

19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 374 419**

51 Int. Cl.:  
**C12N 15/54** (2006.01)  
**C12P 19/18** (2006.01)  
**C12N 9/10** (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

- 96 Número de solicitud europea: **06721794 .3**  
96 Fecha de presentación: **07.04.2006**  
97 Número de publicación de la solicitud: **1869184**  
97 Fecha de publicación de la solicitud: **26.12.2007**

54 Título: **IDENTIFICACIÓN DE UNA BETA-1,3-N-ACETILGALACTOSAMINILTRANSFERASA (CGTE) DE CAMPYLOBACTER JEJUNI LIO87.**

30 Prioridad:  
**11.04.2005 US 670608 P**

45 Fecha de publicación de la mención BOPI:  
**16.02.2012**

45 Fecha de la publicación del folleto de la patente:  
**16.02.2012**

73 Titular/es:  
**NATIONAL RESEARCH COUNCIL OF CANADA  
BUILDING M-58, RM EG-12, 1200 MONTREAL  
ROAD  
OTTAWA, ONTARIO K1A OR6, CA**

72 Inventor/es:  
**GILBERT, Michel;  
WAKARCHUK, Warren y  
HOULISTON, Scott**

74 Agente: **Arias Sanz, Juan**

ES 2 374 419 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

Identificación de una beta-1,3-N-acetilgalactosaminiltransferasa (CGTE) de *Campylobacter jejuni* LIO87.

## Referencias cruzadas a solicitudes relacionadas

Esta solicitud reivindica el beneficio de la solicitud provisional estadounidense n.º 60/670.608, presentada el 11 de abril de 2005.

## Antecedentes de la invención

Los hidratos de carbono se reconocen ahora como de gran importancia en muchos acontecimientos de reconocimiento célula-célula, particularmente la adhesión de bacterias y virus a células de mamífero en la patogénesis y la interacción leucocito-célula endotelial a través de selectinas en la inflamación (Varki (1993) *Glycobiology* 3: 97-130). Además, se cree que los glicoconjugados sialilados que se encuentran en bacterias (Preston *et al.* (1996) *Crit. Rev. Microbiol.* 22:139-180; Reuter *et al.* (1996) *Biol. Chem. Hoppe-Seyler* 377:325-342) imitan a los oligosacáridos que se encuentran en los glicolípidos de mamíferos para evadir la respuesta inmunitaria del huésped (Moran *et al.* (1996) *FEMS Immunol. Med. Microbiol.* 16:105-115). La imitación molecular de estructuras del huésped por la parte de sacárido del lipopolisacárido (LPS) se considera que es un factor de virulencia de diversos patógenos de la mucosa, que usan esta estrategia para evadir la respuesta inmunitaria del huésped (Moran *et al.* (1996) *FEMS Immunol. Med. Microbiol.* 16: 105-115; Moran *et al.* (1996) *J. Endotoxin Res.* 3: 521-531).

Las estructuras de oligosacáridos que participan en estos y otros procesos son posibles agentes terapéuticos, pero requieren mucho tiempo y son caros de preparar por medios químicos tradicionales. Una vía muy prometedora para la producción de estructuras de oligosacáridos específicas es a través del uso de las enzimas que las producen *in vivo*, las glicosiltransferasas. Tales enzimas pueden usarse como catalizadores regio y estereoselectivos para la síntesis *in vitro* de oligosacáridos (Ichikawa *et al.* (1992) *Anal. Biochem.* 202: 215-238).

La síntesis enzimática a gran escala de oligosacáridos depende de la disponibilidad de suficientes cantidades de las glicosiltransferasas requeridas. Sin embargo, la producción de glicosiltransferasas en cantidades suficientes para su uso en la preparación de estructuras de oligosacáridos ha sido problemática. Se ha logrado la expresión de muchas glicosiltransferasas de mamífero con la participación de expresión en huéspedes eucariotas, lo que puede implicar medios de cultivo tisular caros y rendimientos de proteína sólo moderados (Kleene *et al.* (1994) *Biochem. Biophys. Res. Commun.* 201: 160-167; Williams *et al.* (1995) *Glycoconjugate J.* 12: 755-761). Se ha logrado la expresión en *E. coli* para glicosiltransferasas de mamífero, pero estos intentos han producido principalmente formas insolubles de la enzima a partir de las cuales ha sido difícil recuperar enzima activa en grandes cantidades (Aoki *et al.* (1990) *EMBO. J.* 9:3171- 3178; Nishiu *et al.* (1995) *Biosci. Biotech. Biochem.* 59 (9): 1750-1752). Además, debido a la actividad biológica de sus productos, las sialiltransferasas de mamífero actúan generalmente en tejidos específicos, compartimentos celulares y/o etapas del desarrollo para crear estructuras de glicano precisas. La identificación de glicosiltransferasas que pueden usarse en la síntesis enzimática de oligosacáridos comercialmente valiosos y que pueden producirse en grandes cantidades sería por tanto útil en el desarrollo de estas tecnologías. La presente invención satisface ésta y otras necesidades.

La base de datos (UNIPROT [en línea] versión de lanzamiento 01-02-2005 1 de octubre de 2002 (01-10-2002), XP002510530 recuperado del número de registro de EBI UNIPROT: Q8KWRO da a conocer la secuencia de una galactosiltransferasa supuesta. La base de datos EMBL [en línea] 12 de enero de 2005, XP002510532 recuperado del número de registro de EBI: EMBL:CP000025 da a conocer el genoma completo de *Campylobacter jejuni* RMI221. El documento US2004/259140 describe glicosiltransferasas procariontas, incluyendo una sialiltransferasa bifuncional que tiene actividad tanto  $\alpha$ 2,3 como  $\alpha$ 2,8. Se describen también una  $\beta$ 1,4-GalNAc transferasa y una  $\beta$ 1,3-galactosiltransferasa, así como otras glicosiltransferasas y enzimas implicadas en la síntesis de lipooligosacárido (LOS).

## Breve resumen de la invención

En un aspecto, la presente invención proporciona el uso de un polipéptido de  $\beta$ -1,3-N-acetilgalactosaminiltransferasa recombinante o aislado para una reacción de N-acetilgalactosaminilación, en el que el polipéptido de  $\beta$ -1,3-N-acetilgalactosaminiltransferasa comprende una secuencia de aminoácidos con una identidad de al menos el 90% con SEQ ID NO: 2. El polipéptido de  $\beta$ -1,3-N-acetilgalactosaminiltransferasa transfiere un resto de galactosa de un sustrato donador a un sustrato aceptor. Un resto de galactosa puede ser, por ejemplo, o bien galactosa o bien GalNAc. En una realización preferida, el resto de galactosa es GalNAc. En otra realización, el polipéptido de  $\beta$ -1,3-N-acetilgalactosaminiltransferasa comprende la secuencia de aminoácidos de SEQ ID NO:2.

En otro aspecto, la presente invención proporciona una mezcla de reacción para producir el producto de sacárido N-acetilgalactosaminilado en el que un resto de N-acetilgalactosaminilo está unido a un resto de oligosacárido mediante un enlace  $\beta$ -1,3, comprendiendo la mezcla de reacción: un polipéptido de  $\beta$ -1,3-N-acetilgalactosaminiltransferasa recombinante o aislado; un sustrato donador que comprende un resto de N-acetilgalactosaminilo; y un sustrato aceptor que comprende un resto de oligosacárido, en el que el polipéptido de  $\beta$ -

1,3-N-acetilgalactosaminiltransferasa comprende una secuencia de aminoácidos con una identidad de al menos el 90% con SEQ ID NO: 2.

5 En otro aspecto, la presente invención proporciona un método de producción de un producto de sacárido N-acetilgalactosaminilado, comprendiendo el método las etapas de: a) poner en contacto un sustrato aceptor con un sustrato donador que comprende un resto de GalNAc y un polipéptido de  $\beta$ -1,3-N-acetilgalactosaminiltransferasa recombinante o aislado con una identidad de al menos el 90% o el 95% con SEQ ID NO:2; y b) permitir que se produzca la transferencia del resto de GalNAc al sacárido aceptor, produciendo de ese modo el producto de sacárido N-acetilgalactosaminilado. En una realización, el polipéptido de  $\beta$ -1,3-N-acetilgalactosaminiltransferasa comprende la secuencia de aminoácidos de SEQ ID NO:2. En una realización preferida, el método de galactosilación se realiza a una escala de producción comercial. En una realización preferida adicional, el sustrato aceptor es un glicopéptido o una glicoproteína.

### Breve descripción de los dibujos

La figura 1 proporciona la estructura de GalNAc $\beta$ -1,3-Gal $\alpha$ -1,4-Gal $\beta$ -1,4-GlcNAc-p-nitrofenilo sintetizado a partir de Gal $\alpha$ -1,4-Gal $\beta$ -1,4-GlcNAc-p-nitrofenilo usando CgtE.

15 La figura 2 proporciona el espectro de  $^1\text{H}$ - $^{13}\text{C}$  HSQC del compuesto de p-nitrofenilo de tetrasacárido. Los picos cruzados se marcan tal como sigue: *a* para  $\beta$ GlcNAc, *b* para  $\beta$ Gal(1-4), *c* para  $\alpha$ Gal(1-4) y *d* para  $\beta$ GalNAc(1-3). Las resonancias del carbono en las posiciones a4, b4 y c3 presentan desplazamiento a campo bajo en comparación con valores de monosacárido, lo que concuerda con su participación en un enlace glicosídico.

20 La figura 3 proporciona una alineación entre los residuos 5-167 de la proteína CgtE (línea superior) y los residuos 1-161 de la secuencia consenso glycosyl\_trans\_f2 (línea inferior). Los residuos idénticos están en negrita y los residuos conservados están subrayados.

### Descripción detallada de la invención

#### I. Introducción

25 El locus de biosíntesis de lipooligosacárido (LOS) se ha secuenciado en diversas cepas de *Campylobacter jejuni* como parte de un proyecto sobre la genómica comparativa de este locus. Véanse, por ejemplo, Gilbert, *et al.*, J. Biol. Chem. 275:3896-3906 (2000); Gilbert, *et al.*, J. Biol. Chem. 277:327-337 (2002); y Gilbert, *et al.*, en *Campylobacter: Molecular and Cellular Biology*. (Horizon Bioscience, Editors: J.M. Ketley y M. E. Konkel), capítulo 11 (2005). *C. jejuni* LIO87 es una cepa tipo del sistema de serotipado LIOR (termolábil). La organización del locus LOS de L1087 (clase "D", número de registro de GenBank AF400669) es distinto de la mayoría de los locus LOS de *C. jejuni* caracterizados hasta la fecha (clases "A", "B" y "C", Gilbert *et al.* 2002). El locus LOS de *C. jejuni* LIO87 carece de la agrupación de genes implicados en la biosíntesis de ácido siálico y en la expresión de núcleos externos de LOS que imitan gangliósidos. El locus LOS de *C. jejuni* LIO87 incluye 10 marcos de lectura abiertos (ORF). Búsquedas de homología de secuencia indicaron que cuatro de estos ORF (los ORF n.º 1, n.º 2, n.º 3 y n.º 10) están implicados en la biosíntesis del núcleo interno o el lípido. No fue posible deducir las especificidades donadoras oceptoras de las proteínas codificadas por los otros seis marcos de lectura abiertos basándose en la información de secuencia.

35 La presente invención demuestra por primera vez que el producto del gen de CgtE tiene actividad  $\beta$ -1,3-N-acetilgalactosaminiltransferasa. Además, la enzima puede transferir galactosa a una molécula aceptor. La proteína CgtE puede transferir o bien galactosa o bien Gal NAc a un residuo de GalNAc o una galactosa terminal en un azúcar aceptor.

#### 40 II. Definiciones

Se usan las siguientes abreviaturas en el presente documento:

Ara = arabinosilo;

Fru = fructosilo;

Fuc = fucosilo;

45 Gal = galactosilo;

GalNAc = N-acetilgalactosaminilo;

Glc = glucosilo;

GlcNAc = N-acetilglucosaminilo;

Man = manosilo; y

NeuAc = sialilo (N-acetilneuraminilo).

Un “sustrato aceptor” o un “sacárido aceptor” para una glicosiltransferasa, por ejemplo, un polipéptido de CgtE, es un resto de oligosacárido que puede actuar como un aceptor para una glicosiltransferasa particular. Cuando el sustrato aceptor se pone en contacto con la correspondiente glicosiltransferasa y el sustrato donador de azúcar, y otros componentes necesarios de la mezcla de reacción, y se incuba la mezcla de reacción durante un periodo de tiempo suficiente, la glicosiltransferasa transfiere residuos de azúcar desde el sustrato donador de azúcar hasta el sustrato aceptor. El sustrato aceptor puede variar para diferentes tipos de una glicosiltransferasa particular. Por consiguiente, la expresión “sustrato aceptor” se toma en contexto con la glicosiltransferasa particular de interés para una aplicación particular. Se describen en el presente documento sustratos aceptores para GalNActransferasas, por ejemplo, CgtE de *C. jejuni* LIO87 y glicosiltransferasas adicionales.

Un “sustrato donador” para glicosiltransferasas es un azúcar de nucleótido activado. Tales azúcares activados consisten generalmente derivados de uridina, guanosina y citidina monofosfato de los azúcares (UMP, GMP y CMP, respectivamente) o derivados de difosfato de los azúcares (UDP, GDP y CDP, respectivamente) en los que el nucleósido monofosfato o difosfato sirve como grupo saliente. Por ejemplo un sustrato donador para fucosiltransferasas es GDP-fucosa. Un sustrato donador para proteínas CgtE incluye, por ejemplo, UDP-GalNAc o UDP-Gal. Sustratos donadores para sialiltransferasas, por ejemplo, son nucleótidos de azúcar activados que comprenden el ácido siálico deseado. Por ejemplo, en el caso de NeuAc, el azúcar activado es CMP-NeuAc. Sistemas bacterianos, vegetales y fúngicos pueden usar algunas veces otros azúcares de nucleótidos activados.

Se considera que los oligosacáridos tienen un extremo reductor y un extremo no reductor, sea no o de hecho el sacárido en el extremo reductor un azúcar reductor. Según la nomenclatura aceptada, los oligosacáridos se representan en el presente documento con el extremo no reductor a la izquierda y el extremo reductor a la derecha. Todos los oligosacáridos descritos en el presente documento se describen con el nombre o la abreviatura para el sacárido no reductor (por ejemplo, Gal), seguido por la configuración del enlace glicosídico ( $\alpha$  o  $\beta$ ), el enlace de anillo, la posición en el anillo del sacárido reductor implicado en el enlace, y luego el nombre o la abreviatura del sacárido reductor (por ejemplo, GlcNAc). La unión entre dos azúcares puede expresarse, por ejemplo, como 2,3, 2→3, o (2,3). Cada sacárido es una piranosa o furanosa.

Tal como se usa en el presente documento, un “resto de galactosa” se refiere a una molécula que incluye galactosa o que puede derivarse de galactosa. Los restos de galactosa son habitualmente monosacáridos, por ejemplo, galactosa o GalNAc.

Tal como se usa en el presente documento, un “producto de sacárido galactosilado” se refiere a un oligosacárido, un polisacárido o un resto de hidrato de carbono, o bien no conjugado o bien conjugado con un glicolípido o una glicoproteína, por ejemplo, una biomolécula, que incluye un resto de galactosa. Puede usarse cualquiera de los restos de galactosa anteriores, por ejemplo, galactosa o GalNAc. En realizaciones preferidas el resto de galactosa transferido por CgtE es GalNAc.

En algunas realizaciones, otros restos de azúcar, por ejemplo, fucosa, ácido siálico, glucosa o GluNAc, se añaden también al sustrato aceptor a través de la acción de glicosiltransferasas adicionales para producir el producto de sacárido galactosilado. En algunas realizaciones, el sustrato aceptor comprende un resto de galactosa y la proteína CgtE se usa para añadir un resto de galactosa adicional, produciendo el producto de sacárido galactosilado.

La expresión “ácido siálico” o “resto de ácido siálico” se refiere a cualquier miembro de una familia de azúcares carboxilados de nueve carbonos. El miembro más común de la familia de ácido siálico es el ácido N-acetilneuramínico (ácido 2-ceto-5-acetamido-3,5-didesoxi-D-glicero-D-galactononulopiranos-1-ónico (abreviado a menudo como Neu5Ac, NeuAc, o NANA). Un segundo miembro de la familia es el ácido N-glicolil-neuramínico (Neu5Gc o NeuGc), en el que el grupo N-acetilo de NeuAc está hidroxilado. Un tercer miembro de la familia del ácido siálico es el ácido 2-ceto-3-desoxi-nonulosónico (KDN) (Nadano *et al.* (1986) *J. Biol. Chem.* 261: 11550-11557; Kanamori *et al.*, *J. Biol. Chem.* 265: 21811-21819 (1990)). También se incluyen ácidos siálicos 9-sustituidos tales como 9-O-acilo C<sub>1</sub>-C<sub>6</sub>-Neu5Ac como 9-O-lactil-Neu5Ac o 9-O-acetil-Neu5Ac, 9-desoxi-9-fluoro-Neu5Ac y 9-azido-9-desoxi-Neu5Ac. Para una revisión de la familia del ácido siálico, véase, por ejemplo, Varki, *Glycobiology* 2: 25-40 (1992); *Sialic Acids: Chemistry, Metabolism and Function*, R. Schauer, Ed. (Springer-Verlag, Nueva York (1992)). La síntesis y el uso de compuestos de ácido siálico en un procedimiento de sialilación se da a conocer en la solicitud internacional WO 92/16640, publicada el 1 de Octubre de 1992.

