



19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA

11 Número de publicación: **2 366 975**

51 Int. Cl.:

**C12N 9/10** (2006.01)

**C12N 15/54** (2006.01)

**C12N 15/82** (2006.01)

**A01H 5/00** (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Número de solicitud europea: **07818545 .1**

96 Fecha de presentación : **28.09.2007**

97 Número de publicación de la solicitud: **2066790**

97 Fecha de publicación de la solicitud: **10.06.2009**

54

Título: **Galactosiltransferasa.**

30

Prioridad: **29.09.2006 EP 06450139**

45

Fecha de publicación de la mención BOPI:  
**27.10.2011**

45

Fecha de la publicación del folleto de la patente:  
**27.10.2011**

73

Titular/es: **GREENOVATION BIOTECH GmbH**  
**Böttingerstrasse 29B**  
**79111 Freiburg, DE**

72

Inventor/es: **Launhardt, Heike;**  
**Stemmer, Christian;**  
**Jost, Wolfgang;**  
**Gorr, Gilbert;**  
**Reski, Ralf y**  
**Rensing, Stefan**

74

Agente: **Curell Aguilá, Marcelino**

ES 2 366 975 T3

Aviso: En el plazo de nueve meses a contar desde la fecha de publicación en el Boletín europeo de patentes, de la mención de concesión de la patente europea, cualquier persona podrá oponerse ante la Oficina Europea de Patentes a la patente concedida. La oposición deberá formularse por escrito y estar motivada; sólo se considerará como formulada una vez que se haya realizado el pago de la tasa de oposición (art. 99.1 del Convenio sobre concesión de Patentes Europeas).

## DESCRIPCIÓN

Galactosiltransferasa.

5 La presente invención se refiere a polinucleótidos que codifican glucosiltransferasas. Además, la presente invención se refiere a polinucleótidos parciales de los mismos así como a vectores que comprenden estos polinucleótidos para la expresión o la destrucción génica de los mismos, a células hospedadoras recombinantes, tejido u organismos transfectedados con los polinucleótidos o partes de los mismos o a ADN procedente de los mismos, así como a las glucoproteínas producidas en estas células hospedadoras, tejido u organismos. Además, la presente invención se refiere a la utilización del producto de expresión del mismo *in vitro* así como *in vivo*.

En el pasado, se han producido proteínas heterólogas utilizando una variedad de sistemas celulares transformados, tales como los procedentes de bacterias, hongos, tales como estirpes celulares de levadura, insectos, plantas o mamíferos.

15 Las proteínas producidas en organismos procarióticos pueden no modificarse después de la traducción de una manera similar a la de las proteínas eucarióticas producidas en sistemas eucarióticos, por ejemplo pueden no glucosilarse con azúcares apropiados en particular restos de aminoácidos, tales como restos (N) de ácido aspártico (N unido por glucosilación). Además, el plegamiento de las proteínas eucarióticas producidas por bacterias puede ser inapropiado debido a, por ejemplo, la incapacidad de la bacteria para formar puentes disulfuro de cisteína. Además, las proteínas recombinantes producidas por bacterias frecuentemente se agregan y se acumulan como cuerpos de inclusión insolubles.

25 Los sistemas de células eucarióticas son más adecuados para la producción de proteínas glucosiladas encontradas en varios organismos eucarióticos, tales como seres humanos, ya que dichos sistemas celulares pueden efectuar modificaciones después de la traducción, tales como la N-glucosilación de proteínas producidas. Sin embargo, un problema encontrado en los sistemas de células eucarióticas que han sido transformadas con genes heterólogos adecuados para la producción de secuencias de proteínas destinadas para la utilización, por ejemplo, como productos farmacéuticos, consiste en que el modelo de glucosilación de dichas proteínas con frecuencia adquiere un modelo natural, es decir, del sistema de células eucarióticas en el que la proteína se ha producido: se producen proteínas glucosiladas que comprenden modelos no animales de glucosilación y éstas a su vez pueden ser inmunógenas y/o alérgicas si se aplican en animales, incluyendo los seres humanos. En las plantas esta limitación ha sido superada por la eliminación de los restos de azúcar 1,2-xilosa y  $\alpha$ -1,3 fucosa específicos para las plantas que en las plantas están generalmente unidos a la estructura nuclear de los N-glucanos (Lerouge *et al.* 1998 *Plant Mol. Biol.* 38, 31-48; Rayon *et al.* 1998 *J. Experimental Bot.* 49, 1463-1472). En el caso de *Arabidopsis thaliana* (Strasser *et al.* 2004 *FEBS Lett.* 561, 132-136) y en el caso de la briofita *Physcomitrella patens* (patente EP1431394; Koprivova *et al.* 2004 *Plant Biotechnol. J.* 2, 517-523) se generaron mutantes que presentan modelos de N-glucano que carecen completamente de restos nucleares de  $\alpha$ -1,3 fucosa y 1,2-xilosa. Sorprendentemente, a pesar de la modificación del modelo del complejo tipo N-glucanos no se observaron alteraciones o cambios morfológicos en la viabilidad en estos mutantes.

45 Aparte de la adición de los dos restos específicos de la planta descritos anteriormente las etapas de maduración de la glucoproteína en la ER y en el cis-Golgi son idénticas en los vegetales y en los mamíferos hasta la acción de GlcNAc-transferasa I, GlcNAc-transferasa II y Golgi  $\alpha$ -manosidasa (Lerouge *et al.* 1998 *Plant Mol. Biol.* 38, 31-48). Además el alargamiento del N-glucano se realiza de manera diferente en los dos reinos. Aunque en los mamíferos los restos terminales de GlcNAc están inmediatamente protegidos por la acción de  $\beta$ 1,4- (o raras veces,  $\beta$ 1,3-)-galactosiltransferasa - con la notable excepción de IgG donde esta etapa se produce sólo parcialmente - el alargamiento en las plantas es exclusivamente por  $\beta$ 1,3-galactosilación pero sólo una muy pequeña parte de los glucanos parecen experimentar esta modificación como puede deducirse de la abundancia relativa de varios tipos estructurales. Los restos de galactosa en mamíferos pueden protegerse por el ácido siálico y sólo muy raramente sustituirse por la fucosa. De nuevo, las plantas son diferentes ya que están desprovistas de sialilación y en el caso que un resto terminal de galactosa unido en 1,3 esté acoplado esencialmente siempre fucosilan el penúltimo resto de GlcNAc, formando de este modo un determinante de Lewis a (LeA). Aparentemente, la  $\beta$ 1,3-galactosiltransferasa es la enzima de restricción mientras que la mayoría de las células vegetales contienen suficiente actividad de  $\alpha$ 1,4-Fuc-transferasa para asegurar que cada antena que contiene Gal se fucosila. La estructura de LeA es un determinante del grupo sanguíneo humano. Es raro como tal en los adultos sanos pero como sialil-Lewis a (sLeA) se encuentra principalmente en los tejidos malignos tal como en el cáncer de colon.

60 De todos modos, las glucoproteínas que contienen LeA están raramente aisladas en las plantas y en el caso de *Physcomitrella* presentan una cantidad de sólo hasta el cinco por ciento de las glucoproteínas totalmente solubles, independientemente de si se aíslan de las plantas naturales o se aíslan de mutantes modificados genéticamente por gluco que carecen de fucosa y xilosa del núcleo (Koprivova *et al.* 2003 *Plant Biol.* 5, 582-591; Koprivova *et al.* 2004 *Plant Biotechnol. J.* 2, 517-523).

65 Mientras que algunas investigaciones se realizaron en relación con la  $\alpha$ 1,4-flucosiltransferasa que está implicada en la generación de estructuras de glucano de tipo Lewis a en las plantas (Joly *et al.* 2002 *J. Experimental Bot.* 53,

1429-1436; Bakker *et al.* FEBS Lett. 507, 307-312) no existe ninguna información disponible con respecto a una  $\beta$ 1,3 galactosiltransferasa específica que está implicada en el alargamiento de las estructuras de N-glucano en las plantas.

5 En eucariotas las  $\beta$ 1,3-galactosiltransferasas presentan un amplio espectro de especificidades de aceptor así como distintos modelos de expresión tisular (Hennet 2002 *Cell. Mol. Life Sci.* 59, 1081-1095; Amado *et al.* 1998 *J. Biol. Chem.* 21, 12770-12778). Entre los diferentes miembros de la familia de la  $\beta$ 1,3-galactosiltransferasa de seres humanos para la  $\beta$ 1,3-galactosiltransferasa 2 se ha demostrado *in vitro* que esta enzima era activa en la transferencia de los restos de galactosa para las estructuras de N-glucano de tipo complejo que representan GlcNAc $\beta$  y ovoalbúmina de huevo como sustratos aceptores (Amado *et al.* 1998 *J. Biol. Chem.* 21, 12770-12778).

15 Según la existencia de una familia de  $\beta$ 1,3-galactosiltransferasas homólogas en seres humanos los análisis de la base de datos pusieron de manifiesto que en diferentes especies vegetales por ejemplo *Arabidopsis thaliana* y *Oryza sativa* existen grandes familias de genes similares de genes de  $\beta$ 1,3-galactosiltransferasa. Ninguno de los miembros de estos genes de  $\beta$ 1,3-galactosiltransferasa se describe que codifican una enzima que comprenden la capacidad de transferir la galactosa desde la UDP-galactosa a sustratos aceptores con restos terminales de GlcNAc no reductores por ejemplo a restos terminales no reductores del complejo de tipo N-glucanos ni *in vitro* ni *in vivo*.

20 Dunaeva *et al.* (*Eur. J. Biochem.* 2001 (268): 5521-5529) describe un despliegue diferenciado del ARN de *Arabidopsis thaliana* durante las condiciones de estrés en las que se ha identificado una proteína ("LSR5") que tiene una determinada similitud con una  $\beta$ 1,3-galactosiltransferasa en seres humanos.

El documento WO 2004/035798 describe genes que son aumentan o disminuyen en plantas transgénicas, en particular en *Arabidopsis thaliana*.

25 Decker *et al.* (*Current Opinion in Plant Biology* 2004(7):166-170) describe epítomos alérgenos en proteínas vegetales glucosiladas, en particular habida cuenta de los productos  $\alpha$ -1,3-fucosiltransferasa y  $\beta$ -1,2-xilosiltransferasa.

30 Un objetivo de la presente invención consiste en identificar, clonar y secuenciar uno o más genes - incluyendo las secuencias genómicas correspondientes no codificadoras - que codifican  $\beta$ 1,3-galactosiltransferasas vegetales, y preparar vectores que comprenden los genes, fragmentos del ADN de los mismos o de un ADN alterado o un ADN procedente de los mismos o del ADN que comprende eliminaciones de los mismos. Un objetivo adicional consiste en generar células hospedadoras, tejido y organismos que contienen uno o más de estos vectores, para producir glucoproteínas que carecen completamente de estructuras de N-glucano de tipo Lewis a. Un objetivo adicional consiste en generar células hospedadoras, tejido u organismos que comprenden uno o más de estos vectores, para producir glucoproteínas con estructuras mejoradas de N-glucano de tipo Lewis a. Un objetivo adicional consiste en proporcionar secuencias nucleotídicas que codifican dominios membrenarios para dirigir enzimas para las cisternas de Golgi recientes.

40 Por consiguiente, la presente invención suministra moléculas de ADN que contienen una secuencia según la SEC. ID. nº 1 con un marco de lectura abierto desde el par de bases 513 al par de bases 2417 o una secuencia según la SEC. ID. nº 24 con un marco de lectura abierto desde el par de bases 321 al par de bases 2387, o tienen por lo menos 80% de identidad con por lo menos una de las secuencias completas anteriores, o comprenden una secuencia que se degenera a las secuencias anteriores debido al código genético, teniendo las secuencias que codifican proteínas vegetales actividad de  $\beta$ 1,3-galactosiltransferasa (actividad de  $\beta$ 1,3-GalT) o siendo complementarias de ésta. Se da a conocer también i) una molécula de ADN que contiene una secuencia según la SEC. ID. nº 1 que tiene un marco de lectura abierto desde el par de bases 513 al par de bases 2417 o que tiene por lo menos 50% de identidad con la secuencia mencionada anteriormente o que comprende una secuencia que se ha degenerado a la secuencia de ADN anterior debido al código genético, codificando la secuencia una proteína vegetal que tiene actividad de  $\beta$ 1,3-galactosiltransferasa o es complementaria de ésta, ii) una molécula de ADN que contiene una secuencia según la SEC. ID. nº 2 que tiene un marco de lectura abierto desde el par de bases 1 al par de bases 1902 o que tiene por lo menos 50% de identidad con la secuencia mencionada anteriormente o que comprende una secuencia que se ha degenerado a la secuencia de ADN anterior debido al código genético, codificando la secuencia una proteína vegetal que tiene actividad de  $\beta$ 1,3-galactosiltransferasa o es complementaria de ésta, iii) una molécula de ADN que contiene una secuencia según la SEC. ID. nº 24 que tiene un marco de lectura abierto desde el par de bases 321 al par de bases 2387 o que tiene por lo menos 50% de identidad con la secuencia mencionada anteriormente o que comprende una secuencia que se ha degenerado a la secuencia de ADN anterior debido al código genético, codificando la secuencia una proteína vegetal que tiene actividad de  $\beta$ 1,3-galactosiltransferasa o es complementaria de ésta, iv) una molécula de ADN que contiene una secuencia según la SEC. ID. nº 25 que tiene un marco de lectura abierto desde el par de bases 1 al par de bases 2052 o que tiene por lo menos 50% de identidad con la secuencia mencionada anteriormente o que comprende una secuencia que se ha degenerado a la secuencia de ADN anterior debido al código genético, codificando la secuencia una proteína vegetal que tiene actividad de  $\beta$ 1,3-galactosiltransferasa o es complementaria de ésta, v) una molécula de ADN que contiene una secuencia según la SEC. ID. nº 3 que representa la estructura del ADN genómico desde el par de bases 1 al par de bases 6187 incluyendo las secuencias de intrón y las secuencias de exón correspondientes a las SEC. ID. nº 1 que permiten la generación de montajes modificados genéticamente con secuencias genómicas, vi)

una molécula de ADN que contiene una secuencia según la SEC. ID. nº 4 que representa la estructura del ADN genómico desde el par de bases 1 al par de bases 4087 incluyendo las secuencias de intrón y las secuencias de exón correspondientes a las SEC. ID. nº 2 que permiten la generación de montajes modificados genéticamente con secuencias genómicas.

5 Como la familia de glucosiltransferasas es muy divergente (figura 1) y solamente las regiones conservadas (en  
 10 negrilla en la figura 1) son muy similares, la presente memoria da a conocer también una molécula de ADN que  
 15 contiene una secuencia que tiene por lo menos 20% de identidad total con una secuencia según una cualquiera de  
 20 las SEC. ID. nº 1, SEC. ID. nº 2, SEC. ID. nº 24 o SEC. ID. nº 25 y que tiene por lo menos el 80% de identidad con  
 25 una secuencia de los siete dominios conservados de SEC. ID. nº 1 o la SEC. ID. nº 2 que codifican los aminoácidos  
 387-392 (DLFIGI o ELFVGI), 402-409 (RMAVRKTW), 425-428 (FVAL), 455-465 (DRYDIVVLKTV), 479-489  
 (YIMKDDDTFV o HVMKDDDTFV), 536-548 (YPIYANGPGYILS o YPTYANG-PGYILS) y 570-576 (EDVSVGI) de  
 la proteína SEC. ID. nº 19 o las SEC. ID. nº 20, o que comprende una secuencia que se degenera a la secuencia  
 anterior debido al código genético, teniendo actividad de  $\beta$ 1,3-galactosiltransferasa la secuencia que codifica  
 proteínas vegetales o siendo complementaria de ésta. También se da a conocer la molécula de ADN que comprende  
 una secuencia que tiene por lo menos el 20% de identidad total con una secuencia según una cualquiera de SEC.  
 ID. nº 1, SEC. ID. nº 2, SEC. ID. nº 24 o SEC. ID. nº 25 y que codifica por lo menos el 95%, preferentemente todos,  
 los aminoácidos conservados de los siete dominios conservados de las SEC. ID. nº 1 o SEC. ID. nº 2 seleccionadas  
 de entre los aminoácidos 388 (L), 402 (R), 404 (A), 406 (R), 408 (T), 409 (W), 425 (F), 455 (D), 457 (Y), 463 (K), 464  
 (T), 481 (M), 482 (K), 484 (D), 486 (D), 488 (F), 489 (V), 536 (Y), 537 (P), 542 (G), 544 (G), 545 (Y), 548 (S), 570 (E),  
 571 (D), 572 (V), 575 (G) y 576 (I) de la proteína de la SEC. ID. nº 19 o la SEC. ID. nº 20, o que comprende una  
 secuencia que se degenera a la secuencia anterior debido al código genético, teniendo la secuencia de codificación  
 de las proteínas vegetales actividad de  $\beta$ 1,3-galactosiltransferasa o siendo complementaria de ésta. Se dan a  
 conocer las identidades de secuencias totales de por lo menos 25%, por lo menos 30%, por lo menos 35%, por lo  
 menos 40%, o por lo menos 45%. La identidad de la secuencia para los dominios conservados puede ser por lo  
 menos del 90%, por lo menos del 95% o del 100%.

El marco de lectura abierto de la secuencia que tiene los códigos de la SEC. ID. nº 1 para una proteína con 634  
 30 aminoácidos (figura 2, SEC. ID. nº 19). La proteína codificada por la SEC. ID. nº 1 contiene un dominio  
 transmembranario en la región entre Leu20 y Leu39, y contiene los siete dominios conservados - presentes en las  
 $\beta$ 1,3-galactosiltransferasas humanas - descritas por Hennet (2002 *Cell. Mol. Life Sci.* 59, 1081-1095, figura 2) así  
 como la mayoría de los aminoácidos conservados situados en el terminal C descritos por Amado *et al.* (1988 *J. Biol.*  
*Chem.* 21, 12770-12778, figura 2).

El marco de lectura abierto de la secuencia que tiene las SEC. ID. nº 2 codifica una proteína con 633 aminoácidos  
 35 (figura 3, SEC. ID. nº 20). La proteína codificada por las SEC. ID. nº 2 contiene un dominio transmembranario en la  
 región entre Leu20 y Leu39, y codifica los siete dominios conservados - presentes en las  $\beta$ 1,3-galactosiltransferasas  
 humanas - descritas por Hennet (2002 *Cell. Mol. Life Sci.* 59, 1081-1095, figura 2) así como la mayoría de los  
 40 aminoácidos conservados situados en el terminal C descritos por Amado *et al.* (1988 *J. Biol. Chem.* 21, 12770-  
 12778, figura 2).

El marco de lectura abierto de la secuencia que tiene la SEC. ID. nº 24 codifica una proteína con 688 aminoácidos  
 (figura 4; SEC. ID. nº 26), que es una variante de corte y empalme alternativa a la proteína SEC. ID. nº 1.

El marco de lectura abierto de la secuencia que tiene la SEC. ID. nº 25 codifica una proteína con 683 aminoácidos  
 45 (figura 5; SEC. ID. nº 27), que es una variante de corte y empalme alternativa a la proteína SEC. ID. nº 2.

La presente memoria describe también las secuencias genómicas de este gen tal como el suministrado por la SEC.  
 50 ID. nº 3 o nº 4, desde luego, como todas las demás moléculas de ADN o proteínas según la presente invención (si  
 no se describe explícitamente de otra manera) en forma aislada.

La actividad de las  $\beta$ 1,3-galactosiltransferasas vegetales puede analizarse por diferentes métodos.

Según Amado *et al.* (1988 *J. Biol. Chem.* 21, 12770-12778) los montajes que codifican las formas solubles  
 55 segregadas - que carecen de dominio transmembranario - de las  $\beta$ 1,3-galactosiltransferasas pueden clonarse en  
 vectores de expresión por ejemplo apropiados para la transfección de Baculovirus y ampliarse en células Sf9; en los  
 productos de expresión resultantes pueden purificarse y analizarse posteriormente la actividad de  $\beta$ 1,3-  
 galactosiltransferasas.

Otro enfoque debido a los análisis de la actividad específica puede ser sobreexpresión de las  $\beta$ 1,3-  
 60 galactosiltransferasas en un hospedador apropiado por ejemplo como *Physcomitrella patens* preparando montajes  
 de expresión diseñados para codificar los marcos de lectura abierta completos de las  $\beta$ 1,3-galactosiltransferasas  
 según la presente invención y mediante la generación de cepas transgénicas de *Physcomitrella* por lo menos para  
 uno de los genes de  $\beta$ 1,3-galactosiltransferasas según la presente invención. Las cepas transgénicas generadas  
 65 presentan contenidos mejorados de N-glucanos galactosilados. Los modelos de N-glucano procedentes de  
*Physcomitrella* pueden aislarse y analizarse como describe Koprivova *et al.* (2003 *Plant Biol.* 5, 582-591) y Koprivova

*et al.* (2004 *Plant Biotechnol. J.* 2, 517-523).

Las actividades de  $\beta$ 1,3-galactosiltransferasas según la presente invención pueden analizarse indirectamente mediante la destrucción dirigida de los genes responsables en un hospedador apropiado por ejemplo *Physcomitrella patens* que produce la inhibición de las actividades de  $\beta$ 1,3-galactosiltransferasas en relación con la transferencia de la galactosa desde la UDP-galactosa a los restos de GlcNAc del terminal no reductor en N-glucanos y por consiguiente con la pérdida de la galactosilación terminal. De nuevo, los modelos N-glucano de *Physcomitrella* pueden aislarse y analizarse como describe Koprivova *et al.* (2003 *Plant Biol.* 5, 582-591) y Koprivova *et al.* (2004 *Plant Biotechnol. J.* 2, 517-523). Preferentemente, la  $\beta$ 1,3-galactosiltransferasa según la presente invención es una GlcNAc- $\beta$ 1,3-galactosiltransferasa.

Alternativamente la reducción de la actividad de  $\beta$ 1,3-galactosiltransferasa puede conseguirse por procedimientos que se utilizan corrientemente para este finalidad por ejemplo la estrategia complementaria bien conocida, la estrategia transcrita, la tecnología de ribozimas, la tecnología del APN o la estrategia de la interferencia del ARN.

Según la presente invención una célula hospedadora, tejido u organismo se transfecta con las secuencias nucleotídicas que comprenden al menos las secuencias SEC. ID. nº 1 o SEC. ID. nº 24 que codifican una  $\beta$ 1,3-galactosiltransferasa. En una forma de realización preferida de la presente invención las secuencias de codificación están unidas a secuencias reguladoras tales como secuencias activadoras y de terminación que permiten la expresión de los genes de  $\beta$ 1,3-galactosiltransferasa que producen los productos de expresión que presentan actividades de  $\beta$ 1,3-galactosiltransferasa. En cuanto al tejido u organismo de la célula hospedadora las secuencias reguladoras operativamente conectadas a la secuencia de codificación de  $\beta$ 1,3-galactosiltransferasa pueden ser homólogas debido al hospedador utilizado. Las secuencias reguladoras operativamente conectadas a la secuencia de codificación de  $\beta$ 1,3-galactosiltransferasa pueden ser heterólogas. En otra forma de realización, las secuencias reguladoras operativamente conectadas a la secuencia de codificación de  $\beta$ 1,3-galactosiltransferasa pueden ser suministradas por el vector utilizado para la transfección o pueden crearse *in vivo* introduciendo la secuencia de codificación de  $\beta$ 1,3-galactosiltransferasa por integración dirigida por ejemplo recombinación homóloga en un locus apropiado resultante en un ensamble operativamente funcional de la secuencia de codificación de  $\beta$ 1,3-galactosiltransferasa con las secuencias reguladoras endógenas de la célula hospedadora, tejido u organismo.

En una forma de realización preferida de la presente invención la expresión producto o partes del mismo por ejemplo una forma soluble que carece de dominios transmembranarios que comprende  $\beta$ 1,3-galactosiltransferasa puede utilizarse para el alargamiento de N-glucanos en glucolípidos o glucoproteínas *in vitro* o *in vivo*. En una forma de realización más los N-glucanos resultantes que comprenden restos de galactosa terminales unidos en 1,3 pueden alargarse más *in vitro* o *in vivo* con restos de azúcar adicionales, tales como restos de fucosa, galactosa o ácido siálico. Por consiguiente, la presente invención se refiere a nuevas glucoproteínas con estructura de azúcar de N-glucanos que comprende el complejo tipo N-glucanos que contiene restos de azúcar terminales, tal como galactosa, una fucosa adicional, ácido siálico o combinaciones de los mismos. En una forma de realización más preferida, estas glucoproteínas son proteínas superficiales que presentan el complejo de tipo N-glucano al medio externo de la célula, por ejemplo permitiendo contactos proteína/proteína (tales como contactos con anticuerpos, otras células, etc.) o proteínas secretoras, por ejemplo anticuerpos o eritropoyetina. Dichas glucoproteínas producidas según la presente invención son muy adecuadas para la vacunación, especialmente de personas, tanto *in vitro* como *in vivo*.

Se dan a conocer secuencias nucleotídicas según la SEC. ID. nº 1, SEC. ID. nº 2, SEC. ID. nº 24 o SEC. ID. nº 25 que codifican dominios transmembranarios para dirigir una proteína heteróloga a las cisternas de Golgi recientes. En una forma de realización preferida las  $\beta$ -1,4-galactosiltransferasas o las sialiltransferasas que presentan actividad para el alargamiento de N-glucanos se dirigen a las cisternas de Golgi recientes por intercambio de los dominios transmembranarios naturales con éstas de las  $\beta$ 1,3-galactosiltransferasas según la presente invención.

Según la presente invención se suministra una célula hospedadora transformada que comprende por lo menos una secuencia nucleotídica disfuncional de  $\beta$ 1,3-galactosiltransferasa.

En una forma de realización preferida de la invención la célula hospedadora se selecciona de entre plantas, por ejemplo de la especie *Lemna*, de la especie *Wolffia*, de las especies de arroz, zanahoria, trigo, maíz y tabaco. En una forma de realización más preferida de la invención la célula hospedadora se selecciona de briofitas incluyendo musgos y hepáticas, de especies de los géneros *Physcomitrella*, *Funaria*, *Sphagnum*, *Ceratodon*, *Marchantia* y *Sphaerocarpos*. La célula de briofita es preferentemente de *Physcomitrella patens*.

Un hospedador preferido según la presente invención es una briofita, especialmente *Physcomitrella patens*, una planta terrestre haploide no vascular, puede utilizarse para la producción de proteínas recombinantes gluco-transgénicas (documento WO 01/25456). Se han detectado en *Physcomitrella patens* así como en otras estructuras vegetales tipo a de Lewis (Koprivova *et al.* 2003 *Plant Biol.* 5, 582-591; Koprivova *et al.* 2004 *Plant Biotechnol. J.* 2, 517-523). Aunque en las plantas no se ha identificado nada de  $\beta$ 1,3-galactosiltransferasa que presente actividad específica en la prolongación de estructuras de N-glucano se seleccionó *Physcomitrella* como una supuesta fuente de esta clase desconocida de gluco-siltransferasa.

El ciclo vital de los musgos está dominado por la generación fotoautótrofa de gameto ófítico. El ciclo vital es completamente diferente al de las plantas superiores en las que el esporofito es la generación dominante y existen notablemente muchas diferencias que deben observarse entre las plantas superiores y las briofitas.

5 El gametofito de las briofitas incluyendo los musgos se caracteriza por dos fases de desarrollo distintas. El protonema que se desarrolla por crecimiento apical, crece en una red filamentosa de sólo dos tipos de células (células de cloronema y caulonema). La segunda fase, denominada gametófora, se diferencia por el crecimiento caulinario de un sistema apical sencillo. Ambas fases son fotoautótrofas activas. El cultivo del protonema sin diferenciación en el gametóforo más complejo se ha demostrado para cultivos en suspensión en matraces así como  
10 para cultivos en biorreactor (documento WO 01/25456). El cultivo de tejido multicelular completamente diferenciado y fotoautótrofo activo que contiene solamente unos pocos tipos de células no está descrito para plantas superiores. La estabilidad genética del sistema celular del musgo proporciona una ventaja importante sobre los cultivos de células vegetales.

15 Existen algunas diferencias importantes entre las briofitas (plantas no vasculares) y las plantas superiores (plantas vasculares) a nivel bioquímico. La asimilación del sulfato en *Physcomitrella patens* difiere significativamente de la de las plantas superiores. La enzima clave de la asimilación del sulfato en las plantas superiores es la adenosina 5'-fosfosulfato reductasa. En *Physcomitrella patens* coexiste una ruta alternativa a través de fosfoadenosina 5'-fosfosulfato reductasa (Koprivova *et al.* (2002) *J. Biol. Chem.* 277, 32195-32201). Esta ruta no se ha caracterizado  
20 en las plantas superiores.

Además, muchos miembros de las familias de briofitas, algas y helechos producen una gran variedad de grasas de ácidos grasos poliinsaturados (Dembitsky (1993) *Prog. Lipid Res.* 32, 281-356). Por ejemplo, el ácido araquidónico y el ácido eicosapentaenoico se cree que son producidos solamente por las plantas inferiores y no por las plantas  
25 superiores. Algunas enzimas del metabolismo de los ácidos grasos poliinsaturados, (grupo delta-6 acil desaturasa) (Girke *et al.* (1998), *Plant J.*, 15, 39-48) y un componente de una delta 6 elongasa (Zank *et al.* (2002) *Plant J.*, 31, 255-268), se han clonado a partir de *Physcomitrella patens*. Se han descubierto genes no correspondientes en las plantas superiores. Este hecho parece confirmar que existen diferencias fundamentales entre las plantas superiores y las plantas inferiores a nivel bioquímico.

30 Por otra parte, las briofitas presentan recombinación homóloga muy eficiente en su ADN nuclear, una característica única de las plantas, lo que permite la destrucción dirigida de genes (Girke *et al.* (1998) *Plant J.*, 15, 39-48; Strepp *et al.* (1998) *Proc. Natl. Acad. Sci. USA* 95: 4368-4373; Koprivova (2002) *J. Biol. Chem.* 277, 32195-32201; revisado por Reski (1999) *Planta* 208, 301-309; Schaefer y Zryd (2001) *Plant Phys.* 127, 1430-1438; Schaefer (2002) *Annu. Rev. Plant Biol.* 53, 477-501; Koprivova *et al.* 2004 *Plant Biotechnol. J.* 2, 517-523; Brucker *et al.* 2005 *Planta* 220, 864-874) que ilustran más diferencias fundamentales con las plantas superiores. Sin embargo, en algunos casos la utilización de este mecanismo para alterar el modelo de glucosilación ha demostrado ser problemática, como se muestra en la presente memoria en los ejemplos. La disgregación de la N-acetilglucosaminiltransferasa I (GNT1) en  
35 *Physcomitrella patens* dio como resultado la pérdida del transcrito específico, pero sólo en pequeñas diferencias del modelo de N-glucosilación. Estos resultados están en contraste directo con la pérdida de glucanos del complejo Golgi modificado en una planta mutante de *Arabidopsis thaliana* que carece de GNT1 observado por von Schaeuwen *et al.* (1993) *Plant Physiol.* 102, 1109-1118). Por lo tanto, la modificación genética en *Physcomitrella patens* no dio como resultado la modificación prevista del modelo N-glucosilación.

45 Aunque la estrategia de la modificación genética no tuvo éxito para la glucosiltransferasa GNT1, en relación con las alteraciones de los genes que codifican la  $\beta$ 1,2-xilosiltransferasa y  $\alpha$ 1,3-galactosiltransferasa se llevaron a cabo con éxito en *Physcomitrella patens*.

Además la integración de la  $\beta$ 1,4-galactosiltransferasa humana en el genoma de una planta *Physcomitrella patens* modificada genéticamente dos veces dio lugar a un modelo de glucosilación de N ligado como en los mamíferos sin los restos fucosil y xilosil específicos de vegetales y con los restos 1,4-galactosilo terminales como en los mamíferos. Se descubrió que la galactosiltransferasa era activa.

