaaccggata aagcgatcgg cacatttaca 720

acagaagcgc aagaaattgg cgcaggcaaa aaaacagaaa tcagcgcaac agtcccggtt 780

aaaaatccgg aactgtggtc agttgaaaat ccggcactgt atacaattcg cacagaagtt 840

aaagcaggcg ataaactgct ggatagctat gatacagaat atggctttca ttatctgaac 900

tttgatacag aaacaggctt tcagctgaat ggcaaaaacg ttaaactgaa aggcgtttgc 960

atgcatcatg atcaaggcgc acttggcgca gttgcaaata gaagagcaat tgaacgccaa 1020

gtcgaaattc tgcaagaaat gggctgcaat agcattagag tcacacataa tccggcaagc 1080

aaagatctga ttgaagtctg caacgaaaaa ggcattctgg tcattgaaga agtttttgac 1140

ggctggcata gagcaaaaaa tggcaacagc aacgattata gcgtctggtt tgaaaaagcg 1200

atcgaagaag ataacgcgat tctgggaaaa gaagcggata tgacttgggc agaatatgat 1260

ctgaaagcga ttatgaaacg cgatcaaaat gcaccgagca ttattgaatg gtcactgggc 1320

aatgaaattc aagaaggcgc aggcggatca ggctatgcag aaagagcgga taaactgatc 1380

aaatgggcga aagaagcaga cgcaacaaaa acactgacaa ttggcagcaa tgcagttaaa 1440

agaggcgatt gggaacaagt tagcatcggc gataaactta caaaagcagg cggaacatca 1500

ggcacaaatt attcagatgg cgcatcatat gataaaattc ataaagaaca tccggattgg 1560

aaactgtatg gctcagaaac agcatcatca gttaatagcc gtggcattta ttcagttaca 1620

ggcaatcaag aagcaacaag cgatcaacaa ctgacagcgt atgataatag cagagttaat 1680

tggggagcac tggcatcaca agcatggtat gatgttatcc agagagattt tgtcgcaggc 1740

gaatatgttt ggacaggctt tgattatatc ggcgaaccga caccgtggaa tggcacagat 1800

ccgggagcaa aaggcacatg gccgtcaccg aaaaacagct actttggcat tatcgataca 1860

gcaggctttc cgaaagattc atattatttt tatcagagcc agtggaatga agaagtcaat 1920

acactgcacg ttcttccggc atggaatgaa gatgtcgtca aaaaaaactc agatggcaca 1980

gttccggttg ttgtttattc agatgcgaaa gaagtcgaac tgtttttttac accggcaaat 2040

ggcggagaaa aaaaaagcct gggaaaaaaa acatttaaaa cagaaacaac aaaagctggc 2100

tatagctatc aagttctgga aaacggcaaa aaaaaacata aagatctgta tatggaatgg 2160

caagttccgt atgaagcagg cacacttgaa gcagttgcga aagatgcaaa aggcaacgtc 2220

attaaagata cagaaggcag aagcgtcgtt aaaacaacag gcgaagaagc aaaactgtca 2280

gcaaaaacgg atcgcaatag cattcaagca gatggcaaag atctgtcata tattacagtc 2340

gatgtcacag ataaagatgg caatattgtt ccggatgcag caaatagagt cacatttgat 2400

gtccaaggcg caggaaaact ggttggcgtt gataatggct catcaccgga tcatgatagc 2460

tataaagcgg ataaccgcaa agcattttca ggcaaagttc tggcaattgt tcagtcaaca 2520

gaaaaagcag gcgaaattac agttacagca aaagcagatg gcctggaatc aagcacagtc 2580

aaaatcacaa caacaccggt taaagaagaa ccgagcgaaa gatatgtcga aagctataaa 2640

tacagcaaaa gctattatgt gaaaacaggc acaaaaccgc aactgccgaa aaaaattgaa 2700

gcgcagtata gcgatcgcac aaaagaggat gttgcggtca aatgggatga aatctcagat 2760

gaacaaatta gcaaaacagg cagctttaca gttgaaggca cagttggcaa aagagatatc 2820

acagtcaaca ttaacatgat cgatgatgtt gcagcactgc tgaattattc aggcgcaaca 2880

caaaaaggcg ttaaaccgca acttccggat gttagaccgg cagttctgcc tgatggcaca 2940

gtcctggcag catcatttcc ggttcagtgg gatgaaaaag atgcggatac atttcagaaa 3000

ccggatgaaa ttgttacagt taacggcagc gcagatatct ttggcaaaac aattccggtt 3060

acagcaagca ttagagtgca gaaagaagat atcaaaattg gcagcagcgt tacaaatgtt 3120

gcaaaactga gccaaaatat tcaaggcagc gatacactgg aagcaatcaa agatggcaaa 3180

acagaaatga gcctgaataa tgatggcgga ccgaatgaat cagcatggtc aaattgggat 3240

gcatcacaga aaggcacaaa agaagccgaa ctgacattta catttgatac acagcaacgc 3300

attggcgaaa ttgtcattca ttttgcgaaa gataacaact caatcagatt tccggatgct 3360

ggcacaacag aaatctttgt ttcagaaaca ggcaaagatg gcacatggga aaaagttgaa 3420

gtcaaagagc atattggcga agaaaaagat cgcgtcaaag catatcgcta tgaaattgca 3480

ccggttacag cgacatatgt taaagttaaa gtcgtcaatg cgaacgcgac agatacaggc 3540

aatagaaaac cgtgcacagc aattacagaa gtcgaactga aaaaagcaga aggcagcttt 3600

aaagtcaacg aaacagcaga actggaagaa gttaaagttg gcgaacgtgt tctgccgaat 3660

gcagcatatg cactggattc atattcagtt ccggaaacgg atgcagcagt tacagcaaaa 3720

acaaaagata atgcgagcct gacaatcctg ccgaaacatg aaaatgtcgt cagaatgatt 3780

ctggaaagcg aagaccataa agcgacgaaa aactttgcag ttagaatggg cgaagaagaa 3840

