

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 605 677**

51 Int. Cl.:

C12N 9/10 (2006.01)

C12N 15/54 (2006.01)

C12P 21/02 (2006.01)

A61K 38/14 (2006.01)

C12Q 1/68 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **10.11.2010 PCT/US2010/056230**

87 Fecha y número de publicación internacional: **19.05.2011 WO11060069**

96 Fecha de presentación y número de la solicitud europea: **10.11.2010 E 10830656 (4)**

97 Fecha y número de publicación de la concesión europea: **14.09.2016 EP 2499245**

54 Título: **Glucosiltransferasa de hámster chino y métodos relacionados**

30 Prioridad:

11.11.2009 US 260232 P

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

15.03.2017

73 Titular/es:

**MOMENTA PHARMACEUTICALS, INC. (100.0%)
675 West Kendall Street
Cambridge, MA 02142, US**

72 Inventor/es:

MEADOR III, JAMES, W.

74 Agente/Representante:

LEHMANN NOVO, María Isabel

ES 2 605 677 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

Glucosiltransferasa de hámster chino y métodos relacionados

ANTECEDENTES DE LA INVENCION

5 El glucano de galactosa- α 1,3-galactosa terminal (denominado en el presente documento Gal- α -Gal) puede ser una modificación funcionalmente relevante a las proteínas N-glucosiladas dado su potencial inmunogénico en seres humanos cuando se presenta en terapéuticos biológicos heterológamente derivados. La presencia de este epítipo de hidrato de carbono sobre proteínas endógenas parece ser altamente específico de especie, por ejemplo, detectado en cerdo, ratón y rata, pero ausente en primates. Así, el epítipo puede estar presente en biológicos recombinantes producidos en sistemas de expresión celular derivados de ciertos organismos (por ejemplo, células NS0 o SP/2 de ratón, cerdos transgénicos). El anticuerpo monoclonal cetuximab (Erbix®) es un ejemplo tal de un producto farmacéutico de proteína comercial producido en una línea celular derivada de ratón y que también se ha informado que contiene el epítipo de hidrato de carbono Gal- α -Gal (Chung et al. (2008) N Engl J Med 358:1109-1117). La biosíntesis enzimática del glucano de Gal- α 1,3-Gal se ha atribuido a 1,3-glucosiltransferasa-1 (denominado en el presente documento Ggta1) (Taylor et al. (2003) Glycobiology 13:327-337; Smith et al. (1990) J Biol Chem 265:6225-6234).

SUMARIO DE LA INVENCION

La invención se basa, al menos en parte, en el descubrimiento de secuencias de genes de α -1,3-glucosiltransferasa-1 (Ggat1) en tanto el hámster chino (*Cricetulus griseus*) del que se derivan las células CHO como en una línea celular CHO usada para la producción de proteínas recombinantes. Por consiguiente, la invención se refiere a composiciones de ácido nucleico y de polipéptido relacionadas con los genes descubiertos, y vectores y células relacionados (por ejemplo, células CHO genéticamente manipuladas para reducir, eliminar o aumentar la actividad de Ggat1 y vectores útiles para producir tales células). Además, la invención se refiere al uso de las secuencias identificadas en diversos métodos, por ejemplo, métodos de detección de actividad de Ggat1 en células CHO (incluyendo la actividad transcripcional y/o actividad enzimática de Ggat1), por ejemplo, para cribar células CHO para la capacidad para producir estructuras de Gal- α -Gal, o para identificar y cuantificar la expresión de Ggat1 o actividad en una célula CHO, por ejemplo, una célula usada para la producción de una glucoproteína terapéutica.

Las secuencias de Ggat1 desveladas en el presente documento se enumeran en la Tabla 1:

FUENTE (NOMBRE)	TIPO DE SECUENCIA	SEQ ID NO
Tejido de ovario de hámster chino (Ggat1-a)	ADN; secuencia codificante de longitud completa (Figura 3)	SEQ ID NO: 1
Tejido de ovario de hámster chino (Ggat1-a)	polipéptido; secuencia codificante de longitud completa traducida (Figura 3)	SEQ ID NO: 2
Tejido del bazo de hámster chino (Ggat1-b)	ADN; secuencia codificante de longitud completa (Figura 4)	SEQ ID NO: 3
Tejido del bazo de hámster chino (Ggat1-b)	polipéptido; secuencia codificante de longitud completa traducida (Figura 4)	SEQ ID NO: 4
Línea celular de ovario de hámster chino (CHO) (Ggat1-c)	ADN; secuencia de ADN genómico correspondiente a los exones 8 y 9 e intrón intermedio (Figura 5)	SEQ ID NO: 5
Línea celular de ovario de hámster chino (CHO) (Ggat1-c)	ADN; secuencia codificante de los exones 8 y 9 (Figura 6)	SEQ ID NO: 6
Línea celular de ovario de hámster chino (CHO) (Ggat1-c)	polipéptido; secuencia traducida de los exones 8 y 9 (Figura 6)	SEQ ID NO: 7

30 La presente invención proporciona una molécula de ácido nucleico aislada; una construcción de ácido nucleico; una célula huésped aislada; un método de modulación de la estructura de glucano de una glucoproteína terapéutica recombinante producida en una célula CHO; un polipéptido aislado; un anticuerpo aislado o fragmento de unión al antígeno del mismo; un método de detección de la expresión o actividad de Ggat1 en una población de células CHO; una matriz bidimensional; y un método de análisis de una muestra de una célula CHO usando la matriz bidimensional; cada uno como se define en las reivindicaciones adjuntas.

Más generalmente, en un primer aspecto, la presente divulgación presenta una molécula de ácido nucleico aislada que comprende una secuencia de nucleótidos que codifica un polipéptido Ggta1 de hámster chino o de células CHO, o una porción activa del mismo que tiene una o más actividad de Ggta1, por ejemplo, una actividad de Ggta1 descrita en el presente documento.

5 En una realización, la molécula de ácido nucleico aislada comprende una secuencia de nucleótidos que tiene al menos aproximadamente el 75 % de identidad de secuencias, o al menos aproximadamente el 80 % de identidad de secuencias, o al menos aproximadamente el 85 % de identidad de secuencias, o al menos aproximadamente el 90 % de identidad de secuencias, o al menos aproximadamente el 92 % de identidad de secuencias, o al menos aproximadamente el 93 % de identidad de secuencias, o al menos aproximadamente el 94 % de identidad de secuencias, o al menos aproximadamente el 95 % de identidad de secuencias, o al menos aproximadamente el 96 % de identidad de secuencias, o al menos aproximadamente el 97 % de identidad de secuencias, o al menos aproximadamente el 98 % de identidad de secuencias, o al menos aproximadamente el 99 % de identidad de secuencias con (a) una molécula de ADN que tiene una secuencia que codifica la secuencia de restos de aminoácidos de SEQ ID NO: 2, o SEQ ID NO: 4 o SEQ ID NO: 7 o (b) el complemento de la molécula de ADN de (a). La molécula de ácido nucleico aislada puede codificar un polipéptido que tiene una o más actividad de Ggta1. En una realización, la molécula de ácido nucleico aislada codifica la secuencia de SEQ ID NO: 2, o SEQ ID NO: 4 o SEQ ID NO: 7.

20 En una realización, la molécula de ácido nucleico aislada tiene al menos aproximadamente el 90 % de identidad de secuencias, o al menos aproximadamente el 91 % de identidad de secuencias, o al menos aproximadamente el 92 % de identidad de secuencias, o al menos aproximadamente el 93 % de identidad de secuencias, o al menos aproximadamente el 94 % de identidad de secuencias, o al menos aproximadamente el 95 % de identidad de secuencias, o al menos aproximadamente el 96 % de identidad de secuencias, o al menos aproximadamente el 97 % de identidad de secuencias, o al menos aproximadamente el 98 % de identidad de secuencias, o al menos aproximadamente el 99 % de identidad de secuencias con la secuencia de SEQ ID NO: 1 o SEQ ID NO: 3, SEQ ID NO: 5 o SEQ ID NO: 6. La molécula de ácido nucleico aislada puede codificar un polipéptido que tiene una o más actividad de Ggta1.

En una realización, la molécula de ácido nucleico aislada se hibrida bajo condiciones de alta rigurosidad con una molécula de ácido nucleico que comprende la secuencia de nucleótidos de SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5 o SEQ ID NO: 6.

30 En una realización, la molécula de ácido nucleico aislada comprende la secuencia de nucleótidos de SEQ ID NO: 1 o SEQ ID NO: 3, SEQ ID NO: 5 o SEQ ID NO: 6.

Una molécula de ácido nucleico aislada de la presente divulgación puede corresponderse con una molécula de ácido nucleico que existe de forma natural, por ejemplo, con una secuencia de ácidos nucleicos de CHO que existe de forma natural que codifica una proteína Ggta1 expresada. En algunas realizaciones, la molécula de ácido nucleico puede codificar un polipéptido que tiene una o más actividad de Ggta1, sin importar si la molécula de ácido nucleico existe de forma natural. Por ejemplo, la molécula de ácido nucleico aislada es una variante que no existe de forma natural de una secuencia de Ggta1 de hámster chino o de células CHO. En una realización, la molécula de ácido nucleico aislada es una molécula de ácido nucleico antisentido, una molécula de iARN (por ejemplo, una ARNbc o ARNip), una ribozima, o un ácido nucleico peptídico, en la que la molécula antisentido inhibe la producción de Ggta1 descrita en el presente documento.

En una realización, la molécula de ácido nucleico aislada codifica un polipéptido Ggta1 descrito en el presente documento asociado a una secuencia de aminoácidos heteróloga, por ejemplo, codifica una proteína de fusión de Ggta1, por ejemplo, un Ggta1 marcado con epítope.

45 En un segundo aspecto, la divulgación presenta un oligonucleótido que comprende al menos 10 nucleótidos consecutivos (por ejemplo, al menos 12, 15, 18, 20, 25, 30, 35, 40, 45 o 50 nucleótidos consecutivos) y menos de 500 nucleótidos consecutivos (por ejemplo, menos de 450, 400, 350, 300, 250, 200, 180, 160, 140, 130, 125, 120, 100, 90, 80, 75, 60, o 50 nucleótidos consecutivos) de SEQ ID NO: 1 o su complemento, SEQ ID NO: 3 o su complemento, SEQ ID NO: 5 o su complemento, o SEQ ID NO: 6 o su complemento. En una realización, el oligonucleótido tiene entre 10 y 200 nucleótidos de longitud, entre 20 y 200 nucleótidos de longitud, entre 20 y 100 nucleótidos de longitud, entre 15 y 100 nucleótidos de longitud, entre 10 y 100 nucleótidos de longitud, entre 20 y 75 nucleótidos de longitud, entre 25 y 50 nucleótidos de longitud, entre 20 y 50 nucleótidos de longitud, entre 15 y 50 nucleótidos de longitud, entre 10 y 50 nucleótidos de longitud, entre 10 y 40 nucleótidos de longitud, entre 10 y 30 nucleótidos de longitud, o entre 10 y 25 nucleótidos de longitud. Tales oligonucleótidos pueden usarse, por ejemplo, como cebadores, marcas o sondas en métodos de detección y/o medición de la expresión de Ggta1 en células CHO (por ejemplo, usando métodos tales como PCR, chips de ADN, o análisis en serie de la expresión génica (SAGE)). En una realización, el oligonucleótido tiene o incluye una secuencia mostrada en la Tabla 2 o la Tabla 4.

En una realización, el oligonucleótido tiene una o más de las siguientes características: contenido de GC del 50-65 %, carece de estructura secundaria sustancial, punto de fusión (Tm) entre 50 y 60 °C, sin repeticiones de GC más largas de 3 bases, pares de GC en los extremos, abarca un límite intrón-exón.

- 5 En una realización, la divulgación presenta un conjunto de dos de tales oligonucleótidos (por ejemplo, un par de cebadores) que son capaces juntos de amplificar al menos una porción de un ADN de Ggta1 derivado de células CHO, por ejemplo, por PCR. Un conjunto de cebadores a modo de ejemplo incluye un cebador directo que se hibrida con la hebra codificante de una molécula de ácido nucleico de Ggta1 y un cebador inverso que se hibrida con la hebra no codificante de una molécula de ácido nucleico de Ggta1.
- 10 En otra realización, un oligonucleótido de la divulgación está operativamente enlazado a un elemento regulador (por ejemplo, operativamente enlazado a un promotor en una construcción de vector). En algunas realizaciones, el oligonucleótido está en la orientación sentido. En otras realizaciones, el oligonucleótido está en la orientación antisentido.
- 15 En un tercer aspecto, la divulgación presenta una construcción de ácido nucleico tal como un vector (por ejemplo, un plásmido, cósmido, partícula viral o fago) que incluye la secuencia de una molécula de ácido nucleico u oligonucleótido descrita en el presente documento, operativamente enlazado a al menos un elemento regulador para la expresión o clonación, por ejemplo, a un promotor. El vector puede incluir elementos adicionales, tales como uno o más de: un potenciador, una secuencia señal, un origen de replicación, uno o más genes marcadores, y una secuencia de terminación de la transcripción.
- 20 En una realización, el vector se diseña para la expresión de un polipéptido Ggta1 de hámster chino o de células CHO descrito en el presente documento, en una célula huésped (por ejemplo, una célula procarionta (por ejemplo, *E. coli*) o eucariota (por ejemplo, una célula de mamífero tal como una célula CHO)).
- En una realización, el vector incluye una molécula de ADN de la divulgación clonada en el vector de expresión en una orientación antisentido.
- 25 En una realización, el vector contiene una molécula de ácido nucleico descrita en el presente documento, por ejemplo, una molécula de ácido nucleico de Ggta1 de hámster chino o de células CHO descrita en el presente documento, configurada para permitir que se recombine en un sitio específico del genoma de una célula huésped, por ejemplo, configurada para modificar, alterar o inactivar un gen de Ggta1 endógeno en la célula huésped.
- 30 En un cuarto aspecto, la divulgación presenta células huésped aisladas transfectadas con una molécula de ácido nucleico, oligonucleótido, o construcción de ácido nucleico descritos en el presente documento. Las células huésped pueden ser procariontas (por ejemplo, *E. coli*) o eucariotas (por ejemplo, células vegetales, células de levadura o células de mamífero tales como las células CHO). En una realización, las células huésped se manipulan además genéticamente para expresar una glucoproteína terapéutica recombinante.
- 35 En una realización, las células huésped son células CHO transfectadas con una molécula de ácido nucleico, oligonucleótido, o construcción de ácido nucleico descritos en el presente documento (por ejemplo, un vector que incluye una secuencia codificante de Ggta1-a, Ggta1-b o Ggta1-c descrita en el presente documento), en la que las células CHO expresan niveles elevados de Ggta1 con respecto a las células CHO originales. En tales realizaciones, las células huésped pueden ser capaces de producir una glucoproteína (por ejemplo, una glucoproteína terapéutica recombinante) con niveles elevados de glucanos de galactosa- α 1,3-galactosa terminal en comparación con las células parentales. Métodos de producción de una glucoproteína (por ejemplo, una glucoproteína terapéutica recombinante) que tiene niveles elevados de glucanos de galactosa- α 1,3-galactosa terminal empleando tales células huésped para producir la glucoproteína también se proporcionan por la presente divulgación.
- 40 En otra realización, las células huésped son células CHO transfectadas con una molécula de ácido nucleico, oligonucleótido, o construcción de ácido nucleico descritos en el presente documento (por ejemplo, transfectadas con un ARNbc de Ggta1-a, Ggta1-b o Ggta1-c, ARNip, o vector inactivado), en la que la célula CHO tiene expresión reducida de una Ggta1 descrita en el presente documento con respecto a una célula CHO sin transformar, por ejemplo, las células huésped son células silenciadas o inactivadas para Ggta1. En tales realizaciones, las células huésped pueden ser capaces de producir una glucoproteína (por ejemplo, una glucoproteína terapéutica recombinante) con niveles más bajos de glucanos de galactosa- α 1,3-galactosa terminal en comparación con las células parentales de las células huésped. Métodos de producción de una glucoproteína (por ejemplo, una glucoproteína terapéutica recombinante) que tiene niveles más bajos de glucanos de galactosa- α 1,3-galactosa terminal empleando tales células huésped para producir la glucoproteína también se proporcionan por la presente divulgación.
- 45 En un quinto aspecto, la divulgación proporciona un proceso de modulación de la estructura de glucanos de una glucoproteína, por ejemplo, una glucoproteína terapéutica recombinante, por ejemplo, modulación del nivel de glucanos de galactosa- α 1,3-galactosa terminal presentes en una glucoproteína terapéutica recombinante. El método incluye cultivar las células huésped descritas en el presente documento (por ejemplo, las células huésped transfectadas con Ggta1 o silenciadas o inactivadas para Ggta1) en condiciones adecuadas para la expresión de la glucoproteína.
- 50 En un sexto aspecto, la divulgación proporciona un polipéptido aislado que incluye la secuencia de aminoácidos de una Ggta1 descrita en el presente documento, o un fragmento activo de la misma. En una realización, el polipéptido
- 55

incluye una secuencia de aminoácidos codificada por cualquiera de las secuencias de ácidos nucleicos aisladas descritas en el presente documento.

En una realización específica, el polipéptido incluye una secuencia de aminoácidos que comprende la secuencia de SEQ ID NO: 2, SEQ ID NO: 4 o SEQ ID NO: 7.

- 5 En otra realización, el polipéptido incluye una secuencia de aminoácidos que tienen al menos aproximadamente el 90 % de identidad de secuencias, por ejemplo, al menos aproximadamente el 91 % de identidad de secuencias, por ejemplo, al menos aproximadamente el 92 % de identidad de secuencias, por ejemplo, al menos aproximadamente el 93 % de identidad de secuencias, por ejemplo, al menos aproximadamente el 94 % de identidad de secuencias, por ejemplo, al menos aproximadamente el 95 % de identidad de secuencias, por ejemplo, al menos aproximadamente el 96 % de identidad de secuencias, por ejemplo, al menos aproximadamente el 97 % de identidad de secuencias, por ejemplo, al menos aproximadamente el 98 % de identidad de secuencias, por ejemplo, al menos aproximadamente el 99 % de identidad de secuencias con la secuencia de restos de aminoácidos de SEQ ID NO: 2, SEQ ID NO: 4 o SEQ ID NO: 7. El polipéptido puede tener una o más actividad de Ggta1.

- 15 En otra realización, el polipéptido Ggta1, o fragmento del mismo, se diferencia de la secuencia correspondiente de SEQ ID NO: 2, SEQ ID NO: 4 o SEQ ID NO: 7. En una realización se diferencia en al menos uno, pero en menos de 20, 15, 10 o 5 restos de aminoácidos. En otra se diferencia de la secuencia correspondiente en SEQ ID NO: 2, SEQ ID NO: 4 o SEQ ID NO: 7 en al menos un resto, pero menos del 10 % o 5 % de los restos. En una realización, las diferencias incluyen sustituciones conservativas. En otra realización, las diferencias incluyen sustituciones no conservativas. El polipéptido puede tener una o más actividad de Ggta1.

- 20 En una realización, un polipéptido Ggta1 descrito en el presente documento está enlazado a una secuencia de aminoácidos heteróloga, por ejemplo, el polipéptido es una proteína de fusión de Ggta1, por ejemplo, una Ggta1 marcada con epítotope.

- 25 En un séptimo aspecto, la divulgación proporciona un anticuerpo o fragmento de unión al antígeno del mismo, que se une específicamente a un polipéptido Ggta1 de célula CHO o de hámster chino descrito en el presente documento. Opcionalmente, el anticuerpo es un anticuerpo monoclonal, un fragmento funcional de anticuerpo (por ejemplo, un fragmento Fab) o un anticuerpo monocatenario.

- 30 En una realización, el anticuerpo o fragmento funcional del mismo se une a una Ggta1 de célula CHO o de hámster chino con una mayor afinidad de unión que a una Ggta1 que existe de forma natural de uno o más de: ratón, rata, vaca o perro. Por ejemplo, el anticuerpo o fragmento funcional del mismo se une a una Ggta1 de célula CHO o de hámster chino con una afinidad del 20 %, 25 %, 30 %, 40 %, 50 %, 60 % o mayor que a una Ggta1 que existe de forma natural de uno o más de: ratón, rata, vaca o perro. En una realización, el anticuerpo o fragmento funcional del mismo se une a una Ggta1 de célula CHO o de hámster chino descrita en el presente documento y no se une a una Ggta1 que existe de forma natural de uno o más de: ratón, rata, vaca o perro. La afinidad de unión puede determinarse por un método conocido en la técnica, por ejemplo, un radioinmunoensayo (RIA), ELISA, o unión en un formato de soporte de columna. La disociación se realiza aumentando las condiciones desnaturalizantes, o por competición con un ligando relacionado. La constante de disociación, K_d , se determina por una representación de Scatchard.

- 35 En un octavo aspecto, la divulgación presenta métodos de detección, y opcionalmente de cuantificación, de la expresión y/o actividad de Ggta1 en células CHO.

- 40 En una realización, el método incluye poner en contacto una muestra de ácido nucleico de una población de células CHO con una molécula de oligonucleótido o de ácido nucleico de Ggta1 de hámster chino o de células CHO aislada descrita en el presente documento para obtener un valor (por ejemplo, un valor relativo o absoluto) para el nivel de expresión o actividad de Ggta1 en la población de CHO. En una realización, el método incluye emplear una sonda o cebador o molécula de ácido nucleico de Ggta1 descrita en el presente documento en un método de hibridación (por ejemplo, transferencia Southern o Northern), hibridación *in situ* (por ejemplo, hibridación *in situ* con fluorescencia o FISH), hibridación de matrices, análisis de PCR (por ejemplo, RT-PCR, qPCR), análisis en serie de la expresión génica (SAGE), protección por RNasa, hibridación de ácidos nucleicos de sándwich de ADN ramificado (por ejemplo, sistema Quantigene®). En una realización, las células CHO expresan una glucoproteína terapéutica recombinante. Tales métodos pueden usarse, por ejemplo, para cribar clones de células CHO para la expresión diferencial del gen Ggta1.

- 50 En otra realización, el método incluye poner en contacto una muestra de proteína de una población de células CHO con un polipéptido Ggta1 de hámster chino o de células CHO o anticuerpo descrito en el presente documento para obtener un valor (por ejemplo, un valor relativo o absoluto) para la presencia del nivel de Ggta1, expresión o actividad en la población de CHO. En una realización, el método incluye emplear un anticuerpo contra Ggta1 o fragmento del mismo descrito en el presente documento en un método de detección basado en anticuerpo, por ejemplo, ensayo de inmunoadsorción (ELISAs), inmunoprecipitaciones, inmunofluorescencia, inmunoensayo enzimático (EIA), radioinmunoensayo (RIA), análisis de transferencia Western, resonancia de plasmones superficiales (SPR). En una realización, las células CHO expresan una glucoproteína terapéutica recombinante.

Tales métodos pueden usarse, por ejemplo, para cribar clones de células CHO para la presencia diferencial, expresión o actividad de Ggta1.