Puede encontrarse gran parte de la nomenclatura y los procedimientos de laboratorios generales requeridos en esta solicitud en Sambrook, *et al.*, *Molecular Cloning: A Laboratory Manual* (2ª ed.), vol. 1-3, Cold Spring Harbor Laboratory, Cold Spring Harbor, Nueva York, 1989. El manual se denomina a continuación en el presente documento “Sambrook *et al.*”

Las expresiones “CgtE de *C. jejuni* LIO87”, “CgtE,” o un ácido nucleico que codifica para “CgtE de *C. jejuni* LIO87” o “CgtE” se refieren a ácidos nucleicos y variantes polimórficas de polipéptidos, alelos, mutantes y homólogos entre especies que: (1) tienen una secuencia de aminoácidos que tiene una identidad de secuencia de aminoácidos de al menos el 60%, una identidad de secuencia de aminoácidos del 65%, el 70%, el 75%, el 80%, el 85%, el 90%,

preferiblemente el 91%, el 92%, el 93%, el 94%, el 95%, el 96%, el 97%, el 98% o el 99% o más, preferiblemente a lo largo de una región de al menos aproximadamente 25, 50, 100, 200, 500, 1000 o más aminoácidos, con una secuencia de aminoácidos codificada por un ácido nucleico de CgtE de *C. jejuni* LIO87 (para una secuencia de ácido nucleico de CgtE de *C. jejuni* LIO87, véase, por ejemplo, SEQ ID NO:1) o con una secuencia de aminoácidos de una proteína CgtE de *C. jejuni* LIO87 (para una secuencia de proteína de CgtE de *C. jejuni* LIO87, véase, por ejemplo, SEQ ID NO:2); (2) se unen a anticuerpos, por ejemplo, anticuerpos policlonales, generados contra un inmunógeno que comprende una secuencia de ácido nucleico de una proteína CgtE de *C. jejuni* LIO87, y variantes modificadas de manera conservativa de las mismas; (3) se hibridan específicamente en condiciones de hibridación rigurosas con una hebra antisentido que corresponde a una secuencia de ácido nucleico que codifica para una proteína CgtE de *C. jejuni* LIO87, y variantes modificadas de manera conservativa de la misma; (4) tienen una secuencia de ácido nucleico que tiene una identidad de secuencia de nucleótidos de al menos el 90%, preferiblemente al menos el 91%, el 92%, el 93%, el 94%, el 95%, el 96%, el 97%, el 98%, el 99%, o más, preferiblemente a lo largo de una región de al menos aproximadamente 25, 50, 100, 200, 500, 1000 o más nucleótidos, con un ácido nucleico de CgtE de *C. jejuni* LIO87, por ejemplo, SEQ ID NO:1, o un ácido nucleico que codifica para el dominio catalítico. Preferiblemente, el dominio catalítico tiene una identidad de aminoácidos de al menos el 90%, preferiblemente de al menos el 91%, el 92%, el 93%, el 94%, el 95%, el 96%, el 97%, el 98%, el 99% con el dominio catalítico de CgtE de *C. jejuni* LIO87 de SEQ ID NO:2. Una secuencia de polinucleótido o polipéptido es normalmente de una bacteria incluyendo, pero sin limitarse a, *Campylobacter*, *Haemophilus* y *Pasteurella*. Los ácidos nucleicos y proteínas de la invención incluyen tanto moléculas que se producen de manera natural como recombinantes. Una proteína CgtE de *C. jejuni* LIO87 tiene normalmente actividad N-acetilgalactosaminiltransferasa y galactosiltransferasa. Pueden realizarse ensayos de N-acetilgalactosaminiltransferasa y galactosiltransferasa según métodos conocidos por los expertos en la técnica, usando sustratos donadores y sustratos aceptores apropiados, tal como se describe en el presente documento.

“Escala comercial” se refiere a la producción a escala de gramos de un producto galactosilado en una única reacción. En realizaciones preferidas, escala comercial se refiere a la producción de más de aproximadamente 50, 75, 80, 90, 100, 125, 150, 175 ó 200 gramos de producto galactosilado.

Tal como se usa en el presente documento, un “polipéptido de CgtE truncado” o variantes gramaticales se refiere a un polipéptido de CgtE que se ha manipulado para eliminar al menos un residuo de aminoácido, en relación con un polipéptido de CgtE de tipo natural que se produce en la naturaleza, siempre que el polipéptido de CgtE truncado conserve la actividad enzimática. Por ejemplo, una proteína CgtE que carece de los 28 aminoácidos del extremo C-terminal conserva la actividad.

“Variantes modificadas de manera conservativa” se aplica a secuencias tanto de aminoácidos como de ácido nucleico. Con respecto a secuencias de ácido nucleico particulares, variantes modificadas de manera conservativa se refiere a los ácidos nucleicos que codifican para secuencias de aminoácidos idénticas o esencialmente idénticas, o cuando el ácido nucleico no codifica para una secuencia de aminoácidos, se refiere a secuencias esencialmente idénticas. Debido a la degeneración del código genético, un gran número de ácidos nucleicos funcionalmente idénticos codifican para cualquier proteína dada. Por ejemplo, los codones GCA, GCC, GCG y GCU codifican todos para el aminoácido alanina. Por tanto, en cada posición en la que se especifica una alanina mediante un codón, el codón puede alterarse a cualquiera de los codones correspondientes descritos sin alterar el polipéptido codificado. Tales variaciones de ácido nucleico son “variaciones silenciosas”, que son una especie de variaciones modificadas de manera conservativa. Cada secuencia de ácido nucleico en el presente documento que codifica para un polipéptido describe también cada posible variación silenciosa del ácido nucleico. Un experto reconocerá que cada codón en un ácido nucleico (excepto AUG, que es habitualmente el único codón para la metionina, y TGG que es habitualmente el único codón para el triptófano) puede modificarse para producir una molécula funcionalmente idéntica. Por consiguiente, cada variación silenciosa de un ácido nucleico que codifica para un polipéptido está implícita en cada secuencia descrita con respecto al producto de expresión, pero no con respecto a las secuencias de sonda reales.

Como con las secuencias de aminoácidos, un experto reconocerá que sustituciones, deleciones o adiciones individuales en una secuencia de ácido nucleico, péptido, polipéptido o proteína que altera, añade o delecciona un único aminoácido o un pequeño porcentaje de aminoácidos en la secuencia codificada es una “variante modificada de manera conservativa” en la que la alteración da como resultado la sustitución de un aminoácido por un aminoácido químicamente similar. Se conocen bien en la técnica tablas de sustituciones conservativas que proporcionan aminoácidos funcionalmente similares. Tales variantes modificadas de manera conservativa son además de y no excluyen variantes polimórficas, homólogos entre especies y alelos de la invención.

Los expertos reconocerán que muchos aminoácidos pueden sustituirse por otros en una proteína sin afectar a la función de la proteína, es decir, una sustitución conservativa puede ser la base de una variante modificada de manera conservativa de una proteína tal como las proteínas CgtE dadas a conocer. A continuación se facilita una lista incompleta de sustituciones de aminoácidos conservativas. Los siguientes ocho grupos contienen cada uno aminoácidos que son sustituciones conservativas entre sí: 1) Alanina (A), Glicina (G); 2) Ácido aspártico (D), Ácido glutámico (E); 3) Asparagina (N), Glutamina (Q); 4) Arginina (R), Lisina (K); 5) Isoleucina (I), Leucina (L), Metionina (M), Valina (V), Alanina (A); 6) Fenilalanina (F), Tirosina (Y), Triptófano (W); 7) Serina (S), Treonina (T), Cisteína (C); y 8) Cisteína (C), Metionina (M) (véase, por ejemplo, Creighton, Proteins (1984)).

Los métodos de la invención son útiles para producir un producto galactosilado, generalmente transfiriendo un resto de galactosa de un sustrato donador a una molécula aceptora. Los métodos de la invención también son útiles para producir un producto de azúcar galactosilado que comprende residuos de azúcar adicionales, generalmente transfiriendo un monosacárido adicional o grupos sulfato desde un sustrato donador a una molécula aceptora. La adición tiene lugar generalmente en el extremo no reductor de un oligosacárido, polisacárido (por ejemplo, heparina, carragenina y similares) o un resto de hidrato de carbono en un glicolípido o glicoproteína, por ejemplo, una biomolécula. Las biomoléculas tal como se definen en el presente documento incluyen pero no se limitan a moléculas biológicamente significativas tales como hidratos de carbono, oligosacáridos, péptidos (por ejemplo, glicopéptidos), proteínas (por ejemplo, glicoproteínas) y lípidos (por ejemplo, glicolípidos, fosfolípidos, esfingolípidos y gangliósidos).

Las proteínas recombinantes de la invención pueden construirse y expresarse como una proteína de fusión con una "etiqueta de purificación" molecular en un extremo, que facilita la purificación o identificación de la proteína. Tales etiquetas pueden usarse también para la inmovilización de una proteína de interés durante la reacción de glicosilación. Las etiquetas adecuadas incluyen "etiquetas de epítipo", que son una secuencia de proteína que se reconoce específicamente por un anticuerpo. Se incorporan generalmente etiquetas de epítipo en proteínas de fusión para permitir el uso de un anticuerpo fácilmente disponible para detectar o aislar de manera inequívoca la proteína de fusión. Una "etiqueta FLAG" es una etiqueta de epítipo comúnmente usada, reconocida específicamente por un anticuerpo monoclonal anti-FLAG, que consiste en la secuencia AspTyrLysAspAspAsp AspLys o una variante sustancialmente idéntica de la misma. Los expertos en la técnica conocen otras etiquetas adecuadas e incluyen, por ejemplo, una etiqueta de afinidad tal como un péptido de hexahistidina, que se unirá a iones de metal tales como iones cobalto o níquel o una etiqueta myc. Las proteínas que comprenden etiquetas de purificación pueden purificarse usando una pareja de unión que se une a la etiqueta de purificación, por ejemplo, anticuerpos frente a la etiqueta de purificación, iones níquel o cobalto o resinas, y amilosa, maltosa o una ciclodextrina. Las etiquetas de purificación también incluyen dominios de unión a maltosa y dominios de unión a almidón. Los expertos en la técnica conocen la purificación de proteínas con dominios de unión a maltosa. Se describen dominios de unión a maltosa en el documento WO 99/15636, incorporado en el presente documento como referencia. La purificación por afinidad de una proteína de fusión que comprende un dominio de unión a almidón usando una resina derivatizada con betaciclodextrina (BCD) se describe en el documento WO 2005/014779, publicado el 17 de febrero de 2005.

La expresión "ácido nucleico" se refiere a un polímero de desoxirribonucleótidos o ribonucleótidos en forma o bien mono o bien bicatenaria, y a menos que se limite de otra forma, abarca análogos conocidos de nucleótidos naturales que hibridan con ácidos nucleicos de manera similar a nucleótidos que se producen de manera natural. A menos que se indique lo contrario, una secuencia de ácido nucleico particular incluye las secuencias complementarias de la misma. Las expresiones "ácido nucleico", "secuencia de ácido nucleico" y "polinucleótido" se usan de manera intercambiable en el presente documento.

La expresión "operativamente unido" se refiere a la unión funcional entre una secuencia control de la expresión de ácido nucleico (tal como un promotor, una secuencia señal o una serie de sitios de unión de factores de transcripción) y una segunda secuencia de ácido nucleico, en la que la secuencia control de la expresión afecta a la transcripción y/o traducción del ácido nucleico correspondiente a la segunda secuencia.

El término "recombinante", cuando se usa con referencia a una célula, indica que la célula replica un ácido nucleico heterólogo, o expresa un péptido o una proteína codificada por un ácido nucleico heterólogo. Las células recombinantes pueden contener genes que no se encuentran dentro de la forma nativa (no recombinante) de la célula. Las células recombinantes pueden contener también genes que se encuentran en la forma nativa de la célula en la que los genes se modifican y se reintroducen en la células por medios artificiales. El término también abarca células que contienen un ácido nucleico endógeno para la célula que se ha modificado sin eliminar la forma de ácido nucleico de la célula; tales modificaciones incluyen las obtenidas mediante sustitución génica, mutación específica de sitio y técnicas relacionadas.

Un "ácido nucleico recombinante" se refiere a un ácido nucleico que se construyó artificialmente (por ejemplo, formado uniendo dos fragmentos de ácido nucleico sintéticos o que se producen de manera natural). Este término también se aplica a ácidos nucleicos que se producen mediante replicación o transcripción de un ácido nucleico que se construyó artificialmente. Un "polipéptido recombinante" se expresa mediante transcripción de un ácido nucleico recombinante (es decir, un ácido nucleico que no es nativo para la célula o que se ha modificado con respecto a su forma que se produce de manera natural), seguido por la traducción del transcrito resultante.

Un "polinucleótido heterólogo" o un "ácido nucleico heterólogo", tal como se usa en el presente documento, es uno que se origina a partir de una fuente foránea con respecto a la célula huésped particular o, si es a partir de la misma fuente, se modifica con respecto a su forma original. Por tanto, un gen de glicosiltransferasa heterólogo en una célula huésped procarionta incluye un gen de glicosiltransferasa que es endógeno para la célula huésped particular pero que se ha modificado. La modificación de la secuencia heteróloga puede producirse, por ejemplo, tratando el ADN con una enzima de restricción para generar un fragmento de ADN que puede estar operativamente unido a un promotor. Técnicas tales como mutagénesis dirigida al sitio también son útiles para modificar una secuencia heteróloga.

Una “subsecuencia” se refiere a una secuencia de ácidos nucleicos o aminoácidos que comprenden una parte de una secuencia más larga de ácidos nucleicos o aminoácidos (por ejemplo, polipéptido) respectivamente.

Un “casete de expresión recombinante” o simplemente un “casete de expresión” es una construcción de ácido nucleico, generada de manera recombinante o sintética, con elementos de ácido nucleico que pueden afectar a la expresión de un gen estructural en huéspedes compatibles con tales secuencias. Los casetes de expresión incluyen al menos promotores y opcionalmente señales de terminación de la transcripción. Normalmente, el casete de expresión recombinante incluye un ácido nucleico que va a transcribirse (por ejemplo, un ácido nucleico que codifica para un polipéptido deseado), y un promotor. Pueden usarse también factores adicionales necesarios o útiles para afectar a la expresión tal como se describe en el presente documento. Por ejemplo, un casete de expresión puede incluir también secuencias de nucleótidos que codifican para una secuencia señal que dirige la secreción de una proteína expresada desde la célula huésped. También pueden incluirse en un casete de expresión señales de terminación de la transcripción, potenciadores y otras secuencias de ácido nucleico que influyen en la expresión génica.

Un “polipéptido de CgtE de fusión” o un “polipéptido de glicosiltransferasa de fusión” de la invención es un polipéptido que contiene un dominio catalítico de glicosiltransferasa o CgtE y un segundo dominio catalítico de una enzima auxiliar (por ejemplo, UDP-GalNAc 4' epimerasa o una UDP-glucosa 4'epimerasa). El polipéptido de fusión puede catalizar la síntesis de un nucleótido de azúcar (por ejemplo, UDP-GalNAc o UDP-Galactosa) así como la transferencia del residuo de azúcar desde el nucleótido de azúcar hasta una molécula aceptora. Normalmente, los dominios catalíticos de los polipéptidos de fusión serán al menos sustancialmente idénticos a los de las glicosiltransferasas y proteínas de fusión de las que se derivan los dominios catalíticos. En algunas realizaciones, se usan un polipéptido de CgtE y una epimerasa, por ejemplo, UDP-GalNAc 4' UDP-glucosa 4', polipéptido se fusionan para formar un único polipéptido.

Una “enzima auxiliar”, tal como se denomina en el presente documento, es una enzima que está implicada en la catalización de una reacción que, por ejemplo, forma un sustrato u otro reactante para una reacción de glicosiltransferasa. Una enzima auxiliar, por ejemplo, puede catalizar la formación de un azúcar de nucleótido que se usa como resto donador de azúcar por una glicosiltransferasa. Una enzima auxiliar también puede ser una que se usa en la generación de un nucleótido trifosfato que se requiere para la formación de un azúcar de nucleótido, o en la generación del azúcar que se incorpora en el azúcar de nucleótido.

Un “dominio catalítico” se refiere a una parte de una enzima que es suficiente para catalizar una reacción enzimática que normalmente lleva a cabo la enzima. Por ejemplo, un dominio catalítico de un polipéptido de CgtE incluirá una parte suficiente de la CgtE para transferir un resto de galactosa desde un donador de azúcar hasta un sacárido aceptor. Un dominio catalítico puede incluir una enzima completa, una subsecuencia de la misma, o puede incluir secuencias de aminoácidos adicionales que no están unidas a la enzima o subsecuencia tal como se encuentra en la naturaleza.

El término “aislado” se refiere a material que está sustancial o esencialmente libre de componentes que interfieren con la actividad de una enzima. Para células, sacáridos, ácidos nucleicos y polipéptidos de la invención, el término “aislado” se refiere a material que está sustancial o esencialmente libre de componentes que normalmente acompañan al material tal como se encuentra en su estado nativo. Normalmente, los sacáridos, proteínas o ácidos nucleicos aislados de la invención son puros al menos aproximadamente al 50%, el 55%, el 60%, el 65%, el 70%, el 75%, el 80% o al 85%, habitualmente puros al menos aproximadamente al 90%, el 91%, el 92%, el 93%, el 94%, el 95%, el 96%, el 97%, el 98% o el 99% tal como se mide mediante la intensidad de bandas en un gel teñido con plata y otro métodos para determinar la pureza. La pureza u homogeneidad pueden indicarse por varios medios bien conocidos en la técnica, tales como electroforesis en gel de poli(acrilamida) de una muestra de ácido nucleico o proteína, seguido por visualización tras teñir. Para ciertos fines se necesitará una alta resolución y se utilizará HPLC o un medio similar para la purificación. Para oligonucleótidos, u otros productos galactosilados, la pureza puede determinarse usando, por ejemplo, cromatografía en capa fina, HPLC o espectrometría de masas.

Los términos “idéntico” o porcentaje de “identidad”, en el contexto de dos o más secuencias de ácido nucleico o polipéptido, se refieren a dos o más secuencias o subsecuencias que son iguales o tienen un porcentaje especificado de residuos de aminoácido o nucleótidos que son iguales, cuando se comparan y alinean para una correspondencia máxima, tal como se mide usando uno de los siguientes algoritmos de comparación de secuencias o mediante inspección visual.

La frase “sustancialmente idéntico”, en el contexto de dos ácido nucleicos o polipéptidos, se refiere a dos o más secuencias o subsecuencias tienen una identidad de residuos de aminoácidos o nucleótidos de al menos el 60%, preferiblemente el 80% o el 85%, lo más preferiblemente al menos el 90%, el 91%, el 92%, el 93%, el 94%, el 95%, el 96%, el 97%, el 98% o el 99%, cuando se comparan y alinean para correspondencia máxima, tal como se mide usando uno de los siguientes algoritmos de comparación de secuencias o mediante inspección visual. Preferiblemente, la identidad sustancial existe a lo largo de una región de las secuencias que tiene al menos aproximadamente 50 residuos de longitud, más preferiblemente a lo largo de una región de al menos aproximadamente 100 residuos, y lo más preferiblemente las secuencias son sustancialmente idénticas a lo largo de al menos aproximadamente 150 residuos. En la realización más preferida, las secuencias son sustancialmente

idénticas a lo largo de toda la longitud de las regiones codificantes.

Para la comparación de secuencias, normalmente una secuencia actúa como secuencia de referencia, con la que se comparan las secuencias de prueba. Cuando se usa un algoritmo de comparación de secuencias, las secuencias de prueba y de referencia se introducen en un ordenador, se designan las coordenadas de subsecuencias, si es necesario, y se designan los parámetros del programa de algoritmo de secuencias. El algoritmo de comparación de secuencias calcula entonces el porcentaje de identidad de secuencia para la(s) secuencia(s) de prueba en relación con la secuencia de referencia, basándose en los parámetros de programa designados.