acagttctgc cggatgatga ttcaagagat tatccggtcg aaaaaatcac agcaacagca 3900

ggctcagaat ataaaccggg aacagcaaat gaaggaccgg ttaaatatgt tctggatggc 3960

aaagcagaaa cacattggca tacaaattgg tcagtttcag gcgaaggctc aaaaccggaa 4020

catagaacag ttacactgca actgggcaat gatgaagaag aagcaccgat gattgacgca 4080

ctgagatata tgccgagatc aaatggcgca aatggcagag ttacggaata tgaaattcag 4140

tatagcctgg atggcgataa atggcaaaca gcagcaacag gcgaaatcga taaaaaacaa 4200

acaggctgga tgatcctggg ctttgaagaa ccggttcaag caaaatatgt ccgctttatt 4260

ggcacacata caacatcaga tcagggcaat gataaacata tggcagtttc agaactgaga 4320

gcaagagttg caacagaagc accggcaccg tcagagaagt atacaattac agcgaacgtc 4380

aacgataaaa caatgggagc agttacactt gatagcgaaa caggcgaata tgaaaaaggc 4440

acgaaagcaa cactgacagc agttccgaaa gaaggctttg catttgtcaa ctggacaatt 4500

gatggccaag aagtctcaaa agaaaacccg tatatccata cagttgaaac ggatgcgaca 4560

atcacagcga attttgaacg cattgaagtc gaaaatgaag gctgggttca aacagaaaat 4620

ggctgggaat attatgagaa tggccaaaaa gttgtcggct ggaaagaagt ttcaggcaaa 4680

tggtactact ttgaagaaaa tggcctgatg caaacaggat gggtctttgt taacaaccat 4740

tggtattata tggatcagtg gggggcaatg tgcattggct gggttgcagt tgatggccat 4800

tggtactaca tggaccaatg gggtgctatg tgtacaggct gggttagcgt caatggacat 4860

tggtatcata tggaccaatg gggagccatg caaacaggct gggcactggt tgattcaaat 4920

tggtattacc tgaatacgga tggctcaatg gcaattggat gggtcgcagt gaacggccac 4980

tggtattaca tggatcaatg gggagctatg cagacgggat gggctcttgt tgatagcaac 5040

tggtattatc ttaacacaga tggcagcatg gcaatcggct gggtggcggt taatggacac 5100

tggtactata tggatcaatg gggtgcaatg cagacaggct gggttctggt cggcagcgat 5160

tggtactatt taaacacgga tggatctatg gcatcaagcc aatggattga tggctattat 5220

gttgatgcaa gcggcaagat gaag 5244

<210> 15
<211> 5235
<212> ADN
<213> Ruminococcus lactaris

<400> 15

gcaggcgttt cagttccggc actggcacaa caagcagtta gaacagaaag ccaaacacaa 60

atgtcatcag atccggaact ggtctatgtg aataactata gcagcacagc acaaagaagc 120

cagaacttta acagcaactg gaaattctac ttcggagatg cgggaaatgc acaaggcgca 180

acatttgatg atagcaaatg ggaacaagtt tcactgccgc atgattattc aatcagccaa 240

gaatatagca aatcaatgga agcagaatca ggctatcttg gcggaggcac aggctggtat 300

cgcaaaaatt ttacactgag cagcgataca caaggcaaaa gagtccgcat tgattttgat 360

ggcgtctata tgaatgcaac agtttgggtt aatggccatg aagttggcac acatccgtat 420

ggctatacaa gctttagctt tgatatcaca gattatgtga aatatgatgg cgaaaacaca 480

attgcagtca aagtcgtcaa taatacaccg tcaagcagat ggtattcagg ctcaggcatt 540

tatagagatg tcgatctgac aatcacagat gatgttcatg ttgatctgaa cggcacaaaa 600

gttacaacac cgaacctgga aacagaaaaa ggcagcacag tcaatacaga tgttacagca 660

acagttgcga atgattcaga tgcagcaaaa tcagttgcag ttcgccatac agtttttccg 720

aaagatggca gcgcagatca atcaattggc acagtcacaa caaatgcaca atcaattgca 780

gcaggcgcaa cagcagaaat tcaagcaacg gttccggttt caaatcctga actgtggtca 840

gttgaaaatc cgtcactgta tacagtcaga acagaagttc tggtcgacgg ccaagtcaca 900

gatacatatg atacagaata tggctttcgc tattttaact ttgatagcaa cacaggcttt 960

tcactgaatg gcgaaaatat gaaactgaaa ggcgtctgca tgcatcatga tcaaggctca 1020

cttggcgcag cagcatacga ctcagcaatt gatcgccagg tcaaaatcct gaaagaaatg 1080

ggctgcaata gcattagagt cacacataat ccggcagcac aagatctgat tgatgcgtgc 1140

aatgaacaag gcattctggt tgttgaagaa gcgtttgata cttggacaag accgaaaaat 1200

ggcaacagca acgattatag cgtctggttt aatcagacag ttgcgagcga taatgaaatt 1260

ctgggagcga caaatggcga aacatgggca caatttgatc tggaaagcat gatctcacgc 1320

gattataatg caccgtcagt cattatgtgg tcactgggca atgaagttat ggaaggcatt 1380

agcggaggca cagatgcaga atatgaagcg acagcgacga aactgattaa ctgggcgtat 1440

gatgcggata atacacgtcc gatgacaatt ggcgataaca aactgaaagc gaactggcag 1500

atctcaaaaa catttgcgag actgctgaca gaaaaaggcg gaacagtggg ctttaattat 1560

gcagatggca gagttctgga ttcatatcat agcagcaata gcaattggct gctgtatggc 1620

tcagaaacag catcagcgat taatagccgt ggcatctatt atagaacaac aggcggaggc 1680

caaacatcag ataaacagct gacaagctat gataattcaa atgttggctg gggagcaaca 1740

