

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 535 358**

51 Int. Cl.:

C07K 14/705 (2006.01)

C12N 15/62 (2006.01)

C12N 1/15 (2006.01)

C12N 1/21 (2006.01)

C12N 5/10 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **07.12.2007 E 12157790 (2)**

97 Fecha y número de publicación de la concesión europea: **25.02.2015 EP 2479189**

54 Título: **Bibliotecas de proteínas de fusión de transferrina**

30 Prioridad:

12.12.2006 US 934951 P

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
08.05.2015

73 Titular/es:

**BIOREXIS PHARMACEUTICAL CORPORATION
(100.0%)
235 East 42nd Street
New York, NY 10017-5755, US**

72 Inventor/es:

**TURNER, ANDREW JOHN y
WANG, BAIYANG**

74 Agente/Representante:

CARPINTERO LÓPEZ, Mario

ES 2 535 358 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

Bibliotecas de proteínas de fusión de transferrina

Solicitudes relacionadas

- 5 La presente solicitud está relacionada con la Solicitud Internacional PCT PCT/US2006/023742, presentada el 19 de junio de 2006, que reivindica el beneficio de la Solicitud Provisional de Estados Unidos 60/691.229, presentada el 17 de junio de 2005. La presente solicitud también está relacionada con la Solicitud de Patente de Estados Unidos 10/515.429, presentada el 23 de noviembre de 2004; la Solicitud Provisional de Estados Unidos 60/485.404, presentada el 9 de julio de 2003; la Solicitud de Patente de Estados Unidos 10/384.060 presentada el 10 de marzo de 2003; y la Solicitud Provisional de Estados Unidos 60/406.977, presentada el 30 de agosto de 2002.

10 Campo de la invención

La presente invención se refiere a proteínas de fusión, bibliotecas de proteínas de fusión.

Antecedentes de la invención

Sistemas de despliegue de la superficie celular

- 15 Los procedimientos combinatorios de cribado y selección de bibliotecas han llegado a ser herramientas de investigación comunes (Phizicky *et al.* (1995) Microbiological Reviews 59: 94 - 123). Una de las técnicas más extendidas es el despliegue en fago, mediante el cual una proteína se expresa en forma de una fusión de polipéptidos a una proteína de recubrimiento de bacteriófago y posteriormente se seleccionan mediante la unión a un ligando inmovilizado o biotinilado soluble. La presentación de péptidos aleatorios a menudo se lleva a cabo mediante la construcción de proteínas quiméricas expresadas sobre la superficie exterior de bacteriófagos filamentosos tales como M13, fd y f1. El despliegue en fago se ha aplicado de manera exitosa a anticuerpos, proteínas de unión a ADN, inhibidores de proteasa y enzimas. Véase Hoogenboom *et al.* (1997) Trends in Biotechnol. 15: 62 - 70; Ladner (1995) Trends in Biotechnol. 13: 426 - 430; Lowman *et al.* (1991) Biochemistry 30: 10832 - 10838; Markland *et al.* (1996) Biochemistry 35: 8045 - 8057; y Matthews *et al.* (1993) Nucleic Acids Res. 21: 1727 - 1734.

- 25 Además del despliegue en fago, se han desarrollado varios procedimientos de despliegue en la superficie celular bacteriana. Véase Georgiou *et al.* (1997) Nat. Biotechnol. 15: 29 - 34. Un planteamiento tomado en los procedimientos de despliegue en la superficie celular bacteriana ha sido usar una proteína de fusión que comprende una proteína pilina (TraA) o una parte de la misma y un polipéptido heterólogo que despliega el péptido de la biblioteca en la superficie exterior de una célula huésped bacteriana capaz de formar pelo. Véase la Patente de Estados Unidos 5.516.637.

- 30 La biblioteca de péptidos aleatorios FLITRX™ (Invitrogen™ Life Technologies) usa la proteína flagelar bacteriana, FliC, and tioredoxina, TrxA, para desplegar una biblioteca de péptidos aleatorios de dodecámeros sobre la superficie de *E. coli* de una manera conformacional restringida. Véase Lu *et al.* (1995) BioTechnology 13: 366. Estos sistemas se han aplicado al mapeo de epítomos de anticuerpo, el desarrollo y construcción de sistemas de distribución de vacunas bacterianas y la generación de bioabsorbentes de células enteras para los propósitos de depuración ambiental y diagnóstico. Las secuencias de péptidos que se unen a dianas específicas de tumor sobre las células epiteliales derivadas de tumor también se han identificado usando el sistema FLITRX™. Véase Brown *et al.* (2000) Annals of Surgical Oncology, 7(10): 743.

- 40 Se han desarrollado sistemas de despliegue sobre la superficie celular de levaduras para la selección de bibliotecas y han superado con éxito algunas de las limitaciones de los sistemas de despliegue en fago y en bacterias. Los sistemas de despliegue sobre la superficie de levaduras, tal como el Kit pYD1 Yeast Display Vector (Invitrogen™ Life Technologies), usan el receptor de α -aglutinina de *S. cerevisiae* para desplegar proteínas extrañas sobre la superficie de la célula. El receptor de α -aglutinina consta de dos subunidades codificadas por los genes AGA1 y AGA2. La proteína Aga1 (Aga1p, 725 aminoácidos) se secreta de la célula y se llega a unir de manera covalente a β -glucano en la matriz extracelular de la pared celular de levadura. La proteína Aga2 (Aga2p, 69 aminoácidos) se une a Aga1p mediante dos enlaces disulfuro y después de la secreción se mantiene unida a la célula mediante su contacto con Aga1p. La parte N-terminal de Aga2p se requiere para la unión a Aga1p, mientras que las proteínas y péptidos se pueden fusionar al extremo C para la presentación sobre la superficie celular de levaduras. La aglutinina es una proteína de levadura natural que normalmente funciona como un contacto de unión específico para fusionar células de levadura durante el apareamiento. Como tal, se ha desarrollado la unión proteína-proteína sin excesivo impedimento estérico de los componentes de la pared celular. Boder *et al.* en "Yeast Surface Display for Directed Evolution of Protein Expression, Affinity, and Stability", Applications of Chimeric Genes and Hybrid Proteins, (Jeremy Thorner *et al.*), Academic Press, 2000, Vol. 328, páginas 430 - 439; documentos de US 6.699.658; y US 6.423.538.

- 55 Sin embargo, uno de los inconvenientes de este sistema, es que, como la proteína de fusión Aga2p -y Aga1p se requieren para formar un enlace disulfuro con el fin de que la proteína Aga2p se sujete a la pared celular, la eficacia del despliegue es relativamente baja, desplegando de manera eficaz solamente 40% al 60% de las células de levadura la proteína sobre la superficie. Véase Feldhause *et al.* (2003) Nat. Biotechnol. 21(2): 163 - 70. Existe una necesidad de un sistema de despliegue en levadura que presente la mayoría de, si no todas, las proteínas de una biblioteca sobre una superficie celular.

- 60 Otro inconveniente del sistema de despliegue en levadura de Aga1p y Aga2p es que este requiere que el ligando a seleccionar esté unido al extremo C de Aga2p. Como resultado, el sistema no se puede usar para seleccionar péptidos en los que se requiera un extremo N libre para la unión y/o se requiera para actividad. De acuerdo con lo anterior, existe una necesidad de un sistema de despliegue flexible que no requiera la unión del extremo N del ligando a una proteína celular de levadura.

Proteína de fusión de transferrina

La transferrina (Tf) de suero es una glucoproteína monomérica con un peso molecular de 80.000 daltons que se une al hierro en la circulación y lo transporta a diversos tejidos mediante el receptor de transferrina (TfR) (Aisen *et al.* (1980) Ann. Rev. Biochem. 49: 357 - 393; MacGillivray *et al.* (1981) J. Biol. Chem. 258: 3543 - 3553; y la Patente de Estados Unidos 5.026.651). Tf es una de las moléculas en suero más comunes, que comprende hasta aproximadamente 5 - 10% de las proteínas en suero totales. La transferrina deficiente en carbohidratos se produce en niveles elevados en la sangre de individuos alcohólicos y muestra una semivida más larga (aproximadamente 14 - 17 días) que la de la transferrina glucosilada (aproximadamente 7 - 10 días). Véase van Eijk *et al.* (1983) Clin. Chim. Acta 132: 167 - 171; Stibler (1991) Clin. Chem. 37: 2029 - 2037; Arndt (2001) Clin. Chem. 47(1): 13 - 27; y Stibler *et al.* en "Carbohydrate-deficient consumption", Advances in the Biosciences, (Ed Nordmann *et al.*), Pergamon, 1988, Vol. 71, páginas 353 - 357). La estructura de Tf se ha caracterizado bien y se han elucidado los mecanismos de unión al receptor, unión a hierro y liberación y unión del ion carbonato. Véanse las Patentes de Estados Unidos 5.026.651, 5.986.067 y MacGillivray *et al.* (1983) J. Biol. Chem. 258(6): 3543 - 3546.

Las mucinas son una familia de proteínas altamente glucosiladas. Las mucinas y proteínas de tipo mucina se han usado para elevar un dominio de ligando de una proteína de fusión a una distancia sustancial de una microdisposición. Se ha propuesto la hipótesis de que elevar un ligando una distancia significativa de un sustrato incrementa la unión del ligando a un receptor desplegado en las células que expresan el receptor. Véase el documento WO 01/46698.

En la presente invención se han desarrollado previamente bibliotecas de proteínas de fusión de transferrina. Véase la Solicitud de Patente de Estados Unidos 10/515,429. La presente invención proporciona una proteína de fusión de transferrina que contiene un resto de tipo tallo, tal como una mucina o una proteína de tipo mucina, diseñada para reducir el impedimento estérico e incrementar la unión al ligando. La proteína de fusión se puede expresar y desplegar sobre la superficie de una célula huésped, tal como levadura, de manera que la proteína de fusión de transferrina se pueda usar como una plataforma de selección de péptidos. Además, la transferrina y la parte de ligando de la proteína de fusión se puede escindir y usar como un agente terapéutico. Esto puede que no sea posible llevarlo a cabo con la tecnología de despliegue en levadura existente ya que la retirada de la proteína Aga2 fusionada N-terminal afectaría probablemente la conformación de un pequeño ligando unido a transferrina.

Sumario de la invención

Como se describe en más detalle más adelante, la presente divulgación incluye una proteína de fusión con un resto de transferrina (Tf). En una realización de la divulgación, la proteína de fusión comprende un resto de transferrina, un resto del tallo y un grupo de unión a la pared celular. En esta realización de la divulgación, la proteína de fusión se expresa sobre una superficie celular de la célula y se une a la pared celular mediante el grupo de unión a la pared celular. En otra divulgación de la invención, la proteína de fusión comprende un resto de transferrina que está unido a una membrana celular mediante un anclaje, por ejemplo, un anclaje GPI.

El resto de transferrina de la divulgación contiene una proteína de transferrina o una porción de la misma. Por ejemplo, el resto de transferrina puede ser una parte del dominio N, es decir lóbulo, de la proteína de transferrina. El resto Tf puede ser una proteína Tf modificada de manera que la parte Tf de la proteína de fusión muestra glucosilación reducida comparada con la Tf de tipo natural. En una divulgación, la parte de transferrina de la proteína de fusión no muestra glucosilación. En otra realización de la presente invención, el resto de transferrina de la proteína de fusión está modificada de manera que muestre afinidad reducida al hierro, bicarbonato y/o afinidad reducida a un receptor de transferrina comparada con la transferrina de tipo natural. El resto de transferrina se puede modificar de manera que sea incapaz de unirse a un receptor de transferrina, a hierro o a bicarbonato. De acuerdo con lo anterior, la presente invención incluye restos de transferrina modificados en los que el resto de transferrina está modificado en uno o más sitios a partir del grupo que consiste en un sitio de glucosilación, sitio de unión a hierro, sitio de articulación, sitio de bicarbonato y sitio de unión al receptor.

El ligando de la invención reivindicada puede formar complejos o fusionarse con el resto de transferrina de diversas maneras. Además, un resto de transferrina puede tener más de un ligando asociado con él. El resto de ligando se puede fusionar al extremo N, al extremo C del resto de transferrina, o se puede localizar dentro del resto de transferrina. En una divulgación, el ligando está insertado en una o más posiciones de aminoácidos del lóbulo N (N₁ o N₂) seleccionados entre el grupo que consiste en las posiciones de aminoácido Asp33, Asn55, Asn75, Asp90, Gly257, Lys280, His289, Ser298, Ser105, Glu141, Asp166, Gln184, Asp197, Lys217, Thr231 y Cys241.

La divulgación también incluye un ligando que se localiza sobre un bucle expuesto del resto de transferrina. La proteína de fusión se puede construir de manera que el resto del ligando esté en fase con el resto de transferrina, por ejemplo, mediante la expresión en una célula huésped de un vector que codifica la proteína de fusión de transferrina y ligando.

En otra divulgación, el resto de ligando puede comprender uno o más restos de aminoácidos expuestos de superficie aleatorios con el resto de transferrina. Por ejemplo, el resto de transferrina puede contener al menos aproximadamente 3, aproximadamente 4, aproximadamente 5, aproximadamente 6, aproximadamente 7, aproximadamente 8, aproximadamente 9 o aproximadamente 10 o más restos de aminoácidos expuestos de superficie aleatorios. En una realización, una biblioteca de proteínas de transferrina de fusión o proteínas de transferrina se construye usando aproximadamente seis restos de aminoácido expuestos de superficie. Los aminoácidos expuestos de superficie de transferrina se pueden aleatorizar usando procedimientos de mutagénesis conocidos en la técnica. Los restos de aminoácidos distribuidos al azar pueden ser secuenciales o estar en racimos en una región expuesta. Las regiones expuestas de superficie de transferrina incluyen, pero no se limitan a los aminoácidos en aproximadamente la posición 85 a aproximadamente la posición 92, aproximadamente la posición 276 a aproximadamente la posición 298 o aproximadamente la posición 207 a aproximadamente la posición 217.

El ligando puede tomar muchas formas, incluyendo, pero sin limitación a, un anticuerpo de una sola cadena, anticuerpo, fragmento de anticuerpo, región variable de anticuerpo, péptido aleatorio, o región determinante de la complementariedad del anticuerpo (CDR). Los ligandos pueden contener una región variable o aleatoria y una región no variable. En una divulgación, el ligando es un péptido aleatorio. Un resto de ligando de péptido aleatorio expresado con un resto de transferrina se puede crear mediante muchos procedimientos conocidos en la técnica incluyendo, pero no limitados a, recuperación de PCR y ADN propensa a error. Un resto de ligando también se puede añadir a una proteína de fusión de transferrina después que esta última se haya ya trasladado.

El ligando puede ser un ligando de interés o un ligando en una biblioteca de ligandos. El ligando puede ser capaz de unirse a uno o más receptores o agentes tales como un péptido, antígeno, receptor, anticuerpo, toxina, metabolito y ácido nucleico.

En una realización de la invención, el ligando es una secuencia de unión a ligando conocida que contiene uno o más aminoácidos distribuidos al azar dentro de la secuencia de unión al ligando o en proximidad estrecha a la secuencia de unión al ligando. Por ejemplo, el ligando puede ser la secuencia de unión a integrina Arg-Gly-Asp (RGD) con uno o más restos de aminoácidos distribuidos al azar que flanquean o bien un lado o ambos lados de la secuencia de RGD. En una divulgación, el ligando es CXXRGRDXXC, donde X es un resto de aminoácido aleatorizado. En una divulgación adicional, el ligando es CXXRGRDXXX.

Como se ha descrito anteriormente, en una realización, la proteína de fusión comprende un resto de tallo. El resto del tallo se puede orientar de manera que su extremo N esté fusionado al resto de transferrina y su extremo C localizado en la célula, por ejemplo, en la pared celular. En una realización, el extremo C del resto del tallo está fusionado a un resto de anclaje. En una realización, el resto del tallo de la presente invención atraviesa la pared celular de una célula de levadura y en general es un péptido moderada a altamente glucosilado. Al atravesar la pared celular, el resto del tallo puede actuar como un miembro de unión a la pared celular para atar la proteína de fusión a través de la pared celular. En otra realización de la invención, el resto del tallo se extiende sobre la pared celular y se despliega sobre la superficie celular. La composición del resto del tallo puede proporcionar una conformación de tipo bastón que reduce el impedimento estérico que de otra manera existiría entre la proteína de fusión, notablemente el ligando y la célula huésped.

El resto del tallo puede contener o constar de una mucina, variante de mucina o fragmento de la misma. El mucidominio N puede incluir, por ejemplo, MUC1, MUC2, MUC3, MUC4, MUC5AC, MUC5B, MUC6, MUC7, MUC8, MUC9, MUC10, MUC11, MUC12, MUC13, MUC14, MUC15, MUC16, MUC17, MUC18, MUC19, MUC20, MUC21 y sus variantes. En una divulgación, el resto del tallo contiene un dominio MUC1 humano tal como el péptido que corresponde a la secuencia de ácido nucleico de la SEC ID NO: 5 o un fragmento de la misma. En otra realización, el resto del tallo comprende dos o más repeticiones de una mucina, por ejemplo, dos o más repeticiones de MUC1 o MUC3. En una realización adicional, el resto del tallo comprende dos o más proteínas de mucina o variantes o fragmentos de la misma a partir del grupo que consiste en MUC1, MUC2, MUC3, MUC4, MUC5AC, MUC5B, MUC6, MUC7, MUC8, MUC9, MUC10, MUC11, MUC12, MUC13, MUC14, MUC15, MUC16, MUC17, MUC18, MUC19, MUC20 y MUC21. Las variantes de mucina y variantes de otras proteínas de tallo se pueden modificar por ingeniería genética mediante procedimientos conocidos en la técnica, por ejemplo, mediante mutagénesis al azar.

El resto del tallo también puede contener o constar de otras proteínas que están moderada a altamente glucosiladas, incluyendo proteínas nativas de la pared de levadura. Por ejemplo, en una realización de la invención, el resto del tallo contiene o consta de Aga1, una variante de of Aga 1, o uno de sus fragmentos.

Las proteínas de fusión de la presente invención también pueden incluir un miembro de unión a la pared celular que actúa para inmovilizar o atar la proteína de fusión a una célula huésped. El miembro de unión a la pared celular puede unir de manera covalente o no covalente la proteína de fusión de la invención a la pared celular de levadura. En una realización de la invención, el resto del tallo de la proteína de fusión es el miembro de unión a la pared celular. Por ejemplo, O-glucanos del resto del tallo pueden reticularse a beta glucanos de la pared celular. Otros miembros de unión a la pared celular, incluyen, pero no se limitan a, péptidos que contienen restos de cisteína libres. Por ejemplo, un resto del tallo o resto de anclaje que contiene uno o más residuos de cisteína no emparejados pueden formar un(os) enlace(s) disulfuro con uno o más restos de cisteína no emparejados de proteínas en la pared celular.

La proteína de fusión de la invención puede opcionalmente contener un resto de anclaje que también actúa para inmovilizar o atar la proteína de fusión de transferrina a la célula huésped. El resto de anclaje puede ser un miembro de unión a la pared celular o puede atar la proteína de fusión a una membrana celular.

Un dominio de anclaje capaz de atar la proteína de fusión de la presente invención a una membrana celular de levadura, entre otras, es un anclaje de péptido glucosil-fosfatidil-inositol (GPI) que se añade mediante la modificación de la proteína después de la traducción al sitio ω en la secuencia del péptido de señal GPI, tal como la secuencia del péptido de señal proporcionada en la SEC ID NO: 15. Un anclaje de péptido GPI también se puede usar para anclar una proteína de fusión de la invención a una membrana celular de un eucariótico superior, *por ejemplo*, una membrana celular de mamífero o una membrana celular de plantas. En una realización de la invención, un anclaje tal como el proporcionado por una secuencia de señal modificada de GPI ata de manera transitoria la proteína de fusión a una membrana de la célula huésped o pared celular antes de que se escinda. Una vez escindida, la proteína de fusión permanece atada a la célula mediante el miembro de unión a la pared celular como resultado de glucanos del resto del tallo que se está reticulando en los beta glucanos de la pared celular.

En otra realización de la invención, el anclaje es un dominio transmembrana. El dominio transmembrana (TMD) puede ser la región de una sola proteína de membrana de tipo I o tipo II de un solo paso o una cualquiera de las varias regiones transmembrana de una proteína de membrana de multiextensión.

La presente invención también incluye la molécula de ácido nucleico que codifica la proteína de fusión reivindicada.

El ácido nucleico puede estar insertado en un vector y usarse para transformar una célula huésped tal como levadura. Una vez que se ha transformado con el ácido nucleico de la presente invención, la célula huésped puede expresar la proteína de fusión. La inducción de la expresión de la proteína de fusión se puede controlar mediante procedimientos conocidos en la técnica, por ejemplo, mediante el uso de un promotor que se puede inducir. La presente invención incluye una biblioteca de proteínas de fusión expresadas en una colección de células huésped, por ejemplo, una colección de células de levadura que expresan la proteína de fusión de la invención que despliega péptidos distribuidos al azar.

En otra realización de la presente invención, la proteína de fusión se usa para seleccionar la actividad de unión de un ligando o agente. Una biblioteca de células huésped capaces de expresar la proteína de fusión reivindicada se puede exponer a un agente, que incluye pero no se limita a, un antígeno o receptor y después se seleccionan por la actividad de unión. Las bibliotecas de despliegue sobre la superficie celular se pueden seleccionar usando los procedimientos conocidos en la técnica, que incluyen, pero no se limitan a, FACS y perlas magnéticas.

Breve descripción de los dibujos

La FIGURA 1 muestra una biblioteca de péptidos aleatorios o CDR desplegada sobre una proteína de fusión de transferrina y la unión del ligando con una diana.

La FIGURA 2 proporciona la secuencia del péptido de anclaje de GPI de la levadura YIR019C y resalta los aminoácidos responsables de la unión a la membrana celular.

La FIGURA 3 proporciona el mapa del vector para pREX0549.

La FIGURA 4 proporciona el mapa del vector para pREX0995.

La FIGURA 5 proporciona el mapa del vector para pREX0667.

La FIGURA 6 proporciona el mapa del vector para pREX1012.

La FIGURA 7 proporciona el mapa del vector para pREX0759.

La FIGURA 8 muestra la presencia de levadura marcada con Flag después de dos rondas de separación de MACS.

La FIGURA 9 proporciona el mapa del vector para pREX0855.

La FIGURA 10 proporciona el mapa del vector para pREX1087.

La FIGURA 11 proporciona el mapa del vector para pREX1106.

La FIGURA 12 muestra el análisis de FACS con MUC1 y AGA1.

La FIGURA 13 muestra fragmentos A y B obtenidos mediante la amplificación de pREX 1106 con los cebadores P1980/P2181 y P2127/P1173.

La FIGURA 14 muestra la unión directa de transcuerpos RGD a integrina.

La FIGURA 15 muestra el análisis de unión competitiva de unión de trans-cuerpo RGD.

La FIGURA 16 muestra la inhibición de la agregación plaquetaria de ratón mediante un clon de transcuerpo de RGD.

La FIGURA 17 es un análisis FACS de variantes hMUC3 expresadas en levadura.

La FIGURA 18 proporciona el mapa del vector para pREX0757.

La FIGURA 19 proporciona el mapa del vector para pREX1234.

La FIGURA 20 proporciona el mapa del vector para pREX1235.

La FIGURA 21 es una transferencia de Western preparada usando lisados de células de mamífero que expresan la transferrina etiquetada con Flag y probada con anticuerpos anti-Flag o anti-transferrina.

La FIGURA 22 es una transferencia de Western de proteínas de membrana de plasma expuestas.

La FIGURA 23 es un análisis FACS de pREX1235.

Descripción detallada

Descripción General

En la presente invención se ha desarrollado una proteína de fusión multifuncional que se puede usar por ejemplo, como parte de un sistema de despliegue sobre la superficie celular para seleccionar bibliotecas, *por ejemplo*, bibliotecas de péptidos aleatorios o bibliotecas CDR. La proteína de fusión incluye un resto de transferrina que forma complejo con o se fusiona a uno o más ligandos. El resto de transferrina también puede servir como un ligando mediante restos de aminoácidos expuestos de superficie distribuidos al azar.

La invención incluye una proteína de fusión que contiene una proteína diferente de la transferrina mientras que la otra proteína es soluble y es capaz de conferir una semivida en suero incrementada a los uno o más ligandos

fusionados cuando se escinde del remanente de la proteína de fusión, o cuando se reconstruye sin el remanente de la proteína de fusión. Por ejemplo, albúmina o una variante o fragmento de la misma se puede usar en lugar de transferrina.

En una divulgación, el resto de transferrina de la proteína de fusión está fusionado a un resto del tallo, que está moderada a altamente glucosilado. En esta realización de la invención, la proteína de fusión contiene un miembro de unión a la pared celular que es capaz de unir de manera covalente o no covalente la proteína de fusión a la pared celular de una célula de levadura.

En otra realización de la invención, la proteína de fusión contiene un resto de anclaje tal como un dominio transmembrana. La proteína de fusión también puede estar anclada a una membrana celular mediante un anclaje de GPI. Tales anclajes se pueden usar para expresar la proteína de fusión de la invención sobre una célula de mamífero. Aunque una proteína de fusión de la invención puede contener tanto un resto del tallo como un resto de anclaje, la invención incluye proteínas de fusión que comprenden un resto del tallo y ningún resto de anclaje así como un resto de anclaje y ningún resto del tallo.

En una realización de la invención, la proteína de fusión se expresa en levadura. Tal proteína de fusión ofrece ventajas sobre la técnica anterior que incluye proporcionar un incremento en porcentaje de clones con los péptidos desplegados sobre la superficie celular comparado con el sistema de despliegue sobre levadura de Aga1p y Aga2p. La proteína de fusión de la invención también ofrece la flexibilidad de selección de ligandos que requieren un extremo N para unión. De manera adicional, mediante el anclaje de la proteína de fusión de la invención a una membrana celular de mamífero, la proteína de fusión de la invención también es capaz de expresarse y presentarse sobre células de mamífero.

La presente invención también incluye composiciones terapéuticas que comprenden las proteínas de fusión o partes de las mismas y procedimientos de tratamiento, prevención, o mejora de enfermedades o trastornos mediante la administración de las proteínas de fusión o partes de las mismas a un sujeto en necesidad de tal agente terapéutico. Una proteína de fusión de la invención incluye al menos un fragmento o variante de una proteína terapéutica supuesta como un resto de ligando. En una realización de la invención, la transferrina y ligando, es *decir*, la parte terapéutica de la proteína de fusión, se puede escindir de la célula y usarse para preparar un compuesto biofarmacéutico o vacuna. Por ejemplo, la parte terapéutica de la proteína de fusión se puede escindir del resto del tallo o anclaje.

Definiciones

Como se usa en el presente documento, el término “actividad biológica” se refiere a una función o conjunto de actividades realizadas por una molécula terapéutica, resto de ligando, proteína o péptido en un contexto biológico, es *decir*, en un organismo o un facsímile *in vitro* del mismo. Las actividades biológicas pueden incluir, pero no se limitan a, las funciones de la parte de la molécula terapéutica de las proteínas de fusión reivindicadas, tales como, pero no se limitan a, la inducción de secreción de la matriz extracelular de las líneas celulares sensibles, la inducción de la secreción de hormonas, la inducción de quimiotaxis, la inducción de mitogénesis, la inducción de la diferenciación, o la inhibición de una división celular de las células sensibles. Una proteína de fusión o péptido de la invención se considera que es biológicamente activo si muestra una o más actividades biológicas de su homólogo nativo de la proteína terapéutica.

Como se usa en el presente documento, un “aminoácido correspondiente a” o un “aminoácido equivalente” en una secuencia de transferrina se identifica mediante alineación para maximizar la identidad o similitud entre una primera secuencia de transferrina y al menos una segunda secuencia de transferrina. El número usado para identificar un aminoácido equivalente en una segunda secuencia de transferrina se basa en el número usado para identificar el aminoácido correspondiente en la primera secuencia de transferrina. En ciertos casos, estas frases se pueden usar para describir los residuos de aminoácidos en la transferrina humana comparados con ciertos residuos en transferrina de suero de conejo.

Como se usa en el presente documento, los términos “resto Tf”, “fragmento de una proteína Tf” o “proteína Tf,” o “parte de una proteína Tf” se refieren a una secuencia de aminoácidos que comprende al menos aproximadamente el 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 96%, 97%, 98%, 99% o 100% de una proteína Tf de origen natural o mutante de la misma.

Como se usa en el presente documento, el término “gen” se refiere a cualquier segmento de ADN asociado a una función biológica. De este modo, los genes incluyen, pero no se limitan a, secuencias de codificación y/o las secuencias reguladoras requeridas para su expresión. Los genes también pueden incluir segmentos de ADN no expresados que, por ejemplo, forman secuencias de reconocimiento para otras proteínas. Los genes se pueden obtener a partir de una diversidad de fuentes, incluyendo clonación a partir de una fuente de interés y o síntesis a partir de la información de secuencia conocida o predicha y pueden incluir secuencias diseñadas para tener los efectos deseados.

Como se usa en el presente documento, un “polinucleótido heterólogo” o un “ácido nucleico heterólogo” o un “gen heterólogo” o una “secuencia heteróloga” o un “segmento de ADN exógeno” se refiere a un polinucleótido, ácido nucleico o segmento de ADN que se origina a partir de una fuente extraña para la célula huésped particular, o, si es a partir de la misma fuente, se modifica a partir de su forma original. Un gen heterólogo en una célula huésped incluye un gen que es endógeno para la célula huésped particular, pero se ha modificado. De este modo, los términos se refieren a un segmento de ADN que es extraño o heterólogo a la célula, u homólogo a la célula pero en una posición dentro del ácido nucleico de la célula huésped en la que el elemento no se encuentra de manera ordinaria. Como un ejemplo, una secuencia señal nativa a una célula de levadura pero unida a una secuencia de Tf humana es heteróloga.

Como se usa en el presente documento, una secuencia de ácido nucleico “aislada” se refiere a una secuencia de ácido nucleico que está esencialmente libre de otras secuencias de ácido nucleico, *por ejemplo*, al menos aproximadamente 20% puro, preferiblemente al menos aproximadamente 40% puro, más preferiblemente aproximadamente 60% puro, incluso más preferiblemente aproximadamente 80% puro, lo más preferiblemente aproximadamente 90% puro e incluso lo más preferiblemente aproximadamente 95% puro, como se determina mediante electroforesis en gel de agarosa. Por ejemplo, una secuencia de ácido nucleico aislada se puede obtener mediante procedimientos de clonación convencionales usados en ingeniería genética para volver a localizar la secuencia de ácido nucleico a partir de su localización natural a un sitio diferente donde se reproducirá. Los procedimientos de clonación pueden implicar la escisión y aislamiento de un fragmento de ácido nucleico deseado que comprende la secuencia de ácido nucleico que codifica el polipéptido, la inserción del fragmento en una molécula vector e incorporación del vector recombinante en una célula huésped donde se replicarán copias múltiples o clones de la secuencia de ácido nucleico. La secuencia de ácido nucleico puede ser de, ADNc, ARN, genómico de origen semisintético, sintético, o cualquier combinación de los mismos.

Como se usa en el presente documento, dos o más secuencias que codifican ADN se dice que están “unidas” o “fusionadas” cuando, como resultado de las fusiones en fase entre las secuencias que codifican ADN, las secuencias que codifican ADN se traducen en un polipéptido de fusión. El término “fusión” con relación a la proteína de fusión comprende un resto de ligando, resto del tallo y resto de anclaje. Una proteína Tf de fusión es una fusión de un resto de transferrina a un resto del tallo y contiene un miembro de unión a la pared celular.

“Transferrina modificada” como se usa en el presente documento se refiere a una molécula de transferrina que muestra al menos una modificación de su secuencia de aminoácidos, comparado con la transferrina de tipo natural.

“Proteína de fusión de transferrina modificada” como se usa en el presente documento se refiere a una proteína formada por la fusión de al menos una molécula de transferrina modificada (o un fragmento o variante de la misma) que forma complejo o fusionada a un ligando, que está condensado a un resto del tallo.

Como se usa en el presente documento, los términos “ácido nucleico” o “polinucleótido” se refieren a desoxirribonucleótidos o ribonucleótidos y polímeros de los mismos en forma o bien de una sola cadena o de doble cadena. Salvo que se limite específicamente, los términos abarcan ácidos nucleicos que contienen análogos de nucleótidos naturales que tienen propiedades de unión similares al ácido nucleico de referencia y se metabolizan de una manera similar a los nucleótidos de origen natural. Salvo que se indique de otra manera, una secuencia de ácido nucleico particular también abarca de manera implícita sus variantes modificadas de manera conservadora (*por ejemplo*, sustituciones de codón degenerado) y secuencias complementarias así como la secuencia indicada de manera explícita. De manera específica, las sustituciones de codón degenerado se pueden lograr mediante la generación de secuencias en las que la tercera posición de uno o más (o todos) los codones seleccionados está sustituida con restos de base mezclados y/o desoxiinosina (Batzner *et al.* (1991) Nucleic acid Res. 19: 5081; Ohtsuka *et al.* (1985) J. Biol. Chem. 260: 2605 - 2608; Cassol *et al.* (1992); Rossolini *et al.* (1994) Mol. Cell. Probes 8: 91 - 98). El término ácido nucleico se usa indistintamente con gen, ADNc y ARNm codificado por un gen.

Como se usa en el presente documento, un segmento de ADN se califica como “unido de manera operativa” cuando está puesto en relación funcional con otro segmento de ADN. Por ejemplo, ADN para una secuencia de señal está unido de manera operativa a ADN que codifica una proteína de fusión de la invención si se expresa como una preproteína que participa en la secreción de la proteína de fusión; un promotor o potenciador está unido de manera operativa a una secuencia codificante si estimula la transcripción de la secuencia. En general, las secuencias de ADN que están unidas de manera operativa son contiguas y en el caso de una secuencia de señal o proteína de fusión tanto contiguas como en fase de lectura. Sin embargo, los potenciadores no necesitan ser contiguos con las secuencias codificantes cuya transcripción controlan. Unión, en este contexto, se lleva a cabo mediante la unión a sitios de restricción convenientes o a adaptadores o engarces insertados en su lugar.

Como se usa en el presente documento, el término “promotor” se refiere a una región de ADN implicada en la unión de la ARN polimerasa para iniciar la transcripción.

Como se usa en el presente documento, el término “recombinante” se refiere a una célula, tejido u organismo que experimenta la transformación con ADN recombinante.

Como se usa en el presente documento, una entidad de dirección, proteína, polipéptido o péptido se refiere a una molécula que se une de manera específica a un tipo de célula particular, *por ejemplo*, célula normal, tal como un linfocito, o célula anormal, tal como una célula cancerosa, por lo tanto se pueden usar para dirigir una proteína Tf de fusión o compuesto (fármaco, o agente citotóxico) al tipo de célula de manera específica.