En una realización, el método incluye realizar una reacción de PCR en tiempo real cuantitativa (qPCR) para evaluar la expresión de Ggta1 en una preparación de células CHO. En una realización, el método de qPCR emplea un conjunto de oligonucleótidos Ggta1 descritos en el presente documento, por ejemplo, emplea un par de cebadores descritos en el presente documento (por ejemplo, un par de cebadores descritos en la Tabla 2 o la Tabla 4).

En una realización, el valor obtenido para el nivel de expresión de Ggta1 o actividad en la población de CHO se compara con un valor de referencia, por ejemplo, un valor de control, un valor predeterminado, o un valor para la expresión o actividad de Ggta1 en una segunda población de células CHO. Por ejemplo, el valor obtenido puede compararse con un valor obtenido de células CHO cultivadas bajo diferentes condiciones de bioproceso tales como diferentes medios de formulación y/o escala.

En un noveno aspecto, la divulgación presenta un método de evaluación de una muestra derivada de una célula CHO. El método incluye proporcionar una muestra de ácido nucleico derivada de una célula CHO, por ejemplo, de una célula CHO que expresa una glucoproteína terapéutica recombinante, y determinar un perfil de expresión génica de la muestra, en el que el perfil incluye un valor que representa el nivel de expresión de Ggta1 en la célula CHO, y un valor que representa el nivel de expresión de otro gen (por ejemplo, otra enzima) en la célula CHO. El método puede incluir poner en contacto la muestra con una o más moléculas de ácidos nucleicos de Ggta1 o sondas o cebador descritos en el presente documento. El método pueden incluir además comparar el valor o perfil (es decir, múltiples valores) con un valor de referencia o perfil de referencia. El perfil de expresión génica de la muestra puede obtenerse por cualquiera de los métodos descritos en el presente documento (por ejemplo, proporcionando un ácido nucleico de la muestra y poniendo en contacto el ácido nucleico con una matriz que incluye una sonda de Ggta1 descrita en el presente documento). El método puede usarse para determinar el nivel de expresión o actividad de Ggta1 en una célula CHO, entre el nivel de expresión o actividad de otros genes.

En un décimo aspecto, la divulgación presenta una matriz bidimensional que tiene una pluralidad de direcciones, siendo cada dirección de la pluralidad posicionalmente distinguible de cada otra dirección de la pluralidad, y teniendo cada dirección de la pluralidad una sonda de captura única, por ejemplo, una secuencia de ácidos nucleicos o de péptidos o anticuerpo. Al menos una dirección de la pluralidad tiene una sonda de captura que reconoce una molécula de Ggta1 descrita en el presente documento (por ejemplo, una molécula de Ggta-1, Ggta1-b o Ggta1-c descrita en el presente documento). En una realización, la sonda de captura es un ácido nucleico, por ejemplo, una sonda complementaria a una secuencia de ácidos nucleicos de Ggta1. En otra realización, la sonda de captura es un anticuerpo, por ejemplo, un anticuerpo específico para un polipéptido Ggta1 descrito en el presente documento. También se presenta un método de análisis de una muestra (por ejemplo, una muestra de una célula CHO) poniendo en contacto la muestra con la matriz anteriormente mencionada y detectando la unión de la muestra a la matriz. Otras características y ventajas serán evidentes de la siguiente descripción detallada, y de las reivindicaciones.

BREVE DESCRIPCIÓN DE LAS FIGURAS

La **Figura 1** es una fotografía de una separación electroforética en gel de agarosa de productos de PCR de Ggta1 amplificados de conjuntos de ADNc preparados a partir de cerebro (carril 3), riñón (carril 4), ovario (carril 5) y bazo (carril 6) de hámster chino. Los carriles 1 y 2 muestran patrones de tamaño en kB. El ADN se visualiza por tinción con bromuro de etidio.

La **Figura 2** es un alineamiento de múltiples secuencias de Ggta1 de rata (NM_145674, con el exón 5 y 6 delecionados), ratón (NM_010283, longitud completa), vaca (NM_177511, longitud completa), perro (XM_548478, longitud completa), ovario de hámster chino (secuencia consenso), bazo de hámster chino (secuencia consenso) y exones 8 y 9 de células CHO a nivel de aminoácidos. Los números de acceso de NCBI se indican entre paréntesis.

La **Figura 3** es la secuencia de Ggta1-a clonada de ADNc derivado de ovario de hámster chino. La secuencia tiene 1110 pares de bases (SEQ ID NO: 1) y se traduce en 370 aminoácidos (SEQ ID NO: 2), una proteína con un peso molecular predicho de 44.002 Daltons.

La **Figura 4** es la secuencia de Ggta1-b clonada de ADNc derivado de hámster chino. La secuencia tiene 1110 pares de bases (SEQ ID NO: 3) y se traduce en 370 aminoácidos (SEQ ID NO: 4), una proteína con un peso molecular predicho de 44.016 Daltons.

La **Figura 5** es la secuencia genómica de Ggta1 amplificada a partir de la línea de células CHO DHFR-. La secuencia consiste en los exones 8 y 9 (MAYÚSCULAS) con un intrón intermedio (en minúsculas) (SEQ ID NO: 5).

La **Figura 6** es la secuencia de aminoácidos traducida para el exón 8 y 9 de Ggta1 de la línea de células CHO. La secuencia tiene 276 aminoácidos (SEQ ID NO: 6), con un peso molecular predicho de 32.597 Daltons.

La **Figura 7** muestra perfiles de ciclos de PCR de clones seleccionados cribados por qPCR para evaluar la expresión diferencial del gen *Ggta1* en los clones de la línea de células CHO (7A) y análisis de T_m como evaluación de la especificidad del cebador (7B).

5 La **Figura 8** es una fotografía de una separación electroforética en gel de agarosa de productos de PCR de *Ggta1* amplificados. Se usaron los cebadores 863 y 830 para amplificar el ADN genómico de la línea de células CHO (DHFR-) (carriles 2-5) o ADNc derivado de la misma línea celular (carriles 6 y 7).

10 La **Figura 9** es una fotografía de una separación electroforética en gel de agarosa de productos de PCR de *Ggta1* amplificados. Se usaron los cebadores 875 (directo) y 855 (inverso) específicos de *Ggta1* que engloban dos exones (exón 8 y 9) y diseñados para diferenciar entre copias génicas y genómicas expresadas para perfilar tres clones amplificados diferentes de la línea de células CHO DHFR- que expresa CTLA4-Ig. Los clones se aislaron por dilución tras la amplificación por metotrexato. El tamaño esperado del producto de PCR de *Ggta1* (amplificado específicamente a partir de ADNc) es 324 pb.

DESCRIPCIÓN DETALLADA

15 La presencia de residuos de galactosa-alfa-1,3-galactosa terminal (Gal- α -Gal) en glucoproteínas recombinantes representa en general una modificación post-traducciona importante en gran parte debido al potencial inmunogénico del epítipo y (por consiguiente) la importancia reguladora en el desarrollo y la fabricación de productos biológicos. Sin embargo, no se ha informado que se produzca ni expresión de un gen *Ggta1* o producto génico, ni actividad enzimática, en ninguna línea de células CHO, históricamente la plataforma de expresión en mamíferos preferida para la producción comercial de productos biológicos terapéuticos. En realidad, la bibliografía científica sugiere lo opuesto, concretamente que las células CHO son evidentemente incapaces de producir N-glucanos con estructuras de Gal- α -Gal terminal debido a tanto una ausencia de una copia funcional del gen *Ggta1* en el genoma del hámster chino como a su expresión a niveles suficientes (Smith et al. (1990) J Biol Chem 265:6225-6234,4); Takeuchi et al. (1989) Proc Natl Acad Sci USA 86:7819-7822).

25 La presencia inesperada de residuos de Gal- α -Gal terminal en glucoproteínas producidas en células CHO en particular se describe en la solicitud internacional número de serie PCT/US2009/031678 (Ref. de publicación WO 2010/085251). La identificación y clonación de 1,3-glucosiltransferasa-1 (*Ggta1*) de hámster chino y de células CHO, como se describen en el presente documento, proporciona una valiosa herramienta para usar en los métodos descritos en el presente documento para detectar, medir y controlar los glucanos que contienen Gal- α -Gal en un producto biológico.

30 Definiciones

Como se usa en el presente documento, el término "molécula de ácido nucleico" incluye moléculas de ADN (por ejemplo, un ADNc o ADN genómico), moléculas de ARN (por ejemplo, un ARNm) y análogos de ADN o ARN. Puede sintetizarse un análogo de ADN o ARN a partir de análogos de nucleótidos. La molécula de ácido nucleico puede ser monocatenaria o bicatenaria.

35 El término "molécula de ácido nucleico aislada" o "molécula de ácido nucleico purificada" incluye moléculas de ácidos nucleicos que están separadas de otras moléculas de ácidos nucleicos presentes en la fuente natural del ácido nucleico. Por ejemplo, con respecto al ADN genómico, el término "aislado" incluye moléculas de ácidos nucleicos que están separadas del cromosoma con el que el ADN genómico está naturalmente asociado. Preferentemente, un ácido nucleico "aislado" está libre de secuencias que naturalmente flanquean el ácido nucleico (es decir, secuencias localizadas en los extremos 5' y/o 3' del ácido nucleico) en el ADN genómico del organismo del que se deriva el ácido nucleico. Por ejemplo, en diversas realizaciones, la molécula de ácido nucleico aislada puede contener menos de aproximadamente 5 kb, 4 kb, 3 kb, 2 kb, 1 kb, 0,5 kb o 0,1 kb de secuencias de nucleótidos de 5' y/o 3' que flanquean naturalmente la molécula de ácido nucleico en el ADN genómico de la célula de la que se deriva el ácido nucleico. Además, una molécula de ácido nucleico "aislada", tal como una molécula de ADNc, puede estar sustancialmente libre de otro material celular, o medio de cultivo cuando se produce por técnicas recombinantes, o sustancialmente libre de precursores químicos u otros productos químicos cuando se sintetiza químicamente.

50 Como se usa en el presente documento, el término "se hibrida bajo alta rigurosidad" describe condiciones para la hibridación y el lavado en la realización de reacciones de hibridación. Tales condiciones pueden encontrarse en Current Protocols in Molecular Biology, John Wiley & Sons, N.Y. (2007), Unidad 6.3. Se describen en esa referencia métodos acuosos y no acuosos y puede usarse cualquiera. Condiciones de hibridación específicas referidas en el presente documento son las siguientes: 1) Condiciones de hibridación de baja rigurosidad en 6X cloruro sódico/citrato de sodio (SSC) a aproximadamente 45 °C, seguido de dos lavados en 0,2X SSC, 0,1 % de SDS al menos a 50 °C (la temperatura de los lavados puede aumentarse a 55 °C para condiciones de baja rigurosidad); 2) condiciones de hibridación de rigurosidad media en 6X SSC a aproximadamente 45 °C, seguido de uno o más lavados en 0,2X SSC, 0,1 % de SDS a 60 °C; 3) condiciones de hibridación de alta rigurosidad en 6X SSC a aproximadamente 45 °C, seguido de uno o más lavados en 0,2X SSC, 0,1 % de SDS a 65 °C; y 4) condiciones de hibridación de rigurosidad muy alta son fosfato de sodio 0,5 M, 7 % de SDS a 65 °C, seguido de uno o más lavados a 0,2X SSC, 1 % de SDS a 65 °C.

Como se usa en el presente documento, una molécula de ácido nucleico "que existe de forma natural" se refiere a una molécula de ARN o de ADN que tiene una secuencia de nucleótidos que se produce en la naturaleza. Por ejemplo, una molécula de ácido nucleico que existe de forma natural puede codificar una proteína natural expresada por una célula no mutante.

- 5 Como se usa en el presente documento, los términos "gen" y "gen recombinante" se refieren a moléculas de ácidos nucleicos que incluyen al menos un marco de lectura abierto que codifica una proteína Ggta1. El gen puede opcionalmente además incluir secuencias no codificantes, por ejemplo, secuencias reguladoras e intrones.

10 Un polipéptido "aislado" o "purificado" o proteína está sustancialmente libre de material celular u otras proteínas contaminantes de la fuente de células o de tejido de la que se deriva la proteína, o sustancialmente libre de precursores químicos u otros productos químicos cuando se sintetiza químicamente. "Sustancialmente libre" significa que una preparación de proteína Ggta1 es al menos el 50 % de proteína Ggta1 en peso. En una realización preferida, la preparación de proteína Ggta1 tiene menos de aproximadamente el 30 %, 20 %, 10 % y más preferentemente el 5 % (en peso seco) de proteína no Ggta1 (también denominada en el presente documento una "proteína contaminante"), o de precursores químicos o productos químicos no Ggta1. Cuando la proteína Ggta1 o porción biológicamente activa de la misma se produce recombinantemente, también está preferentemente sustancialmente libre de medio de cultivo, es decir, el medio de cultivo representa menos de aproximadamente el 20 %, más preferentemente menos de aproximadamente el 10 %, y lo más preferentemente menos de aproximadamente el 5 % del volumen de la preparación de proteína.

20 Un "polipéptido Ggta1", como se usa el término en el presente documento, se refiere a un polipéptido que (1) muestra una identidad de secuencias global con uno o más de SEQ ID NO: 2, 4 o 7 de al menos aproximadamente el 90 %; (2) incluye al menos un elemento de secuencia característico encontrado en SEQ ID NO: 2, 4 o 7; y/o (3) comparte al menos una actividad biológica encontrada en un polipéptido de SEQ ID NO: 2, 4 o 7. En algunas realizaciones, un polipéptido Ggta1 muestra una identidad de secuencias global con uno o más de SEQ ID NO: 2, 4 o 7 de al menos aproximadamente el 90 %, al menos aproximadamente el 91 %, al menos aproximadamente el 92 %, al menos aproximadamente el 93 %, al menos aproximadamente el 94 %, al menos aproximadamente el 95 %, al menos aproximadamente el 96 %, al menos aproximadamente el 97 %, al menos aproximadamente el 98 %, al menos aproximadamente el 99 % o más. En algunas realizaciones, un polipéptido Ggta1 incluye al menos un elemento de secuencia característico de elemento de secuencia encontrado en SEQ ID NO: 2, 4 o 7 y también muestra una identidad de secuencias global con SEQ ID NO: 2, 4 o 7 que es al menos aproximadamente al menos aproximadamente el 90 %, al menos aproximadamente el 91 %, al menos aproximadamente el 92 %, al menos aproximadamente el 93 %, al menos aproximadamente el 94 %, al menos aproximadamente el 95 %, al menos aproximadamente el 96 %, al menos aproximadamente el 97 %, al menos aproximadamente el 98 %, al menos aproximadamente el 99 % o más. En algunas realizaciones, un polipéptido Ggta1 tiene una longitud que es al menos aproximadamente al menos aproximadamente el 75 %, al menos aproximadamente el 80 %, al menos aproximadamente el 85 %, al menos aproximadamente el 90 %, al menos aproximadamente el 91 %, al menos aproximadamente el 92 %, al menos aproximadamente el 93 %, al menos aproximadamente el 94 %, al menos aproximadamente el 95 %, al menos aproximadamente el 96 %, al menos aproximadamente el 97 %, al menos aproximadamente el 98 %, al menos aproximadamente el 99 % o más de la longitud de SEQ ID NO: 2, 4 o 7. En algunas realizaciones, un polipéptido Ggta1 es un fragmento de SEQ ID NO: 2, 4 o 7. En algunas realizaciones, un polipéptido Ggta1 muestra al menos aproximadamente el 90 %, al menos aproximadamente el 91 %, al menos aproximadamente el 92 %, al menos aproximadamente el 93 %, al menos aproximadamente el 94 %, al menos aproximadamente el 95 %, al menos aproximadamente el 96 %, al menos aproximadamente el 97 %, al menos aproximadamente el 98 %, al menos aproximadamente el 99 % o más identidad con un fragmento de SEQ ID NO: 2, 4 o 7; en algunos tales realizaciones, el fragmento tiene al menos aproximadamente 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 150, 200, 250, 300, 350, o más aminoácidos de longitud.

50 Una "sustitución de aminoácidos conservativa" es una en la que el resto de aminoácido se sustituye con un resto de aminoácido que tiene una cadena lateral similar. Familias de restos de aminoácidos que tienen cadenas laterales similares se han definido en la materia. Estas familias incluyen aminoácidos con cadenas laterales básicas (por ejemplo, lisina, arginina, histidina), cadenas laterales ácidas (por ejemplo, ácido aspártico, ácido glutámico), cadenas laterales polares sin carga (por ejemplo, glicina, asparagina, glutamina, serina, treonina, tirosina, cisteína), cadenas laterales no polares (por ejemplo, alanina, valina, leucina, isoleucina, prolina, fenilalanina, metionina, triptófano), cadenas laterales beta-ramificadas (por ejemplo, treonina, valina, isoleucina) y cadenas laterales aromáticas (por ejemplo, tirosina, fenilalanina, triptófano, histidina). Así, un resto de aminoácido no esencial predicho en una proteína Ggta1 se sustituye preferentemente con otro resto de aminoácido de la misma familia de cadena lateral. Alternativamente, en otra realización, pueden introducirse mutaciones al azar a lo largo de toda o parte de una secuencia codificante de Ggta1, tal como por mutagénesis por saturación, y los mutantes resultantes pueden cribarse para la actividad de Ggta1 para identificar mutantes que retienen la actividad. Tras la mutagénesis, la proteína codificada puede expresarse recombinantemente y puede determinarse la actividad de la proteína.

60 "Porcentaje (%) de identidad de secuencias" con respecto a las secuencias identificadas en el presente documento se define como el porcentaje de restos de aminoácidos o nucleótidos en una secuencia candidata que son idénticos a los restos de aminoácidos o nucleótidos en la secuencia de referencia, después de alinear las secuencias e

introducir huecos, si fuera necesario, para lograr el máximo porcentaje de identidad de secuencias (por ejemplo, los huecos pueden introducirse en uno o ambos de una primera y una segunda secuencia de aminoácidos o de ácidos nucleicos para el alineamiento óptimo y pueden ignorarse secuencias no homólogas para fines de comparación). El alineamiento para fines de determinación del porcentaje de identidad de secuencias puede lograrse de diversas formas que están dentro de la experiencia en la materia, por ejemplo, usando software informático públicamente disponible tal como el software BLAST, ALIGN o Megalign (DNASTAR). Aquellos expertos en la materia pueden determinar parámetros apropiados para medir el alineamiento, que incluyen cualquier algoritmo necesario para lograr el máximo alineamiento sobre la longitud completa de las secuencias que se comparan. En una realización, la longitud de una secuencia de referencia alineada para fines de comparación es al menos el 30 %, por ejemplo, al menos el 40 %, por ejemplo, al menos el 50 %, 60 %, 70 %, 80 %, 90 % o el 100 % de la longitud de la secuencia de referencia. Entonces se comparan los restos de aminoácidos o nucleótidos en posiciones de aminoácidos o posiciones de nucleótidos correspondientes. Cuando una posición en la primera secuencia es ocupada por el mismo resto de aminoácido o nucleótido que la posición correspondiente en la segunda secuencia, entonces las moléculas son idénticas en esa posición.

Un "elemento de secuencia característico", como el término se usa en el presente documento, se refiere a un conjunto de aminoácidos cuya identidad y posiciones relativas la una con respecto a la otra confieren a un polipéptido particular una o más actividades o características de interés. En algunas realizaciones, un elemento de secuencia característico es uno que es único para el polipéptido particular porque no se encuentra en polipéptidos conocidos que carecen de la actividad o característica. En algunas realizaciones, un elemento de secuencia característico comprende al menos 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30 o más aminoácidos. En algunas realizaciones, algunos o todos de los aminoácidos que participan en un elemento de secuencia característico son contiguos entre sí; en algunas realizaciones, ciertos residuos dentro de un elemento de secuencia característico pueden estar separados de uno o varios de otros residuos en orden lineal. En algunas realizaciones, un elemento de secuencia característico comprende residuos que están localizados cerca entre sí en el espacio tridimensional cuando se pliega una cadena de polipéptidos.

Una "preparación de células", como se usa en el presente documento, se refiere a una preparación de células *in vitro*. En el caso de células de organismos multicelulares (por ejemplo, plantas y animales), una preparación de células purificada es un subconjunto de células obtenido del organismo, no el organismo intacto entero. En el caso de microorganismos unicelulares (por ejemplo, células cultivadas y células microbianas), consiste en una preparación de al menos el 10 % y más preferentemente del 50 % de células del sujeto.

Diversos aspectos y realizaciones de la presente divulgación se describen en más detalle a continuación.

Moléculas de ácidos nucleicos de Ggta1

En un aspecto, la divulgación proporciona moléculas de ácidos nucleicos de Ggta1 aisladas o purificadas como se describe en el presente documento. Una molécula de ácido nucleico proporcionada puede codificar el polipéptido Ggta1 descrito en el presente documento, por ejemplo, una proteína Ggta1 de longitud completa o un fragmento activo de la misma (por ejemplo, un fragmento catalíticamente activo de la misma). También están incluidas moléculas de ácidos nucleicos de Ggta1 o fragmentos de las mismas que pueden no codificar una proteína funcional, pero pueden usarse, por ejemplo, para construir un vector para dirigir un gen Ggta1 endógeno de hámster chino o de células CHO, por ejemplo, para silenciar o modificar la secuencia de un gen Ggta1 endógeno.

En algunas realizaciones, las moléculas de ácidos nucleicos de Ggta1 proporcionadas se derivan de hámster chino o de células CHO. En algunas realizaciones, las moléculas de ácidos nucleicos de Ggta1 proporcionadas tienen una secuencia de nucleótidos que es idéntica a la de una molécula de ácido nucleico de Ggta1 derivada de hámster chino o de células CHO. En algunas realizaciones, las moléculas de ácidos nucleicos de Ggta1 proporcionadas codifican un polipéptido Ggta1 que se deriva de hámster chino o de células CHO.

En una realización, una molécula de ácido nucleico aislada de la divulgación tiene una secuencia de nucleótidos de SEQ ID NO: 1, 3, 5 o 6, o una porción de cualquiera de estas secuencias de nucleótidos. En algunas realizaciones, la molécula de ácido nucleico incluye secuencia codificante solo. En otras realizaciones, la molécula de ácido nucleico incluye secuencia no codificante, tal como secuencias no traducidas de 5' o secuencia de intrón. Están incluidas tanto secuencias sentido como antisentido.