gcatcaaatg catggtatac agttctgaca agagattttg cggcaggcga atatgtttgg 1800

acaggctttg attatctggg cgaaccgaca ccgtggaatg gcacaggctc aggcgcagtt 1860

ggctcatggc cgtcaccgaa aaattcttat tttggcatta tcgatacagc aggcttcgca 1920

aaagatagct attattttta tcagagccag tggaatgatg atgttacaac actgcatgtt 1980

cttccggcat ggaataataa tgtcgtcagc aaagattcat caggcaatgt tccggttgtt 2040

gtttattcag atgcggcatc agtcgaactg ttttttcaag caaaaggcag cgatacaaaa 2100

acaagcctgg gcaaaaaaac atttacacag aaaacaacag acgcaggcta tacatatcag 2160

atctatgaag gctcagataa aaacagcaca acagacaaaa acctgtatct gacatggaat 2220

gttccgtatg cagatggaac agtttcagca gttgcgtata atagcaacgg ccagaaaatt 2280

acagatacag ttggccagtc ctcagttaca acaacaggca aagcgtcaaa actgaaagca 2340

tcagcggatc ataaaaaaat tgcagcggat ggcgaatcac tgtcatatat cacagtcgat 2400

gtcacagatg cgaatggcaa tattgttccg gatgcagaaa atcgcgtcaa atttacagtt 2460

gaaggcgatg gcgaactggt tggcgttgat aatggctcat caccggatca tgattcatat 2520

caagcggata accgcaaagc attttcaggc aaagttctgg caattgtgaa aagcacaaaa 2580

gaagctggca caattacagt tacagcatca gcagatggcc tggattcagc atcagtcaaa 2640

atcacaacaa cagcagtcga taatggcagc acagaaaaac aaatcgatag ctttaaaatg 2700

agccgcacat attatgttaa agttggcagc acaccggaac tgccggaaaa aattgtcaca 2760

cgctatacag atggcacatc agaagaactg cctgttactt gggatgcaat tacagaagat 2820

caaattgcag cagcaggctc atttcaagtt aaaggcacag tcaaaggcgg atattcagtt 2880

gcagtcaacg tcaacatgat tgatgaagtt ggcggactgc tgaattattc aacaaataca 2940

gcagttggcg ttgcaccggt tctgccgaca tcaagaccgg cagttctgca agatggcaca 3000

gttatggatg ttacatttcc ggtcacatgg gaagataaag cagcaagcgc atatgataaa 3060

gcaggcacag tgacagtcaa tggcacagca aatgttctgg gcaaagaaat tgcagttaca 3120

gcgagcgtta gagttcagga agaaacaatc acaattggag attcagtttc agcggatgca 3180

ctgaatctga cacaaagcgt tccggcagat aaacaaagcg atacactgaa cgcaattaaa 3240

gatggctcaa caacaattag ctcaaataca agcggaggcg caaatccgac agtttggagc 3300

aactatgact atagccagga tggcaatacg acagcggata tcatttttga atatgcgaca 3360

gaacaaagac tgggccaaat cgttacacat tttgcgagag atagctggtc aatgagatat 3420

cctgatgcag gcgctacaga aatttatgtc tcaccggatg gcacaaattg ggcaaaactg 3480

gatacaacag aaacaattgg cacagaaagc ggcaatgtta aaccgtatac atatgatttt 3540

gcaccggttg gcgcaacatt tgttaaattt catctgacaa acagcacaca agcaacaggc 3600

acaacagcaa aagcatgcac aggcattaca gaaattgaac tgaaagttgc aacaggctca 3660

cgcacaacaa atacaacagc agaactgcaa acactgacag ttaatggcaa agaagttccg 3720

caaacagcac tggatagcaa agtttataca acaccggcaa ttctggcaga aattgaagca 3780

acagcgaaag ataatgcaag cgttacagtt cttccggcat ataatgatgt cattcgcatt 3840

attgtcgaaa gcgaagatca tcaaacacgc aatacatatg aagtcagact gaatgaagcg 3900

gaacaaacaa caccggattc agattcaaga gattatccgg ttagcaaact gacagcatca 3960

gcaggctcag aacaatcaac aacaggcgtt gaaggaccgg catcaaatgc aaaagacggt 4020

gatgaatcaa cactgtggca tacaagatgg tcagcaccgg cagcaacatc agatcaactg 4080

tggtttacat atgaactgga agaagaaacg gtactggacg cactgagata tctgccgaga 4140

caaggcacag cagatggcca aaataatggc agagttaatg aatatcgcgt cgaagttagc 4200

acagatggca gcacatggac aacagtttca acaggcaatt gggaagatag ccaagattgg 4260

aaactggcag aatttacaga accggttgca gcaaaatatg tcagactgac aggcgttcat 4320

acatatggct catcagcagc aaacgtcgat aaatacatga gcgcagcaga aattagactg 4380

agaatggcag aaagcaaaac ggatattgca gatgcagcaa atggcgttac agttacagca 4440

ccggattcaa ttgaagttgc aaaagcagat gcagaaaacc cggttatgtt tgatctgagc 4500

gatattgttg tcaaagcagg cgatacaaca ctgagatatg gcgttgatta tgtcattagc 4560

tatgaaaaca acacagattt tggcacagcg aaactggtca ttaaaggcat tgatggctat 4620

acaggcacac tggaacatga attcacaatc acgcagaaag ccaaagtcat gacaggcatc 4680

acatggaata caaaaccgga aaaagtcatt tatacggaag gtgaaacgct ggatgttaca 4740

ggcctggtta ttaatgtcgt ctatgatgat gatagcacag aagcagttgc atatagcgaa 4800

gcaaatgcgg atgaatttac attttcaccg gcactggata caaaactggc agcgacagat 4860