Como se usa en el presente documento, “proteína terapéutica” se refiere a proteínas, polipéptidos, anticuerpos, péptidos o fragmentos o variantes de los mismos, que tienen una o más actividades terapéuticas y/o biológicas. Las proteínas terapéuticas abarcadas por la invención incluyen pero no se limitan a proteínas, polipéptidos, péptidos, anticuerpos y compuestos biológicos. Los términos péptidos, proteínas y polipéptidos se usan de manera indistinta en el presente documento. De manera adicional, el término “proteína terapéutica” se puede referir a la correspondiente endógena o de origen natural de una proteína terapéutica. Mediante un polipéptido que despliega una “actividad terapéutica” o una proteína que es “terapéuticamente activa” se quiere decir un polipéptido que posee una o más actividades biológicas y/o terapéuticas conocidas asociadas a una proteína terapéutica tal como una o más de las proteínas terapéuticas descritas en el presente documento o conocidas de otra manera en la técnica. Como un ejemplo no limitante, una “proteína terapéutica” es una proteína que es útil para tratar, prevenir o mejorar una enfermedad, afección o trastorno. Tal enfermedad, afección o trastorno puede estar en seres humanos o en un animal no humano, *por ejemplo*, uso veterinario.

Como se usa en el presente documento, el término “transformación” se refiere a la transferencia de ácido nucleico, es decir, un polímero de nucleótido, en una célula. Como se usa en el presente documento, el término

“transformación genética” se refiere la transferencia e incorporación de ADN, especialmente ADN recombinante, en una célula.

Como se usa en el presente documento, el término “transformante” se refiere a una célula, tejido u organismo que ha experimentado transformación.

- 5 Como se usa en el presente documento, el término “transgén” se refiere a un ácido nucleico que está insertado en un organismo, célula huésped o vector de una manera que asegura su función.

Como se usa en el presente documento, el término “transgénico” se refiere a células, cultivos de células, organismos, bacterias, hongos, animales, plantas y progenie de cualquiera de los anteriores, que han recibido un gen extraño o modificado y en particular un gen que codifica una proteína de fusión Tf modificada mediante uno de los diversos procedimientos de transformación, en los que el gen extraño o modificado procede de especies iguales o diferentes de las especies del organismo que recibe el gen extraño o modificado .

10 “Variantes o variante” se refiere a un polinucleótido o ácido nucleico que difiere de un ácido nucleico o polipéptido de referencia, pero que retiene sus propiedades esenciales. En general, las variantes son en conjunto similares y en muchas regiones, idénticas al ácido nucleico o polipéptido de referencia. Como se usa en el presente documento, “variante” se refiere a una parte de de proteína terapéutica de una proteína de fusión de transferrina de la invención, que difiere en secuencia de una proteína terapéutica nativa pero que retiene al menos una propiedad funcional y/o terapéutica de la misma como se describe en cualquier parte en el presente documento o se conoce de otra manera en la técnica.

20 Como se usa en el presente documento, el término “vector” se refiere de manera amplia a cualquier plásmido, fagémido o virus que codifica un ácido nucleico exógeno. El término también se considera que incluye compuestos no plásmidos, no fagémidos y no virales que facilitan la transferencia de ácido nucleico en viriones o células, tales como, por ejemplo, compuestos de polilisina y similares. El vector puede ser un vector viral que es adecuado como un vehículo de distribución para la distribución del ácido nucleico, o mutante del mismo, a una célula, o el vector puede ser un vector no viral que es adecuado para el mismo propósito. Los ejemplos de vectores virales y no virales para la distribución de ADN a las células y tejidos se conocen bien en la técnica y se describen, por ejemplo, en Ma *et al.* (1997, *Proc. Natl. Acad. Sci. EE.UU.* 94: 12744 - 12746). Los ejemplos de vectores virales incluyen, pero no se limitan a, un virus vaccinia recombinante, un adenovirus recombinante, un retrovirus recombinante, un virus adeno-asociado recombinante, un poxvirus aviar recombinante y similares (Cranage *et al.*, 1986, *EMBO J.* 5:3057 - 3063; Solicitud de patente Internacional N° WO94/17810, publicada el 18 de agosto de 1994; Solicitud de patente Internacional N° WO94/23744, publicada el 27 octubre de 1994. Los ejemplos de los vectores no virales incluyen, pero no se limitan a, liposomas, poliamina derivados de ADN y similares.

Como se usa en el presente documento, el término “tipo natural” se refiere a una secuencia de polinucleótidos o polipéptidos que es de origen natural.

35 Como se usa en el presente documento, “proteína de estructura”, “polipéptido de estructura”, o “estructura” se refiere a una proteína a la que se puede fusionar las secuencias de aminoácidos tales como péptidos aleatorios. Los péptidos son exógenos a la estructura.

40 Como se usa en el presente documento, “secuencia de péptido aleatorio” se refiere a una secuencia de aminoácidos compuesta de dos o más monómeros aminoácidos y construida mediante un procedimiento estocástico o al azar. Un péptido aleatorio puede incluir restos de marco conservado o de estructura, que pueden comprender secuencias invariantes. Una secuencia de péptido aleatorio puede contener una parte no variante, es decir, aminoácidos no aleatorios.

Como se usa en el presente documento “biblioteca de péptidos aleatorios” se refiere a un conjunto de secuencias de polinucleótidos que codifican un conjunto de péptidos aleatorios y al conjunto de péptidos aleatorios codificados por esas secuencias de polinucleótidos, así como a las proteínas de fusión que contienen los péptidos aleatorios.

45 Como se usa en el presente documento, el término “pseudoaleatorio” se refiere a un conjunto de secuencias que tienen variabilidad limitada, de manera que por ejemplo, el grado de variabilidad de resto en una posición es diferente del grado de variabilidad del resto en otra posición, pero cualquier posición pseudoaleatoria permite algún grado de variación de resto, sin embargo circunscrita.

50 Como se usa en el presente documento, el término “marco conservado de secuencia definida” se refiere a un conjunto de secuencias definidas que se seleccionan en una base no aleatoria, en general en base a los datos experimentales o datos estructurales, por ejemplo, un marco conservado de secuencia definida puede comprender un conjunto de secuencias de aminoácidos que se predice que forman una estructura de hoja β o puede comprender un resto de repetición de cremallera en grupos de siete de leucina, un dominio de dedo de cinc, entre otras variaciones. Una “secuencia definida interna” es un conjunto de secuencias que abarcan un alcance limitado de variabilidad. Mientras una secuencia completamente aleatoria de 10-meros de los 20 aminoácidos convencionales puede ser cualquiera de las $(20)^{10}$ secuencias y una secuencia pseudoaleatoria de 10-meros de los 20 aminoácidos convencionales puede ser cualquiera de las $(20)^{10}$ secuencias pero mostrarán un sesgo para ciertos restos en ciertas posiciones y/o en conjunto, una secuencia definida interna es un subconjunto de secuencias que es menor que el número máximo de secuencias potenciales si se permite que cada posición de resto sea cualquiera de los 20 aminoácidos convencionales (y/o que permiten amino/iminoácidos no convencionales). Una secuencia interna definida en general comprende posiciones de residuos variantes e invariantes y/o comprende posiciones de residuos variantes que pueden comprender un residuo seleccionado entre un subconjunto definido de residuos de aminoácidos y similares, o bien de manera segmentada o sobre la longitud completa de la secuencia del miembro de la biblioteca seleccionada individual. Las secuencias internas definidas se refieren a bien las secuencias de aminoácidos o bien las secuencias de polinucleótidos.

Como se usa en el presente documento, “engarce” o “espaciador” se refiere a una molécula o grupo de moléculas que conecta dos moléculas, tal como una proteína de unión a ADN y un péptido aleatorio y sirve para colocar las dos moléculas en una configuración deseable, *por ejemplo*, de manera que el péptido aleatorio se pueda unir a un receptor con mínimo impedimento estérico de la proteína de unión a ADN.

- 5 Como se usa en el presente documento, el término “segmento variable” se refiere a una parte de un péptido naciente que comprende una secuencia aleatoria, pseudoaleatoria, o interna definida. Un segmento variable puede comprender tanto posiciones de resto variantes como invariantes y el grado de variación de resto en una posición de resto variante puede estar limitada; ambas opciones se seleccionan a la discreción del profesional. Típicamente, los segmentos variables son de aproximadamente 3 a 20 restos de aminoácidos de longitud, *por ejemplo*, 8 a 10 aminoácidos de longitud, aunque segmentos variables pueden ser más largos y pueden comprender partes de anticuerpo o proteínas receptoras, tales como un fragmento de anticuerpo, una proteína de unión a ácido nucleico, una proteína receptora y similares.

- 15 Como se usa en el presente documento, el término “epítipo” se refiere a la parte de un antígeno u otra macromolécula capaz de formar una interacción de unión que interactúa con el bolsillo de unión a la región variable de un anticuerpo. De manera típica, tal interacción de unión se manifiesta en forma de un contacto intermolecular con uno o más restos de aminoácidos de una CDR.

- 20 Como se usa en el presente documento, el término “receptor,” “diana,” o “agente” se refiere a una molécula que tiene una afinidad por un ligando dado. Los receptores pueden ser moléculas de origen natural o sintéticas. Los receptores se pueden emplear en estado no alterado o agregados con otras especies. Los receptores pueden estar unidos, de manera covalente o no covalente, a un miembro de unión, es *decir*, ligando, bien directamente o bien mediante una sustancia de unión. Los ejemplos de receptores incluyen, pero no se limitan a, anticuerpos, que incluyen anticuerpos monoclonales y reactivos antiseros con determinantes antigénicos específicos (tales como sobre virus, células, u otros materiales), receptores de la membrana celular, antígenos, moléculas que contienen epítipos, carbohidratos complejos y glucoproteínas, enzimas y receptores de hormona.

- 25 Como se usa en el presente documento, el término “ligando” o “resto de ligando” se refiere a una molécula, tal como un péptido aleatorio o una secuencia de segmento variable, que se reconoce por un receptor o agente particular. Como los expertos en la técnica reconocerán, una molécula (o complejo macromolecular) puede ser tanto un receptor como un ligando.

- 30 Como se usa en el presente documento, “fusionado”, “que forma complejo” o “unido de manera operativa” significa que el péptido aleatorio y la proteína de estructura están unidos conjuntamente, de tal manera que se minimice la interrupción con la estabilidad de la estructura.

- 35 Como se usa en el presente documento, el término “anticuerpo de una sola cadena” se refiere a un polipéptido que comprende un dominio V_H y un dominio V_L en el enlace de polipéptido, en general unido mediante un péptido espaciador (*por ejemplo*, [Gly-Gly-Gly-Gly-Ser]_x SEC ID NO: 17) y que puede comprender secuencias adicionales de aminoácidos en los extremos amino- y/o carboxi. Por ejemplo, un anticuerpo de una sola cadena puede comprender un segmento de atadura para unirse al polinucleótido de codificación. Como un ejemplo, un scFv es un anticuerpo de una sola cadena. Los anticuerpos de una sola cadena son en general proteínas que constan de uno o más segmentos de polipéptidos de al menos 10 aminoácidos contiguos codificados sustancialmente por los genes de la superfamilia de inmunoglobulina (*por ejemplo*, véase The Immunoglobulin Gene Superfamily, A. F. Williams y A. N. Barclay, en Immunoglobulin Genes, T. Honjo, F. W. Alt y T. H. Rabbitts, eds., (1989) Academic Press: San Diego, Calif., p. 361 - 387), la mayoría codificada por una secuencia de genes de cadena pesada o ligera de roedor, de primate no humano, aviar, porcino, bovino, ovino, cabra, o humana. Un anticuerpo de una sola cadena funcional en general contiene una parte suficiente de un producto génico de la superfamilia de inmunoglobulina de manera que retenga la propiedad de unirse a una molécula diana, de manera típica un receptor o antígeno (epítipo).

- 45 Como se usa en el presente documento, el término “región determinante de complementariedad” y “CDR” se refieren al término reconocido en la técnica como se ejemplifica por las definiciones de CDR de Kabat y Chothia también en general conocidas como regiones hipervariables o bucles hipervariables. Véase Chothia y Lesk (1987) J. Mol. Biol. 196: 901; Chothia *et al.* (1989) Nature 342: 877; E. A. Kabat *et al.*, Sequences of Proteins of Immunological Interest (National Institutes of Health, Bethesda, Md.) (1987); y Tramontano *et al.* (1990) J. Mol. Biol. 215: 175. Los dominios de región variable de manera típica comprenden los aproximadamente 105 - 115 aminoácidos amino-terminal de una cadena de inmunoglobulina de origen natural, *por ejemplo*, aminoácidos 1 - 110, aunque los dominios variables algo más cortos o más largos son también adecuados para formar anticuerpos de una sola cadena.

- 55 Una región variable de cadena ligera o pesada consta de una región de “marco” interrumpida por tres regiones hipervariables también llamadas CDR. El grado de la región de marco conservado y las CDR se han definido de manera precisa. Véase, “Sequences of Proteins of Immunological Interest”, E. Kabat *et al.*, 4ª Ed., Departamento de Salud y Servicios Humanos, Bethesda, Md. (1987). Las secuencias de las regiones de marco conservado de diferentes cadenas ligeras y pesadas están relativamente conservadas dentro de una especie. Como se usa en el presente documento, una “región de marco humana” es una región de marco que es sustancialmente idéntica (aproximadamente 85% o más, usualmente 90 - 95% o más) a la región de marco de una inmunoglobulina humana de tipo natural. La región de marco de un anticuerpo, que es las regiones de marco combinadas de las cadenas ligera y pesada constituyentes, sirve para posicionar y alinear las CDR. Las CDR son principalmente responsables de la unión a un epítipo de un antígeno.

- 65 Salvo que se defina de otra manera, todas las técnicas y términos científicos usados en el presente documento tienen el mismo significado que comúnmente se entiende por los expertos en la técnica a los que la invención pertenece. Aunque cualesquiera procedimientos y materiales similares o equivalentes a los descritos en el presente documento se pueden usar en la práctica o ensayo de la presente invención, se describen los procedimientos y materiales preferidos.

Transferrina y modificaciones de transferrina

Las proteínas de fusión de la presente divulgación incluyen una proteína de transferrina (Tf) o parte de la misma que es capaz de presentar un ligando tal como un péptido aleatorio o CDR a un receptor o agente. El resto Tf está fusionado al extremo N del resto del tallo. La proteína Tf o parte de la misma de la proteína de fusión se puede denominar como una "parte", "región" o "resto" de Tf de la proteína de fusión. Como se usa en el presente documento, una proteína de fusión de transferrina es una proteína de transferrina o resto fusionado al resto del tallo y contiene un miembro de unión a la pared celular y opcionalmente un resto de anclaje, o es una proteína de transferrina o resto fusionado directamente a un anclaje de membrana celular.

Cualquier transferrina se puede usar para preparar las proteínas de fusión Tf modificadas de la invención. Como un ejemplo, una Tf humana de tipo natural (Tf) es una proteína de 679 aminoácidos, de aproximadamente 75 kDa (no responsable de la glucosilación), con dos lóbulos o dominios principales, N (aproximadamente 330 aminoácidos) y C (aproximadamente 340 aminoácidos), que parece originarse a partir de una duplicación génica. Véanse los números de acceso del GenBank NM001063, XM002793, M12530, XM039845, XM 039847 y S95936 (www.ncbi.nlm.nih.gov), todos los cuales se incorporan en el presente documento como referencia en su totalidad, así como las SEC ID NOS: 1, 2 y 3. Los dos dominios han divergido con el tiempo pero mantienen un gran grado de identidad / similitud.

Cada uno de los dominios N y C además se divide en dos subdominios, N1 y N2, C1 y C2. La función de Tf es transportar hierro a las células del cuerpo. Este proceso está mediado por el receptor de Tf (TfR), que se expresa en todas las células, particularmente células que crecen activamente. TfR reconoce la forma de hierro unido de Tf (dos moléculas de las cuales están unidas por el receptor), después se produce endocitosis mediante la cual el complejo TfR/Tf se transporta al endosoma, momento en el que la caída localizada de pH da como resultado la liberación del hierro unido y el reciclado del complejo TfR/Tf a la superficie celular y liberación de Tf (conocida como apoTf en su forma de hierro no unida). La unión del receptor es principalmente a través del dominio C de Tf. Los dos sitios de glucosilación en el dominio C no parecen estar implicados en la unión del receptor ya que Tf unida a hierro no glucosilado no se une al receptor.

Cada molécula de Tf puede llevar dos iones de hierro (Fe^{3+}). Éstos forman complejos en el espacio entre los subdominios N1 y N2, C1 y C2 que dan como resultado un cambio conformacional en la molécula.

En la transferrina humana, los sitios de unión a hierro comprenden al menos los aminoácidos Asp 63 (Asp 82 de la SEC ID NO: 2 que incluye la secuencia de señal natural de Tf), Asp 392 (Asp 411 de la SEC ID NO: 2), Tyr 95 (Tyr 114 de la SEC ID NO: 2), Tyr 426 (Tyr 445 de la SEC ID NO: 2), Tyr 188 (Tyr 207 de la SEC ID NO: 2), Tyr 514 o 517 (Tyr 533 o Tyr 536 SEC ID NO: 2), His 249 (His 268 de la SEC ID NO: 2) e His 585 (His 604 de la SEC ID NO: 2) de la SEC ID NO: 3. Las regiones de articulación comprenden al menos restos de aminoácidos del dominio N 94 - 96, 245 - 247 y/o 316 - 318 así como los restos de aminoácidos del dominio C 425 - 427, 581 - 582 y/o 652 - 658 de la SEC ID NO: 3. Los sitios de unión al carbonato comprenden al menos los aminoácidos Thr 120 (Thr 139 de la SEC ID NO: 2), Thr 452 (Thr 471 de la SEC ID NO: 2), Arg 124 (Arg 143 de la SEC ID NO: 2), Arg 456 (Arg 475 de la SEC ID NO: 2), Ala 126 (Ala 145 de la SEC ID NO: 2), Ala 458 (Ala 477 de la SEC ID NO: 2), Gly 127 (Gly 146 de la SEC ID NO: 2) y Gly 459 (Gly 478 de la SEC ID NO: 2) de la SEC ID NO: 3.

En una divulgación, las proteínas de fusión incluyen una transferrina humana modificada, aunque se puede usar cualquier molécula de Tf animal para producir las proteínas de fusión de la invención, incluyendo las variantes Tf humanas, vaca, cerdo, oveja, perro, conejo, rata, ratón, hámster, equidna, ornitorrinco, pollos, rana, gusano cornudo, mono, simio, así como otras especies bovinas, caninas y aves. Todas estas secuencias de Tf están fácilmente disponibles en GenBank y otras bases de datos públicas. Está disponible la secuencia de nucleótidos de Tf (véanse las SEC ID NOS: 1, 2 y 3 y los números de acceso descritos anteriormente y disponibles en www.ncbi.nlm.nih.gov) y se pueden usar para preparar fusiones genéticas entre Tf o un dominio de Tf y la molécula terapéutica de elección. También se pueden preparar fusiones a partir de moléculas relacionadas tales como lactotransferrina (lactoferrina) GenBank Acc: NM_002343) o melanotransferrina.

Lactoferrina (Lf), una proteína de unión a hierro de defensa natural, se ha encontrado que posee actividad antibacteriana, antimicótica, antiviral, antineoplásica y antiinflamatoria. La proteína está presente en secreciones exocrinas que están de manera común expuestas a la flora normal: leche, lágrimas, exudado nasal, saliva, moco bronquial, fluidos gastrointestinales, moco cervico-vaginal y fluido seminal. De manera adicional, Lf es un constituyente principal de los gránulos específicos secundarios de neutrófilos polimorfonucleares circulantes (PMN). La apoproteína se libera tras la desgranulación de los PMN en áreas sépticas. Una función principal de Lf es la de depuración de hierro libre en los fluidos y áreas inflamadas de manera que supriman el daño mediado por radical libre y disminuye la disponibilidad para que el metal invada las células microbianas y neoplásicas. En un estudio que examinó la relación de recambio de ^{125}I Lf en adultos, se mostró que Lf es rápidamente captada por el hígado y la radioactividad persistía durante varias semanas en el hígado y bazo (Bennett *et al.* (1979), *Clin. Sci. (Lond.)* 57: 453 - 460).

En una divulgación, la parte de transferrina de la proteína de fusión de la invención incluye una variante de ajuste de transferrina. En un ejemplo, una variante de ajuste de transferrina puede ser una variante de ajuste de transferrina humana. En una realización específica, la variante de ajuste de transferrina humana puede ser la del nº de acceso de Genbank AAA61140.

En otra realización, la parte de transferrina de la proteína de fusión de la invención incluye una variante de ajuste de lactoferrina. En un ejemplo, una variante de ajuste de lactoferrina de suero humana puede ser una variante de ajuste novedosa de una lactoferrina humana. En una realización específica, la variante de ajuste de lactoferrina neutrófila puede ser la de nº de acceso de Genbank AAA59479. En otra realización específica, la variante de ajuste de lactoferrina neutrófila puede comprender la siguiente secuencia de aminoácidos EDCIALKGEADA (SEC ID NO: 4), que incluye la región novedosa de la variante de ajuste.

La fusión también se puede realizar con melanotransferrina (GenBank n° de acceso NM_013900, melanotransferrina murina). La melanotransferrina es una proteína glucosilada encontrada a altos niveles en células de melanoma maligno y se denominó originalmente antígeno p97 de melanoma humano (Brown *et al.*, 1982, Nature, 296: 171 - 173). Posee alta homología de secuencia con la transferrina de suero humana, lactoferrina humana y transferrina de pollos (Brown *et al.*, 1982, Nature, 296: 171 - 173; Rose *et al.*, Proc. Natl. Acad. Sci., 1986, 83: 1261 - 1265). Sin embargo, a diferencia de estas proteínas, no se ha identificado ningún receptor celular para melanotransferrina. La melanotransferrina se une de manera reversible a hierro y existe en dos formas, una de las cuales se une a las membranas celulares mediante un anclaje de glucosil fosfatidilinositol mientras la otra forma es tanto soluble como secretada activamente (Baker *et al.*, 1992, FEBS Lett, 298: 215 - 218; Alemany *et al.*, 1993, J. Cell Sci., 104: 1155 - 1162; Food *et al.*, 1994, J. Biol. Chem. 274: 7011 - 7017).

Las fusiones de Tf modificadas se pueden realizar con cualquier proteína Tf, fragmento, dominio, o dominio modificado. Por ejemplo, se pueden producir proteínas de fusión usando la secuencia de Tf de longitud completa, con o sin la secuencia de señal de Tf nativa. También se pueden preparar transcuerpos usando un dominio de Tf individual, tal como un dominio N o C individual. También se pueden preparar los transcuerpos con un dominio de Tf doble, tal como un dominio N doble o un dominio C doble. En alguna realización, se pueden producir las fusiones de una proteína terapéutica a un dominio C individual, en las que el dominio C está alterado para reducir, inhibir o evitar la glucosilación, unión de hierro y/o unión al receptor de Tf. En otras realizaciones, el uso de un dominio de N individual es ventajoso ya que los sitios de glucosilación de Tf residen en el dominio C y el dominio N, sobre el mismo, no se une al hierro o al receptor de Tf. En una realización la proteína de fusión Tf tiene un dominio N individual que se expresa a alto nivel.

Como se usa en el presente documento, un dominio C terminal o lóbulo modificado para funcionar como un dominio de tipo N se modifica para mostrar patrones de glucosilación o propiedades de unión a hierro sustancialmente similares a las del dominio N o lóbulo nativo o de tipo natural. En una realización, el dominio C o lóbulo se modifica de manera que no está glucosilado y no está unido a hierro mediante la sustitución de las regiones relevantes del dominio C o aminoácidos con las presentes en las regiones o sitios correspondientes de un dominio N nativo o de tipo natural.

Como se usa en el presente documento, un resto Tf que comprende "dos dominios N o lóbulos" incluye una molécula de Tf que se modifica para reemplazar el dominio C o lóbulo nativo con un segundo dominio N o lóbulo nativo o de tipo natural o un dominio N o lóbulo modificado o contiene un dominio C que se ha modificado para funcionar de manera sustancial como un dominio N de tipo natural o modificado. Véase la Solicitud de Provisional de Estados Unidos 60/406.977.

El análisis de los dos dominios mediante la superposición de la estructura de tres dimensiones de los dos dominios (Swiss PDB Viewer 3.7b2, Iterative Magic Fit) y la alineación directa de aminoácidos (alineación múltiple ClustalW) revela que los dominios han divergido con el tiempo. La alineación de aminoácido muestra un 42% de identidad y 59% de similitud entre los dos dominios. Sin embargo, aproximadamente 80% del dominio N coincide con el dominio C para equivalencia estructural. El dominio C también tiene varios enlaces disulfuro extra comparado con el dominio N.

La alineación de los modelos moleculares para el dominio N y dominio C revela los siguientes equivalentes estructurales:

Domini- nio N (1-330)	4- 24	36- 72 75- 88	94- 136	138 - 139	149 - 164	168 - 173	178 - 198 200 - 214	219 - 255	259 - 260	263 - 268	271 - 275	279 - 280	283 - 288 290 - 304	309 - 327
Domini- nio C (340- 679)	340 - 361	365 - 415	425 - 437 439 - 468	470 - 471	475 - 490	492 - 497	507 - 542	555 - 591	593 - 594	597 - 602	605 - 609	614 - 615	620 - 640	645 - 663

Los enlaces disulfuro para los dos dominios se alinean como sigue:

N	C
	C339-C596
C9-C48	C345-C377
C19-C39	C355-C368
	C402-C674
	C418-C637

(continuación)

N	C
C118-C194	C450-C523
C137-C331	
	C474-C665
C158-C174	C484-C498
C161-C179	
C171-C177	C495-C506
C227-C241	C563-C577
	C615-C620

En una divulgación, la parte de transferrina de la proteína de fusión incluye al menos dos lóbulos N terminal de transferrina. En realizaciones adicionales, la parte de transferrina de la proteína de fusión incluye al menos dos lóbulos N terminales de transferrina derivada de transferrina de suero humana.

- 5 En otra divulgación, la parte de transferrina de la proteína de fusión incluye, comprende, o consta de al menos dos lóbulos N terminales de transferrina que tienen una mutación en al menos un resto de aminoácido seleccionado entre el grupo que consiste en Asp63, Gly65, Tyr95, Tyr188 y His249 de la SEC ID NO: 3.

En otra divulgación, la parte de transferrina de la proteína de fusión modificada incluye un mutante de lóbulo N-terminal de transferrina de suero humana recombinante que tiene una mutación en Lys206 o His207 de la SEC ID NO: 3.

- 10 En otra divulgación, la parte de transferrina de la proteína de fusión incluye, comprende, consta esencialmente de, o consta de al menos dos lóbulos C terminales de transferrina. En realizaciones adicionales, la parte de transferrina de la proteína de fusión incluye al menos dos lóbulos C terminales de transferrina derivados de transferrina de suero humano.

- 15 En una divulgación adicional, el mutante del lóbulo C terminal además incluye una mutación de al menos uno de Asn413 y Asn611 de la SEC ID NO: 3 que no permite glucosilación.

- 20 En otra divulgación, la parte de transferrina incluye al menos dos lóbulos C terminales de transferrina que tienen una mutación en al menos un resto de aminoácido seleccionado entre el grupo que consiste en Asp392, Tyr426, Tyr514, Tyr517 e His585 de la SEC ID NO: 3, en la que el mutante retiene la capacidad de unirse a iones metálicos. En una divulgación alternativa, la parte de transferrina incluye al menos dos lóbulos C terminales de transferrina que tienen una mutación en al menos un resto de aminoácido seleccionado entre el grupo que consiste en Tyr426, Tyr514, Tyr517 y His585 de la SEC ID NO: 3, en el que el mutante tiene una capacidad reducida de unirse a iones metálicos. En otra divulgación, la parte de transferrina incluye al menos dos lóbulos C terminales de transferrina que tienen una mutación en al menos un resto de aminoácido seleccionado entre el grupo que consiste en Asp392, Tyr426, Tyr517 y His585 de la SEC ID NO:3, en el que el mutante no mantienen la capacidad de unirse a iones metálicos y funciona sustancialmente como un dominio N.

- 30 En alguna divulgación, la Tf o parte de Tf tendrá suficiente longitud para incrementar la semivida circulatoria *in vivo*, estabilidad en suero, estabilidad en solución *in vitro* o biodisponibilidad del ligando, es decir, será terapéutica, cuando la Tf o parte de Tf y el ligando de la proteína de fusión se escinden del resto de la proteína de fusión comparada con la semivida circulatoria *in vivo*, estabilidad en suero (semivida), estabilidad *in vitro* o biodisponibilidad del ligando en un estado no fusionado, es decir, no fusionado a Tf. Tal incremento en estabilidad, semivida circulatoria *in vivo* o biodisponibilidad puede ser aproximadamente un 30%, 50%, 70%, 80%, 90% o más de incremento sobre la región del resto de ligando no fusionado. En algunos casos, el resto de ligando que comprende la transferrina modificada muestra una semivida en suero de aproximadamente 1 o más días, 1-2 o más días, 3-5 o más días, 5-10 o más días, 10-15 o más días, 10-20 o más días, aproximadamente 12-18 días o aproximadamente 14-17 días comparado con el ligando en un estado no fusionado.

- 40 Cuando el dominio C de Tf es parte de la proteína de fusión, los dos sitios de glucosilación unidos a N, restos de aminoácido que corresponden a N413 y N611 de la SEC ID NO: 3 se pueden mutar para la expresión en un sistema de levadura para evitar la glucosilación o hipermanosilación y ampliar la semivida en suero de la proteína de fusión (para producir asialo-, o en algunos casos, monosialo-Tf o disialo-Tf). Además de los aminoácidos de Tf que corresponde a N413 y N611, mutaciones para los restos dentro de o adyacentes al sitio de glucosilación N-X-S/T evitan o sustancialmente reducen la glucosilación. Véase la Patente de Estados Unidos n° 5.986.067 de Funk *et al.* Por ejemplo, la invención incluye modificaciones en los aminoácidos S415 y T613. En una realización, el resto de transferrina contiene las modificaciones S415A y T613A. En esta divulgación, las modificaciones de aminoácidos corresponden a S415 y T613 de la SEC ID NO: 3.

- 45 También se ha reseñado que el dominio N de Tf expresado en *Pichia pastoris* llega a estar glucosilado unido a o con una sola hexosa en S32 que también se puede mutar o modificar para evitar tal glucosilación. Además, la glucosilación unida a O se puede reducir o eliminar en una célula huésped de levadura con mutaciones en uno o más de los genes *PMT*.

De acuerdo con lo anterior, en una divulgación, la proteína de fusión incluye una molécula de transferrina modificada en la que la transferrina muestra una glucosilación reducida, que incluye pero no se limita a formas asialo-monosialo- y disialo- de Tf. En otra divulgación, la parte de transferrina de la proteína de fusión incluye un mutante de transferrina recombinante que se muta para evitar la glucosilación. En otra divulgación, la parte de transferrina de la proteína de fusión incluye un mutante de transferrina recombinante que está totalmente glucosilado. En una divulgación adicional, la parte de transferrina de la proteína de fusión incluye un mutante de transferrina de suero humana recombinante que está mutado para evitar la glucosilación, en el que al menos uno de Asn413 y Asn611 de la SEC ID NO: 3 está mutado a un aminoácido que no permite glucosilación. En otra realización, la parte de transferrina de la proteína de fusión incluye un mutante de transferrina de suero humana recombinante que está mutado para evitar o sustancialmente reducir la glucosilación, en el que las mutaciones pueden estar dentro del sitio de glucosilación N-X-S/T. Además, la glucosilación se puede reducir o evitar mediante mutación del resto de serina o treonina. Por ejemplo, en una divulgación de la invención, la parte de transferrina de la proteína de fusión incluye un mutante de transferrina de suero humana recombinante que está mutado para evitar la glucosilación, en la que al menos uno de Ser415 y Thr613 de la SEC ID NO: 3 está mutado a un aminoácido que no permite la glucosilación. Además, el cambio de X a prolina se sabe que inhibe la glucosilación.

Como se describe en más detalle más adelante, las proteínas de fusión Tf modificadas, que comprenden una Tf modificada, de la invención también se puede modificar por ingeniería genética para unir y/o no unir el hierro al receptor Tf. En otras realizaciones de la invención, la unión del hierro se mantiene y la capacidad de unión al hierro de Tf se puede usar para distribuir una proteína terapéutica o péptido(s) al interior de una célula y/o mediante la barrera hematoencefálica (BBB). El dominio N solo no se une a TfR cuando está cargado con hierro y el hierro unido al dominio C se unirá a TfR pero no con la misma afinidad que la molécula global.

En otra divulgación, la parte de transferrina de la proteína de fusión de transferrina, incluye un mutante de transferrina recombinante que tiene una mutación en la que el mutante no mantiene la capacidad de unirse a iones metálicos. En una divulgación alternativa, la parte de transferrina de la proteína de fusión de transferrina incluye un mutante de transferrina recombinante que tiene una mutación en la que el mutante tiene una afinidad de unión más débil para iones metálicos que la transferrina de suero de tipo natural. En una divulgación alternativa, la parte de transferrina de la proteína de fusión de transferrina incluye un mutante de transferrina recombinante que tiene una mutación en la que el mutante tiene una afinidad de unión más fuerte por los iones metálicos que la transferrina de suero tipo natural.

En otra divulgación, la parte de transferrina incluye un mutante de transferrina recombinante que tiene una mutación en la que el mutante no retiene la capacidad de unirse a al receptor de transferrina. En una divulgación alternativa, la parte de transferrina incluye un mutante de transferrina recombinante que tiene una mutación en la que el mutante tiene una afinidad e unión más débil por el receptor de transferrina que la transferrina de suero de tipo natural. En una divulgación alternativa, la parte de transferrina incluye un mutante de transferrina recombinante que tiene una mutación en la que el mutante tiene una afinidad de unión más fuerte para el receptor de transferrina que la transferrina de suero de tipo natural.

En otra divulgación, la parte de transferrina incluye un mutante de transferrina recombinante que tiene una mutación en la que el mutante no mantiene la capacidad de unirse a los iones carbonato. En una divulgación alternativa, la parte de transferrina incluye un mutante de transferrina recombinante que tiene una mutación en la que el mutante tiene una afinidad de unión más débil para los iones carbonato que la transferrina de suero de tipo natural. En una divulgación alternativa, la parte de transferrina incluye un mutante de transferrina recombinante que tiene una mutación en la que el mutante tiene una afinidad de unión más fuerte para los iones carbonato que la transferrina de suero de tipo natural.

En otra divulgación, la parte de transferrina incluye un mutante de transferrina de suero humana recombinante que tiene una mutación en al menos un resto de aminoácido seleccionado entre el grupo que consiste en Asp63, Gly65, Tyr95, Tyr188, His249, Asp392, Tyr426, Tyr514, Tyr517 y His585 de la SEC ID NO: 3, en la que el mutante mantiene la capacidad de unirse a iones metálicos. En una divulgación alternativa, un mutante de transferrina de suero humana recombinante que tiene una mutación en al menos un resto de aminoácido seleccionado entre el grupo que consiste en Asp63, Gly65, Tyr95, Tyr188, His249, Asp392, Tyr426, Tyr514, Tyr517 y His585 de la SEC ID NO: 3, en la que el mutante tiene una capacidad reducida para unirse a iones metálicos. En otra divulgación, un mutante de transferrina de suero humana recombinante que tiene una mutación en al menos un resto de aminoácido seleccionado entre el grupo que consiste en Asp63, Gly65, Tyr95, Tyr188, His249, Asp392, Tyr426, Tyr517 y His585 de la SEC ID NO: 3, en la que el mutante no mantiene la capacidad de unirse a iones metálicos.