Puede realizarse la alineación óptima de secuencias para comparación, por ejemplo, mediante el algoritmo de homología local de Smith y Waterman, *Adv. Appl. Math.* 2:482 (1981), mediante el algoritmo de alineación de homología de Needleman y Wunsch, *J. Mol. Biol.* 48:443 (1970), mediante el método de búsqueda de similitud Pearson y Lipman, *Proc. Nat'l. Acad. Sci. USA* 85:2444 (1988), mediante implementaciones informatizadas de estos algoritmos (GAP, BESTFIT, FASTA y TFASTA en el Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr., Madison, WI), o mediante inspección visual (véase en general, *Current Protocols in Molecular Biology*, F.M. Ausubel *et al.*, eds., Current Protocols, una empresa conjunta entre Greene Publishing Associates, Inc. y John Wiley & Sons, Inc., (suplemento de 1995) (Ausubel)).

Ejemplos de algoritmos que son adecuados para determinar el porcentaje de identidad de secuencia y similitud de secuencia son los algoritmos BLAST y BLAST 2.0, que se describen en Altschul *et al.* (1990) *J. Mol. Biol.* 215: 403-410 y Altschul *et al.* (1977) *Nucleic Acids Res.* 25: 3389-3402, respectivamente. El software para realizar análisis BLAST está disponible públicamente a través del National Center for Biotechnology Information ([www.ncbi.nlm.nih.gov/](http://www.ncbi.nlm.nih.gov/)). Este algoritmo implica en primer lugar identificar pares de secuencias de alta puntuación (HSP) identificando palabras cortas de longitud W en la secuencia de consulta, que o bien coinciden o bien satisfacen algún valor umbral de valor positivo T cuando se alinean con una palabra de la misma longitud en una secuencia de la base de datos. T se denomina umbral de puntuación de palabra vecina (Altschul *et al.*, citado anteriormente). Estos aciertos de palabras vecinas iniciales actúan como simientes para iniciar búsquedas para encontrar HSP más largos que los contienen. Los aciertos de palabras se extienden entonces en ambas direcciones a lo largo de cada secuencia durante tanto como pueda aumentarse la puntuación de alineación acumulativa. Se calculan las puntuaciones acumulativas usando, para secuencias de nucleótidos, los parámetros M (puntuación de recompensa para un par de residuos coincidentes; siempre > 0) y N (puntuación de penalización para residuos no coincidentes; siempre < 0). Para secuencias de aminoácidos, se usa una matriz de puntuación para calcular la puntuación acumulativa. La extensión de los aciertos de palabras en cada dirección se detiene cuando: la puntuación de alineación acumulativa disminuye en la cantidad X con respecto a su valor logrado máximo; la puntuación acumulativa llega a cero o inferior, debido a la acumulación de una o más alineaciones de residuos que puntúan negativo; o se alcanza el extremo de cualquier secuencia. Los parámetros del algoritmo BLAST W, T y X determinan la sensibilidad y velocidad de la alineación. El programa BLASTN (para secuencias de nucleótidos) usa por defecto una longitud de palabra (W) de 11, una expectativa (E) de 10, M=5, N=-4 y una comparación ambas hebras. Para secuencias de aminoácidos, el programa BLASTP usa por defecto una longitud de palabra (W) de 3, una expectativa (E) de 10 y la matriz de puntuación BLOSUM62 (véase Henikoff y Henikoff, *Proc. Natl. Acad. Sci. USA* 89:10915 (1989)).

Además de calcular el porcentaje de identidad de secuencia, el algoritmo BLAST también realiza un análisis estadístico de la similitud entre dos secuencias (véase, por ejemplo, Karlin y Altschul, *Proc. Nat'l. Acad. Sci. USA* 90:5873-5787 (1993)). Una medida de similitud proporcionada por el algoritmo BLAST es la probabilidad de suma más pequeña (P(N)), que proporciona una indicación de la probabilidad por la cual una coincidencia entre dos secuencias de nucleótidos o aminoácidos pudiese producirse por casualidad. Por ejemplo, un ácido nucleico se considera similar a una secuencia de referencia si la probabilidad de suma más pequeña en una comparación del ácido nucleico de prueba con el ácido nucleico de referencia es inferior a aproximadamente 0,1, más preferiblemente inferior a aproximadamente 0,01 y lo más preferiblemente inferior a aproximadamente 0,001.

Una indicación adicional de que dos secuencia de ácido nucleico o polipéptidos son sustancialmente idénticos es que el polipéptido codificado por el primer ácido nucleico reaccione inmunológicamente de manera cruzada con el polipéptido codificado por el segundo ácido nucleico, tal como se describe a continuación. Por tanto, normalmente, un polipéptido es sustancialmente idéntico a un segundo polipéptido, por ejemplo, cuando los dos péptidos difieren sólo por sustituciones conservativas. Otra indicación de que dos secuencias de ácido nucleico son sustancialmente idénticas es que las dos moléculas hibridan entre sí en condiciones rigurosas, tal como se describe a continuación.

La frase "que hibrida específicamente con" se refiere a la unión, formación de dúplex o hibridación de una molécula sólo con una secuencia de nucleótidos particular en condiciones rigurosas cuando la secuencia está presente en una mezcla compleja de ADN o ARN (por ejemplo, celular total).

La expresión "condiciones rigurosas" se refiere a condiciones en las que una sonda hibridará con su subsecuencia diana, pero no con otras secuencias. Las condiciones rigurosas dependen de la secuencia y serán diferentes en circunstancias diferentes. Secuencias más largas hibridarán específicamente a temperaturas superiores. Generalmente, se selecciona que las condiciones rigurosas sean aproximadamente 5°C inferiores al punto de fusión térmico (T<sub>m</sub>) para la secuencia específica a un pH y fuerza iónica definidos. El T<sub>m</sub> es la temperatura (a

concentración de ácido nucleico, pH y fuerza iónicas definidos) a la que el 50% de las sondas complementarias a la secuencia diana se hibridan con la secuencia diana en equilibrio. (Como las secuencias diana están presentes generalmente en exceso, a  $T_m$ , el 50% de las sondas están ocupadas en equilibrio). Normalmente, condiciones rigurosas serán aquellas en las que la concentración salina es inferior a aproximadamente  $1,0 \text{ M}$   $\text{Na}^+$ , normalmente una concentración de  $\text{Na}^+$  de aproximadamente  $0,01$  a  $1,0 \text{ M}$  (u otras sales) a pH de  $7,0$  a  $8,3$  y la temperatura es al menos de aproximadamente  $30^\circ\text{C}$  para sondas cortas (por ejemplo, de  $10$  a  $50$  nucleótidos) y al menos de aproximadamente  $60^\circ\text{C}$  para sondas largas (por ejemplo, mayores de  $50$  nucleótidos). También pueden lograrse condiciones rigurosas con la adición de agentes desestabilizantes tales como formamida. Para la amplificación por PCR de alta rigurosidad, es normal una temperatura de aproximadamente  $62^\circ\text{C}$ , aunque las temperaturas de apareamiento de alta rigurosidad pueden oscilar entre aproximadamente  $50^\circ\text{C}$  y aproximadamente  $65^\circ\text{C}$ , dependiendo de la longitud y especificidad del cebador. Las condiciones de ciclo típicas para amplificaciones tanto de alta como de baja rigurosidad incluyen una fase de desnaturalización de  $90$ - $95^\circ\text{C}$  durante  $30$ - $120$  s, una fase de apareamiento que dura  $30$ - $120$  s y una fase de extensión de aproximadamente  $72^\circ\text{C}$  durante  $1$ - $2$  min. Están disponibles protocolos y directrices para reacciones de amplificación de baja y alta rigurosidad, por ejemplo, en Innis, *et al.* (1990) PCR Protocols: A Guide to Methods and Applications Academic Press, N.Y.

Las frases “se une específicamente a” o “inmunorreactivo específicamente con”, cuando se hace referencia a un anticuerpo, se refieren a una reacción de unión que determina la presencia de la proteína u otro antígeno en presencia de una población heterogénea de proteínas, sacáridos y otros compuestos biológicos. Por tanto, en las condiciones de inmunoensayo designadas, los anticuerpos especificados se unen preferentemente a un antígeno particular y no se unen en una cantidad significativa a otras moléculas presentes en la muestra. La unión específica a un antígeno en tales condiciones requiere un anticuerpo que se selecciona por su especificidad para un antígeno particular. Puede usarse una variedad de formatos de inmunoensayo para seleccionar anticuerpos que inmunorreactan específicamente con un antígeno particular. Por ejemplo, se usan de manera rutinaria inmunoensayos ELISA en fase sólida para seleccionar anticuerpos monoclonales específicamente inmunorreactivos con un antígeno. Véase Harlow y Lane (1988) Antibodies, A Laboratory Manual, Cold Spring Harbor Publications, Nueva York, para una descripción de formatos de inmunoensayo y condiciones que pueden usarse para determinar la inmunoreactividad específica.

“Anticuerpo” se refiere a un polipéptido que comprende una región de entramado de un gen de inmunoglobulina o fragmentos del mismo que se une y reconoce específicamente un antígeno. Los genes de inmunoglobulina reconocidos incluyen los genes de regiones constantes kappa, lambda, alfa, gamma, delta, épsilon y mu, así como los miles de genes de regiones variables de inmunoglobulinas. En una realización preferida, se producen anticuerpos que se unen específicamente a una proteína CgE. Las cadenas ligeras se clasifican como o bien kappa o bien lambda. Las cadenas pesadas se clasifican como gamma, mu, alfa, delta o épsilon, que a su vez definen las clases de inmunoglobulinas, IgG, IgM, IgA, IgD e IgE, respectivamente. Normalmente, la región de unión a antígeno de un anticuerpo será la más crítica en especificidad y afinidad de unión.

Una unidad estructural de inmunoglobulina (anticuerpo) a modo de ejemplo comprende un tetrámero. Cada tetrámero está compuesto por dos pares idénticos de cadenas polipeptídicas, teniendo cada par una cadena “ligera” (aproximadamente  $25 \text{ kD}$ ) y una cadena “pesada” (aproximadamente  $50$ - $70 \text{ kD}$ ). El extremo N-terminal de cada cadena define una región variable de aproximadamente  $100$  a  $110$  o más aminoácidos principalmente responsables del reconocimiento del antígeno. Las expresiones cadena ligera variable ( $V_L$ ) y cadena pesada variable ( $V_H$ ) se refieren a estas cadenas ligera y pesada respectivamente.

Los anticuerpos existen, por ejemplo, como inmunoglobulinas intactas o como varios fragmentos bien caracterizados producidos mediante digestión con diversas peptidasas. Por tanto, por ejemplo, la pepsina digiere un anticuerpo por debajo de los puentes disulfuro en la región de bisagra para producir  $F(ab)_2$ , un dímero de Fab que en sí mismo es una cadena ligera unida a  $V_H$ - $C_H1$  mediante un puente disulfuro. El  $F(ab)_2$  puede reducirse en condiciones suaves para romper el punto disulfuro en la región de bisagra, convirtiendo de ese modo el dímero  $F(ab)_2$  en un monómero Fab'. El monómero Fab' es esencialmente Fab con parte de la región de bisagra (véase Fundamental Immunology (Paul ed., 3ª ed. 1993). Mientras que diversos fragmentos de anticuerpo se definen en cuanto a la digestión de un anticuerpo intacto, un experto apreciará que tales fragmentos pueden sintetizarse *de novo* o bien químicamente o bien usando metodología de ADN recombinante. Por tanto, el término anticuerpo, tal como se usa en el presente documento, también incluye fragmentos de anticuerpo o bien producidos por la modificación de los anticuerpos completos, o los sintetizados *de novo* usando metodologías de ADN recombinante (por ejemplo, Fv de cadena sencilla) o bien los identificados usando bibliotecas de presentación en fago (véase, por ejemplo, McCafferty *et al.*, Nature 348:552-554 (1990))

Para la preparación de anticuerpos, por ejemplo, anticuerpos recombinantes, monoclonales o policlonales, pueden usarse muchas técnicas conocidas en la técnica (véase, por ejemplo, Kohler y Milstein, Nature 256:495-497 (1975); Kozbor *et al.*, Immunology Today 4: 72 (1983); Cole *et al.*, págs. 77-96 en Monoclonal Antibodies and Cancer Therapy, Alan R. Liss, Inc. (1985); Coligan, Current Protocols in Immunology (1991); Harlow y Lane, Antibodies, A Laboratory Manual (1988); y Goding, Monoclonal Antibodies: Principles and Practice (2ª ed. 1986)). Los genes que codifican para las cadenas pesada y ligera de un anticuerpo de interés pueden clonarse a partir de una célula, por ejemplo, los genes que codifican para un anticuerpo monoclonal pueden clonarse a partir de un híbrido y usarse para producir un anticuerpo monoclonal recombinante. También pueden prepararse bibliotecas génicas que

codifican para cadenas pesadas y ligeras de anticuerpos monoclonales a partir de células plasmáticas o de hibridoma. Combinaciones al azar de los productos génicos de cadenas pesadas y ligeras generan un gran conjunto de anticuerpos con especificidad antigénica diferente (véase, por ejemplo, Kuby, Immunology (3ª ed. 1997)). Pueden adaptarse técnicas para la producción de anticuerpos de cadena sencilla o anticuerpos recombinantes (patente estadounidense n.º 4.946.778, patente estadounidense n.º 4.816.567) para producir anticuerpos frente a polipéptidos de esta invención. Además, pueden usarse ratones transgénicos, u otros organismos tales como otros mamíferos no humanos, para expresar anticuerpos humanos o humanizados (véanse, por ejemplo, las patentes estadounidenses n.ºs 5.545.807; 5.545.806; 5.569.825; 5.625.126; 5.633.425; 5.661.016, Marks *et al.*, Bio /Technology 10:779-783 (1992); Lonberg *et al.*, Nature 368:856-859 (1994); Morrison, Nature 368:812-13 (1994); Fishwild *et al.*, Nature Biotechnology 14:845-51 (1996); Neuberger, Nature Biotechnology 14:826 (1996); y Lonberg y Huszar, Intern. Rev. Immunol. 13:65-93 (1995)). Alternativamente, puede usarse la tecnología de presentación en fago para identificar anticuerpos y fragmentos Fab heteroméricos que se unen específicamente a antígenos seleccionados (véase, por ejemplo, McCafferty *et al.*, Nature 348:552-554 (1990); Marks *et al.*, Biotechnology 10:779-783 (1992)). También pueden prepararse anticuerpos biespecíficos, es decir, que pueden reconocer dos antígenos diferentes (véanse, por ejemplo, el documento WO 93/08829, Traunecker *et al.*, EMBO J. 10:3655-3659 (1991); y Suresh *et al.*, Methods in Enzymology 121: 210 (1986)). Los anticuerpos también pueden ser heteroconjugados, por ejemplo, dos anticuerpos covalentemente unidos o inmunotoxinas (véanse, por ejemplo, la patente estadounidense n.º 4.676.980, el documento WO 91/00360; el documento WO 92/200373; y el documento EP 03089).

Un anticuerpo puede conjugarse con un resto "efector". El resto efector puede ser cualquiera de varias moléculas, incluyendo restos de marcaje tales como marcadores radiactivos o marcadores fluorescentes para su uso en ensayos de diagnóstico.

La frase "se une específicamente (o selectivamente)" a un anticuerpo o "específicamente (selectivamente) inmunorreactivo con", cuando se hace referencia a una proteína o un péptido, se refiere a una reacción de unión que determina la presencia de la proteína, a menudo en una población heterogénea de proteínas y otros compuestos biológicos. Por tanto, en las condiciones de inmunoensayo designadas, los anticuerpos especificados se unen a una proteína particular al menos dos veces el fondo y más normalmente más de 10 a 100 veces el fondo. La unión específica a un anticuerpo en tales condiciones requiere un anticuerpo que se selecciona por su especificidad para una proteína particular. Por ejemplo, pueden seleccionarse anticuerpos policlonales generados frente a proteína de IgE, variantes polimórficas, alelos, ortólogos y variantes modificadas de manera conservativa, o variantes de corte y empalme, o partes de los mismos, para obtener sólo los anticuerpos policlonales que son específicamente inmunorreactivos con proteínas de IgE y no con otras proteínas. Esta selección puede lograrse restando anticuerpos que reaccionan de manera cruzada con otras moléculas. Puede usarse una variedad de formatos de inmunoensayo para seleccionar anticuerpos específicamente inmunorreactivos con una proteína particular. Por ejemplo, se usan de manera rutinaria inmunoensayos ELISA en fase sólida para seleccionar anticuerpos específicamente inmunorreactivos con una proteína (véase, por ejemplo, Harlow y Lane, Antibodies, A Laboratory Manual (1988) para una descripción de formatos y condiciones de inmunoensayo que pueden usarse para determinar la inmunorreactividad específica).

Un "antígeno" es una molécula reconocida por y a la que se une un anticuerpo, por ejemplo, péptidos, hidratos de carbono, moléculas orgánicas o moléculas más complejas tales como glicolípidos y glicoproteínas. La parte del antígeno que es la diana de la unión del anticuerpo es un determinante antigénico y un grupo funcional pequeño que corresponde a un único determinante antigénico se denomina hapteno.

Un "marcador" es una composición detectable por medios espectroscópicos, fotoquímicos, bioquímicos, inmunoquímicos o químicos. Por ejemplo, marcadores útiles incluyen <sup>32</sup>P, <sup>125</sup>I, tintes fluorescentes, reactivos electrodenso, enzimas (por ejemplo, tal como se usan comúnmente en un ELISA), biotina, digoxigenina o haptenos y proteínas para los que están disponibles antisueros o anticuerpos monoclonales (por ejemplo, el polipéptido de SEQ ID NO:2 puede hacerse detectable, por ejemplo, incorporando un radiomarcador en el péptido, y usarse para detectar anticuerpos específicamente reactivos con el péptido).

El término "inmunoensayo" es un ensayo que usa un anticuerpo para unir específicamente un antígeno. El inmunoensayo se caracteriza por el uso de propiedades de unión específicas de un anticuerpo particular para aislar, seleccionar como diana y/o cuantificar el antígeno.

La expresión "molécula portadora" significa una molécula inmunogénica que contiene determinantes antigénicos reconocidos por células T. Una molécula portadora puede ser una proteína o puede ser un lípido. Una proteína portadora se conjuga con un polipéptido para hacer que el polipéptido sea inmunogénico. Las proteínas portadoras incluyen hemocianina de lapa californiana, hemocianina de cangrejo de herradura y albúmina sérica bovina.