aaaacagtca cagttacata taaaggcgca agcctgattt atgatattac agtcaacccg 4920

aaaaaagtcg atccgacaga tccggatcag cctgataaac cggatacacc ggataatggc 4980

aatgataacg gcaacgataa taatggcaac ggcaataaca acggcacaga tgatggcaaa 5040

aaagatccgg gacaatcagg cgttacagat aacaaaaatc agggcaataa cagcaataat 5100

ggaacagcag caggcaataa agcaaatgca gcagcaaaaa caggcgatac agcaaatatg 5160

ctgctgccga tgattgcagc aatgctggca ggcacagcag ttgttggcac aatttcaatt 5220

cgcagacgca gacgc 5235

<210> 16
<211> 3375
<212> ADN
<213> Ruminococcus hansenii

<400> 16

aaagcagata gccaaacaca aatgtcatca gaaccggaac aagttgcggt taaagattat 60

ggctcaaata gcgcacgcac acagaatttt gatagcgatt ggaaatttaa cctgggagat 120

gttagcaatg cacagacacc gacatttgat gattcaaaat ggcgcacact gtcactgccg 180

catgattata gcatcgaaca ggaatattca caatcactgg aagcagaatc aggctatctt 240

ccgggaggcg ttggctggta tcgcaaaaat tttacactgg gcgaagaagc gaaaggcaaa 300

cgcattcgca ttgattttga tggcgtctat atgaatgcaa cagtctatgt gaatggcaaa 360

gaagttggca cacatccgta tggctataca ccgtttagct ttgatatcac agattatatc 420

agctatgata aagaaaacac aattgcggtc aaagtcgatc atcaaacacc gtcatcaaga 480

tggtattcag gcagcggcat ttatagatca gtcaacctga caacaacaaa tgatgtccat 540

gtcgatctga atggcattaa agtcgaaagc aacaacctgg aaaaagaagc aggcaaaaca 600

gtcaacacag atgtgaaaac aacagttgtg aacggctcaa aagaagcgaa aaacatcaca 660

attacacata cagtctttaa aaaaggcgaa aaaccggata aagcgatcgg cacatttaca 720

acagaagcgc aagaaattgg cgcaggcaaa aaaacagaaa tcagcgcaac agtcccggtt 780

aaaaatccgg aactgtggtc agttgaaaat ccggcactgt atacaattcg cacagaagtt 840

aaagcaggcg ataaactgct ggatagctat gatacagaat atggctttca ttatctgaac 900

tttgatacag aaacaggctt tcagctgaat ggcaaaaacg ttaaactgaa aggcgtttgc 960

atgcatcatg atcaaggcgc acttggcgca gttgcaaata gaagagcaat tgaacgccaa 1020

gtcgaaattc tgcaagaaat gggctgcaat agcattagag tcacacataa tccggcaagc 1080

aaagatctga ttgaagtctg caacgaaaaa ggcattctgg tcattgaaga agtttttgac 1140

ggctggcata gagcaaaaaa tggcaacagc aacgattata gcgtctggtt tgaaaaagcg 1200

atcgaagaag ataacgcgat tctgggaaaa gaagcggata tgacttgggc agaatatgat 1260

ctgaaagcga ttatgaaacg cgatcaaaat gcaccgagca ttattgaatg gtcactgggc 1320

aatgaaattc aagaaggcgc aggcggatca ggctatgcag aaagagcgga taaactgatc 1380

aaatgggcga aagaagcaga cgcaacaaaa acactgacaa ttggcagcaa tgcagttaaa 1440

agaggcgatt gggaacaagt tagcatcggc gataaactta caaaagcagg cggaacatca 1500

ggcacaaatt attcagatgg cgcatcatat gataaaattc ataaagaaca tccggattgg 1560

aaactgtatg gctcagaaac agcatcatca gttaatagcc gtggcattta ttcagttaca 1620

ggcaatcaag aagcaacaag cgatcaacaa ctgacagcgt atgataatag cagagttaat 1680

tggggagcac tggcatcaca agcatggtat gatgttatcc agagagattt tgtcgcaggc 1740

gaatatgttt ggacaggctt tgattatatc ggcgaaccga caccgtggaa tggcacagat 1800

ccgggagcaa aaggcacatg gccgtcaccg aaaaacagct actttggcat tatcgataca 1860

gcaggctttc cgaaagattc atattatttt tatcagagcc agtggaatga agaagtcaat 1920

acactgcacg ttcttccggc atggaatgaa gatgtcgtca aaaaaaactc agatggcaca 1980

gttccggttg ttgtttattc agatgcgaaa gaagtcgaac tgtttttttac accggcaaat 2040

ggcggagaaa aaaaaagcct gggaaaaaaa acatttaaaa cagaaacaac aaaagctggc 2100

tatagctatc aagttctgga aaacggcaaa aaaaaacata aagatctgta tatggaatgg 2160

caagttccgt atgaagcagg cacacttgaa gcagttgcga aagatgcaaa aggcaacgtc 2220

attaaagata cagaaggcag aagcgtcgtt aaaacaacag gcgaagaagc aaaactgtca 2280

gcaaaaacgg atcgcaatag cattcaagca gatggcaaag atctgtcata tattacagtc 2340

gatgtcacag ataaagatgg caatattgtt ccggatgcag caaatagagt cacatttgat 2400

gtccaaggcg caggaaaact ggttggcgtt gataatggct catcaccgga tcatgatagc 2460

tataaagcgg ataaccgcaa agcattttca ggcaaagttc tggcaattgt tcagtcaaca 2520

gaaaaagcag gcgaaattac agttacagca aaagcagatg gcctggaatc aagcacagtc 2580

aaaatcacaa caacaccggt taaagaagaa ccgagcgaaa gatatgtcga aagctataaa 2640

tacagcaaaa gctattatgt gaaaacaggc acaaaaccgc aactgccgaa aaaaattgaa 2700

gcgcagtata gcgatcgcac aaaagaggat gttgcggtca aatgggatga aatctcagat 2760