En otra divulgación, la parte de transferrina incluye un mutante de transferrina de suero humana recombinante que tiene una mutación en Lys206 o His207 de la SEC ID NO: 3, en la que el mutante tiene una afección de unión más fuerte para los iones metálicos que la transferrina de suero tipo natural humana (véase la Patente de Estados Unidos 5.986.067). En una divulgación alternativa, la parte de transferrina incluye un mutante de transferrina de suero humana recombinante que tiene una mutación en Lys206 o His207 de la SEC ID NO: 3, en la que el mutante tiene una afección de unión más débil para los iones metálicos que la transferrina de suero tipo natural humana. En una divulgación adicional, la parte de transferrina incluye un mutante de transferrina de suero humana recombinante que tiene una mutación en Lys206 o His207 de la SEC ID NO: 3, en la que el mutante no se une a iones metálicos.

Se puede usar cualquier técnica disponible para producir la proteína de fusión de la invención, que incluye pero no se limita a técnicas moleculares disponibles comúnmente, por ejemplo, las descritas en Sambrook *et al.* Molecular Cloning: A Laboratory Manual, 2ª Ed., Cold Spring Harbor Laboratory Press, 1989. Cuando se llevan a cabo las sustituciones de nucleótidos usando las técnicas para llevar a cabo mutagénesis específica de sitio que se conocen bien en la técnica, los cambios de aminoácidos son preferiblemente de una naturaleza secundaria, esto es, sustituciones de aminoácido conservadoras, aunque también se contemplan otras sustituciones no conservadoras, particularmente cuando se producen una parte de transferrina modificada, *por ejemplo*, una proteína de fusión modificada que muestra glucosilación reducida, unión a hierro reducida y similares. De manera específica se

contemplan sustituciones de aminoácidos, pequeñas supresiones o inserciones, de manera típica de uno a aproximadamente 30 aminoácidos; las inserciones entre dominios de transferrina; pequeñas extensiones amino- o carboxilo-terminales, tales como un resto de metionina amino-terminal, o pequeños péptidos de engarce de menos de 50, 40, 30, 20 o 10 restos entre dominios de transferrina o unión de una proteína de transferrina de y proteína o péptido terapéutico, ligando, o una región de anticuerpo variable o región de tallo; o una pequeña extensión que facilita la purificación, tal como un espacio de poli-histidina, un epítopo antigénico o un dominio de unión.

Los ejemplos de sustituciones de aminoácidos conservadoras son sustituciones realizadas dentro del mismo grupo tales como sustituciones dentro del grupo de aminoácidos básicos (tales como arginina, lisina, histidina), aminoácidos ácidos (tales como ácido glutámico y ácido aspártico), aminoácidos polares (tales como glutamina y asparagina), aminoácidos hidrófobos (tales como leucina, isoleucina, valina), aminoácidos aromáticos (tales como fenilalanina, triptófano, tirosina) y aminoácidos pequeños (tales como glicina, alanina, serina, treonina, metionina).

Las sustituciones no conservadoras abarcan sustituciones de aminoácidos de un grupo por aminoácidos de otro grupo. Por ejemplo, una sustitución no conservadora incluiría la sustitución de un aminoácido hidrófobo por un aminoácido polar. Para una descripción general de sustitución de nucleótidos, véase, *por ejemplo*, Ford *et al.* (1991), *Prot. Exp. Pur.* 2: 95 - 107. Las sustituciones no conservadoras, supresiones e inserciones son particularmente útiles para producir proteínas de fusión Tf, preferiblemente transcuerpos, de la invención que no muestran unión o muestran unión reducida de hierro y/o no muestran unión o muestran unión reducida de la proteína de fusión al receptor Tf.

En el polipéptido y proteínas de la invención, se sigue el siguiente sistema para designar aminoácidos de acuerdo con la siguiente lista convencional:

TABLA DE AMINOÁCIDOS

AMINOACIDO	SÍMBOLO DE UNA LETRA	SÍMBOLO DE TRES LETRAS
Alanina	A	Ala
Arginina	R	Arg
Asparagina	N	Asn
Acido aspártico	D	Asp
Cisteína	C	Cys
Glutamina	Q	Gln
Acido glutámico	E	Glu
Glicina	G	Gly
Histidina	H	His
Isoleucina	I	Ile
Leucina	L	Leu
Lisina	K	Lys
Metionina	M	Met
Fenilalanina	F	Phe
Prolina	P	Pro
Serina	S	Ser
Treonina	T	Thr
Triptófano	W	Trp
Tirosina	Y	Tyr
Valina	V	Val

La unión a hierro y/o unión al receptor se puede reducir o interrumpir por mutación, que incluye supresión de, sustitución de o inserción en, restos de aminoácido que corresponden a uno o más de restos de dominio N de Tf Asp63, Tyr95, Tyr188, His249 y/o restos de dominio C Asp 392, Tyr 426, Tyr 514 y/o His 585 de la SEC ID NO: 3. la unión a hierro también puede estar afectada por mutación a los aminoácidos Lys206, His207 o Arg632 de la SEC ID NO: 3. La unión a carbonato se puede reducir o interrumpir por mutación, incluyendo supresión de, sustitución de o

inserción en, restos de aminoácido que corresponden a uno o más de los residuos de dominio N de Tf Thr120, Arg124, Ala126, Gly 127 y/o residuos de dominio C Thr 452, Arg 456, Ala 458 y/o Gly 459 de la SEC ID NO: 3. Una reducción o interrupción de unión a carbonato puede afectar de manera adversa a la unión a hierro y/o receptor.

La unión al receptor de Tf receptor se puede reducir o interrumpir mediante mutación incluyendo, supresión de, sustitución de o inserción en, restos de aminoácidos que corresponden a uno o más restos de dominio N de Tf descritos anteriormente para la unión al hierro.

Como se ha descrito anteriormente, la glucosilación se puede reducir o evitar mediante mutación, que incluye supresión de, sustitución de o inserción en, restos de aminoácidos correspondientes a uno o más restos de dominio C de Tf dentro de los sitios N-X-S/T correspondientes a los restos del dominio C N413 y/o N611. Véase la Patente de Estados Unidos N° 5.986.067. Por ejemplo, el N413 y/o N611 pueden estar mutados a restos Glu como pueden ser los aminoácidos adyacentes. En una realización, S415 y/o T613 están mutados a los restos Ala.

En los casos donde las proteínas de fusión de Tf de la invención no estén modificadas para evitar la glucosilación, la unión a hierro, unión a carbonato y/o unión al receptor, glucosilación, los iones hierro y/o carbonato se pueden eliminar o escindir de la proteína de fusión. Por ejemplo, se pueden usar las desglucosilasas disponibles para escindir los restos de glucosilación de la proteína de fusión, en particular los restos de azúcar unidos a la parte Tf, levadura deficiente en enzimas de glucosilación se pueden usar para evitar la glucosilación y/o células recombinantes se pueden hacer crecer en la presencia de un agente que evite la glucosilación, *por ejemplo*, tunicamicina.

Los carbohidratos en la proteína de fusión también se pueden reducir o eliminar completamente mediante enzimas tratando la proteína de fusión con desglucosilasas. Las desglucosilasas se conocen bien en la técnica. Los ejemplos de desglucosilasas incluyen, pero no se limitan a, galactosidasa, PNGasa A, PNGasa F, glucosidasa, mannosidasa, fucosidasa y Endo H desglucosilasa.

Se pueden realizar mutaciones adicionales en Tf para alterar la estructura de tres dimensiones de Tf, tales como modificaciones en la región de articulación para evitar el cambio conformacional necesario para la unión al hierro y reconocimiento del receptor de Tf. Por ejemplo, se pueden realizar mutaciones en o alrededor de los restos de aminoácidos 94 - 96, 245 - 247 y/o 316 - 318 del dominio N así como los C restos de aminoácidos 42 - 427, 581 - 582 y/o 652 - 658 del dominio. Además, se pueden realizar mutaciones en o alrededor de las regiones flanqueantes de estos sitios para alterar la estructura y función de Tf.

En una divulgación, la proteína de fusión puede funcionar como una proteína vehículo para ampliar la semivida o biodisponibilidad del ligando así como, en algunos casos, la distribución del ligando en el interior de las células y mantiene la capacidad de cruzar la barrera hematoencefálica. En una divulgación alternativa, la proteína de fusión incluye una molécula de transferrina modificada en la que la transferrina no mantiene la capacidad de cruzar la barrera hematoencefálica.

En otra divulgación, la proteína de fusión incluye una molécula de transferrina modificada en la que la molécula de transferrina mantiene la capacidad de unirse al receptor de transferrina y transportar la región variable de anticuerpo en el interior de las células. En una divulgación alternativa, la proteína de fusión incluye una molécula de transferrina modificada en la que la molécula de transferrina no mantiene la capacidad de unirse al receptor de transferrina y transportar la región variable de anticuerpo en el interior de las células.

En divulgación adicional, la proteína de fusión incluye una molécula de transferrina modificada en la que la molécula de transferrina mantiene la capacidad de unirse al receptor de transferrina y transportar la región variable de anticuerpo en el interior de las células, pero no mantiene la capacidad de cruzar la barrera hematoencefálica. En una divulgación alternativa, la proteína de fusión incluye una molécula de transferrina modificada en la que la molécula de transferrina mantiene la capacidad de cruzar la barrera hematoencefálica, pero no mantiene la capacidad de unirse al receptor de transferrina y transportar la región variable de anticuerpo en el interior de las células.

Proteínas de fusión de transferrina

Las proteínas de fusión de la divulgación pueden contener una o más copias del ligando, región variable de anticuerpo o péptido aleatorio unido al extremo N y/o al extremo C de la Proteína Tf. En una divulgación, el resto de ligando está unido al extremo N de la proteína Tf. En algunas divulgaciones, el ligando, región variable o péptido se une al Extremo N- y C de la proteína Tf y la proteína de fusión puede contener uno o más equivalentes de estas regiones sobre cualquiera o ambos extremos de Tf.

En otras divulgaciones, el uno más ligandos están insertados en el péptido de transferrina, por ejemplo en dominios conocidos de la proteína Tf tales como en uno o más de los bucles de Tf. Véase Ali *et al.* (1999) *J. Biol. Chem.* 274 (34): 24066 - 24073.

En una divulgación, el ligando está insertado en el lóbulo N de transferrina. Por ejemplo, la invención también incluye que puedan estar realizadas una o más inserciones en o alrededor de otras posiciones en los dominios N₁ y N₂ del lóbulo N mostrados en la tabla a continuación

N ₁	N ₂
Asp33	Ser105
Asn55	Glu141
Asn75	Asp166

(continuación)

N ₁	N ₂
Asp90	Gln184
Gly257	Asp197
Lys280	Lys217
His289	Thr231
Ser298	Cys241

En todavía otra divulgación, el ligando comprende residuos de aminoácidos distribuidos al azar en los residuos de aminoácidos de superficie expuesta sobre la proteína de transferrina. Los residuos de aminoácidos distribuidos al azar se pueden localizar en una región de aminoácidos expuestos de superficie o en una o más región de aminoácidos expuestos de superficie. Por ejemplo, los aminoácidos distribuidos al azar pueden estar contenidos en una o más de las regiones constituidas por aproximadamente 85 residuos de aminoácido a aproximadamente 92 residuos de aminoácido, aproximadamente 276 residuos de aminoácido a aproximadamente 298 residuos de aminoácido, o aproximadamente 207 residuos de aminoácido a aproximadamente 217 residuos de aminoácido de la SEC ID NO: 3. En una divulgación, los residuos de aminoácido distribuidos al azar se localizan en las posiciones Y85, G86, S87, E89, D90 y Q92 de la SEC ID NO: 3. En otra divulgación, los residuos de aminoácido distribuidos al azar se localizan en las posiciones K276, D277, K280, Q283, S286 y D297 y de manera opcional S298 de la SEC ID NO: 3. En todavía otra realización, los residuos de aminoácidos distribuidos al azar se localizan en las posiciones H207, S208, F211, E212 y A215 y de manera opcional N216 y/o K217 de la SEC ID NO: 3.

Aunque el número de restos de aminoácidos distribuidos al azar por región puede variar, preferiblemente al menos aproximadamente 3, al menos aproximadamente 4, al menos aproximadamente 5, al menos aproximadamente 6, al menos aproximadamente 7, al menos aproximadamente 8, al menos aproximadamente 9 o al menos aproximadamente 10 o más restos de aminoácidos están distribuidos al azar por región.

En una divulgación, el resto de ligando contiene una secuencia de unión de ligando conocida tal como una secuencia de ligando capaz de unirse a una integrina, que incluye, pero no se limita a, una alfa 4 integrina (*por ejemplo*, a4b1 y a4b7). Tales ligandos incluyen, pero no se limitan a, VCAM y MadCAM. Véase Jackson, 2002, *Current Pharmaceutical Design*. *: 1229 - 1253.

Los procesos inflamatorios que conducen a un daño de tejido y enfermedad están mediados en parte por las integrinas alfa 4, a4b1 y a4b7, expresadas sobre la superficie de la célula de leucocitos. Estos receptores de glucoproteína modulan la adhesión de la célula mediante la interacción con sus ligandos primarios, molécula de adhesión de la célula vascular (VCAM) y molécula de adhesión de la célula de adreína mucosal (MAdCAM), expresada en el tejido afectado. Tras la unión, las interacciones de integrina/CAM en la superficie de la célula dan como resultado la adhesión firme del leucocito al vaso seguida por la entrada en el tejido afectado. La expresión de la molécula de adhesión de la célula (CAM) en diversos órganos está ligada a varias enfermedades autoinmunes. En una realización de la invención, un transcuerpo (*es decir*, una proteína de fusión que comprende transferrina y un ligando) que contiene uno o más péptidos capaces de unirse a alfa-4 integrinas o sus ligandos CAM se puede administrar a un sujeto para tratar un sujeto a tratar y/o mediar inflamación. En una divulgación, un transcuerpo que contiene uno o más péptidos capaces de unirse a alfa-4 integrinas o sus ligandos CAM se pueden administrar a un sujeto para el tratamiento de una enfermedad autoinmune.

En otra divulgación, el resto de ligando comprende una secuencia de unión a ligando conocida y restos de aminoácidos distribuidos al azar bien dentro de la secuencia de unión a ligando o bien sobre cualquiera o sobre ambos lados de la secuencia de unión a ligando. Por ejemplo, la invención incluye una proteína de fusión de transferrina que contiene la secuencia de unión a integrina RGD rodeada de restos de aminoácido distribuidos al azar. Tal proteína de fusión se puede insertar en el resto de transferrina, por ejemplo, entre los aminoácidos 289 y 290 de la SEC ID NO: 3. En una realización de la invención, la secuencia de unión a integrina resto de péptido aleatorizado de ligando es CXXXRGDXXXC, en el que X es un resto de aminoácido aleatorizado. En otra divulgación, la secuencia de unión a integrina y resto de péptido aleatorizado de ligando es XXXRGDXXX. La divulgación también incluye, pero no se limita a, la secuencia de unión a integrina KGD rodeada de péptidos distribuidos al azar y la secuencia de unión a integrina LDV rodeada de péptidos distribuidos al azar.

En otra divulgación, el resto de ligando es una región variable de anticuerpo. En esta divulgación, en general, la proteína de fusión de transferrina de la invención puede tener una región derivada de transferrina modificada y una región variable de anticuerpo. Sin embargo, más de una región variable de anticuerpo se puede usar para preparar una proteína de fusión de transferrina de la invención, produciendo por lo tanto una Proteína de fusión Tf multi-funcional modificada.

En una divulgación, la proteína de fusión de la invención contiene una región variable de anticuerpo o parte del mismo fusionada a una molécula de transferrina o parte de la misma. En otra divulgación, la proteína de fusión de la invención contiene una región variable de anticuerpo fusionada al extremo N de una molécula de transferrina. En una divulgación alternativa, la proteína de fusión de la invención contiene una región variable de anticuerpo fusionada al extremo C de una molécula de transferrina. En una divulgación adicional, la proteína de fusión de la invención contiene una molécula de transferrina fusionada al extremo N de una región variable de anticuerpo. En una divulgación alternativa, la proteína de fusión de la invención contiene una molécula de transferrina fusionada al

extremo C de una región variable de anticuerpo.

La presente invención también proporciona una proteína de fusión que contiene una región variable de anticuerpo o parte de la misma fusionada a una molécula de transferrina modificada o parte de la misma.

En otra divulgación, la proteína de fusión de la invención contiene una región variable de anticuerpo fusionada a tanto el extremo N como el extremo C de transferrina modificada. En otra divulgación, la región variable de anticuerpos fusionada a los extremos N- y C- se unen a los mismos antígenos. También, las regiones variables de anticuerpos que se unen al mismo antígeno se pueden derivar de anticuerpos diferentes y de este modo, se unen diferentes epitopos sobre la misma diana. En una divulgación alternativa, la región variable de anticuerpos fusionada a los extremos N- y C- se unen a diferentes antígenos. En otra divulgación alternativa, la región variable de anticuerpos fusionada a los extremos N- y C- se unen a diferentes antígenos que pueden ser útiles para activar dos células diferentes para el tratamiento o prevención de enfermedad, trastorno, o afección. En otra divulgación, la región variable de anticuerpos fusionada a los extremos N- y C- se unen a diferentes antígenos que pueden ser útiles para unión por puentes de dos antígenos diferentes para el tratamiento o prevención de enfermedades o trastornos que se conocen en la técnica en la técnica que se producen de manera común en pacientes de manera simultánea.

De manera adicional, la proteína de fusión de transferrina de la invención también se puede producir mediante la inserción de la región variable de anticuerpo de interés, *por ejemplo*, un anticuerpo de una sola cadena que se une a una proteína terapéutica o un fragmento o variante de la misma, en una región interna de la transferrina modificada. Las regiones internas de la transferrina modificada incluyen, pero no se limitan a, las regiones de bucle, el sitio de unión a hierro, las regiones de articulación, los sitios de unión a bicarbonato o el dominio de unión al receptor.

Dentro de la secuencia de proteína de la molécula de transferrina modificada existen un cierto número de bucles o curvas, que están estabilizados por enlaces disulfuro. Estos bucles son útiles para la inserción, o fusión interna, de péptidos terapéuticamente activos, preferiblemente regiones variables de anticuerpos, particularmente las que requieren una estructura secundaria para ser funcionales, o proteínas terapéuticas, preferiblemente región variable de anticuerpos, para generar una molécula de transferrina modificada con actividad biológica específica.

Cuando los ligandos tales como regiones variables de anticuerpos, preferiblemente CDR, se insertan en o reemplazan al menos un bucle de una molécula de Tf, se pueden realizar inserciones dentro de cualquiera de las regiones de bucles expuestas de superficie, además de otros áreas de Tf. Por ejemplo, se pueden realizar inserciones dentro de los bucles que comprenden los aminoácidos 32 - 33, 74 - 75, 256 - 257, 279 - 280 y 288 - 289 de Tf. Véase Ali *et al.*, *supra*. Como se ha descrito anteriormente, se pueden realizar inserciones dentro de otras regiones de Tf tal como los sitios para unión a hierro y a bicarbonato, las regiones de articulación y la unión al dominio del receptor como se describe en más detalle más adelante. También se pueden usar bucles en la secuencia de proteína Tf que son responsables de la modificación/reemplazo de la inserción de proteínas o péptidos para el desarrollo de de una biblioteca que se puede seleccionar de inserciones de péptidos aleatorias. Se pueden usar cualesquiera procedimientos para producir inserciones de ácido nucleico para la generación de bibliotecas de péptido, incluyendo sistemas de visualización de fago y bacterianos, antes de clonar en un dominio Tf y/o fusión a los extremos de Tf.

El extremo N de Tf está libre y apunta hacia fuera de la proteína de fusión. Las fusiones de un ligando o ligandos sobre el extremo N de transferrina es una realización de la invención. Tales fusiones pueden incluir una región de engarce, tal como pero sin limitación a un tramo de poli-glicina o un engarce PEAPTD (SEC ID NO: 18) para separar el ligando de Tf.

El extremo C de Tf parece que está enterrado o parcialmente enterrado y asegurado por un enlace disulfuro de 6 aminoácidos del extremo C. En Tf humana, el aminoácido C-terminal es una prolina que, dependiendo de la forma en que esté orientada, bien apuntará una proteína de fusión hacia fuera o bien en el cuerpo de la molécula. Un engarce o resto espaciador en el extremo C se puede usar en algunas realizaciones de la invención. También existe una prolina cerca del extremo N. En una divulgación, la prolina en los extremos N- y/o el C- se puede modificar o sustituir con otro aminoácido. En otra divulgación, el enlace disulfuro C-terminal se puede eliminar para no atar el extremo C.

Resto del tallo

El resto del tallo de la divulgación está fusionado en su extremo N a un resto de transferrina o ligando y puede de manera opcional estar fusionado a un resto de anclaje en su extremo C. Cuando se expresa en una célula de levadura, el extremo C del resto del tallo se localiza dentro de la célula, por ejemplo, dentro de la pared celular. En una realización de la invención, el resto del tallo actúa como un miembro de unión a la pared celular para unir de manera covalente o no covalente la proteína de fusión a la pared celular de una célula de levadura.

El resto del tallo de la presente invención tiene una conformación de tipo bastón o de tipo brocha. Este tipo de conformación es típica de un péptido moderadamente a altamente glucosilado. El resto del tallo de la invención contiene N-glucanos u O-glucanos. Véase el documento U.S. 6.114.147. La presencia de O-glucanos se prefiere sobre los N-glucanos debido a que los O-glucanos permiten que el resto del tallo tome la forma de una más de una conformación extendida, de tipo bastón cuando se compara con los N-glucanos. El resto del tallo también puede contener moderada a alta glucosilación de sitios de glucosilación de serina y treonina.

El resto del tallo de la proteína de fusión de la invención contiene un porcentaje moderado a alto de restos de serina o treonina. Por ejemplo, la invención incluye un resto del tallo con al menos aproximadamente 5% o más residuos de serina y/o treonina, al menos aproximadamente 10% o más residuos de serina y/o treonina, al menos aproximadamente 20% o más o más residuos de serina y/o treonina, al menos aproximadamente 30% o más o más residuos de serina y/o treonina, al menos aproximadamente 40% o más o más residuos de serina y/o treonina, al

menos aproximadamente 50% o más o más residuos de serina y/o treonina, al menos aproximadamente 60% o más o más residuos de serina y/o treonina, al menos aproximadamente 70% o más o más residuos de serina y/o treonina, al menos aproximadamente 80% o más o más residuos de serina y/o treonina, o al menos aproximadamente 90% o más o más residuos de serina y/o treonina. En una realización de la invención, el residuo del tallo contiene aproximadamente 20 - 30% residuos de serina y/o treonina, aproximadamente 20 - 40% residuos de serina y/o treonina, aproximadamente 30 - 40% residuos de serina y/o treonina, aproximadamente 20 - 50% residuos de serina y/o treonina, aproximadamente 30 - 50% residuos de serina y/o treonina, aproximadamente 20 - 60 residuos de serina y/o treonina o aproximadamente 30 - 60% residuos de serina y/o treonina.

El resto del tallo puede contener al menos aproximadamente 5% o más N- u O-glucanos en peso, al menos aproximadamente 10% o más N- u O-glucanos en peso, al menos aproximadamente 20% o más N- u O-glucanos en peso, al menos aproximadamente 30% o más N- u O- glucanos en peso, al menos aproximadamente 40% o más N- u O- glucanos en peso, al menos aproximadamente 50% o más N- u O-glucanos en peso, al menos aproximadamente 60% o más N- u O-glucanos en peso, al menos aproximadamente 70% o más N- u O-glucanos en peso, al menos aproximadamente 80% o más N- u O-glucanos en peso, o al menos aproximadamente 90% o más N- u O-glucanos en peso. En una realización de la invención, el resto del tallo contiene aproximadamente 20-30% O-glucanos en peso, aproximadamente 20 - 40% O-glucanos en peso, aproximadamente 30-40% O-glucanos en peso, aproximadamente 20 - 50% O-glucanos en peso, aproximadamente 30-50% O-glucanos en peso, aproximadamente 20 - 60% O-glucanos en peso o aproximadamente 30 - 60% O- glucanos. En otra realización, la presencia de glucanos, en particular O- glucanos, permite que el resto del tallo se reticule con los beta glucanos presentes en proteínas de la pared celular. Como tal, el resto del tallo de la invención es capaz de funcionar como miembro de unión a la pared celular.

El resto del tallo puede comprender una proteína de mucina o parte de una proteína de mucina, *es decir* un miembro de las proteínas de tipo MUC. Las mucinas de tipo MUC son una familia de las moléculas relacionadas de manera estructural que están altamente glucosiladas y se expresan en epitelio de los tractos respiratorios, gastrointestinal y reproductor, *por ejemplo*, MUC1 (GenBank Número de acceso AF125525), MUC2 (GenBank Número de acceso L21998), MUC3 (GenBank Número de acceso AF113616), MUC4 (GenBank Número de acceso AJ000281), MUC5AC (GenBank Número de acceso U83139), MUC5B (GenBank Número de acceso AJ001402), MUC6 (GenBank Número de acceso U97698), MUC7 (GenBank Número de acceso L13283), MUC8 (GenBank Número de acceso U14383), MUC9 (GenBank Número de acceso AW271430). En una divulgación, el resto del tallo contiene hMUC1 o una parte de la proteína hMUC1, *por ejemplo*, SEC ID NO: 71 codificada por el ácido nucleico de la SEC ID NO: 70 así como el polipéptido codificado por el ácido nucleico de la SEC ID NO: 5. En otra divulgación, el resto del tallo contiene hMUC3 o una parte de la proteína hMUC3. *Por ejemplo*, la invención incluye el tallo hMUC3 de la SEC ID NO: 69 que está codificado por el ácido nucleico de la SEC ID NO: 68.

La proteína de fusión de la invención también incluye tallos que comprenden variantes tales como análogos y derivados de proteínas de mucina, proteínas de tipo mucina y partes de las mismas. Las variantes se pueden crear para optimizar el despliegue de las proteínas sobre la superficie de una célula. Las variantes se pueden crear mediante procedimientos conocidos en la técnica. *Por ejemplo*, las variantes se pueden modificar por ingeniería genética eliminando los restos de aminoácidos repetitivos y/o sometiendo los péptidos a mutagénesis aleatoria y selección.

El resto del tallo de la divulgación también se puede derivar de proteínas glucosiladas diferentes de mucina, que incluyen, pero no se limitan a, AGA1 (*por ejemplo*, SEC ID NO: 73, codificada por la secuencia de ácido nucleico de la SEC ID NO: 72), MAdCAM-1, GlyCAM-1, CD34; repeticiones de consenso de E-selectina, P-selectina, o L-selectina; o fijación de glucoproteína virales (tales como virus de influenza, herpes simplex, inmunodeficiencia humana, o del mosaico de tabaco) y variantes y fragmentos de los mismos. Véase el documento WO 01/46698, Girard *et al.* (1995) Immunity 2: 113 - 123 y Van Kinken *et al.* (1998) Anal. Biochem. 265: 103 - 116, todas las cuales se incorporan en el presente documento como referencia en su totalidad. La invención incluye repeticiones de dos o más proteínas glucosiladas o fragmentos de las mismas así como las combinaciones de dos o más tipos de proteínas glucosiladas.

En otra realización de la invención, el tallo se modifica por ingeniería genética para que contenga uno o más restos de cisteína libres. El uno o más restos de cisteína libres son capaces de formar enlaces disulfuro con restos de cisteína libres de proteínas en la pared celular de una célula de levadura. La formación de uno o más enlaces disulfuro dentro de la pared celular representa otro procedimiento que se puede usar para modificar por ingeniería genética un resto del tallo capaz de funcionar como un miembro de unión a la pared celular.

Si la proteína de fusión de la invención se va a expresar en una célula de levadura, el resto del tallo es preferiblemente de suficiente longitud para atravesar la pared celular entera de una célula de levadura. Preferiblemente, el extremo N del resto del tallo se sitúa en el exterior de la pared celular, lo más preferiblemente, se extiende en una configuración de tipo bastón fuera de la célula de levadura para reducir el impedimento estérico entre el resto de transferrina y ligando y la célula de levadura huésped. El resto del tallo debe ser de al menos aproximadamente 25 aminoácidos, de al menos aproximadamente 50 aminoácidos, de al menos aproximadamente 75 aminoácidos, de al menos aproximadamente 100 aminoácidos, de al menos aproximadamente 125 aminoácidos, de al menos aproximadamente 150 aminoácidos, de al menos aproximadamente 175 aminoácidos, de al menos aproximadamente 200 aminoácidos, de al menos aproximadamente 225 aminoácidos, de al menos aproximadamente 250 aminoácidos, de al menos aproximadamente 275 aminoácidos, de al menos aproximadamente 300 aminoácidos, de al menos aproximadamente 325 aminoácidos, de al menos aproximadamente 350 aminoácidos, de al menos aproximadamente 375 aminoácidos, de al menos aproximadamente 400 aminoácidos, de al menos aproximadamente 425 aminoácidos, de al menos aproximadamente 450 aminoácidos, de al menos aproximadamente 475 aminoácidos de longitud, de al menos aproximadamente 500 aminoácidos de longitud, de al menos aproximadamente 525 aminoácidos de longitud, de al menos aproximadamente 550 aminoácidos de longitud, de al menos aproximadamente 575 aminoácidos de longitud, de al menos aproximadamente 600 aminoácidos de longitud, de al menos aproximadamente 625 aminoácidos de

longitud, o al menos aproximadamente 650 aminoácidos de longitud. En una realización, el resto del tallo es aproximadamente 500 aminoácidos de longitud. En otra realización, el resto del tallo es aproximadamente 300 a 600 aminoácidos de longitud.

Resto de anclaje

El resto de anclaje de la proteína de fusión de la presente invención es una parte de la proteína de fusión que ata de manera física la proteína de fusión a una superficie de la superficie de la célula huésped o superficie de sustrato. En una realización de la invención, el resto de anclaje ata la proteína de fusión a una membrana celular tal como, pero sin limitación a, una membrana celular de mamífero. En otra realización de la invención, el resto de anclaje puede atar o inmovilizar la proteína de fusión a una pared celular tal como, pero sin limitación a, una pared de la célula de levadura. Cuando el anclaje ata la proteína de fusión a una pared celular es un miembro de unión a la pared celular.

El resto de anclaje puede de manera transitoria unir una proteína de fusión a una pared celular o membrana celular. En una realización de la invención, el resto de anclaje de manera transitoria une una proteína de fusión a una pared celular o membrana celular que proporciona una oportunidad para que el resto del tallo se una de manera covalente o no covalente a la pared celular. Por ejemplo, la atadura transitoria de un anclaje en una célula puede permitir que los O-glucanos de un resto del tallo se reticulen con los beta glucanos de la pared celular.

En una realización de la presente invención, el resto de anclaje se adhiere a membranas celulares o paredes de microorganismos, por ejemplo, eucariotas inferiores tales como levaduras y hongos así como células de mamífero tales como líneas celulares de mamífero. El resto de anclaje puede tener un largo extremo C que lo ancla en la membrana celular o pared celular con aminoácidos tal como prolina (Kok (1990) FEMS Microbiology Reviews 87: 15 - 42).

Un resto de anclaje se puede anclar a una célula mediante el uso de un anclaje de glucosil fosfatidilinositol (GPI). Véase Conzelmann *et al.* EMBO 9: 653 - 661 y Lipke y Ovalle (1998) J. Bacteriol. 180: 3735 - 3740. Un péptido de secuencia señal GPI, tal como los péptidos señal GPI descritos en el presente documento, señala para la unión de GPI al extremo C de la proteína de fusión. La propia señal de GPI tiene tres dominios: la región que contiene el sitio de unión de GPI (el sitio ω) más el primer y segundo aminoácidos cadena abajo del sitio ω , un espaciador de 5 a 10 aminoácidos y un tramo hidrófobo de 10 a 15 aminoácidos. Una proteína que contiene la señal de GPI se escinde en el sitio ω y el extremo carboxi resultante de la proteína se une de manera covalente al resto GPI. Esta reacción se produce en el retículo endoplásmico. Estando asociados con membranas mediante el resto de GPI, las proteínas unidas a GPI después se transportan a la superficie de la célula y se mantiene sobre la membrana plasmática como proteínas ancladas a GPI si las proteínas contienen restos básicos (R y/o K) en la región menos corta de ω . Las proteínas asociadas a GPI con V, I, o L en el sitio -4/-5 de ω e Y o N en el sitio -2 de ω se incorporan en la membrana celular. Véase Hamada *et al.* (1999) J. Bacteriol. 181: 3886 - 3889; Nuoffer *et al.* (1993) J. Biol. Chem. 268: 10558 - 10563; De Nobel *et al.* (1994) Trends Cell Biol. 4: 42-45.; Hamada *et al.* (1998) Mol. Gen. Genet. 258: 53 - 59; y Van Der Vaart *et al.* (1998) Biotechnol. Genet. Eng. Rev. 15: 387 - 411.

En una realización de la invención, se usa levadura GPI YIR019C para proporcionar el resto de anclaje de la proteína de fusión de transferrina. La figura 2 proporciona un diagrama de la GPI YIR019C. El sitio ω en la secuencia de aminoácidos (SEC ID NO: 15) es glicina y se ilustra por tener un espacio sobre cualquier lado de él. Los espacios son indicativos de regiones de espaciador sobre cualquier lado del sitio ω . Los aminoácidos I e Y impresos en letra negrita enfrentada son los sitios -5/-4 y -2 de ω , respectivamente.

Se conocen en la técnica varios restos de anclajes de *Saccharomyces* y se pueden usar para construir las proteínas de fusión de la presente invención. Otros ejemplos de proteínas de señal de GPI de levadura incluyen, pero no se limitan a, YDR534C, YNL327W, YOR214C, YDR134C, YPL130, YOR009W, YER150W, YDR077W, YOR383C, YJR151C, YJR004, YJL078C, YLR110C y YNL300W. Además, las proteínas de señal de GPI se pueden usar de otros organismos tal como el GPI de EPA1 de *Canidida glabrata*, Hwp1p de *Candida albicans*, o VSG de *Trypanosoma brucei*.

En una realización de la invención, el resto de anclaje es un resto de mamífero o derivado o fragmento del mismo. En otra realización de la invención, un péptido de señal de GPI es una proteína de señal de GPI de mamífero. Por ejemplo, la presente invención incluye derivados de proteína de señal de MDP GPI humana tales como los divulgados en la Tabla1 (véase el Ejemplo 5).

La invención también incluye una proteína de fusión que comprende un resto de anclaje con uno o más restos de cisteína no unidos. Los restos de cisteína pueden actuar para atar la proteína de fusión a la célula formando enlaces disulfuro con restos de cisteína de proteínas en la pared celular.

La invención incluye proteínas de fusión que comprenden unos dominios transmembrana (TMD) como un resto de anclaje. En una realización de la invención, el TMD es una región de un solo tipo I de paso I o proteína de membrana. Por ejemplo, la invención incluye, pero no se limita a, residuos 70 - 98 de FUS1.