El término "adyuvante" significa una sustancia que potencia de manera no específica la respuesta inmunitaria a un antígeno. Los adyuvantes incluyen adyuvante de Freund, o bien completo o bien incompleto; adyuvante de oro Titermax; alumbre; y LPS bacteriano.

La expresión "poner en contacto" se usa en el presente documento de manera intercambiable con las siguientes: combinado con, añadido a, mezclado con, hecho pasar sobre, incubado con, hecho fluir sobre, etc.

### III. Polipéptidos de CgtE

Los polipéptidos de CgtE polipéptidos de las invenciones comprenden una secuencia de aminoácidos que está relacionada con un dominio de proteína conservado, el dominio glycosyl\_transf\_2, número de registro PF00535. La familia de proteínas de glicosiltransferasa 2 es un grupo de proteínas que se identificaron como tales basándose en comparaciones de secuencias. Véase, por ejemplo, Campbell *et al.*, Biochem. J. 326: 929-942 (1997). Se proporciona una alineación de los residuos de aminoácido 5-167 de CgtE con la secuencia del dominio glycosyl\_transf\_2 en la figura 3. Se encuentra un dominio relacionado con glycosyl\_transf\_2 en proteínas que transfieren azúcar desde UDP-glucosa, UDP-N-acetilgalactosamina, GDP-manosa o CDP-abecucosa hasta una gama de sustratos incluyendo celulosa, dolicol fosfato y ácidos teicoicos. Véase, por ejemplo, [www.sanger.ac.uk/cgi-bin/Pfam/getacc?PF00535](http://www.sanger.ac.uk/cgi-bin/Pfam/getacc?PF00535) y Coutinho, P.M. & Henrissat, B. (1999) Carbohydrate-Active Enzymes server at URL: [afmb.cnrs-mrs.fr/CAZY/](http://afmb.cnrs-mrs.fr/CAZY/).

En la actualidad se han identificado 2395 proteínas que contienen un dominio relacionado con glycosyl\_transf\_2 en eucariotas, por ejemplo, mamíferos, metazoos, anfibios, artrópodos, plantas, hongos y bacterias. El dominio relacionado con glycosyl\_transf\_2 corresponde a un pliegue tridimensional que se cree que se produce en cada una de las proteínas y, por tanto, está disponible un gran número de secuencias para comparaciones de secuencia y estructura, además del dominio glycosyl\_transf\_2. Véase por ejemplo, [www.sanger.ac.uk/cgi-bin/Pfam/getacc?PF00535](http://www.sanger.ac.uk/cgi-bin/Pfam/getacc?PF00535). La estructura del dominio tridimensional conservado se conoce y está disponible para análisis de función estructural. Véase, por ejemplo, protein databank (PDB) números 1h71, 1h7q, 1qgq, 1qg8, y 1qgs; y Tarbouriech, *et al.*, J. Mol. Biol. 314:655- (2001).

Programas informáticos que comparan secuencias previamente desconocidas tales como las secuencias de aminoácidos de CgtE con secuencias conocidas, tales como el dominio glycosyl\_transf\_2, están libremente disponibles para los expertos. Un programa de este tipo es Cn3D que puede descargarse de [www.ncbi.nlm.nih.gov/Structure/CN3D/cn3d.shtml](http://www.ncbi.nlm.nih.gov/Structure/CN3D/cn3d.shtml). El Cn3D correlaciona la información estructural y de secuencia: por ejemplo, un científico puede encontrar rápidamente los residuos en una estructura cristalina que corresponden a mutaciones patológicas conocidas, o residuos de sitios activos conservados a partir de una familia de homólogos de secuencia. El Cn3D presenta visualmente alineaciones de estructura-estructura junto con sus alineaciones de secuencia basadas en la estructura, para enfatizar qué regiones de un grupo de proteínas relacionadas son las más conservadas en estructura y secuencia. Por tanto, usando un programa tal como Cn3D, los expertos pueden identificar residuos conservados en el dominio relacionado con glycosyl\_transf\_2 de CgtE y, además, pueden predecir cambios en residuos de aminoácido que probablemente no afectarían a la actividad de la proteína. Además, usando el programa Cn3D, los expertos podrían predecir también cambios en residuos de aminoácido que serían perjudiciales para la actividad de CgtE y evitarlos.

Como mínimo, los polipéptidos de CgtE comprenden los residuos de aminoácido 5-167 de SEQ ID NO:2, es decir, el dominio relacionado con glycosyl\_transf\_2 tal como se describió anteriormente.

### IV. Aislamiento de ácidos nucleicos que codifican para polipéptidos de CgtE

Los ácidos nucleicos que codifican para polipéptidos de CgtE incluyen ácidos nucleicos que codifican para los polipéptidos de CgtE descritos anteriormente, es decir, SEQ ID NO:2, y variantes modificadas de manera conservativa de esa secuencia. Los polipéptidos de CgtE de la invención catalizan la transferencia de un resto de galactosa desde un sustrato donador hasta un sustrato aceptor.

Los expertos en la técnica conocen ácidos nucleicos que codifican para polipéptidos de CgtE adicionales basándose en la información dada a conocer en el presente documento, y métodos de obtención de tales ácidos nucleicos. Pueden clonarse ácidos nucleicos adecuados (por ejemplo, ADNc, genómico o subsecuencias (sondas)), o amplificarse mediante métodos *in vitro* tales como la reacción en cadena de la polimerasa (PCR), la reacción en cadena de la (LCR), el sistema de amplificación basado en transcripción (TAS) o el sistema de replicación de secuencias autosostenida (SSR). Los expertos conocen bien una amplia variedad de metodologías de clonación y amplificación *in vitro*. Se encuentran ejemplos de estas técnicas e instrucciones suficientes para dirigir a personas con experiencia a través de muchos ejercicios de clonación en Berger y Kimmel, Guide to Molecular Cloning Techniques. Methods in Enzymology 152 Academic Press, Inc., San Diego, CA (Berger); Sambrook *et al.* (1989) Molecular Cloning - A Laboratory Manual (2ª ed.) vol. 1-3, Cold Spring Harbor Laboratory, Cold Spring Harbor Press, NY, (Sambrook *et al.*); Current Protocols in Molecular Biology, F.M. Ausubel *et al.*, eds., Current Protocols, una empresa conjunta entre Greene Publishing Associates, Inc. y John Wiley & Sons, Inc., (suplemento de 1994) (Ausubel); Cashion *et al.*, patente estadounidense número 5.017.478; y Carr, patente europea n.º 0.246.864.

Puede prepararse un ADN que codifica para un polipéptido de CgtE polipéptido, o subsecuencias del mismo, mediante cualquier método adecuado descrito anteriormente, incluyendo, por ejemplo, clonación y restricción de secuencias apropiadas con enzimas de restricción. En una realización preferida, se aíslan ácidos nucleicos que codifican para polipéptidos de CgtE mediante métodos de clonación rutinarios. Una secuencia de nucleótidos de un polipéptido de CgtE tal como se proporciona en, por ejemplo, SEQ ID NO:2, puede usarse para proporcionar sondas

que hibridan específicamente con un gen que codifica para un polipéptido de CgtE en una muestra de ADN genómico; o con un ARNm, que codifica para un polipéptido de CgtE que comprende, en una muestra de ARN total (por ejemplo, en una transferencia de tipo Southern o Northern). Una vez que se identifica el ácido nucleico que codifica para un polipéptido de CgtE, puede aislarse según métodos convencionales conocidos por los expertos en la técnica (véase, por ejemplo, Sambrook *et al.* (1989) *Molecular Cloning: A Laboratory Manual*, 2ª Ed., vols. 1-3, Cold Spring Harbor Laboratory; Berger y Kimmel (1987) *Methods in Enzymology*, vol. 152: *Guide to Molecular Cloning Techniques*. San Diego: Academic Press, Inc.; o Ausubel *et al.* (1987) *Current Protocols in Molecular Biology*, Greene Publishing y Wiley-Interscience, Nueva York). Además, los ácidos nucleicos aislados pueden escindirse con enzimas de restricción para crear ácidos nucleicos que codifican para el polipéptido de CgtE de longitud completa, o subsecuencias del mismo, por ejemplo, que contienen subsecuencias que codifican para al menos una subsecuencia de un dominio catalítico de un polipéptido B de CgtE. Estos fragmentos de enzimas de restricción, que codifican para un polipéptido de CgtE o subsecuencias del mismo, pueden ligarse entonces, por ejemplo, para producir un ácido nucleico que codifica para una proteína CgtE.

Un ácido nucleico que codifica para un polipéptido de CgtE, o una subsecuencia del mismo, puede caracterizarse sometiendo a ensayo el producto expresado. Pueden usarse ensayos basados en la detección de las propiedades físicas, químicas o inmunológicas de la proteína expresada. Por ejemplo, puede identificarse un ácido nucleico B de CgtE clonado por la capacidad de una proteína codificada por el ácido nucleico para catalizar la transferencia de un resto de galactosa desde un sustrato donador hasta un sustrato aceptor. En un método, se emplea electroforesis capilar para detectar los productos de reacción. Este ensayo altamente sensible implica el uso de derivados de aminofenilo de o bien sacárido o bien disacárido que se marcan con fluoresceína tal como se describe en Wakarchuk *et al.* (1996) *J. Biol. Chem.* 271 (45): 28271-276. Para someter a ensayo la actividad de CgtE, se usa Gal $\alpha$ -Lac-FCHASE o GalNAc- $\alpha$ -FCHASE como sustrato. Pueden detectarse los productos de reacción de otras glicosiltransferasas usando electroforesis capilar, por ejemplo, para someter a ensayo una enzima IgtC de *Neisseria*, pueden usarse o bien FCHASE-AP-Lac o bien FCHASE-AP-Gal, mientras que para la enzima IgtB de *Neisseria* un reactivo apropiado es FCHASE-AP-GlcNAc (Wakarchuk, citado anteriormente). Para someter a ensayo  $\alpha$ 2,8-sialiltransferasa, se usa GM3-FCHASE como sustrato. Véase, por ejemplo, la patente estadounidense n.º 6.503.744. Otros métodos para la detección de productos de reacción de oligosacárido incluyen cromatografía en capa fina y CG/EM y se dan a conocer en la patente estadounidense n.º 6.503.744.

Además, un ácido nucleico que codifica para un polipéptido de CgtE, o una subsecuencia del mismo, puede sintetizarse químicamente. Los métodos adecuados incluyen el método de fosfotriéster de Narang *et al.* (1979) *Me/A. Enzymol.* 68: 90-99; el método de fosfodiéster de Brown *et al.* (1979) *Meth. Enzymol.* 68: 109-151; el método de dietilfosforamida de Beaucage *et al.* (1981) *Tetra. Lett.*, 22: 1859-1862; y el método de soporte sólido de la patente estadounidense n.º 4.458.066. La síntesis química produce un oligonucleótido monocatenario. Éste puede convertirse en ADN bicatenario mediante hibridación con una secuencia complementaria, o mediante polimerización con una ADN polimerasa que usa el oligonucleótido monocatenario como molde. Un experto reconoce que aunque la síntesis química de ADN se limita a menudo a secuencias de aproximadamente 100 bases, pueden obtenerse secuencias más largas mediante la ligación de secuencias más cortas.

Pueden clonarse ácidos nucleicos que codifican para polipéptidos de CgtE, o subsecuencias de los mismos, usando métodos de amplificación de ADN tales como reacción en cadena de la polimerasa (PCR). Por tanto, por ejemplo, la secuencia o subsecuencia de ácido nucleico se amplifica por PCR, usando un cebador homocodificado que contiene un sitio de enzima de restricción (por ejemplo, *NdeI*) y un cebador antisentido que contiene otro sitio de enzima de restricción (por ejemplo, *HindIII*). Esto producirá un ácido nucleico que codifica para el polipéptido de CgtE deseado o una subsecuencia y que tiene sitios de enzimas de restricción terminales. Este ácido nucleico puede ligarse entonces fácilmente en un vector que contiene un ácido nucleico que codifica para la segunda molécula y que tiene los sitios de enzimas de restricción correspondientes apropiados. El experto en la técnica puede determinar cebadores de PCR adecuados usando la información de secuencia proporcionada en GenBank u otras fuentes. También pueden añadirse sitios de enzimas de restricción apropiados al ácido nucleico que codifica para la proteína CgtE o una subsecuencia de proteína de la misma mediante mutagénesis dirigida al sitio. El plásmido que contiene la secuencia o subsecuencia de nucleótidos que codifica para la proteína CgtE se escinde con la endonucleasa de restricción apropiada y se liga entonces en un vector apropiado para su amplificación y/o expresión según métodos convencionales. Se encuentran ejemplos de técnicas suficientes para dirigir a personas de experiencia a través de métodos de amplificación *in vitro* en Berger, Sambrook *et al.* y Ausubel (tal como se hizo referencia anteriormente, véase el párrafo 68), así como Mullis *et al.*, (1987) patente estadounidense n.º 4.683.202; PCR Protocols A Guide to Methods and Applications (Innis *et al.*, eds) Academic Press Inc. San Diego, CA (1990) (Innis); Arnheim & Levinson (1 de octubre de 1990) *C&EN* 36-47; *The Journal of NIH Research* (1991) 3: 81-94; (Kwoh *et al.* (1989) *Proc. Natl. Acad. Sci. USA* 86: 1173; Guatelli *et al.* (1990) *Proc. Natl. Acad. Sci. USA* 87, 1874; Lomell *et al.* (1989) *J. Clin. Chem.*, 35: 1826; Landegren *et al.*, (1988) *Science* 241: 1077-1080; Van Brunt (1990) *Biotechnology* 8: 291-294; Wu y Wallace (1989) *Gene* 4: 560; y Barringer *et al.* (1990) *Gene* 89: 117.

Algunos ácidos nucleicos que codifican para proteínas CgtE bacterianas pueden amplificarse usando cebadores de PCR basados en la secuencia de ácidos nucleicos de CgtE dados a conocer en el presente documento. Los ejemplos de cebadores de PCR que pueden usarse para amplificar el ácido nucleico que codifica para proteínas CgtE incluyen los siguientes pares de cebadores:

cebador 5-prima con un sitio *NdeI*: CJ-640: TTTAAGAAAACATATGCCTAAAAATTTCAATCATC

cebador 3-prima con un sitio *SalI*: CJ-641: GGTAATCTAGTCGACAATTATAACACATTC.

En algunas bacterias, pueden aislarse ácidos nucleicos que codifican para la proteína CgtE amplificando un locus cromosómico específico, por ejemplo, el locus LOS de *C. jejuni*, e identificando entonces un ácido nucleico de CgtE que se encuentra normalmente en ese locus (véase, por ejemplo, la patente estadounidense n.º 6.503.744). Los ejemplos de cebadores de PCR que pueden usarse para amplificar un locus LOS que comprende ácidos nucleicos que codifican para una proteína CgtE incluyen los siguientes pares de cebadores:

CJ42: Cebador en heptosilTasa-II

5' GC CAT TAC CGT ATC GCC TAA CCA GG 3' 25 meros

CJ43: Cebador en heptosilTasa-I

5' AAA GAA TAC GAA TTT GCT AAA GAG G 3' 25 meros

Otras propiedades físicas de un polipéptido de CgtE recombinante expresado a partir de un ácido nucleico particular pueden compararse con propiedades de polipéptidos de CgtE conocidos para proporcionar otro método de identificación de secuencias o dominios adecuados del polipéptido de CgtE que son determinantes de la especificidad de sustrato aceptor y/o actividad catalítica. Alternativamente, puede mutarse un polipéptido de CgtE supuesto o un polipéptido de CgtE recombinante, y establecerse su papel como glicosiltransferasa, o el papel de secuencias o dominios particulares detectando una variación en la estructura de un hidrato de carbono producido de manera normal por el polipéptido de CgtE no mutado, que se produce de manera natural o control. Los expertos reconocerán que la mutación o modificación de polipéptidos de CgtE de la invención puede facilitarse mediante técnicas de biología molecular para manipular los ácidos nucleicos que codifican para los polipéptidos de CgtE, por ejemplo, PCR.

Pueden identificarse dominios funcionales de polipéptidos de CgtE recién identificados usando métodos convencionales para mutar o modificar los polipéptidos y someterlos a prueba para detectar actividades tales como actividad de sustrato aceptor y/o actividad catalítica, tal como se describe en el presente documento. Los dominios funcionales de los diversos polipéptidos de CgtE pueden usarse para construir ácidos nucleicos que codifican para polipéptidos de CgtE y los dominios funcionales de uno o más polipéptidos de CgtE. Estas proteínas de fusión de multi-CgtE pueden someterse a prueba entonces para detectar la actividad catalítica o de sustrato aceptor deseada.

En un enfoque a modo de ejemplo para clonar ácidos nucleicos que codifican para proteínas CgtE, las secuencias de aminoácidos o ácido nucleico conocidas de polipéptidos de CgtE clonados se alinean y comparan para determinar la cantidad de identidad de secuencia entre diversos polipéptidos de CgtE. Esta información puede usarse para identificar y seleccionar dominios de proteínas que confieren o modulan las actividades de CgtE, por ejemplo, actividad de sustrato aceptor y/o actividad catalítica basándose en la cantidad de identidad de secuencia entre las proteínas CgtE de interés. Por ejemplo, pueden usarse dominios que tienen identidad de secuencia entre las proteínas CgtE de interés, y que están asociadas con una actividad conocida, para construir proteínas CgtE que contienen ese dominio, y que tiene la actividad asociada con ese dominio (por ejemplo, especificidad de sustrato aceptor y/o actividad catalítica).

#### V. Expresión de polipéptidos de CgtE en células huésped

Pueden expresarse proteínas CgtE de la invención en una variedad de células huésped, incluyendo *E. coli*, otros huéspedes bacterianos y levaduras. Las células huésped son preferiblemente microorganismos, tales como, por ejemplo, células de levaduras, células bacterianas o células fúngicas filamentosas. Los ejemplos de células huésped adecuadas incluyen, por ejemplo, *Azotobacter sp.* (por ejemplo, *A. vinelandii*), *Pseudomonas sp.*, *Rhizobium sp.*, *Erwinia sp.*, *Escherichia sp.* (por ejemplo, *E. coli*), *Bacillus*, *Pseudomonas*, *Proteus*, *Salmonella*, *Serratia*, *Shigella*, *Rhizobia*, *Vitreoscilla*, *Paracoccus* y *Klebsiella sp.*, entre muchas otras. Las células pueden ser de cualquiera de varios géneros, including *Saccharomyces* (por ejemplo, *S. cerevisiae*), *Candida* (por ejemplo, *C. utilis*, *C. parapsilosis*, *C. krusei*, *C. versatilis*, *C. lipolytica*, *C. zeylanoides*, *C. guilliermondii*, *C. albicans* y *C. humicola*), *Pichia* (por ejemplo, *P. farinosa* y *P. ohmeri*), *Torulopsis* (por ejemplo, *T. candida*, *T. sphaerica*, *T. xylinus*, *T. famata* y *T. versatilis*), *Debaryomyces* (por ejemplo, *D. subglobosus*, *D. cantarellii*, *D. globosus*, *D. hansenii* y *D. japonicus*), *Zygosaccharomyces* (por ejemplo, *Z. rouxii* y *Z. bailii*), *Kluyveromyces* (por ejemplo, *K. marxianus*), *Hansenula* (por ejemplo, *H. anomala* y *H. jadinii*) y *Brettanomyces* (por ejemplo, *B. lambicus* y *B. anomalus*). Los ejemplos de bacterias útiles incluyen, pero no se limitan a, *Escherichia*, *Enterobacter*, *Azotobacter*, *Erwinia*, *Klebsiella*, *Bacillus*, *Pseudomonas*, *Proteus* y *Salmonella*.