gaacaaatta gcaaaacagg cagctttaca gttgaaggca cagttggcaa aagagatatc 2820

acagtcaaca ttaacatgat cgatgatgtt gcagcactgc tgaattattc aggcgcaaca 2880

caaaaaggcg ttaaaccgca acttccggat gttagaccgg cagttctgcc tgatggcaca 2940

gtcctggcag catcatttcc ggttcagtgg gatgaaaaag atgcggatac atttcagaaa 3000

ccggatgaaa ttgttacagt taacggcagc gcagatatct ttggcaaaac aattccggtt 3060

acagcaagca ttagagtgca gaaagaagat atcaaaattg gcagcagcgt tacaaatgtt 3120

gcaaaactga gccaaaatat tcaaggcagc gatacactgg aagcaatcaa agatggcaaa 3180

acagaaatga gcctgaataa tgatggcgga ccgaatgaat cagcatggtc aaattgggat 3240

gcatcacaga aaggcacaaa agaagccgaa ctgacattta catttgatac acagcaacgc 3300

attggcgaaa ttgtcattca ttttgcgaaa gataacaact caatcagatt tccggatgct 3360

ggcacaacag aaatc 3375

<210> 17
<211> 3450
<212> ADN
<213> Ruminococcus lactaris

<400> 17

gcaggcgttt cagttccggc actggcacaa caagcagtta gaacagaaag ccaaacacaa 60

atgtcatcag atccggaact ggtctatgtg aataactata gcagcacagc acaaagaagc 120

cagaacttta acagcaactg gaaattctac ttcggagatg cgggaaatgc acaaggcgca 180

acatttgatg atagcaaatg ggaacaagtt tcactgccgc atgattattc aatcagccaa 240

gaatatagca aatcaatgga agcagaatca ggctatcttg gcggaggcac aggctggtat 300

cgcaaaaatt ttacactgag cagcgataca caaggcaaaa gagtccgcat tgattttgat 360

ggcgtctata tgaatgcaac agtttgggtt aatggccatg aagttggcac acatccgtat 420

ggctatacaa gctttagctt tgatatcaca gattatgtga aatatgatgg cgaaaacaca 480

attgcagtca aagtcgtcaa taatacaccg tcaagcagat ggtattcagg ctcaggcatt 540

tatagagatg tcgatctgac aatcacagat gatgttcatg ttgatctgaa cggcacaaaa 600

gttacaacac cgaacctgga aacagaaaaa ggcagcacag tcaatacaga tgttacagca 660

acagttgcga atgattcaga tgcagcaaaa tcagttgcag ttcgccatac agtttttccg 720

aaagatggca gcgcagatca atcaattggc acagtcacaa caaatgcaca atcaattgca 780

gcaggcgcaa cagcagaaat tcaagcaacg gttccggttt caaatcctga actgtggtca 840

gttgaaaatc cgtcactgta tacagtcaga acagaagttc tggtcgacgg ccaagtcaca 900

gatacatatg atacagaata tggctttcgc tattttaact ttgatagcaa cacaggcttt 960

tcactgaatg gcgaaaatat gaaactgaaa ggcgtctgca tgcatcatga tcaaggctca 1020

cttggcgcag cagcatacga ctcagcaatt gatcgccagg tcaaaatcct gaaagaaatg 1080

ggctgcaata gcattagagt cacacataat ccggcagcac aagatctgat tgatgcgtgc 1140

aatgaacaag gcattctggt tgttgaagaa gcgtttgata cttggacaag accgaaaaat 1200

ggcaacagca acgattatag cgtctggttt aatcagacag ttgcgagcga taatgaaatt 1260

ctgggagcga caaatggcga aacatgggca caatttgatc tggaaagcat gatctcacgc 1320

gattataatg caccgtcagt cattatgtgg tcactgggca atgaagttat ggaaggcatt 1380

agcggaggca cagatgcaga atatgaagcg acagcgacga aactgattaa ctgggcgtat 1440

gatgcggata atacacgtcc gatgacaatt ggcgataaca aactgaaagc gaactggcag 1500

atctcaaaaa catttgcgag actgctgaca gaaaaaggcg gaacagtggg ctttaattat 1560

gcagatggca gagttctgga ttcatatcat agcagcaata gcaattggct gctgtatggc 1620

tcagaaacag catcagcgat taatagccgt ggcatctatt atagaacaac aggcggaggc 1680

caaacatcag ataaacagct gacaagctat gataattcaa atgttggctg gggagcaaca 1740

gcatcaaatg catggtatac agttctgaca agagattttg cggcaggcga atatgtttgg 1800

acaggctttg attatctggg cgaaccgaca ccgtggaatg gcacaggctc aggcgcagtt 1860

ggctcatggc cgtcaccgaa aaattcttat tttggcatta tcgatacagc aggcttcgca 1920

aaagatagct attattttta tcagagccag tggaatgatg atgttacaac actgcatgtt 1980

cttccggcat ggaataataa tgtcgtcagc aaagattcat caggcaatgt tccggttgtt 2040

gtttattcag atgcggcatc agtcgaactg ttttttcaag caaaaggcag cgatacaaaa 2100

acaagcctgg gcaaaaaaac atttacacag aaaacaacag acgcaggcta tacatatcag 2160

atctatgaag gctcagataa aaacagcaca acagacaaaa acctgtatct gacatggaat 2220

gttccgtatg cagatggaac agtttcagca gttgcgtata atagcaacgg ccagaaaatt 2280

acagatacag ttggccagtc ctcagttaca acaacaggca aagcgtcaaa actgaaagca 2340

tcagcggatc ataaaaaaat tgcagcggat ggcgaatcac tgtcatatat cacagtcgat 2400

gtcacagatg cgaatggcaa tattgttccg gatgcagaaa atcgcgtcaa atttacagtt 2460

gaaggcgatg gcgaactggt tggcgttgat aatggctcat caccggatca tgattcatat 2520