También están incluidos oligonucleótidos o fragmentos de las moléculas de ácidos nucleicos que son adecuados para su uso como sondas, marcas o cebadores, por ejemplo, para detectar y/o cuantificar la expresión de Ggta1. Tales oligonucleótidos o fragmentos pueden tener al menos aproximadamente el 10, 12 o 15, aproximadamente 20 o 25, o aproximadamente 30, 35, 40, 45, 50, 55, 60, 65 o 75 nucleótidos consecutivos de SEQ ID NO: 1, 3, 5 o 6. En ciertas realizaciones, los oligonucleótidos tienen menos de aproximadamente 500, 400, 300, 200, 150, 120 o 100 nucleótidos de longitud. En una realización, una sonda o cebador se une a un soporte sólido, por ejemplo, un soporte sólido descrito en el presente documento. En otra realización se proporciona un conjunto de cebadores, por ejemplo, cebadores adecuados para su uso en una reacción de PCR, que pueden usarse para amplificar una región seleccionada de una secuencia de Ggta1, por ejemplo, un dominio, región, sitio u otra secuencia descrita en el presente documento. Los cebadores pueden tener al menos 5, 10 o 50 pares de bases de longitud y menos de 100,

o menos de 200, pares de bases en longitud. Los cebadores deben ser idénticos, o diferenciarse por una base de una secuencia desvelada en el presente documento o de una variante que existe de forma natural. Por ejemplo, cebadores adecuados para amplificar toda o una porción de una secuencia de Ggta1 descrita en el presente documento se desvelan en la Tabla 2 y la Tabla 4. En una realización, un kit de cebadores incluye un cebador directo que se hibrida con una hebra codificante y un cebador inverso que se hibrida con la hebra no codificante de SEQ ID NO: 1, 3, 5 o 6.

Pueden marcarse los oligonucleótidos del ácido nucleico, por ejemplo, sondas, marcas o cebadores descritos en el presente documento. Normalmente, tales marcas son quimioluminiscentes, fluorescentes, radiactivas o colorimétricas.

10 Un ácido nucleico de la divulgación puede producirse recombinantemente, o producirse sintéticamente.

Variantes de ácidos nucleicos de Ggta1:

La divulgación engloba moléculas de ácidos nucleicos que se diferencian de las secuencias de nucleótidos desveladas en el presente documento como SEQ ID NO: 1, 3, 5 o 6. Por ejemplo, una molécula de ácido nucleico aislada de la divulgación puede tener una secuencia relacionada con SEQ ID NO: 1, 3, 5 o 6, o con una porción de cualquiera de estas secuencias de nucleótidos, por ejemplo, relacionada por un nivel especificado de identidad de secuencias o capacidad para hibridarse en condiciones de alta rigurosidad. Un ácido nucleico de la divulgación puede tener una secuencia que existe de forma natural (por ejemplo, la secuencia de una variante alélica o mutante que existe de forma natural), o puede tener una secuencia que no existe de forma natural. En una realización, las diferencias pueden ser debidas, por ejemplo, a la degeneración del código genético (y producen un ácido nucleico que codifica las mismas proteínas Ggta1 que aquellas codificadas por la secuencia de nucleótidos desvelada en el presente documento). Los ácidos nucleicos de la divulgación pueden elegirse para tener codones que son preferidos, o no preferidos, para un sistema de expresión particular. Por ejemplo, el ácido nucleico puede ser uno en el que al menos un codón, por ejemplo, al menos el 10 %, o el 20 % de los codones, ha sido alterado de forma que la secuencia se optimice para la expresión en un sistema particular, por ejemplo, células de *E. coli*, de levadura, humanas, de insecto, o CHO.

Las variantes de ácido nucleico pueden existir de forma natural, tales como variantes alélicas (mismo locus), homólogos (diferente locus), o pueden no existir de forma natural. Las variantes que no existen de forma natural pueden prepararse por técnicas de mutagénesis, que incluyen aquellas aplicadas a polinucleótidos, células u organismos. Las variantes pueden contener sustituciones, deleciones, inversiones e inserciones de nucleótidos. Puede producirse variación en cualquiera o ambas de las regiones codificantes y no codificantes. Las variaciones pueden producir tanto sustituciones de aminoácidos conservativas como no conservativas (como se compara en el producto codificado). Se conocen en la técnica técnicas de modificación y mutagénesis de ácidos nucleicos.

Variantes alélicas de Ggta1, por ejemplo, Ggta1 de CHO, incluyen tanto proteínas funcionales como no funcionales. Variantes alélicas funcionales son secuencias de nucleótidos que existen de forma natural que codifican variantes de la proteína Ggta1 que mantienen la capacidad para proporcionar la función enzimática. Las variantes alélicas funcionales normalmente contendrán la sustitución conservativa de uno o más aminoácidos de SEQ ID NO: 2, 4 o 7, o sustitución, deleción o inserción de restos no críticos en regiones de la proteína no críticas. Variantes alélicas no funcionales son variantes de secuencias de nucleótidos que existen de forma natural que codifican un polipéptido Ggta1 que no preserva la función enzimática. Las variantes alélicas no funcionales normalmente contendrán una sustitución no conservativa, una deleción, o inserción, o truncación prematura de la secuencia de aminoácidos de SEQ ID NO: 2, 4 o 7, o una sustitución, inserción, o deleción en restos críticos o regiones críticas de la proteína.

Moléculas antisentido de Ggta1:

También se desvelan moléculas de ácido nucleico aisladas que son antisentido para una Ggta1 de hámster chino o de células CHO descrita en el presente documento. Las clases principales de agentes antisentido son oligonucleótidos antisentido (ODNs), ribozimas, ADNzimas e interferencia por ARN (iARN). Tales moléculas pueden usarse en métodos para inhibir la actividad de Ggta1, por ejemplo, en células CHO. Un ácido nucleico "antisentido" tiene una secuencia de nucleótidos que es complementaria a un ácido nucleico diana. Un ácido nucleico antisentido puede diseñarse de forma que sea complementario a la región codificante entera de un ARNm de Ggta1, pero más preferentemente es un oligonucleótido que es antisentido a solo una porción de la región codificante o no codificante de un ARNm de Ggta1. Por ejemplo, el oligonucleótido antisentido puede ser complementario a la región que rodea el sitio de inicio de la traducción del ARNm de Ggta1, por ejemplo, entre las regiones -10 y +10 de la secuencia de nucleótidos del gen diana de interés. Un oligonucleótido antisentido puede tener, por ejemplo, aproximadamente 7, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, o más nucleótidos de longitud. Una molécula antisentido inhibirá la producción de Ggta1 descrita en el presente documento

Un ácido nucleico antisentido de la divulgación puede construirse usando síntesis química y reacciones de ligación enzimática usando procedimientos conocidos en la técnica. Por ejemplo, un ácido nucleico antisentido (por ejemplo, un oligonucleótido antisentido) puede sintetizarse químicamente usando nucleótidos que existen de forma natural o nucleótidos modificados de diversas maneras diseñados para aumentar la estabilidad biológica de las moléculas o

para aumentar la estabilidad física del dúplex formado entre los ácidos nucleicos antisentido y sentido, por ejemplo, pueden usarse derivados de fosforotioato y nucleótidos sustituidos con acridina. El ácido nucleico antisentido también puede producirse biológicamente usando un vector de expresión en el que un ácido nucleico se ha subclonado en una orientación antisentido. Las moléculas de ácidos nucleicos antisentido pueden administrarse a células usando vectores descritos en el presente documento. Para lograr concentraciones intracelulares suficientes de las moléculas antisentido, las construcciones de vector en las que la molécula de ácido nucleico antisentido se coloca pueden estar bajo el control de un promotor pol II o pol III fuerte.

Se conocen en la técnica métodos de preparación y uso de las moléculas antisentido para modular actividades biológicas, véanse, por ejemplo: Pan and Clawson, Antisense applications for biological control (2006) J. Cell Biochem. 98(1):14-35; Sioud and Iversen, Ribozymes, DNAzymes and small interfering RNAs as therapeutics (2005) Curr Drug Targets 6(6):647-53; Bhindi et al., Brothers in arms: DNA enzymes, short interfering RNA, and the emerging wave of small-molecule nucleic acid-based gene-silencing strategies (2007) Am J Pathol. 171(4):1079-88.

Polipéptidos Ggta1

En otro aspecto, la divulgación presenta polipéptidos Ggta1 aislados o proteínas (o fragmentos biológicamente activos de los mismos) como se describen en el presente documento. Normalmente, un fragmento biológicamente activo de una proteína Ggta1 tiene una o más de las siguientes características: tiene la capacidad de catalizar la síntesis de un glucano Gal- α 1,3-Gal; utiliza UDP-Gal para transferir un resto de galactosa sobre un glucano enlazado en N existente que contiene Gal para generar una estructura de Gal- α -Gal y UDP; utiliza UDP-Gal para transferir un resto de galactosa sobre un glucano enlazado en O existente que contiene Gal para generar una estructura de Gal- α -Gal y UDP; utiliza UDP-Gal para transferir un resto de galactosa sobre un glucolípido existente que contiene Gal para generar una estructura de Gal- α -Gal y UDP; se une a UDP-Gal y un glucano que termina en galactosa; se une a estructuras que contienen Gal- α -Gal; hidroliza UDP-Gal a Gal; está confinado en el compartimento de Golgi si se transfecta y se expresa en una célula eucariota; se une a un anticuerpo anti-Ggta1 específico de hámster chino o de células CHO; puede utilizar análogos de UDP-Gal químicamente modificados (por ejemplo, tioazúcar, trinitrofenilo, 2-desoxi, 2-azido, radioquímicamente marcados tales como 3H, C14, 33P, 32P) para transferir un resto de galactosa sobre un glucano enlazado en N u O existente.

Los polipéptidos o proteínas Ggta1 (o fragmentos de los mismos) pueden aislarse de células o fuentes de tejido usando técnicas de purificación de proteínas estándar. En algunas realizaciones, los polipéptidos o proteínas Ggta1 se aíslan de hámster chino o de células CHO. Los polipéptidos o proteínas Ggta1 (o fragmentos de los mismos) pueden producirse por técnicas de ADN recombinante o sintetizarse químicamente.

Los polipéptidos de la divulgación incluyen aquellos que surgen como resultado de la existencia de múltiples genes, eventos de transcripción alternativos, eventos de corte y empalme de ARN alternativos, y eventos traduccionales y post-traduccionales alternativos. El polipéptido puede expresarse en sistemas, por ejemplo, células cultivadas, que producen sustancialmente las mismas modificaciones post-traduccionales presentes cuando el polipéptido se expresa en una célula nativa, o en sistemas que producen la alteración u omisión de modificaciones post-traduccionales, por ejemplo, glucosilación o escisión, presentes cuando se expresa en una célula nativa.

En algunas realizaciones, un polipéptido Ggta1 de la divulgación tiene la secuencia de SEQ ID NO: 2, 4 o 7 o un fragmento funcional de la misma. Otras realizaciones incluyen una proteína que contiene al menos un resto de aminoácido diferente de SEQ ID NO: 2, 4 o 7, pero no más del 10 %. Las diferencias pueden ser en restos de aminoácidos que no son esenciales para la actividad. En algunas realizaciones, las diferencias son sustituciones conservativas o diferencias o cambios en un resto no esencial. También están incluidas secuencias de polipéptidos que están altamente relacionadas con SEQ ID NO: 2, 4 o 7 por un porcentaje de identidad de secuencias especificado.

La divulgación incluye proteínas quiméricas o de fusión de Ggta1. Como se usa en el presente documento, una "proteína quimérica" o "proteína de fusión" de Ggta1 incluye un polipéptido Ggta1 asociado a un polipéptido no Ggta1. El polipéptido no Ggta1 puede fusionarse con el extremo N o extremo C del polipéptido Ggta1, o puede fusionarse internamente. En un ejemplo, la proteína de fusión puede incluir un resto que tiene una alta afinidad por un ligando. Por ejemplo, la proteína de fusión puede ser una proteína de fusión GST-Ggta1 en la que las secuencias de Ggta1 están fusionadas con el extremo C de las secuencias de GST. Tales proteínas de fusión pueden facilitar la purificación de Ggta1 recombinante. Están comercialmente disponibles vectores de expresión que codifican un resto de fusión (por ejemplo, un polipéptido GST). Un ácido nucleico que codifica Ggta1 puede clonarse en un vector de expresión tal de forma que el resto de fusión se enlace en marco a la proteína Ggta1.

Alternativamente, la proteína de fusión puede ser una proteína Ggta1 que contiene una secuencia señal heteróloga en su extremo N. En ciertas células huésped (por ejemplo, células huésped de mamífero), la expresión y/o secreción de Ggta1 pueden aumentarse mediante el uso de una secuencia señal heteróloga.

Las proteínas de fusión pueden incluir toda o una parte de una proteína del suero, por ejemplo, una región constante de IgG, o albúmina de suero humano.

Variantes de polipéptidos Ggta1:

En otro aspecto, la divulgación también presenta una variante de un polipéptido Ggta1, por ejemplo, que funciona como agonista (mimético) o como antagonista. Pueden generarse variantes de las proteínas Ggta1 por mutagénesis, por ejemplo, mutación puntual discreta, la inserción o delección de secuencias o la truncación de una proteína Ggta1.

5 Un agonista de las proteínas Ggta1 puede retener sustancialmente las mismas, o un subconjunto de, actividades biológicas de la forma que existe de forma natural de una proteína Ggta1. Un antagonista de una proteína Ggta1 puede inhibir una o más de las actividades de la forma que existe de forma natural de la proteína Ggta1, por ejemplo, modulando competitivamente una actividad mediada por Ggta1 de una proteína Ggta1.

En algunas realizaciones, una variante de un polipéptido Ggta1 comparte la secuencia de aminoácidos de un polipéptido Ggta1 "original" (o polipéptido Ggta1 de variante), pero incluye una o más modificaciones covalentes en comparación con el polipéptido original. Por ejemplo, en algunas realizaciones, las variantes están PEGiladas, glucosiladas, fosforiladas, etc.

10

Pueden identificarse variantes de una proteína Ggta1 cribando bibliotecas combinatorias de mutantes, por ejemplo, mutantes de truncación, de una proteína Ggta1 para actividad de agonista o de antagonista.

15 Pueden producirse variantes de polipéptidos Ggta1, como polipéptidos Ggta1, como proteínas de fusión.

Anticuerpos contra Ggta1 y producción de anticuerpos

En otro aspecto, la presente divulgación proporciona un anticuerpo que se une específicamente a un polipéptido Ggta1 (y/o a una variante de polipéptido Ggta1) como se describe en el presente documento; la presente divulgación también proporciona métodos de producción de tal anticuerpo. El término "anticuerpo" se refiere a una proteína que incluye al menos un dominio variable de inmunoglobulina o secuencia del dominio variable de inmunoglobulina capaz de unirse a antígeno de Ggta1 de hámster chino o CHO. Por ejemplo, un anticuerpo puede incluir una región variable de la cadena pesada (H) (abreviada en el presente documento VH) y una región variable de la cadena ligera (L) (abreviada en el presente documento VL). En otro ejemplo, un anticuerpo incluye dos regiones variables de la cadena pesada (H) y dos regiones variables de la cadena ligera (L). Como tal, el término "anticuerpo" engloba anticuerpos policlonales, mono-específicos, monoclonales, monovalentes, quiméricos, humanizados, humanos, biespecíficos, y heteroconjugados, además de fragmentos de unión al antígeno de estos.

20

25

El término "fragmento de unión al antígeno" de un anticuerpo de longitud completa, como se usa en el presente documento, se refiere a uno o más fragmentos de un anticuerpo de longitud completa que retienen la capacidad para unirse específicamente a una diana de interés. Ejemplos de tales fragmentos incluyen (i) un fragmento Fab, un fragmento monovalente que consiste en los dominios VL, VH, CL y CH1; (ii) un fragmento F(ab')₂, un fragmento bivalente que incluye dos fragmentos Fab enlazados por un puente de disulfuro en la región bisagra; (iii) un fragmento Fd que consiste en los dominios VH y CH1; (iv) un fragmento Fv que consiste en los dominios VL y VH de un solo brazo de un anticuerpo, (v) un fragmento dAb, que consiste en un dominio VH; y (vi) una región determinante de la complementariedad (CDR) aislada que retiene funcionalidad.

30

Pueden usarse proteínas Ggta1 aisladas (incluyendo variantes), o fragmentos antigénicos de las mismas, como antígenos para producir, cribar o seleccionar tales anticuerpos. Pueden obtenerse anticuerpos y fragmentos de unión al antígeno de los mismos usando cualquier técnica apropiada, que incluye técnicas de hibridomas convencionales, técnicas recombinantes, métodos combinatorios, técnicas de presentación en fagos, y otras conocidas para aquellos expertos en la materia. Por ejemplo, véanse *Antibody Engineering: Methods and Protocols*, Benny K.C. Lo, Ed. (Humana Press, 2003); *Making and Using Antibodies: A Practical Handbook*, Gary Howard and Matthew Kaser, Eds. (CRC, 2006); *Antibody Phage Display: Methods and Protocols*, Philippa O'Brien and Robert Aitken, Eds. (Humana Press, 2001); *Monoclonal Antibodies: Methods and Protocols*, Maher Albitar, Ed. (Humana Press, 2007).

35

40

Un anticuerpo de la divulgación puede acoplarse a un segundo resto funcional, por ejemplo, el anticuerpo puede acoplarse a una marca, por ejemplo, una marca fluorescente, marca radiactiva, agente de obtención de imágenes.

En algunas realizaciones, puede usarse un anticuerpo anti-Ggta1 para detectar un polipéptido Ggta1 (por ejemplo, en un lisado celular o sobrenadante de células) con el fin de detectar y opcionalmente medir la presencia, nivel, abundancia o patrón de expresión de la proteína. La detección puede facilitarse acoplando (es decir, enlazando físicamente) el anticuerpo con una sustancia detectable (es decir, marcado del anticuerpo). Ejemplos de sustancias detectables incluyen diversas enzimas, grupos prostéticos, materiales fluorescentes, materiales luminiscentes, materiales bioluminiscentes y materiales radiactivos. Ejemplos de enzimas adecuadas incluyen peroxidasa de rábano picante, fosfatasa alcalina, beta-galactosidasa o acetilcolinesterasa; ejemplos de complejos de grupos prostéticos adecuados incluyen estreptavidina/biotina y avidina/biotina; ejemplos de materiales fluorescentes adecuados incluyen umbeliferona, fluoresceína, isotiocianato de fluoresceína, rodamina, diclorotriazinilamina fluoresceína, cloruro de dansilo o ficoeritrina; un ejemplo de un material luminiscente incluye luminol; ejemplos de materiales bioluminiscentes incluyen luciferasa, luciferina y aequorina.

45

50

55

Vectores de expresión recombinantes

En otro aspecto, la divulgación incluye vectores (por ejemplo, vectores de expresión, vectores inactivados, o vectores que conducen secuencias antisentido) que contienen un ácido nucleico descrito en el presente documento. Como se usa en el presente documento, el término "vector" se refiere a una molécula de ácido nucleico capaz de transportar otro ácido nucleico con el que se ha enlazado y puede incluir un plásmido, cósmido o vector viral. El vector puede ser capaz de replicación autónoma o puede integrarse en un ADN de huésped.

Un vector puede incluir un ácido nucleico de Ggta1 de hámster chino o de células CHO en una forma adecuada para la expresión del ácido nucleico en una célula huésped. Preferentemente, el vector de expresión recombinante incluye una o más secuencias reguladoras operativamente enlazadas a la secuencia de ácidos nucleicos que va a expresarse. El término "secuencia reguladora" incluye promotores, potenciadores y otros elementos de control de la expresión (por ejemplo, señales de poliadenilación). Secuencias reguladoras incluyen aquellas que dirigen la expresión constitutiva de una secuencia de nucleótidos, además de secuencias reguladoras y/o inducibles específicas de tejido. El diseño del vector de expresión puede depender de factores tales como la elección de la célula huésped que va a transformarse, el nivel de expresión de la proteína deseada, y similares. Los vectores de expresión desvelados en el presente documento pueden introducirse en células huésped para así producir proteínas o polipéptidos, que incluyen proteínas o polipéptidos de fusión, codificados por ácidos nucleicos como se describe en el presente documento (por ejemplo, proteínas Ggta1, formas mutantes de proteínas Ggta1, proteínas de fusión, y similares).

Los vectores de expresión recombinantes de la divulgación pueden diseñarse para la expresión de polipéptidos Ggta1 en células procariotas o eucariotas. Por ejemplo, los polipéptidos de la divulgación pueden expresarse en *E. coli*, células de insecto (por ejemplo, usando vectores de expresión de baculovirus), células de levadura o células de mamífero, preferentemente células CHO. Cuando se usa en células de mamífero, las funciones de control del vector de expresión pueden proporcionarse por elementos reguladores virales. Por ejemplo, promotores comúnmente usados se derivan de polioma, adenovirus 2, citomegalovirus y virus 40 simio.

La expresión de proteínas en procariotas es casi siempre llevada a cabo en *E. coli* con vectores que contienen promotores constitutivos o inducibles que dirigen la expresión de tanto proteínas de fusión como no de fusión. Los vectores de fusión añaden varios aminoácidos a una proteína codificada en ellos, normalmente al extremo amino de la proteína recombinante. Tales vectores de fusión normalmente cumplen tres fines: 1) aumentar la expresión de proteína recombinante; 2) aumentar la solubilidad de la proteína recombinante; y 3) ayudar en la purificación de la proteína recombinante actuando de ligando en la purificación por afinidad. Frecuentemente, un sitio de escisión proteolítica se introduce en la intersección del resto de fusión y la proteína recombinante para permitir la separación de la proteína recombinante del resto de fusión posterior a la purificación de la proteína de fusión. Tales enzimas, y sus secuencias de reconocimiento relacionadas, incluyen factor Xa, trombina y enterocinasa. Vectores de expresión de fusión típicos incluyen pGEX (Farmacia Biotech Inc), pMAL (New England Biolabs, Beverly, Masa.) y pRIT5 (Farmacia, Piscataway, NJ) que fusionan glutatión S-transferasa (GST), proteína de unión a maltosa E, o proteína A, respectivamente, con la proteína recombinante diana. Pueden usarse proteínas de fusión purificadas en ensayos de actividad de Ggta1 (por ejemplo, ensayos directos o ensayos competitivos), o para generar anticuerpos específicos para proteínas Ggta1.

En otra realización, un promotor es un promotor inducible, por ejemplo, un promotor regulado por una hormona esteroidea, por una hormona polipeptídica (por ejemplo, por medio de una vía de transducción de señales), o por un polipéptido heterólogo (por ejemplo, los sistemas inducibles por tetraciclina, "Tet-On" y "Tet-Off" de Clontech Inc., CA).

La divulgación proporciona además un vector de expresión recombinante que comprende una molécula de ADN de la divulgación clonado en el vector de expresión en una orientación antisentido. Pueden elegirse secuencias reguladoras (por ejemplo, promotores y/o potenciadores virales) operativamente enlazadas a un ácido nucleico clonado en la orientación antisentido que dirigen la expresión constitutiva, específica de tejido o específica de tipo de célula de ARN antisentido en una variedad de tipos de células. El vector de expresión antisentido puede estar en forma de un plásmido recombinante, fagémido o virus atenuado.

Otro aspecto de la divulgación proporciona un vector que contiene una molécula de ácido nucleico descrita en el presente documento, por ejemplo, una molécula de ácido nucleico de Ggta1 de hámster chino o de células CHO descrita en el presente documento, configurada para permitir que se recombine homológamente en un sitio específico de un genoma de la célula huésped, por ejemplo, configurada para modificar, alterar o inactivar un gen Ggta1 endógeno en la célula huésped.