50 La célula de briofitas, tal como una célula de *Physcomitrella patens*, puede ser cualquier célula adecuada para la transformación según los procedimientos de la invención como se describe en este documento, y puede ser una célula de protoplasto de musgo, una célula que se encuentra en el tejido de protonema u otro tipo de célula. De hecho, el destinatario experimentado apreciará que la mayor parte del tejido vegetal de musgo que comprende poblaciones de células de briofitas transformadas según la invención, tal como el tejido de las, como el tejido de protonema transformado constituye un aspecto de la presente invención.

60 El término "disfuncional", tal como se utiliza en la presente memoria, significa que las secuencias de nucleotídicas nombradas de transferasa de  $\beta$ 1,3-galactosiltransferasa ( $\beta$ 1,3-GalT) son sustancialmente incapaces de codificar ARNm que codifica las proteínas funcionales  $\beta$ 1,3-GalT que son capaces de modificar los glucanos unidos a N con restos terminales de galactosa unidos por 1,3. En un prefermento, las secuencias nucleotídicas de la transferasa vegetal  $\beta$ 1,3-GalT disfuncional comprenden inserciones dirigidas de secuencias nucleotídicas exógenas en endógenas, es decir, genómicas, genes de  $\beta$ 1,3-GalT naturales comprendidos en el genoma nuclear de briofitas (si  
65

se trata de un genoma de briofitas verdaderamente natural, que se encuentra en las células de briofitas que no han sido transformadas con anterioridad por el hombre con otras secuencias de ácido nucleico, o en un genoma nuclear transformado de briofitas en el que las inserciones de la secuencia de ácido nucleico se han realizado previamente de las secuencias deseadas de ácido nucleico), que inhibe o reprime substancialmente la transcripción de ARNm que codifica la actividad funcional de  $\beta$ 1,3-GalT.

Otro aspecto de la invención se refiere a un vector biológicamente funcional que comprende una de las moléculas de ADN anteriormente indicadas o a las partes de las mismas de diferentes longitudes con al menos 20 pares de bases. Para la transfección en las células hospedadoras, es necesario un vector independiente capaz de ampliación, en el que, en función de la célula hospedadora, del mecanismo de la transfección, de la tarea y el tamaño de la molécula de ADN, se puede utilizar un vector adecuado. Dado que se conoce un gran número de diferentes vectores, una enumeración de los mismos iría más allá de los límites de la presente solicitud y por lo tanto se prescinde aquí, sobre todo porque los vectores son muy bien conocidos por los expertos en la materia (en cuanto a los vectores, así como todas las técnicas y los términos utilizados en esta memoria que son conocidos por los expertos en la materia, véase, también Sambrook Maniatis). En el mejor de los casos, el vector tiene una masa de moléculas pequeñas y debería incluir genes seleccionables a fin de dar lugar a un fenotipo fácilmente reconocible en una célula para así permitir una fácil selección de células hospedadoras que contienen vector y sin vector. Para obtener un alto rendimiento de ADN y de productos génicos correspondientes, el vector debería comprender un activador potente, así como un potenciador, señales de ampliación génica y secuencias reguladoras. Para una replicación autónoma del vector, además, es importante un origen de replicación. Las zonas de poliadenilación son responsables del tratamiento correcto del ARNm y de las señales de corte y empalme para las transcripciones del ARN. Si se utilizan como vectores fagos, virus o partículas víricas, las señales de encapsulación controlarán la encapsulación del vector ADN. Por ejemplo, para la transcripción en vegetales, los plásmidos TI son apropiados, y para la transcripción en células de insecto, baculovirus y en insectos, respectivamente, transposones, tal como el elemento P.

Si el vector de la invención descrito anteriormente se inserta en un vegetal o en una célula vegetal, una supresión tras la transcripción de la expresión génica del gen endógeno de la  $\beta$ 1,3-galactosiltransferasa se alcanza por transcripción de un transgén homólogo a éste o de partes del mismo, en la orientación de la transcripción. Para esta técnica de transcripción, además, se hace referencia a las publicaciones de Baucombe 1996, *Plant. Mol. Biol.*, 9:373-382, y Brigneti *et al.*, 1998, *EMBO J.* 17: 6739-6746. Esta estrategia de "silenciamiento génico" es una manera eficaz de suprimir la expresión del gen de la  $\beta$ 1,3-galactosiltransferasa, véase también Waterhouse *et al.*, 1998, *Proc. Natl. Acad. Sci. USA* 95:13959-13964.

Además, la invención se refiere a un vector biológicamente funcional que comprende una molécula de ADN según una de las modalidades descritas anteriormente, o partes de la misma, de diferentes longitudes en orientación inversa al activador. Si este vector se transfecta en una célula hospedadora, un "ARNm complementario" se leerá que es complementario del ARNm de la  $\beta$ 1,3-galactosiltransferasa y acompleja este último. Este enlace impedirá el correcto tratamiento, el transporte, la estabilidad o, evitando la hibridación de los ribosomas, impedirá la traducción y por lo tanto la expresión normal del gen de la  $\beta$ 1,3-galactosiltransferasa.

A pesar de que toda la secuencia de la molécula de ADN podría insertarse en el vector, las secuencias parciales de la misma debido a su menor tamaño pueden ser mejores para determinados fines. Con el aspecto complementario, por ejemplo, es importante que la molécula de ADN sea lo suficientemente grande como para formar un ARNm complementario suficientemente grande que se una a la ARNm transferasa. Una molécula de ARN complementario adecuada comprende, por ejemplo, de 50 a 200 nucleótidos, ya que muchas de las moléculas de ARN complementario naturales conocidas, contienen aproximadamente 100 nucleótidos.

Para una inhibición particularmente eficaz de la expresión de una  $\beta$ 1,3-galactosiltransferasa activa, es adecuada una combinación de la técnica transcrita y la técnica complementaria (Waterhouse *et al.*, 1998, *Proc. Natl. Acad. Sci., USA.*, 95: 13.959 a 13.964).

De manera ventajosa, se utilizan moléculas de ARN que se hibridan rápidamente. La eficacia de las moléculas de ARN complementario que tienen un tamaño de más de 50 nucleótidos dependerá de la cinética de hibridación *in vitro*. Así, por ejemplo, las moléculas de ARN complementario que se hibridan rápidamente presentan una inhibición mayor de la expresión de la proteína que las moléculas de ARN que se hibridan lentamente (Wagner *et al.*, 1994, *Annu. Rev. Microbiol.*, 48:713-742; Rittner *et al.*, 1993, *Nucl. Acids Res.*, 21:1381-1387). Dichas moléculas de ARN que se hibridan rápidamente comprenden en particular un gran número de bases externas (extremos libres y secuencias de conexión), un gran número de subdominios estructurales (componentes), así como un bajo nivel de los bucles (Patzel *et al.* 1998, *Nature Biotechnology*, 16, 64-68). Las estructuras secundarias hipotéticas de la molécula de ARN complementaria pueden determinarse, por ejemplo, con ayuda de un programa de ordenador, según el cual se selecciona una secuencia de ARN ADN complementario adecuada.

Pueden insertarse diferentes regiones de la secuencia de la molécula de ADN en el vector. Una posibilidad consiste, por ejemplo, en insertar en el vector sólo la parte que es responsable de la hibridación de ribosomas. El bloqueo en esta región del ARNm bastará para interrumpir toda la traducción. Una eficacia particularmente elevada de las moléculas complementarias también resulta para las regiones 5' y 3' no traducidas del gen.

Preferiblemente, la molécula de ADN según la invención incluye una secuencia que comprende una mutación por eliminación, inserción y/o sustitución. El número de nucleótidos mutante es variable y varía de uno solo a varios nucleótidos eliminados, insertados o sustituidos. También es posible que el marco de lectura se desplace por la mutación. En dicho "gen modificado genéticamente" es meramente importante que la expresión de una  $\beta$ 1,3-galactosiltransferasa se altere, y se evita la formación de una enzima funcional activa. De este modo, el sitio de la mutación es variable, siempre y cuando se evite la expresión de una proteína enzimáticamente activa. Preferentemente, la mutación está en la zona catalítica de la enzima que se encuentra en la zona C-terminal. Los procedimientos de inserción de mutaciones en las secuencias de ADN son muy conocidos por los expertos en la materia, y por lo tanto, las distintas posibilidades de mutagenia no necesitan ser expuestas en la presente memoria con detalle. Puede emplearse en este caso la mutagenia fortuita, así como, en particular mutagenia dirigida, por ejemplo, la mutagenia dirigida, mutagenia controlada por oligonucleótidos o mutagenia con ayuda de enzimas de restricción.

Por otra parte, las técnicas de ribozima o ARNsi se pueden aplicar para reducir o eliminar la actividad de  $\beta$ 1,3-GalT en las células que tienen actividad de  $\beta$ 1,3-GalT. La adaptación de las técnicas de ARNsi a la presente invención es sencilla basándose en la experiencia existente en la materia (por ejemplo, Nat. Reviews: RNA interference collection (octubre de 2005)).

La invención proporciona además una molécula de ADN que codifica una ribozima que se compone de dos porciones de secuencia de por lo menos 10 a 15 pares de bases cada una, que son complementarias de las porciones de la secuencia de una molécula de ADN de la invención como se describió anteriormente, para los complejos de ribozima y escinde el ARNm que se transcribe a partir de una molécula de ADN  $\beta$ 1,3-galactosiltransferasa natural. La ribozima reconocerá el ARNm de la  $\beta$ 1,3-galactosiltransferasa, por emparejamiento de bases complementarias con el ARNm. Posteriormente, la ribozima escindirán y destruirá el ARN de una manera específica para la secuencia, antes de que se traduzca la enzima. Después de la disociación del sustrato escindido, la ribozima se hibridará en repetidas ocasiones con las moléculas de ARN y actuará como endonucleasa específica. En general, pueden producirse ribozimas específicamente por inactivación de un determinado ARNm, incluso si no se conoce toda la secuencia completa del ADN que codifica la proteína. Las ribozimas son particularmente eficaces si las ribosomas se mueven lentamente a lo largo del ARNm. En ese caso es más fácil que la ribozima encuentre un sitio sin ribosomas en el ARNm. Por esta razón, los mutantes lentos de ribosomas son también adecuados como sistema para ribozimas (J. Burke, 1997, *Nature Biotechnology*, 15, 414- 415). Esta molécula de ADN es particularmente ventajosa para la regulación por disminución y la inhibición, respectivamente, de la expresión de las  $\beta$ 1,3-galactosiltransferasas.

Una forma posible consiste también en utilizar una forma variada de ribozima, es decir, una minizima. Las minizimas son particularmente eficaces para escindir moléculas más grandes de ARNm. Una minizima es una ribozima cabeza de martillo que tiene un enlazador oligonucleotídico corto en lugar del tronco/bucle II. Las dímero-minizimas son particularmente eficaces (Kuwabara *et al.*, 1998, *Nature Biotechnology*, 16.; 961-965).

Por consiguiente, la invención se refiere además a un vector biológicamente funcional que comprende una de las dos moléculas de ADN recién mencionadas (molécula de mutación o ribozima-ADN). Lo que se ha dicho anteriormente con relación a los vectores se aplica también en este caso. Dicho vector puede insertarse, por ejemplo, en un microorganismo y puede utilizarse para la producción de altas concentraciones de las moléculas de ADN descritas anteriormente. Además dicho vector es particularmente adecuado para la inserción de una molécula de ADN específica en un organismo vegetal para reducir o inhibir completamente la producción de  $\beta$ 1,3-galactosiltransferasa en este organismo. Todos los vectores descritos anteriormente pueden prepararse también con secuencias genómicas de genes de  $\beta$ 1,3-GalT, tal como la SEC. ID. n° 3.

Las células de briofitas de la invención o de sus progenitores pueden ser cualquiera que se haya transformado anteriormente con genes heterólogos de interés que codifican las principales secuencias de proteínas de interés que están glucosiladas con los patrones de glucosilación de mamíferos como los descritos en la presente memoria. Preferiblemente, los patrones de glucosilación son de tipo humano. Alternativamente, la célula de briofitas se puede transformar individualmente, es decir, al mismo tiempo o con el tiempo con secuencias nucleotídicas que codifican por lo menos una secuencia primaria de proteínas de interés, por lo general al menos una proteína de interés farmacéutico para su utilización en personas o mamíferos como las especies de ganado, incluidas las especies bovina, ovina, equina y porcina, que requieren patrones de glucosilación de mamíferos para ser colocado en ellas según los métodos de la invención como se describe en la presente memoria. Dichas glucoproteínas farmacéuticas para su utilización en mamíferos, incluido el hombre incluyen, pero no se limitan a proteínas tales como el VEGF, interferones tales como interferón  $\alpha$ , interferón  $\beta$ , interferón gamma, factores de coagulación de la sangre seleccionados de entre el factor VII, VIII, IX, X, XI y XII, hormonas de la fertilidad como la hormona luteinizante, la hormona estimulante del foliculo, factores de crecimiento como el factor de crecimiento epidérmico, factor de crecimiento derivado de plaquetas, factor estimulante de colonias de granulocitos y similares, prolactina, oxitocina, hormona estimulante de la tiroides, hormona adrenocorticotrópica, calcitonina, hormona paratiroidea, somatostatina, eritropoyetina (EPO), enzimas tales como la  $\beta$ -gluco-cerebrosidasa, hemoglobina, colágeno, proteínas de fusión, tales como la proteína de fusión del ligando receptor del TNF $\alpha$  que une el dominio con el fragmento Fc de la IgG y



similares. Por otra parte, el procedimiento de la invención puede utilizarse para la producción de inmunoglobulinas tales como los anticuerpos, tales como los anticuerpos monoclonales específicos o fragmentos activos de los mismos.

5 Información detallada sobre el cultivo de musgos que son adecuados para utilizar en la invención, tales como *Leptobryum pyriforme* y *Sphagnum magellanicum* en biorreactores, se conoce en técnica anterior (véase, por ejemplo, E. Wilbert, "Biotechnological studies concerning the mass culture of mosses with particular consideration of the arachidonic acid metabolism", tesis doctoral, Universidad de Mainz (1991); H. Rudolph y S. Rasmussen, Studies on secondary metabolism of *Sphagnum* cultivated in bioreactors, *Crypt. Bot.*, 3, págs. 67-73 (1992)). Para los fines  
10 de la presente invención se prefiere especialmente la utilización de *Physcomitrella patens*, ya que en este organismo se practican técnicas de biología molecular (para un estudio véase R. Reski, Development, genetics and molecular biology of mosses, *Bot. Acta*, 111, págs. 1-15 (1998)).

15 Se han desarrollado sistemas de transformación adecuados para la explotación biotecnológica de *Physcomitrella* destinados a la producción de proteínas heterólogas. Por ejemplo, se han llevado a cabo con éxito transformaciones por transferencia directa de ADN en tejido de protonema utilizando bombardeo de partículas. La transferencia de ADN mediada por PEG en protoplastos de musgo se ha logrado también con éxito. El procedimiento de transformación mediada por PEG se ha descrito muchas veces para *Physcomitrella patens* y conduce tanto a transformantes temporales como estables (véase, por ejemplo, K. Reutter y R. Reski, Production of a heterologous protein in bioreactor cultures of fully differentiated moss plants, P1. *Tissue culture and Biotech.*, 2, págs. 142-147 (1996)).

25 En otra forma de realización de la presente invención, se proporciona un procedimiento para producir por lo menos una célula de briofita en el que la actividad de  $\beta$ 1,3-GalT se reduce sustancialmente, lo que comprende la introducción en dicha célula, i) una primera secuencia de ácido nucleico que se dirige específicamente a la secuencia nucleotídica que codifica  $\beta$ 1,3 endógena según la SEC. ID. n° 1 y ii) una segunda secuencia de ácido nucleico que se dirige específicamente a la secuencia nucleotídica que codifica  $\beta$ 1,3 endógena según la SEC. ID. n° 2.

30 El destinatario experto apreciará que el orden de introducción de dichas primera y segunda secuencias de ácido nucleico transferasa en la célula de briofitas no sea importante: se pueda realizar en cualquier orden. La primera y segunda secuencias de ácido nucleico pueden dirigirse a partes específicas de los genes de  $\beta$ 1,3-GalT endógena, natural localizados en el genoma nuclear de la célula de briofitas definidas por las secuencias específicas de la enzima de restricción de los mismos, por ejemplo, según los ejemplos proporcionados en la presente memoria. Al dirigir específicamente las secuencias de los genes de  $\beta$ 1,3-GalT natural con las secuencias nucleotídicas que se integran específicamente con los genes de transferasa natural objetivo de interés, la expresión de dichas secuencias se altera sustancialmente si no se destruye completamente.

40 Preferentemente todas las proteínas de mamíferos glucosiladas mencionadas en la presente memoria anteriormente son de tipo humano. Otras proteínas que se contemplan para la producción en la presente invención incluyen proteínas para su utilización en la atención veterinaria y puede corresponder a homólogos animales de las proteínas humanas mencionadas en la presente memoria.

45 Un activador exógeno es el que representa un activador que se introduce enfrente de una secuencia de ácido nucleico de interés y está operativamente asociado a ésta. Por lo tanto un activador exógeno es el que se ha colocado enfrente de un componente del ácido nucleico seleccionado como se define en la presente memoria y no está constituido por el activador natural asociado normalmente al componente del ácido nucleico de interés como se encuentra en circunstancias naturales. Por lo tanto un activador puede ser natural para una célula de briofita de interés pero puede no estar asociado operativamente al ácido nucleico de interés enfrente en las células de briofita naturales. Por lo general, un activador exógeno es el que se transfiere a una célula hospedadora de briofita procedente de otra fuente aparte de la célula hospedadora.

50 Con respecto a la producción de estructuras de N-glucano con  $\beta$ 1,3-galactosilación mejorada los ADNc que codifican las proteínas  $\beta$ 1,3-GalT, las proteínas glucosiladas y de mamíferos tal como se describe en la presente memoria contienen por lo menos un tipo de activador que es operable en una célula de briofita, por ejemplo, un activador inducible o constitutivo operativamente unido a la secuencia de ácido nucleico de  $\beta$ 1,3-GalT y/o una segunda secuencia de ácido nucleico para una proteína de mamífero glucosilada tal como se define en la presente memoria y como proporciona la presente invención. Como se expuso, esto permite el control de la expresión del gen o genes.

60 El término "inducible" tal como se aplica a un activador es bien comprendido por los expertos en la materia. En esencia, la expresión bajo control de un activador inducible está "conectada" o ha aumentado la respuesta a un estímulo aplicado (que puede generarse dentro de una célula o proporcionarse exógenamente). La naturaleza del estímulo varía entre los activadores. Algunos activadores inducibles dan lugar a niveles pequeños o indetectables de expresión (o sin expresión) en ausencia del estímulo apropiado. Otros activadores inducibles producen expresión constitutiva detectable en ausencia del estímulo. Cualquiera que sea el nivel de expresión en ausencia del estímulo, la expresión de cualquier activador inducible aumenta en presencia del estímulo correcto. La situación preferible es

cuando el nivel de expresión aumenta durante la aplicación del estímulo apropiado mediante una cantidad eficaz para alterar una característica fenotípica. De este modo puede utilizarse un activador inducible (o "conectable"), el cual produce un nivel básico de expresión en ausencia del estímulo, cuyo nivel es demasiado bajo para provocar un fenotipo deseado (y puede ser de hecho cero). En la aplicación del estímulo, aumenta la expresión (o se conecta) a un nivel, que ocasiona el fenotipo deseado.

Tal como se alude en la presente memoria, los sistemas de expresión de briofitas son también conocidos por los expertos en la materia. Un activador de briofitas, en particular un activador de *Physcomitrella patens*, es cualquier secuencia de ADN capaz de unir una ARN polimerasa dependiente del ADN hospedador e iniciar la transcripción corriente abajo (3') de una secuencia de codificación (por ejemplo, gen estructural) en el ARNm. Un activador tendrá una zona de iniciación de la transcripción que está normalmente colocada próxima al extremo 5' de la secuencia de codificación. Esta zona de iniciación de la transcripción incluye normalmente una ARN polimerasa que une la secuencia ("secuencia TATA") y una zona de iniciación de la transcripción. Un activador de briofitas puede tener también un segundo dominio denominado secuencia activadora corriente arriba (UAS), que, si está presente, está normalmente distal del gen estructural. La UAS permite la expresión regulada (inducible). La expresión constitutiva se produce en ausencia de una UAS. La expresión regulada puede ser positiva o negativa, aumentando o reduciendo de este modo la transcripción.

El destinatario que sea experto en la materia apreciará que las secuencias activadoras de briofitas que codifican enzimas en las rutas metabólicas de las briofitas puedan suministrar secuencias activadoras particularmente útiles.

Además, los activadores sintéticos que no son naturales pueden funcionar también como activadores de briofitas. Por ejemplo, las secuencias UAS de un activador de briofitas pueden unirse con la región de activación de la transcripción de otro activador de briofitas, creando un activador híbrido sintético. Un ejemplo de activador adecuado es el utilizado en el sistema de expresión TOP 10 para *Physcomitrella patens* por (Zeidler *et al.* (1996) *Plant. Mol. Biol.* 30, 199-205). Además, un activador de briofitas puede incluir activadores naturales de origen distinto que las briofitas que tienen capacidad para unirse a una ARN polimerasa dependiente del ADN de briofitas e iniciar la transcripción. Ejemplos de dichos activadores incluyen los descritos, entre otros, el activador P-Actina 1 del arroz y el activador RbcS de *Chlamydomonas* (Zeidler *et al.* (1999) *J. Plant Physiol.* 154, 641-650), Cohen *et al.* *Proc Natl. Acad. Sci. USA*, 77: 1078, 1980; Henikoff *et al.* *Nature*, 283: 835, 1981; Hollenberg *et al.*, *Curr. Topics Microbiol. Immunol.*, 96: 119, 1981; Hollenberg *et al.*, "The Expression of Bacterial Antibiotic Resistance Genes in the Yeast *Saccharomyces cerevisiae*", en: *Plasmids of Medical, Environmental and Commercial Importance* (eds. K.N. Timms y A. Puhler), 1979; Mercerau-Puigalon *et al.*, *Gene*, 11: 163, 1980; Panthier *et al.*, *Curr. Genet.*, 2: 109, 1980.

Las moléculas de ADN según la presente invención pueden expresarse intracelularmente en briofitas. Una secuencia activadora puede unirse directamente a la molécula de ADN, en cuyo caso el primer aminoácido en el terminal N de la proteína recombinante será siempre una metionina, que está codificada por el codón de iniciación AUG del ARNm. Si se desea, la metionina en el terminal N puede escindirse de la proteína por incubación *in vitro* con bromuro de cianógeno.

Alternativamente, pueden ser segregadas también proteínas extrañas de las células de briofitas en el medio de cultivo creando moléculas de ADN híbrido que codifican una proteína de fusión compuesta por un fragmento de secuencia principal que proporciona la secreción dentro o fuera de las células de briofita de la proteína extraña. Preferentemente, existen zonas de tratamiento codificadas entre el fragmento principal y el gen extraño que pueden escindirse, ya sea *in vivo* o *in vitro*. El fragmento de la secuencia principal codifica normalmente un péptido señal compuesto por aminoácidos hidrófobos que dirigen la secreción de la proteína desde la célula.

El ADN que codifica secuencias señal adecuadas puede proceder de genes para las proteínas de briofitas segregadas, tales como las principales de origen distinto de briofitas, tal como un VEGF principal, existen las que pueden proporcionar también secreción en células de briofitas.

Las secuencias de terminación de la transcripción que son reconocidas por células de briofitas y funcionales en éstas son regiones reguladoras situadas en 3' respecto al codón de terminación de la traducción, y por tanto junto con el activador flanquean la secuencia de codificación. Estas secuencias dirigen la transcripción de un ARNm que puede traducirse en el polipéptido codificado por el ADN. Un ejemplo de una secuencia de terminación adecuada que actúa en *Physcomitrella patens* es la zona de terminación del virus del mosaico de la coliflor.

Por lo general, los componentes, que comprenden un activador, principal (si se desea), la secuencia de codificación de interés y la secuencia de terminación de la transcripción, se colocan juntos en los montajes de expresión de la invención. Los montajes de expresión se mantienen con frecuencia en un ADN plásmido, que es un elemento extracromosómico capaz de mantenimiento estable en un hospedador, tal como una bacteria. El ADN plásmido puede tener dos orígenes de replicación, permitiéndole de este modo que se mantenga, por ejemplo, en una briofita para la expresión y en un hospedador procariótico para la clonación y ampliación. Generalmente hablando es suficiente si el plásmido tiene un origen de replicación para la clonación y ampliación en una célula hospedadora procariótica. Además, un ADN plásmido puede ser un plásmido con un gran o pequeño número de copias. Un plásmido con un gran número de copias generalmente tendrá un número de copias que oscila entre

aproximadamente 5 y aproximadamente 200, y normalmente entre aproximadamente 10 y aproximadamente 150. Un hospedador que contiene un plásmido con gran número de copias tendrá preferentemente por lo menos aproximadamente 10, y más preferentemente por lo menos aproximadamente 20. Un vector con un gran o pequeño número de copias puede seleccionarse, dependiendo del efecto del vector y de la proteína extraña en el hospedador (véase, por ejemplo, Brake *et al.*, anteriormente).

Alternativamente, los montajes de expresión pueden integrarse en el genoma de briofitas con un vector de integración. Los vectores de integración normalmente contienen por lo menos una secuencia homóloga a un cromosoma de briofitas que permite que el vector se integre, y preferentemente contienen dos secuencias homólogas que flanquean el montaje de expresión. Un vector de integración puede dirigirse a un locus específico en el musgo seleccionando la secuencia homóloga apropiada para la inclusión en el vector tal como se describe y se ejemplifica en la presente memoria. Puede integrar uno o más montajes de expresión. Las secuencias cromosómicas incluidas en el vector pueden tener lugar ya sea como un sólo segmento en el vector, lo que produce la integración de todo el vector, o dos segmentos homólogos a los segmentos adyacentes en el cromosoma y que flanquean el montaje de expresión en el vector, lo que da como resultado la integración estable de solamente el montaje de expresión.

Normalmente, los montajes de expresión extracromosómicos e integradores pueden contener marcadores seleccionables que permitan la selección de células de briofitas que han sido transformadas.

Los marcadores seleccionables pueden incluir genes biosintéticos que pueden expresarse en el hospedador del musgo, tal como el G418 o los genes con resistencia a la higromicina B, que proporcionan resistencia en las células de briofitas G418 y a la higromicina B, respectivamente. Además, un marcador seleccionable adecuado puede proporcionar también células de briofitas con capacidad para crecer en presencia de compuestos tóxicos, tales como metales.

Alternativamente, algunos de los componentes descritos anteriormente pueden colocarse juntos en vectores de transformación. Los vectores de transformación están normalmente compuestos por un marcador seleccionable que se mantiene en un ADN plásmido o se desarrolla en un vector de integración, como se describió anteriormente.

Alternativamente, al conseguir altos rendimientos de casos de transformación como se observa en *Physcomitrella* puede evitarse la utilización de marcadores para la selección de los casos de transformación.

Los procedimientos de introducción de ADN exógeno en células de briofitas son bien conocidos en la técnica y son descritos entre otros por Schaefer D. G. "Principles and protocols for the moss *Physcomitrella patens*", (mayo de 2001) Institute of Ecology, Laboratory of Plant Cell Genetics, Universidad de Lausana; Reutter K. y Reski R., Plant Tissue Culture and Biotechnology septiembre de 1996, Vol. 2, nº 3; Zeidler M. *et al.*, (1996), Plant Molecular Biology 30:199-205.

Los expertos en la materia son perfectamente capaces de construir vectores y diseñar protocolos para la secuencia del ácido nucleico recombinante o la expresión de genes como se describió anteriormente. Los vectores adecuados pueden elegirse o construirse, conteniendo secuencias reguladoras adecuadas, incluyendo secuencias de activador, fragmentos de terminador, secuencias de poliadenilación, secuencias de potenciador, genes marcadores y otras secuencias, según corresponda. Para más detalles, véase, por ejemplo, Molecular Cloning: a Laboratory Manual: 2ª edición, Sambrook *et al.*, 1989, Cold Spring Harbor Laboratory Press. Muchas de las técnicas y protocolos conocidos para la manipulación de ácidos nucleicos, por ejemplo en la preparación de montajes de ácido nucleico, mutagenia, secuenciación, introducción de ADN en las células y la expresión génica y el análisis de las proteínas, se describen con detalle en Current Protocols in Molecular Biology, Segunda Edición, Ausubel *et al.* eds., John Wiley & Sons, 1992. Los descubrimientos de Sambrook *et al.* y Ausubel *et al.* se incorporan en la presente memoria por referencia.

Como se describió anteriormente, los marcadores genéticos seleccionables pueden facilitar la selección de células transgénicas de briofitas y éstas pueden consistir en genes híbridos que confieren fenotipos seleccionables según se refiere en este documento.

En la introducción de secuencias seleccionadas de ácido nucleico que codifican glucosiltransferasa y secuencias de polipéptido que comprenden actividad de glucosiltransferasa en una célula de briofitas, se deben tener en cuenta algunas consideraciones, muy conocidas por los expertos en la materia. El/los ácido(s) nucleico(s) que ha(n) de insertarse debería(n) ensamblarse en un montaje, que contiene elementos de regulación eficaces, que dirigirán la transcripción. Debe estar disponible un método de transporte del montaje en la célula. Una vez que el montaje está dentro de la membrana celular, la integración en el material cromosómico endógeno se producirá o no.

La invención abarca además una célula hospedadora transformada con vectores o montajes como se indicó anteriormente, especialmente una célula de briofitas o microbiana. Por lo tanto, se proporciona una célula hospedadora, tal como una célula de briofitas, incluyendo las secuencias nucleotídicas de la invención como se indica en la presente memoria. Dentro de la célula, la secuencia nucleotídica pueden incorporarse dentro del cromosoma.

Además, según la invención, se proporciona una célula de briofitas que ha incorporado en su genoma al menos una secuencia nucleotídica, particularmente secuencias nucleotídicas heterólogas, según proporciona la presente invención bajo control operativo de las secuencias reguladoras para control de la expresión como se describe en la presente memoria. La secuencia de codificación puede estar operativamente unida a una o más secuencias reguladoras que pueden ser heterólogas o extrañas a las secuencias de ácido nucleico empleadas en la invención, tales como, las no asociadas de forma natural a la(s) secuencia(s) de ácido nucleico para su expresión o sus expresiones. La secuencia nucleotídica según la invención puede colocarse bajo el control de un activador sumamente inducible para colocar la expresión bajo el control del usuario. Un aspecto adicional de la presente invención proporciona un procedimiento de preparación de dicha célula de briofita, específicamente una célula de *Physcomitrella patens* que conlleva la introducción de una(s) secuencia(s) de ácido nucleico contempladas para su utilización en la invención o por lo menos de un vector adecuado que incluye la(s) secuencia(s) contempladas para su utilización en la invención en una célula de briofitas y que da lugar o permite la recombinación entre el vector y el genoma de la célula de briofitas para introducir dichas secuencias en el genoma. La invención se extiende a las células de briofitas, particularmente a las células de *Physcomitrella patens* que contiene un nucleótido GalT y/o una secuencia nucleotídica que codifica una secuencia polipeptídica destinada a la adición de un patrón de glucosilación de mamífero a la misma y adecuada para su utilización de la invención como resultado de la introducción de una secuencia nucleotídica en una célula progenitora.

El término "heterólogo" se puede utilizar para indicar que el gen/secuencia de nucleótidos en cuestión se han introducido en las células de briofitas o un progenitor de las mismas, utilizando ingeniería genética, es decir, mediante intervención humana. Una célula transgénica de briofitas, es decir, transgénica para la secuencia de nucleótidos en cuestión, puede suministrarse. El transgén puede estar en un vector extragenómico o incorporarse, preferiblemente de forma estable, en el genoma. Un gen heterólogo puede reemplazar a un gen endógeno equivalente, es decir, uno que normalmente realiza la misma función o una función similar, o la secuencia insertada puede ser adicional al gen endógeno o a otra secuencia. Una ventaja de la introducción de un gen heterólogo es la capacidad para colocar la expresión de una secuencia bajo el control de un promotor de elección, con el fin de poder influir en la expresión según la preferencia. Las secuencias de nucleótidos heterólogas, exógenas o extrañas, a una célula de briofita pueden no ser naturales en las células de ese tipo, cepa o especie. De este modo, una secuencia nucleotídica puede incluir una secuencia de codificación o derivado de un tipo específico de una célula de briofita, tal como de una célula de *Physcomitrella patens*, colocada en el contexto de una célula de briofita de un tipo o especie diferente. Una posibilidad más es la de una secuencia nucleotídica que ha de colocarse en una célula de briofita, en la que ella o un homólogo se encuentra de forma natural, pero en la que la secuencia de nucleótidos está unida y/o adyacente al ácido nucleico que no es natural dentro de la célula o células de este tipo, especie o cepa, como por ejemplo operativamente unida a una o más secuencias reguladoras, tales como una secuencia activadora, para el control de la expresión. Una secuencia en una briofita u otra célula hospedadora puede ser heterólogas, exógena o extraña identificable.