Una vez expresados en una célula huésped, los polipéptidos de CgtE pueden usarse para producir productos galactosilados. Por ejemplo, los polipéptidos de CgtE pueden aislarse usando técnicas de purificación de proteínas convencionales y usarse en reacciones *in vitro* descritas en el presente documento para preparar productos galactosilados. También pueden usarse polipéptidos de CgtE parcialmente purificados en reacciones *in vitro* para preparar productos galactosilados tal como pueden hacer las células huésped permeabilizadas. Las células huésped

pueden usarse también en un sistema *in vivo* (por ejemplo, producción fermentativa) para producir productos galactosilados.

Normalmente, el polinucleótido que codifica para los polipéptidos de CgtE se coloca bajo el control de un promotor que es funcional en la célula huésped deseada. Se conocen bien una variedad extremadamente amplia de promotores, y pueden usarse en los vectores de expresión de la invención, dependiendo de la aplicación particular. Habitualmente, el promotor seleccionado depende de la célula en la que el promotor va a ser activo. También se incluyen opcionalmente otras secuencias de control de la expresión tales como sitios de unión a ribosomas, sitios de terminación de la transcripción y similares. Construcciones que incluyen una o más de estas secuencias de control se denominan "casetes de expresión". Por consiguiente, la invención proporciona casetes de expresión en los que se incorporan los ácidos nucleicos que codifican para proteínas de fusión para expresión a alto nivel en una célula huésped deseada.

Se obtienen a menudo secuencias de control de la expresión que son adecuadas para su uso en una célula huésped particular clonando un gen que se expresa en esa célula. Secuencias de control procariontas comúnmente usadas, que se definen en el presente documento que incluyen promotores para la iniciación de la transcripción, opcionalmente con un operador, junto con secuencias de sitios de unión a ribosomas, incluyen promotores comúnmente usados tales como los sistemas de promotor de beta-lactamasa (penicilinas) y lactosa (*lac*) (Change *et al.*, *Nature* (1977) 198: 1056), el sistema de promotor de triptófano (*trp*) (Goeddel *et al.*, *Nucleic Acids Res.* (1980) 8: 4057), el promotor *tac* (DeBoer, *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* (1983) 80:21-25); y el promotor P<sub>L</sub> derivado de lambda y el sitio de unión a ribosomas del gen N (Shimatake *et al.*, *Nature* (1981) 292: 128). El sistema de promotor particular no es crítico para la invención, puede usarse cualquier promotor disponible que funcione en procariontas.

Para la expresión de proteínas Cgte en células procariontas distintas de *E. coli*, se requiere un promotor que funcione en la especie procarionta particular. Tales promotores pueden obtenerse a partir de genes que se han clonado a partir de las especies, o pueden usarse promotores heterólogos. Por ejemplo, el promotor *trp-lac* híbrido funciona en *Bacillus* además de en *E. coli*.

Se incluye convenientemente un sitio de unión a ribosomas (RBS) en los casetes de expresión de la invención. Un RBS en *E. coli*, por ejemplo, consiste en una secuencia de nucleótidos de 3-9 nucleótidos de longitud ubicada 3-11 nucleótidos en el sentido de 5' del codón de iniciación (Shine y Dalgarno, *Nature* (1975) 254: 34; Steitz, en *Biological regulation and development: Gene expression* (ed. R.F. Goldberger), vol. 1, p. 349, 1979, Plenum Publishing, NY).

Para la expresión de las proteínas Cgte en levaduras, los promotores convenientes incluyen GAL 1-10 (Johnson y Davies (1984) *Mol. Cell. Biol.* 4:1440-1448) ADH2 (Russell *et al.* (1983) *J. Biol. Chem.* 258:2674-2682), PHO5 (EMBO J. (1982) 6:675-680) y MF $\alpha$  (Herskowitz y Oshima (1982) en *The Molecular Biology of the Yeast Saccharomyces* (eds. Strathern, Jones, y Broach) Cold Spring Harbor Lab., Cold Spring Harbor, N.Y., págs. 181-209). Otro promotor adecuado para su uso en levaduras es el promotor híbrido ADH2/GAPDH tal como se describe en Cousens *et al.*, *Gene* 61:265-275 (1987). Para hongos filamentosos tales como, por ejemplo, cepas del hongo *Aspergillus* (McKnight *et al.*, patente estadounidense n.º 4.935.349), los ejemplos de promotores útiles incluyen los derivados de genes glicolíticos de *Aspergillus nidulans*, tales como el promotor ADH3 (McKnight *et al.*, *EMBO J.* 4: 2093-2099 (1985)) y el promotor *tpiA*. Un ejemplo de un terminador adecuado es el terminador ADH3 (McKnight *et al.*).

Pueden usarse en la presente invención promotores o bien constitutivos o bien regulados. Los promotores regulados pueden ser ventajosos porque las células huésped pueden hacerse crecer a altas densidades antes de que se induzca la expresión de las proteínas de fusión. Un alto nivel de expresión de proteínas heterólogas permite ralentizar el crecimiento celular en algunas situaciones. Un promotor inducible es un promotor que dirige la expresión de un gen en el que el nivel de expresión puede alterarse por factores del desarrollo o medioambientales tales como, por ejemplo, temperatura, pH, condiciones aerobias o anaerobias, luz, factores de transcripción y productos químicos. Tales promotores se denominan en el presente documento promotores "inducibles", que permiten controlar el ritmo de expresión de la glicosiltransferasa o enzima implicada en la síntesis de azúcares de nucleótidos. Para *E. coli* y otras células huésped bacterianas, los expertos en la técnica conocen promotores inducibles. Estos incluyen, por ejemplo, el promotor *lac*, el promotor P<sub>L</sub> del bacteriófago lambda, el promotor híbrido *trp-lac* (Amann *et al.* (1983) *Gene* 25: 167; de Boer *et al.* (1983) *Proc. Nat'l. Acad. Sci. USA* 80: 21), y el promotor del bacteriófago T7 (Studier *et al.* (1986) *J. Mol. Biol.*; Tabor *et al.* (1985) *Proc. Nat'l. Acad. Sci. USA* 82: 1074-8). Estos promotores y su uso se tratan en Sambrook *et al.*, citado anteriormente. Un promotor inducible particularmente preferido para su expresión en procariontas es un promotor doble que incluye un componente de promotor *tac* unido a un componente de promotor obtenido de un gen o genes que codifican para enzimas implicadas en el metabolismo de la galactosa (por ejemplo, un promotor de un gen de UDPgalactosa 4-epimerasa (*galE*)). El promotor *tac-gal* doble, que se describe en la publicación de solicitud de patente PCT n.º WO98/20111.

Una construcción que incluye un polinucleótido de interés operativamente unido a señales de control de la expresión génica que, cuando se coloca en una célula huésped apropiada, dirige la expresión del polinucleótido se denomina "casete de expresión". Casetes de expresión que codifican para las proteínas de fusión de la invención se colocan a menudo en vectores de expresión para su introducción en la célula huésped. Los vectores incluyen normalmente, además de un casete de expresión, una secuencia de ácido nucleico que permite que el vector se replique

- independientemente en una o más células huésped seleccionadas. Generalmente, esta secuencia es una que permite que el vector se replique independientemente del ADN cromosómico del huésped, e incluye orígenes de replicación o secuencias de replicación autónoma. Tales secuencias se conocen bien para una variedad de bacterias. Por ejemplo, el origen de replicación del plásmido pBR322 es adecuado para la mayoría de las bacterias Gram-negativas. Alternativamente, el vector puede replicarse integrándose en el complemento genómico de la célula huésped y replicándose cuando la célula experimenta replicación del ADN. Un vector de expresión preferido para la expresión de las enzimas es en células bacterianas pTGK, que incluye un promotor *tac-gal* doble y se describe en la publicación de solicitud de patente PCT n.º WO98/20111.
- La construcción de construcciones de polinucleótido requiere generalmente el uso de vectores que pueden replicarse en bacterias. Está disponible comercialmente una plétora de kits para la purificación de plásmidos a partir de bacterias (véanse, por ejemplo, EasyPrepJ, DexiPrepJ, ambos de Pharmacia Biotech; StrataCleanJ, de Stratagene; y, QIAexpress Expression System, Qiagen). Los plásmidos aislados y purificados pueden manipularse entonces adicionalmente para producir otros plásmidos, y usarse para transfectar células. También es posible la clonación en *Streptomyces* o *Bacillus*.
- Se incorporan a menudo marcadores seleccionables en los vectores de expresión usados para expresar los polinucleótidos de la invención. Estos genes pueden codificar para un producto génico, tal como una proteína, necesaria para la supervivencia o el crecimiento de células huésped transformadas hechas crecer en un medio de cultivo selectivo. Las células huésped no transformadas con el vector que contiene el gen de selección no sobrevivirán en el medio de cultivo. Genes de selección típicos codifican para proteínas que confieren resistencia a antibióticos u otras toxinas, tales como ampicilina, neomicina, kanamicina, cloranfenicol o tetraciclina. Alternativamente, los marcadores seleccionables pueden codificar para proteínas que complementan deficiencias auxótrofas o suministran nutrientes críticos no disponibles a partir de medios complejos, por ejemplo el gen que codifica para la D-alanina racemasa para bacilos. A menudo, el vector tendrá un marcador seleccionable que es funcional en, por ejemplo, *E. coli*, u otras células en las que el vector se replica antes de introducirse en la célula huésped. Los expertos en la técnica conocen varios marcadores seleccionables y se describen por ejemplo en Sambrook *et al.*, citado anteriormente.
- La construcción de vectores adecuados que contienen uno o más de los componentes anteriormente enumerados emplea técnicas de ligamiento convencionales descritas en las referencias citadas anteriormente. Los fragmentos de ADN o plásmidos aislados se escinden, adaptan y vuelven a ligarse en la forma deseada para generar los plásmidos requeridos. Para confirmar las secuencias correctas en los plásmidos construidos, los plásmidos pueden analizarse mediante técnicas convencionales tales como mediante digestión con endonucleasas de restricción, y/o secuenciación según métodos conocidos. Se conocen en la técnica, técnicas de clonación moleculares para lograr estos fines. Los expertos conocen bien una amplia variedad de métodos de amplificación *in vitro* y clonación adecuados para la construcción de ácidos nucleicos recombinantes. Se encuentran ejemplos de estas técnicas e instrucciones suficientes para dirigir a los expertos a través de muchos ejercicios de clonación en Berger y Kimmel, *Guide to Molecular Cloning Techniques, Methods in Enzymology*, volumen 152, Academic Press, Inc., San Diego, CA (Berger); y *Current Protocols in Molecular Biology*, F.M. Ausubel *et al.*, eds., Current Protocols, una empresa conjunta entre Greene Publishing Associates, Inc. y John Wiley & Sons, Inc., (suplemento de 1998) (Ausubel).
- Se conocen bien en la técnica una amplia variedad de vectores comunes adecuados para su uso como materiales de partida para construir los vectores de expresión de la invención. Para clonar en bacterias, los vectores comunes incluyen vectores derivados de pBR322 tales como pBLUESCRIPT™ y vectores derivados del fago λ. En levaduras, los vectores incluyen plásmidos de integración de levaduras (por ejemplo, Ylp5) y plásmidos de replicación de levaduras (los plásmidos de la serie YRp) y pGPD-2. Puede lograrse la expresión en células de mamífero usando una variedad de plásmidos comúnmente disponibles, incluyendo pSV2, pBC12BI y p91023, así como vectores de virus líticos (por ejemplo, virus vaccinia, adenovirus y baculovirus), vectores de virus episomales (por ejemplo, papilomavirus bovino) y vectores retrovirales (por ejemplo, retrovirus murinos).
- Los métodos para introducir los vectores de expresión en una célula huésped elegida no son particularmente críticos, y tales métodos los conocen los expertos en la técnica. Por ejemplo, los vectores de expresión pueden introducirse en células procariontas, incluyendo *E. coli*, mediante transformación con cloruro de calcio, y en células eucariotas mediante tratamiento con fosfato de calcio o electroporación. También son adecuados otros métodos de transformación.
- Puede usarse acoplamiento traduccional para potenciar la expresión. La estrategia usa un marco de lectura abierto en el sentido de 5' corto derivado de un gen altamente expresado nativo para el sistema de traducción, que se coloca en el sentido de 3' del promotor, y un sitio de unión a ribosomas seguido tras unos cuantos codones de aminoácidos por un codón de terminación. Justo antes del codón de terminación está un sitio de unión de ribosomas, y tras el codón de terminación está un codón de iniciación para la iniciación de la traducción. El sistema deshace la estructura secundaria en el ARN, permitiendo la iniciación eficaz de la traducción. Véase Squires, *et al.* (1988), *J. Biol. Chem.* 263: 16297-16302.
- Los polipéptidos de CgtE pueden expresarse intracelularmente, o pueden secretarse de la célula. La expresión intracelular a menudo da como resultado altos rendimientos. Si es necesario, la cantidad de proteína de fusión

soluble, activa puede aumentarse realizando procedimientos de replegamiento (véase, por ejemplo, Sambrook *et al.*, citado anteriormente.; Marston *et al.*, Bio/Technology (1984) 2: 800; Schoner *et al.*, Bio/Technology (1985) 3: 151). En realizaciones en las que los polipéptidos de CgtE se secretan de la célula, o bien en el periplasma o bien en el medio extracelular, la secuencia de ADN se une a una secuencia de péptido señal escindible. La secuencia señal dirige la translocación de la proteína de fusión a través de la membrana celular. Un ejemplo de un vector adecuado para su uso en *E. coli* que contiene una unidad de promotor-secuencia señal es pTA1529, que tiene la secuencia señal y el promotor *phoA* de *E. coli* (véanse, por ejemplo, Sambrook *et al.*, citado anteriormente.; Oka *et al.*, Proc. Natl. Acad. Sci. USA (1985) 82: 7212; Talmadge *et al.*, Proc. Natl. Acad. Sci. USA (1980) 77: 3988; Takahara *et al.*, J. Biol. Chem. (1985) 260: 2670). En otra realización, las proteínas Cgte se fusionan con una subsecuencia de proteína A o albúmina sérica bovina (BSA), por ejemplo, para facilitar la purificación, secreción o estabilidad.

Los polipéptidos de CgtE de la invención también pueden unirse adicionalmente a otras proteínas bacterianas. Este enfoque da como resultado a menudo altos rendimientos, porque secuencias de control procariontas normales dirigen la transcripción y traducción. En *E. coli*, se usan a menudo fusiones de *lacZ* para expresar proteínas heterólogas. Están fácilmente disponibles vectores adecuados, tales como la serie pUR, pEX y pMR100 (véase, por ejemplo, Sambrook *et al.*, citado anteriormente). Para ciertas aplicaciones, puede ser deseable escindir los aminoácidos de la enzima distinta de glicosiltransferasa y/o auxiliar de la proteína de fusión tras la purificación. Esto puede lograrse mediante cualquiera de varios métodos conocidos en la técnica, incluyendo escisión mediante bromuro de cianógeno, una proteasa o mediante el factor X<sub>a</sub> (véanse, por ejemplo, Sambrook *et al.*, citado anteriormente.; Itakura *et al.*, Science (1977) 198: 1056; Goeddel *et al.*, Proc. Natl. Acad. Sci. USA (1979) 76: 106; Nagai *et al.*, Nature (1984) 309: 810; Sung *et al.*, Proc. Natl. Acad. Sci. USA (1986) 83: 561). Pueden diseñarse por ingeniería genética sitios de escisión en el gen para la proteína de fusión en el punto de escisión deseado.

Puede expresarse más de una proteína recombinante en una única célula huésped colocando múltiples casetes transcripcionales en un único vector de expresión, o utilizando diferentes marcadores seleccionables para cada uno de los vectores de expresión que se emplean en la estrategia de clonación.

Un sistema adecuado para obtener proteínas recombinantes a partir de *E. coli* que mantiene la integridad de sus extremos N-terminales se ha descrito por Miller *et al.* Biotechnology 7:698-704 (1989). En este sistema, el gen de interés se produce como una fusión C-terminal con los primeros 76 residuos del gen de ubiquitina de levaduras que contiene un sitio de escisión de peptidasa. La escisión en la unión de los dos restos da como resultado la producción de una proteína que tiene un residuo N-terminal auténtico intacto.

## 30 VI. Purificación de polipéptidos de CgtE

Las proteínas Cgte de la presente invención pueden expresarse como proteínas intracelulares o como proteínas que se secretan de la célula, y pueden usarse en esta forma, en los métodos de la presente invención. Por ejemplo, puede usarse un extracto celular bruto que contiene el polipéptido de CgtE intracelular o expresado en los métodos de la presente invención.

Alternativamente, el polipéptido de CgtE puede purificarse según procedimientos convencionales en la técnica, incluyendo precipitación con sulfato de amonio, columnas de afinidad, cromatografía en columna, electroforesis en gel y similares (véanse, en general, R. Scopes, Protein Purification, Springer-Verlag, N.Y. (1982), Deutscher, Methods in Enzymology vol. 182: Guide to Protein Purification., Academic Press, Inc. N.Y. (1990)). Se prefieren composiciones sustancialmente puras de una homogeneidad de al menos aproximadamente el 70, el 75, el 80, el 85, el 90%, y lo más preferido es una homogeneidad del 92, el 95, el 98 al 99% o más. Las proteínas purificadas pueden usarse también, por ejemplo, como inmunógenos para la producción de anticuerpos.