Puede introducirse ADN de vector en células huésped mediante técnicas de transformación o transfección convencionales. Como se usa en el presente documento, los términos "transformación" y "transfección" pretenden referirse a una variedad de técnicas reconocidas en la técnica para introducir ácido nucleico extraño (por ejemplo, ADN) en una célula huésped, que incluye co-precipitación con fosfato de calcio o cloruro de calcio, transfección mediada por DEAE-dextrano, lipofección, o electroporación.

Alternativamente, un vector de expresión recombinante puede transcribirse y traducirse *in vitro*, por ejemplo usando secuencias reguladoras del promotor T7 y T7 polimerasa.

Métodos de preparación y uso de vectores que incluyen los ácidos nucleicos descritos en el presente documento se conocen en la técnica. Por ejemplo, se proporcionan métodos en Current Protocols in Molecular Biology (2007, John Wiley and Sons, Inc., Print ISSN: 1934-3639).

Células huésped/células genéticamente manipuladas

Una célula huésped (por ejemplo, una célula huésped que contiene cualquier vector descrito en el presente documento) puede ser cualquier célula procariota o eucariota. Por ejemplo, una proteína Ggta1 puede expresarse en células bacterianas (tales como *E. coli*), células vegetales, células de insecto, células de levadura o de mamífero (tales como células CHO). Otras células huésped adecuadas son conocidas para aquellos expertos en la materia.

Una célula aislada de la divulgación puede ser una célula genéticamente manipulada para expresar niveles elevados de Ggta1 de hámster chino o de células CHO con respecto a la célula parental. Por consiguiente, la divulgación también proporciona métodos de producción de una proteína Ggta1 usando las células huésped de la divulgación. En una realización, el método incluye cultivar la célula huésped de la divulgación (en la que se ha introducido un vector de expresión recombinante que codifica una proteína Ggta1) en un medio adecuado de forma que se produzca una proteína Ggta1. En otra realización, el método incluye además aislar una proteína Ggta1 del medio o la célula huésped.

En otro aspecto, la divulgación presenta una célula o preparación purificada de células que incluyen un transgén de Ggta1 de hámster chino o de células CHO, o que de otro modo expresan erróneamente Ggta1. Por ejemplo, una célula aislada de la divulgación es una célula (por ejemplo, una célula CHO) genéticamente manipulada para expresar niveles más bajos de Ggta1 que la célula parental, por ejemplo, la célula es una célula CHO inactivada o silenciada para Ggta1. La generación de CHO inactivadas o silenciadas para reducir o inhibir la expresión de proteínas endógenas es una técnica conocida. Por ejemplo, recientemente se describió una célula inactivada para CHO en Yamane-Ohnuki et al. (2004) Biotechnol. Bioeng 87:614-622.

Una preparación de células puede consistir en células humanas o no humanas aisladas, por ejemplo, células de roedor, por ejemplo, células derivadas de hámster, de ratón o de rata, células de conejo o células de cerdo o líneas celulares, por ejemplo, una línea de células CHO. Células CHO útiles como células huésped de la divulgación incluyen células de cualquier cepa de CHO, que incluye CHO K1 (ATCC CCL-61), CHO pro3-, CHO DG44, CHO-S, CHO P12 o la línea de células CHO dhfr- DUK-BII (Chassin et al., PNAS 77, 1980, 4216-4220).

Se conocen en la técnica métodos de preparación y uso de células huésped de la divulgación, y de preparación de glucoproteínas terapéuticas en tales células huésped. Por ejemplo, se proporcionan métodos en Current Protocols in Cell Biology (2007, John Wiley and Sons, Inc., Print ISSN: 1934-2500); Current Protocols in Protein Science (2007, John Wiley and Sons, Inc., Print ISSN: 1934-3655); Wurm, Production of recombinant protein therapeutics in cultivated mammalian cells (2004) Nature Biotech. 22:1393-1398; Therapeutic Proteins: Methods and Protocols, Smales and James, eds. (2005, Humana Press, ISBN-10: 1588293904).

Métodos de detección de la expresión o actividad de Ggta1

La capacidad para monitorizar la expresión y/o actividad de α -1,3-glucosiltransferasa-1 (Ggta1) responsable de la adición enzimática del epítipo de Gal-alfa-Gal sobre N-glucanos es una valiosa herramienta para predecir las posibilidades de la existencia de epítopos en una línea celular particular o población clonal. Por consiguiente, la divulgación también presenta métodos de evaluación de la expresión de Ggta1 en células CHO. El método incluye proporcionar una muestra de una población de células CHO, y poner en contacto la muestra con un ácido nucleico de Ggta1 o anticuerpo descrito en el presente documento.

Estos métodos pueden usarse para trazar un perfil la expresión de Ggta1 relativa en células de referencia o fenotipo clonal posiblemente diferente. Asimismo, pueden usarse para evaluar cambios en la expresión génica cuando tales células se cultivan bajo diferentes condiciones de bioproceso tales como formulación y/o escala de medios. Por ejemplo, tales métodos pueden usarse para evaluar la actividad de Ggta1 o presencia de Gal-alfa-Gal o nivel en una población de células CHO que expresa o produce una glucoproteína terapéutica. En algunas realizaciones, la población de células CHO es una población en un biorreactor, por ejemplo, un biorreactor comercial. Tales métodos pueden usarse, por ejemplo, para monitorizar la expresión de Ggta1 durante un proceso de fabricación. En otras realizaciones, la población de células CHO es una de una pluralidad de poblaciones de células CHO que se criban para actividad de Ggta1. La pluralidad de células CHO que se criban puede diferenciarse en cualquier número de formas, por ejemplo, puede diferenciarse en contextos genéticos, condiciones de crecimiento, etc. En tales métodos de cribado, células CHO pueden seleccionarse basándose en el nivel de actividad de Ggta1, por ejemplo, seleccionarse como huéspedes adecuados para producir una glucoproteína terapéutica particular. La actividad de Ggta1 en las células CHO que se evalúan puede compararse con un nivel de referencia de actividad de Ggta1, por ejemplo, un nivel de control, un nivel pre-especificado, o un nivel correspondiente al nivel en una segunda población de células CHO.

Métodos de detección basados en ácidos nucleicos engloban los ensayos de hibridación o amplificación que incluyen, pero no se limitan a, análisis Southern o Northern, análisis de reacción en cadena de la polimerasa (por ejemplo, PCR cuantitativa), análisis SAGE, matrices de sondas o matrices de oligonucleótidos.

5 Un método para la detección de la expresión de Ggta1 implica poner en contacto una muestra de ARNm aislado de una población de células CHO con una molécula de ácido nucleico (por ejemplo, sonda) descrita en el presente documento que puede hibridarse con el ARNm codificado por el gen Ggta1. Otro método para la detección de la expresión de Ggta1 implica poner en contacto una muestra de ARNm o ADNc de una población de células CHO con un par de cebadores descrito en el presente documento para amplificar una secuencia de Ggta1. El nivel de ARNm de Ggta1 en una muestra puede evaluarse, por ejemplo, por rtPCR, qPCR, reacción en cadena de la ligasa, 10 replicación de secuencias auto-sostenida, replicación en círculo rodante, o cualquier otro método de amplificación de ácidos nucleicos, seguido de la detección de las moléculas amplificadas usando técnicas conocidas en la técnica. La PCR cuantitativa en tiempo real (o qPCR) representa un ejemplo de un método de perfilado molecular específico y altamente sensible para cuantificar los niveles de expresión génica de Ggta1 en un número limitado de células.

15 Como se usa en el presente documento, un par de cebadores apropiado se define como un par de oligonucleótidos que puede hibridarse con las regiones 5' y 3' de un gen (hebras más y menos, respectivamente, o viceversa) que definen una región entremedias. En general, los cebadores de amplificación son de aproximadamente 10 a 30 nucleótidos de longitud y flanquean una región de aproximadamente 50 a 200 nucleótidos de longitud. Bajo condiciones apropiadas y con reactivos apropiados, tales cebadores permiten la amplificación de una molécula de ácido nucleico que comprende la secuencia de nucleótidos flanqueada por los cebadores.

20 El nivel de expresión génica de Ggta1 en una célula (por ejemplo, una célula CHO) puede determinarse tanto por formatos *in situ* como *in vitro*. En un formato, se inmoviliza ARNm (o ADNc) sobre una superficie y se pone en contacto con las sondas, por ejemplo, haciendo migrar el ARNm aislado sobre un gel de agarosa y transfiriendo el ARNm del gel a una membrana, tal como nitrocelulosa. En un formato alternativo, las sondas se inmovilizan sobre una superficie y el ARNm (o ADNc) se pone en contacto con las sondas, por ejemplo, en un matriz de chip génico 25 bidimensional descrita más adelante. Un experto puede adaptar los métodos de detección de ARNm conocidos para su uso en detectar el nivel de ARNm codificado por un gen Ggta1. Para métodos *in situ*, una célula o muestra de tejido puede prepararse/procesarse e inmovilizarse sobre un soporte, normalmente un portaobjetos de vidrio, y entonces ponerse en contacto con una sonda que puede hibridarse con un gen Ggta1 que se analiza.

30 En otra realización, los métodos incluyen adicionalmente poner en contacto una muestra de control con un compuesto o agente capaz de detectar ARNm de Ggta1, o ADN genómico, y comparar la presencia de ARNm de Ggta1 o ADN genómico en la muestra de control con la presencia de ARNm de Ggta1 o ADN genómico en la muestra de prueba. En otra realización adicional, se usa análisis en serie de la expresión génica, por ejemplo, como se describe en la patente de EE.UU. N.º 5.695.937, para detectar los niveles de transcrito de Ggta1.

35 Las técnicas basadas en anticuerpos para la detección de proteína Ggta1 incluyen enzoinmunoanálisis de adsorción (ELISA), inmunoprecipitaciones, inmunofluorescencia, inmunoensayo enzimático (EIA), radioinmunoensayo (RIA), análisis de transferencia Western, resonancia de plasmones superficiales. Otros métodos pueden incluir la detección de péptidos Ggta1 o fragmentos de los mismos usando métodos basados en espectrometría de masas, que incluyen, pero no se limitan a, CL-EM, EM/EM, EM/EM/EM, MALDI-EM, monitorización de múltiples reacciones (MRM).

40 En una realización, los métodos de detección descritos en el presente documento son parte de la determinación de un perfil de expresión génica de la muestra, en los que el perfil incluye un valor que representa el nivel de expresión de Ggta1, entre al menos otro valor para la expresión de al menos otro gen. El método puede incluir además comparar el valor o el perfil (es decir, múltiples valores) con un valor de referencia o perfil de referencia. El perfil de expresión génica de la muestra puede obtenerse por cualquiera de los métodos descritos en el presente documento 45 (por ejemplo, proporcionando un ácido nucleico de la muestra y poniendo en contacto el ácido nucleico con una matriz). El método puede usarse para evaluar o cribar células CHO.

50 En otro aspecto, la divulgación presenta un medio informático que tiene una pluralidad de registros de datos digitalmente codificados. Cada registro de datos incluye un valor que representa el nivel de expresión de hámster chino o Ggta1 en una muestra, y un descriptor de la muestra. El descriptor de la muestra puede ser un identificador de la muestra, por ejemplo, el tipo de célula del que la muestra se derivó (por ejemplo, una cepa de células CHO), o una condición de cultivo celular bajo la que se cultivó la célula que es la fuente de la muestra. En una realización, el registro de datos incluye además valores que representan el nivel de expresión de genes distintos de Ggta1 (por ejemplo, otros genes asociados a la síntesis de glucano, u otros genes sobre una matriz). El registro de datos puede estar estructurado como una tabla, por ejemplo, una tabla que es parte de una base de datos tal como una base de 55 datos relacional (por ejemplo, una base de datos SQL de los entornos de bases de datos Oracle o Sybase).

Matrices y usos de las mismas

La divulgación también presenta una matriz bidimensional que tiene una pluralidad de direcciones, siendo cada dirección de la pluralidad posicionalmente distinguible de cada otra dirección de la pluralidad. Cada dirección de la

5 pluralidad tiene una sonda de captura única, por ejemplo, una secuencia de ácidos nucleicos o de polipéptidos o anticuerpo descrita en el presente documento. Al menos una dirección de la pluralidad tiene una sonda de captura que reconoce una molécula de Ggta1 de hámster chino o de células CHO descrita en el presente documento (por ejemplo, una molécula de Ggta-1, Ggta1-b o Ggta1-c descrita en el presente documento). En una realización, la sonda de captura es un ácido nucleico, por ejemplo, una sonda complementaria a una secuencia de ácidos nucleicos de Ggta1. En otra realización, la sonda de captura es un anticuerpo, por ejemplo, un anticuerpo específico para un polipéptido Ggta1 descrito en el presente documento. También se presenta un método de análisis de una muestra (por ejemplo, una muestra de una célula CHO) poniendo en contacto la muestra con la matriz anteriormente mencionada y detectando la unión de la muestra a la matriz.

10 La matriz puede tener una densidad de al menos 10, 50, 100, 200, 500, 1.000, 2.000 o 10.000 o más direcciones/cm², e intervalos intermedios. En una realización preferida, la pluralidad de direcciones incluye al menos 10, 100, 500, 1.000, 5.000, 10.000, 50.000 direcciones. En una realización preferida, la pluralidad de direcciones incluye igual a o menos de 10, 100, 500, 1.000, 5.000, 10.000 o 50.000 direcciones. El sustrato puede ser un sustrato bidimensional tal como un portaobjetos de vidrio, una oblea (por ejemplo, sílice o plástico), una placa de espectroscopía de masas, o un sustrato tridimensional tal como una almohadilla de gel. Pueden disponerse direcciones, además de la dirección de la pluralidad, sobre la matriz.

15 Una matriz puede generarse por diversos métodos, por ejemplo, por métodos fotolitográficos (véanse, por ejemplo, las patentes de EE.UU. N.º 5.143.854; 5.510.270; y 5.527.681), métodos mecánicos (por ejemplo, métodos de flujo dirigido como se describen en la patente de EE.UU. N.º 5.384.261), métodos basados en aguja (por ejemplo, como se describen en la patente de EE.UU. N.º 5.288.514) y técnicas basadas en perlas (por ejemplo, como se describen en PCT US/93/04145).

20 En otro aspecto, la divulgación presenta un método de análisis de la expresión de Ggta1. El método incluye proporcionar una matriz como se ha descrito anteriormente; poner en contacto la matriz con una muestra y detectar la unión de una molécula de Ggta1 (por ejemplo, ácido nucleico o polipéptido) a la matriz. En una realización preferida, la matriz es una matriz de ácido nucleico. Opcionalmente, el método incluye además amplificar ácido nucleico de la muestra antes o durante el contacto con la matriz.

25 En otra realización, la matriz puede usarse para ensayar la expresión génica en una población de células (por ejemplo, una población de células CHO), particularmente la expresión de Ggta1. Si se analiza un número suficiente de diversas muestras, puede usarse agrupamiento (por ejemplo, agrupamiento jerárquico, agrupamiento K-means, agrupamiento bayesiano y similares) para identificar otros genes que están co-regulados con Ggta1. Por ejemplo, la matriz puede usarse para la cuantificación de la expresión de múltiples genes. Así, no solo puede determinarse la especificidad por el tipo de célula, sino también el nivel de expresión de una batería de genes en el tejido. Pueden usarse datos cuantitativos para reunir (por ejemplo, agrupar) genes basándose en su expresión en tejido por ella misma y el nivel de expresión en ese tejido.

30 En un ejemplo, puede usarse el análisis de matrices de la expresión génica para evaluar el efecto de diferentes condiciones de cultivo celular sobre la expresión de Ggta1. Puede cultivarse una primera población de células bajo un primer conjunto de condiciones, y puede cultivarse una segunda población de células bajo un segundo conjunto de condiciones y el ácido nucleico de la pluralidad de poblaciones puede analizarse en una matriz tal. En este contexto, puede determinarse el efecto de las condiciones de cultivo sobre la respuesta biológica. Similarmente, pueden compararse células CHO que tienen diferentes contextos genéticos.

Métodos de modulación de la estructura de glucano de glucoproteínas

35 La divulgación presenta métodos de producción de una glucoproteína en células CHO, donde la glucoproteína tiene niveles alterados de estructuras de glucano de Gal-alfa-Gal, por ejemplo, con respecto a una glucoproteína en el estado de la técnica, por ejemplo, con respecto a una glucoproteína comercializada como una glucoproteína terapéutica. El método incluye cultivar las células huésped descritas en el presente documento (por ejemplo, las células huésped transfectadas con Ggta1 o silenciadas o inactivadas para Ggta1) en condiciones adecuadas para la expresión de la glucoproteína. La glucoproteína puede seleccionarse de aquellas descritas en la Tabla 5.

40 En una realización, se produce una glucoproteína que tiene niveles elevados de estructuras de glucano de Gal-alfa-Gal con respecto a una glucoproteína en el estado de la técnica que tiene la misma secuencia de polipéptidos o altamente similar (por ejemplo, al menos el 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % de secuencia de polipéptidos idéntica). El método incluye cultivar una célula CHO huésped descrita en el presente documento que ha sido genéticamente manipulada para producir niveles elevados de Ggta1 de hámster chino o de células CHO con respecto a la célula parental (por ejemplo, una Ggta1 que expresa en exceso célula CHO), donde la célula huésped también contiene un transgén que codifica una glucoproteína terapéutica, y purificar la glucoproteína terapéutica a partir de las células cultivadas resultantes.

45 En otra realización, se produce una glucoproteína que tiene niveles reducidos de estructuras de glucano de Gal-alfa-Gal con respecto a una glucoproteína en el estado de la técnica que tiene la misma secuencia de polipéptidos o altamente similar (por ejemplo, al menos el 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % de

- 5 secuencia de polipéptidos idéntica). El método incluye cultivar una célula huésped CHO descrita en el presente documento que ha sido genéticamente manipulada para producir niveles reducidos de Ggta1 de hámster chino o de células CHO con respecto a la célula parental (por ejemplo, una célula CHO inactivada o silenciada para Ggta1 descrita en el presente documento), donde la célula huésped contiene un transgén que codifica una glucoproteína terapéutica, y purificar la glucoproteína terapéutica a partir de las células cultivadas resultantes. Tales glucoproteínas así producidas pueden proporcionar versiones mejoradas de glucoproteínas terapéuticas comercializadas.

TABLA 5

Producto de proteína	Fármaco de referencia
interferón gamma-1b	Actimmune®
alteplasa; activador tisular del plasminógeno	Activase®/Cathflo®
factor antihemofílico recombinante	Advate
albúmina humana	Albutein®
laronidasa	Aldurazyme®
interferón alfa-N3, derivado de leucocitos humanos	Alferon N®
factor antihemofílico humano	Alphanate®
factor de coagulación IX humano filtrado de virus	AlphaNine® SD
Alefacept; recombinante, proteína de fusión LFA3-Ig dimérica	Amevive®
bivalirudina	Angiomax®
darbepoetina alfa	Aranesp™
bevacizumab	Avastin™
interferón beta-1a; recombinante	Avonex®
factor de coagulación IX	BeneFix™
Interferón beta-1b	Betaseron®
Tositumomab	Bexxar®
factor antihemofílico	Bioclone™
hormona de crecimiento humana	BioTropin™
toxina botulínica tipo A	Botox®
alemtuzumab	Campath®
acritumomab; marcada con tecnecio-99	CEA-Scan®
alglucerasa; forma modificada de beta-glucocerebrosidasa	Ceredase®
imiglucerasa	Cerezyme®
Fab inmune polivalente para crótalos, ovino	CroFab™
Fab inmune de digoxina, ovina	DigiFab™
rasburicasa	Elitek®
etanercept	Enbrel®
epoyetina alfa	Epogen®
cetuximab	Erbitux™

TABLA 5	
Producto de proteína	Fármaco de referencia
algasidasa beta	Fabrazyme®
urofolitropina	Fertinex™
folitropina beta	Follistim™
teriparatida	Forteo®
somatropina humana	GenoTropin®
glucagón	GlucaGen®
folitropina alfa	Gonal-F®
factor antihemofílico	Helixate®
factor antihemofílico; factor XIII	Hemofil®
insulina	Humalog®
complejo de factor antihemofílico / factor de von Willebrand	Humate-P®
somatotropina	Humatrope®
adalimumab	HUMIRA™
insulina humana	Humulin®
hialuronidasa humana recombinante	Hylenex™
interferón alfacon-1	Infergen®
Eptifibatida	Integrilin™
alfa-interferón	Intron A®
palifermina	Kepivance
anakinra	Kineret™
factor antihemofílico	Kogenate®FS
insulina glargina	Lantus®
factor estimulante de colonias de granulocitos-macrófagos	Leukine®/Leukine® Liquid
lutropina alfa, para inyección	Luveris
lipoproteína OspA	LYMERix™
ranibizumab	Lucentis®
gemtuzumab ozogamicina	Mylotarg™
galsulfasa	Naglazyme™
nesiritida	Natrecor®
pegfilgrastim	Neulasta™
oprelvekin	Neumega®
filgrastim	Neupogen®
fanolesomab	NeuroSpec™ (antiguamente LeuTech®)

TABLA 5	
Producto de proteína	Fármaco de referencia
somatropina [ARNr]	Norditropin®/Norditropin Nordiflex®
insulina; suspensión de cinc;	Novolin L®
insulina; suspensión de isófana	Novolin N®
insulina, regular;	Novolin R®
insulina	Novolin®
factor de coagulación VIIa	NovoSeven®
somatropina	Nutropin®
inmunoglobulina intravenosa	Octagam®
PEG-L-asparaginasa	Oncaspar®
abatacept, proteína de fusión soluble completamente humana	Orencia™
muromomab-CD3	Orthoclone OKT3®
gonadotropina coriónica humana	Ovidrel®
peginterferón alfa-2a	Pegasys®
versión pegilada de interferón alfa-2b	PEG-Intron™
Abarelix; antagonista de la hormona liberadora de gonadotropina	Plenaxis™
epoyetina alfa	Procrit®
aldesleucina	Proleukin, IL-2®
somatrem	Protropin®
dornasa alfa	Pulmozyme®
Efalizumab; bloqueante de linfocitos T reversible selectivo	Raptiva™
combinación de ribavirina e interferón alfa	Rebetron™
Interferón beta 1 a	Rebif®
factor antihemofílico	Recombinate®
rAHF/factor antihemofílico	ReFacto®
lepirudina	Refludan®
infiximab	Remicade®
abciximab	ReoPro™
reteplasa	Retavase™
rituximab	Rituxan™
interferón alfa-2a	Roferon-A®
somatropina	Saizen®
secretina porcina sintética	SecreFlo™
basiliximab	Simulect®

TABLA 5	
Producto de proteína	Fármaco de referencia
eculizumab	Soliris®
pegvisomant	Somavert®
Palivizumab; mAb humanizado recombinantemente producido	Synagis™
tirotropina alfa	Thyrogen®
tenecteplasa	TNKase™
natalizumab	Tysabri®
globulina intravenosa inmune humana	Venoglobulin-S®
interferón alfa-n1, linfoblastoide	Wellferon®
drotrecogina alfa	Xigris™
Omalizumab	Xolair®
daclizumab	Zenapax®
ibritumomab tiuxetan	Zevalin™
Somatotropina	Zorbtive™ (Serostim®)

Otras características y ventajas de la invención serán evidentes a partir de la siguiente descripción detallada, y de las reivindicaciones.