La presente invención también abarca el producto deseado de la expresión de polipéptidos de la combinación de moléculas de ácido nucleico de acuerdo con la invención como se describe en este documento o que puede obtenerse según la información y sugerencias en este documento. También se proporcionan procedimientos de elaboración de dicho producto de expresión mediante la expresión de secuencias nucleotídicas que codifican por lo tanto, en condiciones adecuadas en las células hospedadoras adecuadas, por ejemplo *E. coli*. Los expertos en la materia están capacitados para construir vectores y diseñar protocolos y sistemas para la expresión y la recuperación de productos recombinantes de expresión génica recombinante.

Un polipéptido según la presente invención puede ser un alelo, variante, fragmento, derivado, mutante u homólogo de los polipéptidos (a) como se ha mencionado en la presente memoria. El alelo, variante, fragmento, derivado, mutante u homólogo puede tener sustancialmente la misma función de los polipéptidos aludida anteriormente y como se demuestra en este documento o puede ser un mutante funcional del mismo. En el contexto de las proteínas farmacéuticas como se describe en la presente memoria para su utilización en seres humanos, el destinatario experto valorará que la secuencia primaria de dichas proteínas y su patrón de glucosilación se parezca o, preferentemente, sea idéntica a la encontrada en los seres humanos.

"Identidad" en relación con una secuencia de ácido nucleico o una secuencia de aminoácidos de la invención puede utilizarse para referirse a la identidad de toda la secuencia o de partes esenciales de la misma. Como ya se ha señalado anteriormente, el alto nivel de identidad de aminoácidos puede limitarse a los dominios o regiones funcionalmente importantes, por ejemplo, cualquiera de los dominios identificados en este documento.

En particular, homólogos de las secuencias de polipéptidos específicas derivadas de briofitas proporcionados en este documento, son proporcionados por la presente invención, en cuanto que son mutantes, variantes, fragmentos y derivados de dichos homólogos. Por lo tanto la presente invención también se extiende a los polipéptidos que incluyen secuencias de aminoácidos con función de  $\beta$ 1,3-galactosiltransferasas tal como se define en este documento y como puede obtenerse utilizando información de la secuencia como se describe en la presente memoria. La  $\beta$ 1,3-galactosiltransferasa según la presente invención puede, en el nivel de aminoácidos tener identidad con las secuencias de aminoácidos de las secuencias dadas a conocer en la presente memoria,

especialmente de PpGalT1 o las pGalT1a (figuras 2, 4, de por lo menos 80% de identidad, o por lo menos 85%, o por lo menos 88% de identidad, o por lo menos 90% de identidad y más preferiblemente por lo menos 95% o mayor identidad, siempre que estas proteínas tengan actividad de  $\beta$ -1,3-galactosiltransferasa que se inscribe en el contexto de la presente invención. El % de identidad mencionado debería darse de preferencia en la región que comprende los siete dominios conservados como se representa en la figura 1 (incluido "-" apropiado que es obvio que ocurra a los expertos en la materia), al comparar las secuencias en cuestión, por ejemplo, con PpGalT1, PpGalT2, las PpGalT1as o las PpGalT2as.

Dominios o regiones funcionalmente significativos de diferentes polipéptidos pueden combinarse para la expresión de ácido nucleico codificador como proteína de fusión. Por ejemplo, propiedades particularmente ventajosas o deseables de diferentes homólogos pueden combinarse en una proteína híbrida, de modo que el producto de expresión resultante, con función de  $\beta$ 1,3-galactosiltransferasa, puede incluir fragmentos de varias proteínas originales, si procede.

La identidad puede calcularse fácilmente como % del valor de las secuencias alineadas (incluyendo "-" inteligente). La similitud de las secuencias de aminoácidos puede ser como se define y determina por el programa TBLASTN, de Altschul *et al.* (1990) *J. Mol. Biol.* 215: 403-10, que está en uso convencional en la técnica. En particular, TBLASTN 2.0 puede utilizarse con Matrix BLOSUM62 y penalizaciones GAP (por hueco): existencia: 11, extensión: 1. Otro programa convencional que puede utilizarse es BestFit, que forma parte del paquete Wisconsin, versión 8, septiembre de 1994 (Genetics Computer Group, 575 Science Drive, Madison, Wisconsin, USA, Wisconsin 53711). BestFit realiza una alineación óptima del mejor segmento de similitud entre dos secuencias. Las alineaciones óptimas se encuentran insertando huecos para maximizar el número de emparejamientos utilizando el algoritmo de identidad local de Smith y Waterman (*Adv. Appl. Math.* (1981) 2:482-489). Otros algoritmos incluyen GAP, que utiliza el algoritmo de Needleman y Wunsch para alinear dos secuencias completas que maximiza el número de emparejamientos y minimiza el número de huecos. Como con cualquier algoritmo, generalmente se utilizan los parámetros por defecto, que para GAP son una penalización = 12 por creación del hueco y penalización = 4 por extensión del hueco. Alternativamente, puede utilizarse una penalización de 3 por creación del hueco y una penalización de 0,1 por extensión del hueco. El algoritmo FASTA (que utiliza el método de Pearson y Lipman (1988) *PNAS USA* 85: 2444-2448) es otra alternativa.

Un procedimiento ventajoso para producir células hospedadoras recombinantes, en particular células vegetales, o plantas, respectivamente, consiste en que la molécula de ADN según la presente invención, que comprende especialmente una mutación inactivadora se inserta en el genoma de la célula hospedadora, o de la planta, respectivamente, en lugar de la secuencia homóloga no mutante (Schaefer *et al.*, 1997, *Plant. J.*; 11(6):1995-1206). Este procedimiento por lo tanto no funciona con un vector, sino con una molécula de ADN pura. La molécula de ADN según la presente invención se inserta en el hospedador por ejemplo por bombardeo génico, microinyección o transferencia directa de ADN mediada por PEG, para mencionar sólo tres ejemplos. Esta molécula de ADN se une a la secuencia homóloga en el genoma del hospedador de modo que resulte en el genoma una recombinación homóloga y por lo tanto la recepción de la mutación, por eliminación, inserción o sustitución, respectivamente: La expresión de la  $\beta$ 1,3-galactosiltransferasa puede por ejemplo suprimirse o bloquearse completamente, respectivamente.

Otro aspecto de la invención se refiere a vegetales, tejidos vegetales o células vegetales, respectivamente siendo su actividad de  $\beta$ 1,3-galactosiltransferasa inferior al 50%, en particular inferior al 20%, prefiriéndose particularmente el 0%, de la actividad de  $\beta$ 1,3-galactosiltransferasa que se producen en las plantas naturales o en las células vegetales. La ventaja de estas plantas o células vegetales, respectivamente, es que las glicoproteínas producidas por ellas no contienen ninguno o apenas comprenden ninguna galactosa unida por  $\beta$ 1,3. Si los productos de estas plantas, respectivamente, son ingeridos por cuerpos humanos o vertebrados, no habrá ninguna reacción inmunitaria para el epítipo de galactosa unido por  $\beta$ 1,3.

Preferentemente, se proporcionan plantas o células vegetales recombinantes, respectivamente, que se han preparado por uno de los procedimientos descritos anteriormente, suprimiéndose o bloqueándose, respectivamente, su producción de  $\beta$ 1,3-galactosiltransferasa.

La invención se refiere también a una molécula de APN que comprende una secuencia de bases complementaria a la secuencia de la molécula de ADN según la invención así como a secuencias parciales de la misma. El APN (ácido péptido nucleico) es una secuencia similar al ADN, estando las bases nucleotídicas unidas a un pseudoesqueleto peptídico. El APN generalmente de híbrida con oligómeros de ADN, ARN o APN complementarios con emparejamiento de bases de Watson-Crick y formación de la hélice. El eje central del péptido garantiza una resistencia mayor a la degradación enzimática. La molécula de APN por lo tanto es un agente complementario mejorado. Ni las nucleasas ni las proteasas son capaces de atacar una molécula de APN. La estabilidad de la molécula de APN, si está unida a una secuencia complementaria, comprende un bloqueo estérico suficiente de las ADN y ARN polimerasas, transcriptasa inversa, telomerasa y ribosomas. Si la molécula de APN comprende la secuencia mencionada anteriormente, se unirá al ADN o a una secuencia del ADN, respectivamente, que codifica la  $\beta$ 1,3-galactosiltransferasa y de esta manera es capaz de inhibir la transcripción de esta enzima. Ya que ni se transcribe ni se traduce, la molécula de APN se preparará por síntesis, por ejemplo con ayuda de la técnica t-Boc.

Mejor, se proporciona una molécula de APN que comprende una secuencia de bases que corresponde a la secuencia de la molécula de ADN de la invención así como a secuencias parciales de la misma. Esta molécula de APN acomplejará al ARNm o a una secuencia del ARNm de la  $\beta$ 1,3-galactosiltransferasa de modo que la traducción de la enzima se inhibirá. Argumentos similares a los indicados para el ARN complementario se aplican en este caso. Por lo tanto, por ejemplo, una región acomplejante particularmente eficaz es la región de iniciación de la traducción o también las regiones 5' no traducidas del ARNm.

Otro aspecto de la presente invención se refiere a un procedimiento de preparación de plantas, tejidos o células, respectivamente, en particular células vegetales que comprenden una expresión bloqueada de la  $\beta$ 1,3-galactosiltransferasa en el nivel de transcripción o traducción, respectivamente, que se caracteriza porque las moléculas de APN de la invención se insertan en las células. Para insertar la molécula de APN o las moléculas de APN, respectivamente, en la célula, de nuevo se utilizan procedimientos convencionales, tales como, por ejemplo, la electroporación o la microinyección. Resulta particularmente eficaz la inserción si los oligómeros del APN se unen a péptidos de penetración en la célula, por ejemplo, transportán o pAntp (Pooga *et al.* 1998, *Nature Biotechnology*, 16: 857-861).

La invención proporciona un procedimiento de preparación de glucoproteína recombinantes que se caracteriza porque las plantas o las células vegetales recombinantes de la invención, respectivamente, cuya producción de  $\beta$ 1,3-galactosiltransferasa se suprime o se bloquea completamente o las plantas, tejidos o células, respectivamente, en las que las moléculas de APN se han insertado según el procedimiento de la invención, se transfectan con el gen que expresa la glucoproteína de modo que se expresan las glucoproteínas recombinantes. De este modo, como ya se ha descrito anteriormente, los vectores que comprenden genes para las proteínas deseadas se transfectan en el hospedador o en las células hospedadoras, respectivamente, como ya se ha descrito también anteriormente. Las células vegetales transfectadas expresarán las proteínas deseadas, y no presentan o apenas ninguna galactosa unida por  $\beta$ 1,3. Por lo tanto, no desencadenan las reacciones inmunitarias ya mencionadas anteriormente en el cuerpo humano o de vertebrados.

Algunas proteínas pueden producirse en estos sistemas.

Ventajosamente, se proporciona un procedimiento de preparación de glucoproteínas humanas recombinantes que se caracteriza porque las plantas o células vegetales recombinantes, respectivamente, cuya producción de  $\beta$ 1,3-galactosiltransferasa se suprime o se bloquea completamente o las plantas, tejidos o células, respectivamente, en las que las moléculas de APN se han insertado según el procedimiento de la invención, se transfectan con el gen que expresa la glucoproteína de modo que se expresan las glucoproteínas recombinantes. Por este procedimiento será posible producir proteínas humanas en plantas (células vegetales) que, si son ingeridas por cuerpo humano, no desencadenan ninguna reacción inmunitaria dirigida contra los restos de galactosa unidos en  $\beta$ 1,3. En la presente memoria, es posible utilizar tipos de plantas para producir las glucoproteínas recombinantes que sirven como artículos alimenticios, por ejemplo plátanos, patatas y/o tomates. Los tejidos de esta planta comprenden la glucoproteína recombinante de modo que, por ejemplo por extracción de la glucoproteína recombinante del tejido y posterior administración, o directamente comiendo el tejido vegetal, respectivamente, la glucoproteína recombinante se ingiere en el cuerpo humano. Preferentemente, un procedimiento de preparación de glucoproteínas humanas recombinantes para uso médico se proporciona, en el que las plantas o células vegetales recombinantes de la invención, respectivamente, cuya producción  $\beta$ 1,3-galactosiltransferasa está suprimida o completamente bloqueada, respectivamente, o las plantas, tejidos o células, respectivamente, en las que las moléculas de APN se han insertado según el procedimiento de la invención, se transfectan con el gen que expresa la glicoproteína de modo que las glucoproteínas recombinantes se expresan. De este modo, puede utilizarse cualquier proteína que sea de interés médico.

Por otra parte, la presente invención se refiere a glucoproteínas recombinantes según un procedimiento descrito anteriormente, en el que se han preparado en sistemas vegetales y en el que su secuencia peptídica comprende menos del 50%, en particular menos del 20%, particularmente preferido el 0% de los restos de galactosa unidos por  $\beta$ 1,3 que se producen en las proteínas expresadas en sistemas vegetales no reducidos de galactosiltransferasa. Naturalmente, se prefieren las glucoproteínas que no comprenden restos de galactosa unidos por  $\beta$ 1,3. La cantidad de galactosa unida por  $\beta$ 1,3 dependerá del grado de supresión de la  $\beta$ 1,3-galactosiltransferasa descrito anteriormente. Preferentemente, las glucoproteínas recombinantes humanas que se han producido en sistemas vegetales según un procedimiento descrito anteriormente comprenden menos del 50%, en particular menos del 20%, prefiriendo particularmente el 0%, de los restos de galactosa unidos en  $\beta$ 1,3 que se producen en las proteínas expresadas en plantas o sistemas no reducidos por galactosiltransferasa.

En la presente memoria se describen glucoproteínas humanas recombinantes para su uso en medicina que han sido preparadas en sistemas vegetales según un procedimiento descrito anteriormente y cuya secuencia peptídica comprende menos del 50%, en particular menos del 20%, siendo preferido el 0%, de los restos de galactosa unidos en  $\beta$ 1,3 que se producen en las proteínas expresadas en sistemas vegetales no reducidos por galactosiltransferasa.

Se describe además una composición farmacéutica que comprende las glucoproteínas producidas según el procedimiento de la invención. Además de las glucoproteínas, la composición farmacéutica comprende más

adiciones normales de dichas composiciones. Éstas son, por ejemplo, agentes de dilución adecuados de varios contenidos de tampón por ejemplo, Tris-HCl, acetato, fosfato, pH y fuerza iónica, aditivos, tales como tensioactivos y disolventes (por ejemplo Tween 80, Polisorbato 80), conservantes (por ejemplo Thimerosal, alcohol bencílico), adyuvantes, antioxidantes (por ejemplo ácido ascórbico, metabisulfito sódico), emulsionantes, cargas (por ejemplo lactosa, manitol), enlaces covalentes de polímeros, tales como polietilenglicol, para la proteína, incorporación del material en composiciones en partículas de compuestos poliméricos, tales como ácido poliláctico, ácido poliglicólico, etc. o en liposomas, agentes auxiliares y/o sustancias transportadoras que son adecuadas en el tratamiento respectivo. Dichas composiciones influirán en el estado físico, la estabilidad, la velocidad de liberación *in vivo* y la velocidad de excreción *in vivo* de las glucoproteínas de la invención.

La invención proporciona además un procedimiento de selección de moléculas de ADN que codifican una  $\beta$ 1,3-galactosiltransferasa, en una muestra, en la que las moléculas de ADN marcadas de la invención se mezclan con la muestra, que se une a las moléculas de ADN que codifican una  $\beta$ 1,3-galactosiltransferasa. Las moléculas de ADN híbridadas pueden detectarse, cuantificarse y seleccionarse. Para la muestra que contiene ADN monocatenario con la que las moléculas de ADN marcadas pueden hibridarse, la muestra se desnaturaliza, por ejemplo calentando.

Una posible manera consiste en separar el ADN que ha de analizarse, posiblemente después de la adición de endonucleasas, por electroforesis en gel en un gel de agarosa. Una vez se han transferido a una membrana de nitrocelulosa, las moléculas de ADN marcadas según la invención se mezclan las cuales se hibridan con la molécula correspondiente de ADN homólogo ("inmunotransferencia Southern").

Otra posible manera consiste en el descubrimiento de genes homólogos en otras especies por procedimientos dependientes de RCP que utilizan cebadores específicos y/o degenerados, derivados de la secuencia de ADN según la invención.

Preferentemente, la muestra del procedimiento de la invención identificada anteriormente comprende ADN genómico de un organismo vegetal. Mediante este procedimiento, se analiza un gran número de plantas de manera muy rápida y eficaz la presencia del gen  $\beta$ 1,3-galactosiltransferasa. De esta manera, es posible respectivamente seleccionar plantas que no comprendan este gen o suprimir o bloquear completamente, respectivamente, la expresión de la  $\beta$ 1,3-galactosiltransferasa en dichas plantas que contienen este gen, por un procedimiento de la invención descrito anteriormente, de modo que posteriormente pueden utilizarse para la transfección y producción de glucoproteínas (humanas).

La presente memoria se refiere además a moléculas de ADN que codifican una  $\beta$ 1,3-galactosiltransferasa que se ha seleccionado según los dos últimos procedimientos mencionados y posteriormente se han aislado de la muestra. Estas moléculas pueden utilizarse para análisis adicionales. Pueden secuenciarse y a su vez pueden utilizarse como sondas de ADN para encontrar  $\beta$ 1,3-galactosiltransferasas. Estas moléculas marcadas de ADN funcionarán en organismos, las cuales están relacionadas con los organismos de los que se han aislado, de manera más eficaz como sondas que las moléculas de ADN de la invención.

La invención se refiere además a un procedimiento de preparación de unidades de carbohidrato "plantificadas" de glucoproteínas humanas y de otros vertebrados, en las que las unidades de fucosa así como la  $\beta$ 1,3-galactosiltransferasa codificada por una molécula de ADN descrita anteriormente se mezclan con una muestra que contiene una unidad de carbohidrato o una glucoproteína, respectivamente, de modo que la galactosa  $\beta$ 1,3 estará unida por la  $\beta$ 1,3-galactosiltransferasa a la unidad de carbohidrato o a la glucoproteína, respectivamente. Mediante el procedimiento según la invención para clonar la  $\beta$ 1,3-galactosiltransferasa es posible producir grandes cantidades de enzima purificada. Para obtener una transferasa totalmente activa, se proporcionan condiciones de reacción adecuadas.

La invención se explicará con más detalle mediante los siguientes ejemplos y las figuras en dibujo a las que, desde luego, no se limitan.

La figura 1 presenta una alineación de aminoácidos de  $\beta$ 1,3-GalT. Los siete dominios conservados de las  $\beta$ 1,3-galactosiltransferasas se indican en letras en negrilla. Los restos de aminoácidos conservados están indicados por asteriscos. Las similitudes según la secuencia de referencia de humanos (CAA75344,  $\beta$ 1,3-galactosiltransferasa de seres humanos) están predichas de la forma siguiente BAD17182 ( $\beta$ 1,3-galactosiltransferasa supuesta de *Oryza sativa*) = 17%; NP174003 (supuesta  $\beta$ 1,3-galactosiltransferasa de *Arabidopsis thaliana*) = 16%; PpGalT1 ( $\beta$ 1,3-galactosiltransferasa 1 de *Physcomitrella patens*) = 15%; PpGalT2 ( $\beta$ 1,3-galactosiltransferasa 2 de *Physcomitrella patens*) = 16%;

la figura 2 presenta la secuencia de proteínas predichas a partir de la secuencia de ADN de codificación del gen de  $\beta$ 1,3-galactosiltransferasa 1 de *Physcomitrella patens*. Este dominio transmembranario está indicado en letras en negrilla; y

la figura 3 presenta la secuencia de proteínas predichas de la secuencia de ADN de codificación del gen de  $\beta$ 1,3-galactosiltransferasa 2 de *Physcomitrella patens*. Este dominio transmembranario está indicado en letras en negrilla.

La figura 4 presenta la secuencia de proteínas de una variante de corte y empalme alternativa del gen para la  $\beta$ 1,3-galactosiltransferasa 1 de *Physcomitrella patens*. La inserción por corte y empalme de 50 aminoácidos está indicada en letras en negrilla.

La figura 5 presenta la secuencia de proteínas de una variante de corte y empalme alternativa del gen para la  $\beta$ 1,3-galactosiltransferasa 2 de *P. patens*. La inserción por corte y empalme de 50 aminoácidos está indicada en letras en negrilla.

## Ejemplos

### Procedimientos y materiales

#### Material vegetal

Se utilizó la cepa transgénica de *Physcomitrella patens* gluco-modificada genéticamente dos veces que carece de restos de fucosa y xilosa en la estructura nuclear de N-glucanos (Koprivova *et al.* 2004 *Plant Biotechnol. J.* 2, 517-523).

#### Condiciones normales de cultivo

Se cultivaron plantas en cultivo axénico en condiciones estériles en un medio de Knop modificado líquido inorgánico vegetal (1000 mg/l de  $\text{Ca}(\text{NO}_3)_2 \cdot 4\text{H}_2\text{O}$ , 250 mg/l KCl, 250 mg/l de  $\text{KH}_2\text{PO}_4$ , 250 mg/l de  $\text{MgSO}_4 \cdot 7 \text{H}_2\text{O}$  y 12,5 mg/l  $\text{FeSO}_4 \cdot 7 \text{H}_2\text{O}$ ; pH 5,8 (Reski y Abel (1985) *Planta* (165, 354-358). Las plantas se cultivaron en matraces Erlenmeyer de 500 ml que contenían 200 ml de medio de cultivo y los matraces se agitaron en un agitador Certomat R (B. Braun Biotech International, Alemania) ajustado a 120 rpm. Las condiciones en la cámara de cultivo eran  $25 \pm 3^\circ\text{C}$  y un régimen de luz-oscuridad de 16:8 h. Los matraces se iluminaron desde arriba mediante dos tubos fluorescentes (Osram L 58 W/25) que proporcionan  $35 \text{ micromoles}^{-1} \text{ m}^{-2}$ . Los cultivos se subcultivaron una vez a la semana por disgregación utilizando un homogeneizador con Ultra-Turrax (IKA, Staufen, Alemania) e inoculación de dos nuevos matraces Erlenmeyer de 500 ml que contienen 100 ml de medio Knop reciente.

#### Aislamiento del protoplasto

Después de la filtración los protonemas de musgo se incubaron previamente en manitol 0,5 M. Después de 30 min., se añadió a la suspensión Driselasa al 4% (Sigma, Deisenhofen, Alemania). La Driselasa se disolvió en manitol 0,5 M (pH 5,6-5,8), se centrifugó a 3600 rpm durante 10 min. y se esterilizó mediante el paso a través de un filtro de 0,22 micras (Millex Gp, Millipore Corporation, EE.UU.). La suspensión, que contiene Driselasa al 1% (concentración final), se incubó a la oscuridad a t.a. y se agitó suavemente (se consiguieron mejores rendimientos de protoplastos después de 2 horas de incubación) (Schaefer, "Principles and protocols for the moss *Physcomitrella patens*", (mayo de 2001) Instituto de Ecología, Laboratorio de Genética de células vegetales, Universidad de Lausana. La suspensión se pasó a través de tamices (Wilson, CLF, Alemania) con tamaños de poro de 100 micras y 50 micras. La suspensión se centrifugó en tubos de centrifugadora estériles y se sedimentaron los protoplastos a t.a. durante 10 min. a 55 g (aceleración de 3; disminuyendo a 3; Multifuge 3 S-R, Kendro, Alemania) (Schaefer, anteriormente). Se volvieron a poner en suspensión los protoplastos en medio 3M ( $\text{MgCl}_2 \cdot 2 \text{H}_2\text{O}$ , 15 mM MES 0,1%; Manitol 0,48M; pH 5,6; 540 mOsm; se esterilizó por filtración, (Schaefer *et al.* (1991) *Mol Gen Genet* 226, 418-424). La suspensión se centrifugó otra vez a t.a. durante 10 min. a 55 g (aceleración de 3; se disminuyó a 3; Multifuge 3 S-R, Kendro, Alemania). Los protoplastos se volvieron a poner en suspensión suavemente en medio 3M ( $\text{MgCl}_2 \cdot 2 \text{H}_2\text{O}$  15 mM, MES 0,1%; Manitol 0,48 M; pH 5,6; 540 mOsm; se esterilizó por filtración, (Schaefer *et al.* (1991) *Mol Gen Genet* 226, 418-424). Para el recuento de los protoplastos se transfirió un pequeño volumen de la suspensión a una cámara de Fuchs-Rosenthal.

#### Protocolo de transformación.

Para la transformación de los protoplastos se incubaron en hielo en la oscuridad durante 30 minutos. Posteriormente, los protoplastos se sedimentaron por centrifugaron a t.a. durante 10 min. a 55 g (aceleración de 3; disminución a 3; Multifuge 3 S-R, Kendro). Los protoplastos se volvieron a poner en suspensión en medio 3M ( $\text{MgCl}_2 \cdot 2 \text{H}_2\text{O}$  15 mM, MES al 0,1%; manitol 0,48 M; pH 5,6; 540 mOsm; se esterilizó por filtración (Schaefer *et al.* (1991) *Mol Gen Genet* 226, 418-424) a una concentración de  $1,2 \times 10^6$  protoplastos/ml (Reuter y Reski (1996) Production of a heterologous protein in bioreactor cultures of fully differentiated moss plants, P1. Tissue culture and Biotech. 2, págs. 142-147). 25 microlitros de esta suspensión de protoplastos se dispensaron en un nuevo tubo de centrifugadora esterilizado, 5 microlitros de solución de ADN (ADN purificado en columna en  $\text{H}_2\text{O}$  (Qiagen, Hilden, Alemania); 10-100 microlitros; se añadió una cantidad óptima de ADN de 16 microgramos) y por último se añadieron 25 microlitros de solución de PEG (PEG 4000 al 40%; manitol 0,4 M;  $\text{Ca}(\text{NO}_3)_2$  0,1 M; pH 6 tras la esterilización en autoclave). La suspensión se mezcló inmediata pero suavemente y a continuación se incubó durante 6 min. a t.a. con mezclado suave ocasional. La suspensión se diluyó progresivamente añadiendo 1, 2, 3 y 4 ml de medio 3M. La suspensión se centrifugó  $20^\circ\text{C}$  durante 10 minutos a 55 g (aceleración de 3; disminución a 3; Multifuge 3 S-R,



Kendro). El sedimento se volvió a poner en suspensión en 3 ml de medio de regeneración (medio Knop modificado; glucosa al 5%; manitol al 3%; 540 mOsm; pH 5,6 - 5,8). La regeneración se llevó a cabo como describe Strepp *et al.* (1998) *Proc. Natl. Acad. Sci. USA* 95, 4368-4373). Se identificaron clones transgénicos por identificación molecular.

## 5 MS MALDI-TOF de glucanos del musgo.

Se cultivó material vegetal en cultivo líquido, se aisló por filtración, se congeló en nitrógeno líquido y se almacenó a -80°C. El material se transportó bajo nieve carbónica. Los análisis por MS de MALDI-TOF se realizaron en el laboratorio del Prof. Dr. F. Altmann, División de Glucobiología, Instituto de Química, Universität für Brodenkultu, Viena, Austria.

Se digirió con pepsina 0,2 a 0,5 g de peso fresco de material transgénico de *Physcomitrella patens*. Se obtuvieron N-glucanos del digesto tal como describe Wilson *et al.* (2001). Esencialmente los glucanos se liberaron por tratamiento con péptido:N-glucosidasa A y se analizaron por espectrometría de masas MALDI-TOF en un DYNAMO (Thermo BioAnalysis, Santa Fe, NM).

### 1. Identificación de genes que codifican $\beta$ 1,3-galactosiltransferasa.

Aunque las  $\beta$ 1,3-galactosiltransferasas ( $\beta$ 1,3galT) con funcionalidad biológica de seres humanos en relación con el alargamiento de las estructuras de N-glucano no estaban descritas se seleccionó la secuencia de la  $\beta$ 1,3-galT 2 (nº de registro: CAA75344) de seres humanos como secuencia de partida. Basándose en los que siete dominios conservados descritos por Henet (2002 *Cell. Mol. Life Sci.* 59, 1081-1095) y en combinación con los aminoácidos conservados descritos por Amado *et al.* (1998 *J. Biol. Chem.* 273, 12770-12778) se realizó un cribado de la base de datos. Debido a esta estrategia se identificaron una secuencia de *Arabidopsis Thaliana* (nº de registro: NP174003) y una secuencia de *Oryza Sativa* (nº de registro: BAD17812) descrito como supuestas  $\beta$ 1,3-galactosiltransferasas. Aunque para ambas especies numerosas secuencias de proteínas de las supuestas  $\beta$ 1,3-galactosiltransferasas se listaron en las bases de datos públicas únicamente estas dos similitudes presentadas por una parte en los siete dominios conservados y por la otra en varios de los aminoácidos adicionales muy conservados. Sin embargo, si se compara con CAA75344 la identidad global fue muy baja en ambos, en el caso NP174003 fue del 16%, en el caso de BAD17812 fue del 17% (figura 1).

Las tres secuencias de proteínas se utilizaron para la identificación de una base de datos no pública "etiqueta de la secuencia expresada" (EST) de *Physcomitrella patens*. Se identificó una etiqueta de la secuencia expresada que codifica una secuencia peptídica que comprende algunas similitudes con los siete dominios conservados de las  $\beta$ 1,3-galactosiltransferasas. Esta EST se utilizó para diseñar cebadores para clonación y para la identificación adicional en relación con una familia de genes con beta1,3-galactosiltransferasa de una base de datos que contienen secuencias genómicas de *Physcomitrella patens*.

Las secuencias resultantes comprendían dos supuestos genes con  $\beta$ 1,3-galactosiltransferasa incluyendo secuencias con intrón y exón y las estructuras génicas ( $\beta$ 1,3-galT 1 corresponde a la SEC. ID. nº 1 y SEC. ID. nº 3 y  $\beta$ 1,3-galT 2 corresponde a las SEC. ID. nº 2 y SEC. ID. nº 4). Las secuencias proteicas predichas de los marcos de lectura abiertos ( $\beta$ 1,3-galT 1 (figura 2) y  $\beta$ 1,3-galT 2 (figura 3) comprendían dominios transmembranarios, los siete dominios conservados y numerosos aminoácidos conservados (figura 1).