En otra realización de la invención, el TMD comprende una o más de las varias regiones transmembrana de una proteína de membrana multiextendida. En una realización de la invención, el TMD es una región de una proteína de membrana multiextendida que comprende aproximadamente 10 a 60 aminoácidos, aproximadamente 15 a 60 aminoácidos, aproximadamente 20 a 60 aminoácidos, aproximadamente 30 a 60 aminoácidos o aproximadamente 25 a 50 aminoácidos. Por ejemplo, la invención incluye, pero no se limita a, uno o más TMD de STE6 de *Saccharomyces* del grupo que consiste en los residuos 25 - 30, 73 - 100, 171 - 198, 249 - 277, 714 - 742, 761 - 789, 838 - 858, 864 - 884, 940 - 967 y 979-1000 (*Saccharomyces* anotación de Base de Datos del genoma).

En otra realización, el resto de anclaje se usa para atar la proteína de fusión de transferrina a un sustrato sólido tal como una microdisposición. El resto de anclaje es preferiblemente una etiqueta de epítipo corto (es decir una

secuencia reconocida por un anticuerpo, de manera típica un anticuerpo monoclonal) tal como señalización de polihistidina, SEAP, o M1 y M2. Véanse Bush *et al.* (1991) J. Biol. Chem. 266: 13811 - 13814, Berger *et al.* (1988) Gene 66: 1 - 10, Patente de Estados Unidos 5.011.912, Patente de Estados Unidos 4.851.341, Patente de Estados Unidos 4.703.004 y Patente de Estados Unidos 4.782.137, todas las cuales se incorporan por referencia en su totalidad. En una realización, el dominio del tallo se ata a un sustrato mediante un anticuerpo de secuencia anti tallo tal como un anticuerpo anti-mucina.

Albúmina

La divulgación también incluye una proteína de fusión que emplea una proteína o fragmento de proteína diferente de transferrina para "presentar" un ligando a una diana. Las proteínas adecuadas son unas que son solubles y de al menos aproximadamente 50 aminoácidos de longitud o más largas. En una realización de la invención, la proteína o fragmento de proteína contiene una estructura secundaria similar a la de la transferrina.

Es preferible que la proteína o fragmento de la misma sea capaz de incrementar la semivida del ligando cuando se escinde de la parte de tallo de la proteína de fusión y se usa como agente terapéutico. Por ejemplo, la presente divulgación contempla el uso de una proteína de fusión que contiene un resto de albúmina, un resto del tallo y un miembro de unión a la pared celular. La presente invención también contempla el uso de una proteína de fusión que contiene un resto de albúmina y un resto de anclaje. El resto de albúmina es capaz de conferir una semivida incrementada en suero para el ligando, *es decir*, agente terapéutico, cuando la parte de albúmina y de ligando de la proteína de fusión se escinde del remanente de la proteína de fusión y se administra a un paciente en necesidad del ligando como agente terapéutico.

Una proteína de fusión que contiene un resto de albúmina puede contener una proteína de albúmina, una variante de albúmina o fragmento de la misma. En una divulgación, la proteína de albúmina comprende la secuencia de aminoácidos de la SEC ID NO: 67 que está codificada por la secuencia de ácido nucleico de la SEC ID NO: 66. La invención incluye modificaciones de albúmina que se conocen en la técnica.

Ácidos nucleicos

Las moléculas de ácido nucleico también se proporcionan por la presente invención. Éstas codifican una proteína de fusión Tf modificada divulgada que comprende una proteína de transferrina o una parte de una proteína de transferrina unida o acoplada de manera covalente a un resto de ligando. La proteína de fusión puede además comprender una región de enlace, por ejemplo un engarce de menos de aproximadamente 50, 40, 30, 20, o 10 restos de aminoácido. El engarce puede estar de manera covalente a y entre la proteína de transferrina o parte de la misma y la parte de ligando. Las moléculas de ácido nucleico de la invención pueden estar purificadas o no.

También se proporcionan las células huésped y vectores para replicar las moléculas de ácido nucleico y que expresan las proteínas de fusión modificadas. Se pueden usar cualesquiera vectores o células huésped, bien procarióticos o bien eucarióticos, pero se pueden preferir los sistemas de expresión procarióticos, en particular los sistemas de expresión de levadura y sistemas de expresión de mamífero. Muchos vectores y células huésped se conocen en la técnica para tales propósitos. Está también dentro de la experiencia en la técnica seleccionar un conjunto apropiado para la aplicación deseada.

Secuencias de ADN que codifican transferrina, partes de transferrina y proteínas terapéuticas de interés se pueden clonar a partir de una diversidad de bibliotecas genómicas o de ADNc conocidas en la técnica. Las técnicas para aislar tales secuencias de ADN que usan procedimientos basados en sondas son técnicas convencionales y se bien conocen bien por los expertos en la técnica. Las sondas para aislamiento de tales secuencias de ADN se pueden basar en secuencias de ADN o de proteínas (véase, por ejemplo, Baldwin, G.S. (1993) Comparison of Transferrin Sequences from Different Species. Comp. Biochem. Physiol. 106B/1:203-218). De manera alternativa, se puede usar el procedimiento de la reacción en cadena de la polimerasa (PCR) descrito por Mullis *et al.* (Patente de Estados Unidos Nº 4.683.195) y Mullis (Patente de Estados Unidos Nº 4.683.202). La elección de biblioteca y selección de sondas para el aislamiento de tales secuencias de ADN está dentro de la experiencia ordinaria en la técnica.

Como se sabe en la técnica, "similitud" entre dos polinucleótidos o polipéptidos se determina por comparación de la secuencia de nucleótidos o aminoácidos y sus sustitutos de nucleótidos o aminoácidos de un polinucleótido o polipéptido con la secuencia de un segundo polinucleótido o polipéptido. Como se conoce en la técnica es "identidad" que significa el grado de relación de secuencia entre dos secuencias de polipéptido o dos de polinucleótidos como se determina por la identidad de la coincidencia entre dos hebras de tales secuencias. Tanto la identidad como la similitud se pueden calcular fácilmente

(Computational Molecular Biology, Lesk, A. M., ed., Oxford University Press, Nueva York, 1988; Biocomputing: Informatics and Genome Projects, Smith, D. W., ed., Academic Press, Nueva York, 1993; Computer Analysis of SEQUENCE Data, Part I, Griffin, A. M., and Griffin, H. G., eds., Humana Press, Nueva Jersey, 1994; SEQUENCE Analysis in Molecular Biology, von Heinje, G., Academic Press, 1987; y SEQUENCE Analysis Primer, Gribskov, M. y Devereux, J., eds., M Stockton Press, Nueva York, 1991).

Aunque existen numerosos procedimientos para medir la identidad y similitud entre dos secuencias de de polinucleótidos o polipéptidos, los términos "identidad" y "similitud" son bien conocidos por los expertos en la técnica (Sequence Analysis in Molecular Biology, von Heinje, G., Academic Press, 1987; Sequence Analysis Primer, Gribskov, M. and Devereux, J., eds., M Stockton Press, Nueva York, 1991; y Carillo, H. y Lipman, D., SIAM J. Applied Math., 48: 1073 (1988). Los procedimientos empleados de manera común para determinar la identidad o similitud entre dos secuencias incluyen, pero no se limitan a los descritos en Guide to Huge Computers, Martin J. Bishop, ed., Academic Press, San Diego, 1994 y Carillo, H. y Lipman, D., SIAM J. Applied Math. 48: 1073 (1988).

Los procedimientos preferidos para determinar la identidad se diseñan para proporcionar la mayor coincidencia entre

las dos secuencias ensayadas. Los procedimientos para determinar la identidad y similitud se codifican en programas de ordenador. Los procedimientos de programas de ordenador preferidos para determinar la identidad y similitud entre dos secuencias incluyen, pero no se limitan a, paquete de programa GCG (Devereux, *et al.*, Nucl. Acid Res. 12 (1):387 (1984)), BLASTP, BLASTN, FASTA (Atschul, *et al.*, J. Mol. Biol. 215: 403 (1990)). El grado de similitud o identidad mencionado antes se determina como el grado de identidad entre las dos secuencias, indicando a menudo una derivación de la primera secuencia de la segunda. El grado de identidad entre dos secuencias de ácido nucleico se puede determinar mediante programas de ordenador conocidos en la técnica tales como GAP proporcionado en el paquete de programa GCG (Needleman and Wunsch J. Mol. Biol. 48: 443 - 453 (1970)). Para los propósitos de determinar el grado de identidad entre dos secuencias de ácidos nucleicos para la presente invención, GAP se usa con las siguientes situaciones: fallo de creación de GAP de 5,0 y fallo de extensión de GAP de 0.3.

Optimización de codón

La degeneración del código genético permite variaciones de la secuencia de nucleótidos de una proteína de transferrina y/o proteína terapéutica de interés, aunque todavía se esté produciendo un polipéptido que tiene la secuencia de aminoácidos idéntica a la del polipéptido codificado por la secuencia de ADN nativo. El procedimiento, conocido como "optimización de codón" (descrito en la Patente de Estados Unidos 5.547.871) proporciona un medio de diseñar tal secuencia de ADN alterada. El diseño de genes de codón optimizado debe tener en cuenta una diversidad de factores, incluyendo la frecuencia de uso de codón en un organismo, lo más cerca de las secuencias vecinas, estabilidad de ARN, el potencial de la formación de la estructura secundaria, la vía de síntesis y las manipulaciones de ADN futuras propuestas de ese gen. En particular, los procedimientos disponibles se pueden usar para alterar los codones que codifican una proteína de fusión dada con los que se reconocen más fácilmente por la levadura cuando se usan los sistemas de expresión en levadura.

La degeneración del código genético permite que la misma secuencia de aminoácidos se codifique y traduzca de muchas maneras diferentes. Por ejemplo, leucina, serina y arginina están cada una de ellas codificada por seis codones diferentes, mientras que valina, prolina, treonina, alanina y glicina están cada una de ellas codificada por cuatro codones diferentes. Sin embargo, la frecuencia de uso de tales codones sinónimos varía de genoma en genoma entre eucariotas y procariotas. Por ejemplo, los patrones de elección de codones sinónimos entre mamíferos son muy similares, mientras que los organismos distantes de manera evolutiva tales como levadura (*S. cerevisiae*), bacterias (tal como *E. coli*) e insectos (tal como *D. melanogaster*) revelan un patrón claramente diferente de frecuencias de uso de codón genómico (Grantham, R., *et al.*, Nucl. Acid Res., 8, 49 - 62 (1980); Grantham, R., *et al.*, Nucl. Acid Res., 9, 43 - 74 (1981); Maruyama, T., *et al.*, Nucl. Acid Res., 14, 151 - 197 (1986); Aota, S., *et al.*, Nucl. Acid Res., 16, 315 - 402 (1988); Wada, K., *et al.*, Nucl. Acid Res., 19 Sup., 1981 - 1985 (1991); Kurland, C. G., FEBS Lett., 285, 165 - 169 (1991)). Estas diferencias en patrón de elección de codón parecen contribuir a los niveles de expresión global de genes individuales modulando las velocidades de elongación de péptidos. (Kurland, C. G., FEBS Lett., 285, 165 - 169 (1991); Pedersen, S., EMBO J., 3, 2895 - 2898 (1984); Sorensen, M. A., J. Mol. Biol., 207, 365 - 377 (1989); Randall, L. L., *et al.*, Eur. J. Biochem., 107, 375 - 379 (1980); Curran, J. F. y Yarus, M., J. Mol. Biol., 209, 65 - 77 (1989); Varenne, S., *et al.*, J. Mol. Biol., 180, 549 - 576 (1984); Varenne, S., *et al.*, J. Mol. Biol., 180, 549 - 576 (1984); Garel, J.-P., J. Theor. Biol., 43, 211 - 225 (1974); Ikemura, T., J. Mol. Biol., 146, 1 - 21 (1981); Ikemura, T., J. Mol. Biol., 151, 389 - 409 (1981)).

Las frecuencias de uso de codón para un gen sintético deben reflejar los usos de codón de genes nucleares derivados del genoma exacto (o tan estrechamente relacionados como sea posible) de la célula/organismo que se pretende usar para la expresión de la proteína recombinante, particularmente la de especies de levaduras. Como se ha descrito anteriormente, en una realización la secuencia de Tf humana es de codón optimizado, antes o después de la modificación como se describe en el presente documento para la expresión de levadura como puede ser la(s) secuencia(s) de nucleótidos de proteína(s) terapéutica(s).

Vectores

Las unidades de expresión para uso en la presente invención comprenderán en general los siguientes elementos, unidos de manera operativa en una orientación 5' a 3': un promotor transcripcional, una secuencia de señal secretora, una secuencia de ADN que codifica una proteína de fusión Tf modificada que comprende proteína de transferrina o una parte de una proteína de transferrina unida a una secuencia de ADN que codifica una proteína terapéutica o péptido de interés y un terminador transcripcional. Como se ha descrito anteriormente, cualquier disposición de la proteína terapéutica o péptido fusionado a o dentro de la parte Tf se puede usar en los vectores de la invención. La selección de promotores adecuados, secuencias de señal y terminadores se determinarán por la célula huésped seleccionada y será evidente para los expertos en la técnica y se describen más de manera específica más adelante.

Los vectores de levadura adecuados para uso en la presente invención se describen en la Patente de Estados Unidos 6.291.212 e incluyen YRp7 (Struhl *et al.*, Proc. Natl. Acad. Sci. EE.UU. 76: 1035 - 1039, 1978), YEpl3 (Broach *et al.*, Gene 8: 121 - 133, 1979), pJDB249 y pJDB219 (Beggs, Nature 275: 104 - 108, 1978), pPPC0005, pSeCHSA, pScNHSA, pC4 y sus derivados. Los vectores de plásmido de levadura útiles también incluyen pRS403-406, pRS413-416 y los vectores de *Pichia* disponibles de Stratagene Cloning Systems, La Jolla, CA 92037, Estados Unidos. Los plásmidos pRS403, pRS404, pRS405 y pRS406 son plásmidos que se integran en levadura (Ylps) e incorporan los marcadores que se pueden seleccionar de levadura *HIS3*, *TRP1*, *LEU2* y *URA3*. Los plásmidos pRS413-416 son plásmidos de centrómero de levadura (YCps).

Tales vectores incluirán en general un marcador que se puede seleccionar, que puede ser uno de cualquier número de genes que muestran un fenotipo dominante para el que existe un ensayo fenotípico para permitir que los transformantes se seleccionen. Los marcadores que se pueden seleccionar preferidos son los que complementan la auxotrofia de la célula huésped de complemento, proporcionan resistencia a antibióticos o permiten que una célula utilice fuentes de carbono e incluyen *LEU2* (Broach *et al.* *ibid.*), *URA3* (Botstein *et al.*, Gene 8: 17, 1979), *HIS3*

(Struhl *et al.*, *ibid.*) o *POT1* (Kawasaki y Bell, EP 171,142). Otros marcadores que se pueden seleccionar adecuados incluyen el gen *CAT*, que confiere resistencia a cloranfenicol sobre células de levadura. Los promotores preferidos para uso en levadura incluyen promotores de genes glucolíticos de levadura (Hitzeman *et al.*, J Biol. Chem. 225: 12073 - 12080, 1980; Alber y Kawasaki, J. Mol. Appl. Genet. 1: 419 - 434, 1982; Kawasaki, Patente de Estados Unidos N° 4.599.311) o genes de alcohol deshidrogenasa (Young *et al.*, en Genetic Engineering of Microorganisms for Chemicals, Hollaender *et al.*, (eds.), p. 355, Plenum, N.Y., 1982; Ammerer, Meth. Enzymol. 101: 192 - 201, 1983). A este respecto, los promotores que se pueden usar son el promotor *TP11* (Kawasaki, Patente de Estados Unidos N° 4.599.311) y el promotor *ADH2-4^c* (véase la Patente de Estados Unidos 6.291.212 (Russell *et al.*, Nature 304: 652 - 654, 1983). Las unidades de expresión también pueden incluir terminador transcriptional. Un terminador transcriptional es el terminador *TPI1* (Alber and Kawasaki, *ibid.*).

Además de levadura, las proteínas de fusión modificadas de la presente invención se pueden expresar en hongos filamentosos, por ejemplo, cepas de los hongos *Aspergillus*. Los ejemplos de promotores útiles incluyen los derivados de genes glucolíticos de *Aspergillus nidulans*, tal como el promotor *adh3* (McKnight *et al.*, EMBO J. 4: 2093 - 2099, 1985) y el promotor *tpiA*. Un ejemplo de un terminador adecuado es el terminador *adh3* (McKnight *et al.*, *ibid.*). Las unidades de expresión que utilizan tales componentes se pueden clonar en vectores que son capaces de inserción en el ADN cromosómico de *Aspergillus*, por ejemplo.

Los vectores de expresión de mamíferos para uso para llevar a cabo la presente divulgación incluirán un promotor capaz de dirigir la transcripción de la proteína de fusión Tf modificada. Los promotores incluyen, pero no se limitan a, promotores virales y promotores celulares. Por ejemplo, promotores virales incluyen el promotor tardío principal de adenovirus 2 (Kaufman and Sharp, Mol. Cell. Biol. 2: 1304 - 1319, 1982) y el promotor SV40 (Subramani *et al.*, Mol. Cell. Biol. 1: 854 - 864, 1981). Los promotores celulares incluyen el promotor de metalotioneína 1 de ratón (Palmiter *et al.*, Science 222: 809 - 814, 1983) y un promotor V6 de ratón (véase la Patente de Estados Unidos 6.291.212) (Grant *et al.*, Nuc. Acids Res. 15: 5496, 1987). Tal promotor es un promotor *V_H* de ratón (véase la Patente de Estados Unidos 6.291.212) (Loh *et al.*, *ibid.*). Tales vectores de expresión también pueden contener un conjunto de sitios de ajuste de ARN localizados cadena abajo del promotor y cadena arriba de la secuencia de ADN que codifica la proteína de fusión de transferrina. Los sitios de ajuste de ARN se pueden obtener, por ejemplo, a partir de genes de adenovirus y/o inmunoglobulina.

También contenida en los vectores de expresión es una señal de poliadenilación localizada cadena abajo de la secuencia codificante de interés. Las señales de poliadenilación incluyen las señales de poliadenilación tempranas o tardías de SV40 (Kaufman y Sharp, *ibid.*), la señal de poliadenilación de la región E1B de adenovirus 5 y el terminador del gen de la hormona de crecimiento humano (DeNoto *et al.*, Nucl. Acid Res. 9: 3719 - 3730, 1981). Una tal señal de poliadenilación es el terminador del gen *V_H* (véase la Patente de Estados Unidos 6.291.212) (Loh *et al.*, *ibid.*). Los vectores de expresión pueden incluir una secuencia guía viral no codificante, tal como la guía tripartita de adenovirus 2, localizada entre el promotor los sitios de ajuste de ARN. Los vectores preferidos también pueden incluir secuencias potenciadoras, tal como el potenciador SV40 y el potenciador de ratón: (véase la Patente de Estados Unidos 6.291.212) (Gillies, Cell 33: 717 - 728, 1983). Los vectores de expresión también pueden incluir secuencias que codifican los ARN de adenovirus VA.

Transformación

Las técnicas para transformar hongos se conocen bien en la bibliografía y se han descrito, por ejemplo, por Beggs (*ibid.*), Hinnen *et al.* (Proc. Natl. Acad. Sci. EE.UU. 75: 1929 - 1933, 1978), Yelton *et al.*, (Proc. Natl. Acad. Sci. EE.UU. 81: 1740 - 1747, 1984) y Russell (Nature 301: 167 - 169, 1983). El genotipo de la célula huésped en general contendrá un defecto genético que se complementa por el marcador que se puede seleccionar presente sobre el vector de expresión. La elección de un huésped y marcador que se puede seleccionar particular está dentro del nivel de la experiencia ordinaria en la técnica.

Las secuencias de ADN que comprenden las proteínas de fusión de Tf modificadas de la invención se pueden introducir en células de mamífero cultivadas mediante, por ejemplo, la transfección mediada por fosfato de calcio (Wigler *et al.*, célula 14: 725, 1978; Corsaro and Pearson, Somatic Cell Genetics 7: 603, 1981; Graham y Van der Eb, Virology 52: 456, 1973). Se pueden usar también otras técnicas para introducir secuencias de ADN clonadas en células de mamífero, tales como electroporación (Neumann *et al.*, EMBO J. 1: 841 - 845, 1982), o lipofección. Con el fin de identificar las células que tienen integrado el ADN clonado, un marcador que se puede seleccionar se introduce en general en las células junto con el gen o ADNc de interés. Los marcadores que se pueden seleccionar preferidos para uso en las células de mamífero cultivadas incluyen genes que confieren resistencia a fármacos, tales como neomicina, higromicina y metotrexato. El marcador que se puede seleccionar puede ser un marcador que se puede seleccionar que se puede amplificar. Un marcador que se puede seleccionar que se puede amplificar es el gen DHFR. Un marcador que se puede amplificar es DHFR^r (véase la Patente de Estados Unidos 6.291.212) ADNc (Simonsen and Levinson, Proc. Natl. Acad. Sci. EE.UU. 80: 2495 - 2499, 1983). Los marcadores que se pueden seleccionar están recapitulados por Thilly (Mammalian Cell Technology, Butterworth Publishers, Stoneham, Mass.) y la elección de marcadores que se pueden seleccionar está dentro de la experiencia ordinaria en la técnica.

Células huésped

La presente invención también incluye una célula, por ejemplo, una célula de levadura o una célula de mamífero, transformada para expresar una proteína de fusión de transferrina modificada de la invención. Además de las propias células huésped transformadas, la presente invención también incluye un cultivo de las células, por ejemplo, un cultivo monoclonal (clonalmente homogéneo), o un cultivo derivado de un cultivo monoclonal en un medio nutritivo. Si el polipéptido se secreta, el medio contendrá el polipéptido, con las células, o sin las células si se han separado por filtración o centrifugación.

Las células huésped para uso en la práctica de la presente invención incluyen células eucarióticas, en algunos casos células procarióticas, capaces de transformarse o transfectarse con ADN exógeno y crecimiento en cultivo, tales

como células de mamífero, de insecto, de hongos, de plantas y de bacterias.

Las células de hongos, que incluyen especies de levadura (por ejemplo, *Saccharomyces* spp., *Schizosaccharomyces* spp., *Pichia* spp.) se pueden usar como células huésped dentro de la presente invención. Los géneros ejemplares de levadura contemplados que son útiles en la práctica de la presente invención como huéspedes para la expresión de la proteína de fusión de transferrina de la invención son *Pichia* (incluyendo las especies anteriormente clasificadas como *Hansenula*), *Saccharomyces*, *Kluyveromyces*, *Aspergillus*, *Candida*, *Torulopsis*, *Torulaspora*, *Schizosaccharomyces*, *Citeromyces*, *Pachysolen*, *Zygosaccharomyces*, *Debaromyces*, *Trichoderma*, *Cephalosporium*, *Humicola*, *Mucor*, *Neurospora*, *Yarrowia*, *Metschnikowia*, *Rhodospiridium*, *Leucosporidium*, *Botryosascus*, *Sporidiobolus*, *Endomycopsis* y similares. Son ejemplos de *Saccharomyces* spp. *S. cerevisiae*, *S. italicus* y *S. rouxii*. Los ejemplos de *Kluyveromyces* spp. son *K. lactis* y *K. marxianus*. Una especie adecuada es *T. delbrueckii*. Son ejemplos de *Pichia* (*Hansenula*) spp. *P. angusta* (antes *H. polymorpha*), *P. anomala* (antes *H. anomala*) y *P. pastoris*.

Las células huésped particularmente útiles para producir las proteínas de fusión de Tf de la divulgación son la *Pichia pastoris* metanoltrófica (Steinlein et al. (1995) *Protein Express. Purif.* 6:619 - 624). *Pichia pastoris* ha sido desarrollada para ser un huésped destacado para la producción de proteínas extrañas ya que su promotor de la alcohol oxidasa se aisló y se clonó; su transformación se reseñó primero en 1985. *P. pastoris* puede utilizar metanol como una fuente de carbono en ausencia de glucosa. El sistema de expresión de *P. pastoris* puede usar el promotor de la alcohol oxidasa (AOX1) inducida por metanol, que controla el gen que codifica la expresión de la alcohol oxidasa, la enzima que cataliza la primera etapa en el metabolismo de metanol. Este promotor se ha caracterizado e incorporado en una serie de vectores de expresión de *P. pastoris*. Ya que las proteínas producidas en *P. pastoris* están de manera típica plegadas correctamente y secretadas en el medio, la fermentación de *P. pastoris* modificada por ingeniería genética proporciona una excelente alternativa a los sistemas de expresión de *E. coli*. Se han producido numerosas proteínas usando este sistema, incluyendo fragmento de toxina de tétanos, pertactina de *Bordetella pertussis*, albúmina de suero humano y lisozima.

La transformación de *F. oxysporum* puede, por ejemplo, llevarse a cabo como se describe por Malardier et al. (1989) *Gene* 78: 147 - 156.

Las cepas de la levadura *Saccharomyces cerevisiae* son otro huésped preferido. En una realización, se usa una célula de levadura, o de manera más específica, una célula huésped de *Saccharomyces cerevisiae* que contiene una deficiencia genética en un gen requerido para la glucosilación ligada a asparagina de glucoproteínas. Células huésped de *S. cerevisiae* que tienen tales defectos se pueden preparar usando técnicas convencionales de mutación y selección, aunque muchas cepas de levadura se han modificado para prevenir o reducir la glucosilación o hipermanosilación. Ballou et al. (*J. Biol. Chem.* 255: 5986 - 5991, 1980) han descrito el aislamiento de mutantes de la síntesis de manoproteína que son defectuosos en genes que afectan a la glucosilación asociada a asparagina. Gentzsch y Tanner (*Glycobiology* 7:481 - 486, 1997) han descrito una familia de al menos seis genes (*PMT1-6*) que codifican las enzimas responsables de la primera etapa en la O-glucosilación de proteínas en levadura. Los mutantes defectuosos en uno o más de estos genes muestran glucosilación ligada a O reducida y/o especificidad alterada de O-glucosilación.

Para optimizar la producción de las proteínas heterólogas, se puede preferir que la cepa huésped lleve una mutación, tal como la mutación *S. cerevisiae pep4* (Jones, *Genetics* 85: 23 - 33, 1977), que da como resultado la actividad proteolítica reducida. Las cepas huésped que contienen mutaciones en otras regiones que codifican proteasas son particularmente útiles para producir grandes cantidades de proteínas de fusión de Tf de la invención.

Las células huésped que contienen construcciones de ADN de la presente invención se desarrollan en un medio de crecimiento apropiado. Como se usa en el presente documento, el término "medio de crecimiento apropiado" significa un medio que contiene nutrientes requerido para el crecimiento de las células. Los nutrientes requeridos para el crecimiento de la célula pueden incluir una fuente de carbono, una fuente de nitrógeno, aminoácidos esenciales, vitaminas, minerales y factores de crecimiento. El medio de crecimiento en general seleccionará las células que contienen la construcción de ADN, por ejemplo, por selección de fármaco o deficiencia en un medio nutriente esencial que se complementa por el marcador que se puede seleccionar sobre la construcción de ADN o transfectado de manera conjunta con la construcción de ADN. Las células de levadura, por ejemplo, se desarrollan preferiblemente en un medio definido químicamente, que comprende una fuente de carbono, por ejemplo, sacarosa, una fuente de nitrógeno de no aminoácido, sales inorgánicas, vitaminas y suplementos de aminoácidos esenciales. El pH del medio se mantiene preferiblemente a un pH mayor que 2 y menor que 8, preferiblemente a pH 5,5 a 6,5. Los procedimientos para mantener un pH estable incluyen tamponación y control de pH constante, preferiblemente mediante la adición de hidróxido de sodio. Los agentes de tamponación preferidos incluyen ácido succínico y Bis-Tris (Sigma Chemical Co., San Luis, Mo.). Las células de levaduras que tienen un defecto en un gen requerido para la glucosilación ligada a asparagina se desarrollan preferiblemente en un medio que contiene un estabilizador osmótico. Un tal estabilizador osmótico es sorbitol suplementado en el medio a una concentración entre 0,1 M y 1,5 M., preferiblemente a 0,5 M o 1,0 M.

Las células de mamífero cultivadas se desarrollan en general en medios comercialmente disponibles que contienen suero o sin suero. La selección de un medio apropiado para la línea de células particular usada está dentro del nivel de la experiencia ordinaria en la técnica. Las células de mamífero transfectadas se dejan crecer durante un período de tiempo, de manera típica 1 - 2 días, para comenzar la expresión de la(s) secuencia(s) de ADN de interés. La selección de fármacos se aplica después para seleccionar el crecimiento de las células que expresan el marcador que se puede seleccionar de una manera estable. Para las células que se han transfectado con un marcador que se puede seleccionar que se puede amplificar la concentración de fármaco se puede aumentar de una manera por etapas para seleccionar el número de copias incrementado de las secuencias clonadas, incrementando por lo tanto los niveles de expresión.

Los sistemas de expresión de baculovirus/células de insectos también se pueden usar para producir las proteínas de

fusión de Tf modificadas de la invención. El Sistema de expresión de Baculovirus BacPAK™ (BD Biosciences (Clontech)) expresa las proteínas recombinantes a altos niveles en células huésped de insecto. El gen diana se inserta en un vector de transferencia, que se transfecta de manera conjunta en células huésped de insecto con el ADN viral de BacPAK6 linealizado. El ADN de BacPAK6 ADN está ausente en una parte esencial del genoma de baculovirus. Cuando el ADN se recombina con el vector, se reestablece el elemento esencial y el gen diana se transfiere al genoma del baculovirus. Después de la recombinación, se recogen unas pocas placas virales y se purifican y se verifica el fenotipo recombinante. El virus recombinante recientemente aislado después se puede amplificar y usar para infectar los cultivos de células de insecto para producir grandes cantidades de la proteína deseada.

Secuencias de señal secretoras

Los términos “secuencia de señal secretora” o “secuencia de señal” o “secuencia guía de secreción” se usan indistintamente y se describen en, por ejemplo la Patente de Estados Unidos 6.291.212 y la Patente de Estados Unidos 5.547.871, las cuales se incorporan en el presente documento como referencia en su totalidad. Las secuencias de señal secretoras o secuencias de señal o secuencias guía de secreción codifican péptidos secretores. Un péptido secretor es una secuencia de aminoácidos que actúa para dirigir la secreción de un polipéptido o proteína maduro de una célula. Los péptidos secretores están en general caracterizados por un núcleo de aminoácidos hidrófobos y de manera típica (pero no exclusivamente) se encuentran en los extremos amino de las proteínas recientemente sintetizadas. Muy a menudo el péptido secretor se escinde de la proteína madura durante la secreción. Los péptidos secretores pueden contener sitios de procesamiento que permiten la escisión del péptido de señal de la proteína madura ya que pasa a través de la ruta secretora. Los sitios de procesamiento pueden estar codificados dentro del péptido de señal o se pueden añadir al péptido de señal, por ejemplo, mutagénesis *in vitro*.

Los péptidos secretores se pueden usar para dirigir la secreción de proteínas de fusión de Tf modificadas de la invención. Un péptido secretor tal que se puede usar en combinación con otros péptidos secretores es el tercer dominio de la proteína de barrera de levadura. Las secuencias de señal secretoras o secuencias de señal o secuencias guía de secreción se requieren para una serie compleja de etapas de procesamiento después de la traducción de una proteína. Si una secuencia de señal intacta está presente, la proteína que se expresa entra en el lumen del retículo endoplásmico rugoso y después se transporta a través del aparato de Golgi a vesículas secretoras y finalmente se transporta al exterior de la célula. En general, a la secuencia de señal sigue inmediatamente el codón de inicio y codifica un péptido de señal en el extremo amino-terminal de la proteína a secretar. En la mayoría de los casos, la secuencia de señal se retira por escisión mediante una proteasa específica, llamada una peptidasa de señal. Las secuencias de señal preferidas mejoran el procesamiento y eficacia de exportación de la expresión de la proteína recombinante usando vectores de expresión virales, de mamíferos o levaduras. En algunos casos, la secuencia de señal de Tf nativa se puede usar para expresar y secretar las proteínas de fusión de la invención.

Engarces

El resto Tf y el ligando de las proteínas de transferrina de fusión modificadas de la divulgación se pueden fusionar directamente o usando un péptido de engarce de diversas longitudes para proporcionar mayor separación física y permitir más movilidad espacial entre las proteínas fusionadas y de esta manera maximizar la accesibilidad de la región variable de anticuerpo, por ejemplo, para unirse a su receptor afín. El péptido de engarce puede constar de aminoácidos que son flexibles o más rígidos. Una divulgación incluye un engarce sustancialmente no helicoidal tal como (PEAPTD)_n (SEC ID NO: 18). En otra divulgación, la proteína de fusión de la invención contiene un engarce con un tramo poli-glicina. El engarce puede ser de menos de aproximadamente 50, 40, 30, 20, o 10 restos de aminoácido. El engarce se puede unir de manera covalente a y entre la proteína de transferrina o porción de la misma y la región variable de anticuerpo.

También se pueden usar engarces para unir la región variable de anticuerpos dentro de un ligando o ligandos. Los engarces adecuados para unir la región variable de anticuerpos son los que permiten que las regiones variables de anticuerpos se plieguen en una estructura de tres dimensiones que mantiene la especificidad de unión de un anticuerpo completo.

Procedimientos de selección

El número de moléculas diana posible para el que se pueden identificar ligandos mediante la selección de bibliotecas de proteínas de fusión de la presente invención es virtualmente ilimitado. Por ejemplo, la molécula diana, es decir receptor o agente, puede ser un anticuerpo (o parte de unión del mismo) o antígeno. El antígeno al que se une el anticuerpo puede conocerse y quizás incluso secuenciarse, en cuyo caso la divulgación se puede usar para mapear los epítopos del antígeno. Si no se conoce el antígeno, tal como con ciertas enfermedades autoinmunes, por ejemplo, suero, fluidos, tejido, o células de pacientes con la enfermedad se pueden usar en el presente procedimiento de selección para identificar péptidos y por consiguiente el antígeno, que provocan la respuesta autoinmune. Una vez se ha identificado un péptido, ese péptido puede servir como, o proporcionar la base para, el desarrollo de una vacuna, un agente terapéutico, un reactivo de diagnóstico, etc. Véase el documento WO 01/46698 para una lista de moléculas diana sobre la que se pueden seleccionar los ligandos.

La selección se puede realizar mediante el uso de los procedimientos bien conocidos por los expertos en la técnica, tales como bioafinidad, FACS o MACS. En una realización de la invención, la selección se realiza para la activación del receptor. La diana puede estar bien purificada y en solución o bien unida a superficie o asociada a célula. La diana puede estar marcada, por ejemplo, con biotina o mediante otros procedimientos conocidos en la técnica.

Los polipéptidos y péptidos que tienen la propiedad deseada se pueden aislar e identificar mediante la secuenciación de la secuencia de ácidos nucleicos correspondiente o mediante secuenciación de aminoácidos o espectrometría de masas. La optimización posterior se puede realizar mediante repetición del reemplazo de las subsecuencias por secuencias diferentes, preferiblemente mediante secuencias aleatorias y la etapa de selección una o más veces.