La presente invención se ilustra además por los siguientes ejemplos, que no deben interpretarse como limitantes.

5 EJEMPLOS

Ejemplo 1: Clonación de ADNc de Ggta1 de tejidos de *Cricetulus griseus* y de células CHO

Se diseñaron conjuntos de cebadores de oligonucleótidos originales (véase, por ejemplo, la Tabla 2) basados en la homología con tanto la secuencia codificante de Ggta1 de rata como de ratón. Intentos repetidos iniciales por usar estos cebadores para amplificar por PCR al menos una secuencia de Ggta1 parcial de ADNc derivado de células CHO o ADN genómico (ADNg) fueron no satisfactorios. La presente divulgación, por tanto, engloba el reconocimiento de que enfoques de hibridación estándar pueden ser insuficientes para permitir la clonación de Ggta1 de CHO.

Específicamente, se usó un subconjunto de estos mismos conjuntos de cebadores para cribar cuatro conjuntos de ADNc específico de tejido derivados de hámster chino (en vez de una línea de células CHO): cerebro, bazo, riñón y ovario. Estos conjuntos de ADNc se derivaron de dos animales de sexo diferenciado separados (conjuntos de cerebro y bazo de ADNc de un hámster chino macho y conjunto de riñón y ovario de ADNc de un hámster chino hembra). Los conjuntos de ADNc se validaron por PCR como que eran específicos para hámster chino usando un conjunto de cebadores de oligonucleótidos complementarios a una secuencia de genes específica de CHO de control conocida. Los cuatro conjuntos de ADNc produjeron la amplificación de un producto específico de 1,1 kb (véase la **Figura 1**). Se confirmó por secuenciación de ADN que los productos de PCR se correspondían con el supuesto ortólogo de hámster chino de Ggta1 basado en homología de secuencias con las secuencias traducidas de Ggta1 de ratón y de rata (**Figura 2**).

La secuencia codificante de Ggta1 de ovario de hámster chino (denominada en el presente documento Ggta1-a) se muestra como SEQ ID NO: 1 y su secuencia de aminoácidos como SEQ ID NO: 2 en la **Figura 3**. La secuencia codificante de la Ggta1 de bazo de hámster chino (denominada en el presente documento Ggta1-b) se muestra como SEQ ID NO: 3 y su secuencia de aminoácidos como SEQ ID NO: 4 en la **Figura 4**. Como se ha descrito anteriormente, estas dos secuencias se derivaron de "donantes" de hámster chino separados (concretamente hembra y macho). Los dos clones parecen tener longitud completa e incluyen todos los exones predichos (4 a 9) basándose en una comparación con la estructura del exón del gen Ggta1 de ratón.

Las secuencias codificantes de Ggta1-a y Ggta1-b son más del 99 % idénticas entre sí, discrepando solo en 6 nucleótidos de la secuencia codificante de 1110 pb. Cinco de estas diferencias de nucleótidos son silenciosas; solo

una produce un cambio de aminoácido conservativo (Ile69Val). Por consiguiente, las secuencias de aminoácidos de Ggta1-a y Ggta1-b son más del 99,5 % idénticas.

Las secuencias de nucleótidos de Ggta1-a y Ggta1-b son el 79,4 y el 79,6 % idénticas, respectivamente, a la secuencia de rata; 87,2 % y 87,5 % idénticas, respectivamente, a la secuencia de ratón; 77,5 % y 77,7 % idénticas, respectivamente, a la secuencia de vaca, y 82,2 % y 82,3 % idénticas, respectivamente, a la secuencia de perro. A nivel de aminoácidos, las secuencias de Ggta1-a y Ggta1-b son el 81 % idénticas a la secuencia de rata; 86 % idénticas a la secuencia de ratón, 73 % idénticas a la secuencia de vaca, y 77 % idénticas a la secuencia de perro.

Las regiones más altamente conservadas son aquellas aparentemente requeridas para unión al ligando y la actividad catalítica, correspondientes a los restos de aminoácidos Iso-82 a Val-370 de SEQ ID NO: 2 o 4 (véase Biochimica et Biophysics Acta (2000) 1480:222-234). Véase la Figura 2. Al nivel de aminoácidos, las secuencias de Ggta1-a y Ggta1-b son el 91 % idénticas a la secuencia de rata; 90,3 % idénticas a la secuencia de ratón, el 83 % idénticas a la secuencia de vaca, y 79 % idénticas a la secuencia de perro con respecto a la región catalítica. Las regiones catalíticas de Ggta1-a y Ggta1-b son más del 95 % idénticas a la región catalítica de Ggta1-c.

Similar a los genes de las secuencias de ratón y de rata, también se observó corte y empalme alternativo del gen Ggta1 de hámster chino (datos no mostrados). En particular, parece que en al menos un clon, el exón 6 se escindió (correspondiente a Iso-40 a Gly-60 de SEQ ID NO: 2). Esta delección fue perceptible por PCR, produciendo un fragmento de ADN más pequeño amplificado cuando se visualizó por electroforesis en gel de agarosa (obsérvese el doblete en la Figura 1, carriles 3-6).

Basándose en las secuencias de Ggta1 de hámster chino anteriormente descritas derivadas de conjuntos de ADN específico de tejido, se diseñaron conjuntos de cebadores de oligonucleótidos de qPCR adicionales con intención de usarlos para amplificar por PCR una Ggta1 de ADNc generado a partir de una línea de células CHO. Sin embargo, los esfuerzos para hacer esto no fueron inicialmente satisfactorios. Más sorprendentemente, estos cebadores también dejaron de amplificar cualquier copia genómica del gen (si se usa ADN genómico aislado de la misma fuente de células CHO que se usó para el cribado de ADNc). Sin desear ceñirse a teoría particular alguna, los presentes inventores propusieron que el problema podría resultar de la homogeneidad de secuencias de ADN presente en la copia genómica del gen Ggta1 de la línea de células CHO en comparación con el genoma del hámster chino parental del que los conjuntos de ADNc (usados para clonar secuencias de Ggta1-a y Ggta1-b) se derivaron independientemente. Además, el exón 9, que codifica la mayoría de la región catalítica, presentó características que inicialmente condujeron a los presentes inventores a creer que era tóxico para *E. coli* (las células bacterianas usadas para clonar), suponiendo además un reto a los esfuerzos de clonación satisfactorios de células CHO.

La amplificación satisfactoria del ADN genómico (ADNg) de células CHO de una secuencia de genes parcial de 5.877 pares de bases y que comprende los exones 8, 9 (y el intrón intermedio) se logró por último lugar usando un conjunto racionalmente diseñado de cebadores de PCR (Tabla 2). Esta secuencia genómica de Ggta1 parcial se muestra en la **Figura 5** (SEQ ID NO: 5), y la secuencia de aminoácidos correspondiente en la **Figura 6** (SEQ ID NO: 6). Un análisis de secuencias de varios clones confirmó la presencia de polimorfismos de ADN (con respecto a las secuencias de Ggta1-a y Ggta1-b originales). En particular, se observaron 13 cambios de aminoácidos (de los 275 restos) en la región del exón 8-9 del gen Ggta1 con respecto a la secuencia originalmente clonada a partir del conjunto de ADNc de hámster chino de ADNc derivado de ovario. Este polimorfismo es equivalente a aproximadamente una diferencia del 5 % en la secuencia de proteínas (6,5 % de diferencia al nivel de nucleótidos) cuando se comparan las dos secuencias de genes. A diferencia, una secuencia de CHO de control sin relacionar asimismo amplificada por PCR a partir de los cuatro conjuntos de ADNc de hámster chino específico de tejido presentaron todos el 100 % identidad de pares de bases cuando se comparó directamente con la secuencia de ADNc derivada de células CHO. Esta observación sugiere un sesgo específico del gen en polimorfismos genéticos.

En resumen, la clonación de Ggta1 de células CHO fue técnicamente desafiante. Mientras que las secuencias de genes Ggta1 para varias especies, que incluyen ratón, rata, vaca, cerdo, perro, gato, y varias otras, estuvieron públicamente accesibles, el uso de estas secuencias relacionadas para obtener el gen Ggta1 de células CHO por clonación por homología rutinaria no fue satisfactorio. Acoplado a la falta de apoyo en la bibliografía para la presencia de una actividad de Ggta1 en células CHO, otros podrían entonces haber llegado a la conclusión de que una Ggta1 de CHO no existió. En su lugar, los inventores reconocieron las fuentes de problemas que encontraron, y continuaron clonando los genes Ggta1 de tejidos y ADN genómico de hámster chino con el fin de proporcionar las herramientas que por último lugar permitieran la clonación de ADNc de Ggta1 de células CHO. Entre otras cosas, los retos presentados a los investigadores que investigan la clonación de ADNc de Ggta1 de células CHO incluyeron (1) consenso general en la bibliografía de que las células CHO no incluyeron la actividad de Ggta1; (2) divergencia poco usual de la homología de la secuencia de ADN entre secuencias de roedor (por ejemplo, ratón frente a rata en la que las dos secuencias de genes Ggta1 respectivas divergen aproximadamente el 9 %); y (3) polimorfismos genéticos de la secuencia de genes Ggta1 de células CHO. Los presentes inventores identificaron estos retos, desarrollaron estrategias para cumplirlos, y clonaron satisfactoriamente secuencias de Ggta1 de CHO.

Ejemplo 2: Uso de perfil transcripcional por qPCR para cribar la expresión diferencial del gen Ggta1 en clones de la línea de células CHO

El uso satisfactorio de PCR requiere consideraciones rigurosas, concretamente el uso de cebadores de oligonucleótidos cuidadosamente diseñados con secuencias de nucleótidos complementarias a la secuencia de genes diana que se detecta. El diseño de estos cebadores que coinciden en secuencia para tratar un perfil específicamente de la expresión génica de Ggta1 en células CHO no fue inicialmente posible sin primero determinar la secuencia exacta del gen Ggta1 de CHO, como se ha descrito anteriormente. Basándose en el conocimiento de la secuencia de Ggta1 de células CHO clonada como se ha descrito anteriormente, cebadores adicionales se diseñaron entonces cuidadosamente para cuantificar por qPCR los niveles de expresión relativa de genes Ggta1 en la población clonal de células CHO que expresa CTLA4-Ig. Estos cebadores se diseñaron para hibridarse específicamente en una región de la secuencia del exón 9 de Ggta1 que se mostró mediante un alineamiento de múltiples secuencias que estaba absolutamente conservada en sus secuencias de nucleótidos respectivas (**Tabla 2**). Los cebadores se clasificaron basándose en la eficiencia, especificidad y sensibilidad usando un molde de ADN de control. El intervalo dinámico de estos cebadores superó **8** órdenes de magnitud, que indica su capacidad en principio para detectar niveles de ARNm de Ggta1 hasta la copia de un único dígito por célula.

Los resultados de qPCR se resumen en la **Figura 7**. Se aislaron clones por dilución tras la amplificación basada en DHFR de una línea de células CHO que expresa CTLA4-Ig. La expresión de Ggta1 se cuantificó por qPCR usando cebadores específicos del exón 9 845 (directo) y 846 (inverso) (secuencias mostradas en la Tabla 2). La Figura 7A muestra los perfiles de ciclos de PCR de clones seleccionados. La Figura 7B muestra una evaluación de la especificidad por cebador por análisis de Tm. La **Tabla 3** ilustra los valores de ciclo umbral (Ct) para tres clones ilustrativos. Estos resultados indican la amplificación de un producto específico (basado en el análisis de la temperatura de fusión del producto amplificado, **Figura 7B**). Además, estos resultados también demuestran el uso de estos cebadores para diferenciar clones únicos derivados de una única línea de células CHO parental basándose en los niveles de expresión relativa del gen Ggta1 (**Tabla 3**).

Además, los presentes inventores diseñaron cebadores que se diferencian entre copias genómicas frente a transcripcionales (ADNc) del gen Ggta1 por mapeo en exones contiguos (es decir, cebador directo, el exón 8; *cebador inverso, el exón 9*). La amplificación diferencial de la secuencia de Ggta1 correspondiente al molde de ADNc frente a un molde de ADN genómico (ADNg) se ilustra en la **Figura 8**. Estos conjuntos de cebadores no amplificarán copias genómicas de Ggta1 bajo las condiciones de qPCR usadas debido a la presencia de una gran secuencia de intrón intermedio (>5 kb), que físicamente separa los dos exones en la copia genómica, pero *posteriormente* se corta y empalma para formar el transcrito de ARNm de Ggta1 maduro. Otros cebadores desvelados en el presente documento se usaron asimismo para cribar clones de células CHO para la expresión diferencial del gen Ggta1. La **Figura 9** muestra que el producto de PCR de Ggta1 (amplificado específicamente a partir de ADNc) tiene 324 pb. Tales cebadores tienen valor particular en ciertas aplicaciones, tales como la necesidad de detectar y/o cuantificar inequívocamente bajos niveles de expresión génica de Ggta1 en un modo específico que excluye cualquier posibilidad de un resultado positivo falso debido a ADN genómico sobrante.

Fin	Conjunto de cebadores	Secuencia
Clonación de ADNc de la secuencia codificante de Ggta1 de longitud completa de tejidos de hámster chino	807 (directo)	TGGATCACAGGAGAAAATAATGAA (SEQ ID NO: 8)
	806 (inverso)	AAGTTTCCATCACAAATTTGAAGTCAGA (SEQ ID NO: 9)
Clonación de la secuencia de genes Ggta1 (exón 8, 9 e intrón) de ADN genómico de células CHO	863 (directo)	GAACCGCCAGAAAGTTTTGACAGTGACC (SEQ ID NO: 10)
	830 (inverso)	GTCAGACATTACTCCTAACCAAATTATAC (SEQ ID NO: 11)
Perfil transcripcional de la expresión de Ggta1 en línea de células CHO	845 (directo)	ATGCCTCCAGAATGCCTTT (SEQ ID NO: 12)
	846 (inverso)	ATCCACGTCCATACAGAAGA (SEQ ID NO: 13)
	875 (directo)	GCCAGACAGAAAATCACCG (SEQ ID NO: 14)
	855 (inverso)	AAAGACCTGATCCACGTCCAT (SEQ ID NO: 15)
	897 (directo)	TGGGAAGGCACTTATGACAG (SEQ ID NO: 16)

Tabla 2		
Cebadores de oligonucleótidos de Ggta1 usados para la clonación y el perfil transcripcional		
Fin	Conjunto de cebadores	Secuencia
	892 (inverso)	TTGTAAGGAATGCAGAGGAC (SEQ ID NO: 17)

Tabla 3						
Valores de Ct de clones ilustrativos cribados por qPCR usando los cebadores 845 y 846						
Pares de cebadores	Clon 33	Clon 33 (sin RT)	Clon 34	Clon 34 (sin RT)	Clon 57	Clon 57 (sin RT)
pr845+pr846	32,74	ND	34,26	ND	ND	ND

Los clones se obtuvieron por dilución de la línea de células CHO DHFR- amplificada por MTX que expresa CTLA4-Ig;
Sin RT= control de transcriptasa inversa menos para excluir la amplificación de copias genómicas del gen Ggta1;
ND=no detectado

Tabla 4				
Oligonucleótidos/cebadores/sondas a modo de ejemplo				
nombre	secuencia	hebra	nt	fuelle
pr854	GATGCCTCCAGAATGCCTT (SEQ ID NO: 18)	directa	19	CHO1
pr874	CTCATTCTTGAAGCTATGCC (SEQ ID NO: 19)	directa	20	CHO1
pr875	GCCAGACAGAAAATCACCG (SEQ ID NO: 20)	directa	19	CHO1
pr876	CTTACGGGGACACATGCTA (SEQ ID NO: 21)	inversa	19	CHO1
pr877	GGCACAGAAGGGAAAGAC (SEQ ID NO: 22)	directa	18	CHO1
pr878	GTCAACCATCTAAGCTACAGT (SEQ ID NO: 23)	inversa	21	CHO1
pr886	AAATCTGACCCTAATCCTGTG (SEQ ID NO: 24)	directa	21	CHO1
pr887	CTCCTTGCTTTCTATGGCTT (SEQ ID NO: 25)	inversa	20	CHO1
pr888	GCCAGACAGAAAATCACCC (SEQ ID NO: 26)	directa	18	CHO1
pr895	GTCTCCTCCTAGTAACTCAA (SEQ ID NO: 27)	inversa	22	CHO1
pr896	CTAGCATGTGTCCCCGTAA (SEQ ID NO: 28)	directa	19	CHO1
pr898	GCTGGTGGTTCCCGAG (SEQ ID NO: 29)	directa	16	CHO1
pr855	AAAGACCTGATCCACGTCCAT (SEQ ID NO: 30)	inversa	21	Exón 8-9
pr889	GAAGTTTTGACAGTGACCCC (SEQ ID NO: 31)	directa	20	Exón 8-9
pr891	GTTTTTGCTGTGGGAAAGTACAT (SEQ ID NO: 32)	directa	23	Exón 8-9
pr892	TTGTAAGGAATGCAGAGGAC (SEQ ID NO: 33)	inversa	20	Exón 8-9
pr893	AAAGACCTGATCCACGTCCA (SEQ ID NO: 34)	inversa	20	Exón 8-9
pr894	GATCTCAAACACTTGTAAGGAATG (SEQ ID NO: 35)	inversa	24	Exón 8-9
pr897	TGGGAAGGCACTTATGACAG (SEQ ID NO: 36)	directa	20	Exón 8-9

pr822	AAATTCCAGAGATTGGTGACAGC (SEQ ID NO: 37)	directa	23	Ovario CH
pr823	GTCATAAGTGCCTTCCCACAC (SEQ ID NO: 38)	inversa	21	Ovario CH
pr824	TTATTACCACGCAGCCATTT (SEQ ID NO: 39)	directa	20	Ovario CH
pr825	TTCCATCACAATTTGAAGTCAGA (SEQ ID NO: 40)	inversa	23	Ovario CH
pr830	GTCAGACATTACTCCTAACCAAATTATAC (SEQ ID NO: 41)	inversa	43	Ovario CH
pr831	AATGAATGTCAAGGGAAAAGTGGTC (SEQ ID NO: 42)	directa	36	Ovario CH
pr842	TGATAGTCCCAACAGTACTCT (SEQ ID NO: 43)	inversa	21	Ovario CH
pr843	CACATCCTGACCCACATACA (SEQ ID NO: 44)	directa	20	Ovario CH
pr844	GACAGTTCCCGCCTCTCATA (SEQ ID NO: 45)	inversa	20	Ovario CH
pr845	ATGCCTCCAGAATGCCTTT (SEQ ID NO: 46)	directa	19	Ovario CH
pr846	ATCCACGTCCATACAGAAGA (SEQ ID NO: 47)	inversa	20	Ovario CH
pr849	GTACATTGAGCATTATTTGGAAG (SEQ ID NO: 48)	directa	23	Ovario CH
pr850	GTACATCGAGCATTATTTGGAAG (SEQ ID NO: 49)	directa	23	Ovario CH
pr863	GAACCGCCCAGAAAGTTTGGACAGTGACC (SEQ ID NO: 50)	directa	42	Ovario CH
pr899	GCTGGTGGTTCCCGAG (SEQ ID NO: 51)	directa	30	Ovario CH
pr900	GTCTCCTCCTAGTAACTCAA (SEQ ID NO: 52)	inversa	36	Ovario CH
pr922	ATGAAAATCCAGAGATTGGTGA (SEQ ID NO: 53)	directa	35	Ovario CH
pr923	GGGcACCCACAGTTATCAAGAA (SEQ ID NO: 54)	directa	36	Ovario CH
pr924	GTCAGAATTGAgGAGCCTCACTT (SEQ ID NO: 55)	directa	38	Ovario CH

Extensiones y alternativas

5 En el supuesto caso de que uno o más de la bibliografía citada y materiales similares se diferencien de o contradigan esta memoria descriptiva, que incluye pero no se limita a los términos definidos, uso de términos, técnicas descritas, o similares, controla esta memoria descriptiva. Los encabezados de sección usados en el presente documento son para fines organizativos solo y no deben interpretarse de ningún modo como limitantes de la materia descrita. Aunque la presente invención se ha descrito conjuntamente con diversas realizaciones y ejemplos, no se pretende que la invención se limite a tales realizaciones o ejemplos. Por el contrario, la invención engloba diversas alternativas, modificaciones y equivalentes que se encuentran dentro del alcance de las reivindicaciones adjuntas, como será apreciado por aquellos expertos en la materia.

LISTADO DE SECUENCIAS

<110> Meador, James W.