### 45 1.1 Clonación de la secuencia de codificación del gen para $\beta$ 1,3-galactosiltransferasa 1 de *Physcomitrella patens*

Ampliación de la secuencia de nucleótidos que codifica  $\beta$ 1,3-galactosiltransferasa de *Physcomitrella patens*

(SEC ID nº: 1:

5'AGTTGTCGATTTGTTGTTTTGATATGTAAGCGGTTGCCTTCGCGCCGTGCTTGATTGTAATTGTAATTCAATC  
TGGAGTGTGAGATATATATATATATATATATAGCGAGAGGGAGAGAGAAAGAGAGAGAGAGGGAGAGAGAAA  
GAGAGAGAGAGGGAGAGAGAGAGATGGCTTGTGTATGAGGGCCATGCGAGGAGGAGGCTGTGTTTGTGCCC  
GAAGAGATGGGATGGTTTATGTGTAGTGCAGGGGTTGGATGTGAAGCACCTGTTTGAAGGAGTCTGCGAGAGTT  
TGAATTCGGATTACAGAGTCCGCGATCGATGGTGAACGTTGTTAGCAGTGATTGTTTTCGCCAACAGAACTG  
55 ACATCATTTGGATTTTTTTACGCGTGGATGTGCCCTCTTTTAAAAAATTTCCGCGTGAANAGAGACGGGGGT  
TTGTAATGGAGGCAGGCTGTGGTCATCACCCCTAGTATAGCCTGTCAAGAGAGTTCAAATTCGGTAATATGAAGA  
GGGGTTCGAGACTACCGGATATGGCGTGTACAGGGCGGCAAGAAATGATCTTATCCTAGTTGCAATTGTTTGC  
TTGTTTTTATGGTGATATTCATCCCACCATATCTCCAAATGAACTCACTTCCGGACATTTAGTTCTCCTGATTCGG  
ACAAGAAATCATCAAGCTACTCGAAAAAACCCTCTAGAAGCCAATAGTAAGGAGGAACGCCGTAGTCCGGGG  
60 AATACCACAGGCGACATTGTTTCTCTGGATGATGTGATAGATCGTGCCTGGTCTGCTGGTGCCAAAGCGTGGGA  
AGAAGTGGAACTGCGTTAAGAAATGGAGAAGGTGTCTCAAAGAATGTCAGTAATGCCACTGCAAATGCTGATCC  
GTGTCCAGCATCACTCTCTGCAGCAGGGAAGTTAGACGAAATGGGTTAAAGTCTTCCCCTTGCCCTGTGGTC  
TAATGTTTGGGTACGCCATTACTCTGATTGGAAAGCCTCGAGAGGCTCACATGGAGTACAACCCCAATCGCC  
AGAGTTGGGAAGCGTCTCCATATGTCATGGTTCCAGTTCTTAGTAGATTACAAGGCTTAAAGGTGGTG  
65 AAAGGTGAAGATCCTCCTCGAATTCTACACTGAATCCTCGACTTCGTGGTGAATTGGAGCTGAAACCCATCATT  
GAGCACAACACTTGTATCGGAACCAGTGGGGTCTGCCACCGATGCGAGGGTTGGCAAGTGCCTGAATACG

AAGAAACTGTTGACGGTCTTCCCAAGTGCAGAGAAGTGGCTTCGAGATGATGGCAAGAAACCTGCTTCAACGCAA  
 AAATCTTGGTGGCTTGAAGATTAGTTGGTCTTCTGACAAGGAGACGCTTGAATGGGAGTACCCATTATCTGAG  
 GGTCCGGAGTTCGTTCTCACCATTTCGAGCAGGTGTTGAAGGGTTTCATGTGACTATCGATGGTCCGTCACATCAG  
 CTCGTTTCCTTATCGTGTGGGTTACGCTGTGGAAGAAACAACGGGGATATTAGTAGCAGGAGACGTTGATGTGA  
 5 TGTCTATCACAGTGACATCCCTACCCTAACACATCTAGCTACTACCCTGAGTTAGTTTTGGAATCGGGGGACA  
 TTTGGAAGGCACCACCTGTCCCAGCTACCAAGATAGATTTATTTATTTGGGATCATGTCCAGCAGTAACCAATTTG  
 CAGAACGGATGGCAGTAAGGAAGACGTGGTTTCAATCTAAAGCTATTCAATCTTCGACGGCCGTGGCTCGCTTC  
 TTTGTAGCTCTGCATGCAAACAAGGATATCAATATGCAGTTGAAGAAGGAGGCAGACTATTATGGCGATATTATA  
 10 ATCCTGCCTTTCATCGACAGATATGATATAGTGGTCTCAAGACCGTTGAAATTTGCAAGTTTGGGGTCCAGAAT  
 GTCACAGCTAAGTATATTATGAAGTGTGACGATGACACTTTTGTGAGGATTGATAGCGTTCTCGAAGAGATTCTGA  
 ACTACTTCAATATCACAAGGCCCTTACATGGGTAGCATGAATGAGTTTACAGGCCTCTTCGTTCTGGAAAGTGG  
 GCCGTGACTGCCGAGGAATGGCCTGAGCGAATTTACCCAATATATGCTAATGGACCAGGATATATCCTGTCAGA  
 GGATATTGTGCATTTTCATTGTGGAGATGAATGAGAGAGGCAGTTTGCAGTTATTTAAGATGGAGGACGTCAGTGT  
 15 CCGGTTGTATACCGAAATACTTGACAGCTATTACCAATCGCGCGTCAAATGCTGTGTCTGTGGGACAAGGTA  
 CTTGCTCATGACGATGGGAAATGCTGCAACTTGT**AG**GGAAAATACATACAATGAATGTGTTCAACGGTCTTTACC  
 AGACAGAATTACTTTGGGTCGGGAACCAGATATAGCAGACAGCTCACATTCAATTCAGCCGTGTTGATCCAGAG  
 GGGTAATTGATAGTTTCTTGTCCCCTACCCTCTCTAGAGGTGGAGATCTTACAACCTAATCAAATGATCCTCTGC  
 AATGTCACCTTGTCAACAATACTTAGTATAGCTCAAAAT**TT**GGCCACGGATATTCAGGAATGTTTCATCTTGAAGGTCG  
 20 AGACTTGTGAGTAAATGGTTGGGTGGTGTGCGATGGCATGGTGTCTTATCAATCCCTCTTAGCATCAGTATGATCGTC  
 AGAATCAGTGTTCGACACTCCCCGGTGGAGTATTTTTTCGATTCTCTTATTGATTCCACTCAAGTACTAGCTTA  
 TATTTAGTGAGGCTCGAACCCAAGTAGTTAGTTACGTCTGCCTTTTGGCGAAATGAGTAGAGTAATTTGT  
 GGCAGTAGTTGGTGAAGAGACATGGTTAGGATTTAGTGTTCAAAATCTG 3'; los codones de inicio y terminación  
 están indicados en negrilla) se realizó por RCP con ADNc y los cebadores MOB1251 (SEC ID nº: 5: 5'-  
 25 CTGAATATCCGTGGCCAA-3') y el cebador MOB1410 (SEC ID nº: 6: 5'-  
 TTCGAGCTCATGAAGAGGGGTCGAGACT-3'). El producto de ampliación se digirió con Sac I y Msc I y se clonó  
 en el vector pRT101 digerido con Sac I/Sma I (Toepfer *et al.* 1987 NAR 15, 5890). La secuencia clonada se verificó  
 por secuenciado.

30 1.2 Clonación de la secuencia de codificación del gen para  $\beta$ 1,3-galactosiltransferasa 2 de *Physcomitrella patens*

Ampliación de la secuencia de nucleótidos que codifica  $\beta$ 1,3-galactosiltransferasa de *Physcomitrella patens*

(SEC ID nº: 2: 5'-**ATGA**AGAGGGGTGTGAGACCACCGGGTGTGGGATGTACAGGGCGGCAAGAAACAATCTAAT  
 35 CATAGTGGCAATCATATGTTTGGTTTTTATAGCGATATTCATCCCACCGTTTCTTGAATGAATTCACTTCCCGATA  
 TTGATCCCCTGTTTTGGAGAAGAAAGTATCAAGCTATTTGAAAAAGTCACTCTGGAAACTTACAGTAAAGAGGA  
 ACGCCGTAGTCCAGGGAACACAACAGGTGACATTGTTTCGCTGGAAGATGTGATAGATCGCGCCTGGTCTGCC  
 GGCGCCAAAGCTTGGGAAGAGCTGGAAATTGCATTACAGACAGGGAGAACATTTTTCGAAGAAGGACAATAATGC  
 CAATGCAACTGCAGATCCATGCCAGCATCACTCTTTACAACAGGAAAGGAATTGGACAATTTAGGAAGGGTCTT  
 40 CCCACTGCCTTGTGGTCTAATGTTTGGATCAGCCATAACTCTCATTGGAAAGCCACGGGAAGCTCACATGGAGT  
 ACAAACCGCCAATCGCCAGAGTTGGGGAAGGTGTCTCTCCATACGTATGGTGTCCCAGTTTATAATGGAGTTA  
 CAGGGTGAAGGTGATAAAAGTGAAGATCTCTAGAACTCCACATAAACCCCTGACTCCGTGGTGCATG  
 GAGCTGGAACCCATCATTGAGCATAATACATGCTATCGAAACCAGTGGGGCCAGCTCATCGGTTGTAAGGTT  
 45 GGCAAGTACCTGAATACGAAGAAACCGTGGACGGTCTTCCCAAGTGCAGAAAGTGGCTTCGAGGGCGATGACAA  
 AAAACCTGCTTCGACCCAAAAATCCTGGTGGCTTGGGCGATTAGTTGGTCAATCCGACAAGGAGACGCTTGAAT  
 GGGAGTATCCATTGTCCGAAGGTCCGGAGTTTGTCTCACCATTTCGAGCAGGTGTAGAAGGATTTCACTTAACTA  
 TTGATGGTCCGCACATCAGTTCCGTTCCCTTATCGTGCGGGTTATGCTATGGAAGAAGCAACAGGAATATCAGTG  
 GCAGGAGACGTCGATGTTCTTTCGATGACAGTAACATCATTACCTTTAACACATCCCAGCTACTACCCTGAGTTG  
 50 GTTTTGGATTCCGGGTGATATCTGGAAGGCACCACCTTTACCAACAGGCAAGATAGAGTTATTTGTTGGAATCATG  
 TCAAGCAGCAATCACTTTGCAGAACGTATGGCAGTAAGAAGACGTGGTTTCAGTCTCTGGTTATCCAATCTCC  
 CAAGCGGTGGCTCGCTTCTTTGTAGCTCTGCATGCAAACAAGGATATCAATCTGCAGCTGAAGAAAGAGGCTGA  
 CTATTACGGCGATATGATAATTTTACCTTTCATCGACAGATATGATATAGTGGTCTTAAAGACCGTTGAAATTTTCA  
 AGTTTGGGGTCCAGAATGTTACAGTTAGCCACGTCATGAAATGTGACGATGACACATTTGTAAGGATTGACAGCG  
 TTCTTGAAGAGATTGGAACGACGTCAGTAGGACAGGGCCTTTACATGGGCAGCATGAATGAGTTTCATAGACCC  
 55 CTTCGTTCTGGGAAGTGGGCCGTGACAGTTGAGGAGTGGCCTGAGCEPGCATTACCCAACATACGCAAATGGT  
 CCAGGATACATCTTTTCGGAAGATATTGTGCATTTTATAGTGGAGGAGAGCAAAGAAATAATTTGAGGTTATTTA  
 AGATGGAGACGTCAGCTAGGTATATGAGTACCGGATATGCAAAGATGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT  
 GTACGGTTTGTCAAGCCGGTTGTATACCTAACTACCTGACAGCGCACTATCAATCGCCGCGTCAAATGCTGTGT  
 60 CTGTGGGACAAGGTGCTTGTACCAATGACGGCAAGTGTGACCTTGT**GA**-3'; los codones de inicio y  
 terminación están indicados en negrilla) se realizó por RCP con ADNc y los cebadores Pp $\beta$ 1-3GalT 2 para (SEC ID  
 nº: 7: 5'-TACGAGCTCATGAAGAGGG-GTGAGACC-3') y el cebador Pp $\beta$ 1-3 GalT 2 rev (SEC ID nº: 8: 5'-  
 GTAGAGCTCTACAAGGTGCAGCACTTG-3'). El producto de ampliación se digirió con Sac I y se clonó en el  
 vector pRT101 digerido con Sac I (Toepfer *et al.* 1987 NAR 15, 5890). La secuencia clonada se verificó por  
 secuenciado.

65 2.1 Creación del montaje transgénico del gen para  $\beta$ 1,3-galactosiltransferasa 1 de *Physcomitrella patens*

El montaje transgénico para la destrucción dirigida del gen para  $\beta$ 1,3-galactosiltransferasa 1 de *Physcomitrella patens* se generó por RCP realizada con ADN genómico procedente de *Physcomitrella patens*. En un cebador MOB1336 de RCP (SEC. ID. nº 9: 5'-TACGGATCCAACCTTCGAGTTCGTGTCTGTA-3' y el cebador MOB1333 (SEC. ID. nº 10: 5'-ACACT**AAGCTT**CCTAATCAATGTCCGGAAGTGAG-3') se utilizaron para ampliar la parte 5' del montaje transgénico. En un segundo cebador MOB1334 de RCP (SEC. ID. nº 11: 5'-TTAG**AAGCTT**AGTGTACGCTGAGT-GTCTACATTG-3') y el cebador MOB1335 (SEC. ID. nº 12: 5'-CATTGTGACCCTACACAGCTCTTAACGCTAC-3') se utilizaron para ampliar el fragmento 3' del montaje transgénico. Ambos montajes ampliados se digirieron con Hind III (las enzimas de restricción se indican en las secuencias de cebador MOB1333 y MOB1334 en letras en negrilla) y se ligaron en una reacción de ligadura posterior utilizando T4 ADN ligasa. La secuencia resultante de ADN ligada y purificada se utilizó como plantilla para una RCP posterior con cebador MOB1336 y MOB1335. El producto de ampliación resultante  $\beta$ 1,3-GalT1 ko (SEC ID nº: 13: 5'CAACTTCGAGTTCGATGTCTGTATGAAGAAGTCCACGGGTTCAATGTGTTAAGACTTAGGCATTTCCCTTCA GCTTTGCGTAGTGGATATGCGTATTTTTGATTGTGAGGATCCGGTTCCTTAGACCATGATTGGTTATTACAG TGGTCATTCAAATCCTATTTGATTTGAGAATGATTTACTTCGTTGTGTTGGGAGATGATTGTTCCCTCGAATTCTA TGCGGTAGCTACCGCTTCTTTGTAATGAAGACCTTTGAAGTTCACATAGACTTCAAGAAGAATGCTATTTGTGTT TTTGTGATTGTGTGTTCAAGTTTGGTGCAGTATTGTTAAATTTGGGTGATGACTAAGTACACTTTATGCGGCCCA AGTAGTCAAGTTGAGCATTGTAAATGCTGAAATGAGTTAGGCTGACGGTAAATGTCTGTGGATGTAGCCTAGTG ATGATTTGATCTCGGCATAATCTTCAGTGATCAATACAAATAATTCAAGAAAGAGGGGTCAATGTGTTCCCTGCGA GTACCTCTCGCAGTGTCAACGTGAAGTGAATTAATTAAGCTGAGCAACATAGACCTTCTTGCTGTTGACAGA TTCAAATTCGGTAATATGAAGAGGGGTCGAGACTCCGATGAGGCTACATGGAGTACAAACCGCCAAAGAAATGATC TTATCCTAGTTGCAATTGTTGCTTGTTTTTTATGGTATATTCATCCACCATATCTCCAAATGAACTCACTTCCG GACATTGATTAG**AAGCTT**AGTGTACGCTGAGTGTCTACATTGTGATTGAATGTTCCCTTAGAATTGTTTGTGTT TATGTTTTTATTTTTATATTTCTGCCGGCTATTGAGGAAGAATACATTCAAATGTTTCAGGATTCCGGACAAGAAATC ATCAAGCTACTCGAAAAAACCACTCTAGAAGCCAATAGTAAGGAGGAACGCCGTAGTCCGGGGAATACCACAG GCGACATTGTTTCTCTGGATGATGTGATAGATCGTGCCTGGTCTGCTGGTGCCAAAGCGTGGGAAGAAGTGGAA ACTGCGTTAAGAAATGGAGAAGGTGTCTCAAAGAATGTCAGTAATGCCACTGCAAAATGCTGATCCGTGTCCAGC ATCACTCTCTGCAGCAGGAAAAAGTTAGACGAATTGGGTAAAGTCTTCCCCTTGCCCTGTGGTCTAATGTTTGG GTCAGCCATTAATCTGATTGGAAAGCCTCGAGAGGCTCACATGGAGTACAAACCGCCAAATCGCCAGAGTTGGGG AAGGCGTCTCCATATGTCATGGTTTCCAGTTCTTAGTAGAGTTACAAGGCTTAAAGGTGGTGAAGGTGAAG ATCCTCCTCGAATTCTACACTTGAATCCTCGACTTCGTGGTATTGGAGCTGGAAACCCATCATTGAGCACAACA CTTGTTATCGGAACCAGTGGGGTCTGCCACCGATGCGAGGGTTGGCAAGTGCCTGAATACGAAGAAACTGG TGAGTGCTGATTCCACCGCACCAGTTTGTGTTTTTATGCTGACACTATGCTTCTCAGGTTTGTAGACGTTAAGAG CTGTGTAGG-3'; la enzima de restricción Hind III se indica en letras en negrilla) comprendía una eliminación de 270 pb con respecto a la secuencia genómica del gen para  $\beta$ 1,3-galactosiltransferasa 1 de *Physcomitrella patens* que además inician un codón de terminación en el fragmento 5' inicial del ADNc correspondiente. Por lo tanto, dando como resultado un gen de  $\beta$ 1,3-galactosiltransferasa disfuncional cuando se integra por recombinación homóloga en el genoma de *Physcomitrella patens*. Este montaje transgénico se utilizó para la transformación de *Physcomitrella patens* sola o en combinación con el montaje  $\beta$ 1,3-GalT2ko transgénico (véase el apartado 2.2).

La identificación de las supuestas plantas transformadas se realizó por RCP utilizando combinaciones de cebadores apropiadas.

## 2.2 Creación del montaje transgénico del gen para $\beta$ 1,3-galactosiltransferasa 2 de *Physcomitrella patens*

El montaje transgénico para la destrucción dirigida del gen para  $\beta$ 1,3-galactosiltransferasa 2 de *Physcomitrella patens* se generó por RCP realizada con ADN genómico procedente de *Physcomitrella patens*. En un cebador MOB1339 de RCP (SEC. ID. nº 14: 5'- TGGCACGATACAGTGGCATGA - 3' y el cebador MOB1337 (SEC. ID. nº 15: 5'- TGGAAATTCATTCAAGAAACGGTGGGATGA - 3') se utilizaron para ampliar la parte 5' del montaje transgénico. En un segundo cebador MOB1338 de RCP (SEC. ID. nº 16: 5' - TGAATTCATAACGAAGACACCGTCTA - 3') y el cebador MOB1313 (SEC. ID. nº 17: 5' - CAAGCAGCGGAGACCTTGCAATGC - 3') se utilizaron para ampliar la parte 3' del montaje transgénico. Ambos montajes ampliados se digirieron con Eco RI (las enzimas de restricción se indican en las secuencias de cebador MOB1337 y MOB1338 en letras en negrilla) y se ligaron en una reacción de ligadura posterior utilizando T4 ADN ligasa. La secuencia de ADN resultante ligada y purificada se utilizó como plantilla para una RCP posterior con cebador MOB1339 y MOB1313. El producto de ampliación resultante  $\beta$ 1,3-GalT2ko (SEC ID nº: 18: 5'TGGCAGCAGTACAGTGGCATGAGATTTATCGCTGCCAAACTGTGGACAATGATGTTTGA AAACAGTCTATTCATCACTGTTGGCAAATCTATGTACAGGGCTAAAAGGGCCAAACTGAGCTTAAACAGCAGTG ATCGAGTTCTTGAGCAGGATCAGCGCAAGGGTAAAGTTGCTTAGGACCGCTTCAACCTGGTGAGTTAGACACT CAAAATAATTACGAAACAGTGACATTTATAAGCTTTGTGTCGCTACTACTTTGAGCCTTCAGAGTACATTTATAGG TGGTGAATTCGTTAATGATGTTAAAAATATGAGGTGAGGACATGTCTTCTTGTGATTAGAGTGATCACTTTGATCC TTTTGCAAACGCTGAAAGGAGTAAGTCTGATTGTCAACAGAAATGTTTTGGTTGACGCTGGCTAATATTATTGG TCTCAGTTCAATTTTCGATGGAGTGGCGTACAAGTATCCAGAAAGCAAGAATCATGGATTTCTACAATTTTCATT TAGATTTTTCATGTTGTTGAGTTATGCTGATTGATTTGGGAAAGAGGGAGCTTAGCGTTGTATACAGGGTCAA ACACCGTAATATGAAGAGGGGTGTGAGACTACCGGGTGTGGGATGTACAGGGCCAAAGAAACAACTAATC ATAGTGCAATCATATGTTTTGTTTTTATAGCGATTTTCATCCACCCTTTCTTGAAT**GAATTC**CATAACGAAGAC ACCGTCTAAAGCTTCACAGTTAGTGCAGAAATGATTGGTTCGCCCTCGCTATGCCAGTCAGGCTTACTGAGTTT

TACTTGGATCGTTTCTACTTGGATCTTTTATGGCTTCCTAGCAGTCGGAGGTTTCTTTCTGGTTTGAAGAAAGCCAT  
 GTATGGAACGTTTACAGGTTTTGGAGAAGAAAGTATCAAGCTATTTGAAAAAGTCACTCTGGAACTTACAGTAA  
 AGAGGAACGCCGTAGTCCAGGGAACACAACAGGTGACATTGTTTCGCTGGAAGATGTGATAGATCGCGCCTGG  
 TCTGCCGCGCCAAAGCTTGGGAAGAGCTGGAAATTGCATTTCAGACAGGGAGAACATTTTTCGAAGAAGGACAA  
 5 TAATGCCAATGCAACTGCAGATCCATGCCAGCATCACTCTTACAACAGGAAAGGAATTGGACAATTTAGGAAG  
 GGTCTTCCCCTGCCTTGTGGTCTAATGTTTGGATCAGCCATAACTCTCATTGGAAAGCCACGGGAAGCTCACAT  
 GGAGTACAAACCGCAATCGCCAGAGTTGGGGAAGGTGTCTCTCCATACGTCATGGTGTCCCAGTTCATAATGG  
 AGTTACAGGGCTTGAAGGTGGTAAAAGGTGAAGATCCTCCTAGAATCCTCCACATAAACCCCTCGACTCCGTGGT  
 10 GACTGGAGCTGGAAACCCATCATTGAGCATAATACATGCTATCGAAACCAGTGGGGCCCAGCTCATCGGTGTGA  
 AGTTTGGCAAGTACCTGAATACGAAGAAACCGGTGAGTGTGGTCCATCACACTTTATCTTTTCATAGTGACAC  
 GGTCTTTTTAGGTGTACTAGTGTGAAAGCTGTGCATGTTAAATGGTAACCCTAATCAATCTTCTCGCTAATTT  
 CGCATTGCAAGGTCTCCGCTGCTTG -3';

la enzima de restricción Eco RI está indicada en letras en negrilla) comprendía una eliminación de 148 pb con  
 15 respecto a la secuencia genómica del gen para  $\beta$ 1,3-galactosiltransferasa 2 de *Physcomitrella patens* que además  
 inician un codón de terminación en la parte 5' inicial del ADNc correspondiente. Por lo tanto, dando como resultado  
 un gen para  $\beta$ 1,3-galactosiltransferasa disfuncional cuando se integra por recombinación homóloga en el genoma de  
*Physcomitrella patens*. Este montaje transgénico se utilizó para la transformación de *Physcomitrella patens* sola o en  
 combinación con el montaje  $\beta$ 1,3-GalT1ko transgénico (véase el apartado 2.1).

La identificación de las supuestas plantas transformadas se realizó por RCP utilizando combinaciones de cebadores  
 apropiadas.

3. Espectrometría de masas MALDI-TOF

Los N-glucanos de la cepa de *Physcomitrella patens* modificada genéticamente con gluco que carecen de los restos  
 de  $\alpha$ 1,3-fucosa y  $\beta$ 1,2-xilosa específicos del vegetal - utilizados en la presente memoria como referencia - presentan  
 las características estructurales típicas de los N-glucanos vegetales procesados en estas cepas tal como se describe  
 en Koprivova *et al.* 2004 *Plant Biotechnol. J.* 2, 517-523); es decir, sin fucosa en el enlace  $\alpha$ 1,3 al GlcNAc unido por  
 30 Asn, y sin xilosa en el enlace  $\beta$ 1,2 al resto  $\beta$ -manosil, epítopos de Lewis a (restos de  $\alpha$ 1,4-fucosil y  $\beta$ 1,3-galactosil  
 unidos a GlcNAc) como elementos terminales no reductores (tab. 1). En cambio no se detectaron epítopos de Lewis  
 a (restos de  $\alpha$ 1,4-fucosil y  $\beta$ 1,3-galactosil unidos a GlcNAc) en los N-glucanos aislados de una cepa de  
*Physcomitrella patens* modificada genéticamente con gluco que además comprendía alteraciones genéticas dirigidas  
 de ambos genes de  $\beta$ 1,3-galactosiltransferasa 1 y  $\beta$ 1,3-galactosiltransferasa 2.

Tabla 1: Estructuras de N-glucano de cepas de *Physcomitrella patens* bitransgénicas y tetratransgénicas. Se  
 aislaron N-glucanos de material vegetal cultivado en las mismas condiciones (matraces de 100 ml., medio de Knop)  
 restos, GF = estructura Lewis a que comprende fucosa y galactosa (unida por  $\beta$ 1,3), Gn = N-acetilglucosamina,  
 M/Man = manosa

	<i>Physcomitrella patens</i> bitransgénicas	<i>Physcomitrella patens</i> tetratransgénicas
	Estructuras de N-glucano que carecen de los restos $\alpha$ -1,3-fucosa y $\beta$ -1,2-xilosa del núcleo	Estructuras de N-glucano que carecen de los restos $\alpha$ -1,3-fucosa, $\beta$ -1,2-xilosa y $\beta$ 1,3-galactosa del núcleo (por consiguiente carecen de epítopos de Lewis a en total)
933	Man3 (MM)	Man3 (MM)
1096	Man4	Man4
1137	MGn/GnM	MGn/GnM
1258	Man5	Man5
1299	Man4Gn	Man4Gn
1340	GnGn	GnGn
1420	Man6	Man6
1582	Man7	Man7
1648	(GF)Gn/Gn(GF)	
1744	Man8	Man8
1907	Man9	Man9
1956	(GF) (GF)	

SEC. ID. nº 1  
 $\beta$ 1,3-GalT1 de ADNc

5' AGTTGTCGATTTGTTGTTTTGATATGTAAGGCGGTTGCCTTCGCGCCGTGCTTGATTGTAAT-  
 TGTAAATCAATCTGGAGTGTGAGATATATATATATATATATAGCGAGAGGGAGAGAGAAAGAGAGAGAGAGG-  
 GAGAGAGAAAGAGAGAGAGAGAGAGAGAGATGGCTTGTGTATGAGGGCCATCGGAGGAGGAGGCTGT-  
 GTTTGTTGCCCGAAGAGATGGGATGGTTTATGTGTAGTGCAGGGGTTGGATGTGAAGCACCTGTTTGAAGGAGTCT-  
 GCGAGAGTTTGAATTCGATTTCAGAGTGCGGCGATCGATGGTGCAACGTTGTTAGCAGTGATTGTTTTCGC-  
 CAACAGAACTGACATCATTGGATTTTTTTACGCGTGGATGTGCCTCTTTTTAAAAAATTTCCGCGTGAANA-  
 GAGACGGGGGTTTGTAAATGGAGGCAGGCTGTGGTCATCACCCCTAGTATAGCCTGTCAAGAGAGTTCAAATTCG-  
 GTAATATGAAGAGGGGGTTCGAGACTACCCGGATATGGCGTGTACAGGGCGCAAAGAAATGATCTTATCCTAGT-  
 TGCAATGTTTGTGTTTATGGTGATATTCATCCACCATATCTCCAATGAACTCACTTCCGGACAT-  
 TGATTCTC  
 CTGATTCCGACAAGAAATCATCAAGCTACTCGAAAAAACCACTCTAGAAGCCATAGTAAGGAGGAACGC-  
 CGTAGTCCGGGAATACCACAGGCGACATTGTTTCTCTGGATGATGTGATAGATCGTGCTGGTCTGCTGGTGC-  
 CAAAGCGTGGGAAGAACTGGAAGTGCCTTAAGAAATGGAGAAGGTCTCAAAGAATGTCAGTAATGCCACT-  
 GCAATGCTGATCCGTGTCCAGCATCACTCTCTGCAGCAGGAAAAAGTTAGACGAATGGGTAAAGTCTTCCCCT-  
 TGCCCTGTGGTCTAATGTTGGGTGAGCCATTACTCTGATGGAAAGCCTCGAGAGGCTCACATGGAGTACAAC-  
 CGCAATCGCCAGAGTTGGGAAGGCGTCTCCATATGTCATGGTTCCAGTCTTAGTAGAGTTACAAGGCT-  
 TAAAGGTGGTGAAGGTGAAGATCCTCCTCGAATTCACACTGAATCCTCGACTTCGTGGTATTGGAGCTG-  
 GAAACCCATCATTGAGCACAACACTTGTATCGGAACAGTGGGGTCTGCCACCAGTGCAGGGTTGGCAAGT-  
 GCCTGAATACGAAGAACTGTTGACGGTCTCCCAAGTGCAGAAAGTGGCTTCGAGATGATGGCAAGAAACCT-  
 GCTTCAACGCAAAAATCTTGGTGGCTTGAAGATTAGTTGGTCTGACAAGGAGACGCTTGAATGGGAGTAC-  
 CCATTATCTGAGGGTCCGGAGTTCGTTCTCACCATTTCGAGCAGGTGTGAAGGGTTTCATGTACTATCGATG-  
 GTCGTACATCAGCTCGTTTCCTTATCGTGTGGTTACGCTGTGGAAGAAACAACGGGATATTAGTAGCAG-  
 GAGACGTTGATGTGATGTCTATCAGATGACATCCCTACCCTAACACATCTAGCTACTACCCTGAGT-  
 TAGTTTTGGAATCGGGGACATTTGGAAGGCACCCTGTCCCAGCTACCAAGATAGATTTATTTATGGGATCAT-  
 GTCACAGCTAACCATTTTCGAGAACGGATGGCAGTAAGGAAGACGTG-  
 GTTCAATCTAAGCTATTCATCTTCGACAGGCGTGGCTCGCTTCTTTGTAGCTCTGCATGCAAACAAGGATA-  
 TCAATATGCAGTTGAAGAAGGAGGAGACTATTATGGCGATATTATAATCCTGCCTTTCATCGACAGATATGATA-  
 TAGTGGTTCTCAAGACCGTTGAAATTTGCAAGTTTGGGGTCCAGAATGTCACAGCTAAGTATATTATGAAGTGT-  
 GACGATGACACTTTTGTGAGGATTGATAGCCTTCTCGAAGAGATTCGAACTACTTCAATATCACAAGGCCTTA-  
 CATGGGTAGCATGAATGAGTTTCACAGCCCTCTCGTTCGAAAGTGGGCGTACTGCCGAGGAATGGCCT-  
 GAGCGAATTTACCCAATATATGCTAATGGACCAGGATATATCCTGTCAGAGGATATTGTGCATTTCAATGTTGGAG-  
 ATGAATGAGAGAGGCAGTTTCAGTTATTTAAGATGGAGGACGTCAGTGTGGAATATGGGTACGCGAATA-  
 TGCGAAGCAAGTGAAGCACGTTCAATACGAACATAGCATAACGGTTTGTCTAAGCCGGTTGTATACCGAAATACT-  
 TGACAGCTCATTACCAATCGCCGCTCAAATGCTGTGTCTGTGGGACAAGGTAAGTCTCATGACGATGGGAAAT-  
 GCTGCAACTTGTGAGGAAAATACATACAATGAATGTGTTCAACGGTCTTTACCAGACAGAATTAATTTGGGTCGG-  
 GAACCAGATATAGCAGACAGCTCAGTTCAATTCAGCCGTGTTGATCCAGAGGGTAATGATAGTTTCCT-  
 TGCCCCCTACCCTCTAGAGGTGGAGATCTTACAACCTAATCAAATGATCCTCTGCAATGTCAGTTGT-  
 CACAATACTTAGTATAGCTCAAAATGGCCACGGATATTCAGGAATGTTTCATCTTGTAAAGTTCGAGCTTGT-  
 GAGTAAATGGTTGGTGGTGTGATGGCATGGTTGCTTATCAATCCCTCTTAGCATCAGTGATCGTCAGAATCAGT-  
 GTTTTCGACACTCCCCGGTGGAGATTTTTTCGATCTCTTGATTCCACTCAAGTGGTACTAGCTTATATTTAGT-  
 GAGGCCCTGGAACCCAGTAGTTAGTTTCAAGTACGCTGCTTTTTGCCGAAATGAGTAGAGTAATTTGTGGCAGTAGT-  
 TGTTGAAGAGACATGGTTAGGATTTAGTCTTCAAATCTG 3'