Para facilitar la purificación de los polipéptidos de CgtE de la invención, los ácidos nucleicos que codifican para las proteínas pueden incluir también una secuencia codificante para un epítipo o "etiqueta" para la que está disponible un reactivo de unión por afinidad, es decir, una etiqueta de purificación. Los ejemplos de epítipos adecuados incluyen los genes indicadores *myc* y *V-5*; vectores de expresión útiles para la producción recombinante de proteínas de fusión que tienen estos epítipos están disponibles comercialmente (por ejemplo, los vectores de Invitrogen (Carlsbad CA) pcDNA3.1/Myc-His y pcDNA3.1/V5-His son adecuados para la expresión en células de mamífero). Los expertos en la técnica conocen vectores de expresión adicionales para unir una etiqueta al polipéptido de CgtE de la invención, y sistema de detección sistemas correspondientes, y varios están disponibles comercialmente (por ejemplo, FLAG" (Kodak, Rochester NY). Otro ejemplo de una etiqueta adecuada es una secuencia de polihistidina, que puede unirse a ligandos de afinidad de quelatos de metal. Normalmente, se usan seis histidinas adyacentes, aunque pueden usarse más o menos de seis. Los ligandos de afinidad de quelatos de metal adecuados que pueden servir como resto de unión para una etiqueta de polihistidina incluyen ácido nitrilotri-acético (NTA) (Hochuli, E. (1990) "Purification of recombinant proteins with metal chelating adsorbents" en Genetic Engineering: Principles and Methods, J.K. Setlow, Ed., Plenum Press, NY; disponible comercialmente de Qiagen (Santa Clarita, CA)). Otras etiquetas de purificación o de epítipo incluyen, por ejemplo, AU1, AU5, DDDDK (EC5), etiqueta E, etiqueta E2, Glu-Glu, un péptido de 6 residuos, EYMPME, derivado de la proteína T media del polioma, HA, HSV, IRS, KT3, Stage, etiqueta S1, etiqueta T7, etiqueta V5, VSV-G, β-galactosidasa, Gal4, proteína fluorescente verde (GFP), luciferasa, proteína C, proteína A, proteína de unión a celulosa, GST (glutación S-transferasa), una etiqueta step, Nus-S, PPI-asas, Pfg 27, proteína de unión a calmodulina, dsb A y fragmentos de la

misma, y granzima B. Anticuerpos y péptidos de epítipo que se unen específicamente a secuencias de epítipo están disponibles comercialmente de, por ejemplo, Covance Research Products, Inc.; Bethyl Laboratories, Inc.; Abcam Ltd.; y Novus Biologicals, Inc.

5 Las etiquetas de purificación también incluyen dominios de unión a maltosa y dominios de unión a almidón. Las proteínas que comprenden etiquetas de purificación pueden purificarse usando una pareja de unión que se une a la etiqueta de purificación, por ejemplo, anticuerpos frente a la etiqueta de purificación, iones níquel o cobalto o resinas, y amilosa, maltosa o una ciclodextrina. Las etiquetas de purificación también incluyen dominios de unión a almidón, dominios de tiorredoxina de *E. coli* (vectores y anticuerpos disponibles comercialmente de por ejemplo, Santa Cruz Biotechnology, Inc. y Alpha Diagnostic International, Inc.), y la mitad carboxilo terminal de la proteína SUMO (vectores y anticuerpos disponibles comercialmente de por ejemplo, Life Sensors Inc.). Se describen dominios de unión a almidón, tales como un dominio de unión a maltosa de *E. coli* y SBD (dominio de unión a almidón) de una amilasa de *A. niger*, en el documento WO 99/15636. La purificación por afinidad de una proteína de fusión que comprende un dominio de unión a almidón usando una resina derivatizada con betaciclodextrina (BCD) se describe en el documento WO 2005/014779, publicado el 17 de febrero de 2005. En algunas realizaciones, un polipéptido de CgtE comprende más de una etiqueta de epítipo o de purificación.

15 Los expertos en la técnica conocen otros haptenos que son adecuados para su uso como etiquetas y se describen, por ejemplo, en el Handbook of Fluorescent Probes and Research Chemicals (6ª Ed., Molecular Probes, Inc., Eugene OR). Por ejemplo, dinitrofenol (DNP), digoxigenina, barbituratos (véase, por ejemplo, la patente estadounidense n.º 5.414.085) y varios tipos de fluoróforos son útiles como haptenos, ya que son derivados de estos compuestos. Están disponibles comercialmente kits para unir haptenos y otros restos a proteínas y otras moléculas. Por ejemplo, cuando el hapteno incluye un tiol, puede usarse un ligador heterobifuncional tal como SMCC para unir la etiqueta a residuos de lisina presentes en el reactivo de captura.

20 Un experto reconocería que pueden hacerse modificaciones en los dominios catalíticos o funcionales del polipéptido de CgtE sin disminuir su actividad biológica. Pueden hacerse algunas modificaciones para facilitar la clonación, expresión o incorporación del dominio catalítico en una proteína de fusión. Tales modificaciones las conocen bien los expertos en la técnica e incluyen, por ejemplo, la adición de codones en cualquier extremo terminal del polinucleótido que codifica para el dominio catalítico para proporcionar, por ejemplo, una metionina añadida al extremo amino terminal para proporcionar un sitio de iniciación o aminoácidos adicionales (por ejemplo, poli His) colocados en cualquier extremo para crear sitios de enzimas de restricción o codones de terminación o secuencias de purificación convenientemente ubicados.

### 25 VII. Proteínas Cgte de fusión

En algunas realizaciones, las células recombinantes de la invención expresan proteínas de fusión que tienen más de una actividad enzimática que está implicada en la síntesis de un oligosacárido galactosilado deseado. Los polipéptidos de fusión pueden estar compuestos, por ejemplo, por un polipéptido de CgtE que está unido a una 35 enzima auxiliar, por ejemplo, [UDP-GalNAc 4' epimerasa o una UDP-glucosa 4' epimerasa. También pueden prepararse proteínas de fusión usando dominios catalíticos u otros truncamientos de las enzimas. Por ejemplo, un polinucleótido que codifica para un polipéptido de CgtE puede unirse, en marco, a un polinucleótido que codifica, por ejemplo, para una UDP-GalNAc 4' epimerasa o una UDP-glucosa 4' epimerasa. La proteína de fusión resultante puede catalizar entonces no sólo la síntesis de la molécula de GalNAc o galactosa activada, sino también la transferencia del resto de galactosa a la molécula aceptora. La proteína de fusión puede ser dos o más enzimas del ciclo de la galactosa unidas en una secuencia de nucleótidos expresable. Los polipéptidos de CgtE de fusión de la presente invención pueden diseñarse y fabricarse fácilmente utilizando diversas técnicas de ADN recombinante bien conocidas por los expertos en la técnica. Se describen proteínas de fusión a modo de ejemplo en la solicitud de patente PCT PCT/CA98/01180, que se publicó como el documento WO99/31224 el 24 de junio de 1999 y que da a 40 conocer CMP-ácido siálico sintasa de *Neisseria* fusionada con una  $\alpha$ 2,3-sialiltransferasa de *Neisseria*. En algunas realizaciones, se expresa más de un polipéptido de CgtE de fusión en la célula. La proteína de fusión también puede comprender etiquetas de epítipo o de purificación tal como se describe en el presente documento.

### 45 VIII. Sustratos donadores y sustratos aceptores

Los sustratos donadores adecuados usados por los polipéptidos de CgtE y otras glicosiltransferasas en los métodos de la invención incluyen, pero no se limitan a, UDP-Glc, UDP-GlcNAc, UDP-Gal, UDP-GalNAc, GDP-Man, GDP-Fuc, UDPGlcUA y CMP-ácido siálico y otros restos de ácido siálico activados. Guo *et al.*, Applied Biochem. and Biotech. 68: 1-20 (1997)

Normalmente, los sustratos aceptores incluyen un residuo de GalNAc o galactosa terminal para la adición de un residuo de GalNAc o galactosa mediante un enlace  $\beta$ 1,3. Los ejemplos de aceptores adecuados incluyen una Gal terminal que está unida a GlcNAc o Glc mediante un enlace  $\beta$ 1,4, y una Gal terminal que está unida en  $\beta$ 1,3 a o bien GlcNAc o bien GalNAc. Los aceptores adecuados incluyen, por ejemplo, aceptores de galactosilo tales como Gal $\beta$ 1,4GlcNAc, Gal $\beta$ 1,4GalNAc, Gal $\beta$ 1,3GalNAc, lacto-N-tetraosa, Gal $\beta$ 1,3G1cNAc, Gal $\beta$ 1,3Ara, Gal $\beta$ 1,6GlcNAc, Gal $\beta$ 1,4Glc (lactosa) y otros aceptores conocidos por los expertos en la técnica. El residuo terminal al que se une el resto de galactosa puede estar unido por sí mismo a, por ejemplo, H, un sacárido, oligosacárido o un grupo aglicona

que tiene al menos un átomo de hidrato de carbono. En algunas realizaciones, el residuo aceptor es una parte de un oligosacárido que está unido a un péptido, una proteína, un lípido o un proteoglicano, por ejemplo.

5 Los sustratos aceptores adecuados usados por los polipéptidos de CgtE y los métodos de la invención incluyen, pero no se limitan a, polisacáridos y oligosacáridos. Los polipéptidos de CgtE descritos en el presente documento pueden usarse también en sistemas multienzimáticos para producir un producto deseado a partir de un material de partida conveniente.

10 Los sustratos aceptores adecuados usados por los polipéptidos de CgtE y los métodos de la invención incluyen, pero no se limitan a, proteínas, lípidos, gangliósidos y otras estructuras biológicas (por ejemplo, células completas) que pueden modificarse mediante los métodos de la invención. Estos sustratos aceptores comprenderán normalmente las moléculas de polisacárido u oligosacárido descritas anteriormente. Las estructuras a modo de ejemplo, que pueden modificarse mediante los métodos de la invención incluyen cualquiera de varios glicolípidos, glicoproteínas y estructuras de hidrato de carbono en células conocidas por los expertos en la técnica tal como se expone en la tabla 1.

**Tabla 1**

|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <p><u>Hormonas y factores de crecimiento</u></p> <ul style="list-style-type: none"> <li>• G-CSF</li> <li>• GM-CSF</li> <li>• TPO</li> <li>• EPO</li> <li>• Variantes de EPO</li> <li>• TNF-<math>\alpha</math></li> <li>• Leptina</li> </ul> <p><u>Enzimas e inhibidores</u></p> <ul style="list-style-type: none"> <li>• t-PA</li> <li>• Variantes de t-PA</li> <li>• Urocinasa</li> <li>• Factores VII, VIII, IX, X</li> <li>• ADNasa</li> <li>• Glucocerebrosidasa</li> <li>• Hirudina</li> <li>• <math>\alpha</math>1 antitripsina</li> <li>• Antitrombina III</li> </ul> <p><u>Citocinas y citocinas quiméricas</u></p> <ul style="list-style-type: none"> <li>• Interleucina-1 (IL-1), 1B, 2,3,4</li> <li>• Interferón-<math>\alpha</math> (IFN-<math>\alpha</math>)</li> <li>• IFN-<math>\alpha</math>-2b</li> <li>• IFN-<math>\beta</math></li> <li>• IFN-<math>\gamma</math></li> <li>• Toxina diftérica quimérica-IL-2</li> </ul> | <p><u>Receptores y receptores quiméricos</u></p> <ul style="list-style-type: none"> <li>• CD4</li> <li>• Receptor de factor de necrosis tumoral (TNF)</li> <li>• Alfa-CD20</li> <li>• AcM-CD20</li> <li>• AcM-alfa-CD3</li> <li>• AcM-receptor de TNF</li> <li>• AcM-CD4</li> <li>• PSGL-1</li> <li>• AcM-PSGL-1</li> <li>• Complemento</li> <li>• GlyCAM o su quimera</li> <li>• N-CAM o su quimera</li> <li>• LFA-3</li> <li>• CTLA-IV</li> </ul> <p><u>Anticuerpos monoclonales (inmunoglobulinas)</u></p> <ul style="list-style-type: none"> <li>• AcM-anti-RSV</li> <li>• AcM-anti-receptor de IL-2</li> <li>• AcM-anti-CEA</li> <li>• AcM-anti-receptor de plaquetas IIb/IIIa</li> <li>• AcM-anti-EGF</li> <li>• AcM-anti-receptor de Her-2 receptor</li> </ul> <p><u>Células</u></p> <ul style="list-style-type: none"> <li>• Glóbulos rojos</li> <li>• Glóbulos blandos (por ejemplo, células T, células B, células dendríticas, macrófagos, células NK, neutrófilos, monocitos y similares)</li> <li>• Células madre</li> </ul> |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

15 La presente invención proporciona polipéptidos de CgtE que se seleccionan por su capacidad para producir oligosacáridos, glicoproteínas y glicolípidos que tienen restos de oligosacáridos deseados. De manera similar, si están presentes, se eligen enzimas auxiliares basándose en un sustrato de azúcar activado o en un azúcar encontrado en el producto oligosacárido.

20 Para la síntesis de glicoproteínas, pueden identificarse fácilmente polipéptidos de CgtE adecuados haciendo reaccionar diversas cantidades de un polipéptido de CgtE de interés (por ejemplo, 0,01-100 mU/mg de proteína) con una glicoproteína (por ejemplo, a 1-10 mg/ml) a la que se une un oligosacárido que tiene un sitio aceptor potencial

para la glicosilación mediante la proteína CgtE de interés. Se comparan las capacidades de las proteínas CgtE recombinantes de la presente invención para añadir un residuo de azúcar en el sitio aceptor deseado, y se selecciona un polipéptido de CgtE que tiene la propiedad deseada (por ejemplo, actividad catalítica o especificidad de sustrato aceptor).

5 En general, la eficacia de la síntesis enzimática de oligosacáridos, glicoproteínas y glicolípidos, que tienen restos de oligosacáridos galactosilados deseados, puede potenciarse a través del uso de polipéptidos de CgtE producidos de manera recombinante de la presente invención. Se requieren técnicas recombinantes que permiten la producción de los polipéptidos de CgtE recombinantes en grandes cantidades para la modificación de glicolípidos, glicoproteínas y oligosacáridos *in vitro* a gran escala.

10 En algunas realizaciones, oligosacáridos, glicoproteínas y glicolípidos adecuados para su uso por los polipéptidos de CgtE y los métodos de la invención pueden ser glicoproteínas y glicolípidos inmovilizados sobre un soporte sólido durante la reacción de glicosilación. La expresión "soporte sólido" también abarca soportes semisólidos. Preferiblemente, el glicolípido o glicoproteína diana se inmoviliza de manera reversible de modo que el glicolípido o glicoproteína respectivo puede liberarse tras completarse la reacción de glicosilación. Los expertos en la técnica conocen muchas matrices adecuadas. Puede emplearse intercambio iónico, por ejemplo, para inmovilizar temporalmente una glicoproteína o glicolípido sobre una resina apropiada mientras se realiza la reacción de glicosilación. También puede usarse un ligando que se une específicamente a la glicoproteína o glicolípido para inmovilización basada en afinidad. Por ejemplo, pueden usarse también anticuerpos que se unen específicamente a una glicoproteína. Además, cuando la glicoproteína de interés es por sí misma un anticuerpo o contiene un fragmento del mismo, puede usarse proteína A o G como resina de afinidad. También son adecuados tintes y otras moléculas que se unen específicamente a una glicoproteína o glicolípido de interés.

Preferiblemente, cuando el azúcar aceptor es una versión truncada de la glicoproteína de longitud completa, incluye preferiblemente la subsecuencia biológicamente activa de la glicoproteína de longitud completa. Las subsecuencias biológicamente activas a modo de ejemplo incluyen, pero no se limitan a, sitios activos de enzimas, sitios de unión a receptor, sitios de unión a ligando, regiones determinantes de complementariedad de antibióticos y regiones antigénicas de antígenos.

#### IX. Producción de productos galactosilados

Pueden usarse polipéptidos de CgtE para preparar productos galactosilados en mezclas de reacciones *in vitro* o mediante reacciones *in vivo*, por ejemplo, mediante crecimiento fermentativo de microorganismos recombinantes que comprenden nucleótidos que codifican para polipéptidos de CgtE.

##### A. Reacciones *in vitro*

Los polipéptidos de CgtE pueden usarse para preparar productos galactosilados en mezclas de reacciones *in vitro*. Las mezclas de reacción *in vitro* pueden incluir microorganismos permeabilizados que comprenden los polipéptidos de CgtE, polipéptidos de CgtE parcialmente purificados o polipéptidos de CgtE purificados; así como sustratos donadores y sustratos aceptores, y tampones de reacción apropiados. Para reacciones *in vitro*, las proteínas glicosiltransferasa recombinantes, tales como polipéptidos de CgtE, sustratos aceptores, sustratos donadores y otros componentes de la mezcla de reacción se combinan mediante mezcla en un medio de reacción acuoso. Pueden usarse glicosiltransferasas adicionales en combinación con los polipéptidos de CgtE, dependiendo del producto galactosilado deseado. El medio tiene generalmente un valor de pH de aproximadamente 5 a aproximadamente 8,5. La selección de un medio se basa en la capacidad del medio para mantener el valor de pH al nivel deseado. Por tanto, en algunas realizaciones, el medio se tampona a un valor de pH de aproximadamente 7,5. Si no se usa un tampón, el pH del medio debe mantenerse a aproximadamente de 5 a 8,5, dependiendo de la glicosiltransferasa particular usada. Para polipéptidos de CgtE, el intervalo de pH se mantiene preferiblemente desde aproximadamente 6,0 hasta 8,0. Para sialiltransferasas, el intervalo es preferiblemente de desde aproximadamente 5,5 hasta aproximadamente 8,0.

Las concentraciones o cantidades de enzima se expresan en unidades de actividad, que es una medida de la velocidad de catálisis inicial. Una unidad de actividad cataliza la formación de 1  $\mu\text{mol}$  de producto por minuto a una temperatura (normalmente 37°C) y un valor de pH (normalmente 7,5) dados. Por tanto, 10 unidades de una enzima es una cantidad catalítica de la enzima en la que 10  $\mu\text{mol}$  de sustrato se convierten en 10  $\mu\text{mol}$  de producto en un minuto a una temperatura de 37°C y un valor de pH de 7,5.

La mezcla de reacción puede incluir cationes de metal divalentes ( $\text{Mg}^{2+}$ ,  $\text{Mn}^{2+}$ ). El medio de reacción puede comprender también detergentes de solubilización (por ejemplo, Triton o SDS) y disolventes orgánicos tales como metanol o etanol, si es necesario. Las enzimas pueden utilizarse libres en disolución o pueden estar unidas a un soporte tal como un polímero. La mezcla de reacción es por tanto sustancialmente homogénea al comienzo, aunque puede formarse algo de precipitado durante la reacción.