<120> GLUCOSILTRANSFERASA DE HÁMSTER CHINO Y MÉTODOS RELACIONADOS

<130> 2007774-0095 (M0061WO)

<150> US 61/260232

< 151> 11-11-2009

<160> 59

<170> PatentIn versión 3.5

5 <210> 1

<211> 780

<212> ADN

<213> Cricetulus griseus

10 <400> 1

```

atgaatgtca agggaaaagt ggtcctgtcg atgctggttg tctcaactgt gcttgtcgtg      60
ttttgggaat atgtcaacag cccagaaggt tccttcttgt ggatatatca catgaaaatt      120
ccagagattg gtgacagcag caggcagagc tggtggttcc cgagctggtt tagcaatggg      180
accacagtt atcaagaaga ggacgtagaa agagggagag aaaagcagcg aaatgcagtc      240
agaattgaag agcctcactt gtgggactgg ttaaatcaa agaaccgcc agaagttttg      300
acagtgacct catggaaggc accaattgtg tgggaaggca cttatgacag agctatgctg      360
gaaaagtatt acgccagaca gaaaatcacc gtgggactga cagtttttgc tgtgggaaag      420
tacattgagc attatttgga agatttcctg gagtctgcta gcaggactt catggttggc      480
cacagggtca tattttatgt catgatggac gatgcctcca gaatgccttt tatacacctg      540
agtcctctgc attccttaca agtgtttgag atcaagtctg agaccagtg gcaagacatt      600
agcatgatgc gcatgaagac aatcgaggag cacatcctga cccacataca gcatgaggtc      660
gacttcctct tctgtatgga cgtggatcag gtctttcaag acaattttgg agtggagact      720
ctgggccagt tggtagctca gctccaggcc tggtggtaca aagctgacca tgacaagttt      780

```

<210> 2

15 <211> 260

<212> PRT

<213> Cricetulus griseus

<400> 2

20

ES 2 605 677 T3

Met Asn Val Lys Gly Lys Val Val Leu Ser Met Leu Val Val Ser Thr
 1 5 10 15

Val Leu Val Val Phe Trp Glu Tyr Val Asn Ser Pro Glu Gly Ser Phe
 20 25 30

Leu Trp Ile Tyr His Met Lys Ile Pro Glu Ile Gly Asp Ser Ser Arg
 35 40 45

Gln Ser Trp Trp Phe Pro Ser Trp Phe Ser Asn Gly Thr His Ser Tyr
 50 55 60

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

Arg Ile Glu Glu Pro His Leu Trp Asp Trp Phe Asn Pro Lys Asn Arg
 85 90 95

Pro Glu Val Leu Thr Val Thr Pro Trp Lys Ala Pro Ile Val Trp Glu
 100 105 110

Gly Thr Tyr Asp Arg Ala Met Leu Glu Lys Tyr Tyr Ala Arg Gln Lys
 115 120 125

Ile Thr Val Gly Leu Thr Val Phe Ala Val Gly Lys Tyr Ile Glu His
 130 135 140

Tyr Leu Glu Asp Phe Leu Glu Ser Ala Ser Arg Tyr Phe Met Val Gly
 145 150 155 160

His Arg Val Ile Phe Tyr Val Met Met Asp Asp Ala Ser Arg Met Pro
 165 170 175

Phe Ile His Leu Ser Pro Leu His Ser Leu Gln Val Phe Glu Ile Lys
 180 185 190

Ser Glu Thr Arg Trp Gln Asp Ile Ser Met Met Arg Met Lys Thr Ile
 195 200 205

Glu Glu His Ile Leu Thr His Ile Gln His Glu Val Asp Phe Leu Phe
 210 215 220

Cys Met Asp Val Asp Gln Val Phe Gln Asp Asn Phe Gly Val Glu Thr
 225 230 235 240

Leu Gly Gln Leu Val Ala Gln Leu Gln Ala Trp Trp Tyr Lys Ala Asp
 245 250 255

His Asp Lys Phe
 260

<210> 3

5 <211> 1110

ES 2 605 677 T3

<212> ADN

<213> Cricetulus griseus

<400> 3

5

```

atgaatgtca agggaaaagt ggtcctgtcg atgctggttg tctcaactgt gcttgtcgtg      60
ttttgggaat atgtcaacag cccagaaggt tccttcttgt ggatataatca catgaaaatt      120
ccagagattg gtgacagcag caggcagagc tggtggttcc cgagctggtt tagcaatggg      180
accacagtt atcaagaaga ggacatagaa agagggagag aaaagcagcg aaatgcagtc      240
agaattgaag agcctcactt gtgggactgg tttaatccaa agaaccgccc agaagttttg      300
acagtgaccc catggaaggc accaattgtg tgggaaggca cttatgacag agctatgctg      360
gaaaagtatt atgccagaca gaaaatcacc gtgggactga cagtttttgc tgtgggaaag      420
tacatcgagc attatttgga agatttccta gagtctgcta gcaggtactt catggttggc      480
cacagggatc tattttatgt catgatggac gatgcctcca gaatgccttt tatacacctg      540
agtcctctgc attccttaca agtgtttgag atcaagtctg agaccaggtg gcaagacatt      600
agcatgatgc gcatgaagac aatcgaggag cacatcctga cccacataca gcatgaggtc      660
gacttcctct tctgtatgga cgtggatcag gtctttcaag acaattttgg agtggagact      720
ctgggccagt tggtagctca gctccaggcc tggtggtaca aagctgacca cgacaagttt      780
acctatgaga ggccgggaact gtcggcggca tacatcccat ttggacaggg ggatttttat      840
taccatgcag ccattttcgg agggacaccc attcatattc tcaacctcac ccgggagtg      900
tttgagggga tcctccagga caagaacaac aacatagaag caaagtggca tgacgaaagt      960
cacctaaaca aatatttcct tttcaacaag cccactaaaa ttttatctcc agagtactgt     1020
tgggactatc agataggctt gccttcagaa attaaaaatg taaagatagc ttggcagaca     1080
aaagagtata atttggttag gagtaatgctc                                     1110

```

<210> 4

10

<211> 370

<212> PRT

<213> Cricetulus griseus

<400> 4

15

ES 2 605 677 T3

Met Asn Val Lys Gly Lys Val Val Leu Ser Met Leu Val Val Ser Thr
1 5 10 15

Val Leu Val Val Phe Trp Glu Tyr Val Asn Ser Pro Glu Gly Ser Phe
 20 25 30

Leu Trp Ile Tyr His Met Lys Ile Pro Glu Ile Gly Asp Ser Ser Arg
 35 40 45

Gln Ser Trp Trp Phe Pro Ser Trp Phe Ser Asn Gly Thr His Ser Tyr
 50 55 60

Gln Glu Glu Asp Ile Glu Arg Gly Arg Glu Lys Gln Arg Asn Ala Val
65 70 75 80

ES 2 605 677 T3

Arg Ile Glu Glu Pro His Leu Trp Asp Trp Phe Asn Pro Lys Asn Arg
85 90 95

Pro Glu Val Leu Thr Val Thr Pro Trp Lys Ala Pro Ile Val Trp Glu
100 105 110

Gly Thr Tyr Asp Arg Ala Met Leu Glu Lys Tyr Tyr Ala Arg Gln Lys
115 120 125

Ile Thr Val Gly Leu Thr Val Phe Ala Val Gly Lys Tyr Ile Glu His
130 135 140

Tyr Leu Glu Asp Phe Leu Glu Ser Ala Ser Arg Tyr Phe Met Val Gly
145 150 155 160

His Arg Val Ile Phe Tyr Val Met Met Asp Asp Ala Ser Arg Met Pro
165 170 175

Phe Ile His Leu Ser Pro Leu His Ser Leu Gln Val Phe Glu Ile Lys
180 185 190

Ser Glu Thr Arg Trp Gln Asp Ile Ser Met Met Arg Met Lys Thr Ile
195 200 205

Glu Glu His Ile Leu Thr His Ile Gln His Glu Val Asp Phe Leu Phe
210 215 220

Cys Met Asp Val Asp Gln Val Phe Gln Asp Asn Phe Gly Val Glu Thr
225 230 235 240

Leu Gly Gln Leu Val Ala Gln Leu Gln Ala Trp Trp Tyr Lys Ala Asp
245 250 255

His Asp Lys Phe Thr Tyr Glu Arg Arg Glu Leu Ser Ala Ala Tyr Ile
260 265 270

Pro Phe Gly Gln Gly Asp Phe Tyr Tyr His Ala Ala Ile Phe Gly Gly
275 280 285

Thr Pro Ile His Ile Leu Asn Leu Thr Arg Glu Cys Phe Glu Gly Ile
290 295 300

Leu Gln Asp Lys Asn Asn Asn Ile Glu Ala Lys Trp His Asp Glu Ser
305 310 315 320

His Leu Asn Lys Tyr Phe Leu Phe Asn Lys Pro Thr Lys Ile Leu Ser
325 330 335

Pro Glu Tyr Cys Trp Asp Tyr Gln Ile Gly Leu Pro Ser Glu Ile Lys
340 345 350

Asn Val Lys Ile Ala Trp Gln Thr Lys Glu Tyr Asn Leu Val Arg Ser
355 360 365

Asn Val
370

ES 2 605 677 T3

<210> 5

<211> 5877

<212> ADN

<213> *Cricetulus griseus*

5

<400> 5

gaaccgcca gaagtttga cagtgacccc atggaaggcg ccaattgtgt ggaaggcac	60
ttatgacaga gctctgctgg aaaagtatta tgccagacag aaaatcaccg tggggctgac	120
ggtttttgct gtgggaaagt aagcaaaaat gacaaaactg cccttgattt tttttttctt	180
gttttaccat caaagtcatt gtgagggacc cgtagcccta cagaagcctc ccttactttg	240
agtactagag aaggagactt ccatttcctt tctgatgaca gctgaggagc aaaatcaata	300
gtaggtaagt ccaccattc ccaccacaga acggagagaa ctgtcttcag tgccaccgga	360
accagagct aagagcatgc ctgccgctcc ctctccctg acctctaggg ctttcttagg	420
gaggcccaa atgccagctg ttaatgcca ggaccaagga cgggggttaga agttgtggag	480
tgtgctctgt gctccatggt ttccaggctg acatgtggct gctgctctct catgggcccc	540
tctggtcctg tctactggac cttcccggag gatttgacct tgggtttgga tgatcagaac	600
ctgctgagcc tggacaggtt ctggctgcag cagcttcac ttccttattt agagctggat	660
gcagaaacat tgtactcatt cttgaagcta tgccattggt tttagtctaa cccattttgt	720
tttcttagtg cattaacat tttcccctaa actaaaacat taggtttata ttttcattct	780
ttttccctt tctttctttc tttctttctt tctttctttc tttcttttct tgaggcagag	840
ttgcattttt aaggatggtt taagaaagga aaagtcagga ttatgagaga gccaccaaca	900
aagtaagaca tattggagag tgtgtgagtg taggcttctc tgagagaatc accctctgct	960
taaaaaaaaa aaaaaaaaaagt tttgtatcta gtgtgtgtgt gtaggtcaga ggatgacttt	1020
caggagttgg ctctcttctt ctactatgtg ggttctgggg atccaactta agtcttcag	1080
cttggcagca agtgccctta catactgaga attcttgctg ttatgtttct ttttctgcc	1140
tctgagcctc tgtgttgga cctcttatga aaaagtgaca aatcaacag gatgttgact	1200
cccctgccat gttcaaaag agtagaaggg gcagtttggc ctgacctga tctttgacaa	1260
gaacctaaaa gacagcaccg tgatgatgcc acagggcccc cacagcggcc agtgctgtct	1320
ttgggaaatt ttagcattgc tcagtgtgct gagtccctgg ggtccaatgg catcccaatg	1380
ggaacagtg gctgtggagt ttggacattt cactagaaga atttgtattt tttagggtgg	1440

ES 2 605 677 T3

gattcaaggc tcaacttagc aaacactttt ggcacagaag ggaaagacag ctatattaca 1500
cattgatttg atccccaaag aacaaaatga acccactgtc tcacaatggc tatcaaccat 1560
ccaattacaa ctttgacgga aaatagaaaa caaaacaata aaaaaaagtc ctacgactgc 1620
ctcccacttc ctggtgaaag ccgtgctccc agctaggcca gtctgcagtg gtgaatccca 1680
gaggaaatgc acctggacct ccgggaggtg tgacttttgc tttgcttttt atctcagtg 1740
ctggagtcaa ctgtgatacc ttaacctctt aaacatcctc agatgtcaca gtgacactgc 1800
cctacactcc cctaagcatt gtttccccga acttcaggca cactagtttt tcttgaagtt 1860
cagagtacag aagtctgtct tgacctccag ccatggagtc atgtcagagc tctcttttac 1920
cactgggata tatttgcccc acccaaagaa tcagtactag aaaccttgag ctttgcaatg 1980
ccaacagtaa aagtgaaatt tatttcgatg gcaagaaca acccaagcat caagaaaaac 2040
aaaacagaga ttggagagat ggctcagtg ttgatagtat tggctgccct tctagtggac 2100
ccagggttca attcccagca ctcacatggt gtgaggaaaa agggcagcca aagaaacaca 2160
cttggtcctt ggaatcagac ccagagaaaag aaatacatcc tggccctaaa attagagcca 2220
aaagactaaa aaggaccagt gaaagaaaca tactttggcc ctgaaaccag agccaaatct 2280
gaccctaate ctgtgcagaa aaccaaccaa tccttgagca ggagatttag ccattcagag 2340
ctcagggatt gaaatctggc caatccccac cctggaaatc tccttgaaa aactccccta 2400
agattcccca tataaacctt gtatctgttc agcttgtggc tgcttgatt cttccctccc 2460
aataaatctc ttcaatgatg tttgaggtga gaccttcttt caccagagag cagatcggag 2520
tagggctgta aactttgtct gctctccttg ccgaactgta acaactttac caggaagcc 2580
ctctcctgca ggagcagagc tgatacactg ggaagccgtt ccctgaaaca gggctgtaag 2640
ctgctatgct tactgtgacc tgtggtattc cttggctcct gattgccaga atacctttcc 2700
atctgagctg taacactcag tattcttcgg ctttcaagtg ccgtaatcca tcccatccga 2760
tctgtcaaaa tgcttacaca tggcagctca caactgtctg taactccagt tccaggggat 2820
ctggcaactt cacacaggca tacacacaag caaaacacaa atgcacataa aaataaatta 2880
ttggaaaaaa aagagaaaat gcaagtattt tttaaaaaag aaaaaaacac tagtcattct 2940
gottggctag ttgtgtgctc agtaaactgt ggtccgtagt ctctgtgtgt aagtttgagg 3000
tttgaaagg gaccatctga cttccccctc atctctgcag tctccttggc atgaattagg 3060
aaaatagtct gtagcttgag cttgtataag gcatcagtat aatgtcactc aatgataaa 3120
tagacaaatt gacaaatctc atctgttaga atggcaaaaa aaaaaaaaaa aaactgccaa 3180
taccaaatgt tgaagagggtg tgaagtaatc agaatccttg ttactgcta gtgggaatat 3240
aatatggtag agccatcttg ggagaaattt tggcagttgc ttgcaaaatt gagcatgctc 3300
ttatcataat ctaatcatca aattccttgg tattaacca agtgagttga aaaagtctta 3360
cacaaaaacc tgcatataga tattcttaaa agctttatctc ataattgcta aaacttggga 3420

ES 2 605 677 T3

gcagacaagt tttcctttga caggcaaatg taaaaaatga attgtggtcc attcagataa 3480
cagaatatta ttcattctgc taaaagtga gctatgaagc catagaaagc aaggagaaat 3540
ttttgtcact aagtgaagaa cctaactctgg gaaggttcct atgctgtcag tattatatca 3600
ccttgtcaca gctggaggca ccagagagga ggagcctcag atcaggctgc agggcatttt 3660
ctaatagacta attgatggag gagagcccag cccattgtgg gtggagctat ccctgggctg 3720
gtggtcctgt attctataag aaagcagact aagacatgag ggacaagcca taagcagcac 3780
acctccatgg cttctgcacg agcttctgcc tccaggttcc taccctgttt gaggctcctgt 3840
catgtcactt cagtgggtta ctatgtgaat gtgtgagcca aataattttc cttctgaact 3900
tgcttttggg gttgggtgtc caccaagggc aacagtaacc ttatgacagc tccatactgc 3960
cctgctttgg attctgattc tgaagaggc aaaactatag tgtaggaac aagattattg 4020
actcaagagc tgggggagag aggaggagaa aagaagagcc tagaagattc ttagggtagt 4080
gaagctattc tgaatgatac cacagcggtg gatatgcctg gcattgtgct tttgcacatt 4140
tcccaaagtg aaccttagta gaaactatgt tgcagtgga agtgacaact attacaaacc 4200
tgctgtctgg agcatggtgg tcctagtagg aaggttgttc atttgtctag gcagaggca 4260
tttgggtact cccaaacttt cagctcagtt ttgctcagaa ttcaaaatac agaagtatta 4320
aaatatagcc agactgtagc ttagatggtt gactaagctt tgcaaggag ctctaggttt 4380
gatcccaggc accacaaaaa ctgcacgtgt tgtggcatgt ttgtaatcct agcaggagta 4440
tgagaagttg aaggctcattc ttggctgcat agtgagtttg aggctagcct gacatacatg 4500
ataacctgic tcaaaaaaaaa aagttaaatt aagagaatca gagactgttc cagtatttgg 4560
atggatttgt aacatcatta tttgtctgga tgatacactg gtgtttgagg taaccagtag 4620
gaggtttcta gaatggccag gcaatgtggg gatttccatc ttccaccatt ccctgcatca 4680
cagaggatga tgggacggtg ccggtgggag tcgatctcca tcctacacac cattcctcgt 4740
atcacaaggg aggggtgggac agtgctatgc ttagacttaa caaaacacag agaagataag 4800
gttgttatga atgcccgtg gacatgggca gcatacttgc tttgcatttg gcctggcagt 4860
agcgacactg gcctcagggt attccagggc tacaccttct ttgatggcag gcttccaaat 4920
ttgagaacag agccttaccg tgaggcttac tgcaaagcgg gtagtgtgaa acacatacct 4980
ctcaggctgg cctgcttcca agagcaggag gacagggcta tgcatgcca gttgttctta 5040
cagagttggc agtggcagaa gttattccac tttccctgag gactgggaaa ggaccagctg 5100
gtggagagga gccagggtgc tgaactagca tgtgtccccg taagggtact agtgatggtt 5160
aaacatgtcc ccctctgttt caggtacata gagcattatt tggagattt tctggagtct 5220
gctaacaggc acttcatggt tggccacagg gtcataatct atgtcatggt ggacgatgcc 5280
tccagaatgc cttctataca cctgagtcct ctgcattcct tacaagtgtt tgagatcaag 5340

ES 2 605 677 T3

tccgagaagc gatggcagga catcagcatg atgcggatga agaccatcgg ggagcacatc 5400
ctgtcccaca tacagcacga ggtcgcacttc ctcttctgca tggacgtgga tcaggtcttt 5460
caagacaatt tcggtgtgga gactctgggc cagttggtag ctcagctcca ggcctggtgg 5520
tacaaggccg atcgtgacaa gtttacctat gagaggcgga aactgtcggc ggcgtacatc 5580
ccgtttgggc agggggattt ttattaccac gcagccattt ttggaggaac acccattcac 5640
attctcaacc tcacccgaga gtgctttgag ggaatcctcc aggacaagag caatgacatc 5700
gaagcagagt ggcacgatga gagccacctc aacaaatact tccttttcaa caaacccact 5760
aaaatcttat ctccagaata ctggtgggac tatcagatag gcttgccttc agaaattaa 5820
aatgtaaagg tagcctggca gacaaaagag tataatttgg ttaggagtaa tgtctga 5877

<210> 6

<211> 828

5 <212> ADN

<213> *Cricetulus griseus*

<400> 6

aaccgcccag aagttttgac agtgacccca tggaaaggcg caattgtgtg ggaaggcact 60
tatgacagag ctctgctgga aaagtattat gccagacaga aaatcacctg ggggctgacg 120
gtttttgctg tgggaaagta catagagcat tatttggag attttctgga gtctgctaac 180
aggacttca tggttggcca cagggtcata ttttatgtca tgggtggacga tgcctccaga 240
atgccttcta tacacctgag tcctctgcat tccttacaag tgtttgagat caagtccgag 300
aagcgatggc aggacatcag catgatgcgg atgaagacca tcggggagca catcctgtcc 360
cacatacagc acgaggtcga cttcctcttc tgcatggacg tggatcaggt ctttcaagac 420
aatttcggtg tggagactct gggccagttg gtactcagc tcaggcctg gtggtacaag 480
gccgatcgtg acaagtttac ctatgagagg cggaaactgt cggcggcgta catcccgttt 540
gggcaggggg atttttatta ccacgcagcc atttttggag gaacacccat tcacattctc 600
aacctcaccg gagagtgtt tgaggaatc ctccaggaca agagcaatga catcgaagca 660
gagtggcagc atgagagcca cctcaacaaa tacttctttt tcaacaaacc cactaaaatc 720
ttatctccag aatactgttg ggactatcag ataggcttgc cttcagaaat taaaaatgta 780
aaggtagcct ggcagacaaa agagtataat ttggttagga gtaatgtc 828