5

SEC ID n°: 2  
 Ppβ1-3Ga1T2 de ADNc

ATGAAGAGGGGTGTGAGACCACCGGGTGTGGGATGTACAGGGCGGCAAAGAAACAATCTAAT-  
 CATAGTGGCAATCATATGTTTGGTTTTTATAGCGATATTCATCCCACCGTTTCTTGAAAT-  
 GAATTCACTTCCCATATGATTCCCTGTTTTTGGAGAAGAAAGTAT-  
 CAAGCTATTTGAAAAAGTCACTCTGAAACTTACAGTAAAGAGGAACGCCGTAGTCCAGG-  
 GAACACAACAGGTGACATGTTTCGCTGGAAGATGTGATAGATCGCGCTGGTCTGCCGGCGC-  
 CAAAGCTTGGGAAGAGCTGAAATTGCATTGACACAGGGAGAACATTTTTCGAAGAAG-  
 GACAATAATGCCAATGCAACTGCAGATCCATGCCCAGCATCACTCTTACAACAGGAAAG-  
 GAATGGACAATTTAGGAAGGGTCTTCCCCTGCTTGTGGTCTAATGTTTGGATCAGC-  
 CATAACTCTCATTGGAAGCCACGGGAAGCTCACATGGAGTACAAACCGCCAATCGCCAGAGT-  
 TGGGGAAGGTGTCTCTCCATACGTCATGGTGTCCCAGTTCATAATGGAGTTACAGGGCT-  
 TGAAGGTG  
 GTAAAAGGTGAAGATCCTCCTAGAACTCCTCCACATAAACCCCTCGACTCCGTGGTACTG-  
 GAGCTGGAAACCCATCATGAGCATAATACATGCTATCGAAACCAGTGGGGCCAGCTCATCG-  
 GTGTGAAGGTGGCAAGTACCTGAATACGAAGAAACCGTGGACGGTCTTCCAAGTGC-  
 GAGAAGTGGCTTCGAGGCGATGACAAAAACCTGCTTCGACCCAAAAATCCTGGTGGCTTGG-  
 GCGATTAGTTGGTCATTCGACAAGGAGACGCTTGAATGGGAGTATCCATTGTCCGAAG-  
 GTCGGGAGTTGTTCTCACCATTGAGCAGGTGTAGAAGGATTTCACTTAACTATTGATG-  
 GTCGGCACATCAGTTCGTTCCCTTATCGTGGGGTTATGCTATGGAAGAAGCAACAGGAATA-  
 TCAGTGGCAGGAGACGTCGATGTTCTTTCGATGACAGTAACATCATTACCTTTAACA-  
 CATCCAGCTACTACCCTGAGTTGGTTTTGGATTGCGGTGATATCTGGAAGGCAC-  
 CACCTTTACCAACAGGCAAGATAGAGTTATTTGTTGGAATCATGTCAAGCAGCAAT-  
 CACTTTGCAGAACGTATGGCAGTAAGAAAGACGTGGTTTTCACTCTCTGGT-  
 TATCCAATCCTCCAAGCGGTGGCTCGCTTCTTTGTAGCTCTGCATGCAAACAAGGATA-  
 TCAATCTGCAGCTGAAGAAAGAGGCTGACTATTACGGCGATA-  
 TGATAATTTTACCTTTCATCGACAGATATGATATAGTGGTCTTAAGACCGT-  
 TGAATTTTCAAGTTTGGGGTCCAGAATGTTACAGTTAGCCACGTCATGAAATGTGACGAT-  
 GACACATTTGTAAGGATTGACAGCGTCTTGAAGAGATTCGAACGACGTCAGTAGGACAGG-  
 GCCTTTACATGGGACGATGAATGAGTTTCATAGACCCCTTCGTTCTGGGAAGTGGGCCGT-  
 GACAGTTGAGGAGTGGCTGAGCGCATTTACCCAACATACGCAAATGGTCCAGGATA-  
 CATCCTTTCGGAAGATATGTGCATTTTATAGTGGAGGAGAGCAAAGAAATAATTTGAGGT-  
 TATTTAAGATGGAGGACGTCAGCGTAGGTATATGGGTACGCGAGTATGCAAAGAT-  
 GAAGTACGTGCAATACGAGCATAGCGTACGGTTTGCTCAAGCCGGTTGTATACCTAACTACCT-  
 GACAGCGCACTATCAATCGCCGCGTCAAATGCTGTGTCTGTGGGACAAGGTGCTTGCTAC-  
 CAATGACGGCAAGTGTGTCACCTTGTGA

5

SEC ID nº: 3  
 β1-3GalT1 de ADN genómico

5' : AGTTGTCGATTTGTGTTTTTGATATGTAAGGCGGTTGCCTTCGCGCCGTGCTTGATTGTAAT-  
 TGTAATCAATCTGGAGTGTGAGATATATATATATATATATAGCGAGAGGGAGAGAAAGAGAGAGAGG-  
 GAGAGAGAAGAGAGAGAGAGGGAGAGAGAGATGGCTTGTGTATGAGGGCCATGCGAGGAGGAGGCTGT-  
 GTTTGTTGCCGAAGAGATGGGATGTTTATGTGTAGTGCAGGGTTGGATGTGAAGCACCTGTTTGAAGGAGTCT-  
 GCGAGAGTTTGAATTCGATTGAGAGTGCAGGATCGATGGTGCACGTTGTTAGCAGTATTGTTTTCGC-  
 CAACAGAAGTGCATgtaaatgaatagtttcgagggcatgatcgcggtttttctcaatttgaaggggtgtttgtgg-  
 gtgatctatgtgcagaagtgtcactgatggtcagattcgatgcttgacaatttgatcctttgtgagtgatgcagC-  
 ATTTGGATTTTTTTACGCGTGGATGTGCCCTCTTTTTAAAAAATTCGCGTGGAAAAGAGACGGGG-  
 GTTTGTAATGGAGGCAGGCTGTGGTCATACCCCTAGTATAGCCTGTCAAGAGgtgagattgacaccctctttgct-  
 caattgtagattttttccttctcagggct-

10



CAAGATAGATTTATTTATTTGGGATCATGTCCAGCAGTAACCATTTTGCAGAACGGATGGCAGTAAGGAAGACGTG-  
 GTTCAATCTAAAGCTATTCAATCTTCGCAGGCCGTGGCTCGTCTTTTGTAGCTCTGgtacttccctcctat-  
 caaatctcattaactttcgaattattagtgatcatctacataagtggtctgttgattgctgaaagtggtg-  
 tgcgtgcctttgcgtaaatgactttccaaattcatttagaacagtggaaacataatttgtgtgtgcgt-  
 tgcgtatttaacttttccggtgaatgtcttattgaattgtgatgtagCATGCAAACAAGGATATCAATATGCAGT-  
 TGAAGAAGGAGGCAGACTATTATGGCGATATTATAATCCTGCCTTTTCATCGACAGATATGATATAGTGGTTCT-  
 CAAGACCGTTGAAATTTGCAAGTTGGGGgtacgtgtgcgaataatggcttcaaagcttgtgacggtgtct-  
 gcaatttggggatggtgataatgaggcttgataccaactgaaggttaggtgacttttaacactaggttctgct-  
 tactgtgcagTCCGAATGTCACAGCTAAGTATATTATGAAGTGTGACGATGACACTTTTGTGAGGAT-  
 TGATAGCGTTCTCGAAGAGATTCGAACTACTTCAATATCACAAGGCCTTTACATGGGTAGCATGAAT-  
 GAGTTTCACAGCCCTCTCGTTCGAAAGTGGGCCGTGACTGCCGAGgtatttttatttttatttttg-  
 gcttttgcgggaacgtgagagaaaccaagatgaatataatcacgatgtgttttttattgcaaggatttatttg-  
 atgctcttgagaaatctgtgtagccataccactcaattggatactagatgtgttcgtccttatgtataaaaat-  
 gaaacatgtgcttttcaggaagattaattcagtttgacttgtacgtctagttagattgatggtgatgaaacaag-  
 gattatctcgcgaattgacaagtgggtgcttgacagGAATGGCCTGAGCGAATTTACCCAATATATGCTAATG-  
 GACCAGGATATATCCTGTGTCAGAGGATATTGTGCATTTTATTGTGGAGATGAATGAGAGAGGCAGTTTGCAGgtag-  
 gttcttttagaactgtgctgctgctattcacgctctacaagtttaaaaattagaactttcttgttg-  
 gcaatttccatccaggaatcttttgcaccgcaagttcgtaaataggagtcggtacattctgtgtgtg-  
 gcatcgtttgttaaatgcatttttcaattttctttgcttaaaaatctctgttgcgatctcctcatgatct-  
 tgcattgtgaacatgagaagatatgaaatgtgaactcaatattcttctatgatcatgtgcagTTATTTAAGATG-  
 GAGGACGTCAGTGTGGAAATATGGGTACGCGAATATGCGAAGCAAGTGAAGCACGTTCAATACGAA-  
 CATAGCATAACGTTTGTCTCAAGCCGTTGTATACCGAAATACCTGACAGCTCATTACCAATCGCCGCTCAAAT-  
 GCTGTGTCTGTGGACAAGGTAATGCTCATGACGATGGGAAATGCTGCAACTTGTGAGGAAAATACATACAAT-  
 GAATGTGTTCAACGGTCTTTACCAGACAGAAATTACTTTGGGTCGGGAACCAGATATAGCAGACAGCTCA-  
 CATTCAATTCAGCCGTTGTATCCAGAGGGTAATTGATAGTTTCCCTGTCCCTACCCTCTCTAGAGGTGGAG-  
 ATCTTACAACCTTAATCAAATGATCCTCTGCAATGTCACCTGTGCAATACTTAGTATAGCTCAAAAATGGCCACG-  
 GATATTCAGGAATGTTTATCTTGTAAAGTGCAGCTTGTGAGTAAATGGTTGGGTGTTGCGATGCCATGGTTGCT-  
 TATCAATCCCTCTTAGCATCAGTGCATCGTCAGAATCAGTGTTCGACACTCCCCGGTG-  
 GAGTATTTTTTCGATTCTCTTGATCCACTCAAGTGGTACTAGCTTATATTTAGTGAGGCCTGGAACCAAGTAGT-  
 TAGTTCAGTACGTCGCCCTTTGCCGAAATGAGTAGAGTAATTTGTGGCAGTAGTTGGTGAAGAGACATGTTAG-  
GATTTAGTGTTCAAAATCTG 3'

5 SEC ID nº: 4  
 Ppβ1-3GalT2 de ADN genómico

ATGAAGAGGGGTGTGAGACCACCGGTTGGGATGTACAGGGCGGCAAAGAAACAATTAATCATAGTGGCAAT-  
 CATATGTTTGGTTTTTATAGCGATATTATCCACCCTTTCTTGAATGAATCACTTCCCGATATTGATCCCT-  
 gtgtataggttagaaggtattaacttcgcttcacatagacgtcgtatcaagaacaggattcacgtgtcagt-  
 tacagtggtatggacagccagatagccatcaactggtgatgaagacataacgaagacac-  
 cgtctaagcttcacaggttagtgcgaaatgatggttcgcccctcgctatgccagtcaggcttactgagttc-  
 tacttggatcgttctacttggatctttatggtctcctagcagtcggaggttctttctggttgaagaaagccat-  
 gtatggaacgtttacagGTTTGGAGAAGAAAGTATCAAGCTATTGAAAAAAGTCACTCTGGAAACT-  
 TACAGTAAAGAGGAACGCCGTAGTCCAGGGAACACAACAGGTGACATTGTTTCGCTGGAAGATGTGATAG-  
 ATCGCGCTGGTCTGCCGGGCCAAAGCTTGGGAAGAGCTGGAAATGCATTTCAGACAGGGAGAA-  
 CATTTTTCGAAGAAGGACAATAATGCCAATGCAACTGCAGATCCATGCCAGCATCACTCTTTACAACAGGAAAG-  
 GAATTGGACAATTTAGGAAGGGTCTCCCACTGCCTTGTGGTCTAATGTTGGATCAGCCATAACTCTCATTG-  
 GAAAGCCACGGGAAGCTCACATGGAGTACAAACCGCCAAATCGCCAGAGTTGGGGAAGGTGTCTCTCCATACGTCAT-  
 GGTGCCAGTTCATAATGGAGTTACAGGGCTTGAAGGTGGTAAAAGGTGAAGATCCTCCTAGAATCCTCCA-  
 CATAAACCTCGACTCCGTGGTACTGGAGCTGGAAACCCATCATTGAGCATAATACATGCTATCGAAACAGT-





ATGCGAGGAGGAGGCTGTGTTGTGCCCCGAAGAGATGGGATGGTTTATG  
 TGTAGTGCAGGGTTGGATGTGAAGCACCTGTTTGAAGGAGTCTGCGAGA  
 GTTTGAAATTCGGATTCAGAGTGCGGCGATCGATGGTCAACGTTGTTAG  
 CAGTGATTGTTTTCGCCAACAGAACTGACATCATTGGATTTTTTTTACG  
 CGTGGATGTGCCCTCTTTTTAAAAAATTTCCGCGTGGAAAAGAGACGGGG  
 GTTTGTAATGGAGGAGGCTGTGGTCATCACCCCTAGTATAGCCTGTCAA  
 GAGAGTTCAAATTCGGTAATATGAAGAGGGGGTTCGAGACTACCGGATATG  
 GCGTGTACAGGGCGCAAAGAAATGATCTTATCCTAGTTGCAATTGTTTG  
 CTTGTTTTTTATGGTGATATTCATCCACCATACTCCAAATGAACTCAC  
 TTCCGGACATTGATTCTCCTGTCGAGAAGCTAGAAGATGATGATGATGCT  
**GTCTTCACTTCTCATAGACGTCGTAACCAAGAGCAGATTTTCAGTTGTCAC**  
**TGACAGTGGTCAGAGACGGACAGTTATGCCATCTTCGACTGGTCCGGAGG**  
**ACGTAACGAATGCACCCTCTAAAGATTCACAGGATTCGGACAAGAAATCA**  
 TCAAGCTACTCGAAAAAACCACTCTAGAAGCCAATAGTAAGGAGGAACG  
 CCGTAGTCCGGGAATACCACAGGGCAGATTGTTTCTCTGGATGATGTGA  
 TAGATCGTGCCTGGTCTGCTGGTGCCAAAGCGTGGGAAGAACTGGAACT  
 GCGTTAAGAAATGGAGAAGGTGTCTCAAAGAATGTCAGTAATGCCACTGC  
 AAATGCTGATCCGTGTCCAGCATCACTCTCTGCAGCAGGGAAAAAGTTAG  
 ACGAATTGGGTAAAGTCTTCCCCTTGCCCTGTGGTCTAATGTTGGGTCA  
 GCCATTACTCTGATTGGAAAGCCTCGAGAGGCTCACATGGAGTACAAACC  
 GCCAATCGCCAGAGTTGGGAAGGCGTCTCTCCATATGTCATGGTTTCCC  
 AGTTCTTAGTAGAGTTACAAGGCTTAAAGGTGGTGAAGGTGAAGATCCT  
 CCTCGAATTCTACACTTGAATCTCGACTTCGTGGTATTGGAGCTGGAA  
 ACCCATCATTGAGCACAACACTTGTATCGGAACCACTGGGGTCCCTGCC  
 ACCGATGCGAGGGTTGGCAAGTGCCTGAATACGAAGAACTGTTGACGGT  
 CTTCCCAAGTCCGAGAAGTGGCTTCGAGATGATGGCAAGAAACCTGCTTC  
 AACGCAAAAATCTTGGTGGCTTGGAAAGATTAGTTGGTCGTTCTGACAAGG  
 AGACGCTTGAATGGGAGTACCCATTATCTGAGGGTCCGGAGTTCGTTCTC  
 ACCATTGAGCAGGTGTTGAAGGGTTTCATGTGACTATCGATGGTCGTC  
 CATCAGCTCGTTTCTTATCGTGTGGTTACGCTGTGGAAGAAACACGG  
 GGATATTAGTAGCAGGAGACGTTGATGTGATGCTATCACAGTGACATCC  
 CTACCCCTAACACATCCTAGCTACTACCCTGAGTTAGTTTTGGAATCGGG  
 GGACATTTGGAAGGCACCACCTGTCCAGCTACCAAGATAGATTTATTTA  
 TTGGGATCATGTCCAGCAGTAACCATTTTGCAGAACGGATGGCAGTAAGG  
 AAGACGTGGTTTCAATCTAAAGCTATTCAATCTTCGCAGGCCGTGGCTCG  
 CTTCTTTGTAGCTCTGCATGCAAACAAGGATATCAATATGCAGTTGAAGA  
 AGGAGGCAGACTATTATGGCGATATTATAATCCTGCCTTTCATCGACAGA  
 TATGATATAGTGGTTCTCAAGACCGTTGAAATTTGCAAGTTTGGGGTCCA  
 GAATGTCACAGCTAAGTATATTATGAAGTGTGACGATGACACTTTTGTGA  
 GGATTGATAGCGTTCTCGAAGAGATTCGAACTACTTCAATATCACAAGGC  
 CTTTACATGGGTAGCATGAATGAGTTTACAGGCCTCTTCGTTCTGGAAA  
 GTGGGCCGTGACTGCCGAGGAATGGCCTGAGCGAATTTACCCAATATATG  
 CTAATGGACCAGGATATATCCTGTGAGGATATGTGCATTTTCATTGTG  
 GAGATGAATGAGAGAGGCAGTTTGCAGTTATTTAAGATGGAGGACGTCAG  
 TGTGGAAATATGGGTACGCGAATATGCGAAGCAAGTGAAGCACGTTCAAT  
 ACGAACATAGCATAACGTTTGTCTAAGCCGGTTGTATACCGAAATACTTG  
 ACAGCTCATTACCAATCGCCGCGTCAAATGCTGTGTCTGTGGGACAAGGT  
 ACTTGCTCATGACGATGGGAATGCTGCAACTTGTGA

- 5 SEC ID nº: 25  
 Inserción por corte y empalme de 150 nucleótidos de variante de corte y empalme alternativa de β1,3-GalT2 de ADNc representada en negrilla (nt151-300)

**ATGAAGAGGGGTGTGAGACCACCGGGTGTGGGATGTACAGGGCGGCAAAG**  
**AAACAATCTAATCATAGTGGCAATCATATGTTTGGTTTTTATAGCGATAT**  
**TCATCCCACCGTTTCTTGAAATGAATTCACTTCCCGATATTGATTCCCCT**  
**GTGTATAGGTTAGAAGGTATTAACCTCGCTTCACATAGACGTCGCTATCA**  
**AGAACAGGATTCACGTGTACGTTACAGTGGCTATGGACAGCCAGATATGC**  
**CATCAACTGGTGATGAAGACATAACGAAGACACCGTCTAAAGCTTCACAG**  
**GTTTTGGAGAAGAAAGTATCAAGCTATTTGAAAAAGTCACTCTGGAAAC**  
**TTACAGTAAAGAGGAACGCCGTAGTCCAGGGAACACAACAGGTGACATTG**  
**TTTCGCTGGAAGATGTGATAGATCGCGCCTGGTCTGCCGGCGCAAAGCT**  
**TGGGAAGAGCTGGAAATGCATTTCAGACAGGGAGAACATTTTCGAAGAA**  
**GGACAATAATGCCAATGCAACTGCAGATCCATGCCAGCATCACTCTTTA**  
**CAACAGGAAAGGAATTGGACAATTTAGGAAGGGTCTTCCCCTGCCTTGT**  
**GGTCTAATGTTTGGATCAGCCATAACTCTCATTTGAAAGCCACGGGAAGC**  
**TCACATGGAGTACAAACCGCCAATCGCCAGAGTTGGGGAAGGTGTCTCTC**  
**CATACGTATGGTGTCCAGTTCATAATGGAGTTACAGGGCTGAAGGTG**  
**GTAAAAGGTGAAGATCCTCCTAGAATCCTCCACATAAAACCCTCGACTCCG**  
**TGGTGACTGGAGCTGGAAACCCATCATTGAGCATAATACATGCTATCGAA**  
**ACCAGTGGGGCCAGCTCATCGGTGTGAAGGTTGGCAAGTACCTGAATAC**  
**GAAGAAACCGTGGACGCTTCCCAAGTGCAGAGAAGTGGCTTCGAGGCGA**  
**TGACAAAAACCTGCTTCGACCCAAAAATCCTGGTGGCTTGGGCGATTAG**  
**TTGGTCATTCCGACAAGGAGACGCTTGAATGGGAGTATCCATTGTCCGAA**  
**GGTCGGGAGTTTGTCTCACCATTTCGAGCAGGTGTAGAAGGATTTCACTT**  
**AACTATTGATGGTCGGCACATCAGTTCGTTCCCTTATCGTGGGGTTATG**  
**CTATGGAAGAAGCAACAGGAATATCAGTGGCAGGAGCGTCGATGTTCTT**  
**TCGATGACAGTAACATCATTACCTTTAACACATCCAGCTACTACCCTGA**  
**GTTGGTTTTGGATTTCGGGTGATATCTGGAAGGCCACCCTTTACCAACAG**  
**GCAAAGATAGAGTTATTTGTTGGAATCATGTCAAGCAGCAATCACTTTGCA**  
**GAACGTATGGCAGTAAGAAAGACGTGGTTTCAGTCTCTGGTTATCCAATC**  
**CTCCCAAGCGGTGGCTCGCTTCTTTGTAGCTCTGCATGCAAACAAGGATA**  
**TCAATCTGCAGCTGAAGAAAGAGGCTGACTATTACGGCGATATGATAATT**  
**TTACCTTTCATCGACAGATATGATATAGTGGTTCTTAAGACCGTTGAAAT**  
**TTTCAAGTTTGGGGTCCAGAATGTTACAGTTAGCCACGTCATGAAATGTG**  
**ACGATGACACATTTGTAAGGATTGACAGCGTTCCTGAAGAGATTCGAACG**  
**ACGTCAGTAGGACAGGGCCTTACATGGGCAGCATGAATGAGTTTCATAG**  
**ACCCCTTCGTTCTGGGAAGTGGGCCGTGACAGTTGAGGAGTGGCCTGAGC**  
**GCATTTACCCAACATACGCAAATGGTCCAGGATACATCTTTTCGGAAGAT**  
**ATTGTGATTTTATAGTGGAGGAGAGCAAAGAAATAATTTGAGGTTATT**  
**TAAGATGGAGGACGTCAGCGTAGGTATATGGGTACCGAGTATGCAAAGA**  
**TGAAGTACGTGCAATACGAGCATAGCGTACGGTTTGTCTAAGCCGGTTGT**  
**ATACCTAACTACCTGACAGCGCACTATCAATCGCCGCGTCAAATGCTGTG**  
**TCTGTGGGACAAGGTGCTTGCTACCAATGACGGCAAGTCTGCACCTTGT**  
**GA**

5

**Listado de secuencias**

- <110> Greenovation Biotech GmbH
- <120> Galactosiltransferasa
- <130> r51014
- <150> EP 06450139.8
- <151> 2006-09-29
- <160> 27

<170> PatentIn versión 3.4

<210> 1

<211> 2957

5 <212> ADN

<213> *Physcomitrella patens*

<220>

<221> misc\_feature

10 <222> (432)..(432)

<223> n e s a , c , g , o t

<400> 1

	agttgtcgat ttgttgtttt tgatatgtaa ggcggttgcc ttcgcccgt gcttgattgt	60
	aattgtaatt caatctggag tgtgagatat atatatatat atatatatag cgagaggag	120
	agagaaagag agagagaggg agagagaaag agagagagag ggagagagag agatggcttg	180
	tgtatgaggg ccatcgagg aggaggctgt gtttgttgcc cgaagagatg ggatggttta	240
	tgtgtagtgc aggggttggg tgtgaagcac ctgtttgaag gagtctgcga gagtttga	300
	ttcggattca gagtgcggcg atcgatggg caacgttgtt agcagtgatt gtttcgcca	360
15	acagaactga catcatttgg attttttta cgcgtggatg tgcctcttt ttaaaaaatt	420
	tcccggtgga anagagacgg gggttttaa tggaggcagg ctgtggtcat caccctagt	480
	atagcctgtc aagagagttc aaattcggta atatgaagag ggggtcgaga ctaccgata	540
	tggcgtgtac agggcgcaa agaaatgac ttatcctagt tgcaattggt tgcttgttt	600
	ttatggtgat attcattcca ccatatctcc aaatgaactc acttccggac attgattctc	660
	ctgattcgga caaaaaatca tcaagctact cgaaaaaac cactctagaa gccaatagta	720
	aggaggaacg ccgtagtccg ggaataacca caggcgacat tgtttctctg gatgatgtga	780
	tagatcgtgc ctggtctgct ggtgccaaag cgtggaaga actggaact cgtttaagaa	840
	atggagaagg tgtctcaaag aatgtcagta atgccactgc aaatgctgat ccgtgtccag	900
	catcactctc tgcagcaggg aaaaagttag acgaattggg taaagtctt ccttgcct	960
	gtggtctaata gtttgggtca gccattactc tgattggaaa gcctcgagag gctcacatgg	1020
	agtacaaacc gccaatcgcc agagtgggg aaggcgtctc tccatagtgc atggtttccc	1080
	agttcttagt agagttacaa ggcttaaagg tggtaaaag tgaagatcct cctcgaattc	1140
	tacacttgaa tcctcgactt cgtggtgatt ggagctggaa acccatcatt gagcacaaca	1200
	cttgttatcg gaaccagtgg ggtctgccc accgatgcga gggttggcaa gtgcctgaat	1260
	acgaagaaac tgttgacggt cttccaagt gcgagaagtg gcttcgagat gatggcaaga	1320
	aacctgctc aacgaaaaa tcttgggtggc ttggaagatt agttggctgt tctgacaagg	1380
	agacgcttga atgggagtc ccatatctg agggtcggga gttcgttctc accattcgag	1440
	cagggttga agggtttcat gtgactatcg atggctgca catcagctcg tttccttacc	1500
	gtgtgggtta cgtgtggaa gaaacaacgg ggatattagt agcaggagac gttgatgtga	1560
	tgctatcac agtgacatcc ctacccttaa cacatcctag ctactaccct gagttagttt	1620
	tggaatcggg ggacatttgg aaggcaccac ctgtcccagc taccaagata gatttattta	1680
	ttgggatcat gtccagcagt aaccattttg cagaacggat ggcagtaagg aagacgtggt	1740
	ttcaatctaa agctattcaa tcttcgagg cctgtgctcg cttctttgta gctctgcatg	1800
	caaacaagga tatcaatatg cagtgaaga aggaggcaga ctattatggc gatattataa	1860
	tcctgccttt catcgacaga tatgatatag tggttctcaa gaccgttga atttgcaagt	1920
	ttgggtcca gaatgcaca gctaagtata ttatgaagtg tgacgatgac acttttga	1980
	ggattgatag cgttctcgaa gagattcgaa ctacttcaat atcacaaggc ctttaccatg	2040
	gtagcatgaa tgagtttacc aggcctcttc gttctggaaa gtgggcccgt actgccgagg	2100
	aatggcctga gcgaatttac ccaatatatg ctaatggacc aggatatact ctgtcagagg	2160

atattgtgca tttcattgtg gagatgaatg agagaggcag tttgcagtta ttaaatgatgg 2220  
 aggacgtcag tgttgaata tgggtacgag aatatgcgaa gcaagtgaag cacgttcaat 2280  
 acgaacatag catacggttt gctcaagccg gttgtatacc gaaatacttg acagctcatt 2340  
 accaatcgcc gcgcaaatg ctgtgtctgt gggacaaggt acttgctcat gacgatggga 2400  
 aatgctgcaa cttgtgagga aaatacatac aatgaatgtg ttcaacggtc tttaccagac 2460  
 agaattactt tgggtcggga accagatata gcagacagct cacattcaat tcagccgtgt 2520  
 tgatccagag gggtaattga tagtttcctt gtcctacc ctctctagag gtggagatct 2580  
 tacaacttaa tcaaatgatc ctctgcaatg tcactgtgca caatacttag tatagctcaa 2640  
 aattggccac ggatattcag gaatgttcat cttgtaaggt cgcagcttgt gagtaaatgg 2700  
 ttgggtggtg tcgatggcat ggttgcttat caatccctct tagcatcagt gatcgtcaga 2760  
 atcagtgttt tcgacactcc ccggtggagt attttttcga ttctcttgat tccactcaag 2820  
 tggactagc ttatatattg tgaggcctgg aaccaagta gttagtccag tacgtctgcc 2880  
 ttttgccgaa atgagtagag taatttgtgg cagtgttgg tgaagagaca tggtaggat 2940  
 ttagtgttca aaatctg 2957

<210> 2  
 <211> 1902  
 <212> ADN  
 <213> *Physcomitrella patens*

<400> 2

10 atgaagaggg gtgtgagacc accgggtgtg ggatgtacag ggcggcaaag aaacaatcta 60  
 atcatagtgg caatcatatg tttggttttt atagcgatat tcacccacc gtttctttaa 120  
 atgaattcac ttcccgatat tgattcccct gttttggaga agaaagtatc aagctatttg 180  
 aaaaaagtca ctctggaaac ttacagtaa gaggaacgcc gtagtccagg gaacacaaca 240  
 ggtgacattg tttcgtgga agatgtgata gatcgcgct ggtctgccgg cgccaaagct 300  
 tgggaagagc tggaaattgc attcagacag ggagaacatt tttcgaagaa ggacaataat 360  
 gccaatgcaa ctgcagatcc atgccagca tcactcttta caacaggaaa ggaattggac 420  
 aatttaggaa gggcttccc actgccttgt ggtctaattg ttggatcagc cataactctc 480  
 attgaaagc cacgggaagc tcacatggag tacaaccgc caatcgccag agttggggaa 540  
 ggtgtctctc catacgtcat ggtgtcccag ttcataatgg agttacaggg cttgaagggtg 600  
 gtaaaagtg aagatcctc tagaatcctc cacataaacc ctgcactccg tggtgactgg 660  
 agctggaaac ccatcattga gcataataca tgctatcga accagtgggg cccagctcat 720  
 cgggtggaag gttgcaagt acctgaatac gaagaaaccg tggacggctt tccaagtgc 780  
 gagaagtggc ttcgagcga tgacaaaaa cctgcttcca ccaaaaatc ctggtggctt 840  
 gggcgattag tgggtcattc cgacaaggag acgcttgaat gggagtatcc attgtccgaa 900  
 ggtcgggagt ttgttctcac cattcgagca ggtgtagaag gatttcactt aactattgat 960  
 ggtcggcaca tcagttcgtt cccttatcgt gcgggttatg ctatggaaga agcaacagga 1020  
 atatcagtgg caggagacgt cgatgttctt tcgatgacag taacatcatt accttaaca 1080  
 catcccagct actaccctga gttggtttt gattcgggtg atatctggaa ggcaccacct 1140  
 ttaccaacag gcaagataga gttatttgtt ggaatcatgt caagcagca tcactttgca 1200  
 gaacgtatgg cagtaagaaa gacgtggtt cagtctctgg ttatccaatc ctccaagcg 1260  
 gtggctcgtc tctttgtagc tctgcatgca aacaaggata tcaatctgca gctgaagaaa 1320  
 gaggctgact attacggcga tatgataatt ttaccttca tcgacagata tgatatagtg 1380  
 gttcttaaga ccggtgaaat tttcaagttt ggggtccaga atgttacagt tagccacgct 1440  
 atgaaatgtg acgatgacac atttgtaagg attgacagc ttcttgaaga gattcgaacg 1500  
 acgtcagtag gacagggcct ttacatgggc agcatgaatg agtttcatag accccttctg 1560  
 15 tctgggaagt gggccgtgac agttgaggag tggcctgagc gcatttacc aacatacgca 1620

aatggtccag gatacatcct ttcggaagat attgtgcatt ttatagtgga ggagagcaaa 1680  
 agaaataatt tgaggttatt taagatggag gacgtcagcg taggtatatg ggtacgag 1740  
 tatgcaaaga tgaagtacgt gcaatacagag catagcgtac ggtttgctca agccggttg 1800  
 atacctaact acctgacagc gcactatcaa tcgccgcgtc aaatgctgtg tctgtgggac 1860  
 aaggtgcttg ctaccaatga cggcaagtgc tgcaccttgt ga 1902