Una vez que se ha construido una biblioteca de péptidos, las células huésped se transforman con los vectores de biblioteca. Los transformantes exitosos se seleccionan de manera típica mediante crecimiento en un medio selectivo o en condiciones selectivas, *por ejemplo*, un medio de crecimiento apropiado u otros dependiendo del vector usado. Esta selección se puede hacer sobre medio de crecimiento sólido o líquido. Para el desarrollo de células bacterianas sobre medio sólido, las células se desarrollan a una alta densidad (aproximadamente 10^8 a 10^9 transformantes por m^2) sobre una gran superficie de, *por ejemplo*, L-agar que contiene el antibiótico selectivo para formar esencialmente un césped confluyente. Para el desarrollo en cultivo líquido, las células se pueden desarrollar en medio de cultivo L (con selección de antibiótico) a través de aproximadamente 10 o más duplicaciones. El desarrollo en cultivo líquido puede ser más conveniente debido al tamaño de las bibliotecas, mientras el desarrollo sobre medio sólido del mismo proporciona menos oportunidad de desviación durante el procedimiento de amplificación.

Si se va a seleccionar una biblioteca de péptidos de proteína de fusión de transferrina mediante despliegue en la superficie de la células de levadura, las células de levadura se transformarán con el vector de expresión que codifica la proteína de fusión de transferrina. Es consistente una gama completa de procedimientos de mutagénesis con la construcción de biblioteca de despliegue en la superficie de levadura tal como reacción en cadena de la polimerasa propensa a error y recomposición de ADN. Véase Boder *et al.* (2000) *Methods of Enzymology* 328: 430 - 444. De manera alternativa, el resto de transferrina de las proteínas de fusión expresadas pueden servir como una estructura para las secuencias de péptido aleatorio o CDR.

Se conocen en la técnica varios planteamientos para identificar los péptidos deseables una vez que se ha creado una biblioteca de péptidos de proteína de fusión de transferrina de célula de levadura. Por ejemplo, los péptidos se pueden distinguir mediante la unión equilibrada con bajas concentraciones de diana marcada de manera fluorescente, *es decir* receptor o agente, en los casos de concentraciones de bastante baja afinidad ($K_d > nM$, o sin afinidad si la biblioteca se está seleccionando para aislar especificidad de unión novedosa). Para las aplicaciones diseñadas para desarrollar proteínas de fuerte unión, pueden ser necesarios grandes volúmenes de soluciones diana diluidas para mantener un exceso molar de ligando, complicando la manipulación de las muestras. En tales casos, las mejoras en la afinidad de unión se pueden aproximar mediante cambios en la cinética de disociación. La competición cinética por una diana estequiométricamente limitante se puede usar para identificar clones mejorados dentro de la población (Hawkins *et al.* (1992) *J. Mol. Biol.* 226: 889); sin embargo, este planteamiento elimina las cantidades predecibles del planteamiento de selección y no se recomienda en general. Véase Boder *et al.* (2000) *Methods of Enzymology* 328: 430 - 444.

Las dianas pueden estar biotiniladas o marcadas de manera fluorescente, o de manera alternativa, se puede marcar un ligando de interés, *es decir* un péptido desplegado sobre transferrina. Preferiblemente, las dianas están marcadas. Las dianas marcadas, *por ejemplo*, dianas biotiniladas, se pueden incubar con una biblioteca de péptidos de proteína de fusión de transferrina. La biblioteca puede tener al menos aproximadamente 10^4 miembros (*es decir* péptidos desplegados), al menos aproximadamente 10^5 miembros, al menos aproximadamente 10^6 miembros, al menos aproximadamente 10^7 miembros, al menos aproximadamente 10^8 miembros, al menos aproximadamente 10^9 miembros, al menos aproximadamente 10^{10} miembros, al menos aproximadamente 10^{11} miembros, al menos aproximadamente 10^{12} miembros, al menos aproximadamente 10^{13} miembros, al menos aproximadamente 10^{14} miembros, al menos aproximadamente 10^{15} miembros, o al menos aproximadamente 10^{16} miembros.

Después de la incubación, las células se pueden marcar con una segunda etiqueta tal como anticuerpos secundarios, una molécula marcada con estreptavidina, u otro procedimiento conocido en la técnica. El anticuerpo secundario puede ser un anticuerpo anti-biotina. Las moléculas marcadas con estreptavidina, incluyen, pero no se limitan a, microperlas de estreptavidina-ficoeritrina o estreptavidina.

Se puede usar citometría de flujo para analizar poblaciones de células como se conoce en la técnica. Cuando se hace esto, solamente se analiza la fracción de despliegue de la población. Véase Boder *et al.* (2000) *Methods of Enzymology* 328: 430 - 444 y Kondo *et al.* (2004) *Appl. Microbiol. Biotechnol.* 64: 28 - 40, las cuales se incorporan en el presente documento como referencia en su totalidad.

De manera alternativa, si se usa un segundo marcado que consiste en perlas marcadas, *es decir* perlas marcadas con anti-biotina o estreptavidina, la mezcla de ligandos y moléculas diana se pueden aislar usando un protocolo de aislamiento como se describe en Yeung *et al.* (2002) *Biotechnol. Prog.* 18: 212 - 220. Se puede usar un kit MACS® MicroBeads con este protocolo de selección (Miltenyi Biotec GmbH). El aislamiento magnético se puede usar junto con FACS.

En una divulgación, es deseable caracterizar un solo ligando de interés expresado en una célula de levadura. La proteína expresada se puede seleccionar de una diversidad de formas. Si la proteína tiene una función se puede ensayar directamente. Por ejemplo, los anticuerpos de una sola cadena expresados sobre la superficie de levadura son completamente funcionales y se pueden seleccionar basándose en la unión a un antígeno. Si la proteína no tiene una función detectable que se pueda ensayar fácilmente, la expresión del ligando se puede controlar usando un anticuerpo. Debido a que una célula de levadura es mucho mayor que el fago, se puede usar citometría de flujo para controlar el fenotipo de la proteína sobre una sola célula de levadura.

En otra realización de la presente invención, la unión del resto de ligando con un receptor o agente se realiza mediante medios conocidos en la técnica, distintos del despliegue sobre la superficie de la célula, tal como mediante ELISA, ensayos de unión competitiva cuando se conoce la pareja de unión nativa de la diana, ensayos de sándwich, ensayos de radioreceptor usando un ligando radiactivo cuyo enlace está bloqueado por la biblioteca de péptido, *etc.* En estos procedimientos, las células huésped transformadas con la biblioteca de péptidos de la proteína de fusión Tf se lisan. Los péptidos de la proteína de fusión Tf están anclados al sustrato de ensayo mediante un resto de anclaje apropiado tal como, pero no limitado a, un anticuerpo anti-MUC1. El procedimiento de selección implica la reacción de la biblioteca de péptidos de Tf con la diana de interés para establecer un nivel de unión de línea de base contra el que las actividades de unión de las bibliotecas de péptidos posteriores se comparan. La naturaleza del ensayo no es crítica mientras sea suficientemente sensible para detectar pequeñas cantidades de unión de péptidos a o

competición de unión para la unión a la diana. Las condiciones de ensayo se pueden variar para tener en cuenta las condiciones de unión óptimas para diferentes sustancias de unión de interés u otras actividades biológicas. De este modo, el pH, temperatura, concentración salina, volumen y duración de unión, *etc.* se pueden usar para lograr la unión del péptido a una diana en condiciones que se parecen a las del ambiente de interés.

Una vez que se determina que la biblioteca de los péptidos de Tf posee un péptido o péptidos que se unen a la diana de interés, los procedimientos de la invención se pueden usar para identificar la secuencia del (de los) péptido(s) en la mezcla. Los péptidos que se despliegan sobre las células que se unen a la diana se pueden aislar de la población general de la biblioteca mediante selección de MACS o FACS. El procedimiento de selección se repite durante 2 a 3 veces sobre los aislamientos iniciales para reducir drásticamente cualesquiera enlaces no específicos. Después se realiza una ronda final de selección mediante aislamiento de FACS basándose en la afinidad de unión. Se recupera el ADN de plásmido a partir de células aisladas y el ADN para la región de la inserción se secuencia para determinar la secuencia de proteínas. Después se pueden determinar los restos comunes entre los aislamientos.

Moléculas de ligando terapéuticas

Los ligandos de la divulgación pueden ser moléculas terapéuticas supuestas o conocidas. Como se usa en el presente documento, una molécula terapéutica es de manera típica una proteína o péptido capaz de ejercer un efecto biológico beneficioso *in vitro* o *in vivo* e incluye proteínas o péptidos que ejercen un efecto beneficioso en relación con la homeostasis normal, fisiología o un estado patológico. Las moléculas terapéuticas no incluyen parejas de fusión usadas comúnmente como marcadores o auxiliares de purificación de proteína, tales como galactosidasas (véase por ejemplo, la Patente de Estados Unidos 5.986.067 y Aldred *et al.* (1984) *Biochem. Biophys. Res. Commun.* 122: 960 - 965). Por ejemplo, un efecto beneficioso relacionado con un estado patológico incluye cualquier efecto que es ventajoso para el sujeto tratado, incluyendo prevención de enfermedad, estabilización de enfermedad, la reducción o alivio de los síntomas de enfermedad o una modulación, alivio o cura del defecto subyacente para producir un efecto beneficioso al sujeto tratado.

Se puede fusionar un ligando terapéutico directamente a un resto de transferrina o indirectamente mediante un resto de engarce como se ha descrito previamente. En una divulgación, puede ser deseable escindir la proteína de fusión para separar la parte de transferrina y ligando de la proteína de fusión del remanente de la proteína de fusión. En otra realización, puede ser deseable escindir el ligando del remanente de la proteína de fusión.

El resto de ligando de la proteína de fusión de la invención puede contener al menos un fragmento o variante de una proteína terapéutica, y/o al menos un fragmento o variante de un anticuerpo. En una divulgación adicional, las proteínas de fusión pueden contener fragmentos de péptido o variantes de péptido de proteínas o anticuerpos en las que la variante o fragmento retiene al menos una actividad terapéutica o biológica. Las proteínas de fusión pueden contener proteínas terapéuticas que pueden ser fragmentos de péptido o variantes de péptido de al menos aproximadamente 3, al menos aproximadamente 4, al menos 5, al menos 6, al menos 7, al menos 8, al menos 9, al menos 10, al menos 11, al menos 12, al menos 13, al menos 14, al menos 15, al menos 20, al menos 25, al menos 30, al menos 35, o al menos aproximadamente 40, al menos aproximadamente 50, al menos aproximadamente 55, al menos aproximadamente 60 o al menos aproximadamente 70 o más aminoácidos de longitud fusionados a los extremos N y/o C, insertados dentro, o insertados en un bucle de una transferrina modificada.

En otra divulgación, el resto de ligando de la proteína de fusión de la presente invención contiene una parte de proteína terapéutica que puede ser fragmentos de una proteína terapéutica que incluyen la proteína de longitud completa así como polipéptidos que tienen uno o más restos suprimidos del extremo amino de la secuencia de aminoácidos.

En otra divulgación, el resto de ligando de la proteína de fusión de la presente invención contiene una parte de la proteína terapéutica que puede ser fragmentos de una proteína terapéutica que incluyen la proteína de longitud completa así como polipéptidos que tienen uno o más restos suprimidos del extremo carboxi de la secuencia de aminoácidos.

En otra divulgación, el resto de ligando de las proteínas de fusión de la presente invención contienen una parte de la proteína terapéutica que pueden tener uno o más aminoácidos suprimidos de ambos extremos amino y carboxi.

En otra divulgación, la proteína de fusión contiene una parte de proteína terapéutica, es *decir* resto de ligando, que es al menos aproximadamente el 80%, 85%, 90%, 95%, 96%, 97%, 98% o 99% idéntica a una proteína terapéutica de referencia establecida en el presente documento, o sus fragmentos. En realizaciones adicionales, las moléculas de fusión de transferrina contienen una parte de la proteína terapéutica que es al menos aproximadamente 80%, 85%, 90%, 95%, 96%, 97%, 98% o 99% idéntica a los polipéptidos que tienen la secuencia de aminoácidos de las supresiones N- y C-terminal como se ha descrito anteriormente.

En otra divulgación, la proteína de fusión contiene la parte de la proteína terapéutica que es al menos aproximadamente el 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% o 100%, idéntica a, por ejemplo, la secuencia de aminoácidos nativa o de tipo natural de una proteína terapéutica. Los fragmentos, de estos polipéptidos también se proporcionan.

Las proteínas terapéuticas que corresponden a una parte de la proteína terapéutica de una proteína de fusión de transferrina modificada de la invención, tal como la superficie de la célula y proteínas secretoras, se pueden modificar mediante unión de uno o más grupos oligosacarídicos. La modificación denominada glucosilación, puede afectar de manera significativa a las propiedades físicas de proteínas y puede ser importante en la estabilidad de la proteína, secreción y localización. La glucosilación se produce en localizaciones específicas junto con la estructura central del polipéptido. Existen usualmente dos tipos principales de glucosilación: glucosilación caracterizada por la oligosacáridos ligados a O, que están unidos a restos de serina o treonina; y glucosilación caracterizada por oligosacáridos ligados a N, que están unidos a restos de asparagina en una secuencia Asn-X-Ser/Thr, donde X

puede ser un aminoácido salvo prolina. Las variables tales como estructura de proteína y tipo de célula influyen en el número y naturaleza de las unidades de carbohidrato dentro de las cadenas en diferentes sitios de glucosilación. Los isómeros de glucosilación son también comunes en el mismo sitio dentro de un tipo de célula dada. Por ejemplo, varios tipos de interferón humano están glucosilados.

Las proteínas terapéuticas que corresponden a una parte de la proteína terapéutica de una proteína de fusión de la divulgación, así como sus análogos y variantes, se pueden modificar de manera que la glucosilación en uno o más sitios se altera como un resultado de la manipulación(es) de su secuencia de ácido nucleico por la célula huésped en las que se expresan, o debido a otras condiciones de su expresión. Por ejemplo, se pueden producir los isómeros de glucosilación mediante abolición de o introducción de los sitios de glucosilación, *por ejemplo*, mediante sustitución o supresión de restos de aminoácido, tales como sustitución de glutamina para asparagina, o proteínas recombinantes no glucosiladas se pueden producir mediante la expresión de las proteínas en las células huésped que no las glucosilarán, *por ejemplo*, en levadura deficiente en glucosilación. Estos planteamientos se conocen en la técnica.

Las proteínas terapéuticas y sus secuencias de ácido nucleico se conocen bien en la técnica y están disponibles en las bases de datos públicas tales como las bases de datos de los servicios de los Resúmenes de Química (*por ejemplo*, el Registro CAS), GenBank y GenSEC.

La presente divulgación se refiere además a proteínas de fusión que comprenden fragmentos de las proteínas terapéuticas descritas en el presente documento. Incluso si la supresión de uno o más aminoácidos del extremo N de una proteína da como resultado la modificación o pérdida de una o más funciones biológicas de la parte de proteína terapéutica, otras actividades y/o actividades funcionales (*por ejemplo*, actividades biológicas, capacidad de multimerizarse, capacidad de unirse a un ligando) se pueden todavía retener. Por ejemplo, la capacidad de los polipéptidos con supresiones N-terminales para inducir y/o unirse a anticuerpos que reconocen las formas completas o maduras de los polipéptidos en general se retendrán con menos de la mayoría de los restos del polipéptido completo del extremo N. Si un polipéptido particular que carece de los restos N-terminales de un polipéptido completo retiene tales actividades inmunológicas se pueden ensayar mediante procedimientos de rutina descritos en el presente documento y conocidos de otra manera en la técnica. No es probable que un mutante con un gran número de residuos de aminoácidos N-terminal suprimidos pueda retener algunas actividades biológicas o inmunogénicas. De hecho, los péptidos compuestos de tan pocos como seis residuos de aminoácidos pueden a menudo inducir una respuesta inmune.

También como se ha mencionado anteriormente, incluso si la supresión de uno o más aminoácidos del extremo N o extremo C de una proteína terapéutica da como resultado la modificación o pérdida de uno o más funciones biológicas de la proteína, otras actividades funcionales, *por ejemplo*, actividades biológicas, capacidad de multimerizarse, capacidad de unirse a un ligando, y/o actividades terapéuticas todavía se pueden retener. Por ejemplo la capacidad de polipéptidos con supresiones C-terminales para inducir y/o unirse a anticuerpos que reconocen las formas completas o maduras del polipéptido en general se retendrán cuando menos de la mayoría de los residuos del polipéptido completo o maduro se eliminan del extremo C. Si un polipéptido particular que carece de los residuos N-terminales y/o, C-terminales de un polipéptido de referencia retiene la actividad terapéutica se puede determinar fácilmente mediante procedimientos descritos en el presente documento y/o de otra manera conocidos en la técnica.

Los fragmentos de péptido de las proteínas terapéuticas pueden ser fragmentos que comprenden, o de manera alternativa, consisten en, una secuencia de aminoácidos que muestra una actividad terapéutica y/o actividad funcional, *por ejemplo*, actividad biológica, de la secuencia de polipéptidos de la proteína terapéutica de la que la secuencia de aminoácidos es un fragmento.

Otros fragmentos de polipéptido son fragmentos biológicamente activos. Los fragmentos biológicamente activos son los que muestran actividad similar, pero no necesariamente idéntica a una actividad de una proteína terapéutica usada en la presente invención. La actividad biológica de los fragmentos puede incluir una actividad deseada mejorada, o una actividad no deseable disminuida.

En general, las variantes de proteínas son en conjunto muy similares y en muchas regiones, idénticas a la secuencia de aminoácidos de la proteína terapéutica que corresponde a una parte de la proteína terapéutica de una proteína de fusión de transferrina de la invención. Los ácidos nucleicos que codifican estas variantes están también abarcados por la invención.

Además los polipéptidos terapéuticos que se pueden usar en la divulgación son polipéptidos codificados por los polinucleótidos que se hibridan al complemento de una molécula de ácido nucleico que codifica una secuencia de aminoácidos de una proteína terapéutica en condiciones de hibridación rigurosas que son conocidas por los expertos en la técnica. Véase, por ejemplo, Ausubel, F.M. *et al.*, eds., 1989 Current protocol in Molecular Biology, Green Publishing Associates, Inc. y John Wiley & Sons Inc., Nueva York. Los polinucleótidos que codifican estos polipéptidos están también abarcados por la invención.

Por un polipéptido que tiene una secuencia de aminoácidos al menos, por ejemplo, 95% "idéntica" a una secuencia de aminoácidos de investigación de la presente invención, se pretende que la secuencia de aminoácidos del polipéptido sujeto sea idéntica a la secuencia de investigación salvo porque la secuencia de polipéptido sujeto puede incluir hasta cinco alteraciones de aminoácido por cada 100 aminoácidos de la secuencia de aminoácidos de investigación. En otras palabras, para obtener un polipéptido que tiene una secuencia de aminoácidos al menos 95% idéntica a una secuencia de aminoácidos de investigación, hasta el 5% de los restos de aminoácidos en la secuencia sujeto se pueden insertar, suprimir, o sustituir con otro aminoácido. Estas alteraciones de la secuencia de referencia se puede producir en las posiciones amino- o carboxi-terminal de la secuencia de aminoácidos de referencia o en cualquier parte entre las posiciones terminales, entremezclados o bien individuales entre los restos en la secuencia de referencia, o en uno o más grupos contiguos dentro de la secuencia de referencia.

Como un tema práctico, si cualquier polipéptido particular es al menos aproximadamente el 80%, 85%, 90%, 95%, 96%, 97%, 98% o 99% idéntico a, por ejemplo, la secuencia de aminoácidos de una proteína de fusión de la invención o un fragmento de la misma (tal, como la parte de la proteína terapéutica de la proteína de fusión o parte de la misma), se puede determinar de manera convencional usando programas de ordenador conocidos. Un procedimiento para determinar la mejor coincidencia global entre una secuencia de investigación (una secuencia de la presente invención) y una secuencia sujeto, también denominada una alineación de secuencia global, se puede determinar usando el programa de ordenador FASTDB basado en el algoritmo de Brufiga *et al.* (Comp. App. Biosci 245- (1990)).

Las variantes polinucleotídicas de la divulgación pueden contener alteraciones en las regiones codificantes, regiones no codificantes, o ambas. Las variantes de polinucleótido que contienen alteraciones que producen sustituciones silenciosas, adiciones o supresiones, pero no alteran las propiedades o actividades del polipéptido codificado se pueden usar para producir restos de ligando modificados. Se pueden utilizar las variantes de nucleótido producidas por las sustituciones silenciosas debidas a la degeneración del código genético. Además, también se pueden utilizar las variantes de polipéptido en las que menos de aproximadamente 50, menos de 40, menos de 30, menos de 20, menos de 10, o 5 - 50, 5 - 25, 5 - 10, 1 - 5, o 1 - 2 aminoácidos están sustituidos, suprimidos, o añadidos en cualquier combinación. Las variantes de polinucleótido se pueden producir por una diversidad de razones, *por ejemplo*, para optimizar la expresión de codones para un huésped particular (los codones de cambio en el ARNm humano a los preferidos por un huésped, tal como, levadura o *E. coli* como se ha descrito anteriormente).

En otras divulgaciones, el resto de proteína terapéutica, es *decir*, resto de ligando, tiene sustituciones conservadoras comparadas con la secuencia de tipo natural. Por "sustituciones conservadoras" se pretenden intercambios dentro de grupos tales como reemplazos de los aminoácidos alifáticos o hidrófobos Ala, Val, Leu e Ile; reemplazo de los residuos hidroxilo Ser y Thr; reemplazo de los residuos ácidos Asp y Glu; reemplazo de los residuos amida Asn y Gln, reemplazo de los residuos básicos Lys, Arg e His; reemplazo de los residuos aromáticos Phe, Tyr y Trp y reemplazo de los aminoácidos de tamaño pequeño Ala, Ser, Thr, Met y Gly. La guía con relación a cómo preparar sustituciones de aminoácidos silenciosos fenotípicamente se proporciona, por ejemplo, en Bowie *et al.*, "Deciphering the Message in Protein Sequences: Tolerance to Amino Acid Substitutions," Science 247:1306 - 1310 (1990). En divulgaciones específicas, los polipéptidos de la invención comprenden, o de manera alternativa, constan de, fragmentos o variantes de la secuencia de aminoácidos de una proteína terapéutica descrita en el presente documento y/o transferrina de suero, y/o proteína de transferrina modificada de la invención, en la que los fragmentos o variantes tienen 1 - 5, 5 - 10, 5 - 25, 5 - 50, 10 - 50 o 50 - 150 adiciones, sustituciones, y/o supresiones de restos de aminoácido cuando se comparan con la secuencia de aminoácidos de referencia. En realizaciones adicionales, las sustituciones de aminoácidos son conservadoras. Los ácidos nucleicos que codifican estos polipéptidos también están abarcados por la divulgación.

Las proteínas de fusión modificadas de la presente invención se pueden componer de aminoácidos unidos entre sí por enlaces peptídicos o enlaces péptidos modificados y pueden contener aminoácidos distintos de los 20 aminoácidos codificados por genes. Los polipéptidos se pueden modificar mediante cualesquiera procedimientos naturales, tales como procesamiento después de la traducción, o mediante técnicas de modificación química que se conocen bien en la técnica. Tales modificaciones se describen bien en los textos básicos y en monografías más detalladas, así como en una bibliografía de investigación voluminosa.

Las modificaciones se pueden producir en cualquier lugar en un polipéptido, incluyendo la estructura central del péptido, las cadenas laterales de aminoácido y los extremos amino o carboxi. Se apreciará que el mismo tipo de modificación en los mismos o en grados variables en varios sitios de un polipéptido dado. También, un polipéptido dado puede contener muchos tipos de modificaciones. Los polipéptidos pueden ser ramificados, por ejemplo, como resultado de la ubiquitinación y pueden ser cíclicos, con o sin ramificación. Los polipéptidos cíclicos, ramificados y cíclicos ramificados se pueden producir a partir de procedimientos naturales de postraducción o se pueden preparar mediante procedimientos sintéticos. Las modificaciones incluyen acetilación, acilación, ADP-ribosilación, amidación, unión covalente de flavina, unión covalente de un resto hemo, unión covalente de un nucleótido o derivado de nucleótido, unión covalente de un lípido o derivado de lípido, unión covalente de fosfatidilinositol, reticulación, ciclación, formación de enlace disulfuro, desmetilación, formación de reticulaciones covalentes, formación de cisteína, glucosilación, formación de anclaje de GPI, hidroxilación, yodación, metilación, miristilación, oxidación, pegilación, procesamiento proteolítico, fosforilación, prenilación, racemización, sulfonación, adición mediada por el ARN de transferencia de aminoácidos a proteínas tal como arginilación y ubiquitinación. (Véase, por ejemplo, PROTEINS - STRUCTURE AND MOLECULAR PROPERTIES, 2ª Ed., T. E. Creighton, W. H. Freeman and Company, Nueva York (1993); POST-TRANSLATIONAL COVALENT MODIFICATION OF PROTEINS, B. C. Johnson, Ed., Academic Press, Nueva York, páginas. 1 - 12 (1983); Seifter *et al.* (1990) *Meth. Enzymol.* 182: 626 - 646; Rattan *et al.*, *Ann. N.Y. Acad. Sci.* 663: 48 - 62.

Las moléculas terapéuticas que se pueden usar como restos de ligando incluyen, pero no se limitan a, hormonas, proteínas de matriz, inmunosupresores, broncodilatadores, agentes cardiovasculares, enzimas, agentes del CNS, neurotransmisores, proteínas o péptidos receptores, hormonas de crecimiento, factores de crecimiento, péptidos antivirales, péptidos inhibidores fusogénicos, citocinas, linfocinas, monocinas, interleucinas, factores estimulantes de colonia, factores de diferenciación de colonias, factores angiogénicos, ligandos receptores, proteínas asociadas a cáncer, antineoplásicos, péptidos virales, péptidos antibióticos, proteínas de sangre, proteínas antagonistas, factores de transcripción, factores anti-angiogénicos, proteínas o péptidos antagonistas, antagonistas de receptores, anticuerpos, anticuerpos de una sola cadena y moléculas de adhesión a células. Diferentes moléculas terapéuticas se pueden combinar en una sola proteína de fusión para producir una molécula bi o multi-funcional terapéutica. También se pueden usar en combinación diferentes moléculas para producir una proteína de fusión con una entidad terapéutica y una entidad de dirección. Las moléculas terapéuticas se pueden fusionar directamente al resto del tallo de la presente invención o, de manera alternativa, fusionarse o insertarse en un resto presentador, tal como un resto Tf o resto de albúmina.

Las citocinas son proteínas solubles liberadas por las células del sistema inmune, que actúan de manera no

enzimática a través de los receptores enzimáticos específicos para regular las respuestas inmunes. Las citocinas se parecen a hormonas en que actúan a bajas concentraciones unidas con alta afinidad a un receptor específico. El término “citocina” se usa en el presente documento para describir las proteínas de origen natural o recombinantes proteínas, sus análogos y sus fragmentos que inducen una respuesta biológica específica en una célula que tiene un receptor para esa citocina. Las citocinas preferiblemente incluyen interleucinas tales como interleucina-2 (IL-2) (GenBank N° de acceso S77834), IL-3 (GenBank N° de acceso M14743), IL-4 (GenBank N° de acceso M23442), IL-5 (GenBank N° de acceso J03478), IL-6 (GenBank N° de acceso M14584), IL-7 (GenBank N° de acceso NM_000880), IL-10 (GenBank N° de acceso NM_000572), IL-12 (GenBank N° de acceso AF180562 y GenBank N° de acceso AF180563), IL-13 (GenBank N° de acceso U10307), IL-14 (GenBank N° de acceso XM_170924), IL-15 (GenBank N° de acceso X91233), IL-16 (GenBank N° de acceso NM_004513), IL-17 (GenBank N° de acceso NM_002190) y IL-18 (GenBank N° de acceso NM_001562), factores hematopoyéticos tales como factor estimulador de la colonia de granulocitos (GM-CSF) (GenBank N° de acceso X03021), factor estimulador de la colonia de granulocitos (G-CSF) (GenBank N° de acceso X03656), factor activador de plaquetas (GenBank N° de acceso NM_000437) y eritropoyetina (GenBank N° de acceso X02158), factores de necrosis tumoral (TNF) tal como TNF α (GenBank N° de acceso X02910), linfocinas tal como linfotóxina- α (GenBank N° de acceso X02911), linfotóxina - β (GenBank N° de acceso L11016), leucorregulina, factor inhibidor de la migración de macrófagos (GenBank N° de acceso M25639) y neuroleucina (GenBank N° de acceso K03515), reguladores de procesos metabólicos tales como leptina (GenBank N° de acceso U43415), interferones tal como interferón α (IFN α) (GenBank N° de acceso M54886), IFN β (GenBank N° de acceso V00534), IFN γ (GenBank N° de acceso J00219), IFN δ (GenBank N° de acceso NM_002177), trombospondina 1 (THBS1) (GenBank N° de acceso NM_003246), THBS2 (GenBank N° de acceso L12350), THBS3 (GenBank N° de acceso L38969), THBS4 (GenBank N° de acceso NM_003248) y quimiocinas. Preferiblemente, la proteína de fusión transferrina-citocina modificada de la presente invención muestra actividad biológica de citocina.

El término “hormona” se usa en el presente documento para describir cualquiera de un número de sustancias biológicamente activas que se producen por ciertas células o tejidos y que provocan que se produzcan cambios o actividades biológicas específicas en otra célula o tejido localizado en otra parte del cuerpo. Las hormonas preferiblemente incluyen proinsulina (GenBank N° de acceso V00565), insulina (GenBank N° de acceso NM_000207), hormona del crecimiento 1 (GenBank N° de acceso V00520), hormona del crecimiento 2 (GenBank N° de acceso F006060), factor de liberación de la hormona del crecimiento (GenBank N° de acceso NM_0210811), factor I de crecimiento de tipo insulina (GenBank N° de acceso M27544), factor II de crecimiento de tipo insulina (GenBank N° de acceso NM_000612), proteína 1 de unión al factor de crecimiento de tipo insulina (IGFBP-1) (GenBank N° de acceso M59316), IGFBP-2 (GenBank N° de acceso X16302), IGFBP-3 (GenBank N° de acceso NM_000598), IGFBP-4 (GenBank N° de acceso Y12508), IGFBP-5 (GenBank N° de acceso M65062), IGFBP-6 (GenBank N° de acceso NM_002178), IGFBP-7 (GenBank N° de acceso NM_001553), cadena β de la gonadotropina coriónica (GenBank N° de acceso NM_033142), cadena α de la gonadotropina coriónica (GenBank N° de acceso NM_000735), hormona β luteinizante (GenBank N° de acceso X00264), hormona β foliculoestimulante (GenBank N° de acceso NM_000510), hormona β estimulante del tiroides (GenBank N° de acceso NM_000549), prolactina (GenBank N° de acceso NM_000948), pro-opiomelanocortina (GenBank N° de acceso V01510), corticotropina (ACTH), β -lipotropina, hormona estimulante de melanocitos α (α -MSH), γ -lipotropina, β -MSH, β -endorfina y péptido del lóbulo intermedio de tipo corticotropina (CLIP).

El término “hormona” también incluye péptido-1 de tipo glucagón (GLP-1) que es una hormona gastrointestinal que regula la secreción de insulina que pertenece al llamado eje enteroinsular así como exendina (*por ejemplo*, exendina-4 y sus variantes) que es un agonista del receptor de GLP-1.

El término “factor de crecimiento” se usa en el presente documento para describir cualquier proteína o péptido que se une a un receptor para estimular la proliferación de las células. Los factores de crecimiento preferiblemente incluyen factor- α de crecimiento derivado de plaquetas (PDGF- α) (GenBank N° de acceso X03795), PDGF- β (GenBank N° de acceso X02811), hormonas esteroideas, factor de crecimiento epidérmico (EGF) (GenBank N° de acceso NM_001963), factores de crecimiento de fibroblastos tal como factor 1 de crecimiento de fibroblasto (FGF1) (GenBank N° de acceso NM_000800), FGF2 (GenBank N° de acceso NM_002006), FGF3 (GenBank N° de acceso NM_005247), FGF4 (GenBank N° de acceso NM_002007), FGF5 (GenBank N° de acceso M37825), FGF6 (GenBank N° de acceso X57075), FGF7 (GenBank N° de acceso NM_002009), FGF8 (GenBank N° de acceso AH006649), FGF9 (GenBank N° de acceso NM_002010), FGF10 (GenBank N° de acceso AB002097), FGF11 (GenBank N° de acceso NM_004112), FGF12 (GenBank N° de acceso NM_021032), FGF13 (GenBank N° de acceso NM_004114), FGF14 (GenBank N° de acceso NM_004115), FGF16 (GenBank N° de acceso AB009391), FGF17 (GenBank N° de acceso NM_003867), FGF18 (GenBank N° de acceso AF075292), FGF19 (GenBank N° de acceso NM_005117), FGF20 (GenBank N° de acceso NM_019851), FGF21 (GenBank N° de acceso NM_019113), FGF22 (GenBank N° de acceso NM_020637) y FGF23 (GenBank N° de acceso NM_020638), angiogenina (GenBank N° de acceso M11567), factor neurotrófico derivado de cerebro (GenBank N° de acceso M61176), factor de crecimiento neurotrófico ciliar (GenBank N° de acceso X60542), factor- α de crecimiento de transformación (TGF- α) (GenBank N° de acceso X70340), TGF- β (GenBank N° de acceso X02812), factor- α de crecimiento nervioso (NGF- α) (GenBank N° de acceso NM_010915), NGF- β (GenBank N° de acceso X52599), inhibidor de tejido de metaloproteína 1 (TIMP1) (GenBank N° de acceso NM_003254), TIMP2 (GenBank N° de acceso NM_003255), TIMP3 (GenBank N° de acceso U02571), TIMP4 (GenBank N° de acceso U76456) y estimulante 1 de macrófagos (GenBank N° de acceso L11924).

El término “proteína de matriz” se usa en el presente documento para describir proteínas o péptidos que normalmente se encuentran en la matriz extracelular. Estas proteínas pueden ser funcionalmente importantes para la fortaleza, filtración, o adhesión. Las proteínas de matriz preferiblemente incluyen colágenos tales como colágeno I (GenBank N° de acceso Z74615), colágeno II (GenBank N° de acceso X16711), colágeno III (GenBank N° de acceso X14420), colágeno IV (GenBank N° de acceso NM_001845), colágeno V (GenBank N° de acceso NM_000393), colágeno VI (GenBank N° de acceso NM_058175), colágeno VII (GenBank N° de acceso L02870), colágeno VIII (GenBank N° de acceso NM_001850), colágeno IX (GenBank N° de acceso X54412), colágeno X (GenBank N° de acceso X60382), colágeno XI (GenBank N° de acceso J04177) y colágeno XII (GenBank N° de acceso U73778),

proteínas de laminina tales como LAMA2 (GenBank N° de acceso NM_000426), LAMA3 (GenBank N° de acceso L34155), LAMA4 (GenBank N° de acceso NM_002290), LAMB1 (GenBank N° de acceso NM_002291), LAMB3 (GenBank N° de acceso L25541), LAMC1 (GenBank N° de acceso NM_002293), nidogeno (GenBank N° de acceso NM_002508), α -tectorina (GenBank N° de acceso NM_005422), β -tectorina (GenBank N° de acceso NM_058222) y fibronectina (GenBank N° de acceso X02761).