La temperatura a la que se lleva a cabo un procedimiento anterior puede oscilar entre justo por encima de la congelación y la temperatura a la que se desnaturaliza la enzima más sensible. Ese intervalo de temperatura es preferiblemente de desde aproximadamente 0°C hasta aproximadamente 45°C, y más preferiblemente a de

aproximadamente 20°C a aproximadamente 37°C.

La mezcla de reacción así formada se mantiene durante un periodo de tiempo suficiente para obtener el alto rendimiento deseado de determinantes de oligosacáridos deseados presentes en grupos oligosacárido unidos a la glicoproteína que va a glicosilarse. Para preparaciones a gran escala, se dejará a menudo que la reacción se realice durante entre aproximadamente 0,5-240 horas, y más normalmente entre aproximadamente 1-36 horas.

Puede llevarse a cabo una o más de las reacciones de glicosiltransferasa como parte de un ciclo de glicosiltransferasa. Se han descrito descripciones y condiciones preferidas de ciclos de glicosiltransferasa. Se describen varios ciclos de glicosiltransferasa (por ejemplo, ciclos de sialiltransferasa, ciclos de galactosiltransferasa y ciclos de fucosiltransferasa) en la patente estadounidense n.º 5.374.541 y el documento WO 9425615 A. Se describen otros ciclos de glicosiltransferasa en Ichikawa *et al.* J. Am. Chem. Soc. 114:9283 (1992), Wong *et al.* J. Org. Chem. 57: 4343 (1992), DeLuca, *et al.*, J. Am. Chem. Soc. 117:5869-5870 (1995), e Ichikawa *et al.* en Carbohydrates and Carbohydrate Polymers. Yaltami, ed. (ATL Press, 1993).

Pueden sustituirse otras glicosiltransferasas en ciclos de transferasas similares tal como se ha descrito en detalle para las fucosiltransferasas y sialiltransferasas. En particular, la glicosiltransferasa puede ser también, por ejemplo, glucosiltransferasas, por ejemplo, Alg8 (Stagljov *et al.*, Proc. Natl. Acad. Sci. USA 91:5977 (1994)) o Alg5 (Heesen *et al.* Eur. J. Biochem. 224:71 (1994)), N-acetilgalactosaminiltransferasas tales como, por ejemplo,  $\alpha(1,3)$ N-acetilgalactosaminiltransferasa,  $\beta(1,4)$ N-acetilgalactosaminiltransferasas (Nagata *et al.* J. Biol. Chem. 267:12082-12089 (1992) y Smith *et al.* J. Biol. Chem. 269:15162 (1994)) y polipéptido N-acetilgalactosaminiltransferasa (Homa *et al.* J. Biol. Chem. 268:12609 (1993)). Las N-acetilglucosaminiltransferasas adecuadas incluyen GnTI (2.4.1.101, Hull *et al.*, BBRC 176:608 (1991)), GnTII, y GnTIII (Ihara *et al.* J. Biochem. 113:692 (1993)), GnTV (Shoreiban *et al.* J. Biol. Chem. 268: 15381 (1993)), N-acetilglucosaminiltransferasa O-unida (Bierhuizen *et al.* Proc. Natl. Acad. Sci. USA 89:9326 (1992)), N-acetilglucosamina-1-fosfato transferasa (Rajput *et al.* Biochem J. 285:985 (1992)) y hialuronano sintasa. Las manosiltransferasas adecuadas incluyen  $\alpha(1,2)$ manosiltransferasa,  $\alpha(1,3)$ manosiltransferasa,  $\beta(1,4)$  manosiltransferasa, Dol-P-Man sintasa, OCh1 y Pmt1.

Para los ciclos de glicosiltransferasa anteriores, las concentraciones o cantidades de los diversos reactivos usados en los procedimientos dependen de numerosos factores incluyendo condiciones de reacción tales como temperatura y valor de pH, y la elección y cantidad de sacáridos aceptores que van a glicosilarse. Debido a que el proceso de glicosilación permite la regeneración de nucleótidos activantes, azúcares donadores activados y la eliminación del PPI producido en presencia de cantidades catalíticas de las enzimas, el proceso está limitado por las concentraciones o cantidades de los sustratos estequiométricos tratados anteriormente. El límite superior para las concentraciones de reactantes que pueden usarse según el método de la presente invención se determina mediante la solubilidad de tales reactantes.

Preferiblemente, las concentraciones de nucleótidos activantes, donador de fosfato, el azúcar donador y las enzimas se seleccionan de manera que la glicosilación avanza hasta que el aceptor se consume. Las consideraciones tratadas a continuación, aunque en el contexto de una sialiltransferasa, pueden aplicarse generalmente a otros ciclos de glicosiltransferasa.

Cada una de las enzimas está presente en una cantidad catalítica. La cantidad catalítica de una enzima particular varía según la concentración del sustrato de esa enzima así como según condiciones de reacción tales como temperatura, tiempo y valor de pH. Los expertos en la técnica conocen bien medios para determinar la cantidad catalítica para una enzima dada en condiciones de reacción y concentraciones de sustrato preseleccionadas.

#### B. Reacciones *in vivo*

Los polipéptidos de CgtE puede usarse para preparar productos galactosilados mediante reacciones *in vivo*, por ejemplo, crecimiento fermentativo de microorganismos recombinantes que comprenden los polipéptidos de CgtE. El crecimiento fermentativo de microorganismos recombinantes puede producirse en presencia de medio que incluye un sustrato aceptor y un sustrato donador o un precursor para un sustrato donador, por ejemplo, galactosa o GalNAc. Véase, por ejemplo, Priem *et al.*, Glycobiology 12:235-240 (2002). El microorganismo capta el sustrato aceptor y el sustrato donador o el precursor para un sustrato donador y la adición del sustrato donador al sustrato aceptor tiene lugar en la célula viva. El microorganismo puede alterarse para facilitar la captación del sustrato aceptor, por ejemplo, expresando una proteína transportadora de azúcar. Por ejemplo, cuando el sacárido aceptor es lactosa, pueden usarse células de *E. coli* que expresan la LacY permeasa. Pueden usarse otros métodos para reducir la descomposición de un sacárido aceptor o para aumentar la producción de un sacárido donador o un precursor del sacárido donador. En algunas realizaciones, la producción de productos galactosilados se potencia mediante la manipulación del microorganismo huésped. Por ejemplo, en *E. coli*, puede minimizarse la descomposición del ácido siálico usando una cepa huésped que carece de CMP-sialato sintasa (*NanA*). (En *E. coli*, CMP-sialato sintasa parece ser una enzima catabólica). Además en *E. coli*, cuando la lactosa es, por ejemplo, el sacárido aceptor o un producto intermedio en la síntesis de producto galactosilado, la descomposición de la lactosa puede minimizarse usando células huésped que son *LacZ*.

#### C. Caracterización y aislamiento de productos galactosilados

La producción de productos galactosilados puede monitorizarse, por ejemplo, determinando que se ha producido la producción del producto deseado o determinando que se ha agotado un sustrato tal como el sustrato aceptor. Los expertos reconocerán que pueden identificarse productos galactosilados tales como oligosacárido usando técnicas tales como cromatografía, por ejemplo, usando placas de CCF o papel, o mediante espectrometría de masas, por ejemplo, espectrometría MALDI-TOF, o mediante espectrometría RMN. Los expertos en la técnica conocen métodos de identificación de productos galactosilados y se encuentran, por ejemplo, en la patente estadounidense n.º 6.699.705, y en Varki *et al.*, Preparation and Analysis of Glycoconjugates, en Current Protocols in Molecular Biology, capítulo 17 (Ausubel *et al.* eds, 1993).

En algunas realizaciones, los polipéptidos de CgtE y los métodos de la presente invención se usan para sintetizar enzimáticamente una glicoproteína o glicolípido que tiene un patrón de glicosilación sustancialmente uniforme. Las glicoproteínas y glicolípidos incluyen un sacárido u oligosacárido que está unido a una proteína, glicoproteína, lípido o glicolípido para el que se desea una alteración de glicofórmula. El sacárido u oligosacárido incluye una estructura que puede funcionar como sustrato aceptor para una glicosiltransferasa. Cuando el sustrato aceptor está glicosilado, se forma el resto de oligosacárido deseado. El resto de oligosacárido deseado es uno que confiere la actividad biológica deseada a la glicoproteína o glicolípido al que está unido. En las composiciones de la invención, el residuo de sacárido preseleccionado se une a al menos aproximadamente el 30% de los posibles sitios aceptores de interés. Más preferiblemente, el residuo de sacárido preseleccionado se une a al menos aproximadamente el 50% de los posibles sustratos aceptores de interés, y todavía más preferiblemente a al menos el 70% de los posibles sustratos aceptores de interés. En situaciones en las que la glicoproteína o glicolípido de partida presenta heterogeneidad en el resto de oligosacárido de interés (por ejemplo, algunos de los oligosacáridos en la glicoproteína o glicolípido de partida ya tienen el residuo de sacárido preseleccionado unido al sustrato aceptor de interés), los porcentajes mencionados incluyen tales residuos de sacáridos previamente unidos.

El término "alterado" se refiere a la glicoproteína o glicolípido de interés que tiene un patrón de glicosilación que, tras la aplicación de los polipéptidos de CgtE y los métodos de la invención, es diferente del observado en la glicoproteína tal como se produjo originalmente. Un ejemplo de tales glicoconjugados son glicoproteínas en las que las glicofórmulas de las glicoproteínas son diferentes de las encontradas en la glicoproteína cuando se produce mediante células del organismo para el que la glicoproteína es nativa. También se proporcionan polipéptidos de CgtE y métodos de uso de tales proteínas para sintetizar enzimáticamente glicoproteínas y glicolípidos en los que el patrón de glicosilación de estos glicoconjugados se modifica en comparación con el patrón de glicosilación de los glicoconjugados tal como se producen originalmente mediante una célula huésped, que puede ser de la misma especie o de una diferente que las células a partir de las que se producen los glicoconjugados nativos.

Pueden evaluarse diferencias en los patrones de glicosilación no sólo mediante análisis estructural de las glicoproteínas y glicolípidos, sino también mediante la comparación de una o más actividades biológicas de los glicoconjugados. Por ejemplo, una glicoproteína que tiene una "glicofórmula alterada" incluye una que presenta una mejora en una o más actividades biológicas de la glicoproteína tras compararse la reacción de glicosilación con la glicoproteína no modificada. Por ejemplo, un glicoconjugado alterado incluye uno que, tras la aplicación de los polipéptidos de CgtE y los métodos de la invención, presenta una afinidad de unión mayor por un ligando o receptor de interés, una semivida terapéutica mayor, antigenicidad reducida y direccionamiento a tejidos específicos. Preferiblemente, la cantidad de mejora observada es estadísticamente significativa, y es más preferiblemente una mejora de al menos aproximadamente el 25%, y todavía más preferiblemente es de al menos aproximadamente el 30%, el 40%, el 50%, el 60%, el 70%, e incluso todavía más preferiblemente es de al menos el 80%, el 90% o el 95%.

Los productos producidos usando polipéptidos de CgtE pueden usarse sin purificación. Sin embargo, pueden usarse técnicas bien conocidas, convencionales, por ejemplo, cromatografía en capa fina o gruesa, cromatografía de intercambio iónico o filtración en membrana para recuperar los sacáridos glicosilados. Además, por ejemplo, puede usarse filtración en membrana utilizando una membrana de ósmosis inversa o nanofiltración tal como se describe en la patente australiana de titularidad compartida n.º 735695. Como ejemplo adicional, puede usarse filtración en membrana en la que las membranas tienen un punto de corte de peso molecular de aproximadamente 1000 a aproximadamente 10.000 Daltons para eliminar proteínas. Como otro ejemplo, puede usarse entonces nanofiltración u ósmosis inversa para eliminar sales. Las membranas de nanofiltro son una clase de membranas de ósmosis inversa que dejan pasar sales monovalentes pero retienen sales polivalentes y solutos no cargados mayores de aproximadamente 200 a aproximadamente 1000 Daltons, dependiendo de la membrana usada. Por tanto, por ejemplo, los oligosacáridos producidos mediante las composiciones y los métodos de la presente invención pueden retenerse en la membrana y las sales contaminantes pasarán a su través.

## **X. Síntesis de oligosacáridos multienzimática**

Tal como se trató anteriormente, en algunas realizaciones, pueden usarse dos o más enzimas para formar un oligosacárido deseado, incluyendo un determinante de oligosacárido en una glicoproteína o glicolípido. Por ejemplo, un determinante de oligosacárido particular podría requerir la adición de una galactosa, un ácido siálico y una fucosa con el fin de que presente una actividad deseada. Por consiguiente, la invención proporciona métodos en los que se usan dos o más glicosiltransferasas, por ejemplo, un polipéptido de CgtE, y otra glicosiltransferasa, tal como una fucosiltransferasa o una sialiltransferasa, para obtener la síntesis de alto rendimiento de un determinante de

oligosacárido deseado.

Los polipéptidos de CgtE, preparados tal como se describe en el presente documento, pueden usarse en combinación con una multitud de glicosiltransferasas. Por ejemplo, puede usarse una combinación de polipéptidos de CgtE recombinantes y fucosiltransferasas recombinantes, por ejemplo, una  $\alpha$ -1,3/4-fucosiltransferasa de *H. pylori*. Por ejemplo, se dan a conocer fucosiltransferasas de *Helicobacter pylori* en las patentes estadounidenses n.ºs 6.534.298 y 6.238.894; el documento WO2004009838, publicado el 29 de enero de 2004; documento USSN 10/764.212, (US 2005/0164338A1) presentado el 22 de enero de 2004. También pueden usarse glicosiltransferasas bacterianas, incluyendo  $\alpha$ 2,3-sialiltransferasas,  $\alpha$ 2,3-2,8-sialiltransferasas bifuncionales,  $\beta$ -1,4-GalNAc transferasas y  $\beta$ -1,3-galactosiltransferasas que se han aislado de *Campylobacter jejuni* y se dan a conocer en la patente estadounidense n.º 6.699.705, concedida el 2 de marzo de 2004. Se dan a conocer sialiltransferasas adicionales en la patente estadounidense n.º 6.096.529, concedida el 1 de agosto de 2000 y en el documento USSN 60/610.807, presentado el 17 de septiembre de 2004. De manera similar, las glicosiltransferasas recombinantes pueden usarse con enzimas auxiliares recombinantes, que pueden estar fusionadas o no con la glicosiltransferasa formando así una proteína de fusión. En otras realizaciones, los polipéptidos de CgtE y glicosiltransferasas adicionales y/o enzimas auxiliares se producen en la misma célula y se usan para sintetizar un producto final deseado.

En algunos casos, un oligosacárido unido a glicoproteína o glicolípido incluirá un sustrato aceptor para la glicosiltransferasa particular de interés tras la biosíntesis *in vivo* de la glicoproteína o glicolípido. Tales glicoproteínas o glicolípidos pueden glicosilarse usando las proteínas de fusión de glicosiltransferasas recombinantes y los métodos de la invención sin modificación previa del patrón de glicosilación de la glicoproteína o glicolípido, respectivamente. En otros casos, sin embargo, una glicoproteína o glicolípido de interés carecerá de un sustrato aceptor adecuado. En tales casos, pueden usarse los métodos de la invención para alterar el patrón de glicosilación de la glicoproteína o glicolípido de modo que los oligosacáridos unidos a glicoproteína o glicolípido incluyen entonces un sustrato aceptor para la unión catalizada por glicosiltransferasa de una unidad de sacárido preseleccionada de interés para formar un resto de oligosacárido deseado.

En primer lugar, pueden “recortarse” opcionalmente oligosacáridos unidos a glicoproteína o glicolípido, o bien en su totalidad o bien en parte, para exponer o bien un sustrato aceptor para la glicosiltransferasa o bien un resto al que pueden añadirse uno o más residuos apropiados para obtener un sustrato aceptor adecuado. Enzimas tales como glicosiltransferasas y endoglicosidasas son útiles para las reacciones de unión y recorte. Por ejemplo, una glicoproteína que presenta oligosacáridos de tipo de “alto contenido en manosa” puede someterse a recorte mediante una manosidasa para obtener un sustrato aceptor que, tras la unión de una o más unidades de sacárido preseleccionadas, forma el determinante de oligosacárido deseado.

Los métodos también son útiles para sintetizar un resto de oligosacárido deseado en una proteína o lípido que no está glicosilado en su forma nativa. Puede unirse un sustrato aceptor adecuado para la glicosiltransferasa correspondiente a tales proteínas o lípidos antes de la glicosilación usando los métodos de la presente invención. Véase, por ejemplo, la patente estadounidense n.º 5.272.066 para métodos de obtención de polipéptidos que tienen aceptores adecuados para la glicosilación.

Por tanto, en algunas realizaciones, la invención proporciona métodos para la sialilación *in vitro* de grupos sacárido presentes en un glicoconjugado que implican en primer lugar modificar el glicoconjugado para crear un aceptor adecuado.

Debe observarse que tal como se usa en el presente documento y en las reivindicaciones adjuntas, las formas singulares “un/una”, y “el/la” incluyen los referentes plurales a menos que el contexto dicte claramente lo contrario.

Las publicaciones tratadas en el presente documento se proporcionan únicamente por su divulgación antes de la fecha de presentación de la presente solicitud. Nada en el presente documento debe interpretarse como una admisión de que la presente invención no tiene derecho a anticipar tal publicación en virtud de la invención anterior. Además, las fechas de publicación proporcionadas pueden ser diferentes de las fechas de publicación reales, lo que puede ser necesario confirmar independientemente.

## Ejemplos

### Ejemplo 1: Clonación y expresión de CgtE de *C. jejuni* LIO87

Se expresaron marcos de lectura abiertos del locus LOS de *C. jejuni* LIO87 en *E. coli*. Se recogieron las células y se prepararon extractos celulares y se sometieron a ensayo para detectar la actividad glicosiltransferasa usando UDP-Glc, UDP-GlcNAc, UDP-Gal y UDPGalNAc como sustratos donadores y derivados fluorescentes de Glc, Gal, Lac, LacNAc y Gal- $\alpha$ -1,4-Lac como sustratos aceptores. El ORF n.º 7 tenía actividad con UDP-Gal y UDP-GalNAc como sustrato donador y el derivado de Gal- $\alpha$ -1,4-Lac (Pk) como sustrato aceptor. El ORF n.º 7 se denominó CgtE (glicosiltransferasa de *Campylobacter*; E es un código interno) y se caracterizó adicionalmente.

Se clonó el ácido nucleico que codifica para CgtE de *C. jejuni* LIO87 en un vector de expresión, pCWori+, o bien solo (construcción CJL-107) o bien como una fusión C-terminal con la proteína de unión a maltosa (MalE) de *E. coli* (construcción CJL-101). Se sometieron a electroporación CJL-101 y CJL-107 en la cepa de *E. coli* AD202. Se

expresó la proteína CgtE con la etiqueta MalE (de 10 a 15 unidades por litro) y se expresó como una proteína sin etiqueta (de 30 a 40 unidades por litro).