caagcggata accgcaaagc attttcaggc aaagttctgg caattgtgaa aagcacaaaa 2580

gaagctggca caattacagt tacagcatca gcagatggcc tggattcagc atcagtcaaa 2640

atcacaacaa cagcagtcga taatggcagc acagaaaaac aaatcgatag ctttaaaatg 2700

agccgcacat attatgttaa agttggcagc acaccggaac tgccggaaaa aattgtcaca 2760

cgctatacag atggcacatc agaagaactg cctgttactt gggatgcaat tacagaagat 2820

caaattgcag cagcaggctc atttcaagtt aaaggcacag tcaaaggcgg atattcagtt 2880

gcagtcaacg tcaacatgat tgatgaagtt ggcggactgc tgaattattc aacaaataca 2940

gcagttggcg ttgcaccggt tctgccgaca tcaagaccgg cagttctgca agatggcaca 3000

gttatggatg ttacatttcc ggtcacatgg gaagataaag cagcaagcgc atatgataaa 3060

gcaggcacag tgacagtcaa tggcacagca aatgttctgg gcaaagaaat tgcagttaca 3120

gcgagcgtta gagttcagga agaaacaatc acaattggag attcagtttc agcggatgca 3180

ctgaatctga cacaaagcgt tccggcagat aaacaaagcg atacactgaa cgcaattaaa 3240

gatggctcaa caacaattag ctcaaataca agcggaggcg caaatccgac agtttggagc 3300

aactatgact atagccagga tggcaatacg acagcggata tcatttttga atatgcgaca 3360

gaacaaagac tgggccaaat cgttacacat tttgcgagag atagctggtc aatgagatat 3420

cctgatgcag gcgctacaga aatttatgtc 3450

REIVINDICACIONES

1. Un polipéptido aislado que tiene actividad transgalactosilante y que comprende una secuencia de aminoácidos que tiene al menos 80% de identidad de secuencia con la secuencia de aminoácidos del polipéptido maduro de SEQ ID NO: 1, cuyo polipéptido tiene una longitud entre 900 y 1180 aminoácidos.

2. El polipéptido según la reivindicación 1 que tiene una relación entre actividad transgalactosilante y actividad β-galactosidasa de al menos 1, al menos 2,5, al menos 3, al menos 4, al menos 5, al menos 6, al menos 7, al menos 8, al menos 9, al menos 10, al menos 11, o al menos 12.

3. El polipéptido según una cualquiera de las reivindicaciones 1-2, en donde la secuencia de aminoácidos tiene al menos 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, o 99%, de identidad de secuencia con la secuencia de aminoácidos madura de SEQ ID NO: 1.

4. El polipéptido según las reivindicaciones 1-3 que contiene el dominio catalítico de glicosil hidrolasa clase 2 (GH2), que preferiblemente contiene uno o más dominios Pfam seleccionados de: Glyco_hydro2N (PF02837), Glyco_hydro (PF00703), Glyco_hydro 2C (PF02836) y dominio similar a Ig bacteriana (grupo 4) (PF07532).

5. El polipéptido según una cualquiera de las reivindicaciones 1-4 que comprende o consiste en la secuencia de aminoácidos de SEQ ID NO: 1.

6. El polipéptido según una cualquiera de las reivindicaciones 1-5 que es un fragmento del polipéptido maduro de SEQ ID NO: 12.

7. El polipéptido según una cualquiera de las reivindicaciones 1-6, que se obtiene a partir de *Ruminococcus hansenii* o *Ruminococcus lactaris*.

8. Un ácido nucleico que codifica un polipéptido según una cualquiera de las reivindicaciones 1-7.

9. Un vector de expresión que comprende un ácido nucleico según la reivindicación 8 y que expresa un polipéptido según una cualquiera de las reivindicaciones 1-7.

10. Una célula que expresa un polipéptido según una cualquiera de las reivindicaciones 1-7.

11. Un método para expresar un polipéptido, comprendiendo el método utilizar una célula según la reivindicación 10 y expresar el polipéptido en la célula, y opcionalmente purificar el polipéptido.

12. Una composición que comprende un polipéptido según una cualquiera de las reivindicaciones 1-7, preferiblemente una composición alimentaria, más preferiblemente un producto lácteo.

13. Un método para producir un producto alimentario tratando un sustrato tal como un sustrato a base de leche que comprende lactosa con un polipéptido como se define en una cualquiera de las reivindicaciones 1-7, y opcionalmente tratando además el sustrato con una beta-galactosidasa hidrolizante.