10

<210> 7

<211> 276

<212> PRT

15 <213> *Cricetulus griseus*

ES 2 605 677 T3

<400> 7

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

<210> 8

<211> 24

<212> ADN

<213> Secuencia artificial

5

<220>

<223> Cebador de Ggta químicamente sintetizado

<400> 8

10 tggatcacag gagaaaataa tga 24

<210> 9

<211> 27

<212> ADN

15

<213> Secuencia artificial

<220>

<223> Cebador de Ggta químicamente sintetizado

20

<400> 9

aagtttccat cacaattga agtcaga 27

<210> 10

<211> 28

25

<212> ADN

<213> Secuencia artificial

<220>

<223> Cebador de Ggta químicamente sintetizado

30

<400> 10

gaaccgccca gaagtttga cagtgacc 28

<210> 11

35

<211> 29

<212> ADN

<213> Secuencia artificial

ES 2 605 677 T3

<220>

<223> Cebador de Ggta químicamente sintetizado

<400> 11

5 gtcagacatt actcctaacc aaattatac 29

<210> 12

<211> 19

<212> ADN

10 <213> Secuencia artificial

<220>

<223> Cebador de Ggta químicamente sintetizado

15 <400> 12

atgcctccag aatgccttt 19

<210> 13

<211> 20

20 <212> ADN

<213> Secuencia artificial

<220>

<223> Cebador de Ggta químicamente sintetizado

25

<400> 13

atccacgtcc atacagaaga 20

<210> 14

30 <211> 19

<212> ADN

<213> Secuencia artificial

<220>

35 <223> Cebador de Ggta químicamente sintetizado

<400> 14

ES 2 605 677 T3

gccagacaga aaatcaccg 19

<210> 15

<211> 21

5 <212> ADN

<213> Secuencia artificial

<220>

<223> Cebador de Ggta químicamente sintetizado

10

<400> 15

aaagacctga tccacgtcca t 21

<210> 16

15 <211> 20

<212> ADN

<213> Secuencia artificial

<220>

20 <223> Cebador de Ggta químicamente sintetizado

<400> 16

tggaaggca cttatgacag 20

25 <210> 17

<211> 20

<212> ADN

<213> Secuencia artificial

30 <220>

<223> Cebador de Ggta químicamente sintetizado

<400> 17

ttgtaaggaa tgcagaggac 20

35

<210> 18

<211> 19

<212> ADN

<213> Secuencia artificial

<220>

5 <223> Cebador de Ggta químicamente sintetizado

<400> 18

gatgcctcca gaatgcctt 19

10 <210> 19

<211> 20

<212> ADN

<213> Secuencia artificial

15 <220>

<223> Cebador de Ggta químicamente sintetizado

<400> 19

ctcattcttg aagctatgcc 20

20

<210> 20

<211> 19

<212> ADN

<213> Secuencia artificial

25

<220>

<223> Cebador de Ggta químicamente sintetizado

<400> 20

30 gccagacaga aaatcaccg 19

<210> 21

<211> 19

<212> ADN

35 <213> Secuencia artificial

<220>

<223> Cebador de Ggta químicamente sintetizado

<400> 21

cttacgggga cacatgcta 19

5

<210> 22

<211> 18

<212> ADN

<213> Secuencia artificial

10

<220>

<223> Cebador de Ggta químicamente sintetizado

<400> 22

15

ggcacagaag ggaaagac 18

<210> 23

<211> 21

<212> ADN

20

<213> Secuencia artificial

<220>

<223> Cebador de Ggta químicamente sintetizado

25

<400> 23

gtcaaccatc taagctacag t 21

<210> 24

<211> 21

30

<212> ADN

<213> Secuencia artificial

<220>

<223> Cebador de Ggta químicamente sintetizado

35

<400> 24

aaatctgacc ctaatcctgt g 21

<210> 25

<211> 20

<212> ADN

<213> Secuencia artificial

5

<220>

<223> Cebador de Ggta químicamente sintetizado

<400> 25

10

ctccttgctt tctatggctt 20

<210> 26

<211> 18

<212> ADN

15

<213> Secuencia artificial

<220>

<223> Cebador de Ggta químicamente sintetizado

20

<400> 26

gccagacaga aaatcacc 18

<210> 27

<211> 22

25

<212> ADN

<213> Secuencia artificial

<220>

<223> Cebador de Ggta químicamente sintetizado

30

<400> 27

gtctccttcc tagtaactca aa 22

<210> 28

35

<211> 19

<212> ADN

<213> Secuencia artificial

ES 2 605 677 T3

<220>

<223> Cebador de Ggta químicamente sintetizado

<400> 28

5 ctagcatgtg tccccgtaa 19

<210> 29

<211> 16

<212> ADN

10 <213> Secuencia artificial

<220>

<223> Cebador de Ggta químicamente sintetizado

15 <400> 29

gctggtgggt cccgag 16

<210> 30

<211> 21

20 <212> ADN

<213> Secuencia artificial

<220>

<223> Cebador de Ggta químicamente sintetizado

25

<400> 30

aaagacctga tccacgtcca t 21

<210> 31

30 <211> 20

<212> ADN

<213> Secuencia artificial

<220>

35 <223> Cebador de Ggta químicamente sintetizado

<400> 31

ES 2 605 677 T3

gaagtttga cagtgacccc 20

<210> 32

<211> 23

5

<212> ADN

<213> Secuencia artificial

<220>

<223> Cebador de Ggta químicamente sintetizado

10

<400> 32

gttttgctg tgggaaagta ca 23

<210> 33

15

<211> 20

<212> ADN

<213> Secuencia artificial

<220>

20

<223> Cebador de Ggta químicamente sintetizado

<400> 33

ttgtaaggaa tgcagaggac 20

25

<210> 34

<211> 20

<212> ADN

<213> Secuencia artificial

30

<220>

<223> Cebador de Ggta químicamente sintetizado

<400> 34

aaagacctga tccacgtcca 20

35

<210> 35

<211> 24

<212> ADN

<213> Secuencia artificial

<220>

5 <223> Cebador de Ggta químicamente sintetizado

<400> 35

gatctcaaac acttgaagg aatg 24

10 <210> 36

<211> 20

<212> ADN

<213> Secuencia artificial

15 <220>

<223> Cebador de Ggta químicamente sintetizado

<400> 36

tggaaggca ctatgacag 20

20

<210> 37

<211> 23

<212> ADN

<213> Secuencia artificial

25

<220>

<223> Cebador de Ggta químicamente sintetizado

<400> 37

30 aaattccaga gattggtgac agc 23

<210> 38

<211> 21

<212> ADN

35 <213> Secuencia artificial

<220>

<223> Cebador de Ggta químicamente sintetizado

<400> 38

gtcataagtg ccttcccaca c 21

5

<210> 39

<211> 20

<212> ADN

<213> Secuencia artificial

10

<220>

<223> Cebador de Ggta químicamente sintetizado

<400> 39

15

ttattaccac gcagccattt 20

<210> 40

<211> 23

<212> ADN

20

<213> Secuencia artificial

<220>

<223> Cebador de Ggta químicamente sintetizado

25

<400> 40

ttccatcaca atttgaagtc aga 23

<210> 41

<211> 29

30

<212> ADN

<213> Secuencia artificial

<220>

<223> Cebador de Ggta químicamente sintetizado

35

<400> 41

gtcagacatt actcctaacc aaattatac 29

<210> 42

<211> 25

<212> ADN

<213> Secuencia artificial

5

<220>

<223> Cebador de Ggta químicamente sintetizado

<400> 42

10 aatgaatgtc aagggaaaag tggtc 25

<210> 43

<211> 21

<212> ADN

15 <213> Secuencia artificial

<220>

<223> Cebador de Ggta químicamente sintetizado

20 <400> 43

tgatagtccc aacagtactc t 21

<210> 44

<211> 20

25 <212> ADN

<213> Secuencia artificial

<220>

<223> Cebador de Ggta químicamente sintetizado

30

<400> 44

cacatcctga cccacataca 20

<210> 45

35 <211> 20

<212> ADN

<213> Secuencia artificial

<220>

<223> Cebador de Ggta químicamente sintetizado

<400> 45

5 gacagtccc gcctctcata 20

<210> 46

<211> 19

<212> ADN

10 <213> Secuencia artificial

<220>

<223> Cebador de Ggta químicamente sintetizado

15 <400> 46

atgcctccag aatgcctt 19

<210> 47

<211> 20

20 <212> ADN

<213> Secuencia artificial

<220>

<223> Cebador de Ggta químicamente sintetizado

25

<400> 47

atccacgtcc atacagaaga 20

<210> 48

30 <211> 23

<212> ADN

<213> Secuencia artificial

<220>

35 <223> Cebador de Ggta químicamente sintetizado

<400> 48

ES 2 605 677 T3

gtacattgag cattattgg aag 23

<210> 49

<211> 23

5 <212> ADN

<213> Secuencia artificial

<220>

<223> Cebador de Ggta químicamente sintetizado

10

<400> 49

gtacatcgag cattattgg aag 23

<210> 50

15 <211> 28

<212> ADN

<213> Secuencia artificial

<220>

20 <223> Cebador de Ggta químicamente sintetizado

<400> 50

gaaccgccca gaagtttga cagtgacc 28

25 <210> 51

<211> 16

<212> ADN

<213> Secuencia artificial

30 <220>

<223> Cebador de Ggta químicamente sintetizado

<400> 51

gctggtggtt cccgag 16

35

<210> 52

<211> 22

<212> ADN

<213> Secuencia artificial

<220>

5 <223> Cebador de Ggta químicamente sintetizado

<400> 52

gtctccttc tagtaactca aa 22

10 <210> 53

<211> 23

<212> ADN

<213> Secuencia artificial

15 <220>

<223> Cebador de Ggta químicamente sintetizado

<400> 53

atgaaaattc cagagattgg tga 23

20

<210> 54

<211> 22

<212> ADN

<213> Secuencia artificial

25

<220>

<223> Cebador de Ggta químicamente sintetizado

<400> 54

30 gggcaccac agttatcaag aa 22

<210> 55

<211> 23

<212> ADN

35 <213> Secuencia artificial

<220>

ES 2 605 677 T3

<223> Cebador de Ggta químicamente sintetizado

<400> 55

gtcagaattg aggagcctca ctt 23

5

<210> 56

<211> 368

<212> PRT

<213> Bos taurus

10

<400> 56

ES 2 605 677 T3

Met Asn Val Lys Gly Lys Val Ile Leu Ser Met Leu Val Val Ser Thr
1 5 10 15

Val Ile Val Val Phe Trp Glu Tyr Ile His Ser Pro Glu Gly Ser Leu
20 25 30

Phe Trp Ile Asn Pro Ser Arg Asn Pro Glu Val Gly Gly Ser Ser Ile
35 40 45

Gln Lys Gly Trp Trp Leu Pro Arg Trp Phe Asn Asn Gly Tyr His Glu
50 55 60

Glu Asp Gly Asp Ile Asn Glu Glu Lys Glu Gln Arg Asn Glu Asp Glu
65 70 75 80

Ser Lys Leu Lys Leu Ser Asp Trp Phe Asn Pro Phe Lys Arg Pro Glu
85 90 95

Val Val Thr Met Thr Lys Trp Lys Ala Pro Val Val Trp Glu Gly Thr
100 105 110

Tyr Asn Arg Ala Val Leu Asp Asn Tyr Tyr Ala Lys Gln Lys Ile Thr
115 120 125

Val Gly Leu Thr Val Phe Ala Val Gly Arg Tyr Ile Glu His Tyr Leu
130 135 140

Glu Glu Phe Leu Thr Ser Ala Asn Lys His Phe Met Val Gly His Pro
145 150 155 160

Val Ile Phe Tyr Ile Met Val Asp Asp Val Ser Arg Met Pro Leu Ile
165 170 175

Glu Leu Gly Pro Leu Arg Ser Phe Lys Val Phe Lys Ile Lys Pro Glu
180 185 190

Lys Arg Trp Gln Asp Ile Ser Met Met Arg Met Lys Thr Ile Gly Glu
195 200 205

His Ile Val Ala His Ile Gln His Glu Val Asp Phe Leu Phe Cys Met
210 215 220

Asp Val Asp Gln Val Phe Gln Asp Lys Phe Gly Val Glu Thr Leu Gly
225 230 235 240

Glu Ser Val Ala Gln Leu Gln Ala Trp Trp Tyr Lys Ala Asp Pro Asn
245 250 255

ES 2 605 677 T3

Asp Phe Thr Tyr Glu Arg Arg Lys Glu Ser Ala Ala Tyr Ile Pro Phe
260 265 270

Gly Glu Gly Asp Phe Tyr Tyr His Ala Ala Ile Phe Gly Gly Thr Pro
275 280 285

Thr Gln Val Leu Asn Ile Thr Gln Glu Cys Phe Lys Gly Ile Leu Lys
290 295 300

Asp Lys Lys Asn Asp Ile Glu Ala Gln Trp His Asp Glu Ser His Leu
305 310 315 320

Asn Lys Tyr Phe Leu Leu Asn Lys Pro Thr Lys Ile Leu Ser Pro Glu
325 330 335

Tyr Cys Trp Asp Tyr His Ile Gly Leu Pro Ala Asp Ile Lys Leu Val
340 345 350

Lys Met Ser Trp Gln Thr Lys Glu Tyr Asn Val Val Arg Asn Asn Val
355 360 365

<210> 57

<211> 370

5 <212> PRT

<213> Canis familiaris

<400> 57

ES 2 605 677 T3

Met Asn Val Lys Gly Lys Val Ile Leu Ser Met Leu Val Val Ser Thr
 1 5 10 15

Val Ile Val Val Phe Trp Glu Tyr Ile Asn Ser Pro Glu Gly Ser Phe
 20 25 30

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

Gln Lys Gly Trp Trp Phe Pro Asn Trp Phe Asn Asn Gly Thr His Phe
 50 55 60

Tyr Gln Glu Glu Glu Asp Ile Asp Asp Glu Asn Glu Gln Gly Glu Glu
 65 70 75 80

Asn Asn Ala Glu Leu Gln Leu Ser Asp Trp Phe Asn Pro Gln Lys Arg
 85 90 95

Pro Glu Val Val Thr Val Thr Arg Trp Lys Ala Pro Val Val Trp Glu
 100 105 110

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

ES 2 605 677 T3

Ile Thr Ile Gly Leu Thr Val Phe Ala Val Gly Arg Tyr Ile Glu His
 130 135 140

Tyr Leu Glu Glu Phe Leu Ile Ser Ala Asn Arg Tyr Phe Met Val Gly
 145 150 155 160

His Lys Val Ile Phe Tyr Ile Met Val Asp Asp Val Ser Arg Met Pro
 165 170 175

Leu Val Glu Leu Gly Pro Leu Arg Ser Phe Lys Val Phe Glu Ile Glu
 180 185 190

Pro Glu Lys Arg Trp Gln Asp Ile Ser Met Met Arg Met Lys Thr Ile
 195 200 205

Gly Glu His Ile Val Ala His Ile Gln His Glu Val Asp Phe Leu Phe
 210 215 220

Cys Met Asp Val Asp Gln Val Phe Gln Asp Ser Phe Gly Val Glu Thr
 225 230 235 240

Leu Gly Gln Ser Val Ala Gln Leu Gln Ala Trp Trp Tyr Lys Ala Asp
 245 250 255

Pro Asp Glu Phe Thr Tyr Glu Arg Arg Lys Glu Ser Ala Ala Tyr Ile
 260 265 270

Pro Phe Gly Gln Gly Asp Phe Tyr Tyr His Ala Ala Ile Phe Gly Gly
 275 280 285

Thr Pro Ile Gln Val Leu Asn Ile Thr Gln Glu Cys Phe Lys Gly Ile
 290 295 300

Leu Gln Asp Lys Lys Asn Asp Ile Glu Ala Glu Trp His Asp Glu Ser
 305 310 315 320

His Leu Asn Lys Tyr Phe Leu Leu Asn Lys Pro Thr Lys Ile Leu Ser
 325 330 335

Pro Glu Tyr Cys Trp Asp Tyr His Ile Gly Leu Pro Ser Asp Ile Lys
 340 345 350

Thr Val Lys Ile Ser Trp Gln Thr Lys Glu Tyr Asn Leu Val Arg Asn
 355 360 365

Asn Ile
 370

<210> 58

<211> 337

5 <212> PRT

<213> Rattus norvegicus

ES 2 605 677 T3

<400> 58

Met Asn Val Lys Gly Lys Ile Ile Leu Ser Val Leu Met Val Ser Thr
 1 5 10 15

Val Leu Val Val Phe Trp Glu Tyr Val Asn Arg Thr His Ser Tyr Gln
 20 25 30

Glu Glu Asp Ile Glu Arg Ala Arg Glu Lys Gly Arg Asn Gly Asp Ser
 35 40 45

Ile Val Glu Pro Gln Leu Trp Asp Trp Phe Asn Pro Lys Asn Arg Pro
 50 55 60

Glu Val Leu Thr Val Thr Pro Trp Lys Ala Pro Ile Val Trp Glu Gly
 65 70 75 80

Thr Tyr Asp Thr Ala Leu Leu Glu Lys Tyr Tyr Ala Arg Gln Lys Ile
 85 90 95

Thr Val Gly Leu Thr Val Phe Ala Val Gly Lys Tyr Ile Glu His Tyr
 100 105 110

Leu Glu Asp Phe Leu Glu Ser Ala Asn Lys Tyr Phe Met Val Gly His
 115 120 125

Arg Val Ile Phe Tyr Val Met Met Asp Asp Thr Ser Arg Met Pro Ala
 130 135 140

Val His Leu Ser Pro Leu His Ser Leu Gln Val Phe Glu Ile Arg Ser
 145 150 155 160

Glu Lys Arg Trp Gln Asp Ile Ser Met Met Arg Met Lys Thr Ile Gly
 165 170 175

Glu His Ile Leu Asp His Ile Gln His Glu Val Asp Phe Leu Phe Cys
 180 185 190

Met Asp Val Asp Gln Val Phe Gln Asp Asn Phe Gly Val Glu Thr Leu
 195 200 205

Gly Gln Leu Val Ala Gln Leu Gln Ala Trp Trp Tyr Lys Ala Ser Pro
 210 215 220

Asp Glu Phe Thr Tyr Glu Arg Arg Glu Leu Ser Ala Ala Tyr Ile Pro

ES 2 605 677 T3

Glu Gly Thr Tyr Asp Thr Ala Leu Leu Glu Lys Tyr Tyr Ala Thr Gln
 115 120 125
 Lys Leu Thr Val Gly Leu Thr Val Phe Ala Val Gly Lys Tyr Ile Glu
 130 135 140
 His Tyr Leu Glu Asp Phe Leu Glu Ser Ala Asp Met Tyr Phe Met Val
 145 150 155 160
 Gly His Arg Val Ile Phe Tyr Val Met Ile Asp Asp Thr Ser Arg Met
 165 170 175
 Pro Val Val His Leu Asn Pro Leu His Ser Leu Gln Val Phe Glu Ile
 180 185 190
 Arg Ser Glu Lys Arg Trp Gln Asp Ile Ser Met Met Arg Met Lys Thr
 195 200 205
 Ile Gly Glu His Ile Leu Ala His Ile Gln His Glu Val Asp Phe Leu
 210 215 220
 Phe Cys Met Asp Val Asp Gln Val Phe Gln Asp Asn Phe Gly Val Glu
 225 230 235 240
 Thr Leu Gly Gln Leu Val Ala Gln Leu Gln Ala Trp Trp Tyr Lys Ala
 245 250 255
 Ser Pro Glu Lys Phe Thr Tyr Glu Arg Arg Glu Leu Ser Ala Ala Tyr
 260 265 270
 Ile Pro Phe Gly Glu Gly Asp Phe Tyr Tyr His Ala Ala Ile Phe Gly
 275 280 285
 Gly Thr Pro Thr His Ile Leu Asn Leu Thr Arg Glu Cys Phe Lys Gly
 290 295 300
 Ile Leu Gln Asp Lys Lys His Asp Ile Glu Ala Gln Trp His Asp Glu
 305 310 315 320
 Ser His Leu Asn Lys Tyr Phe Leu Phe Asn Lys Pro Thr Lys Ile Leu
 325 330 335
 Ser Pro Glu Tyr Cys Trp Asp Tyr Gln Ile Gly Leu Pro Ser Asp Ile
 340 345 350
 Lys Ser Val Lys Val Ala Trp Gln Thr Lys Glu Tyr Asn Leu Val Arg
 355 360 365
 Asn Asn Val
 370

REIVINDICACIONES

1. Una molécula de ácido nucleico aislada que codifica un polipéptido que produce estructuras de Gal-alfa-Gal, en la que la molécula de ácido nucleico aislada está seleccionado del grupo que consiste en:
 - 5 (i) una molécula de ácido nucleico aislada que comprende una secuencia de nucleótidos que tiene al menos el 90 % de identidad de secuencias con (a) una molécula de ADN que tiene una secuencia que codifica la secuencia de restos de aminoácidos de SEQ ID NO: 2, o SEQ ID NO: 4 o SEQ ID NO: 7 o (b) el complemento de la molécula de ADN de (a); y
 - (ii) una molécula de ácido nucleico aislada que tiene al menos el 90 % de identidad de secuencias con la secuencia de SEQ ID NO: 1 o SEQ ID NO: 3 o SEQ ID NO: 5 o SEQ ID NO: 6, o un complemento de la misma.
- 10 2. La molécula de ácido nucleico aislada de la reivindicación 1, que es una molécula de ácido nucleico aislada que comprende una secuencia de nucleótidos que tiene al menos el 95 %, 96 %, 97 %, 98 % o el 99 % de identidad de secuencias con (a) una molécula de ADN que tiene una secuencia que codifica la secuencia de restos de aminoácidos de SEQ ID NO: 2, o SEQ ID NO: 4 o SEQ ID NO: 7 o (b) el complemento de la molécula de ADN de (a).
- 15 3. La molécula de ácido nucleico aislada de la reivindicación 1, que es una molécula de ácido nucleico aislada que codifica la secuencia de SEQ ID NO: 2, o SEQ ID NO: 4 o SEQ ID NO: 7.
4. La molécula de ácido nucleico aislada de la reivindicación 1, que es una molécula de ácido nucleico aislada que comprende la secuencia de nucleótidos de SEQ ID NO: 1 o SEQ ID NO: 3, SEQ ID NO: 5 o SEQ ID NO: 6, o un complemento de la misma.
- 20 5. La molécula de ácido nucleico aislada de una cualquiera de las reivindicaciones 1-4, en la que la molécula de ácido nucleico codifica un polipéptido Ggta1 enlazado a una secuencia de aminoácidos heteróloga.
6. Una construcción de ácido nucleico que comprende la secuencia de una molécula de ácido nucleico de una cualquiera de las reivindicaciones 1-5.
7. Una célula huésped aislada transfectada con una molécula de ácido nucleico o construcción de ácido nucleico de una cualquiera de las reivindicaciones precedentes.
- 25 8. La célula huésped aislada de la reivindicación 7, en la que la célula huésped produce una glucoproteína terapéutica recombinante que tiene: (i) niveles elevados de glucanos de galactosa- α 1,3-galactosa terminal en comparación con la glucoproteína terapéutica recombinante producida por células parentales de la célula huésped; o (ii) niveles más bajos de glucanos de galactosa- α 1,3-galactosa terminal en comparación con la glucoproteína terapéutica recombinante producida por las células parentales de la célula huésped.
- 30 9. Un método de modulación de la estructura de glucano de una glucoproteína terapéutica recombinante producida en una célula CHO, comprendiendo el método cultivar la célula huésped de la reivindicación 8 en condiciones suficientes para expresar la glucoproteína.
10. Un polipéptido aislado que comprende una secuencia de aminoácidos que tienen al menos el 90 % de identidad de secuencias con SEQ ID NO: 2, SEQ ID NO: 4 o SEQ ID NO: 7, en el que el polipéptido produce estructuras de Gal-alfa-Gal.
- 35 11. El polipéptido aislado de la reivindicación 10, que comprende una secuencia de aminoácidos que se diferencia de SEQ ID NO: 2, SEQ ID NO: 4 o SEQ ID NO: 7 por al menos uno, pero menos de 20 restos de aminoácidos.
12. El polipéptido aislado de la reivindicación 10, que comprende la secuencia de aminoácidos de SEQ ID NO: 2, SEQ ID NO: 4 o SEQ ID NO: 7.
- 40 13. Un anticuerpo aislado, o fragmento de unión al antígeno del mismo, que se une específicamente al polipéptido de la reivindicación 12.
14. Un método de detección de la expresión o actividad de Ggta1 en una población de células CHO, comprendiendo el método:
 - 45 (i) poner en contacto un muestra de ácido nucleico de la población de células CHO con la molécula de ácido nucleico de una cualquiera de las reivindicaciones 1-5; o
 - (ii) poner en contacto una muestra de la población de células CHO con un anticuerpo, o fragmento de unión al antígeno del mismo, que se une específicamente al polipéptido de la reivindicación 10, 11 o 12;
 para así detectar la expresión o actividad de Ggta1 en la población de células CHO.

15. El método de la reivindicación 14, en el que el método comprende poner en contacto una muestra de ácido nucleico de la población de células CHO con la molécula de ácido nucleico de una cualquiera de las reivindicaciones 1-5; y amplificar una secuencia de ácidos nucleicos en el muestra de ácido nucleico usando PCR.
- 5 16. El método de la reivindicación 14, que comprende además cuantificar el nivel de expresión de Ggta1 en la población de células CHO, comprendiendo el método opcionalmente además comparar el nivel de expresión de Ggta1 en la población de células CHO con un nivel de referencia, un nivel de control, un nivel de predeterminado o un nivel presentado por una segunda población de células CHO.
- 10 17. Una matriz bidimensional que tiene una pluralidad de direcciones, siendo cada dirección de la pluralidad posicionalmente distinguible de cada otra dirección de la pluralidad, y teniendo cada dirección de la pluralidad una sonda de captura única, en la que al menos una dirección de la pluralidad tiene una sonda de captura que comprende una molécula de ácido nucleico de una cualquiera de las reivindicaciones 1-5, o el anticuerpo o fragmento de unión al antígeno del mismo de la reivindicación 13.
18. Un método de análisis de una muestra de una célula CHO, comprendiendo el método poner en contacto la muestra con la matriz de la reivindicación 17.