El término “proteínas de la sangre” se definen tradicionalmente como las que provienen del plasma, muchas ahora producidas de manera común por medios recombinantes e incluyen, pero no se limitan a proteínas de suero nativas, derivados, fragmentos y mutantes o variantes de las mismas, factores de coagulación de la sangre, derivados, mutantes, variantes y fragmentos (incluyendo los factores VII, VIII, IX, X), inhibidores de proteasa (antitrombina 3, alfa-1 antitripsina), activador de plasminógeno de tipo urocinasa, inmunoglobulinas, factor von Willebrand y mutantes von Willebrand, fibronectina, fibrinógeno, trombina y hemoglobina.

El término “enzima” se usa en el presente documento para describir cualquier proteína o sustancia proteinácea que cataliza una reacción específica sin alterarse o destruirse de manera permanente ella misma. Las enzimas preferiblemente incluyen factores de coagulación tales como F2 (GenBank N° de acceso XM_170688), F7 (GenBank N° de acceso XM_027508), F8 (GenBank N° de acceso XM_013124), F9 (GenBank N° de acceso NM_000133), F10 (GenBank N° de acceso AF503510) y otros, metaloproteinasas de matriz tal como metaloproteinasa I de matriz (GenBank N° de acceso MMP1) (GenBank N° de acceso NM_002421), MMP2 (GenBank N° de acceso NM_004530), MMP3 (GenBank N° de acceso NM_002422), MMP7 (GenBank N° de acceso NM_002423), MMP8 (GenBank N° de acceso NM_002424), MMP9 (GenBank N° de acceso NM_004994), MMP10 (GenBank N° de acceso NM_002425), MMP12 (GenBank N° de acceso NM_002426), MMP13 (GenBank N° de acceso X75308), MMP20 (GenBank N° de acceso NM_004771), adenosina desaminasa (GenBank N° de acceso NM_000022), proteínas cinasas activadas por mitógeno MAPK3 (GenBank N° de acceso XM_055766), MAP2K2 (GenBank N° de acceso NM_030662), MAP2K1 (GenBank N° de acceso NM_002755), MAP2K4 (GenBank N° de acceso NM_003010), MAP2K7 (AF013588) y MAPK12 (NM_002969), cinasas tales como JNKK1 (GenBank N° de acceso U17743), JNKK2 (GenBank N° de acceso AF014401), JAK1 (M64174), JAK2 (NM_004972) y JAK3 (NM_000215) y fosfatasa tales como PPM1A (GenBank N° de acceso NM_021003) y PPM1D (GenBank N° de acceso NM_003620).

El término “factores de transcripción” se usa en el presente documento para describir la proteína o péptido implicado en la transcripción de genes que codifican proteína. Los factores de transcripción pueden incluir Sp1, Sp2 (GenBank N° de acceso NM_003110), Sp3 (GenBank N° de acceso AY070137), Sp4 (GenBank N° de acceso NM_003112) NFYB (GenBank N° de acceso NM_006166), Hap2 (GenBank N° de acceso M59079), GATA-1 (GenBank N° de acceso NM_002049), GATA-2 (GenBank N° de acceso NM_002050), GATA-3 (GenBank N° de acceso X55122), GATA-4 (GenBank N° de acceso L34357), GATA-5, GATA-6 (GenBank N° de acceso NM_005257), FOG2 (NM_012082), Eryf1 (GenBank N° de acceso X17254), TRPS1 (GenBank N° de acceso NM_014112), NF-E2 (GenBank N° de acceso NM_006163), NF-E3, NF-E4, TFCEP2 (GenBank N° de acceso NM_005653), Oct-1 (GenBank N° de acceso X13403), proteínas homeobox tales como HOXB2 (GenBank N° de acceso NM_002145), HOX2H (GenBank N° de acceso X16665), homólogo sin pelo (GenBank N° de acceso NM_005144), madres contra proteínas decapentaplégicas tales como MADH1 (GenBank N° de acceso NM_005900), MADH2 (GenBank N° de acceso NM_005901), MADH3 (GenBank N° de acceso NM_005902), MADH4 (GenBank N° de acceso NM_005359), MADH5 (GenBank N° de acceso AF009678), MADH6 (GenBank N° de acceso NM_005585), MADH7 (GenBank N° de acceso NM_005904), MADH9 (GenBank N° de acceso NM_005905) y transductor de señal y activador de proteínas de transcripción STAT1 (GenBank N° de acceso XM_010893), STAT2 (GenBank N° de acceso NM_005419), STAT3 (GenBank N° de acceso AJ012463), STAT4 (GenBank N° de acceso NM_003151), STAT5 (GenBank N° de acceso L41142) y STAT6 (GenBank N° de acceso NM_003153).

En todavía otra realización de la divulgación, la molécula terapéutica es una proteína no humana o no de mamífero. Por ejemplo, se pueden usar VIH gp120, VIH Tat, proteínas de superficie de otros virus tales como hepatitis, herpes, influenza, adenovirus y RSV, otros componentes de VIH, proteínas de superficie parásitas tales como antígenos de malaria y proteínas de superficie bacterianas. Estas proteínas no humanas se pueden usar, por ejemplo, como antígenos, o porque tienen actividades útiles. Por ejemplo, la molécula terapéutica puede ser estreptocinasa, estafilocinasa, asparaginasa, urocinasa, u otras proteínas con actividades enzimáticas útiles.

En una realización alternativa de la divulgación, la molécula terapéutica es una proteína de unión a ligando con actividad biológica. Tales proteínas de unión a ligando pueden, por ejemplo, (1) bloquear las interacciones receptor-ligando en la superficie de la célula; o (2) neutralizar la actividad biológica de una molécula en la fase fluida de la sangre, evitando por lo tanto que alcance su diana celular. En algunas divulgaciones, las proteínas de transferrina de fusión modificadas incluyen una molécula de transferrina modificada fusionada a un dominio de unión a un ligando de un receptor seleccionado entre el grupo que consiste en, pero sin limitación a, un receptor de lipoproteína de baja densidad (LDL), un receptor de LDL acetilado, un receptor del factor α de necrosis tumoral, un receptor del factor β de crecimiento de transformación, un receptor de citocina, un receptor de la inmunoglobulina Fc, un receptor de hormona, un receptor de glucosa, un receptor de glucolípido y un receptor de glucosaminoglucano. En otras realizaciones, las proteínas de unión a ligando incluyen CD2 (M14362), CD3G (NM_000073), CD3D (NM_000732), CD3E (NM_000733), CD3Z (J04132), CD28 (NM_006139), CD4 (GenBank N° de acceso NM_000616), CD1A (GenBank N° de acceso M28825), CD1B (GenBank N° de acceso NM_001764), CD1C (GenBank N° de acceso NM_001765), CD1D (GenBank N° de acceso NM_001766), CD80 (GenBank N° de acceso NM_005191), GNB3 (GenBank N° de acceso AF501884), CTLA-4 (GenBank N° de acceso NM_005214), moléculas de adhesión intercelular tales como ICAM-1 (NM_000201), ICAM-2 (NM_000873) e ICAM-3 (NM_002162), receptores del factor de necrosis tumoral tales como TNFRSF1A (GenBank N° de acceso X55313), TNFR1SFB (GenBank N° de acceso NM_001066), TNFRSF9 (GenBank N° de acceso NM_001561), TNFRSF10B (GenBank N° de acceso NM_003842), TNFRSF11B (GenBank N° de acceso NM_002546) y TNFRSF13B (GenBank N° de acceso NM_006573) y receptores de interleucina tales como IL2RA (GenBank N° de acceso NM_000417), IL2RG (GenBank N° de acceso NM_000206), IL4R (GenBank N° de acceso AF421855), IL7R (GenBank N° de acceso NM_002185), IL9R (GenBank N° de acceso XM_015989) e IL13R (GenBank N° de acceso X95302). Preferiblemente, la proteína de fusión de

unión a ligando de la presente invención muestra la actividad biológica de la proteína de unión a ligando.

El término “proteínas asociadas a cáncer” se usa en el presente documento para describir proteínas o polipéptidos cuya expresión está asociada con cáncer o el mantenimiento de crecimiento de células controlado, tales como proteínas codificadas por genes supresores de tumor u oncogenes. Las proteínas asociadas a cáncer pueden incluir p16 (GenBank N° de acceso AH005371), p53 (GenBank N° de acceso NM_000546), p63 (GenBank N° de acceso NM_003722), p73 (GenBank N° de acceso NM_005427), BRCA1 (GenBank N° de acceso U14680), BRCA2 (GenBank N° de acceso NM_000059), proteína de interacción de CTBP (GenBank N° de acceso U72066), DMBT1 (GenBank N° de acceso NM_004406), HRAS (GenBank N° de acceso NM_005343), NCYM (GenBank N° de acceso NM_006316), FGR (GenBank N° de acceso NM_005248), myb (GenBank N° de acceso AF104863), raf1 (GenBank N° de acceso NM_002880), erbB2 (GenBank N° de acceso NM_004448), VAV (GenBank N° de acceso X16316), c-fos (V GenBank N° de acceso 01512), c-fes (GenBank N° de acceso X52192), c-jun (GenBank N° de acceso NM_002228), MAS1 (GenBank N° de acceso M13150), pim-1 (GenBank N° de acceso M16750), TIF1 (GenBank N° de acceso NM_003852), c-fms (GenBank N° de acceso X03663), EGFR (GenBank N° de acceso NM_005228), erbA (GenBank N° de acceso X04707), tirosina cinasa c-src (GenBank N° de acceso XM_044659), c-abl (GenBank N° de acceso M14752), N-ras (GenBank N° de acceso X02751), K-ras (GenBank N° de acceso M54968), jun-B (GenBank N° de acceso M29039), c-myc (GenBank N° de acceso AH001511), RB1 (GenBank N° de acceso M28419), DCC (GenBank N° de acceso X76132), APC (GenBank N° de acceso NM_000038), NF1 (GenBank N° de acceso M89914), NF2 (GenBank N° de acceso Y18000) y bcl-2 (GenBank N° de acceso M13994).

“Péptidos inhibidores fusogénicos” se usa en el presente documento para describir péptidos que muestran actividad antiviral, capacidad de fusión anti-membrana, y/o una capacidad para modular los procesos intracelulares, por ejemplo, los que implican estructuras de péptido enrollado-espiral. La actividad antiviral incluye, pero no se limita a, la inhibición de virus VIH-1, VIH-2, RSV, SIV, EBV, sarampión, virus de la influenza, o transmisión de CMV a células no infectadas. De manera adicional, la capacidad antifusogénica, actividad antiviral o actividad moduladora intracelular de los péptidos meramente requiere la presencia de los péptidos y de manera específica no requiere la estimulación de una respuesta inmune huésped dirigida contra tales péptidos. Antifusogénico se refiere a una capacidad de un péptido para inhibir o reducir el nivel de episodios de fusión de membrana entre dos o más restos con relación al nivel de fusión de membrana que se produce entre dichos restos en la ausencia del péptido. Los restos pueden ser, por ejemplo, membranas celulares o estructuras virales, tales como envolturas o pelos virales. El término “péptido antiviral”, como se usa en el presente documento, se refiere a la capacidad del péptido de inhibir la infección viral de células o alguna actividad viral requerida para la infección viral y/o patogénesis viral, mediante, por ejemplo, fusión célula-célula o sin infección por virus. Tal infección puede implicar la fusión de membrana, como sucede en el caso de virus envueltos, o algún otro episodio de fusión que implica una estructura viral y una estructura celular. Los péptidos inhibidores fusogénicos y péptidos antivirales tienen a menudo secuencias de aminoácidos que se derivan de más de una proteína viral (*por ejemplo*, un VIH-1, VIH-2, RSV y polipéptido derivado de SIV).

Ejemplos de péptidos inhibidores fusogénicos y péptidos antivirales se pueden encontrar en los documentos WO 94/2820, WO 96/19495, WO 96/40191, WO 01/64013 y las Patentes de Estados Unidos 6.333.395, 6.258.782, 6.228.983, 6.133.418, 6.093.794, 6.068.973, 6.060.065, 6.054.265, 6.020.459, 6.017.536, 6.013.263, 5.464.933, 5.346.989, 5.603.933, 5.656.480, 5.759.517, 6.245.737; 6.326.004 y 6.348.568; todos los cuales se incorporan en el presente documento como referencia.

Los ejemplos de otros tipos de péptidos incluyen fragmentos de proteínas terapéuticas como se describen en el presente documento, en particular, los fragmentos de proteínas humanas que retienen al menos una actividad de la molécula precursora. Los péptidos que se pueden usar para producir restos de ligando de la invención también incluyen péptidos miméticos y péptidos que muestran una actividad biológica de una proteína terapéutica pero difieren en secuencia o estructura de tres dimensiones de una proteína terapéutica de longitud completa. Como un ejemplo no limitado, los péptidos incluyen péptidos miméticos de eritropoyetina descritos por Johnson *et al.* (2000) *Nephrol. Dial. Transplant* 15 (9): 1274 - 7, Kuai *et al.* (2000) *J. Pept. Res.* 56 (2): 59 - 62, Barbone *et al.* (1999) *Nephrol. Dial. Transplant.* 14 Sup 2: 80 - 4, Middleton *et al.* (1999) *J. Biol. Chem.* 274 (20): 14163 - 9, Johnson *et al.* (1998) *Biochemistry* 37(11): 3699 - 710, Johnson *et al.* (1997) *Chem. Biol.* 12: 939 - 50, Wrighton *et al.* (1997) *Nat. Biotechnol.* 15 (12): 1261 - 5, Livnah *et al.* (1996) *Science* 273: 464 - 71 y Wrighton *et al.*, (1996) *Science* 273: 458 - 64.

Las moléculas terapéuticas también incluyen proteínas alergénicas y los fragmentos digeridos de las mismas. Éstas incluyen alérgenos de polen de ambrosía, centeno, poa de los prados, pata de gallo, césped de primavera dulce, césped de cabeza roja, césped de fleo, acedera, trigo, maíz, artemisa, césped azul, césped anual de California, amaranto, césped de Bermudas, cardo ruso, cedro de montaña, roble, saúco, sicomoro, arce, olmo, *etc.*, ácaros de polvo, veneno de abejas, alérgenos de alimentos, caspa animal y otros venenos de insectos.

Otras moléculas terapéuticas incluyen vacunas microbianas que incluyen vacunas de virus, vacunas bacterianas y protozoarias y sus diversos componentes tales como antígenos de superficie. Éstos incluyen vacunas que contienen glucoproteínas, proteínas o péptidos derivados de estas proteínas. Tales vacunas se preparan a partir de *Staphylococcus aureus*, *Streptococcus pyogenes*, *Streptococcus pneumoniae*, *Neisseria meningitidis*, *Neisseria gonorrhoeae*, *Salmonella spp.*, *Shigella spp.*, *Escherichia coli*, *Klebsiella spp.*, *Proteus spp.*, *Vibrio cholerae*, *Campylobacter pylori*, *Pseudomonas aeruginosa*, *Haemophilus influenzae*, *Bordetella pertussis*, *Mycobacterium tuberculosis*, *Legionella pneumophila*, *Treponema pallidum*, clamidia, toxoide de tétanos, toxoide de difteria, virus de influenza, adenovirus, paramixovirus (paperas, sarampión), virus de rubeola, virus de polio, virus de hepatitis, virus herpes, virus de rabia, virus de VIH-1, VIH-2, RSV y papilomavirus.

Las moléculas de fusión preferidas pueden contener péptidos anti-virus VIH, péptidos anti-RSV, hormona de crecimiento humana, interferones α y/o β , eritropoyetina (EPO), péptidos de tipo EPO, factor estimulante de la colonia de granulocitos (GCSF), factor estimulante de la colonia de granulocitos-macrófagos (GMCSF), insulina, factor de crecimiento de tipo insulina (IGF), trombopoyetina, péptidos correspondientes a CDR de un anticuerpo,

Proteína Asociada a la Neogénesis Insular (INGAP), calcitonina, angiostatina, endostatina, interleucina-2, factor de liberación de la hormona de crecimiento, hormona de paratiroides humana, péptidos de factor de necrosis antitumoral (TNF), receptor de interleucina-1 (IL-1) y/o anticuerpos de una sola cadena.

Las proteínas de fusión de la invención también se pueden preparar para que incluyan péptidos o polipéptidos derivados de bibliotecas de péptidos para seleccionar las moléculas con funciones nuevas o novedosas. Tales bibliotecas de péptidos pueden incluir los disponibles comercialmente o públicamente, *por ejemplo*, American Peptide Co. Inc., Cell Sciences Inc., Invitrogen Corporation, Phoenix Pharmaceuticals Inc., United States Biological, así como los producidos mediante tecnologías disponibles, *por ejemplo*, bibliotecas de despliegue sobre bacteriófagos y bacterias preparadas usando procedimientos convencionales.

En todavía otras realizaciones de la divulgación, las proteínas de fusión se pueden preparar mediante el uso de restos de proteínas terapéuticas conocidos en la técnica y se ejemplifican por los péptidos y proteínas actualmente aprobados por la Administración de Alimentos y Fármacos (www.fda.gov/cber/efoi/approve.htm) así como por las Publicaciones de Patentes PCT números WO 01/79258, WO 01/77137, WO 01/79442, WO 01/79443, WO 01/79444 y WO 01/79480, todas las cuales se incorporan en el presente documento como referencia en su totalidad.

La tabla 1 de la Publicación Internacional PCT N° WO 03/020746, proporciona una lista no exhaustiva de proteínas terapéuticas que corresponde a una parte de proteína terapéutica, es decir resto de ligando, de una proteína de fusión de la invención. La columna "proteína terapéutica X" divulga moléculas proteicas terapéuticas segundas por paréntesis que contienen nombres científicos y de marca comercial que comprenden o de manera alternativa consisten en esa molécula proteica terapéutica o un fragmento o variante de la misma. "Proteína terapéutica X" como se usa en el presente documento se puede referir bien a una molécula proteica terapéutica individual (según se define mediante la secuencia de aminoácidos que se pueden obtener a partir de CAS y números de acceso del Genbank), o bien al grupo completo de proteínas terapéuticas asociadas a una molécula proteica terapéutica dada divulgada en esta columna. La columna "identificador ejemplar" proporciona identificadores de los Números de Registro de Servicios de Resúmenes de Química (CAS) (publicados por la Sociedad Americana Química) y/o Números de acceso del Genbank (*por ejemplo*, Locus ID, NP - XXXXX (Proteína de Secuencia de referencia) y XP-XXXXX (Proteína Modelo) disponibles del Centro nacional para la Información Biotecnológica (NCBI) página web (www.ncbi.nlm.nih.gov) que corresponde a las entradas en el Registro CAS o base de datos del Genbank que contienen una secuencia de aminoácidos de la molécula de proteína o de un fragmento o variante de la molécula de proteína terapéutica. Además los números de acceso de GenSEC y/o citas en las publicaciones de revistas se proporcionan para identificar la secuencia de aminoácidos ejemplares para algunos polipéptidos.

Están disponibles las páginas del sumario asociadas a cada una de estas CAS y números de acceso de Genbank y GenSEC así como las publicaciones citadas de las revistas (*por ejemplo*, número ID de PubMed (PMID)). La columna PCT/Referencia de Patente proporciona los números de las patentes de Estados Unidos, o números de Publicación Internacional PCT correspondientes a patentes y/o solicitudes de patente publicadas que describen la molécula de proteína terapéutica. La columna actividad biológica describe las actividades biológicas asociadas a la molécula de proteína terapéutica. La columna ensayo de actividad ejemplar proporciona referencias que describen ensayos que se pueden usar para ensayar la actividad terapéutica y/o biológica de una proteína terapéutica o una proteína de fusión de transferrina de la invención que comprende una parte de una proteína terapéutica X. La columna "indicación preferida Y" describe enfermedad, trastornos, y/o afecciones que se pueden tratar prevenir, diagnosticar, o mejorar mediante proteína terapéutica X o una proteína de fusión de transferrina de la invención que comprende una parte de proteína terapéutica X. La presente invención incluye las proteínas terapéuticas proporcionadas en el documento WO 03/020746.

Ejemplos

Ejemplo 1 - Preparación de anclaje de GPI, hMUC1 y casete de expresión mTF

El vector pREX0549 que contiene una casete de expresión mTf (SEC ID NO: 16) se digirió con *SaI* y *HindIII*. La figura 3 proporciona un mapa del vector para pREX0549. Los cebadores P0922 y P0923 (SEC ID NO: 7 y SEC ID NO: 8) se hibridaron conjuntamente y se ligaron en pREX0549 en el sitio de digestión *SaI/HindIII*. El engarce formado por P0922 y P0923 contenía los sitios de restricción *SpeI*, *HindIII* y *XbaI* y se diseñó para aceptar una molécula de ácido nucleico que codifica un anclaje de GPI y tallo MUC1. El vector resultante, pREX0628, contenía la casete de expresión mTf con el engarce P0922/P0923.

pREX0628 se digirió con *HindIII* y *XbaI*. Los cebadores P0924 y P0925 (SEC ID NO: 9 y SEC ID NO: 10) se hibridaron para formar el anclaje YIR019c de GPI. YIR019c se ligó en el pREX0628 para crear el vector pREX0634.

hMUC1 ADNc se amplificó mediante RT-PCR a partir de una biblioteca de ARN total de tumor de mama humano (Clontech) usando los cebadores P0958 y P0959 (SEC ID NO: 11 y SEC ID NO: 12). El ADNc resultante se amplificó con los cebadores P1019 y P1020 (SEC ID NO: 13 y SEC ID NO: 14) para crear los sitios de restricción *SpeI* y *HindIII*. El hMUC1 resultante con *SpeI* e *HindIII* se proporciona en la SEC ID NO: 6.

pREX0634 se digirió con *SpeI* e *HindIII* y el hMUC1 con los sitios *SpeI* e *HindIII* se ligó en el vector. El vector resultante, pREX0663, se usó como casete de expresión de visualización (mTf-MUC1-GPI).

pREX0663 se usó para crear vectores de expresión de levaduras de alto y bajo número de copias. Para crear un vector de expresión de levaduras de alto número de copias, la casete de expresión de despliegue de 4,1 kb se retiró de pREX0663 mediante digestión del vector con *NotI*. La casete de expresión se ligó después en un vector pSAC35c digerido y desfosforilado con *NotI*, dando como resultado el vector pREX0667 (vector I de despliegue en levaduras I).

Un vector de expresión de levaduras de bajo número de copias se creó mediante digestión de pREX0663 con *NotI* y

ligando el vector de expresión en un pREX0699 digerido y desfosforilado con *NotI*, dando como resultado despliegue de levaduras pREX0721 (vector de despliegue en levaduras II).

Los vectores de expresión de levadura descritos anteriormente se pueden usar para transformar células de levaduras y células bacterianas como se conoce en la técnica. El vector se puede expresar en levadura como se conoce en la técnica. Además, se puede crear una colección de proteínas de transferrina de fusión expresadas capaces de desplegar una biblioteca de restos de ligando tales como péptido aleatorios o CDR y usar para seleccionar agentes de unión como se conoce en la técnica.

Ejemplo 2 - Construcción de biblioteca aleatoria de 15-meros

Para la selección de variantes de transferrina con características de unión novedosas, se construyó una biblioteca aleatoria de 15-meros en la posición de aminoácido 289 - 290 de transferrina mediante un procedimiento de soldadura de PCR conocido en la técnica (véase Martín y Smith (2006) *Biochem J.* 396 (2): 287 - 95). Se diseñó una biblioteca de 15-meros incluso aunque solamente aproximadamente 7 aminoácidos se necesitaran de manera usual para formar un epítipo de unión y una biblioteca de $\sim 10^9$ solamente cubre una pequeña fracción de la biblioteca diseñada ($3,3 \times 10^9$). Sin embargo, con un tamaño de biblioteca de 10^9 , una biblioteca de 15-meros cubre 6,4 veces más 7-meros que una biblioteca de 7-meros del mismo tamaño.

Después de obtener un fragmento de ADN que contiene la secuencia *BamHI/BspEI* de transferrina usando P1174/P1227, se realizaron dos reacciones de PCR (cada una con un solo cebador – P1172 y P1173) para obtener los ADN de una sola cadena. Los DNAs se aislaron y se hibridaron para formar una biblioteca de 15-meros de soldadura. Esta operación aseguró que la biblioteca mantenía la complejidad original del oligonucleótido sintético. La biblioteca de 15-meros de soldadura de doble cadena se amplificó adicionalmente usando P1174/P1227 para obtener una cantidad suficiente de ADN. El producto de PCR se purificó, se digirió con *BamHI/BspEI* y se clonó en vectores de plásmido apropiados, *por ejemplo*, pREX0995 (figura 4) o pREX0667.

P1172 (SEC ID NO: 19)

Fragmento de directo inverso de soldadura de inserción de péptido aleatorio lib de 15-meros 289 – 290: C CAA CTA TTC AGC TCT CCT 567 567 567 567 567 567 567 567 567 567 567 567 567 567 567 CAT GGG AAG GAC CTG CTG TTT AAG

Con el fin de introducir aleatoriedad en cada posición en la secuencia de ADN, se incorporó una mezcla de nucleótidos (A, G, T y C) en la posición a una relación predeterminada de acuerdo con LaBean y Kauffman (1993) *Protein Sci.* 2: 1249 - 54. La mezcla indicada más adelante minimiza la frecuencia de codón de parada y composición de aminoácido de coincidencia con las proteínas naturales.

5 13% de T, 32% de G, 20% de C, 35% de A

6 24% de T, 24% de G, 22% de C, 30% de A

7 37% de T, 26% de G, 37% de C

P1173 (SEC ID NO: 20)

35 Cebador inverso de soldadura de péptido aleatorio lib de 15-meros 289 - 290 para el fragmento frontal

AGGAGAGCTGAATAGTTGG

P1174 (SEC ID NO: 21)

Cebador directo de soldadura de péptido aleatorio lib de 15-meros 289 - 290 para el fragmento frontal

CTGGATGCAGGTTTGGTGTATG

40 P1227 (SEC ID NO: 22)

Cebador inverso de soldadura de péptido aleatorio lib de 15-meros 289 - 290 para el fragmento trasero

TCATGATCTTGGCGATGCAGTC

Ejemplo 3 - Selección de células de levadura que despliegan Flag

Se estableció un sistema de despliegue en levadura mediante el que el lóbulo N de transferrina se desplegó sobre la superficie de levadura mediante fusión a una región de tallo, huMUC1 y una secuencia de señal de GPI. Para demostrar la utilidad de este sistema en la selección del ligante, una secuencia etiquetada con Flag, DYKDDDDK (SEC ID NO: 23), o una biblioteca de péptido aleatorio de 15 meros se insertó en la posición de aminoácido 289 del lóbulo N de la transferrina. La levadura que despliega la transferrina etiquetada con Flag, pREX1012 (figura 6), después se fijó a un conjunto de levadura que despliega el lóbulo N de transferrina con péptidos aleatorios de 15-meros. A partir de esta población mixta solamente la levadura que despliega el lóbulo N de transferrina etiquetada con Flag se recuperó mediante selección con un anticuerpo anti-Flag.

Para insertar la secuencia con marcador Flag en la posición de aminoácido 289 de transferrina, se sintetizaron los oligos que incorporan la secuencia con marcador Flag y se soldaron mediante PCR en el vector pREX0667 para generar pREX0759 (figura 7). El fragmento *BamHI/BspEI* de pREX0759 que contenía la secuencia de marcador de Flag se usó después para reemplazar el mismo fragmento de restricción de pREX0995 (figura 4). El plásmido

resultante, pREX1012, expresa la proteína de fusión lóbulo N de transferrina etiquetado con Flag -MUC1-GPI.

También se clonó la biblioteca de 15 meros entre los sitios *Bam*HI/*Bsp*EI de pREX0995. La preparación de la biblioteca de 15 meros se describe más adelante. La muestra de ligadura de transformó en *E. coli* DH5 α y la mezcla de transformación se sembró toda sobre placas de agar 2 LB/Amp (50 μ g/ml). Se recogieron todas las colonias y se extrajo el ADN de plásmido usando un protocolo Qiagen plasmid prep que usa varias columnas de miniprep.

El ADN de plásmido tanto para pREX1012 como para la biblioteca de 15-meros se transformó en la cepa DS1101 cir^o de *Saccharomyces cerevisiae*. Se inoculó una sola colonia de pREX1012 en medio mínimo tamponado con sacarosa (BMM/S) y se cultivó durante toda la noche. Todas las colonias de la biblioteca de 15-meros se recogieron y se inocularon BMM/S. Los recuentos de células de los dos cultivos durante toda la noche se determinaron mediante hemocitómetro y las siguientes mezclas de células preparadas:

(A) 10^3 células de levadura pREX1012 mezcladas con 10^9 de células de levadura de la biblioteca de 15-meros ($10:10^7$)

(B) 10^3 pREX1012 células de levadura mezcladas con 10^8 de células de levadura de la biblioteca de 15-meros ($100:10^7$)

Las mezclas de células se incubaron en 1 ml de solución de bloqueo de células (1xPBS/0,05% de Tween-20, 1% de BSA) sobre hielo durante 30 minutos. Después de centrifugación (30 segundos a 13000 rpm), se suspendieron los sedimentos celulares en 1 ml de solución de lavado (1xPBS, 0,5% BSA, 2 mM EDTA) con anticuerpo anti-Flag (Sigma Aldrich, dilución 1:25) y se incuban en hielo durante 30 minutos. Las células se lavaron dos veces con 1 ml de solución de lavado y se suspendieron en 800 μ l (A) o 160 μ l (B) de solución de bloqueo. A las suspensiones de células se añadieron 200 μ l (A) o 40 μ l (B) de microperlas de estreptavidina MACS (Miltenyi Biotec) y se incubaron en hielo durante 30 minutos. Las células marcadas se separaron de las células no marcadas usando una columna MS de acuerdo con las instrucciones del fabricante (Miltenyi Biotec). Se recogieron las células marcadas y se sembraron sobre placas de agarosa BMM/S y se incubaron a 30°C hasta que aparecieron pequeñas colonias.

Se realizó una segunda ronda de selección recogiendo todas las colonias de cada placa y haciéndolas crecer durante toda la noche 30°C en 5 ml de BMM/S. a partir de estos cultivos las células equivalentes a 1,5 de DO₆₀₀ se sometieron a una ronda adicional de separación de MACS como se ha descrito anteriormente. Las células de esta segunda ronda de selección se cultivaron durante toda una noche a 30°C en 5 ml de BMM/S. Los cultivos de células de levadura antes y después de cada selección se analizaron mediante FACS usando un anticuerpo monoclonal anti-Flag (Sigma Aldrich) y anticuerpo de detección anti-ratón de cabra marcado con APC en un Bioanalyzer de Tecnología Agilent. Se realizó el análisis FACS de acuerdo con las instrucciones del fabricante. La presencia de levadura etiquetada con Flag se hizo evidente después de dos rondas de separación de MACS (Figura 8)

Ejemplo 4 - Despliegue en tallo Aga1

Se obtuvo la secuencia de ADN para la región central del gen de levadura AGA1 (restos 116 - 658 de la proteína madura) mediante PCR de ADN genómico de levadura S288c usando los siguientes cebadores:

(a) CAGATCTAGAACAACCGCTATCAGCTCATTATCC (SEC ID NO: 24)

(b) CAGAAAGCTTAGTAGTGGAACCTTCTGTAGTG (SEC ID NO: 25)

AGA1 (secuencia parcial) (SEC ID NO: 74)

ACAACCGCTATCAGCTCATTATCCGAAGTAGGAACACAACCGTGGTATCATCCAGCGCCATTGAACCATCAAGT
GCCTCTATAATCTCACCTGTACCTCTACACTTTTCGAGTACAACATCGTCCAATCCAACCTACTACCTCCCTAAGTT
CGACATCTACATCTCCAAGCTCTACATCTACATCTCCAAGCTCTACATCTACCTCATCAAGTTTCGACATCTACCTC
ATCAAGTTTCGACATCTACCTCATCAAGTTTCGACATCTACATCTCCAAGTTTCGACATCTCAAGTTTGACA
TCCACATCTTCAAGTTCTACATCTACATCTCCAAAGTTCTACATCTACCTCATCAAGTTTCGACATCTACATCTCCA
GCTCTACATCTACCTCATCAAGTTCAACATCTACATCTCCAAGTTCTAAATCTACTTCTGCAAGCTCCACTTCCAC
TTCTTCATATTTCAACATCTACATCTCCCAAGTTTGACTTTCTTCACTCTCCAACTTTGGCTTCCACTTCTCCAAAGTTCAA
TTATATTAGCTCTACTTTTACTGATTCAACTTCTACCTCTTGGCTCCTCTATAGCATCTTTCATCACTGTGTGCA
TTATACAGCCCATCCACACCTGTTTACTCCGTCCTTCGACTTCGTCAAATGTTGCAACTCCTTCTATGACTTCTT
CAACTGTTGAAACAACCTGTTAGTTCACAAAGTTTCGTCTGAATATATCACCAAACTCCTCAATTTCTACTACTATCCCA
TCATTTTCCATGTCTACATATTTCAACCACTGTTAGTGGAAGTCACTACAATGTATACGACATGGTGTCTTATAGCT
CTGAATCTGAGACTAGCAGCATTAACCAAGTATGCATGGAACGTTACAACAGACGCTACAGTCTGCACTCACGAGT
CTTGCAATGCCCCGCGCAGACAACAAGTTTGATTACATCTTCTATAAAAATGTCCACTAAAAACGTCGCAACTTCTGT
AAGCACTCAACGTTGAAATCCTCATATGCATGCTCCACATGTGCTGAAACGTCACACTCGTATTCTTCCGTGCA
AACAGCTTCATCAAGTTCTGTAACACAGCAGACCATCCACAAAGAGTTGGGTAAGTTCAATGACAACCTTCGGA
TGAAGATTTCAATAAGCAGCTACCGGTAAGTATCATGTAACATCTTCAGGTACCTCAACATTTGACTGAGTGA
AGTGAAGCCACGAGTACATCAAGCATTGACTCAGAATCTCAAGAACAATCATCACACTTATTATCGACATCGGTC
CTTTCATCCTCCTCTGTTGCTGCTACATTATCCTCTGACAGTACTATTTGCTATTGAGTTCTGTATCATCACTAAG
TGTCGAACAGTCACCAAGTTACCACTTCAAATTTCTTCAACATCAGAGATTTTACAACCCACTTCTTCCACAGCT
ATTGCTAAATATCTGCCTCTACATCATCAATTTCCGCAACATCTATCTACACCATCTACCTCTGTGGAATCGA
CTATTGAATCTTCATCATTGACTCCGACGGTATCTCTATTTTCTCTCATCATCATCTGCTCCCTCTTCTCTACAA
ACATCTGTTACCACTACAGAAGTTTCCACTACT

(SEC ID NO: 75)

TTAISSLSEVGTTTVVSSSAIEPSSASIIISPVTSTLSSTTSSNPSTTSLSTSTSPSSTSTSPSSTSTSSSTSTSSSTST
SSSTSTSPSSTSTSSSTSTSSSTSTSSSTSTSSSTSTSSSTSTSSSTSTSSSTSTSSSTSTSSSTSTSSSTSTSSSTST
SLTSSSTLASTSPSSTSISSFTDSTSSLGSSIASSTSVSLYSPSTPVYVSPSTSSNVATPSMTSSTVETTSSQSSS

EYITKSSISTTIPFSMSTYFTTVSGVTTMYTTWCOPYSSSESTSTLTSMHETVTTDATVCTHESCMPSQTTSLITSSIKM
STKNVATSVSTSTVESSYACSTCAETSHSYSSVQTASSSSVTQQTSTKSWVSSMTTSDDFNKHATGKYHVTSSGT
STISTSVSEATSTSSIDSESQEQQSHLLSTSVLSSSSLSATLSSDSTILLFSSVSSLSVEQSPVTTLQISSTSEILQPTSST
AIATISASTSSLSATSISTPSTSVESTIESSSLTPTVSSIFLSSSSAPSSLQTSVTTTEVSTT

- 5 Se aisló un producto de PCR de ~1.5 kb y se digirió con *XbaI/HindIII*. Se ligó el fragmento en pREX0855 digerido con *SpeI/HindIII* (figura 9) y se transformó en *E. coli* DH5 α . Se recogieron todas las colonias resultantes y se aisló el ADN de plásmido de las células. La casete de expresión se recuperó mediante digestión con *NotI* y se ligó en pSAC35 proporcionando el vector de expresión de levadura.