Fig. 1

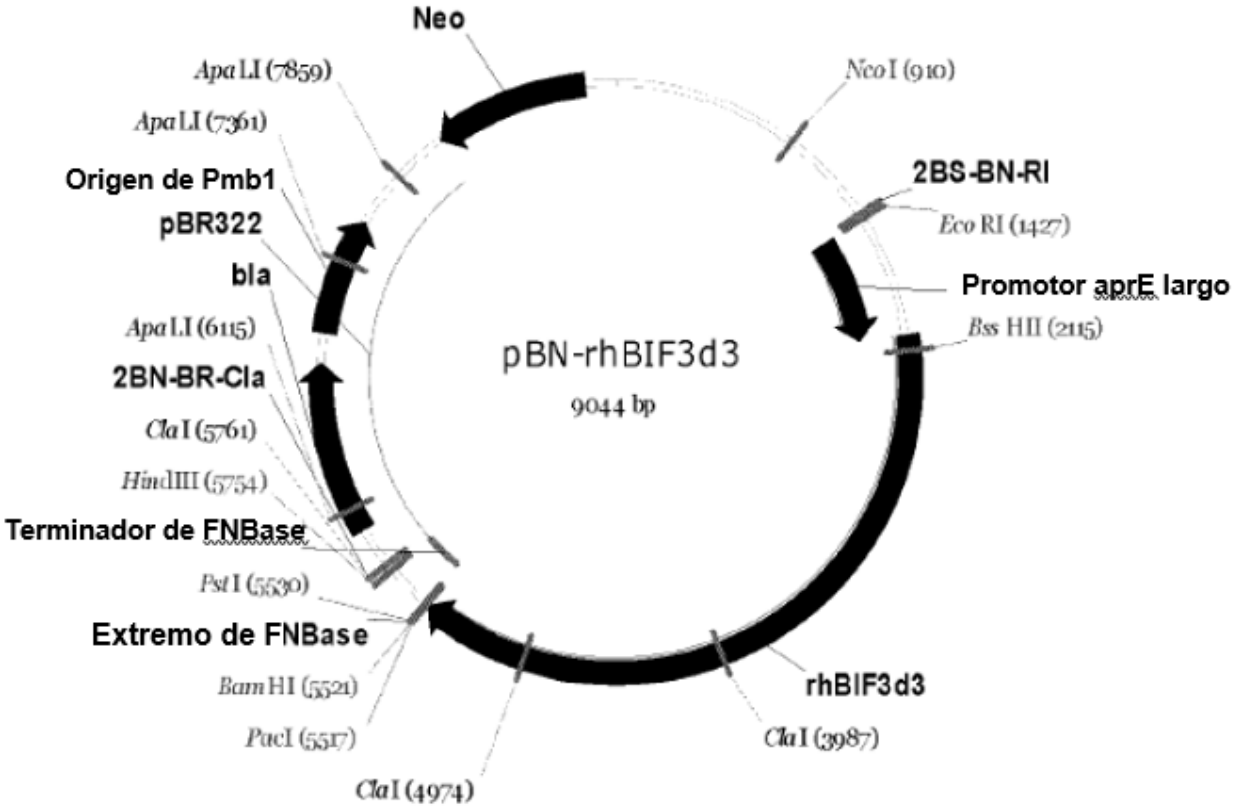
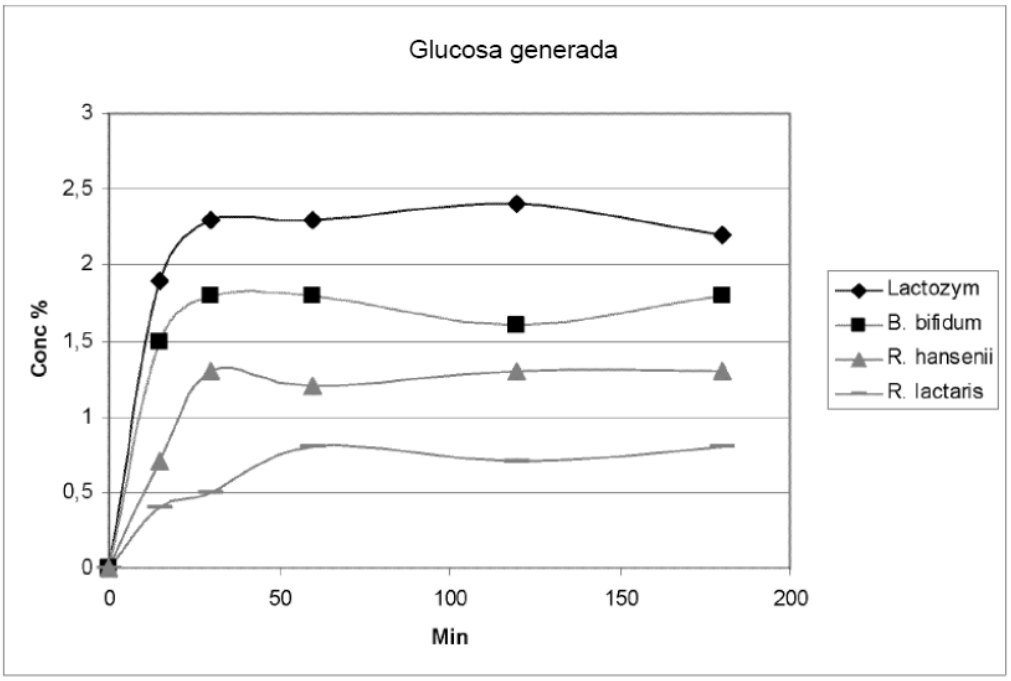

Neo
Apa LI (7859)
Apa LI (7361)
Origen de Pmb1
pBR322
bla
Apa LI (6115)
2BN-BR-Cla
Cla I (5761)
HindIII (5754)
Terminador de FNBase
Pst I (5530)
Extremo de FNBase
Bam HI (5521)
Pac I (5517)
Cla I (4974)
Nco I (910)
2BS-BN-RI
Eco RI (1427)
Promotor aprE largo
Bss HII (2115)
pBN-rhBIF3d3
9044 bp
rhBIF3d3
Cla I (3987)

Fig. 2


Glucosa generada
Conc %
Min
0
0,5
1
1,5
2
2,5
3
0
50
100
150
200
Lactozym
B. bifidum
R. hansenii
R. lactaris

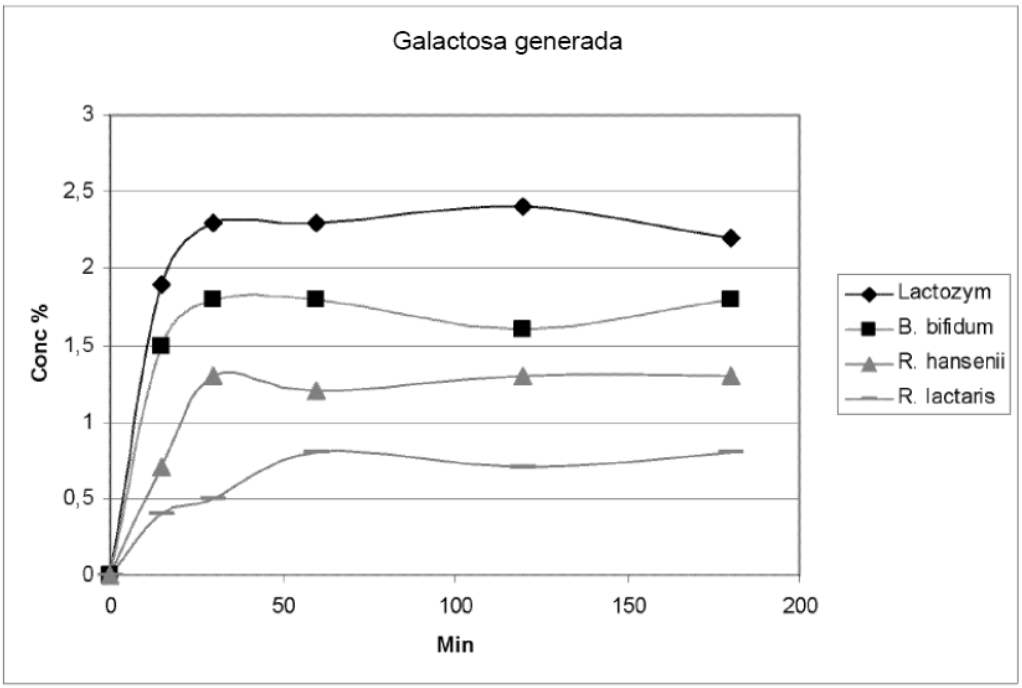

Galactosa generada
Conc %
Min
0
0,5
1
1,5
2
2,5
3
0
50
100
150
200
Lactozym
B. bifidum
R. hansenii
R. lactaris