15

FIGURA 3A-I

Ggtal-a

SEQ ID NO:1: ATG AAT GTC AAG GGA AAA GTG GTC CTG TCG ATG CTG GTT GTC TCA ACT GTG CTT GTC GTG 60
 SEQ ID NO:2: M N V K G K V V L S M L V V S T V L V V 20

61 TTT TGG GAA TAT GTC AAC AGC CCA GAA GGT TCC TTC TTG TGG ATA TAT CAC ATG AAA ATT 120
 21 F W E Y V N S P E G S F L W I Y H M K I 40

121 CCA GAG ATT GGT GAC AGC AGC AGG CAG AGC TGG TGG TTC CCG AGC TGG TTT AGC AAT GGG 180
 41 P E I G D S S R Q S W W F P S W F S N G 60

181 ACC CAC AGT TAT CAA GAA GAG GAC GTA GAA AGA GGG AGA GAA AAG CAG CGA AAT GCA GTC 240
 61 T H S Y Q E E D V E R G R E K Q R N A V 80

241 AGA ATT GAA GAG CCT CAC TTG TGG GAC TGG TTT AAT CCA AAG AAC CGC CCA GAA GTT TTG 300
 81 R I E E P H L W D W F N P K N R P E V L 100

301 ACA GTG ACC CCA TGG AAG GCA CCA ATT GTG TGG GAA GGC ACT TAT GAC AGA GCT ATG CTG 360
 101 T V T P W K A P I V W E G T Y D R A M L 120

361 GAA AAG TAT TAC GCC AGA CAG AAA ATC ACC GTG GGA CTG ACA GTT TTT GCT GTG GGA AAG 420
 121 E K Y Y A R Q K I T V G L T V F A V G K 140

FIGURA 3A-2

421 TAC Att GAG CAT TAT TTG GAA GAT TTC CTg GAG TCT GCT AGC AGG TAC TTC ATG GTT GGC 480
 141 Y I E H Y L E D F L E S A S R Y F M V G 160

481 CAC AGG GTC ATA TTT TAT GTC ATG ATG GAC GAT GCC TCC AGA ATG CCT TTT ATA CAC CTG 540
 161 H R V I F Y V M D D A S R M P F I H L 180

541 AGT CCT CTG CAT TCC TTA CAA GTG TTT GAG ATC AAG TCT GAG ACC AGG TGG CAA GAC ATT 600
 181 S P L H S L Q V F E I K S E T R W Q D I 200

601 AGC ATG ATG CGC ATG AAG ACA ATC GAG GAG CAC ATC CTG ACC CAC ATA CAG CAT GAG GTC 660
 201 S M M R M K T I E E H I L T H I Q H E V 220

661 GAC TTC CTC TTC TGT ATG GAC GTG GAT CAG GTC TTT CAA GAC AAT TTT GGA GTG GAG ACT 720
 221 D F L F C M D V D Q V F Q D N F G V E T 240

721 CTG GGC CAG TTG GTA GCT CAG CTC CAG GCC TGG TGG TAC AAA GCT GAC Cat GAC AAG TTT 780
 241 L G Q Q L V A Q L Q A W Y K A D H D K F 260

FIGURA 3B

781 ACC TAT GAG AGG CGG GAA CTG TCG GCG GCA TAC ATC CCA TTT GGA CAG GGG GAT TTT TAT 840
 261 T Y E R R E L S A A Y I P F G Q G D F Y 280

841 TAC Cacc GCA GCC ATT TTC GGA GGG ACA CCC ATT CAT ATT CTC AAC CTC ACC CGG GAG TGC 900
 281 Y H A A A I F G G T P I H I L N L T R E C 300

901 TTT GAG GGG ATC CTC CAG GAC AAG AAC AAC ATA GAA GCA AAG TGG CAT GAC GAA AGT 960
 301 F E G I L Q D K N N I E A K W H D E S 320

961 CAC CTA AAC AAA TAT TTC CTT TTC AAC AAG CCC ACT AAA ATT TTA TCT CCA GAG TAC TGT 1020
 321 H L N K Y F L F N K P T K I L S P E Y C 340

1021 TGG GAC TAT CAG ATA GGC TTG CCT TCA GAA ATT AAA AAT GTA AAG ATA GCT TGG CAG ACA 1080
 341 W D Y Q I G L P S E I K N V K I A W Q T 360

1081 AAA GAG TAT AAT TTG GTT AGG AGT AAT GTC 1110
 361 K E Y N L V R S N V 370

FIGURA 4A-1

Ggtal-b

SEQ ID NO:3: ATG AAT GTC AAG GGA AAA GTG GTC CTG TCG ATG CTG GTT GTC TCA ACT GTG CTT GTC GTG 60
 SEQ ID NO:4: M N V K G K V L S M L V V S T V L V V 20

61 TTT TGG GAA TAT GTC AAC AGC CCA GAA GGT TCC TTC TTG TGG ATA TAT CAC ATG AAA ATT 120
 21 F W E Y V N S P E G S F L W I Y H M K I 40

121 CCA GAG ATT GGT GAC AGC AGC AGG CAG AGC TGG TGG TTC CCG AGC TGG TTT AGC AAT GGG 180
 41 P E I G D S S R Q S W F P S W F S N G 60

181 ACC CAC AGT TAT CAA GAA GAG GAC ATA GAA AGA GGG AGA AAG CAG CGA AAT GCA GTC 240
 61 T H S Y Q E E D I E R G R E K Q R N A V 80

241 AGA ATT GAA GAG CCT CAC TTG TGG GAC TGG TTT AAT CCA AAG AAC CGC CCA GAA GTT TTG 300
 81 R I E E P H L W D W F N P K N R P E V L 100

301 ACA GTG ACC CCA TGG AAG GCA CCA ATT GTG TGG GAA GGC ACT TAT GAC AGA GCT ATG CTG 360
 101 T V T P W K A P I V W E G T Y D R A M L 120

361 GAA AAG TAT TAT GCC AGA CAG AAA ATC ACC GTG GGA CTG ACA GTT TTT GCT GTG GGA AAG 420
 121 E K Y Y A R Q K I T V G L T V F A V G K 140

FIGURA 4A-2

421 TAC ATC GAG CAT TAT TTG GAA GAT TTC CTA GAG TCT GCT AGC AGG TAC TTC ATG GTT GGC 480
 141 Y I E H Y L E L E D F L E S A S R Y F M V G 160

481 CAC AGG GTC ATA TTT TAT GTC ATG ATG GAC GAT GCC TCC AGA ATG CCT TTT ATA CAC CTG 540
 161 H R V I F Y V M M D D A S R M P F I H L 180

541 AGT CCT CTG CAT TCC TTA CAA GTG TTT GAG ATC AAG TCT GAG ACC AGG TGG CAA GAC ATT 600
 181 S P L H S L Q V F E I K S E T R W Q D I 200

601 AGC ATG ATG CGC ATG AAG ACA ATC GAG GAG CAC ATC CTG ACC CAC ATA CAG CAT GAG GTC 660
 201 S M M R M K T I E E H I L T H I Q H E V 220

661 GAC TTC CTC TTC TGT ATG GAC GTG GAT CAG GTC TTT CAA GAC AAT TTT GGA GTG GAG ACT 720
 221 D F L F C M D V D Q V F Q D N F G V E T 240

FIGURA 4B

721 CTG GGC CAG TTG GTA GCT CAG CTC CAG GCC TGG TGG TAC AAA GCT GAC CAC GAC AAG TTT 780
 241 L G Q L V A Q L Q A W Y K A D H D K F 260

781 ACC TAT GAG AGG CGG GAA CTG TCG GCG GCA TAC ATC CCA TTT GGA CAG GGG GAT TTT TAT 840
 261 T Y E R R E L S A A Y I P F G Q G D F Y 280

841 TAC CAT GCA GCC ATT TTC GGA GGG ACA CCC ATT CAT ATT CTC AAC CTC ACC CGG GAG TGC 900
 281 Y H A A I F G G T P I H I L N L T R E C 300

901 TTT GAG GGG ATC CTC CAG GAC AAG AAC AAC AAC ATA GAA GCA AAG TGG CAT GAC GAA AGT 960
 301 F E G I L Q D K N N I E A K W H D E S 320

961 CAC CTA AAC AAA TAT TTC CTT TTC AAC AAG CCC ACT AAA ATT TTA TCT CCA GAG TAC TGT 1020
 321 H L N K Y F L F N K P T K I L S P E Y C 340

1021 TGG GAC TAT CAG ATA GGC TTG CCT TCA GAA ATT AAA AAT GTA AAG ATA GCT TGG CAG ACA 1080
 341 W D Y Q I G L P S E I K N V K I A W Q T 360

1081 AAA GAG TAT AAT TTG GTT AGG AGT AAT GTC 1110
 361 K E Y N L V R S N V 370

FIGURA 5A-2

tgtttccccgaacttcaggcacactagtgttttcttgaagttcagagtacagaagtctgtcttgacctccagccatggagtcattgtcagagctctcttttaccactggg
 atatatgtgcccccccaagaatcagtagtagaaaccttgagctttgcaatgccaacagtaaaagtgaattttcgatggcaaaagaaaccaaccocaaagcatcaaga
 aaaaacaaacagagattggagagatggctcagtggtgatagttggctgacctctagtggaacccagggttcaattcccagcaactcacatggtgtgaggaaaaaagg
 gcagccaaaagaaacacacttggctccctggaatcagaccccagagaagaaatcacatcctggccctaaaaatagagccaaaagactaaaaaggaccagtgaagaaacacat
 actttggccctgaaacccagagccaaaatctgacctaatcctgtgcagaaaaacccaaatccctgaGcaggagatttagccattcagagctcagggattgaaaatctgg
 ccaatccccaccctggaaatctccctggaaaaactccctaaagattccccatataaacccctgtatc
 tgttcagcttgtggctgcctggattcttccctcccaataaaatctcttcaatgatgtttgaggtgagaccttctttcaccagagagcagatcgggtagggctgtaca
 ctttgctgctccttgcogaaactgtaacaaacttaccagggaaacctctcctgcaggagcagagctgatacactgggaaagcogt
 tccctgaagcagggctgtaaagctgctatgcttactgtgacctgtggtattccttggctcctgattgccagaataacctttccatctgagctgtaacactcagttattctt
 cggctttcaagtggctaatccatccatccgatctgtcaaaatgcttacaactggcagctcaacaactgtctgtaactccagttccaggggatctggcaacttccacac
 aggcatacacacaagcaaaaaacacaaaatgcacataaaaataaaatctattggaaaaaaaagagaaaaatgcaagtatctt

FIGURA 5B-1

tttaaaaaaagaaaaaacactagtcattctgcttggcttagttgtgtgctcagtaaacctgtggtcogtagtctctgtgtgtaagtttgaggttggaaaagggaccatct
 gacttccccctcatctctgcagtcctcttggcatgaattaggaataatgctgagcttgataaggcatcagataaatgtcactcaaatgataaaatagaca
 aattgacaaaatctcatctgttagaatggcaaaaaaaactgccaataccaaaatgttgaaaaggtgtgaagtaaacagaatccttgttactgctagtgg
 gaataataatgggtacagccatcttgggagaaaatcttggcagttgcttgcaaaaatggagcatgctcttaccataaactcaataatcctcaaaatccttggtaaacccaaag
 tgagttgaaaaaagtcttacacaaaaaacctgcatatagataatcttataaagctttattcataaattgctaaaaacttgggagcagacaaagtttcccttggacagggcaaatg
 taaaaaatgaatttggttccattcagataaacagaaatatttcatctctgctaaaaagtgagctatgaagccatagaaaagcaaggagaaaatcttctgcaactaaagtgaag
 aacctaatctgggaaaggttctctatgctgtagtataatcaccttgtcacagctggaggcaccagagaggagcctcagatcaggtgcagggcatttctctaatg
 actaatgtgaggagagcccagccatctgggtggagctatccctgggtggtcctgtattctataagaaagcagactaagacatgagggacaaagccataaag
 cagcacacctccatggctctgcatcagcttctgcctccaggttccctaccctgttggagttcctgctatgtcactcagtggttactatgtgaatgtgtgagccaaa
 taatttcccttgaaacttgcttttgggtgttgggtgttccaccaagggcaacagtaaaccttatgacaggtccatactgcccctgcttggattctgattcttggaagagggc
 aaaaactatagtgtaggaaacaagattattgactcaagagctgggggagagagggagaaaaagagagcctagaaagattcttaggggtagtgaagctattctgaaatgat
 accacagcggfggatagcctggcatgtgcttttgcacatttcccaaaagtgaaccttagtagaaaactatggtgtcagtggaagtgacaactattcaaaaaacctgctgt
 ctggagcatggtgggtcctagtaggaaaggttgttcatctgtctaggcaggagggcatttgggtactccccaaaacttccagctcagttttgctcagaattcaaaaaaacagaa
 gtataaaaaatagccagactgtagcttagatgggtgactaaagtctgagtgaggctctaggtttgatccccaggtaccacaaaaaaactgcacgtgttggcagatgttt
 gtaatcctagcaggagatgagaaagtgaaggtcattcttggctgcatagtgagtttgaggctagcctgacatacatgataacctgtctcaaaaaaaagttaaaatt
 aagaaatcagagactgttccagatattggatggatttghtaacatcattatttctggatgatcacactgggtgtttgaggttaaccagtaggggtttgtagaatggcc
 aggcaatgtggggatttccatcttccaccatccctgcatcacagaggatgatgggacgggtgocgggtggagtcgactccatcctacacacccatccctcgtatccaca
 aaggaggggtgggacagtgctatgcttagacttaacaaaaacacagagaagaataaggttghttatgaatgcccgctggacatgggocagcatc

FIGURA 5B-2

ttgctttgcatbtggcctggcagtagcagactggcctcaggglatccagggtcacaccttcttggatggcaggttccaaatbtgagaacagagccttaccgtgag
 gcttactgcaaaagcgggtagttgaaaaacacatacctctcaggtggcctgctccaagagcagagggacaggggtatgcatgcccagttgttcttacagagttggcag
 tggcagaagtattccacttccctgaggactgggaaaggaccagctggtggagaggaccaggggtgctgaaactagcatgtgtcccccgtaagggtactagtgatgggtt
 aacatgtccccctctgtttcaggtACATAGAGCATTAATTGGAAAGATTTTCTGGAGTCTGTAAACAGGTACTTCATGGTTGGCCACACAGGGTCAATATTTTATGTCATG
 GTGGACGATGCCCTCCAGAAATGCCCTTCTATACACCTGACTCTCTGCATTCCTTACAAGTGTGGAGATCAAGTCCGAGAAAGCGATGGCAGGACATCAGCATGATGCCG
 ATGAAGACCATCGGGAGCACATCCTGTCCCACATACAGCACGAGGTGACTTCTCTTCTGCATGGACCTGATCAGGTCTTTTCAAGACAATTTCCGGTGTGGAGACT
 CTGGCCAGTTGGTAGCTCAGCTCCAGGCCCTGGTGGTACAAGGCCGATCGTGACAAGTTTACCTATGAGAGGCCGAAACTGTCCGGCCGCTACATCCCCGTTTGGGCAG
 GGGATTTTTTATTA CCA CGCAGCCA TTTTGGAGGAA CACCATTCACATTTCAA CCTCACCCGAGAGTGCCTTTGAGGGAA TCCCTCCAGGACAAGCAATGACATC
 GAAGCAGATGGCACGATCAGAGCCACCTCAACAATACTTCTCTTTTCAACAACCCACTAAAATCTTTAICTCCAGAAATCTGTTGGGACTATCAGATAGGCTTGCCT
 TCAGAAATTAATAATGTTAAAGGTAGCCTGGCCAGACAAAAGAGTATAAATTTGGTTAGGAGTAATGTCCTGA

FIGURA 6-1

exón 8 y 9 de Gg1a1

SEQ ID NO:6:AAC CGC CCA GAA GTT TTG ACA GTG ACC CCA TGG AAG GCG CCA ATT GTG TGG GAA GGC ACT 60
 SEQ ID NO:7: N R P E V L T V T P W K A P I V W E G T 20

61 TAT GAC AGA GCT CTG GAA AAG TAT TAT GCC AGA CAG AAA ATC ACC GTG GGG CTG ACG 120
 21 Y D R A L L E K Y A R Q K I T V G L T 40

121 GTT TTT GCT GTG GGA AAG TAC ATA GAG CAT TAT TTG GAA GAT TTT CTG GAG TCT GCT AAC 180
 41 V F A V G K Y I E H Y L E D F L E S A N 60

181 AGG TAC TTC ATG GTT GGC CAC AGG GTC ATA TTT TAT GTC ATG GTG GAC GAT GCC TCC AGA 240
 61 R Y F M V G H R V I F Y V M V D D A S R 80

241 ATG CCT TCT ATA CAC CTG AGT CCT CTG CAT TCC TTA CAA GTG TTT GAG ATC AAG TCC GAG 300
 81 M P S I H L S P L H S L Q V F E I K S E 100

301 AAG CGA TGG CAG GAC ATC AGC ATG ATG CGG ATG AAG ACC ATC GGG GAG CAC ATC CTG TCC 360
 101 K R W Q D I S M R M R M K T I G E H I L S 120

361 CAC ATA CAG CAC GAG GTC GAC TTC CTC TGC ATG GAC GTG GAT CAG GTC TTT CAA GAC 420
 121 H I Q H E V D F L F C M D V D Q V F Q D 140

FIGURA 6-2

421 AAT TTC GGT GTG GAG ACT CTG GGC CAG TTG GTA GCT CAG CTC CAG GCC TGG TGG TAC AAG 480
 141 N F G V E T L G Q L V A Q L Q A W Y K 160

481 GCC GAT CGT GAC AAG TTT ACC TAT GAG AGG CGG AAA CTG TCG GCG GCG TAC ATC CCG TTT 540
 161 A D R D K F T Y E R R K L S A A Y I P F 180

541 GGG CAG GGG GAT TTT TAT TAC CAC GCA GCC ATT TTT GGA GGA ACA CCC ATT CAC ATT CTC 600
 181 G Q G D F Y Y H A A I F G G T P I H I L 200

601 AAC CTC ACC CGA GAG TGC TTT GAG GGA ATC CTC CAG GAC AAG AGC AAT GAC ATC GAA GCA 660
 201 N L T R E C F E G I L Q D K S N D I E A 220

661 GAG TGG CAC GAT GAG AGC CAC CTC AAC AAA TAC TTC CTT TTC AAC AAA CCC ACT AAA ATC 720
 221 E W H D E S H L N K Y F L F N K P T K I 240

721 TTA TCT CCA GAA TAC TGT TGG GAC TAT CAG ATA GGC TTG CCT TCA GAA ATT AAA AAT GTA 780
 241 L S P E Y C W D Y Q I G L P S E I K N V 260

781 AAG GTA GCC TGG CAG ACA AAA GAG TAT AAT TTG GTT AGG AGT AAT GTC 828
 261 K V A W Q T K E Y N L V R S N V 276

FIGURA 7
Perfil transcripcional por qPCR

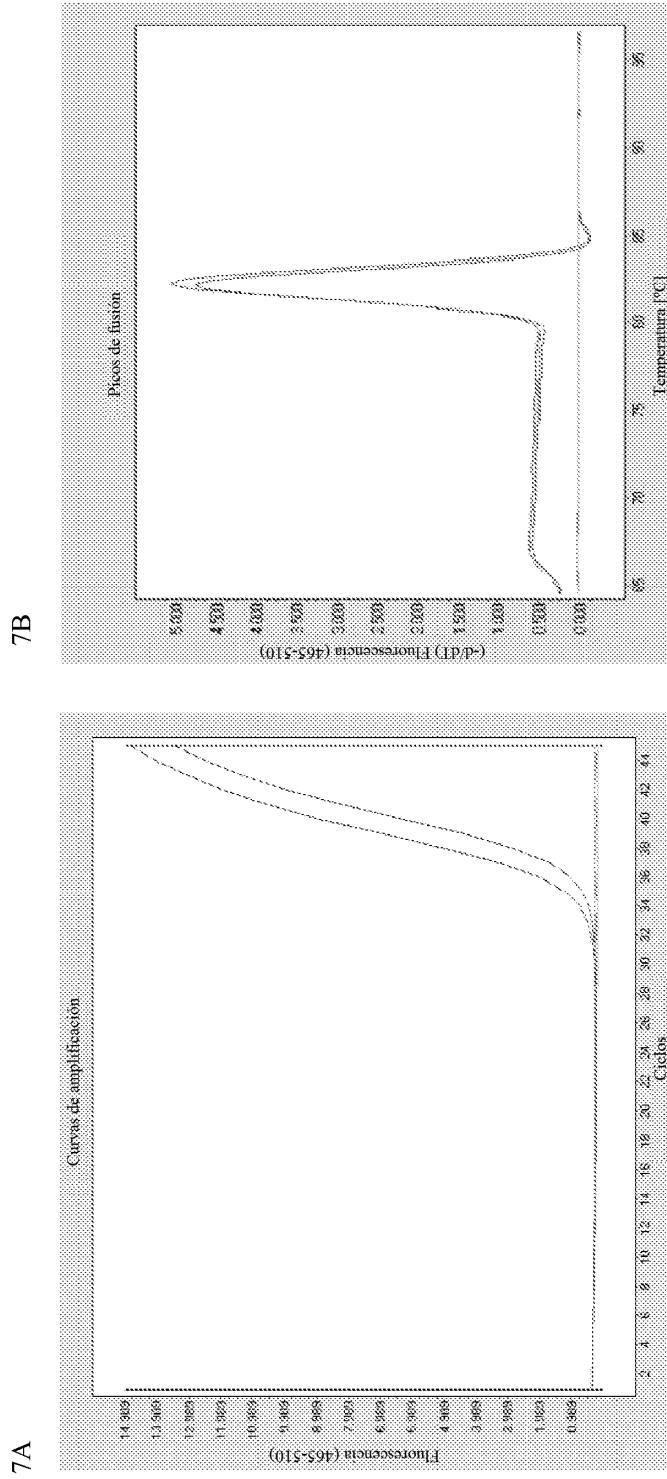


FIGURA 8

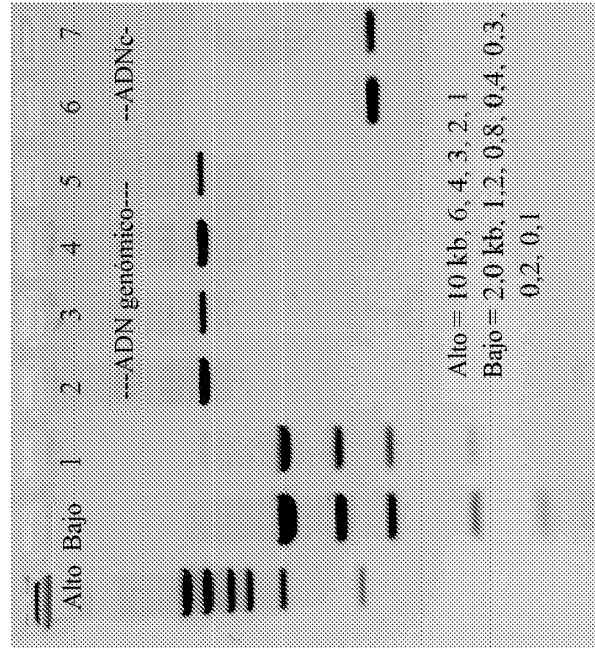


FIGURA 9