<210> 3  
 <211> 6187  
 <212> ADN  
 <213> *Physcomitrella patens*

<400> 3

5  
 10  
 15

agttgtcgat ttgttgtttt tgatatgtaa ggcggttgcc ttcgcgccgt gcttgattgt 60  
 aattgtaatt caatctggag tgtgagatat atatataat atatataatag cgagagggag 120  
 agagaaagag agagagaggg agagagaaag agagagagag ggagagagag agatggcttg 180  
 tgtatgaggg ccatcagagg aggaggctgt gtttgttgcc cgaagagatg ggatggttta 240  
 tgtgtagtgc aggggttggg tgtgaagcac ctgtttgaag gagtctgcga gagtttgaaa 300  
 ttcggattca gagtgcggcg atcgtatggtg caacgttgtt agcagtgatt gttttcgcca 360  
 acagaactga catgtaatga atagtctcga ggcattgatc cggtttttct caatttgaag 420  
 gggttgtttg tgggtgatct atgtgcagaa gtgtcactga tggtcagatt cgatgcttga 480  
 caatttgatc ctttgtgagt gtgcagcatt tggatttttt ttacgcgtgg atgtgccctc 540  
 ttttataaaa atttccgcgt gaaaaagaga cgggggtttg taatggaggc aggtctgtgtg 600  
 catcaccctt agtatagcct gtcaagaggt gagattgaca cctcttttgc tcaattgtag 660  
 atttttttcc ttctcagggc tgaatcccag tttttttttt tttttttttt tttttttcct 720  
 tcttcttcaa cttcaggttc gtgtctgtat gaagaagtcc acgggttcaa tgtgttaaga 780  
 cttaggcatt tcttcagct ttgcctagtg gagatatgag tattttttga ttgtgaggat 840  
 tccggttctt agaccatgat tggtttatta cagtggatcat tcaaatccta tttgatttga 900  
 gaatgtattt acttcgttgt gttgggagat gattgttccc tcgaattcta tgcggtagct 960  
 accgcttctt tcgtaatgaa gacctttgaa gttcacatag acttcaagaa gaatgctatt 1020  
 tgtgtttttg tgattgtgtg ttcaagtttg gtgcagtatt gttaaaattt ggggtgatgac 1080  
 taagtacact ttatgcggcc caagtagtca agttgagcat ttgtaaatgc tgaatgagt 1140  
 taggctgacg gtaaatgtct gtggatgtag cctagtgatg tatttgatct cggcataatc 1200  
 ttcagtgatc aatacaata attcaagaaa gaggggtcaa tgtgttcctg cgagtacctt 1260  
 cgcattgtca acgtgaactg aattatgtta attaagctga gcaacataga ccttcttgct 1320  
 gttgacagag ttcaaatcag gtaatatgaa gaggggttcg agactaccgg atatggcgtg 1380  
 tacagggcgg caaagaaatg atcttctcct agttgcaatt gtttgcctgt tttttatggt 1440  
 gatattcatc ccaccatc tccaaatgaa ctcaactccg gacattgatt ctctgtcga 1500  
 gaagctagaa gatgatgatg atgctgtcct cacttctcat agacgtcgtg accaagagca 1560  
 gatttcagtt gtcaactgaca gtggctcagag acggacagtt atgccatctt cgactggtgc 1620  
 ggaggacgta acgaatgcac cgtcctaaaga ttcacaggtt agacaaaag tagttgacct 1680  
 gaaatgcatg tgtaaatcaa gcactcttgt ccttattcga gcttttattt cttgccatca 1740  
 ggtattttta atacttccct agtgtacgct gagtgtctac attgtgtatt gaatgttctt 1800  
 tagaattggt tgtttgttta tgtttttatt tttatattt tgcggctat tgaggaagaa 1860  
 tacattcaaa ttgttcagga ttcggacaag aaatcatcaa gctactcga aaaaaccact 1920  
 ctagaagcca atagtaagga ggaacgccgt agtccgggga ataccacagg cgacattgtt 1980

tctctggatg atgtgataga tctgctctgg tctgctgggtg ccaaagcgtg ggaagaactg 2040  
gaaactgctg taagaatgg agaaggtgtc tcaaagaatg tcagtaatgc cactgcaaat 2100  
gctgatccgt gtccagcatc actctctgca gcagggaaaa agttagacga attgggtaaa 2160  
gtcttcccct tgcctctggt tctaattgtt gggtcagcca ttactctgat tggaaagcct 2220  
cgagaggctc acatggagta caaacgcca atcgccagag ttggggaagg cgtctctcca 2280  
tatgtcatgg tttcccagtt cttagtagag ttacaaggct taaaggtggt gaaaggtgaa 2340  
gatcctcctc gaattctaca cttgaatcct cgacttctgt gtgattggag ctggaaccc 2400  
atcattgagc acaacacttg ttatcggaac cagtgggtc ctgccaccg atgcgaggg 2460  
tggcaagtgc ctgaatacga agaactggt gagtctgat tccaccgac cagtttgtgt 2520  
ttttatgct gacactatgc ttctcaggt ttagacgtt aagagctgtg taggttccgt 2580  
ggtacttcca attggcactt gccacttctc tcattgtaag ttgtaaatg tctgcatgag 2640  
caataaatc caacactgga tgtgtatttt ctgaaatgat tctgtttctt gtagttgacg 2700  
gtcttcccaa gtgcgagaag tggcttcgag atgatggcaa gaaacctgct tcaacgcaa 2760  
aatcttgggt gcttgaaga ttagttggtc gttctgaca ggagacgctt gaatgggagt 2820  
accattatc tgagggtcgg gagttcgttc tcaccattcg agcaggtgtt gaagggtttc 2880  
atgtgactat cgatggtcgt cacatcagct cgttccctta tctgtgtgta agttgaaaat 2940  
gctatgtaa catataatgc taaagttgac ctcatgtctt tctttttct tttttctt 3000  
tttatttct ggagggggg ggggtaatgc aaatcaactc taaaatttga gtataccagt 3060  
taaattatc atttcaaata taacaataca aatacacatc ttttaattt gtatttttg 3120  
atccctctc tccttacta aaattaataa tatagcaaca ttttggtact acgaaagttc 3180  
atttgtattg cttcatgtc aagatttatt caaaatttct atccctctg tttctgaatt 3240  
acattatcaa caatggaata acaataatga cggcccatc cttcagacac caggaacatt 3300  
acataacca gactacgtc gggtaagtct gaagaattaa ttataacaa gaaactagt 3360  
gtattcactg ttttctttt tacgccatg cgatttatc aagtcttct caatttctta 3420  
ttattcttct ttatttttt aagtttttaa ttattttta agcaacgaat tgataaataa 3480  
ataacatatt aatgtttta actttaaagt tttttcccg tatttagtat aagatttctg 3540  
caaaacgatt aggtgattag atcgaacatt atctaattgc acttactta tatgatatga 3600  
agagtaattt ctcttagcag aagctacatc ctgctattc cttgggaaac ccgattaggt 3660  
ctttcaaatc accctgctt cctctataag tgtaccatga ttgaggttcg ttagggcatt 3720  
agtttaaggg tatcgttggt atgtgtgtct agttagtctt aaaatctgtg caaatcgatt 3780  
cattaacaac tctttctgt agtgtttgt tttgagaact gctatttatc ttccattgtg 3840  
5 cagggttacg ctgtggaaga aacaacgggg atattagtag caggagacgt tgatgtgatg 3900  
tctatcacag tgacatccct acccttaaca catcctagct actaccctga gttagtttg 3960  
gaatcggggg acatttgaa ggcaccacct gtcccagcta ccaagataga tttatttatt 4020  
gggatcatgt ccagcagtaa ccattttgca gaacggatgg cagtaaggaa gacgtggtt 4080  
caatctaaag ctattcaatc ttcgcaggcc gtggctcgtc tctttgtagc tctggtactt 4140  
cctctatca aatctcatta actttcgaat tattagtgat catctacata agtggctgt 4200  
tgattgtga aagggtgctg ttgctgcct ttgctaatg actttccaaa ttcatttaga 4260  
acagtggaaa cataatttgt gtgttcggt gcgtatttaa cttttcggg gaatgtctta 4320  
ttgaattgtg atgtagcatg caacaagga tatcaatag cagttgaaga aggaggcaga 4380  
ctattatggc gatattataa tctgccttt catcgacaga tatgatatag tggttctcaa 4440  
gaccgtgaa atttgcaagt ttgggtacg tgtgtcgaat aatggcttca aagctttgtg 4500  
acggtgtctg caatttggg atggtgataa tgaggcttga taccaactga aggttaggtg 4560  
actttaaca ctagttctg cttactgtgc aggtccagaa tgtcacagct aagtatatta 4620  
tgaagtgtga cgatgacact tttgtgagga ttgatagcgt tctcgaagag attcgaacta 4680  
cttcaatc acaaggcctt tacatgggta gcatgaatga gtttcacagg ccttctcgtt 4740  
ctggaaagtg ggccgtgact gccgaggtat ttttattttt atttttggct tttgtcggga 4800

acgtgagaga aaccaagatg aatataatca cgatgttgtt ttttattgca aggatttatt 4860  
 tgatgctctt gagaaatctg tggtagccat accactcaat ttggatacta gatgtgttcg 4920  
 tccttatgta taaaaatgaa acatgtgctt ttcaggaaga ttaattcagt ttgacttgta 4980  
 cgtctagtta gattgatggt gatgaaacaa gaggattatc tcgcaattg acaagtgggt 5040  
 tgcttggaca ggaatggcct gagcgaattt acccaatata tgctaatagga ccaggatata 5100  
 tcctgtcaga ggatatttg catttcattg tggagatgaa tgagagaggc agtttgcagg 5160  
 taggttcttt tagaactgtg tcgtcgctat tacacgtcta caagttttaa aaattagaaa 5220  
 ctttcttgtt ggcaaattt catccaggaa tctttttgca ccgcaagttc gtaataggag 5280  
 tcggtacatt ctgtgtgtgt gcatcgtttg ttaaatgcat ttttcaattt tcttttgctt 5340  
 aaaatatctc tgttgcgat atctcctcat gatccttgc tgtgaacatg agaagatatg 5400  
 aaatgtgaac tcaatattct tctatgatca tgtgcagtta ttaaatgag aggacgtcag 5460  
 tgttggaaata tgggtacgcg aatatgcaa gcaagtgaag cacgttcaat acgaacatag 5520  
 catacggttt gctcaagccg gttgtatacc gaaatacttg acagctcatt accaatcgc 5580  
 gcgtcaaatg ctgtgtctgt gggacaagg acttgctcat gacgatggga aatgctgcaa 5640  
 cttgtgagga aaatacatac aatgaatgtg ttcaacggtc tttaccagac agaattactt 5700  
 tgggtcggga accagatata gcagacagct cacattcaat tcagccgtgt tgatccagag 5760  
 gggtaattga tagtttctt gtcctctacc ctctctagag tggagatct tacaacttaa 5820  
 tcaaatgatc ctctgcaatg tcacttgc caatacttag tatagctcaa aattggccac 5880  
 ggatattcag gaatgttcat ctgttaaggt cgcagcttgt gagtaaatgg ttgggtggtg 5940  
 tcgatggcat ggttgcctt caatcctct tagcatcagt gatcgtcaga atcagtgttt 6000  
 tcgacactcc ccggtggagt attttttgcg ttctcttgat tccactcaag tggacttagc 6060  
 ttatatttag tgaggcctgg aacccaagta gttagttagc tacgtctgcc ttttgccgaa 6120  
 atgagtagag taatttggc cagtagttgg tgaagagaca tggtaggat ttagtgttca 6180  
 aaatctg . 6187

5

<210> 4  
 <211> 4087  
 <212> ADN  
 <213> *Physcomitrella patens*

10

<400> 4

atgaagaggg gtgtgagacc accgggtgtg ggatgtacag ggcggcaag aaacaatcta 60  
 atcatagtgg caatcatatg tttggtttt atagcgatat tcatcccacc gtttctttaa 120  
 atgaattcac tccccgatat tgattcccct gtgtataggt tagaaggat taacttcgct 180  
 tcatatagac gtcgctatca agaacaggat tcacgtgtca gttacagtgg ctatggacag 240  
 ccagatatgc catcaactgg tgatgaagac ataacgaaga caccgtctaa agcttcacag 300  
 gttagtgcag aaatgattgg ttcgccctcg ctatgccagt caggcttact gagttctact 360  
 tggatcgttc tacttggatc ttttatggct tcctagcagt cggaggtttc tttctggttt 420  
 gaagaaagcc atgtatggaa cgtttacagg ttttgagaa gaagtatca agctatttga 480  
 aaaaagtcac tctgaaact tacagtaaag aggaacgcc tagtccaggg aacacaacag 540  
 gtgacattgt ttcgctggaa gatgtgatag atcgcgcctg gtctgccggc gccaaagctt 600  
 gggagagct ggaattgca ttcagacagg gagaacattt ttcgaagaag gacaataatg 660  
 ccaatgcaac tgcagatcca tgcccagcat cactctttac aacaggaaag gaattggaca 720  
 atttaggaag ggtcttccca ctgccttgg gtctaattgt tggatcagcc ataactctca 780  
 ttgaaagcc acgggaagct cacatggagt acaaacgcc aatcgcaga gttggggaag 840  
 gtgtctctcc atacgtcatg gtgtcccagt tcataatgga gttacagggc ttgaaggtgg 900  
 taaaaggtga agatcctcct agaatcctcc acataaaccc tcgactccgt ggtgactgga 960  
 gctgaaacc catcattgag cataatacat gctatcgaac ccagtggggc ccagctcatc 1020  
 ggtgtgaagg ttggcaagta cctgaatacg aagaaaccgg tgagtgtctg ttccatcaca 1080

15



ctttatcttt tcatagtgac acggttcttt ttaggtgtac tagtggtgaa agctgtgcat 1140  
 gttaaatggt aaccctaatac aatcttctcg ctaattttcg cattgcaagg tctccgctgc 1200  
 ttggacaatc agcactctaa cattggctgt atttactgaa atgattcctt actttgtagt 1260  
 ggacggtcct cccaagtgcg agaagtggct tcgaggcgat gacaaaaaac ctgcttcgac 1320  
 ccaaaaaatcc tggtagcttg ggcgattagt tggcattcc gacaaggaga cgcttgaatg 1380  
 ggagtatcca ttgtccgaag gtcgggagtt tgttctcacc attcgagcag gtgtagaagg 1440  
 atttactta actattgatg gtcggcacat cagttcgctt ccttatcgtg cggtaggttg 1500  
 aaaatactag tttgatattc aatgatgagg tttaccgag gtatatttgg tctcattgtc 1560  
 aagtgtgtgt gtgtgtgttg ttttctttt ttccttttca ttttctgaat cataatgata 1620  
 agaaatcaat tctatgaaac ttacgctcaa tttttaaag ttttattgtt tttgtttgtt 1680  
 tttatttttt tgtgttttgt gtttgtgtt tatttcacaa tacaatgta acaatggaat 1740  
 agaacaatg atggtccac ctcacagaca ccaggtagac tacctacacc agactgcgtc 1800  
 tgagtaagtt taagaacag caaccacaa caatctgatt gtaaatctta aattccttct 1860  
 ccaccagaaa accatgtgat ccgtcttga gttctgcttg cactctacct atatgatcca 1920  
 aagagtaatt cctcttaaca ggagttataa cctgctgggg ttttgaat accgatgagt 1980  
 tcaaatgta aacaacccc ggatctattt caagggtatg aagggttag ctttgtttaa 2040  
 gaataaggtc aagagtatct gtgtgtgtg catcccaaaa tggatgcaa tttgttaatt 2100  
 ggcaactgtt tctgtgtgta tgtttgtga cgcactattt attgtgtatt gtgcagggtt 2160  
 atgctatgga agaagcaaca ggaatatcag tggcaggaga cgctgatgtt ctttcatgta 2220  
 cagtaacatc attaccttta acacatcca gctactacc tgagtgtgtt ttggattcgg 2280  
 gtgatatctg gaaggacca cctttacca caggcaagat agagtattt gttggaatca 2340  
 tgtcaagcag caatcactt gcagaacgta tggcagtaag aaagacgtg tttcagtctc 2400  
 tggttatcca atcctccaa gcggtggctc gcttcttgt agctctgta cttgtcatta 2460  
 tactcttttt tctgtccaag tatctgtaac tcgggaatat ttaaaaagt caaacaaca 2520  
 gtgagctgtt aattgtgaa aattggtgtt ataagtctg atgcagtgc cttccagatt 2580  
 gaccaagtat atcagacctt agaatttga cagcactact tacttaccat ttttaatgaa 2640  
 tcccttgttg ggttgtgatg cagcatgcaa acaaggatat caatctgcag ctgaagaaag 2700  
 aggctgacta ttacggcgat atgataattt tacctttcat cgacagatat gatatagtgg 2760  
 5 tcttaagac cgttgaat tcaagttt gggaagcga attaaaatt gtagtattta 2820  
 caaagtaata ttttaaacg ttgtgaggac atctgcaact tgatatattt ctttctgag 2880  
 gttcgtgct gattaaagct taggtgattt aaaagcacgg tgttcttgc tatgcaggtc 2940  
 cagaatgta cagttagcca cgtcatgaaa tgtgacgat acacatttgt aaggattgac 3000  
 agcgttcttg aagagattcg aacgacgca gtaggacagg gcctttacat gggcagcatg 3060  
 aatgagttc atagaccct tcgttctggg aagtgggccc tgacagttga ggaattttc 3120  
 cctgtacaa attatccaag atttctgtaa ccattgtgtg ccttattcat ttcttctgaa 3180  
 atctcaagaa aaatgaaaa tgcttgagaa acgctcgtag ccgtatcaca ttatgcgaat 3240  
 tccaaaaaag aatgtggaac aaaagtctt gtgaaaataa ttgatattt caaattgtac 3300  
 acatttatgc actaagataa gatattgtca aatagtgcct tccagtgtc tagaaaatgc 3360  
 ttgtttttt ttggaagctt taactttatt tagcttgaac atcttgtttg agggttggtg 3420  
 accaagtaag aaggtccata caagacaata aatggattgg ttcgtgcatg tacaggagtg 3480  
 gcctgagcgc atttaccxaa catacgcaa tggccagga tacatcctt cggaagatat 3540  
 tgtgcatttt atagtggagg agagcaaaag aaataattg aggggtcgtt tttcatagct 3600  
 gtgtcctggt gattaaatgc cccatgttca acattgaaac cttcatctg gacagttttc 3660  
 catcatgta tctcctgtca ttataattgc attatagaac tgttcgctg tacatttctt 3720  
 tcctgttctt ctttttctt tttttttct cttctttct tcatttact ctcctctgt 3780  
 cgatgctttc tgttgacctt atattgtgga tatgtatctt ttcagtacta cggagacgat 3840

ES 2 366 975 T3

	atgaacata agtttgatat tcttctgtga taaagcgag ttatttaaga tggaggacgt	3900
	cagcgtaggt atatgggtac gcgagtatgc aaagatgaag tacgtgcaat acgagcatag	3960
	cgtacggttt gctcaagccg gttgtatacc taactacctg acagcgcaat atcaatcgcc	4020
	gcgtcaaatg ctgtgtctgt gggacaaggt gcttgctacc aatgacggca agtgctgcac	4080
	cttgtga	4087
	<210> 5	
	<211> 18	
5	<212> ADN	
	<213> Artificial	
	<220>	
	<223> cebador	
10	<400> 5	
	ctgaatatcc gtggcaa	18
15	<210> 6	
	<211> 29	
	<212> ADN	
	<213> Artificial	
20	<220>	
	<223> cebador	
	<400> 6	
25	ttcgagctca tgaagagggg gtcgagact	29
	<210> 7	
	<211> 29	
	<212> ADN	
30	<213> Artificial	
	<220>	
	<223> cebador	
35	<400> 7	
	tacgagctca tgaagagggg tgtgagacc	29
	<210> 8	
40	<211> 28	
	<212> ADN	
	<213> Artificial	
	<220>	
45	<223> cebador	
	<400> 8	
	gtagagctct cacaaggtgc agcacttg	28
50	<210> 9	
	<211> 30	
	<212> ADN	
	<213> Artificial	
55	<220>	
	<223> cebador	
	<400> 9	
60		

	tacggatcca acttcgagtt cggtctgta	30
	<210> 10	
	<211> 33	
5	<212> ADN	
	<213> Artificial	
	<220>	
	<223> cebador	
10	<400> 10	
	acactaagct tctaatacat gtccggaagt gag	33
15	<210> 11	
	<211> 34	
	<212> ADN	
	<213> Artificial	
20	<220>	
	<223> cebador	
	<400> 11	
25	ttagaagctt agtgtacgct gagtcttac attg	34
	<210> 12	
	<211> 33	
	<212> ADN	
30	<213> Artificial	
	<220>	
	<223> cebador	
35	<400> 12	
	cattgtcgac cctacacagc tcttaacgtc tac	33
	<210> 13	
40	<211> 1585	
	<212> ADN	
	<213> Artificial	
	<220>	
45	<223> Gen modificado genéticamente GalT	
	<400> 13	
	<b>caacttcgag ttcgtgtctg tatgaagaag tccacgggtt caatgtgtta agacttaggc 60</b>	
	<b>atttccttca gctttgccta gtggagatat gcgtatTTTT tgattgtgag gattccgggtt 120</b>	
	<b>cttagaccat gattggttta ttacagtgggt cattcaaatc ctatttgatt tgagaatgta 180</b>	
	<b>tttacttcgt tgtgttggga gatgattggt ccctcgaatt ctatgcggtta gctaccgctt 240</b>	
	<b>ctttcgtaat gaagaccttt gaagttcaca tagacttcaa gaagaatgct atttgtgttt 300</b>	
	<b>ttgtgattgt gtgttcaagt ttggtgcagt attgttaaaa tttgggtgat gactaagtac 360</b>	
	<b>actttatgcy gcccaagtag tcaagttgag catttgtaa tgctgaaatg agttaggctg 420</b>	
	<b>acggtaaatg tctgtggatg tagcctagtg atgtatttga tctcggcata atcttcagtg 480</b>	
	<b>atcaatacaa ataattcaag aaagaggggt caatgtgttc ctgaggtac ctctcgatgt 540</b>	
	<b>tcaacgtgaa ctgaattatg ttaattaagc tgagcaacat agaccttctt gctgttgaca 600</b>	
	<b>gagttcaaat tcggaatat gaagaggggg tcgagactac cggatatggc gtgtacaggg 660</b>	
50	<b>cggcaaagaa atgatcttat cctagttgca attgtttctg tgttttttat ggtgatattc 720</b>	

atcccacat atctccaaat gaactcactt ccggacattg attagaagct tagtgtacgc 780  
 tgagtgtcta cattgtgat tgaatgttcc ttagaattgt ttgtttgttt atgtttttat 840  
 ttttatattt ctgccggcta ttgaggaaga atacattcaa attgttcagg attcggacaa 900  
 gaaatcatca agctactcga aaaaaaccac tctagaagcc aatagtaagg aggaacgccg 960  
 tagtccgggg aataccacag gcgacattgt ttctctggat gatgtgatag atcgtgcctg 1020  
 gtctgctggt gccaaagcgt gggagaact ggaaactcgc ttaagaaatg gagaagggtg 1080  
 ctcaaagaat gtcagtaatg ccaactgcaa tgctgatccg tgtccagcat cactctctgc 1140  
 agcagggaaa aagttagacg aattgggtaa agtcttcccc ttgccctgtg gtctaattgt 1200  
 tgggtcagcc attactctga ttgaaaagcc tcgagaggct cacatggagt acaaaccgcc 1260  
 aatcgccaga gttggggaag gcgtctctcc atatgtcatg gtttcccagt tcttagtaga 1320  
 gttacaaggc ttaaagggtg tgaaagggtga agatcctcct cgaattctac acttgaatcc 1380  
 tcgacttcgt ggtgattgga gctggaaacc catcattgag cacaacactt gttatcggaa 1440  
 ccagtggggc cctgccacc gatgcgaggg ttggcaagtg cctgaatcag aagaaactgg 1500  
 tgagtgtgta ttccaccgca ccagtttgtg tttttatgc tgacactatg cttctcaggt 1560  
 ttgtagactg taagagctgt gtagg 1585

5 <210> 14  
 <211> 21  
 <212> ADN  
 <213> Artificial

10 <220>  
 <223> cebador  
 <400> 14

15 tggcacgata cagtggcatg a

21

<210> 15  
 <211> 29  
 <212> ADN  
 <213> Artificial

20 <220>  
 <223> cebador

25 <400> 15

tggaattcat tcaagaaacg gtgggatga

29

<210> 16  
 <211> 27  
 <212> ADN  
 <213> Artificial

35 <220>  
 <223> cebador  
 <400> 16

tgaattccat aacgaagaca ccgtcta

27

<210> 17  
 <211> 24  
 <212> ADN  
 <213> Artificial

45 <220>  
 <223> cebador

<400> 17

caagcagcgg agacctgca atgc

24

- 5 <210> 18
- <211> 1656
- <212> ADN
- <213> Artificial

- 10 <220>
- <223> Gen modificado genéticamente GalT

<400> 18

15	<p>                     tggcacgata cagtgggcatg agatttatcg ctgccaact gtggacaatg atgtttgaaa 60                      cagtctattc atcactgggt ggcaaattct atgtacaggg ctaaagggc caaactaggc 120                      ttaacagcag tgatcgaggt tcttgagcag gatcagcgcagggttaagggt tgcttaggac 180                      cgcttcaacc tggtgagtta gacactcaa ataattacga aacagtgaca ttataagct 240                      ttgtgtcgtc actactttga gccttcagag tacatttata ggtggtgact tcgtaaatga 300                      tgttaaaaat atgaggtagg gacatgtcct cttgtgatta gagtgatcac ttgatcctt 360                      ttgcaaacgc tgaaggaggt aagtctgatt gtcaacagaa atgtttttgg ttgcagcctg 420                      gctaataatta ttggtctcag ttcaatttcc gatggagtg cgtacaagt atccagaaag 480                      caagaatcat ggatttccta caatttcatt tagattttcg atgttggtg agttatgctg 540                      attgatattg gaaagagggg gcttagcggt gtatacaggg ttcaaacc gtaatatgaa 600                      gaggggtgtg agaccaccgg gtgtgggatg tacagggcgg caaagaaaca atctaatac 660                      agtggaatc atatgtttgg tttttatagc gatattcacc ccaccgttcc ttgaatgaat 720                      tccataacga agaccaccgc taaagcttca cagggttagtg cagaaatgat tggttcgccc 780                      tcgctatgcc agtcaggctt actgagttct acttggatcg ttctacttgg atcttttatg 840                      gcttcttagc agtcggaggt ttctttctgg tttgaagaaa gccatgatg gaacgtttac 900                      aggtttttga gaagaaagta tcaagctatt tgaaaaaagt cactctggaa acttacagta 960                      aagaggaacg ccgtagtcca gggaaacaaa cagggtgacat tgtttcgctg gaagatgtga 1020                      tagatcgcgc ctggtctgcc ggcgccaag cttgggaaga gctggaaatt gcattcagac 1080                      agggagaaca tttttcgaag aaggacaata atgccaatgc aactgcagat ccattgccag 1140                      catcactctt tacaacagga aaggaattgg acaatttagg aagggtcttc ccaactgcctt 1200                      gtggtctaata gtttgatca gccataactc tcattggaaa gccacgggaa gctcacatgg 1260                      agtacaacc gccaatcgc agagttgggg aagggtgtctc tccatacgtc atggtgtccc 1320                      agttcataat ggagttacag ggcttgaagg tggtaaaagg tgaagatcct cctagaatcc 1380                      tccacataaa ccctcgactc cgtggtgact ggagctggaa accatcatt gagcataata 1440                      catgctatcg aaaccagtgg ggccagctc atcgggtgta aggttgcaaa gtacctgaat 1500                      acgaagaaac cggtgagtg tggttccatc acactttatc ttttcatagt gacacggttc 1560                      tttttagggt tactagtgtt gaaagctgtg catgttaaat ggtaacccta atcaatcttc 1620                      tcgctaattt tcgattgca aggtctccgc tgcttg 1656                 </p>
----	--

- 20 <210> 19
- <211> 634
- <212> PRT
- <213> *Physcomitrella patens*

- 25 <400> 19

Met Lys Arg Gly Ser Arg Leu Pro Asp Met Ala Cys Thr Gly Arg Gln  
 1 5 10 15

Arg Asn Asp Leu Ile Leu Val Ala Ile Val Cys Leu Phe Phe Met Val  
 20 25 30

Ile Phe Ile Pro Pro Tyr Leu Gln Met Asn Ser Leu Pro Asp Ile Asp  
 35 40 45

Ser Pro Asp Ser Asp Lys Lys Ser Ser Ser Tyr Ser Lys Lys Thr Thr  
 50 55 60

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

Gly Asp Ile Val Ser Leu Asp Asp Val Ile Asp Arg Ala Trp Ser Ala  
 85 90 95

Gly Ala Lys Ala Trp Glu Glu Leu Glu Thr Ala Leu Arg Asn Gly Glu  
 100 105 110

Gly Val Ser Lys Asn Val Ser Asn Ala Thr Ala Asn Ala Asp Pro Cys  
 115 120 125

Pro Ala Ser Leu Ser Ala Ala Gly Lys Lys Leu Asp Glu Leu Gly Lys  
 130 135 140

Val Phe Pro Leu Pro Cys Gly Leu Met Phe Gly Ser Ala Ile Thr Leu  
 145 150 155 160

Ile Gly Lys Pro Arg Glu Ala His Met Glu Tyr Lys Pro Pro Ile Ala  
 165 170 175

Arg Val Gly Glu Gly Val Ser Pro Tyr Val Met Val Ser Gln Phe Leu  
 180 185 190

Val Glu Leu Gln Gly Leu Lys Val Val Lys Gly Glu Asp Pro Pro Arg  
 195 200 205

Ile Leu His Leu Asn Pro Arg Leu Arg Gly Asp Trp Ser Trp Lys Pro  
 210 215 220

Ile Ile Glu His Asn Thr Cys Tyr Arg Asn Gln Trp Gly Pro Ala His  
 225 230 235 240

Arg Cys Glu Gly Trp Gln Val Pro Glu Tyr Glu Glu Thr Val Asp Gly  
 245 250 255

Leu Pro Lys Cys Glu Lys Trp Leu Arg Asp Asp Gly Lys Lys Pro Ala  
 260 265 270

Ser Thr Gln Lys Ser Trp Trp Leu Gly Arg Leu Val Gly Arg Ser Asp  
 275 280 285

Lys Glu Thr Leu Glu Trp Glu Tyr Pro Leu Ser Glu Gly Arg Glu Phe  
 290 295 300

Val Leu Thr Ile Arg Ala Gly Val Glu Gly Phe His Val Thr Ile Asp  
 305 310 315 320

Gly Arg His Ile Ser Ser Phe Pro Tyr Arg Val Gly Tyr Ala Val Glu  
 325 330 335

Glu Thr Thr Gly Ile Leu Val Ala Gly Asp Val Asp Val Met Ser Ile  
 340 345 350

5

Thr Val Thr Ser Leu Pro Leu Thr His Pro Ser Tyr Tyr Pro Glu Leu  
355 360 365

Val Leu Glu Ser Gly Asp Ile Trp Lys Ala Pro Pro Val Pro Ala Thr  
370 375 380

Lys Ile Asp Leu Phe Ile Gly Ile Met Ser Ser Ser Asn His Phe Ala  
385 390 395 400

Glu Arg Met Ala Val Arg Lys Thr Trp Phe Gln Ser Lys Ala Ile Gln  
405 410 415

Ser Ser Gln Ala Val Ala Arg Phe Phe Val Ala Leu His Ala Asn Lys  
420 425 430

Asp Ile Asn Met Gln Leu Lys Lys Glu Ala Asp Tyr Tyr Gly Asp Ile  
435 440 445

Ile Ile Leu Pro Phe Ile Asp Arg Tyr Asp Ile Val Val Leu Lys Thr  
450 455 460

Val Glu Ile Cys Lys Phe Gly Val Gln Asn Val Thr Ala Lys Tyr Ile  
465 470 475 480