Figura 3

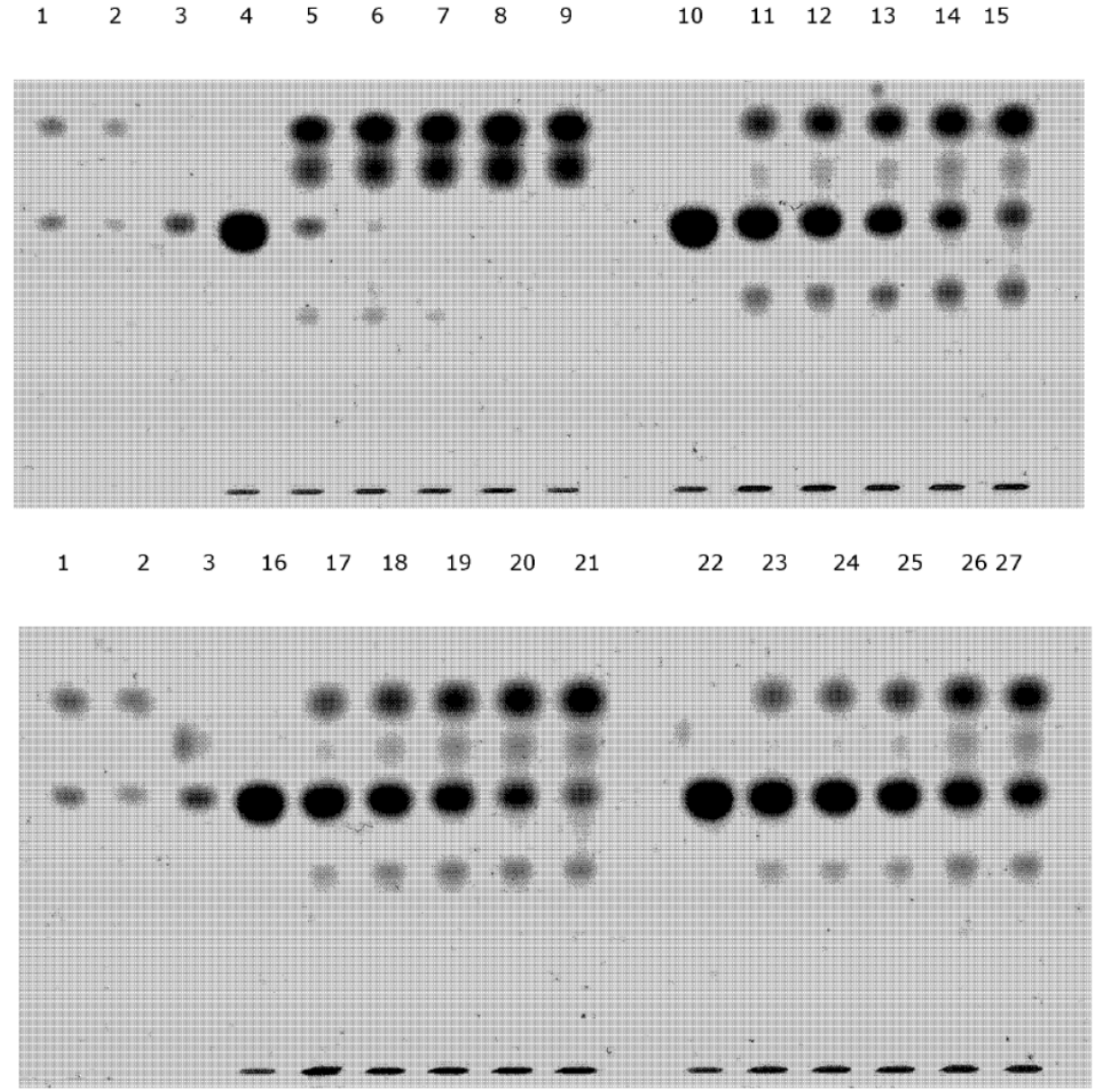
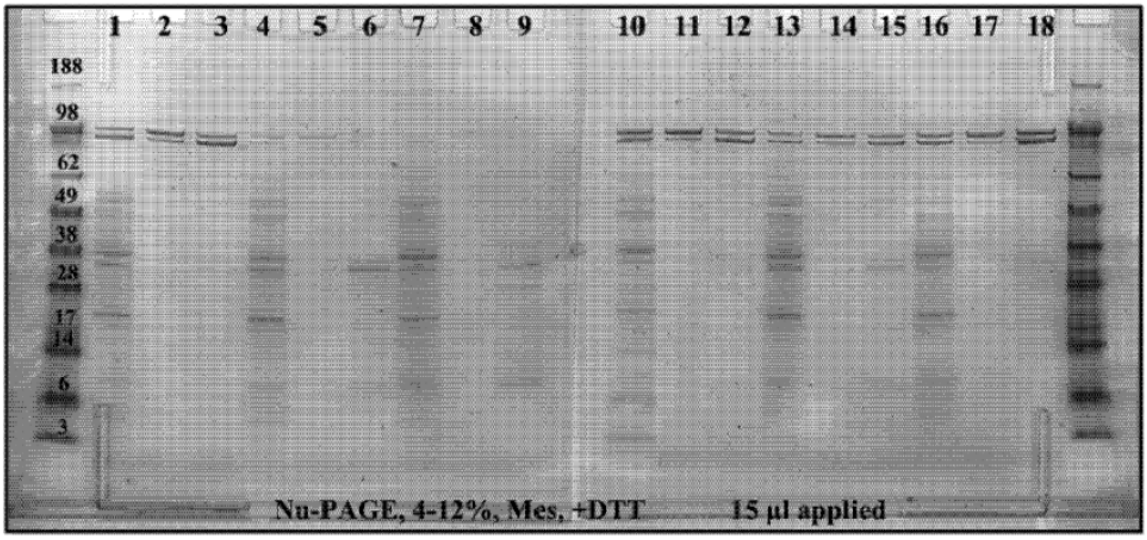


Figura 4


1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18
188
98
62
49
38
28
17
14
6
3
Nu-PAGE, 4-12%, Mes, +DTT
15 µl applied