- 10 Transformación en levadura y FACS con anticuerpo anti-Flag como se ha descrito previamente. Las colonias de levadura mostraron alto nivel de despliegue del lóbulo N, aproximadamente 10 veces mayor que la construcción basada en tallo MUC1 comparable (datos no mostrados).

- 15 Una sola colonia de levadura se aisló y se extrajo el ADN de plásmido de estas células de levadura. La casete de expresión *NotI* se recuperó después de la digestión con *NotI* del ADN del plásmido extraído y se ligó en un pREX0855 digerido con *NotI* proporcionando el plásmido pREX1087 (figura 10), un vector basado en pUC que contenía casete de expresión Flag-N-lóbulo-Aga1-GPI. Una región de la secuencia de ADN que corresponde a Aga1 se secuenció para confirmar su identidad. Esta casete de expresión también se volvió a transferir a pSAC35 proporcionando pREX1106 (figura 11).

Ejemplo 5 - Selección de variantes de GPI de Mamífero que funcionan en Células de levadura

- 20 Las señales de GPI de mamífero juegan papeles que sus homólogos de levadura no juegan, tal como tráfico intracelular, transmisión de señales de transmembrana y endocitosis independiente de clatrina (Biochem J. 1993, 294: 305 - 324). Las células de levadura no tienen solamente membrana celular, sino paredes celulares que no están en las células de mamíferos y muchas de las GPI de levadura tienen secuencias únicas que dirigen las proteínas a la pared celular de levadura (J Bacteriol, 1999, 181: 3886 - 3889). la GPI de la fosfatasa alcalina de placenta humana se ha mostrado que no funciona del todo en células de levadura (Mol Microbiol 1999, 34: 247 - 256). Como un medio para obtener secuencias novedosas que pueden unir proteínas recombinantes expresadas en una pared de la célula de levadura, un vector de despliegue de levadura se basó en pREX0885 (figura 9) pero usando la secuencia de GPI de huMDP (DQLGGSCRTHYGY S GASSLHRHWGLLLASLAPLVLC SLL). Esta secuencia se modificó para incorporar cuatro codones completamente aleatorios (X) así como varias modificaciones racionales (subrayado), XQXGG SXXTIGGY G AASSLQRTIGLLLASLAPLVLASLL (SEC ID NO: 26), en la que X es cualquier aminoácido.

- 30 Una biblioteca de levadura que expresa la siguiente proteína de fusión, Flag tag-N-lóbulo-MUC1 tallo-GPI en la que la secuencia de GPI se modificó como se ha descrito anteriormente se transformó en la cepa DS1101 cir^o de *Saccharomyces cerevisiae*. Cualesquiera células de levadura con la proteína de fusión unida a la pared celular se aisló mediante MACS usando un anticuerpo anti-Flag biotinilado.

- 35 Se hibridaron dos oligos P2035 y P2036 (véase más adelante) y se extendieron usando la Taq polimerasa. Se purificó el fragmento de ADN resultante, se digirió con *HindIII/XbaI* y se ligó a pREX0855 digerido con *HindIII/XbaI* (véase más adelante y la figura 9).

Cebadores

P2035 (SEC ID NO: 27)

- 40 CTACAAGCTTNNKCAANNKGGTGGTTCTNNKNNKACTATTGGTGGTTATTCTGGTGCTGCTTCTTCCTTGCAGAG
AACTATTG

P2036 (SEC ID NO: 28)

GATGTCTAGATTATTATAACAAAGAAGCTAAAACCAATGGAGCTAAAGAAGCCAATAACAAACCAATAGTTCTCTG
CAAGGAAG

HindIII

- 45 -+----

aagcttnnkc aannkgggtg ttctnnknnk actattggtg gttattctgg tgctgcttct

ttcgaannmg ttnnmccacc aagannmnnm tgataaccac caataagacc acgacgaaga

k l x q x g g s x x t i g g y s g a a s

>>.....P2035.....>>

- 50 P2036 <<

tccttgacga gaactattgg ttgttattg gcttcttag ctccattggt ttagcttct

aggaacgtct ctgataacc aaacaataac cgaagaaatc gaggaacca aaatcgaaga

s l g r t i g l l l a s l a p l v l a s

>.....P2035.....>>

<.....P2036.....<

XbaI

-+----

ttgtataat aatctaga (SEC ID NO: 29)

5 aacaatatta ttagatct

I I - - s r (SEC ID NO: 30)

<.....P2036.....<<

10 Se transformó la mezcla de ligadura en *E. coli* DH5 α para obtener aproximadamente 5 x 10⁵ colonias. Se recogieron todas las colonias y se aisló el ADN de plásmido. El ADN de plásmido se digirió con *NotI* para recuperar la casete de expresión y se clonó en SAC35 para crear la biblioteca de expresión de levadura. Esta biblioteca se transformó en células DS1101 cir^o mediante electroporación. Un cultivo durante toda la noche de la biblioteca anteriormente mencionada se sometió a MACS usando un anticuerpo anti-Flag biotinilado. Las células aisladas se purificaron inmediatamente de nuevo mediante MACS con el mismo anticuerpo. Las células resultantes se sembraron sobre placas BMMS y se caracterizaron 24 colonias mediante FACS y análisis de secuenciación de ADN. (Véase descripción de fijación de Flag.)

15 De los 22 clones que proporcionaron una secuencia legible solamente 7 tenían anclajes de GPI de longitud completa (Tabla 1) con niveles variables de despliegue y los mejores de los cuales eran mejores que el vector pREX1003 que expresa la misma proteína de fusión con un anclaje de GPI de levadura.

Tabla 1

Clon Nº.	Secuencia	Aminoácidos	Nivel de despliegue
	NNKCAANNKGGTGGTTCTNNKNNK (SEC ID NO: 31)		
23	TGTCAATAGGGTGGTTCTAGGCCT (SEC ID NO: 32)	CysGlnStop	200
8	TGTCAAATTGGTGGTTCTTAGTGT (SEC ID NO: 33)	CysGlnIleGlyGlySerStop -(SEC ID NO: 34)	150
15	CAGCAATATGGTGGTTCTGTGGAT (SEC ID NO: 35)	GluGlnTyrGlyGlySerValAsp (SEC ID NO: 36)	120
14	TCTCAAGTTGGTGGTTCTACTTGG (SEC ID NO: 37)	SerGlnValGlyGlySerThrTrp (SEC ID NO: 38)	100
5	NNKCAANNKGGTGGTTCTNNKNNK (SEC ID NO.: 39)	Desplazamiento de fase	80
2	CATCAAGGTGGTGGTTCTATTCGG (SEC ID NO: 40)	HisGlnGlyGlyGlySerIleArg (SEC ID NO: 41)	60
6	NNKCAANNKGGTGGTTCTNNKNNK (SEC ID NO: 42)	Desplazamiento de fase	50
12	CATCAATTGGGTGGTTCTGTTACG (SEC ID NO: 43)	HisGlnLeuGlyGlySerValThr (SEC ID NO: 44)	50
18	TATCAATCGGGTGGTTCTGGGACT (SEC ID NO: 45)	TyrGlnSerGlyGlySerGlyThr (SEC ID NO: 46)	50
13	GGGCAATATGGTGGTTCTTAGTGG (SEC ID NO: 47)	GlyGlnTyrGlyGlySerStop -(SEC ID NO: 48)	40
1	GTGCAAGCGGGTGGTTCTGATTAG (SEC ID NO: 49)	ValGlnAlaGlyGlySerAsp (SEC ID NO: 50)	30
4	TAGCAAATGGGTGGTTCTACTAAG (SEC ID NO: 51)	Stop	30
21	TAGCAAACGGGTGGTTCTTCTTAT (SEC ID NO: 52)	Stop	20
3	AAGCAACGGGGTGGTTCTTAGACT (SEC ID NO: 53)	LysGlnProGlyGlySerStop (SEC ID NO: 54)	20

(continuación)

Clon Nº.	Secuencia	Aminoácidos	Nivel de despliegue
7	CTGCAATGTGGTGGTTCTTAGTGG (SEC ID NO: 55)	LeuGlnLysGlyGlySerStop (SEC ID NO: 56)	15
16	TAGCAACTGGGTGGTTCTTTTGGG (SEC ID NO: 57)	Stop	15
17	TAGCAATATGGTGGTTCTGTTCTA (SEC ID NO: 58)	Stop	15
19	CTTCAAGTGGGTGGTTCTTTGTAG (SEC ID NO: 59)	LeuGlnValGlyGlySerLeuStop (SEC ID NO: 60)	15
24	TAGCAATTTGGTGGTTCTCATGCG (SEC ID NO: 61)	Stop	15
9	CGGCAACGGGGTGGTTCTAAGTGG (SEC ID NO: 62)	ArgGlnArgGlyGlySerLysTrp (SEC ID NO: 63)	Bajo
20	TCGCAAACCTGGTGGTTCTGTTGCT (SEC ID NO: 64)	SerGlnThrGlyGlySerValAla (SEC ID NO: 65)	Bajo

Inesperadamente, el 50% de los clones se truncaron mediante un codón de parada en uno de los codones distribuidos al azar suprimiendo de manera eficaz la señal de anclaje de GPI. De éstos 12 clones, se determinó que dos tenían niveles de despliegue significativamente mejores que pREX1003 y se encontró que contenían un resto de cisteína justo antes del codón de parada (CQIGGS* (SEC ID NO: 34) y CQ* donde * = codón de parada) (figura 12). Con toda probabilidad estas construcciones se reticularon en la pared celular mediante enlace disulfuro a un resto de cisteína libre en una proteína de la pared celular.

Ejemplo 6 – Construcción y Selección de una biblioteca de RGD

Para la selección de variantes de transferrina que se unen a integrina $\alpha_{IIb}\beta_3$ e inhiben la agregación de plaquetas, se construyó una biblioteca insertando péptidos que comprenden tres aminoácidos distribuidos al azar sobre cualquier lado de la secuencia de unión a integrina Arg-Gly-Asp (RGD) en la estructura de Tf (CXXXRGDXXXC; X que representa una posición de aleatorización) en la posición de aminoácido 289 - 290 de Transferrina mediante un procedimiento de soldadura de PCR.

Se usaron los cebadores P1980/P2181 y P2127/P1173 para amplificar pREX1106 (figura 11) para obtener el fragmento A y B (figura 13).

P1173

Cebador inverso de soldadura 289 - 290 para el fragmento frontal

AGGAGAGCTGAATAGTTGG

P1980

289 - 290 RGD que contiene cebador directo de soldadura de biblioteca de péptido aleatorio para el fragmento trasero

CCAACCTATTCAGCTCTCCTTGT567567567AGAGGAGAC567567567TGTCATGGGAAGGACCTGCTGTTTAAG

13% de T, 32% de G, 20% de C, 35% de A

24% de T, 24% de G, 22% de C, 30% de A

37% de T, 26% de G, 37% de C

(Mezcla de nucleótidos que minimiza la frecuencia de codón de parada y composición de aminoácidos de coincidencia a la proteína natural; Labean, TH y Kauffman, SA, 1993, Protein Science. 2: 1249 - 1254)

P2127

Cebador directo de soldadura de 289-290 para el fragmento frontal

CCTCCTACCTTGATTGCATCAG

P2181

Cebador inverso de soldadura de 289-290 para el fragmento frontal

GGAGATGAAGAAGTCAAACCTTGGG

P2139

GPI lib de reparación de Gap

GAGGCACTTGATGGTTCAATGG

5 El cebador P1980 introdujo la secuencia distribuida al azar. Los fragmentos de ADN amplificados se purificaron en gel y adicionalmente se amplificaron usando un solo cebador (P2181 para el fragmento A y P2127 para el fragmento B) para obtener un ADN de una sola cadena (ss-A y ss-B). ss-A y ss-B se hibridaron en la presencia de dNTP y enzima Klenow formando un fragmento de doble cadena. Este producto de hibridación se amplificó finalmente usando P2127 y P2139 y se purificó en gel para generar inserciones de bibliotecas de RGD. Esta operación asegura que la biblioteca mantenía la complejidad original del oligonucleótido sintético.

10 Se usaron las inserciones para construir una biblioteca de 7×10^8 mediante reparación de huecos en pREX1106 digerido con *Bam*HI-*Bsp*EI usando el procedimiento de electroporación proporcionado más adelante. Para cada reacción de reparación de huecos, 1,4 μ g de pREX1106 digerido con *Bam*HI/*Bsp*EI se mezclaron con 1 μ g de la biblioteca de RGD mencionada anteriormente y se sometieron a electroporación en célula de levaduras DS1101 cir°.

Electroporación de alta eficiencia de levadura (DS1101 cir°) para la construcción de biblioteca.

15 1) Inocular 100 ml de YEP/S (1% p/v de extracto de levadura, 2% p/v de peptona, 2% p/v sacarosa) hasta $DO_{600} = 0,2$ de un cultivo reciente durante toda la noche de DS1101 cir°.

2) Hacer crecer las células a 30°C, 200 rpm hasta una DO_{600} de 0,8 (aproximadamente 5 horas).

3) Recoger las células mediante centrifugación a 2.000 rpm durante 5 min.

20 4) Decantar el sobrenadante y volver a suspender el sedimento en 18 ml de tampón TE (Tris-HCl 10 mM, pH 7,5 y EDTA 1mM). Después añadir 2 ml de acetato de Litio 1 M.

5) Incubar las células a 30°C en un cilindro rodante durante 45 minutos.

6) Añadir 0,5 ml de DTT 1M.

7) Incubar durante 15 minutos a 30°C en un cilindro rodante.

8) Lavar con 80 ml de agua estéril a temperatura ambiente.

25 9) Lavar el sedimento en 100 ml de agua estéril a temperatura ambiente.

10) Lavar el sedimento en 10 ml de sorbitol 1M enfriado con hielo.

11) Volver a suspender las células en 60 μ l de sorbitol 1M enfriado con hielo (suficiente para 6 transformaciones).

12) Mantener las células en hielo y mezclar en tubos de microcentrifuga estériles:

30 i. 50 μ l de células de levaduras competentes.

ii. 1,4 μ g de ADN (hasta 5 μ l)

iii. 5 μ g de ADNss (0,5 μ l 10 mg/ml)

13) Incubar en hielo durante 5 minutos antes de pulso. Golpear ligeramente la suspensión de células/ADN en el fondo de una cubeta enfriada con hielo de 0,2 cm y pulsar una vez:

35 i. Tensión 1.5Kv

ii. Capacitancia 25 μ F

iii. Resistencia 200 Ω

40 14) Añadir inmediatamente 1 ml de una mezcla 50:50 de YEP/S y de sorbitol 1M, transferir la levadura a unos tubos de microcentrifuga y permitir que la levadura se recupere a 30°C durante una hora (puede no ser necesario recuperar, ensayar si es necesario).

15) Recoger las células mediante centrifugación a 5000 rpm durante 1 minuto y lavar una vez más con 1 ml de sorbitol 1M.

16) Volver a suspender 1 ml de sorbitol 1 M y sembrar con asa sobre placas BMM/S. Incubar a 30°C.

45 Se seleccionaron *Transcuerpos* de alta afinidad hasta $\alpha_{IIb}\beta_3$ entre la biblioteca de despliegue de levadura sobre las placas recubiertas con integrina purificada de plaquetas humanas. La purificación de integrina $\alpha_{IIb}\beta_3$ se llevó a cabo de acuerdo con Hillman *et al.* Con pequeñas modificaciones (Hillman *et al.*, 2002, Protein Expr. Purif. 25(3): 494 - 502). Se diluyó la integrina (dilución 1:30) en tampón de recubrimiento (tampón borato 50 mM, pH 9,5) y se cubrió sobre una placa Petri durante toda una a 4°C. Después de enjuagar tres veces con tampón A (Tris-HCl 20 mM, pH 7,5, NaCl 150 mM, MgCl₂ 1 mM, CaCl₂ 1 mM y MnCl₂ 1 mM), se bloqueó la placa durante 2 horas a TA con tampón de unión + 1% BSA. Se recogieron sobre la placa 40 unidades de DO_{600} de biblioteca de RGD de levadura durante

toda una noche, desarrolladas en BMM/S (2% p/v de sacarosa) a 30°C 200 rpm, mediante centrifugación y se volvieron a suspender en 10 ml de tampón de unión con 1% de BSA y péptido inhibidor de GRGDSP (1 µg/ml para la 1ª selección y 10 µg/ml para la 2ª selección) y se incubaron con integrina recubierta durante 2 horas a TA con agitación suave. Las células no unidas se retiraron por lavado con 3 lavados de tampón de unión y se recogieron las células unidas.

Producción de Proteínas Solubles.

La región codificadora de ADN para las secuencias de RGD que contienen el lóbulo N se recuperó del vector de despliegue de levadura y se clonó en un vector de expresión de levadura produciendo las moléculas de Tf. Estas se expresaron en fermentaciones de lote de alta densidad de células y se purificaron mediante cromatografía en columna. Los ejemplos de secuencias de clones seleccionados en comparación con las secuencias de ligandos $\alpha_{IIb}\beta_3$ naturales se proporciona a continuación:

Genoteca	C	X	X	X	R	G	D	X	X	X	C
n°41		N	S	P				T	P	I	
n°32		I	I	P				R	P	R	
n°14		Q	G	S				T	P	N	
n°16		G	F	H				T	P	D	
n°34		R	Q	H				T	P	A	
n°39		Y	P	S				F	A	H	
n°30		N	S	R				V	P	I	
n°37		H	P	I				T	Y	Y	
Flavonidina	C	R	I	A				F	P	D	C
Quistrina	C	R	I	P				M	P	D	C
Fibronectina		V	T	G				S	P	A	

Análisis de unión a integrina.

La unión directa de *Transcuerpos* de RGD a integrina se midió mediante ELISA. Se diluyó la integrina en tampón de recubrimiento y se revistieron sobre placas Maxisorb ELISA durante toda una noche a 4°C. Las placas se bloquearon con tampón A que contenía 1% de BSA y 0,05% de Tween-20 (tampón de unión) durante 2 horas a temperatura ambiente con agitación. Los *Transcuerpos* diluidos en tampón de unión se dejaron que se unieran a integrina a 4°C durante toda una noche. El material no unido se lavó con tampón A que contenía 0,05% de Tween-20 (tampón de lavado). Se detectó el material unido usando anti-transferrina biotinilada y estreptavidina-HRP en tampón de unión, lavando con tampón de lavado después de cada etapa. La cuantificación de unión se realizó sobre un SpectraMax Gemini EM y software SoftMax usando sustrato fluorescente QuantaBlue. Véase la figura 14

La unión de los *Transcuerpos* también se determinó en formato de ensayo de ELISA de competición. Se revistió la integrina y las placas se bloquearon como antes. Los *Transcuerpos*, o péptido de control y se añadieron fibronectina biotinilada en tampón de unión a los pocillos apropiados y se dejó que se unieran durante toda una noche a 4°C. Se lavaron las placas con tampón de lavado y se detectó la fibronectina unida usando estreptavidina-HRP. La cuantificación de la unión se realizó como antes. Véase la figura 15.

Análisis de agregación plaquetaria.

Los ensayos de agregación plaquetaria se llevó a cabo con plasma rico en plaquetas (PRP). Las plaquetas recientes se obtuvieron de ratones B6. Las plaquetas se diluyeron hasta $2,5 \times 10^5$ /ml para uso. La agregación plaquetaria se midió con un agregómetro a 37°C con agitación (900 rpm). La muestra de ensayo se incubó con PRP durante 10 minutos antes de la adición de ADP (10 µM). El grado de agregación de plaquetas se controló de manera continua durante 6 min mediante turbidimetría y se expresó como el incremento en transmisión de luz. Véase la figura 16.

Ejemplo 7 – Modificación por ingeniería genética del tallo de MUC3

Para el despliegue de proteínas sobre la superficie de levadura, la proteína MUC3 humana (hMUC3) se optimizó sobre un tallo de despliegue. La secuencia de ADN de hMUC3 (Genebank n° de acceso AAC02272) se modificó para retirar algo de la repetitividad y se permitió la construcción mediante codón optimizado de oligonucleótidos sintéticos para levadura. La secuencia de tallo de hMUC3 resultante (véase la secuencia más adelante) se sintetizó mediante GenScript Corporation Piscataway, Nueva Jersey.

El ADN de hMUC3 modificado se probó que era inestable en *E. coli* cuando estaba en un vector de despliegue de levadura. La secuencia de ADN de hMUC3 además se modificó por ingeniería genética mediante mutagénesis al azar y selección de clones que eran estables en *E. coli*. La secuencia de ADN de hMUC3 se sometió a 2 rondas de mutagénesis al azar mediante PCR propensa a error, kit GeneMorph II (Stratagene). Los fragmentos de PCR resultantes se digirieron con *SpeI/HindIII* y se ligaron en pREX0855 digerido con *SpeI/HindIII*. La mezcla de ligadura se usó para transformar *E. coli* XL1-Blue. Se sembraron las células, se desarrollaron durante toda una noche y se recuperó toda la biomasa de las placas. Se recuperó el ADN de plásmido, se digirió con *NotI/XmnI* y se separó mediante electroforesis en gel de agarosa. La banda de ADN de tamaño correcto se cortó del gel y se purificó. El ADN resultante que contenía la casete de despliegue de *NotI* con tallo de hMUC3 se clonó en un vector de despliegue de levadura basado en pSAC35, pREX1003, mediante reparación de huecos en levadura.

Todas las levaduras resultantes se cultivaron en medio BMM/S y se sometieron a selección por afinidad sobre placas recubiertas con anticuerpo anti-FLAG. Un anticuerpo anti-Flag (Sigma F9291) se diluyó (dilución 1:25) en tampón de recubrimiento (50 mM tampón borato, pH 9,5) y se revistió sobre una placa Petri durante toda una noche

a 4°C. Después de enjuagar tres veces con tampón de unión (PBS-T, 2 mM EDTA, 1% de BSA), la placa se bloqueó durante 2 horas a TA con tampón de unión. 40 unidades de DO₆₀₀ de cultivo de levadura durante toda una noche se recogieron y se suspendieron en 10 ml de tampón de unión y se incubó con anticuerpo recubierto durante 2 horas a TA con agitación suave. Las células no unidas se retiraron por lavado con 3 lavados usando tampón de unión y las células se recogieron células.

Los ligantes mostraron un alto nivel de despliegue de proteína como se demostró mediante análisis FACS (figura 17). Los clones individuales se aislaron obteniéndose la secuencia de ADN de hMUC3 que era estable en *E. coli* y proporcionó un alto nivel de despliegue en levadura.

Gen de MUC3 humano modificado (SEC ID NO: 76)

10 TCTACTAGTACACCAAGCTATACTACATCAATTACTAGCACTGAGACACC
ATCTCATTCAACTCCTTCTAGTTGACTTCAATTACAACACTACAGAGACAC
CCAGTCACAGCACCCCATCTTACACTAGCTCTGTCTCCACATCCGAGACT
ACATCACATTCTACTCCATCAGAAACAAGCTCCAGTAGAACAACAGAAAG
CACCTCTTATAGTTCACCTAGCTCCACTTCATCTAACACAATTACTGAGA
15 CAAGCTCACACTCCACTCCTAGTACTGCTACTTCTATTTCTCGACCGAA
ACACCTAGTTCAAGTACACCATCTGTATCATCGTCCATTACTGTTACTGA
GAGTTCATCTCATAGCACTCCTGGAGCTACTTCCACTTTGACATCGAGTG
AACTTCTACTTGGTCAACACCATCTAGCACAAGTTCTATTATGTCAAGC
TCCTACACTTCAGCTGACACTCCATCTGAAACATCAGTTTATACTTCCAG
20 CGAAACCCCATCGTCCTCAAGCCCACTAGCACATCTTTGATTTCTAGTT
CGAAGTCAACATCGACCAGTACACCTTCGTTTACTTCTTCGATTACTAGC
ACTGAGACCTCCTCATATTCTGCTAGTTCCTATACACCTTCAGTTAGTAG
CACAGCAAGTTCTAGCAAGAACAACGAGTTCCACTGCTTCTATAAGCA
GTACAGAGACTGTTAGTTCATCGACTAGCTCTGTCTCTAGTACTATTCT
25 TCTTCTCAATCCACAAGTTATTCTACACCATCATTCTCCAGTTCGGCAAC
AAGCAGTGTTACTCCATTGCATTCAACACCATCTCTACCATCTTGGGTTA
CTACAAGTAAGACCACATCACATATTACACCAGGTCTGACTTCGTCCATG
TCTTCGAGCGAGACCTATAGCCATAGTACTCCAGGTTTTACAAGCTCTAT
TACTTCGACAGAATCGACAAGTGAGTCAACTCCATCATTGTCCAGTTCTA
30 CAATTTATAGTACTGTTTCAACATCTACTACAGCTATTACTTTCACATTTT
ACAATTCTGAGACAGCTGTTACTCCAACACCAGTTACACCCTCTAGCTT
GAGTACAGATATCCCAACTACAAGTTTGAGAACTTTGACTCCTTCTCTG
TTGGTACCTCGACTTCCTTGACAACACTACAAGTACTTCCATCAATTCCA
ACAGACATTAGCACTTTGCCAACAAGAACTCATATTATTTCAAGCTCACC
35 ATCTATTCAATCAACAGAGACTAGCAGTTTGGTTGGTACTACATCTCCAA
CTATGTCAACAGTTAGAATGACTTTGAGAATTACAGAGAACACTCCAATT
TCTTCATTAGTACTTCCATTGTTGTTATTCCAGAGACACCAACTCAAAC
ACCACCAGTTTTGACTTCTGCTACAGGTAACATCACCAGCTCCAA
CTACAGTTACTTTTGGTTCTACTGACTCATCTACATCAACTTTGCATAAG
40 CTTTTG

Ejemplo 8 – Distribución al azar de restos de aminoácido de transferrina expuestos a superficie

Se pueden crear bibliotecas de proteínas de fusión de transferrina mediante distribución al azar de residuos de aminoácido expuestos a superficie, *por ejemplo*, 6 restos expuestos de superficie, *es decir*, no inserciones, en varios

puntos dentro de la molécula de Tf. Estos residuos no necesitan ser secuenciales pero sería preferible que estuvieran agrupados en una región con las cadenas laterales que se proyectan hacia fuera del cuerpo de la molécula en la fase de disolvente. Una biblioteca con 6 residuos expuestos de superficie 6 solamente requerirían ~10⁷ proteínas y como tal solamente requerirían 20 a 40 transformaciones por biblioteca (dependiendo de la calidad del cebador).

Los ejemplos de tres posibles sitios son:

Sitio I	Y85	Sitio II	K276	Sitio III	H207
	G86		D277		S208
	S87		K280		F211
	E89		Q283		E212
	D90		S286		A215
	Q92		D297 (Opcional S298)		N216 (Opcional)
					K217 (Opcional)

Ejemplo 9 - Expresión de proteína de fusión de transferrina sobre la superficie de células de mamíferos

Materiales y procedimientos.

Se desarrollaron células HEK-293 y se mantuvieron en DMEM (HyClone) suplementado con 10% de suero fetal bovino (HyClone), L-glutamina (GIBCO-BRL) y HEPES (GIBCO-BRL).

La secuencia de ADN de GPI de fosfatasa alcalina de placenta humana (ALPP) se generó mediante hibridación de cebadores P0926 y P0927 y después se extiende sobre la *Taq* ADN polimerasa.

P0926

ACTAAGCTTGACCTGGCTCCACAAATTGCTGGCTATACCGACGCCGCGCATCCGGGTAGATCCGTGGTCCCAGCTTTGCTTCCTCT

P0927

TTATCTAGATTATTATGGAGCAGTGGCAGTTTCCAACAGTAACAGAGTACCGGCCAGCAGAGGAAGCAAAGCTGGGAC

Las condiciones de reacción de la PCR fueron 94°C durante 1 min., 25 ciclos 94°C durante 40 seg, 55°C durante 40 seg y 72°C durante 2 minutos seguidos de una extensión de 7 min a 72°C. El producto de PCR se digirió con *HindIII/XbaI*, se clonó en los sitios de *HindIII/XbaI* de pREX0757 (figura 18), que retiraron una secuencia de GPI de levadura y después se secuenció. El plásmido resultante, pREX1234, (figura 19) contenía transferrina modificada (mTf) con una etiqueta FLAG Tag en el extremo N y la secuencia de GPI humana en el extremo C. Esta casete de mTf se modificó en los extremos 5' y 3' para clonación en el vector de expresión de mamífero mediante PCR con los cebadores P2225 y P2226 en las siguientes condiciones: 94°C durante 1 min., 25 ciclos 94°C durante 40 seg, 55°C durante 40 seg y 72°C durante 2 min seguidos de una extensión de 7 min a 72°C.

P2225

CACCATGAGGCTCGCCGTGGGAGCCCTG

P2226

TTATTATGGAGCAGTGGCAGTTTC

Este producto de PCR se clonó en pcDNA3.1 (Invitrogen, Carlsbad, CA) para generar pREX1235 (figura 20). La secuencia de la casete de mTf en pREX1235 se confirmó mediante secuenciación de ADN.

Células HEK-293 se transfectaron con pREX 1235 usando FuGENE 6 (Roche) de acuerdo con el protocolo del fabricante. Se lavaron las células en 1X solución salina tamponada con fosfato (PBS 1X) y se lisaron durante toda una noche a 4°C con tampón de lisis celular (Tris-HCl 20 mM a pH 7,4, NaCl 150 mM, 1% de Tritón X-100, EDTA 1 mM) y una mezcla del cóctel inhibidor de la fosfatasa Halt™ (Pierce) y un cóctel inhibidor de la proteasa (Sigma). Se sedimentaron las células durante 30 min a 14.000 rpm a 4°C. Se recogió el sobrenadante y se dejó unirse a perlas de agarosa anti-Flag (Sigma) durante 2 h a 4°C. Se lavaron los sedimentos una vez en tampón de lisis, dos veces en solución salina tamponada con Tris (TBS) (Tris-HCl 20 mM a pH 7,4, NaCl 150 mM) + 0,5% de Tritón X-100 y una vez en PBS 1X. Las muestras se eluyeron en tampón de carga de muestra de LDS NuPAGE® + agente reductor de muestra (Invitrogen). La proteína recuperada se corrió sobre un NuPAGE™ 4 - 12% de gel Bis-Tris (Invitrogen) y se transfirió a una membrana PVDF (Invitrogen) con tampón de transferencia NuPAGE® según las instrucciones del fabricante. Las manchas se bloquearon en 5% de leche en polvo no grasa en PBS-T (PBS 1X+ 0,2% de Tween-20) durante 1 h a temperatura ambiente y se sondearon con anticuerpos primarios biotinilados bien contra FLAG (1:1000 anti-FLAG BioM2 de ratón; Sigma) o bien contra transferrina humana (1:4000 transferrina anti-humana de pollo biotinilada; Accurate Chemical & Scientific) durante 2 h a temperatura ambiente (TA) seguido de incubación con estreptavidina-peroxidasa de rábano picante (SA-HRP) (1:2500 ImmunoPure® Conjugado de Estreptavidina

peroxidasa de rábano picante; Pierce) durante 1 h a TA. De manera alternativa, para los experimentos de biotilación de superficie, las manchas se sondaron con solamente SA-HRP (1:2500; Pierce). Todas las manchas se desarrollaron usando el SuperSignal West Dura Extended Duration Substrate (Pierce) y se visualizaron sobre Flúor-S™ Multimagar (Bio-Rad) usando el software Quantity One (v.4.5.0; Bio-Rad).

Células HEK-293 se transfectaron durante 24 h como se ha descrito anteriormente. Se lavaron las células tres veces en PBS 1X estéril, enfriada por hielo. Después las células se trataron con 0,5 mg/ml de reactivo Sulfo-NHS-LC-Biotina (Pierce) en PBS 1X y se incubaron sobre una plataforma oscilante durante 2 hr a 4°C. Después de la biotilación de superficie, se lavaron las células tres veces con tampón de inactivación (PBS 1X + glicina 100 mM) y se volvieron a suspender en tapón de lisis celular para la inmunoprecipitación e inmunotransferencia como se ha descrito anteriormente.

Células HEK-293 se transfectaron durante 24 h como se ha descrito anteriormente. Las células se lavaron una vez en PBS 1X estéril, después se volvieron a suspender en PBS 1X adicional y se contaron. $4,25 \times 10^5$ células por muestra se transfirieron a tubos eppendorf de 1,5 ml. Se sedimentaron las células a 4000 rpm durante 3 min a 4°C. Después se bloquearon las células en 1X PBS-T (0,05% de Tween-20) durante 30 min a TA, se centrifugaron, se volvieron a suspender en anticuerpo primario (1:25 - 1:625 anti-FLAG BioM2 de ratón; Sigma) y se dejaron incuban durante 30 min a TA. Se lavaron las células dos veces en 1X PBS-T, se volvieron a suspender en anticuerpo secundario (1:100 anti-ratón-APC de cabra; Molecular Probes) y se dejaron incuban durante 30 min a TA. Después de dos lavados adicionales en PBS-T 1X, las células se volvieron a suspender en tampón de células (Agilent) para llevar a las células hasta una concentración final de 2×10^6 células/ml y se cargaron en un procesador de fluorescencia de células (Kit Cell Fluorescence LabChip®, Agilent) según las instrucciones del fabricante. Todos los análisis de FACS se realizaron usando el bioanalizador Agilent 2100 y software 2100.

Resultados.

Los lisados de las células transfectadas con bien pREX1235 o bien células transfectadas simuladas se sometieron a inmunoprecipitación con perlas anti-FLAG- agarosa y posterior transferencia de western. La sonda de transferencias de western con anticuerpos o bien anti-FLAG o anti-transferrina demuestra claramente la presencia de tanto transferrina como restos FLAG (figura 21).

pREX1235 o células transfectadas simuladas se biotinilaron en superficie antes de la lisis y se sometieron a inmunoprecipitación con anti-FLAG-agarosa y transferencia de western usando estreptavidina-HRP. La biotilación de superficie asegura que solamente las proteínas de membrana de plasma expuesta se detectan en transferencia de western. Se observó una proteína que puede precipitar por FLAG demostrando que la fusión FLAG-transferrina se expresa sobre la superficie de la célula (figura 22).