Met Lys Cys Asp Asp Thr Phe Val Arg Ile Asp Ser Val Leu Glu  
485 490 495

Glu Ile Arg Thr Thr Ser Ile Ser Gln Gly Leu Tyr Met Gly Ser Met  
500 505 510

Asn Glu Phe His Arg Pro Leu Arg Ser Gly Lys Trp Ala Val Thr Ala  
515 520 525

Glu Glu Trp Pro Glu Arg Ile Tyr Pro Ile Tyr Ala Asn Gly Pro Gly  
530 535 540

Tyr Ile Leu Ser Glu Asp Ile Val His Phe Ile Val Glu Met Asn Glu  
545 550 555 560

Arg Gly Ser Leu Gln Leu Phe Lys Met Glu Asp Val Ser Val Gly Ile  
565 570 575

Trp Val Arg Glu Tyr Ala Lys Gln Val Lys His Val Gln Tyr Glu His  
580 585 590

5 Ser Ile Arg Phe Ala Gln Ala Gly Cys Ile Pro Lys Tyr Leu Thr Ala  
595 600 605

His Tyr Gln Ser Pro Arg Gln Met Leu Cys Leu Trp Asp Lys Val Leu  
610 615 620

Ala His Asp Asp Gly Lys Cys Cys Asn Leu  
625 630

<210> 20

<211> 633

10 <212> PRT

<213> *Physcomitrella patens*

<400> 20

Met Lys Arg Gly Val Arg Pro Pro Gly Val Gly Cys Thr Gly Arg Gln  
1 5 10 15

Arg Asn' Asn Leu Ile Ile Val Ala Ile Ile Cys Leu Val Phe Ile Ala  
20 25 30

Ile Phe Ile Pro Pro Phe Leu Glu Met Asn Ser Leu Pro Asp Ile Asp  
35 40 45

Ser Pro Val Leu Glu Lys Lys Val Ser Ser Tyr Leu Lys Lys Val Thr  
50 55 60

15

Leu Glu Thr Tyr Ser Lys Glu Glu Arg Arg Ser Pro Gly Asn Thr Thr  
 65 70 75 80  
 Gly Asp Ile Val Ser Leu Glu Asp Val Ile Asp Arg Ala Trp Ser Ala  
 85 90 95  
 Gly Ala Lys Ala Trp Glu Glu Leu Glu Ile Ala Phe Arg Gln Gly Glu  
 100 105 110  
 His Phe Ser Lys Lys Asp Asn Asn Ala Asn Ala Thr Ala Asp Pro Cys  
 115 120 125  
 Pro Ala Ser Leu Phe Thr Thr Gly Lys Glu Leu Asp Asn Leu Gly Arg  
 130 135 140  
 Val Phe Pro Leu Pro Cys Gly Leu Met Phe Gly Ser Ala Ile Thr Leu  
 145 150 155 160  
 Ile Gly Lys Pro Arg Glu Ala His Met Glu Tyr Lys Pro Pro Ile Ala  
 165 170 175  
 Arg Val Gly Glu Gly Val Ser Pro Tyr Val Met Val Ser Gln Phe Ile  
 180 185 190  
 Met Glu Leu Gln Gly Leu Lys Val Val Lys Gly Glu Asp Pro Pro Arg  
 195 200 205  
 Ile Leu His Ile Asn Pro Arg Leu Arg Gly Asp Trp Ser Trp Lys Pro  
 210 215 220  
 Ile Ile Glu His Asn Thr Cys Tyr Arg Asn Gln Trp Gly Pro Ala His  
 225 230 235 240  
 Arg Cys Glu Gly Trp Gln Val Pro Glu Tyr Glu Glu Thr Val Asp Gly  
 245 250 255  
 Leu Pro Lys Cys Glu Lys Trp Leu Arg Gly Asp Asp Lys Lys Pro Ala  
 260 265 270  
 Ser Thr Gln Lys Ser Trp Trp Leu Gly Arg Leu Val Gly His Ser Asp  
 275 280 285  
 Lys Glu Thr Leu Glu Trp Glu Tyr Pro Leu Ser Glu Gly Arg Glu Phe  
 290 295 300  
 Val Leu Thr Ile Arg Ala Gly Val Glu Gly Phe His Leu Thr Ile Asp  
 305 310 315 320  
 Gly Arg His Ile Ser Ser Phe Pro Tyr Arg Ala Gly Tyr Ala Met Glu  
 325 330 335  
 Glu Ala Thr Gly Ile Ser Val Ala Gly Asp Val Asp Val Leu Ser Met  
 340 345 350  
 Thr Val Thr Ser Leu Pro Leu Thr His Pro Ser Tyr Tyr Pro Glu Leu  
 355 360 365  
 Val Leu Asp Ser Gly Asp Ile Trp Lys Ala Pro Pro Leu Pro Thr Gly  
 370 375 380  
 Lys Ile Glu Leu Phe Val Gly Ile Met Ser Ser Ser Asn His Phe Ala  
 385 390 395 400  
 Glu Arg Met Ala Val Arg Lys Thr Trp Phe Gln Ser Leu Val Ile Gln  
 405 410 415  
 Ser Ser Gln Ala Val Ala Arg Phe Phe Val Ala Leu His Ala Asn Lys  
 420 425 430  
 Asp Ile Asn Leu Gln Leu Lys Lys Glu Ala Asp Tyr Tyr Gly Asp Met  
 435 440 445  
 Ile Ile Leu Pro Phe Ile Asp Arg Tyr Asp Ile Val Val Leu Lys Thr

5



450                      455                      460

Val Glu Ile Phe Lys Phe Gly Val Gln Asn Val Thr Val Ser His Val  
 465                      470                      475                      480

Met Lys Cys Asp Asp Thr Phe Val Arg Ile Asp Ser Val Leu Glu  
                                  485                      490                      495

Glu Ile Arg Thr Thr Ser Val Gly Gln Gly Leu Tyr Met Gly Ser Met  
                                  500                      505                      510

Asn Glu Phe His Arg Pro Leu Arg Ser Gly Lys Trp Ala Val Thr Val  
                                  515                      520                      525

Glu Glu Trp Pro Glu Arg Ile Tyr Pro Thr Tyr Ala Asn Gly Pro Gly  
                                  530                      535                      540

Tyr Ile Leu Ser Glu Asp Ile Val His Phe Ile Val Glu Glu Ser Lys  
                                  545                      550                      555                      560

Arg Asn Asn Leu Arg Leu Phe Lys Met Glu Asp Val Ser Val Gly Ile  
                                  565                      570                      575

Trp Val Arg Glu Tyr Ala Lys Met Lys Tyr Val Gln Tyr Glu His Ser  
                                  580                      585                      590

Val Arg Phe Ala Gln Ala Gly Cys Ile Pro Asn Tyr Leu Thr Ala His  
                                  595                      600                      605

Tyr Gln Ser Pro Arg Gln Met Leu Cys Leu Trp Asp Lys Val Leu Ala  
                                  610                      615                      620

Thr Asn Asp Gly Lys Cys Cys Thr Leu  
                                  625                      630

5

<210> 21  
 <211> 422  
 <212> PRT  
 <213> homo sapiens

10

<400> 21

Met Leu Gln Trp Arg Arg Arg His Cys Cys Phe Ala Lys Met Thr Trp  
 1                      5                      10                      15

Asn Ala Lys Arg Ser Leu Phe Arg Thr His Leu Ile Gly Val Leu Ser  
                                  20                      25                      30

Leu Val Phe Leu Phe Ala Met Phe Leu Phe Phe Asn His His Asp Trp  
                                  35                      40                      45

Leu Pro Gly Arg Ala Gly Phe Lys Glu Asn Pro Val Thr Tyr Thr Phe  
                                  50                      55                      60

Arg Gly Phe Arg Ser Thr Lys Ser Glu Thr Asn His Ser Ser Leu Arg  
                                  65                      70                      75                      80

Asn Ile Trp Lys Glu Thr Val Pro Gln Thr Leu Arg Pro Gln Thr Ala  
                                  85                      90                      95

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

Glu Asn Thr Leu Ser Ala Asn Gly Ser Ile Tyr Asn Glu Lys Gly Thr  
                                  115                      120                      125

Gly His Pro Asn Ser Tyr His Phe Lys Tyr Ile Ile Asn Glu Pro Glu  
 130 135 140

Lys Cys Gln Glu Lys Ser Pro Phe Leu Ile Leu Leu Ile Ala Ala Glu  
 145 150 155 160

Pro Gly Gln Ile Glu Ala Arg Arg Ala Ile Arg Gln Thr Trp Gly Asn  
 165 170 175

Glu Ser Leu Ala Pro Gly Ile Gln Ile Thr Arg Ile Phe Leu Leu Gly  
 180 185 190

Leu Ser Ile Lys Leu Asn Gly Tyr Leu Gln Arg Ala Ile Leu Glu Glu  
 195 200 205

Ser Arg Gln Tyr His Asp Ile Ile Gln Gln Glu Tyr Leu Asp Thr Tyr  
 210 215 220

Tyr Asn Leu Thr Ile Lys Thr Leu Met Gly Met Asn Trp Val Ala Thr  
 225 230 235 240

Tyr Cys Pro His Ile Pro Tyr Val Met Lys Thr Asp Ser Asp Met Phe  
 245 250 255

Val Asn Thr Glu Tyr Leu Ile Asn Lys Leu Leu Lys Pro Asp Leu Pro  
 260 265 270

Pro Arg His Asn Tyr Phe Thr Gly Tyr Leu Met Arg Gly Tyr Ala Pro  
 275 280 285

Asn Arg Asn Lys Asp Ser Lys Trp Tyr Met Pro Pro Asp Leu Tyr Pro  
 290 295 300

Ser Glu Arg Tyr Pro Val Phe Cys Ser Gly Thr Gly Tyr Val Phe Ser  
 305 310 315 320

Gly Asp Leu Ala Glu Lys Ile Phe Lys Val Ser Leu Gly Ile Arg Arg  
 325 330 335

Leu His Leu Glu Asp Val Tyr Val Gly Ile Cys Leu Ala Lys Leu Arg  
 340 345 350

Ile Asp Pro Val Pro Pro Pro Asn Glu Phe Val Phe Asn His Trp Arg  
 355 360 365

Val Ser Tyr Ser Ser Cys Lys Tyr Ser His Leu Ile Thr Ser His Gln  
 370 375 380

Phe Gln Pro Ser Glu Leu Ile Lys Tyr Trp Asn His Leu Gln Gln Asn  
 385 390 395 400

Lys His Asn Ala Cys Ala Asn Ala Ala Lys Glu Lys Ala Gly Arg Tyr  
 405 410 415

Arg His Arg Lys Leu His  
 420

5

10

<210> 22  
 <211> 621  
 <212> PRT  
 <213> *oryza sativa*  
 <400> 22

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

Pro Leu Leu Ile Val His His Leu Ile Val Asn Ser Pro Val Ser Gly  
 20 25 30

Pro Ser Arg Tyr Gln Val Ile His Ser Asn Leu Leu Gly Trp Leu Ser  
 35 40 45

Asp Ser Leu Gly Asn Ser Val Ala Gln Asn Pro Asp Asn Thr Pro Val  
 50 55 60

Glu Val Ile Pro Ala Asp Ala Ser Ala Ser Asn Ser Ser Asp Ser Gly  
 65 70 75 80

Asn Ser Ser Leu Glu Gly Phe Gln Trp Leu Asn Thr Trp Asn His Met  
 85 90 95

Lys Gln Leu Thr Asn Ile Ser Asp Gly Leu Pro His Ala Asn Glu Ala  
 100 105 110

Ile Asp Asn Ala Arg Thr Ala Trp Glu Asn Leu Thr Ile Ser Val His  
 115 120 125

Asn Ser Thr Ser Lys Gln Ile Lys Lys Glu Arg Gln Cys Pro Tyr Ser  
 130 135 140

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

Asp Ile Pro Cys Gly Leu Ile Val Gly Ser Ser Val Thr Ile Ile Gly  
 165 170 175

Thr Pro Gly Ser Leu Ser Gly Asn Phe Arg Ile Asp Leu Val Gly Thr  
 180 185 190

Glu Leu Pro Gly Gly Ser Gly Lys Pro Ile Val Leu His Tyr Asp Val  
 195 200 205

Arg Leu Thr Ser Asp Glu Leu Thr Gly Gly Pro Val Ile Val Gln Asn  
 210 215 220

Ala Phe Thr Ala Ser Asn Gly Trp Gly Tyr Glu Asp Arg Cys Pro Cys  
 225 230 235 240

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

Ser Met Val Gly Arg Glu Glu Lys Arg Ala Ile Asn Ser Lys Gln His  
 260 265 270

Leu Asn Ala Lys Lys Asp Glu His Pro Ser Thr Tyr Phe Pro Phe Lys  
 275 280 285

Gln Gly His Leu Ala Ile Ser Thr Leu Arg Ile Gly Leu Glu Gly Ile  
 290 295 300

His Met Thr Val Asp Gly Lys His Val Thr Ser Phe Pro Tyr Lys Ala  
 305 310 315 320

Gly Leu Glu Ala Trp Phe Val Thr Glu Val Gly Val Ser Gly Asp Phe  
 325 330 335

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

5

Glu Asn Ser Phe Asp Leu Ala Met Leu Lys Ser Ser Pro Ile Pro Glu  
355 360 365

Gly Lys Asp Val Asp Leu Leu Ile Gly Ile Phe Ser Thr Ala Asn Asn  
370 375 380

Phe Lys Arg Arg Met Ala Ile Arg Arg Thr Trp Met Gln Tyr Asp Ala  
385 390 395 400

Val Arg Glu Gly Ala Val Val Val Arg Phe Phe Val Gly Leu His Thr  
405 410 415

Asn Leu Ile Val Asn Lys Glu Leu Trp Asn Glu Ala Arg Thr Tyr Gly  
420 425 430

Asp Ile Gln Val Leu Pro Phe Val Asp Tyr Tyr Ser Leu Ile Thr Trp  
435 440 445

Lys Thr Leu Ala Ile Cys Ile Tyr Gly Thr Gly Ala Val Ser Ala Lys  
450 455 460

Tyr Leu Met Lys Thr Asp Asp Asp Ala Phe Val Arg Val Asp Glu Ile  
465 470 475 480

His Ser Ser Val Lys Gln Leu Asn Val Ser His Gly Leu Leu Tyr Gly  
485 490 495

Arg Ile Asn Ser Asp Ser Gly Pro His Arg Asn Pro Glu Ser Lys Trp  
500 505 510

Tyr Ile Ser Pro Glu Glu Trp Pro Glu Glu Lys Tyr Pro Pro Trp Ala  
515 520 525

His Gly Pro Gly Tyr Val Val Ser Gln Asp Ile Ala Lys Glu Ile Asn  
530 535 540

Ser Trp Tyr Glu Thr Ser His Leu Lys Met Phe Lys Leu Glu Asp Val  
545 550 555 560

Ala Met Gly Ile Trp Ile Ala Glu Met Lys Lys Gly Gly Leu Pro Val  
565 570 575

Gln Tyr Lys Thr Asp Glu Arg Ile Asn Ser Asp Gly Cys Asn Asp Gly  
580 585 590

Cys Ile Val Ala His Tyr Gln Glu Pro Arg His Met Leu Cys Met Trp  
595 600 605

5 Glu Lys Leu Leu Arg Thr Asn Gln Ala Thr Cys Cys Asn  
610 615 620

<210> 23

<211> 643

<212> PRT

10 <213> *Arabidopsis thaliana*

<400> 23

Met Lys Arg Phe Tyr Gly Gly Leu Leu Val Val Ser Met Cys Met Phe  
1 5 10 15

Leu Thr Val Tyr Arg Tyr Val Asp Leu Asn Thr Pro Val Glu Lys Pro  
20 25 30

Tyr Ile Thr Ala Ala Ala Ser Val Val Val Thr Pro Asn Thr Thr Leu  
35 40 45

Pro Met Glu Trp Leu Arg Ile Thr Leu Pro Asp Phe Met Lys Glu Ala  
50 55 60

15 Arg Asn Thr Gln Glu Ala Ile Ser Gly Asp Asp Ile Ala Val Val Ser  
65 70 75 80

Gly Leu Phe Val<sub>85</sub> Glu Gln Asn Val Ser Lys<sub>90</sub> Glu Glu Arg Glu Pro Leu<sub>95</sub>  
 Leu Thr Trp Asn<sub>100</sub> Arg Leu Glu Ser Leu Val<sub>105</sub> Asp Asn Ala Gln Ser Leu<sub>110</sub>  
 Val Asn Gly<sub>115</sub> Val Asp Ala Ile Lys<sub>120</sub> Glu Ala Gly Ile Val<sub>125</sub> Trp Glu Ser  
 Leu Val<sub>130</sub> Ser Ala Val Glu Ala<sub>135</sub> Lys Lys Leu Val Asp Val<sub>140</sub> Asn Glu Asn  
 Gln Thr Arg Lys Gly<sub>150</sub> Lys Glu Glu Leu Cys Pro<sub>155</sub> Gln Phe Leu Ser Lys<sub>160</sub>  
 Met Asn Ala Thr<sub>165</sub> Glu Ala Asp Gly Ser Ser<sub>170</sub> Leu Lys Leu Gln Ile Pro<sub>175</sub>  
 Cys Gly Leu Thr<sub>180</sub> Gln Gly Ser Ser Ile Thr Val Ile Gly Ile Pro Asp<sub>190</sub>  
 Gly Leu Val<sub>195</sub> Gly Ser Phe Arg Ile<sub>200</sub> Asp Leu Thr Gly Gln<sub>205</sub> Pro Leu Pro  
 Gly Glu Pro Asp Pro Pro Ile<sub>215</sub> Ile Val His Tyr Asn Val<sub>220</sub> Arg Leu Leu  
 Gly Asp Lys Ser Thr<sub>230</sub> Glu Asp Pro Val Ile Val<sub>235</sub> Gln Asn Ser Trp Thr<sub>240</sub>  
 Ala Ser Gln Asp Trp<sub>245</sub> Gly Ala Glu Glu Arg Cys Pro Lys Phe Asp Pro<sub>255</sub>  
 Asp Met Asn Lys<sub>260</sub> Lys Val Asp Asp Leu Asp<sub>265</sub> Glu Cys Asn Lys Met Val<sub>270</sub>  
 Gly Gly Glu Ile Asn Arg Thr Ser<sub>280</sub> Ser Thr Ser Leu Gln Ser Asn Thr<sub>285</sub>  
 Ser Arg Gly Val Pro Val Ala<sub>295</sub> Arg Glu Ala Ser Lys His Glu Lys Tyr<sub>300</sub>  
 Phe Pro Phe Lys Gln Gly<sub>310</sub> Phe Leu Ser Val Ala Thr Leu Arg Val Gly<sub>320</sub>  
 Thr Glu Gly Met Gln Met Thr Val Asp Gly<sub>330</sub> Lys His Ile Thr Ser Phe<sub>335</sub>  
 Ala Phe Arg Asp Thr Leu Glu Pro Trp<sub>345</sub> Leu Val Ser Glu Ile Arg Ile<sub>350</sub>  
 Thr Gly Asp Phe Arg Leu Ile Ser Ile Leu Ala Ser Gly Leu Pro Thr<sub>365</sub>  
 Ser Glu Glu Ser Glu His Val Val Asp Leu Glu Ala Leu Lys Ser Pro<sub>370</sub>  
 Thr Leu Ser Pro Leu Arg Pro Leu Asp Leu Val Ile Gly Val Phe Ser<sub>385</sub>  
 Thr Ala Asn Asn Phe<sub>405</sub> Lys Arg Arg Met Ala Val Arg Arg Thr Trp Met<sub>415</sub>  
 Gln Tyr Asp Asp Val Arg Ser Gly Arg Val Ala Val Arg Phe Phe Val<sub>420</sub>  
 Gly Leu His Lys Ser Pro Leu Val Asn Leu Glu Leu Trp Asn Glu Ala<sub>435</sub>  
 Arg Thr Tyr Gly Asp Val Gln Leu Met Pro Phe Val Asp Tyr Tyr Ser<sub>450</sub>  
 Leu Ile Ser Trp Lys Thr Leu Ala Ile Cys Ile Phe Gly Thr Glu Val<sub>465</sub>  
 470 475 480

5

Asp Ser Ala Lys Phe Ile Met Lys Thr Asp Asp Asp Ala Phe Val Arg  
 485 490 495  
 Val Asp Glu Val Leu Leu Ser Leu Ser Met Thr Asn Asn Thr Arg Gly  
 500 505 510  
 Leu Ile Tyr Gly Leu Ile Asn Ser Asp Ser Gln Pro Ile Arg Asn Pro  
 515 520 525  
 Asp Ser Lys Trp Tyr Ile Ser Tyr Glu Glu Trp Pro Glu Glu Lys Tyr  
 530 535 540  
 Pro Pro Trp Ala His Gly Pro Gly Tyr Ile Val Ser Arg Asp Ile Ala  
 545 550 555 560  
 Glu Ser Val Gly Lys Leu Phe Lys Glu Gly Asn Leu Lys Met Phe Lys  
 565 570 575  
 Leu Glu Asp Val Ala Met Gly Ile Trp Ile Ala Glu Leu Thr Lys His  
 580 585 590  
 Gly Leu Glu Pro His Tyr Glu Asn Asp Gly Arg Ile Ile Ser Asp Gly  
 595 600 605  
 Cys Lys Asp Gly Tyr Val Val Ala His Tyr Gln Ser Pro Ala Glu Met  
 610 615 620  
 Thr Cys Leu Trp Arg Lys Tyr Gln Glu Thr Lys Arg Ser Leu Cys Cys  
 625 630 635 640  
 Arg Glu Trp

5  
 <210> 24  
 <211> 2387  
 <212> ADN  
 <213> *Physcomitrella patens*

10  
 <400> 24  
 atgcgaggag gaggctgtgt ttgttgcgcc aagagatggg atggtttatg tgtagtgcag 60  
 gggttggatg tgaagcacct gtttgaagga gtctgcgaga gtttgaaatt cggattcaga 120  
 gtgcggcgat cgatggtgca acgttgttag cagtgtattgt tttcgccaac agaactgaca 180  
 tcatttggat tttttttacg cgtggatgtg ccctcttttt aaaaaatttc cgcgtggaag 240  
 agagacgggg gtttghtaat gaggcaggct gtggatcatca cccctagtat agcctgtcaa 300  
 gagagttcaa attcggtaat atgaagaggg ggtcgagact accggatatg gcgtgtacag 360  
 ggcggcaag aaatgatctt atcctagtgt caattgtttg cttgtttttt atggtgatat 420  
 tcatcccacc atatctccaa atgaactcac ttccggacat tgattctcct gtcgagaagc 480  
 tagaagatga tgatgatgct gtcttcactt ctcatagacg tcgtaaccaa gagcagattt 540  
 cagttgtcac tgacagtggc cagagacgga cagttatgcc atcttcgact ggtgcggagg 600  
 acgtaacgaa tgcaccgtct aaagattcac aggattcggg caagaaatca tcaagctact 660  
 cgaaaaaac cactctagaa gccaatagta aggaggaacg ccgtagtccg gggaaatacca 720  
 caggcgacat tgtttctctg gatgatgtga tagatcgtgc ctggctctgct ggtgccaag 780  
 cgtgggaaga actgaaact gcgttaagaa atggagaagg tgtctcaaag aatgtcagta 840  
 atgccactgc aaatgctgat ccgtgtccag catcactctc tgcagcaggg aaaaagttag 900  
 15 acgaattggg taaagtcttc cccttgcctt gtggctaat gtttgggtca gccattactc 960  
 tgattggaaa gcctcgagag gctcacatgg agtacaacc gccaatcgcc agagttgggg 1020  
 aaggcgtctc tccatattgc atggtttccc agttcttagt agagttacaa ggcttaaagg 1080  
 tggtgaaagg tgaagatcct cctcgaattc tacacttgaa tccctcgaact cgtggtgatt 1140  
 ggagctggaa acccatcatt gagcacaaca cttgttatcg gaaccagtgg ggtcctgccc 1200  
 accgatgcga gggttggcaa gtgcctgaat acgaagaac tgttgacggt cttccaagt 1260

gcgagaagtg gcttcgagat gatggcaaga aacctgcttc aacgcacaaa tcttggtggc 1320  
 ttggaagatt agttggtcgt tctgacaagg agacgcttga atgggagtac ccattatctg 1380  
 agggtcggga gttcgttctc accattcagag cagggtttga agggtttcat gtgactatcg 1440  
 atggctcgtca catcagctcg tttccttacc gtgtgggta cgctgtggaa gaaacaacgg 1500  
 ggatattagt agcaggagac gttgatgtga tgtctatcac agtgacatcc ctacccttaa 1560  
 cacatcctag ctactaccct gagttagttt tggaaatcggg ggacatttgg aaggcaccac 1620  
 ctgtcccagc taccaagata gatttattta ttgggatcat gtccagcagt aaccattttg 1680  
 cagaacggat ggcagtaagg aagacgtggt ttcaatctaa agctattcaa tcttcgagg 1740  
 ccgtggctcg cttctttgta gctctgcatg caaacaagga tatcaatag cagttgaaga 1800  
 aggaggcaga ctattatggc gatattataa tcctgccttt catcgacaga tatgatatag 1860  
 tggttctcaa gaccgtttaa atttgcaagt ttgggtcca gaatgtcaca gctaagtata 1920  
 ttatgaagtg tgacgatgac acttttgtga ggattgatag cgttctcga gagattcga 1980  
 ctacttcaat atcacaaggc ctttacatgg gtagcatgaa tgagtttcac aggcctcttc 2040  
 gttctggaaa gtggccgtg actgccgagg aatggcctga gcgaatttac ccaatatatg 2100  
 ctaatggacc aggatatac ctgtcagagg atattgtgca tttcattgtg gagatgaatg 2160  
 agagaggcag tttgagttta ttaaatgag aggacgtcag tgttgaata tgggtacgag 2220  
 aatatgcaa gcaagtgaag cacgttcaat acgaacatag catacggttt gctcaagccg 2280  
 gttgtatacc gaaatacttg acagctcatt accaatcgcc gcgtcaaatg ctgtgtctgt 2340  
 gggacaaggt acttgctcat gacgatggga aatgctgcaa cttgtga 2387

5

<210> 25  
 <211> 2052  
 <212> ADN  
 <213> *Physcomitrella patens*

10

<400> 25

atgaagaggg gtgtgagacc accgggtgtg ggatgtacag ggcggcaaag aaacaatcta 60  
 atcatagtgg caatcatatg tttggtttt atagcgatat tcatcccacc gtttcttgaa 120  
 atgaattcac tccccgatat tgattcccct gtgtataggt tagaaggtat taacttcgct 180  
 tcacatagac gtcgctatca agaacaggat tcacgtgtca gttacagtgg ctatggacag 240  
 ccagatagc catcaactgg tgatgaagac ataacgaaga caccgtctaa agcttcacag 300  
 gttttggaga agaaagtatc aagctatttg aaaaaagtca ctctggaaac ttacagtaaa 360  
 gaggaacgcc gtagtccagg gaacacaaca ggtgacattg tttcgtgga agatgtgata 420  
 gatcgcgcct ggtctgcccg cgccaaagct tgggaagagc tggaaattgc attcagacag 480  
 ggagaacatt tttcgaagaa ggacaataat gccaatgcaa ctgcagatcc atgccagca 540  
 tcactcttta caacaggaaa ggaattggac aatttaggaa gggcttccc actgccttgt 600  
 ggtctaattg ttggatcagc cataactctc attggaagc cacgggaagc tcacatggag 660  
 tacaaccgc caatcgccag agttggggaa ggtgtctctc catacgtcat ggtgtcccag 720  
 ttcataatgg agttacaggg cttgaagtg gtaaaagtg aagatcctcc tagaatctc 780  
 cacataaacc ctcgactccg tggtgactgg agctggaaac ccatcattga gcataatata 840  
 tgctatcgaa accagtggg cccagctcat cgggtgtaag gttggcaagt acctgaatac 900  
 gaagaaaccg tggacggtct tcccaagtgc gagaagtggc ttcgaggcga tgacaaaaaa 960  
 cctgcttca cccaaaaatc ctggtggcct gggcgattag ttggtcattc cgacaaggag 1020  
 acgcttgaat gggagtatcc attgtccgaa ggtcgggagt ttgttctcac cattcgagca 1080  
 ggtgtagaag gatttcactt aactattgat ggtcggcaca tcagttcgtt cccttatcgt 1140  
 gcgggttatg ctatggaaga agcaacagga atatcagtgg caggagacgt cgatgttctt 1200  
 tcgatgacag taacatcatt accttaaca catcccagct actaccctga gttggtttt 1260

15

gattcgggtg atatctggaa ggcaccacct ttaccaacag gcaagataga gttatttgtt 1320  
 ggaatcatgt caagcagcaa tcactttgca gaacgtatgg cagtaagaaa gacgtgggtt 1380  
 cagtctctgg ttatccaatc ctcccaagcg gtggctcgt tctttgtagc tctgcatgca 1440  
 aacaaggata tcaatctgca gctgaagaaa gaggctgact attacggcga tatgataatt 1500  
 ttacctttca tcgacagata tgatatagtg gttcttaaga ccgttgaaat tttcaagttt 1560  
 ggggtccaga atgttacagt tagccacgtc atgaaatgtg acgatgacac atttgtaagg 1620  
 attgacagcg ttcttgaaga gattcgaacg acgtcagtag gacagggcct ttacatgggc 1680  
 agcatgaatg agtttcatag accccttctg tctgggaagt gggccgtgac agttgaggag 1740  
 tggcctgagc gcatttacc aacatacgca aatgggtccag gatacatcct ttcggaagat 1800  
 attgtgcatt ttatagtga ggagagcaaa agaataaatt tgaggttatt taagatggag 1860  
 gacgtcagcg taggtatag ggtacgcgag tatgcaaaga tgaagtacgt gcaatacgag 1920  
 catagcgtag ggtttgctca agccggttgt atacctaact acctgacagc gcactatcaa 1980  
 tcgccgctc aaatgctgtg tctgtgggac aagggtcttg ctaccaatga cggcaagtgc 2040  
 tgcaccttgt ga 2052