Ejemplo 2: Caracterización de la actividad CgtE de *C. jejuni* LIO87

5 Se caracterizó adicionalmente la actividad enzimática de CgtE. La actividad de CgtE era óptima desde pH 7 hasta 7,5. Eran necesarios cationes divalentes para la actividad proporcionando  $MnCl_2$  la actividad óptima (dos veces superior a la actividad observada con  $MgCl_2$ ).

10 Se observó que CgtE transfería residuos tanto de Gal como de GalNAc a un residuo de  $\alpha$ -Gal terminal. También se sometió a ensayo la capacidad de CgtE para transferir un azúcar a un residuo de  $\alpha$ -GalNAc terminal y los resultados se muestran en la tabla 1. CgtE usó tanto Gal $\alpha$ -Lac-FCHASE como GalNAc- $\alpha$ -FCHASE como sustrato aceptor. La actividad enzimática de CgtE era superior cuando se usó Gal $\alpha$ -Lac-FCHASE como sustrato aceptor.

**Tabla 1:** Especificidad de sustrato de CgtE (CJL-107)

| Aceptor                  | Donador    | Actividad (mU/ml) |
|--------------------------|------------|-------------------|
| Gal $\alpha$ -Lac-FCHASE | UDP-GalNAc | 63,5              |
| Gal $\alpha$ -Lac-FCHASE | UDP-Gal    | 7,0               |
| GalNAc- $\alpha$ -FCHASE | UDP-GalNAc | 0,3               |
| GalNAc- $\alpha$ -FCHASE | UDP-Gal    | 0,6               |

15 Con el fin de determinar la región y estereoespecificidad de la actividad de CgtE, se usó un sobrenadante de 15.000 rpm de CJL-107 para transferir un residuo de GalNAc a Gal $\alpha$ -1,4-Gal $\beta$ -1,4-GlcNAc-p-nitrofenilo. Véase, por ejemplo, la figura 1. En el ensayo, se sintetizaron 6,9 mg de GalNAc $\beta$ -1,3-Gal $\alpha$ -1,4-Gal $\beta$ -1,4-GlcNAc-p-nitrofenilo. Se asignaron las resonancias de  $^1H$  RMN al compuesto de tetrasacárido mediante el uso de espectros COSY y TOCSY homonucleares bidimensionales. Véase, por ejemplo, la tabla 2. Se usaron entonces estas asignaciones para identificar picos cruzados en un espectro de  $^1H$ - $^{13}C$  HSQC que correlaciona el desplazamiento químico de un átomo de protón con su carbono vecino directamente unido. Véase, por ejemplo, la figura 2. Debido a que pueden establecerse conectividades entre residuos a través de enlaces glicosídicos usando una secuencia de pulso de  $^1H$ - $^{13}C$  HMBC, se obtuvo un espectro de HMBC del compuesto de tetrasacárido para establecer los enlaces covalentes entre residuos de azúcar. Los espectros de HMBC confirmaron que la proteína CgtE transfería un residuo de GalNAc a  $\alpha$ -Gal en el precursor de trisacárido a través de un enlace  $\beta$ 1 $\rightarrow$ 3. Concordando con los espectros de HMBC, el átomo de carbono en la posición C3 de  $\alpha$ Gal se desplazaba a campo bajo en comparación con los valores del monosacárido, lo que es un indicador cualitativo de que participa en un enlace glicosídico con el residuo de  $\beta$ GalNAc adyacente. Véase, por ejemplo, la figura 2.

25 Aunque la invención anterior se ha descrito en cierto detalle a modo de ilustración y ejemplo para fines de claridad de comprensión, resultará fácilmente evidente para un experto habitual en la técnica en vista de las enseñanzas de esta invención que pueden hacerse ciertos cambios y modificaciones a la misma sin apartarse del alcance de las reivindicaciones adjuntas.

LISTA DE SECUENCIAS INFORMAL

SEQ ID NO: 1

Ácido nucleico de CgtE

ES 2 374 419 T3

ATGCCTAAAATTTCAATCATCTTGCCAACTTTTAATGTAGAAAAATATATAGCCAAAGCTTTAGAAA  
GTTGTATCAATCAAAGTTTTAAAGATATAGAAATAATTGTAGTGGATGATTGTGGAAGTGATAAAAG  
TATAGATATAGCTAAAGAATATGCCAAAAAAGATGAAAGAATAAAAAATAATTCACAATGAAGAAAAT  
TTAGGTCTTTTTAAGAGCTAGATATGAAGGAGTTAAGGCAGCTGGGGGGGGTATATTTATGTTTTTAG  
ACCCTGATGATTATTTAGAACTTAATGCTTGTGAAGAGTGTGTGAGAATTTTAAACACTGAAAAAGA  
AAGTGATTTTATATGGTTTGATTTTATATATAAAAAGAATTTAGGGGTTATAAATAGGGGAAAATTTT  
TTACAAGATCAAACCTTTACTATTTTTGAATATTGTGAGAATATCATAAATTCAAAATAAAAATATCT  
GTTATTGGAATCTTTGTTCTAAACTTATTAACAGAAATATATTTAGCATCTTTTTCTTTTTTTAGA  
AAAAGAGATATTAATACTAGATTAATTATGGCAGAAGATGCATTAATTTATTTTTTTATTATTTTA  
AATTGTGGAATAACTACTAGTGCAAAAAATATTTATTATTATTGTGAAAATGACAATAGTTCTG  
TTGGGACTAATGATATTGTCAAATGAAAAAACTTACAAGATGAAAAAATGGTTATTGGAATTTT  
ACTAGAATTTTTAAATCATCATAAAAAGAATATTGAATTGTATTTGTATGTTTTTTTTAAAATAATG  
ATTGGTAAATTAATCGTATATAAATTACATCGAGAAATAAACTATATCGATTAAAATATAATTATA  
TGATATATACAATGAAAAAATAAAAAATAAATTTATAATAAAATATTTTTTTATTGAAAAGATTTTT  
GATGGAAAAGTATTTTTAAAATAAATTTAATTGA

SEQ ID NO: 2

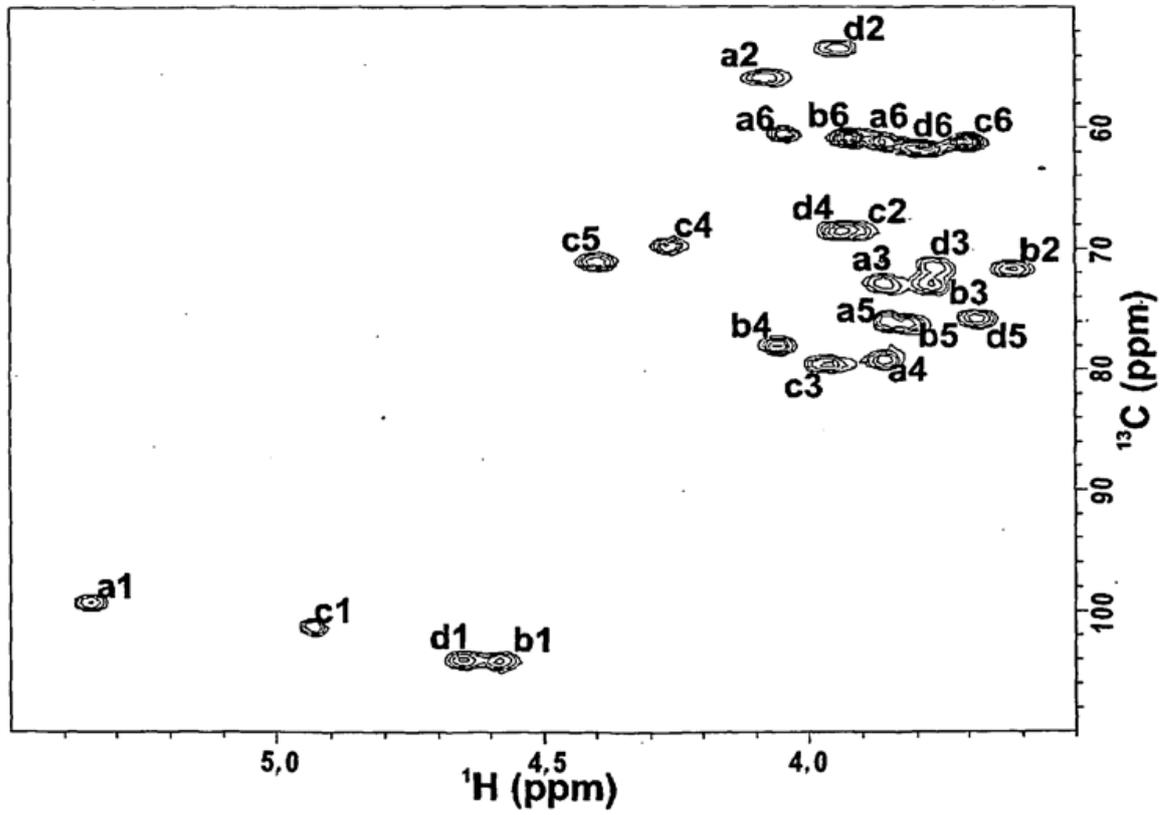
Aminoácidos de CgtE

MPKISIIILPTFNVEKYIAKALESCINQSFKDIEIIVVDDCGSDKSIDIAKEYAKKDERIKIIHNEEN  
LGLLRARYEGVKAAGGGYIMFLDPDDYLELNACEECVRILNTEKESDFIWFDIFYKRISGVINRGNF  
LQDQFTTIFEYCENIIIQNKNICYWNLCSKLIKTEIYLASFSELEKEILNTRLIMAEDALIYFFIIL  
NCGKITTSAKNIYYCENDNSSVGTNDIVKIEKNLQDEKMOVIGILLEFLNHHKKNIELYLYVFLKIM  
IGKLIVYKLRHREIKLYRLKYNMIYTMKKIKNKFIKYFLLKRFLMEKYFKNKFN

## REIVINDICACIONES

1. Uso de un polipéptido de  $\beta$ -1,3-N-acetilgalactosaminiltransferasa recombinante o aislado para una reacción de N-acetilgalactosaminilación, en el que un resto de N-acetilgalactosaminilo está unido a un resto de oligosacárido mediante un enlace  $\beta$ -1,3, en el que el polipéptido de  $\beta$ -1,3-N-acetilgalactosaminiltransferasa comprende una secuencia de aminoácidos con una identidad de al menos el 90% con SEQ ID NO: 2.
- 5
2. Mezcla de reacción para producir un producto de sacárido N-acetilgalactosaminilado en el que un resto de N-acetilgalactosaminilo está unido a un resto de oligosacárido mediante un enlace  $\beta$ -1,3, comprendiendo la mezcla de reacción:
- un polipéptido de  $\beta$ -1,3-N-acetilgalactosaminiltransferasa recombinante o aislado;
- 10 un sustrato donador que comprende un resto de N-acetilgalactosaminilo;
- y un sustrato aceptor que comprende un resto de oligosacárido, en el que el polipéptido de  $\beta$ -1,3-N-acetilgalactosaminiltransferasa comprende una secuencia de aminoácidos con una identidad de al menos el 90% con SEQ ID NO: 2.
- 15
3. Método de producción de un producto de sacárido N-acetilgalactosaminilado, comprendiendo el método la etapa de:
- a) poner en contacto un sustrato aceptor que comprende un resto de oligosacárido con un sustrato donador que comprende un resto de N-acetilgalactosaminilo y un polipéptido de  $\beta$ -1,3-N-acetilgalactosaminiltransferasa recombinante o aislado con una identidad de al menos el 90% con SEQ ID NO:2; y
- 20 b) permitir que se produzca la transferencia del resto de N-acetilgalactosaminilo al sacárido aceptor, en el que un resto de N-acetilgalactosaminilo está unido a un resto de oligosacárido mediante un enlace  $\beta$ -1,3 produciendo de ese modo el producto de sacárido N-acetilgalactosaminilado.
4. Uso según la reivindicación 1, mezcla de reacción según la reivindicación 2, o método según la reivindicación 3, en los que el polipéptido de  $\beta$ -1,3-N-acetilgalactosaminiltransferasa comprende una secuencia de aminoácidos con una identidad de al menos el 95% con SEQ ID NO:2.
- 25
5. Uso según la reivindicación 1, método según la reivindicación 3, en los que el polipéptido de  $\beta$ -1,3-N-acetilgalactosaminiltransferasa comprende una secuencia de aminoácidos de SEQ ID NO: 2.
6. Método según la reivindicación 5, realizándose el método a una escala de producción de gramo en una única reacción.
- 30





**FIG. 2**

### FIG. 3

Longitud de CD = 168 residuos, alineación del 95,8%  
 Puntuación = 108 bits (260), expectativa = 1e-24

|           |     |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             |     |
|-----------|-----|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| Consulta: | 5   | <b>S</b> <u>I</u> <u>I</u> <u>L</u> <u>P</u> <u>T</u> <u>E</u> <u>N</u> <u>V</u> <u>E</u> <u>K</u> <u>Y</u> <u>I</u> <u>A</u> <u>K</u> <u>A</u> <u>L</u> <u>E</u> <u>S</u> <u>C</u> <u>I</u> <u>N</u> <u>Q</u> <u>S</u> <u>E</u> <u>K</u> <u>D</u> <u>I</u> <u>E</u> <u>I</u> <u>I</u> <u>V</u> <u>V</u> <u>D</u> <u>D</u> <u>C</u> <u>G</u> <u>S</u> <u>D</u> <u>K</u> <u>S</u> <u>I</u> <u>D</u> <u>I</u> <u>A</u> <u>K</u> <u>E</u> <u>Y</u> <u>A</u> <u>K</u> <u>K</u> <u>D</u> <u>E</u> <u>R</u> <u>I</u> <u>K</u> <u>I</u> <u>I</u> <u>H</u> <u>N</u> | 64  |
| Objeto:   | 1   | <b>S</b> <u>V</u> <u>I</u> <u>I</u> <u>P</u> <u>T</u> <u>Y</u> <u>N</u> <u>E</u> <u>E</u> <u>K</u> <u>Y</u> <u>L</u> <u>E</u> <u>E</u> <u>T</u> <u>L</u> <u>E</u> <u>S</u> <u>L</u> <u>L</u> <u>A</u> <u>Q</u> <u>T</u> <u>Y</u> <u>P</u> <u>N</u> <u>F</u> <u>E</u> <u>I</u> <u>I</u> <u>V</u> <u>V</u> <u>D</u> <u>D</u> <u>G</u> <u>S</u> <u>T</u> <u>D</u> <u>G</u> <u>T</u> <u>V</u> <u>E</u> <u>I</u> <u>A</u> <u>E</u> <u>E</u> <u>Y</u> <u>A</u> <u>K</u> <u>N</u> <u>D</u> <u>P</u> <u>R</u> <u>I</u> <u>R</u> <u>V</u> <u>I</u> <u>R</u> <u>L</u> | 60  |
| Consulta: | 65  | <b>E</b> <u>E</u> <u>N</u> <u>L</u> <u>G</u> <u>L</u> <u>L</u> <u>R</u> <u>A</u> <u>R</u> <u>Y</u> <u>E</u> <u>G</u> <u>V</u> <u>K</u> <u>A</u> <u>A</u> <u>G</u> <u>G</u> <u>Y</u> <u>I</u> <u>M</u> <u>F</u> <u>L</u> <u>D</u> <u>P</u> <u>D</u> <u>D</u> <u>Y</u> <u>L</u> <u>E</u> <u>L</u> <u>N</u> <u>A</u> <u>C</u> <u>E</u> <u>C</u> <u>V</u> <u>R</u> <u>I</u> <u>L</u> <u>N</u> <u>T</u> <u>E</u> <u>K</u> <u>E</u> <u>S</u> <u>D</u> <u>F</u> <u>I</u> <u>W</u> <u>F</u> <u>D</u> <u>F</u> <u>I</u> <u>Y</u> <u>K</u> <u>R</u>                   | 124 |
| Objeto:   | 61  | <b>E</b> <u>E</u> <u>N</u> <u>L</u> <u>G</u> <u>K</u> <u>A</u> <u>A</u> <u>R</u> <u>N</u> <u>A</u> <u>G</u> <u>L</u> <u>K</u> <u>H</u> <u>A</u> <u>T</u> <u>G</u> <u>D</u> <u>Y</u> <u>I</u> <u>L</u> <u>F</u> <u>L</u> <u>D</u> <u>A</u> <u>D</u> <u>D</u> <u>E</u> <u>V</u> <u>A</u> <u>P</u> <u>D</u> <u>W</u> <u>L</u> <u>E</u> <u>K</u> <u>L</u> <u>V</u> <u>E</u> <u>L</u> <u>L</u> <u>-</u> <u>E</u> <u>K</u> <u>N</u> <u>G</u> <u>A</u> <u>D</u> <u>I</u> <u>V</u> <u>I</u> <u>G</u> <u>S</u> <u>R</u> <u>V</u> <u>A</u> <u>I</u>                   | 118 |
| Consulta: | 125 | <b>I</b> <u>S</u> <u>G</u> <u>V</u> <u>I</u> <u>N</u> <u>R</u> <u>G</u> <u>N</u> <u>F</u> <u>L</u> <u>Q</u> <u>D</u> <u>Q</u> <u>T</u> <u>F</u> <u>T</u> <u>I</u> <u>F</u> <u>E</u> <u>Y</u> <u>C</u> <u>E</u> <u>N</u> <u>I</u> <u>I</u> <u>Q</u> <u>N</u> <u>K</u> <u>N</u> <u>I</u> <u>C</u> <u>Y</u> <u>W</u> <u>N</u> <u>L</u> <u>C</u> <u>S</u> <u>K</u> <u>L</u> <u>I</u> <u>K</u>                                                                                                                                                                   | 167 |
| Objeto:   | 119 | <b>F</b> <u>G</u> <u>E</u> <u>T</u> <u>R</u> <u>L</u> <u>D</u> <u>G</u> <u>R</u> <u>A</u> <u>L</u> <u>R</u> <u>M</u> <u>E</u> <u>L</u> <u>L</u> <u>L</u> <u>L</u> <u>L</u> <u>L</u> <u>G</u> <u>K</u> <u>L</u> <u>G</u> <u>A</u> <u>R</u> <u>S</u> <u>L</u> <u>G</u> <u>L</u> <u>K</u> <u>V</u> <u>L</u> <u>F</u> <u>L</u> <u>I</u> <u>G</u> <u>S</u> <u>N</u> <u>A</u> <u>L</u> <u>Y</u> <u>R</u> <u>R</u>                                                                                                                                                 | 161 |