5 <210> 26  
 <211> 688  
 <212> PRT  
 <213> *Physcomitrella patens*

10 <400> 26

Met Lys Arg Gly Ser Arg Leu Pro Asp Met Ala Cys Thr Gly Arg Gln  
 1 5 10 15

Arg Asn Asp Leu Ile Leu Val Ala Ile Val Cys Leu Phe Phe Met Val  
 20 25 30

Ile Phe Ile Pro Pro Tyr Leu Gln Met Asn Ser Leu Pro Asp Ile Asp  
 35 40 45

Ser Pro Val Glu Lys Leu Glu Asp Asp Asp Ala Val Phe Thr Ser  
 50 55 60

His Arg Arg Arg Asn Gln Glu Gln Ile Ser Val Val Thr Asp Ser Gly  
 65 70 75 80

Gln Arg Arg Thr Val Met Pro Ser Ser Thr Gly Ala Glu Asp Val Thr  
 85 90 95

Asn Ala Pro Ser Lys Asp Ser Gln Asp Ser Asp Lys Lys Ser Ser Ser  
 100 105 110

Tyr Ser Lys Lys Thr Thr Leu Glu Ala Asn Ser Lys Glu Glu Arg Arg  
 115 120 125

Ser Pro Gly Asn Thr Thr Gly Asp Ile Val Ser Leu Asp Asp Val Ile  
 130 135 140

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

15 Ala Leu Arg Asn Gly Glu Gly Val Ser Lys Asn Val Ser Asn Ala Thr  
 165 170 175

Ala Asn Ala Asp Pro Cys Pro Ala Ser Leu Ser Ala Ala Gly Lys Lys  
 180 185 190

Leu Asp Glu Leu Gly Lys Val Phe Pro Leu Pro Cys Gly Leu Met Phe  
 195 200 205

Gly Ser Ala Ile Thr Leu Ile Gly Lys Pro Arg Glu Ala His Met Glu  
 210 215 220



Tyr Lys Pro Pro Ile Ala Arg Val Gly Glu Gly Val Ser Pro Tyr Val  
 225 230 235 240  
 Met Val Ser Gln Phe Leu Val Glu Leu Gln Gly Leu Lys Val Val Lys  
 245 250 255  
 Gly Glu Asp Pro Pro Arg Ile Leu His Leu Asn Pro Arg Leu Arg Gly  
 260 265 270  
 Asp Trp Ser Trp Lys Pro Ile Ile Glu His Asn Thr Cys Tyr Arg Asn  
 275 280 285  
 Gln Trp Gly Pro Ala His Arg Cys Glu Gly Trp Gln Val Pro Glu Tyr  
 290 295 300  
 Glu Glu Thr Val Asp Gly Leu Pro Lys Cys Glu Lys Trp Leu Arg Asp  
 305 310 315 320  
 Asp Gly Lys Lys Pro Ala Ser Thr Gln Lys Ser Trp Trp Leu Gly Arg  
 325 330 335  
 Leu Val Gly Arg Ser Asp Lys Glu Thr Leu Glu Trp Glu Tyr Pro Leu  
 340 345 350  
 Ser Glu Gly Arg Glu Phe Val Leu Thr Ile Arg Ala Gly Val Glu Gly  
 355 360 365  
 Phe His Val Thr Ile Asp Gly Arg His Ile Ser Ser Phe Pro Tyr Arg  
 370 375 380  
 Val Gly Tyr Ala Val Glu Glu Thr Thr Gly Ile Leu Val Ala Gly Asp  
 385 390 395 400  
 Val Asp Val Met Ser Ile Thr Val Thr Ser Leu Pro Leu Thr His Pro  
 405 410 415  
 Ser Tyr Tyr Pro Glu Leu Val Leu Glu Ser Gly Asp Ile Trp Lys Ala  
 420 425 430  
 Pro Pro Val Pro Ala Thr Lys Ile Asp Leu Phe Ile Gly Ile Met Ser  
 435 440 445  
 Ser Ser Asn His Phe Ala Glu Arg Met Ala Val Arg Lys Thr Trp Phe  
 450 455 460  
 Gln Ser Lys Ala Ile Gln Ser Ser Gln Ala Val Ala Arg Phe Phe Val  
 465 470 475 480  
 Ala Leu His Ala Asn Lys Asp Ile Asn Met Gln Leu Lys Lys Glu Ala  
 485 490 495  
 Asp Tyr Tyr Gly Asp Ile Ile Ile Leu Pro Phe Ile Asp Arg Tyr Asp  
 500 505 510  
 Ile Val Val Leu Lys Thr Val Glu Ile Cys Lys Phe Gly Val Gln Asn  
 515 520 525  
 Val Thr Ala Lys Tyr Ile Met Lys Cys Asp Asp Asp Thr Phe Val Arg  
 530 535 540  
 Ile Asp Ser Val Leu Glu Glu Ile Arg Thr Thr Ser Ile Ser Gln Gly  
 545 550 555 560  
 Leu Tyr Met Gly Ser Met Asn Glu Phe His Arg Pro Leu Arg Ser Gly  
 565 570 575  
 Lys Trp Ala Val Thr Ala Glu Glu Trp Pro Glu Arg Ile Tyr Pro Ile  
 580 585 590  
 Tyr Ala Asn Gly Pro Gly Tyr Ile Leu Ser Glu Asp Ile Val His Phe  
 595 600 605

5

Ile Val Glu Met Asn Glu Arg Gly Ser Leu Gln Leu Phe Lys Met Glu  
610 615 620

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

His Val Gln Tyr Glu His Ser Ile Arg Phe Ala Gln Ala Gly Cys Ile  
645 650 655

Pro Lys Tyr Leu Thr Ala His Tyr Gln Ser Pro Arg Gln Met Leu Cys  
660 665 670

Leu Trp Asp Lys Val Leu Ala His Asp Asp Gly Lys Cys Cys Asn Leu  
675 680 685

<210> 27

<211> 683

<212> PRT

<213> *Physcomitrella patens*

<400> 27

5

10

Met Lys Arg Gly Val Arg Pro Pro Gly Val Gly Cys Thr Gly Arg Gln  
1 5 10 15

Arg Asn Asn Leu Ile Ile Val Ala Ile Ile Cys Leu Val Phe Ile Ala  
20 25 30

Ile Phe Ile Pro Pro Phe Leu Glu Met Asn Ser Leu Pro Asp Ile Asp  
35 40 45  
Ser Pro Val Tyr Arg Leu Glu Gly Ile Asn Phe Ala Ser His Arg Arg  
50 55 60

Arg Tyr Gln Glu Gln Asp Ser Arg Val Ser Tyr Ser Gly Tyr Gly Gln  
65 70 75 80

Pro Asp Met Pro Ser Thr Gly Asp Glu Asp Ile Thr Lys Thr Pro Ser  
85 90 95  
Lys Ala Ser Gln Val Leu Glu Lys Lys Val Ser Ser Tyr Leu Lys Lys  
100 105 110

Val Thr Leu Glu Thr Tyr Ser Lys Glu Glu Arg Arg Ser Pro Gly Asn  
115 120 125

Thr Thr Gly Asp Ile Val Ser Leu Glu Asp Val Ile Asp Arg Ala Trp  
130 135 140

Ser Ala Gly Ala Lys Ala Trp Glu Glu Leu Glu Ile Ala Phe Arg Gln  
145 150 155 160

Gly Glu His Phe Ser Lys Lys Asp Asn Asn Ala Asn Ala Thr Ala Asp  
165 170 175

Pro Cys Pro Ala Ser Leu Phe Thr Thr Gly Lys Glu Leu Asp Asn Leu  
180 185 190

Gly Arg Val Phe Pro Leu Pro Cys Gly Leu Met Phe Gly Ser Ala Ile  
195 200 205

Thr Leu Ile Gly Lys Pro Arg Glu Ala His Met Glu Tyr Lys Pro Pro  
210 215 220

Ile Ala Arg Val Gly Glu Gly Val Ser Pro Tyr Val Met Val Ser Gln  
225 230 235 240

Phe Ile Met Glu Leu Gln Gly Leu Lys Val Val Lys Gly Glu Asp Pro  
245 250 255

15

Pro Arg Ile Leu His Ile Asn Pro Arg Leu Arg Gly Asp Trp Ser Trp  
 260 265 270

Lys Pro Ile Ile Glu His Asn Thr Cys Tyr Arg Asn Gln Trp Gly Pro  
 275 280 285

Ala His Arg Cys Glu Gly Trp Gln Val Pro Glu Tyr Glu Glu Thr Val  
 290 295 300

Asp Gly Leu Pro Lys Cys Glu Lys Trp Leu Arg Gly Asp Asp Lys Lys  
 305 310 315 320

Pro Ala Ser Thr Gln Lys Ser Trp Trp Leu Gly Arg Leu Val Gly His  
 325 330 335

Ser Asp Lys Glu Thr Leu Glu Trp Glu Tyr Pro Leu Ser Glu Gly Arg  
 340 345 350

Glu Phe Val Leu Thr Ile Arg Ala Gly Val Glu Gly Phe His Leu Thr  
 355 360 365

Ile Asp Gly Arg His Ile Ser Ser Phe Pro Tyr Arg Ala Gly Tyr Ala  
 370 375 380

Met Glu Glu Ala Thr Gly Ile Ser Val Ala Gly Asp Val Asp Val Leu  
 385 390 395 400

Ser Met Thr Val Thr Ser Leu Pro Leu Thr His Pro Ser Tyr Tyr Pro  
 405 410 415

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

Thr Gly Lys Ile Glu Leu Phe Val Gly Ile Met Ser Ser Ser Asn His  
 435 440 445

Phe Ala Glu Arg Met Ala Val Arg Lys Thr Trp Phe Gln Ser Leu Val  
 450 455 460

Ile Gln Ser Ser Gln Ala Val Ala Arg Phe Phe Val Ala Leu His Ala  
 465 470 475 480

Asn Lys Asp Ile Asn Leu Gln Leu Lys Lys Glu Ala Asp Tyr Tyr Gly  
 485 490 495

Asp Met Ile Ile Leu Pro Phe Ile Asp Arg Tyr Asp Ile Val Val Leu  
 500 505 510

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

His Val Met Lys Cys Asp Asp Asp Thr Phe Val Arg Ile Asp Ser Val  
 530 535 540

Leu Glu Glu Ile Arg Thr Thr Ser Val Gly Gln Gly Leu Tyr Met Gly  
 545 550 555 560

Ser Met Asn Glu Phe His Arg Pro Leu Arg Ser Gly Lys Trp Ala Val  
 565 570 575

Thr Val Glu Glu Trp Pro Glu Arg Ile Tyr Pro Thr Tyr Ala Asn Gly  
 580 585 590

Pro Gly Tyr Ile Leu Ser Glu Asp Ile Val His Phe Ile Val Glu Glu  
 595 600 605

Ser Lys Arg Asn Asn Leu Arg Leu Phe Lys Met Glu Asp Val Ser Val  
 610 615 620

Gly Ile Trp Val Arg Glu Tyr Ala Lys Met Lys Tyr Val Gln Tyr Glu  
 625 630 635 640

5

His Ser Val Arg Phe Ala Gln Ala Gly Cys Ile Pro Asn Tyr Leu Thr  
645 650 655

Ala His Tyr Gln Ser Pro Arg Gln Met Leu Cys Leu Trp Asp Lys Val  
660 665 670

Leu Ala Thr Asn Asp Gly Lys Cys Cys Thr Leu  
675 680

## REIVINDICACIONES

- 5 1. Moléculas de ADN, caracterizadas porque comprenden una secuencia según la SEC. ID. nº 1 con un marco de lectura abierto desde el par de bases 513 al par de bases 2417 o una secuencia según la SEC. ID. nº 24 con un marco de lectura abierto desde del par de bases 321 al par de bases 2387, o presentan por lo menos un 80% de identidad con por lo menos una de las secuencias completas anteriores, o comprenden una secuencia que se degenera a las secuencias anteriores debido al código genético, presentando las secuencias que codifican proteínas vegetales con actividad de  $\beta$ 1,3-galactosiltransferasa (actividad de  $\beta$ 1,3-GalT) o siendo complementarias de ésta.
- 10 2. Molécula de ADN según la reivindicación 1, caracterizada porque codifica una proteína con actividad de GlcNAc- $\beta$ 1,3-galactosiltransferasa, una proteína con actividad con respecto a la transferencia de galactosa desde la UDP-galactosa a restos no reductores de GlcNAc, y/o una proteína con actividad con respecto a la transferencia de galactosa desde la UDP-galactosa a restos no reductores de GlcNAc de estructuras de N-glucano unidos a las proteínas.
- 15 3. Molécula de ADN según la reivindicación 1 ó 2, caracterizada porque presenta por lo menos un 90% de identidad con una de las secuencias según la SEC. ID. nº 1 o la SEC. ID. nº 24 o está degenerada debido al código genético o es complementaria de ésta.
- 20 4. Molécula de ADN según la reivindicación 1 a 3, caracterizada porque está asociada por enlace covalente a una sustancia marcadora detectable.
- 25 5. Vector que transcribe un ARN complementario que es complementario al ARNm de la  $\beta$ 1,3-galactosiltransferasa codificada por la molécula de ADN según cualquiera de las reivindicaciones 1 a 3 y en el que el ARN complementario contiene de 50 a 200 nucleótidos.
- 30 6. Vector de expresión, caracterizado porque comprende una molécula de ADN según cualquiera de las reivindicaciones 1 a 4 o una molécula de ADN según cualquiera de las reivindicaciones 1 a 4 que está inversamente orientada con respecto al activador.
- 35 7. Molécula de ADN que codifica una ribozima, caracterizada porque presenta dos secciones de secuencia, cada una de las cuales presenta un longitud de por lo menos 10 a 15 pares de bases y que son complementarias de las secciones de la secuencia de una molécula de ADN según cualquiera de las reivindicaciones 1 a 4 de modo que dicha ribozima acompleja y corta el ARNm transcrito por una molécula de  $\beta$ 1,3-GalT natural.
- 40 8. Vector biológicamente funcional, caracterizado porque contiene una molécula de ADN según la reivindicación 7.
- 45 9. Procedimiento de clonación de  $\beta$ 1,3-galactosiltransferasa, caracterizado porque una molécula de ADN según cualquiera de las reivindicaciones 1 a 3 o una molécula de ADN según cualquiera de las reivindicaciones 1 a 3 pero que carece de por lo menos la secuencia que codifica la transmembrana se clona en un vector posteriormente transfectado en una célula hospedadora, un tejido del hospedador o un hospedador con estirpes celulares que se obtienen mediante selección y ampliación de células hospedadoras transfectadas, cuyas estirpes celulares expresan la  $\beta$ 1,3-galactosiltransferasa activa.
- 50 10. Proteína con actividad de  $\beta$ 1,3-galactosiltransferasa y por lo menos un 80% de identidad de secuencia con la secuencia de aminoácidos SEC. ID. nº 19 o nº 26.
- 55 11. Vector ADN que contiene una molécula con una secuencia de ácido nucleico según la SEC. ID. nº 3.
- 60 12. Procedimiento de preparación de células hospedadoras recombinantes, particularmente de células vegetales o plantas, respectivamente, en el que la producción de  $\beta$ 1,3-galactosiltransferasa se suprime o se interrumpe completamente, respectivamente, caracterizado porque
- por lo menos uno de los vectores según las reivindicaciones 6, 8 u 11, o un vector que comprende una molécula de ADN según cualquiera de las reivindicaciones 1 a 3, respectivamente, mediante el cual dicha molécula de ADN comprende una mutación por eliminación, inserción y/o sustitución, se inserta en dicha célula hospedadora o planta, respectivamente; o
  - la molécula de ADN según cualquiera de las reivindicaciones 1 a 4, en la que dicha molécula de ADN comprende una mutación por eliminación, inserción y/o sustitución, se inserta en el genoma de dicha célula hospedadora o planta, respectivamente, en la posición de la secuencia homóloga sin mutar.
- 65 13. Plantas o células vegetales recombinantes en las que su producción de  $\beta$ 1,3-galactosiltransferasa se suprime o se interrumpe completamente,
- que comprende por lo menos uno de los vectores según la reivindicación 6, 8 u 11, o un vector que comprende

una molécula de ADN según cualquiera de las reivindicaciones 1 a 4, respectivamente, mediante el cual dicha molécula de ADN comprende una mutación por eliminación, inserción y/o sustitución; o

- 5 • que comprende la molécula de ADN según cualquiera de las reivindicaciones 1 a 4, en la que dicha molécula de ADN comprende una mutación por eliminación, inserción y/o sustitución, en el genoma de las células o plantas, respectivamente, en la posición de la secuencia homóloga sin mutar.

10 14. Molécula de ácido péptido nucleico (APN), caracterizada porque comprende una secuencia de bases complementaria de la secuencia de una molécula de ADN según cualquiera de las reivindicaciones 1 a 3 que codifica la  $\beta$ 1,3-galactosiltransferasa, o comprende una secuencia de bases correspondiente a la secuencia de una molécula de ADN según cualquiera de las reivindicaciones 1 a 3 que codifica la  $\beta$ 1,3-galactosiltransferasa.

15 15. Procedimiento de producción de plantas o células, respectivamente, en particular células vegetales que tienen bloqueada la expresión de  $\beta$ 1,3-galactosiltransferasa a nivel de transcripción o traducción, respectivamente, caracterizado porque las moléculas de APN según la reivindicación 14 se insertan en las células.

20 16. Procedimiento de producción de glucoproteínas recombinantes, caracterizado porque las plantas recombinantes, las células vegetales o los tejidos vegetales, respectivamente, según la reivindicación 13, o las plantas, los tejidos o las células vegetales, respectivamente, en los que se inserta la molécula de APN según la reivindicación 14 y que tienen una expresión bloqueada de la  $\beta$ 1,3-galactosiltransferasa a nivel de transcripción o traducción, respectivamente, se transfectan con el gen que codifica la glucoproteína, de tal modo que las glucoproteínas recombinantes se expresen.

25 17. Procedimiento según la reivindicación 16 caracterizado porque las glucoproteínas recombinantes son glucoproteínas, preferentemente para utilización médica.

30 18. Procedimiento de producción de glucoproteínas con N-glucanos, que comprende el alargamiento *in vitro* o *in vivo* del N-glucano de una glucoproteína con una proteína  $\beta$ 1,3-galactosiltransferasa activa según la reivindicación 10.

19. Utilización de un vector según la reivindicación 5 para inhibir la expresión de la  $\beta$ 1,3-galactosiltransferasa codificada por las moléculas de ADN según cualquiera de las reivindicaciones 1 a 3.

Alineación de secuencia múltiple CLUSTAL W (1.7)

```

PpGalT1      MKR--GSRLPDMACTGRQRNDLILVAIVCLFFMV-----IFI-----P---PYLQM---
PpGalT2      MKR--GVRPPGVGCTGRQRNNLIIVAIICLVFIA-----IFI-----P---PFLEM---
NP 174003    MKRFYGG-----LLVSMCMFLTVYRYVDLNT-----PVEKPYITAAAS
BAD17812     -----MWVTKRLGITVLIVLFLPLLVHH-----LIVNSPVSGP---SRYQVIHS
CAA75344     -----MLQW---

PpGalT1      -----NSL---P---D---IDSPDSDKKSSYSKKT-----LEA
PpGalT2      -----NSL---P---D---IDSPVLEKKVSSYLKKVT-----LET
NP 174003    VVVTN-----TTL---PMEWLR-----ITLPDFMKEARNTQEAISGDDIAVVSGLFVEQ
BAD17812     NLLGWLSDSLGNSVAQNP-----DNTPVEVI PADASASNSSDSGNS---LEG
CAA75344     -----R-----RRHCCFAKMT-----

PpGalT1      ----NSKEERRSPGNTTGDIVSLDDV-----IDRAWSAGAKAWEELETALRNGEGVSK
PpGalT2      ----YSKEERRSPGNTTGDIVSLEDV-----IDRAWSAGAKAWEELEIAFRQGEHFSK
NP 174003    ----NVSKEEREPLLTWNRLESVDNAQSLVNGVDAIKEAGI-VWESLVS AVEAKLV DV
BAD17812     FQWLNTWNHMKQLTNISDGLPHANEA-----IDNA---RTAWENLTSVHNS--TSK
CAA75344     -----WNAKRSLFRT-----
*

PpGalT1      NVSNATANADP-CPASLSAAGK-KLDELGK-VFPLPCGLMFGSAITLIGKPREAHMEYKP
PpGalT2      KDNNANATADP-CPASLFTTGK-ELDNLGR-VFPLPCGLMFGSAITLIGKPREAHMEYKP
NP 174003    NENQTRKGKEELCPQFLSKMNATEADGSSL-KLQIPCGLTQSSITVIGIP-----
BAD17812     QIKKER----Q-CPYSIHRMNA-SKPDGTGDFITIDPCGLIVGSSVTIIGTPGS-----
CAA75344     -----H-LIGVLSL-VLFAMLFFN-----HHDWLP
*

PpGalT1      PIARVGEVSPYVMVSQFLVELQGLKVVKGEDPPRILHLPRLRGD-WSWK-PIIEHNT-
PpGalT2      PIARVGEVSPYVMVSQFIMELQGLKVVKGEDPPRILHINPRLRGD-WSWK-PIIEHNT-
NP 174003    -----DG-----LVGSFRIDL TGQPLPGEPPPIIVHYNVRL LGD-KSTEDPVIQNS-
BAD17812     -----LSGNFRI DLVGTTELPGSGKPIVLHYDVRLTSD ELTGG-PVIVQNAF
CAA75344     GRAGFKENPVTYT-----FRGFRSTKSE-----TNHS-
*

PpGalT1      CYRNQ-WGPAHRCEGWQVPEYEETVDGLPKCEKWL RDGKKPAST--QKSWWLGRLVGRS
PpGalT2      CYRNQ-WGPAHRCEGWQVPEYEETVDGLPKCEKWL RGDDKKPAST--QKSWWLGRLVGRS
NP 174003    WTASQDWGAERCPKFD-PDMNKKVDDLDECNKMVGGEINRTSSTSLQSNTRGVPVAR-
BAD17812     TASNG-WGYEDRCPSCN CNATQ-VDDLERCNSMVGREEKRAINS--KQ-----HLNAKK
CAA75344     SLRNI-W-----KETV-----PQTLRPQTA--TNS-----NNT
*

PpGalT1      DKETLEWEYPLSEGREFVLTIRAGVEGFHVTIDGRHISSFPYRVGYAVEETT GILVAGDV
PpGalT2      DKETLEWEYPLSEGREFVLTIRAGVEGFHVTIDGRHISSFPYRAGYAMEEATGISVAGDV
NP 174003    EASKHEKYFPFKQGFSLVATLRVGTGEMQMTVDGKHITSFAFRD TLEPWL VSEIRITGDF
BAD17812     DEHPSTY-FPFKQGH LAISTLRIGLEGIHMTVDGKHVTSFPYKAGLEAWFVTEVGVSGDF
CAA75344     D-----LSP-----QGV TGLENTLSA---NGSIYN-----EKGTG-----
* * *

PpGalT1      DVMSITVTSPLPLTHPSY--PELVLESGDI--WKAPPV-PATK-IDLFIGIMSSSNHFAE
PpGalT2      DVLSMTVTSPLPLTHPSY--PELVLDSDGI--WKAPPL-PTGK-IELFVGIMSSSNHFAE
NP 174003    RLISILASGLPTSEES----EHVVDL-EA--LKSPTLSPLRP-LDLVIQVFSTANNFKR
BAD17812     KLVSAIASGLPTS-----EDLENSFDLAMLKSSPI-PEGKDVDLLIGIFSTANNFKR
CAA75344     -----HPNSYHFKYIINEPEKC--QEKSPF-----LILLIAAEPGQIEA
*

PpGalT1      RMAVRRTWFQSKAIQSSQAVARFFV---ALHANKDINMQLKKEADY YGDI I L P F I D R Y D
PpGalT2      RMAVRRTWFQSLVIQSSQAVARFFV---ALHANKDINLQLKKEADY YGDMI I L P F I D R Y D
NP 174003    RMAVRRTWMQYDDVRSGRVAVRFFV---GLHKSPLVNLELWNEARTYGDVQLMPFVDYYS
BAD17812     RMAIRRTWMQYDAVREGAVVVRFFV---GLHTNLIVNKELWNEARTYGDIQVLPFVDYYS
CAA75344     RRAIRQTW-GNESLAPGIQITRIFLLGLSIKLNGLYLRQRAILEESRQYHDI IQQEYLDTTY
* * * * *

```

FIG. 1

PpGalT1 IVVLKTVEICKFGVQNVTA-----KYIMKCDDDTFVRIIDSVLEEIRTTSSIS--QGLYMGS  
 PpGalT2 IVVLKTVEIFKFGVQNVTV-----SHVMKCDDDTFVRIIDSVLEEIRTTSSVG--QGLYMGS  
 NP 174003 LISWKTLAICIFGTEVDSA-----KFIMKTDDDAFVRVDEVLLSLSMTNNT--RGLIYGL  
 BAD17812 LITWKTLAICITYGTGAVSA-----KYLKMTDDDAFVRVDEIHSSVKQLNVS--HGLLYGR  
 CAA75344 NLTIKTL---MGMNWVATYCPHIPYVMKTDSDMFVNTEYLINKLLKPDLPFRHNYFTGY  
 \* \* \* \* \*

PpGalT1 MNEFHRPLRS--GKWAVTAEWPERIYPIYANGPGYILSEDIHVHFIVEMNERGSLQLFKM  
 PpGalT2 MNEFHRPLRS--GKWAVTVEEWPERIYPTYANGPGYILSEDIHVHFIVEESKRNNLRLFKM  
 NP 174003 INSDSQPIRNPDSKWIISYEAWPEEKYPPWAHGPGYIVSRDIAESVGKLFKEGNLKMFKL  
 BAD17812 INSDSGPHRNPEKWIISPEEWPEEKYPPWAHGPGYVVSQDIAKEINSWYETSHLKMFKL  
 CAA75344 LMRGYAPNRNKDSKWMPPDLYPSERYPVFVFCSTGYVFSGDIAEKIFKVSLL--GIRRLHL  
 \* \* \* \* \*

PpGalT1 EDVSVGIWVREY----AKQVKHVQYEHISIRFAQAGCIPKY---LTAHYQSPRQMLCLWDK  
 PpGalT2 EDVSVGIWVREY----AK-MKYVQYEHISVRFQAQACIPNY---LTAHYQSPRQMLCLWDK  
 NP 174003 EDVAMGIWIAEL----TKHGLEPHYENDGRIISDGCKDGY---VVAHYQSPAEMTCLWRK  
 BAD17812 EDVAMGIWIAEM----KKGGLPVQYKTDERINSDGCNDGC---IVAHYQEPHMLCMWEK  
 CAA75344 EDVYVGIKLAKLRIIDPVPPPNEFVFNHW-RVSYSSC--KYSHLITSHQFQPSSELIKYWNH  
 \* \* \* \* \*

PpGalT1 VLAHDDGKCCNL-----  
 PpGalT2 VLAHDDGKCCNL-----  
 NP 174003 YQETKRSLLCCREW\*-----  
 BAD17812 LLRTNQATCCN\*-----  
 CAA75344 LQONKHNACANAAKEKAGRYRHRKHL\*  
 \*

FIG. 1 (Continuación)



Proteína  $\beta$ 1-3GalT1: SEC ID NO: 19

MKRGSRLPDMACTGRQRND**LILVAIVCLFFMVFIPPYL**QMNSLPDIDSPDSDKKSS-  
 SYSKKTLEANSKEERRSPGNTTGDIVSLDDVIDRAWSAGAKAWEELETALRNREGVSKNVS-  
 NATANADPCPASLSAAGKKLDELGKVFPLPCGLMFGSAITLIGKPREAHMEYKPPPIARV-  
 GEGVSPYVMVSQFLVELQGLKVVKGEDPPRILHLNPRLRGDWSWKPIIEHNTCYRNQWGPahr-  
 CEGWQVPEYEETVDGLPKCEKWLRRDDGKKPASTQKSWWLGRLVGRSDKETLEWEYPLSE-  
 GREFVLTIRAGVEGFHVTIDGRHISSFPYRVGYAVEETTGILVAGDVDVMSITVTSPLPLTH-  
 PSYYPELVLESGDIWKAPPVPATKIDLFIGIMSSSNHFAERMAVRKTFQSKAIQSSQA-  
 VARFFVALHANKDINMQLKKEADYYGDIILPFIDRYDIVVLKTVEICKFGVQNV-  
 TAKYIMKCDDDTFVRIDSVLEEIRTTISQGLYMGSMNEFHRPLRSGKWAVTAEWPE-  
 RIYPIYANGPGYILSEDIVHFIVEMNERGSLQLFKMEDVSVGIWVREYAKQVKHVQYEHSIR-  
 FAQAGCIPKYLTAHYQSPRQMLCLWDKVLAHDDGKCCNL

**FIG. 2**

Proteína Pp  $\beta$ 1-3GalT2: SEC ID NO: 20

MKRGVRPPGVGCTGRQRNN**LIIVAIICLVFIAIFIPPF**LEMNSLPDIDSPVLEKKVS-  
 SYLKKVTLETYSKEERRSPGNTTGDIVSLEDVIDRAWSAGAKAWEELEIAFRQGEHFSSKDD-  
 NNANATADPCPASLFTTGKELDNLGRVFPLPCGLMFGSAITLIGKPREAHMEYKPPPIARV-  
 GEGVSPYVMVSQFIMELQGLKVVKGEDPPRILHINPRLRGDWSWKPIIEHNTCYRNQWGPahr-  
 CEGWQVPEYEETVDGLPKCEKWLRRDDGKKPASTQKSWWLGRLVGHSDKETLEWEYPLSE-  
 GREFVLTIRAGVEGFHVTIDGRHISSFPYRAGYAMEEATGISVAGDVDVLSMTVTSPLPLTH-  
 PSYYPELVLDGDIWKAPPLPTGKIELFVGIMSSSNHFAERMAVRKTFQSLVIQSSQA-  
 VARFFVALHANKDINLQKKEADYYGDMIIILPFIDRYDIVVLKTVEIFKFGVQNVTVS-  
 HVMKCDDDTFVRIDSVLEEIRTTSVGQGLYMGSMNEFHRPLRSGKWAVTVEEWPERIY-  
 PTYANGPGYILSEDIVHFIVEESKRNNLRLFKMEDVSVGIWVREYAKMKYVQYEHSVRFAQAG-  
 CIPNYLTAHYQSPRQMLCLWDKVLATNDGKCCNL

**Fig. 3**

Proteína Pp  $\beta$ 1,3GalT1as: SEC ID NO: 26

MKRGSRLPDMACTGRQRNDLILVAIVCLFFMVIFIPPYLQMNSLPDIDSP  
**VEKLEDDDDAVFTSHRRRNQEQISVVTD SGQRRTVMPSSTGAEDVTNAPS**  
**KDSQDS**DKKSSSYSKKTTLEANSKEERRSPGNTTGDIVSLDDVIDRAWSA  
 GAKAWEELETALRNGEGVSKNVSANATANADPCPASLSAAGKKLDELGKVF  
 PLPCGLMFGSAITLIGKPREAHMEYKPPIARVGEGVSPYVMVSQFLVELQ  
 GLKVVKGEDPPRILHLNPRLRGDWSWKPIIEHNTCYRNQWGP AHRCEGWQ  
 VPEYEETVDGLPKCEKWLRRDDGKKPASTQKSWWLGRLVGRSDKETLEWEY  
 PLSEGREFVLTIRAGVEGFHVTIDGRHISSFPYRVGYAVEETT GILVAGD  
 VDMSITVTSPLTHPSYYPELVLESGDIWKAPPVPATKIDLFIGIMSSS  
 NHFAERMAVRKTFWQSKAIQSSQAVARFFVALHANKDINMQLKKEADYYG  
 DIIILPFIDRYDIVVLKTVEICKFGVQNV TAKYIMKCDDDTFVRIDSVLE  
 EIRTTISISQGLYM GSMNEFHRPLRSGKWAVTAEWPERIYPIYANGPGYI  
 LSEDIVHFIVEMNERGSLQLFKMEDVSVGIWVREYAKQVKHVQYEHSIRF  
 AQAGCIPKYLT AHYQSPRQMLCLWDKVL AHDDGKCCNL

**Fig. 4**

Proteína Pp  $\beta$ 1,3GalT2as: SEC ID NO: 27

MKRGVRPPGVGCTGRQRNNLIIVAIICLVFIAIFIPPFLEMNSLPDIDSP  
**VYRLEGINFASHRRRYQEQDSRVSYSYGYGQPDMPSTGDEDITKTPSKASQ**  
 VLEKKVSSYLKKVTLETYSKEERRSPGNTTGDIVSLEDVIDRAWSAGAKA  
 WEELEIAFRQGEHFSKKNANANATADPCPASLFTTGKELDNLGRVFPLPC  
 GLMFGSAITLIGKPREAHMEYKPPIARVGEGVSPYVMVSQFIMELQGLKV  
 VKGEDPPRILHINPRLRGDWSWKPIIEHNTCYRNQWGP AHRCEGWQVPEY  
 EETVDGLPKCEKWLRRDDKPPASTQKSWWLGRLVGHSDKETLEWEYPLSE  
 GREFVLTIRAGVEGFHLTIDGRHISSFPYRAGYAMEEATGISVAGDV DVL  
 SMTVTSPLTHPSYYPELVLD SGDIWKAPPLPTGKIELFVGIMSSSNHFA  
 ERMAVRKTFWQSLVIQSSQAVARFFVALHANKDINLQLKKEADYYGDMII  
 LPFIDRYDIVVLKTVEIFKFGVQNVTVSHVMKCDDDTFVRIDSVLEEIRT  
 TSVGQGLYM GSMNEFHRPLRSGKWAVTVEEWPERIYPTYANGPGYILSED  
 IVHFIVEESKRNNLRLFKMEDVSVGIWVREYAKMKYVQYEHSVRFAQAGC  
 IPNYLTAHYQSPRQMLCLWDKVLATNDGKCCTL

**Fig. 5**