De manera adicional, las células transfectadas se sometieron a análisis de aislamiento de células activada por fluorescencia (FACS). Se observó un incremento de 4 veces en la tinción en las células transfectadas con pREX1235 con relación a las células transfectadas simuladas, demostrando de manera evidente la expresión de la superficie celular de la fusión FLAG-transferrina (figura 23).

Listado de secuencias

<110> WANG, Baiyang TURNER, Andrew

<120> Bibliotecas de proteínas de fusión de transferrina

<130> BIOR-022/00US

<160> 76

<170> PatentIn version 3.3

<210> 1

<211> 2318

<212> ADN

<213> Homo sapiens

<220>

<221> CDS

<222> (51)..(2147)

<223> RenBank número de acceso: NM_00163, gen de transferrina y proteína transferrina

<220>

<221> sig_péptido

<222> (51)..(107)

<400> 1

gcacagaagc gagtccgact gtgctcgctg ctcagcgccg cacccggaag atg agg	56
Met Arg	
1	
ctc gcc gtg gga gcc ctg ctg gtc tgc gcc gtc ctg ggg ctg tgt ctg	104
Leu Ala Val Gly Ala Leu Leu Val Cys Ala Val Leu Gly Leu Cys Leu	
5 10 15	
gct gtc cct gat aaa act gtg aga tgg tgt gca gtg tcg gag cat gag	152
Ala Val Pro Asp Lys Thr Val Arg Trp Cys Ala Val Ser Glu His Glu	
20 25 30	
gcc act aag tgc cag agt ttc cgc gac cat atg aaa agc gtc att cca	200
Ala Thr Lys Cys Gln Ser Phe Arg Asp His Met Lys Ser Val Ile Pro	
35 40 45 50	
tcc gat ggt ccc agt gtt gct tgt gtg aag aaa gcc tcc tac ctt gat	248
Ser Asp Gly Pro Ser Val Ala Cys Val Lys Lys Ala Ser Tyr Leu Asp	
55 60 65	
tgc atc agg gcc att gcg gca aac gaa gcg gat gct gtg aca ctg gat	296
Cys Ile Arg Ala Ile Ala Ala Asn Glu Ala Asp Ala Val Thr Leu Asp	
70 75 80	
gca ggt ttg gtg tat gat gct tac ctg gct ccc aat aac ctg aag cct	344
Ala Gly Leu Val Tyr Asp Ala Tyr Leu Ala Pro Asn Asn Leu Lys Pro	
85 90 95	
gtg gtg gca gag ttc tat ggg tca aaa gag gat cca cag act ttc tat	392
Val Val Ala Glu Phe Tyr Gly Ser Lys Glu Asp Pro Gln Thr Phe Tyr	
100 105 110	
tat gct gtt gct gtg gtg aag aag gat agt ggc ttc cag atg aac cag	440
Tyr Ala Val Ala Val Val Lys Lys Asp Ser Gly Phe Gln Met Asn Gln	
115 120 125 130	

ctt cga ggc aag aag tcc tgc cac acg ggt cta ggc agg tcc gct ggg Leu Arg Gly Lys Lys Ser Cys His Thr Gly Leu Gly Arg Ser Ala Gly 135 140 145	488
tgg aac atc ccc ata ggc tta ctt tac tgt gac tta cct gag cca cgt Trp Asn Ile Pro Ile Gly Leu Leu Tyr Cys Asp Leu Pro Glu Pro Arg 150 155 160	536
aaa cct ctt gag aaa gca gtg gcc aat ttc ttc tcg ggc agc tgt gcc Lys Pro Leu Glu Lys Ala Val Ala Asn Phe Phe Ser Gly Ser Cys Ala 165 170 175	584
cct tgt gcg gat ggg acg gac ttc ccc cag ctg tgt caa ctg tgt cca Pro Cys Ala Asp Gly Thr Asp Phe Pro Gln Leu Cys Gln Leu Cys Pro 180 185 190	632
ggg tgt ggc tgc tcc acc ctt aac caa tac ttc ggc tac tcg gga gcc Gly Cys Gly Cys Ser Thr Leu Asn Gln Tyr Phe Gly Tyr Ser Gly Ala 195 200 205 210	680
ttc aag tgt ctg aag gat ggt gct ggg gat gtg gcc ttt gtc aag cac Phe Lys Cys Leu Lys Asp Gly Ala Gly Asp Val Ala Phe Val Lys His 215 220 225	728
tcg act ata ttt gag aac ttg gca aac aag gct gac agg gac cag tat Ser Thr Ile Phe Glu Asn Leu Ala Asn Lys Ala Asp Arg Asp Gln Tyr 230 235 240	776
gag ctg ctt tgc ctg gac aac acc cgg aag ccg gta gat gaa tac aag Glu Leu Leu Cys Leu Asp Asn Thr Arg Lys Pro Val Asp Glu Tyr Lys 245 250 255	824
gac tgc cac ttg gcc cag gtc cct tct cat acc gtc gtg gcc cga agt Asp Cys His Leu Ala Gln Val Pro Ser His Thr Val Val Ala Arg Ser 260 265 270	872
atg ggc ggc aag gag gac ttg atc tgg gag ctt ctc aac cag gcc cag Met Gly Gly Lys Glu Asp Leu Ile Trp Glu Leu Leu Asn Gln Ala Gln 275 280 285 290	920
gaa cat ttt ggc aaa gac aaa tca aaa gaa ttc caa cta ttc agc tct Glu His Phe Gly Lys Asp Lys Ser Lys Glu Phe Gln Leu Phe Ser Ser 295 300 305	968
cct cat ggg aag gac ctg ctg ttt aag gac tct gcc cac ggg ttt tta Pro His Gly Lys Asp Leu Leu Phe Lys Asp Ser Ala His Gly Phe Leu 310 315 320	1016
aaa gtc ccc ccc agg atg gat gcc aag atg tac ctg ggc tat gag tat Lys Val Pro Pro Arg Met Asp Ala Lys Met Tyr Leu Gly Tyr Glu Tyr 325 330 335	1064
gtc act gcc atc cgg aat cta cgg gaa ggc aca tgc cca gaa gcc cca Val Thr Ala Ile Arg Asn Leu Arg Glu Gly Thr Cys Pro Glu Ala Pro 340 345 350	1112
aca gat gaa tgc aag cct gtg aag tgg tgt gcg ctg agc cac cac gag Thr Asp Glu Cys Lys Pro Val Lys Trp Cys Ala Leu Ser His His Glu 355 360 365 370	1160
agg ctc aag tgt gat gag tgg agt gtt aac agt gta ggg aaa ata gag Arg Leu Lys Cys Asp Glu Trp Ser Val Asn Ser Val Gly Lys Ile Glu 375 380 385	1208
tgt gta tca gca gag acc acc gaa gac tgc atc gcc aag atc atg aat	1256

Cys	Val	Ser	Ala	Glu	Thr	Thr	Glu	Asp	Cys	Ile	Ala	Lys	Ile	Met	Asn	
			390					395					400			
gga	gaa	gct	gat	gcc	atg	agc	ttg	gat	gga	ggg	ttt	gtc	tac	ata	gcg	1304
Gly	Glu	Ala	Asp	Ala	Met	Ser	Leu	Asp	Gly	Gly	Phe	Val	Tyr	Ile	Ala	
		405					410					415				
ggc	aag	tgt	ggt	ctg	gtg	cct	gtc	ttg	gca	gaa	aac	tac	aat	aag	agc	1352
Gly	Lys	Cys	Gly	Leu	Val	Pro	Val	Leu	Ala	Glu	Asn	Tyr	Asn	Lys	Ser	
		420				425					430					
gat	aat	tgt	gag	gat	aca	cca	gag	gca	ggg	tat	ttt	gct	gta	gca	gtg	1400
Asp	Asn	Cys	Glu	Asp	Thr	Pro	Glu	Ala	Gly	Tyr	Phe	Ala	Val	Ala	Val	
		435			440					445					450	
gtg	aag	aaa	tca	gct	tct	gac	ctc	acc	tgg	gac	aat	ctg	aaa	ggc	aag	1448
Val	Lys	Lys	Ser	Ala	Ser	Asp	Leu	Thr	Trp	Asp	Asn	Leu	Lys	Gly	Lys	
			455						460					465		
aag	tcc	tgc	cat	acg	gca	gtt	ggc	aga	acc	gct	ggc	tgg	aac	atc	ccc	1496
Lys	Ser	Cys	His	Thr	Ala	Val	Gly	Arg	Thr	Ala	Gly	Trp	Asn	Ile	Pro	
			470					475					480			
atg	ggc	ctg	ctc	tac	aat	aag	atc	aac	cac	tgc	aga	ttt	gat	gaa	ttt	1544
Met	Gly	Leu	Leu	Tyr	Asn	Lys	Ile	Asn	His	Cys	Arg	Phe	Asp	Glu	Phe	
		485					490					495				
ttc	agt	gaa	ggt	tgt	gcc	cct	ggg	tct	aag	aaa	gac	tcc	agt	ctc	tgt	1592
Phe	Ser	Glu	Gly	Cys	Ala	Pro	Gly	Ser	Lys	Lys	Asp	Ser	Ser	Leu	Cys	
		500				505					510					
aag	ctg	tgt	atg	ggc	tca	ggc	cta	aac	ctg	tgt	gaa	ccc	aac	aac	aaa	1640
Lys	Leu	Cys	Met	Gly	Ser	Gly	Leu	Asn	Leu	Cys	Glu	Pro	Asn	Asn	Lys	
		515			520					525					530	
gag	gga	tac	tac	ggc	tac	aca	ggc	gct	ttc	agg	tgt	ctg	gtt	gag	aag	1688
Glu	Gly	Tyr	Tyr	Gly	Tyr	Thr	Gly	Ala	Phe	Arg	Cys	Leu	Val	Glu	Lys	
				535				540						545		
gga	gat	gtg	gcc	ttt	gtg	aaa	cac	cag	act	gtc	cca	cag	aac	act	ggg	1736
Gly	Asp	Val	Ala	Phe	Val	Lys	His	Gln	Thr	Val	Pro	Gln	Asn	Thr	Gly	
			550					555					560			
gga	aaa	aac	cct	gat	cca	tgg	gct	aag	aat	ctg	aat	gaa	aaa	gac	tat	1784
Gly	Lys	Asn	Pro	Asp	Pro	Trp	Ala	Lys	Asn	Leu	Asn	Glu	Lys	Asp	Tyr	
		565					570					575				
gag	ttg	ctg	tgc	ctt	gat	ggt	acc	agg	aaa	cct	gtg	gag	gag	tat	gcg	1832
Glu	Leu	Leu	Cys	Leu	Asp	Gly	Thr	Arg	Lys	Pro	Val	Glu	Glu	Tyr	Ala	
		580				585					590					
aac	tgc	cac	ctg	gcc	aga	gcc	ccg	aat	cac	gct	gtg	gtc	aca	cgg	aaa	1880
Asn	Cys	His	Leu	Ala	Arg	Ala	Pro	Asn	His	Ala	Val	Val	Thr	Arg	Lys	
				600						605					610	
gat	aag	gaa	gct	tgc	gtc	cac	aag	ata	tta	cgt	caa	cag	cag	cac	cta	1928
Asp	Lys	Glu	Ala	Cys	Val	His	Lys	Ile	Leu	Arg	Gln	Gln	Gln	His	Leu	
			615					620						625		
ttt	gga	agc	aac	gta	act	gac	tgc	tcg	ggc	aac	ttt	tgt	ttg	ttc	cgg	1976
Phe	Gly	Ser	Asn	Val	Thr	Asp	Cys	Ser	Gly	Asn	Phe	Cys	Leu	Phe	Arg	
			630					635					640			
tcg	gaa	acc	aag	gac	ctt	ctg	ttc	aga	gat	gac	aca	gta	tgt	ttg	gcc	2024
Ser	Glu	Thr	Lys	Asp	Leu	Leu	Phe	Arg	Asp	Asp	Thr	Val	Cys	Leu	Ala	

ES 2 535 358 T3

645	650	655	
aaa ctt cat gac aga aac aca tat gaa aaa tac tta gga gaa gaa tat			2072
Lys Leu His Asp Arg Asn Thr Tyr Glu Lys Tyr Leu Gly Glu Glu Tyr			
660	665	670	
gtc aag gct gtt ggt aac ctg aga aaa tgc tcc acc tca tca ctc ctg			2120
Val Lys Ala Val Gly Asn Leu Arg Lys Cys Ser Thr Ser Ser Leu Leu			
675	680	685	690
gaa gcc tgc act ttc cgt aga cct taa aatctcagag gtagggctgc			2167
Glu Ala Cys Thr Phe Arg Arg Pro			
695			
caccaaggtg aagatgggaa cgcagatgat ccatgagttt gccctggttt cactggccca			2227
agtggtttgt gctaaccacg tctgtcttca cagctctgtg ttgccatgtg tgctgaacaa			2287
aaaataaaaa ttattattga ttttatattt c			2318

<210> 2

<211> 698

<212> PRT

5 <213> Homo sapiens

<400> 2

ES 2 535 358 T3

Met Arg Leu Ala Val Gly Ala Leu Leu Val Cys Ala Val Leu Gly Leu
1 5 10 15

Cys Leu Ala Val Pro Asp Lys Thr Val Arg Trp Cys Ala Val Ser Glu
20 25 30

His Glu Ala Thr Lys Cys Gln Ser Phe Arg Asp His Met Lys Ser Val
35 40 45

Ile Pro Ser Asp Gly Pro Ser Val Ala Cys Val Lys Lys Ala Ser Tyr
50 55 60

Leu Asp Cys Ile Arg Ala Ile Ala Ala Asn Glu Ala Asp Ala Val Thr
65 70 75 80

Leu Asp Ala Gly Leu Val Tyr Asp Ala Tyr Leu Ala Pro Asn Asn Leu
85 90 95

Lys Pro Val Val Ala Glu Phe Tyr Gly Ser Lys Glu Asp Pro Gln Thr
100 105 110

Phe Tyr Tyr Ala Val Ala Val Val Lys Lys Asp Ser Gly Phe Gln Met
115 120 125

Asn Gln Leu Arg Gly Lys Lys Ser Cys His Thr Gly Leu Gly Arg Ser
130 135 140

Ala Gly Trp Asn Ile Pro Ile Gly Leu Leu Tyr Cys Asp Leu Pro Glu

ES 2 535 358 T3

145		150		155		160
Pro Arg Lys Pro	Leu Glu Lys Ala Val	Ala Asn Phe Phe Ser Gly Ser				
	165	170			175	
Cys Ala Pro Cys	Ala Asp Gly Thr Asp Phe Pro	Gln Leu Cys Gln Leu				
	180	185			190	
Cys Pro Gly Cys	Gly Cys Ser Thr Leu Asn Gln Tyr	Phe Gly Tyr Ser				
	195	200			205	
Gly Ala Phe Lys	Cys Leu Lys Asp Gly Ala Gly Asp	Val Ala Phe Val				
	210	215			220	
Lys His Ser Thr	Ile Phe Glu Asn Leu Ala Asn Lys	Ala Asp Arg Asp				
	225	230			235	240
Gln Tyr Glu Leu	Leu Cys Leu Asp Asn Thr Arg Lys	Pro Val Asp Glu				
	245	250			255	
Tyr Lys Asp Cys	His Leu Ala Gln Val Pro Ser His	Thr Val Val Ala				
	260	265			270	
Arg Ser Met Gly	Gly Lys Glu Asp Leu Ile Trp Glu	Leu Leu Asn Gln				
	275	280			285	
Ala Gln Glu His	Phe Gly Lys Asp Lys Ser Lys Glu	Phe Gln Leu Phe				
	290	295			300	
Ser Ser Pro His	Gly Lys Asp Leu Leu Phe Lys Asp	Ser Ala His Gly				
	305	310			315	320
Phe Leu Lys Val	Pro Pro Arg Met Asp Ala Lys Met	Tyr Leu Gly Tyr				
	325	330			335	
Glu Tyr Val Thr	Ala Ile Arg Asn Leu Arg Glu Gly	Thr Cys Pro Glu				
	340	345			350	
Ala Pro Thr Asp	Glu Cys Lys Pro Val Lys Trp Cys	Ala Leu Ser His				
	355	360			365	
His Glu Arg Leu	Lys Cys Asp Glu Trp Ser Val Asn	Ser Val Gly Lys				
	370	375			380	
Ile Glu Cys Val	Ser Ala Glu Thr Thr Glu Asp Cys	Ile Ala Lys Ile				
	385	390			395	400
Met Asn Gly Glu	Ala Asp Ala Met Ser Leu Asp Gly	Gly Phe Val Tyr				
	405	410			415	

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

Lys Ser Asp Asn Cys Glu Asp Thr Pro Glu Ala Gly Tyr Phe Ala Val
435 440 445

Ala Val Val Lys Lys Ser Ala Ser Asp Leu Thr Trp Asp Asn Leu Lys
450 455 460

Gly Lys Lys Ser Cys His Thr Ala Val Gly Arg Thr Ala Gly Trp Asn
465 470 475 480

Ile Pro Met Gly Leu Leu Tyr Asn Lys Ile Asn His Cys Arg Phe Asp
485 490 495

Glu Phe Phe Ser Glu Gly Cys Ala Pro Gly Ser Lys Lys Asp Ser Ser
500 505 510

Leu Cys Lys Leu Cys Met Gly Ser Gly Leu Asn Leu Cys Glu Pro Asn
515 520 525

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

Glu Lys Gly Asp Val Ala Phe Val Lys His Gln Thr Val Pro Gln Asn
545 550 555 560

Thr Gly Gly Lys Asn Pro Asp Pro Trp Ala Lys Asn Leu Asn Glu Lys
565 570 575

Asp Tyr Glu Leu Leu Cys Leu Asp Gly Thr Arg Lys Pro Val Glu Glu
580 585 590

Tyr Ala Asn Cys His Leu Ala Arg Ala Pro Asn His Ala Val Val Thr
595 600 605

Arg Lys Asp Lys Glu Ala Cys Val His Lys Ile Leu Arg Gln Gln Gln
610 615 620

His Leu Phe Gly Ser Asn Val Thr Asp Cys Ser Gly Asn Phe Cys Leu
625 630 635 640

Phe Arg Ser Glu Thr Lys Asp Leu Leu Phe Arg Asp Asp Thr Val Cys
645 650 655

Leu Ala Lys Leu His Asp Arg Asn Thr Tyr Glu Lys Tyr Leu Gly Glu
660 665 670

Glu Tyr Val Lys Ala Val Gly Asn Leu Arg Lys Cys Ser Thr Ser Ser
675 680 685

Leu Leu Glu Ala Cys Thr Phe Arg Arg Pro
690 695

<210> 3

<211> 679

<212> PRT

5 <213> Homo sapiens

<220>

<221> MISC_FEATURE

<223> Proteína transferrina madura

<400> 3

ES 2 535 358 T3

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

ES 2 535 358 T3

Cys Gly Cys Ser Thr Leu Asn Gln Tyr Phe Gly Tyr Ser Gly Ala Phe
 180 185 190
 Lys Cys Leu Lys Asp Gly Ala Gly Asp Val Ala Phe Val Lys His Ser
 195 200 205
 Thr Ile Phe Glu Asn Leu Ala Asn Lys Ala Asp Arg Asp Gln Tyr Glu
 210 215 220
 Leu Leu Cys Leu Asp Asn Thr Arg Lys Pro Val Asp Glu Tyr Lys Asp
 225 230 235 240
 Cys His Leu Ala Gln Val Pro Ser His Thr Val Val Ala Arg Ser Met
 245 250 255
 Gly Gly Lys Glu Asp Leu Ile Trp Glu Leu Leu Asn Gln Ala Gln Glu
 260 265 270
 His Phe Gly Lys Asp Lys Ser Lys Glu Phe Gln Leu Phe Ser Ser Pro
 275 280 285
 His Gly Lys Asp Leu Leu Phe Lys Asp Ser Ala His Gly Phe Leu Lys
 290 295 300
 Val Pro Pro Arg Met Asp Ala Lys Met Tyr Leu Gly Tyr Glu Tyr Val
 305 310 315 320
 Thr Ala Ile Arg Asn Leu Arg Glu Gly Thr Cys Pro Glu Ala Pro Thr
 325 330 335
 Asp Glu Cys Lys Pro Val Lys Trp Cys Ala Leu Ser His His Glu Arg
 340 345 350
 Leu Lys Cys Asp Glu Trp Ser Val Asn Ser Val Gly Lys Ile Glu Cys
 355 360 365
 Val Ser Ala Glu Thr Thr Glu Asp Cys Ile Ala Lys Ile Met Asn Gly
 370 375 380
 Glu Ala Asp Ala Met Ser Leu Asp Gly Gly Phe Val Tyr Ile Ala Gly
 385 390 395 400
 Lys Cys Gly Leu Val Pro Val Leu Ala Glu Asn Tyr Asn Lys Ser Asp
 405 410 415
 Asn Cys Glu Asp Thr Pro Glu Ala Gly Tyr Phe Ala Val Ala Val Val
 420 425 430
 Lys Lys Ser Ala Ser Asp Leu Thr Trp Asp Asn Leu Lys Gly Lys Lys

435	440	445																	
Ser	Cys	His	Thr	Ala	Val	Gly	Arg	Thr	Ala	Gly	Trp	Asn	Ile	Pro	Met				
450						455					460								
Gly	Leu	Leu	Tyr	Asn	Lys	Ile	Asn	His	Cys	Arg	Phe	Asp	Glu	Phe	Phe				
465					470					475					480				
Ser	Glu	Gly	Cys	Ala	Pro	Gly	Ser	Lys	Lys	Asp	Ser	Ser	Leu	Cys	Lys				
				485					490					495					
Leu	Cys	Met	Gly	Ser	Gly	Leu	Asn	Leu	Cys	Glu	Pro	Asn	Asn	Lys	Glu				
			500					505					510						
Gly	Tyr	Tyr	Gly	Tyr	Thr	Gly	Ala	Phe	Arg	Cys	Leu	Val	Glu	Lys	Gly				
			515				520					525							
Asp	Val	Ala	Phe	Val	Lys	His	Gln	Thr	Val	Pro	Gln	Asn	Thr	Gly	Gly				
	530					535					540								
Lys	Asn	Pro	Asp	Pro	Trp	Ala	Lys	Asn	Leu	Asn	Glu	Lys	Asp	Tyr	Glu				
545					550					555					560				
Leu	Leu	Cys	Leu	Asp	Gly	Thr	Arg	Lys	Pro	Val	Glu	Glu	Tyr	Ala	Asn				
				565					570					575					
Cys	His	Leu	Ala	Arg	Ala	Pro	Asn	His	Ala	Val	Val	Thr	Arg	Lys	Asp				
			580					585					590						
Lys	Glu	Ala	Cys	Val	His	Lys	Ile	Leu	Arg	Gln	Gln	Gln	His	Leu	Phe				
			595				600					605							
Gly	Ser	Asn	Val	Thr	Asp	Cys	Ser	Gly	Asn	Phe	Cys	Leu	Phe	Arg	Ser				
	610					615					620								
Glu	Thr	Lys	Asp	Leu	Leu	Phe	Arg	Asp	Asp	Thr	Val	Cys	Leu	Ala	Lys				
625				630						635					640				
Leu	His	Asp	Arg	Asn	Thr	Tyr	Glu	Lys	Tyr	Leu	Gly	Glu	Glu	Tyr	Val				
				645					650					655					
Lys	Ala	Val	Gly	Asn	Leu	Arg	Lys	Cys	Ser	Thr	Ser	Ser	Leu	Leu	Glu				
			660					665					670						
Ala	Cys	Thr	Phe	Arg	Arg	Pro													
			675																

<210> 4
 <211> 12
 <212> PRT
 <213> Secuencia artificial

5 <220>
 <223> Secuencia variante de ajuste neutrofilico
 <400> 4

Glu Asp Cys Ile Ala Leu Lys Gly Glu Ala Asp Ala
 1 5 10

10 <210> 5
 <211> 732
 <212> ADN
 <213> Homo sapiens
 <220>

15 <221> misc_feature
 <223> MUCI humano
 <400> 5

```

acaggttctg gtcattgcaag ctctacccca ggtggagaaa aggagacttc ggctaccag 60
agaagttcag tgcccagctc tactgagaag aatgctgtga gtatgaccag cagcgtactc 120
tccagccaca gccccgggtc aggtctctcc accactcagg gacaggatgt cactctggcc 180
ccggccacgg aaccagcttc aggttcagct gccacctggg gacaggatgt cacctcggtc 240
ccagtcacca ggccagccct gggctccacc accccgccag cccacgatgt cacctcagcc 300
ccggacaaca agccagcccc gggctccacc gcccccccag cccacgggtg cacctcggcc 360
ccggacaaca ggcccgcctt ggcgtccacc gccctccag tccacaatgt cacctcggcc 420
tcaggctctg catcaggctc agcttctact ctggtgcaca acggcacctc tgccagggct 480
accacaaccc cagccagcaa gagcactcca ttctcaattc ccagccacca ctctgatact 540
cctaccaccc ttgccagcca tagcaccaag actgatgcca gtagcactca ccatagcacg 600
gtacctctc tcacctctc caatcacagc acttctcccc agttgtctac tggggtctct 660
ttctttttcc tgtcttttca catttcaaac ctccagttta attcctctct ggaagatccc 720
agcaccgact ac 732
    
```

<210> 6
 20 <211> 744
 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> hMUCI con sitios SpeI y HindIII

<400> 6

actagtagacag gttctgggtca tgcaagctct accccaggtg gagaaaagga gacttcgggt	60
accagagagaa gttcagtgcc cagctctact gagaagaatg ctgtgagtat gaccagcagc	120
gtactctcca gccacagccc cggttcaggc tcctccacca ctcagggaca ggatgtcact	180
ctggccccgg ccacggaacc agcttcagggt tcagctgcc cctggggaca ggatgtcacc	240
tcgggtcccag tcaccaggcc agccctgggc tccaccaccc cggcagccca cgatgtcacc	300
tcagccccgg acaacaagcc agccccgggc tccaccgccc cccagccca cgggtgtcacc	360
tcggccccgg acaacaggcc cgccttggcg tccaccgccc ctccagtcca caatgtcacc	420
tcggcctcag gctctgcatc aggctcagct tctactctgg tgcacaacgg cacctctgcc	480
agggtacca caaccccagc cagcaagagc actccattct caattcccag ccaccactct	540
gatactccta ccacccttgc cagccatagc accaagactg atgccagtag cactcaccat	600
agcacggtac ctctctcac ctctccaat cacagcactt ctcccagtt gtctactggg	660
gtctctttct ttttcctgtc ttttcacatt tcaaacctcc agtttaattc ctctctggaa	720
gatcccagca ccgactacaa gctt	744

<210> 7

<211> 32

5 <212> ADN

<213> Secuencia artificial

<220>

<223> Cebador

<400> 7

10 tactagtagt aagcttatct ctgataata at 32

<210> 8

<211> 36

<212> ADN

<213> Secuencia artificial

15 <220>

<223> Cebador

<400> 8

agctattatt atctagagat aagcttgata ctgata 36

<210> 9

20 <211> 86

<212> ADN

<213> Secuencia artificial

<220>

<223> Cebador
 <400> 9
 atcaagcttt ctggttctgc agtcgctaca tactctgttc cttctatctc gagtacttac 60
 caaggtgctg ctaatatcaa ggttct 86
 <210> 10
 5 <211> 80
 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> Cebador
 10 <400> 10
 ttatctagat tattagaata caactggaag agcgagtagc aaccacataa agtttccaag 60
 aaccttgata ttagcagcac 80
 <210> 11
 <211> 26
 <212> ADN
 15 <213> Secuencia artificial
 <220>
 <223> Cebador
 <400> 11
 gtagtcggtg ctgggatctt ccagag 26
 20 <210> 12
 <211> 26
 <212> ADN
 <213> Secuencia artificial
 <220>
 25 <223> Cebador
 <400> 12
 gctgctcctc acagtgctta cagttg 26
 <210> 13
 <211> 32
 30 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> Cebador
 <400> 13
 35 cgtactagta caggttctgg tcatgcaagc tc 32
 <210> 14

<211> 31
<212> ADN
<213> Secuencia artificial
<220>
5 <223> Cebador
<400> 14
tctaagcttg tagtcggtgc tgggatcttc c 31
<210> 15
<211> 40
10 <212> PRT
<213> Saccharomyces cerevisiae
<220>
<221> MISC_FEATURE
<223> anclaje GPI de levadura YIR019C
15 <400> 15
Ser Gly Ser Ala Val Ala Thr Tyr Ser Val Pro Ser Ile Ser Ser Thr
1 5 10 15
Tyr Gln Gly Ala Ala Asn Ile Lys Val Leu Gly Asn Phe Met Trp Leu
20 25 30
Leu Leu Ala Leu Pro Val Val Phe
35 40
<210> 16
<211> 2094
<212> ADN
20 <213> Homo sapiens
<220>
<221> misc_feature
<223> Transferrina modificada
<220>
25 <221> misc_feature
<223> casete de expression Tf
<400> 16

atgaggctcg ccgtagggagc cctgctgggc tgcgcgcgtcc tggggctgtg tctggcggta	60
cctgataaaa ctgtgagatg gtgtgcagtg tcggagcatg aggccactaa gtgccagagt	120
ttccgcgacc atatgaaaag cgtcattcca tccgatgggc ccagtgttgc ttgtgtgaag	180
aaagcctcct accttgattg catcagggcc attgcggcaa acgaagcggg tgctgtgaca	240
ctggatgcag gtttggtgta tgatgcttac ctggctccca ataacctgaa gcctgtgggtg	300
gcagagttct atgggtcaaa agaggatcca cagactttct attatgctgt tgctgtgggtg	360
aagaaggata gtggcttcca gatgaaccag cttcgaggca agaagtcctg ccacacgggt	420
ctaggcaggt ccgctgggtg gaacatcccc ataggcttac tttactgtga cttacctgag	480
ccacgtaaac ctcttgagaa agcagtggcc aatttcttct cgggcagctg tgccccctgt	540
gcggatggga cggacttccc ccagctgtgt caactgtgtc caggggtgtgg ctgctccacc	600
cttaaccaat acttcggcta ctcgggagcc ttcaagtgtc tgaaggatgg tgctggggat	660
gtggcctttg tcaagcactc gactatattt gagaacttgg caaacaaggc tgacagggac	720
cagtatgagc tgctttgcct ggacaacacc cggaagccgg tagatgaata caaggactgc	780
cacttggccc aggtcccttc tcataccgtc gtggcccgaa gtatgggcgg caaggaggac	840
ttgatctggg agcttctcaa ccaggcccag gaacattttg gcaaagacaa atcaaaagaa	900
ttccaactat tcagctctcc tcatgggaag gacctgctgt ttaaggactc tgcccacggg	960
tttttaaaaag tcccccccag gatggatgcc aagatgtacc tgggctatga gtatgtcact	1020
gccatccgga atctacggga aggcacatgc ccagaagccc caacagatga atgcaagcct	1080

gtgaagtggg gtgcgctgag ccaccacgag aggctcaagt gtgatgagtg gagtgttaac 1140
 agtgtagggg aaatagagtg tgtatcagca gagaccaccg aagactgcat cgccaagatc 1200
 atgaatggag aagctgatgc catgagcttg gatggagggt ttgtctacat agcgggcaag 1260
 tgtgggtctgg tgcctgtctt ggcagaaaac tacaataagg ctgataattg tgaggataca 1320
 ccagaggcag ggtatcttgc tgtagcagtg gtgaagaaat cagcttctga cctcacctgg 1380
 gacaatctga aaggcaagaa gtcttgccat acggcagttg gcagaaccgc tggctggaac 1440
 atccccatgg gcctgctcta caataagatc aaccactgca gatttgatga atttttcagt 1500
 gaagggttggt cccctgggtc taagaaagac tccagtctct gtaagctgtg tatgggctca 1560
 ggcctaaacc tctgtgaacc caacaacaaa gagggatact acggetacac aggcgctttc 1620
 aggtgtctgg ttgagaaggg agatgtggcc tttgtgaaac accagactgt cccacagaac 1680
 actgggggaa aaaaccctga tccatgggct aagaatctga atgaaaaaga ctatgagttg 1740
 ctgtgccttg atggtactag gaaacctgtg gaggagtatg cgaactgcca cctggccaga 1800
 gccccgaatc acgctgtggg cacacggaaa gataaggaag catgcgtcca caagatatta 1860
 cgtcaacagc agcacctatt tggaagcaac gtagctgact gctcgggcaa cttttgtttg 1920
 ttccggtcgg aaaccaagga ctttctgttc agagatgaca cagtatgttt ggccaaactt 1980
 catgacagaa acacatatga aaaatactta ggagaagaat atgtcaaggc tgttggtaac 2040
 ctgagaaaat gctccacctc atcactcctg gaagcctgca ctttccgtcg acct 2094

<210> 17

<211> 5

<212> PRT

5 <213> Desconocido

<220>

<223> Péptido espaciador

<220>

<221> REPETICIÓN

10 <222> (1)..(5)

<223> La secuencia se puede repetir

<400> 17

Gly Gly Gly Gly Ser

1

5

15 <210> 18

<211> 6

<212> PRT

<213> Desconocido

<220>

<223> Péptido de engarce
 <220>
 <221> REPETICIÓN
 <222> (1)..(6)
 5 <223> Se puede repetir
 <400> 18

Pro Glu Ala Pro Thr Asp
 1 5

<210> 19
 10 <211> 88
 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> Cebador P1172
 15 <220>
 <221> misc_feature
 <222> (20)..(64)
 <223> n es a, c, g o t
 <400> 19

60

20 **ccaactattc agctctcctn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn**
nnnnncatggg aaggacctgc tgtttaag **88**

<210> 20
 <211> 19
 <212> ADN
 <213> Secuencia artificial
 25 <220>
 <223> cebador P1173
 <400> 20

aggagagctg aatagttgg 19

<210> 21
 30 <211> 22
 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> cebador P1174
 35 <400> 21

ctggatgcag gtttggtgta tg 22

<210> 22
 <211> 22

<212> ADN
 <213> Secuencia artificial
 <220>
 <223> cebador P1227
 5 <400> 22
 tcatgatctt ggcgatgcag tc 22
 <210> 23
 <211> 8
 <212> PRT
 10 <213> Secuencia artificial
 <220>
 <223> Secuencia flag-tag
 <400> 23

 Asp Tyr Lys Asp Asp Asp Asp Lys
 15 1 5
 <210> 24
 <211> 34
 <212> ADN
 <213> Secuencia artificial
 20 <220>
 <223> cebador de PCR para el gen de levadurasd AGA1
 <400> 24
 cagatctaga acaaccgcta tcagctcatt atcc 34
 <210> 25
 25 <211> 32
 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> cebador de PCR para el gen de levadurasd AGA1
 30 <400> 25
 cagaaagctt agtagtggaa acttctgtag tg 32
 <210> 26
 <211> 40
 <212> PRT
 35 <213> Secuencia artificial
 <220>
 <223> secuencia de GPI de huMDP modificada
 <220>
 <221> misc_feature

<222> (3)..(3)

<223> Xaa puede ser cualquier aminoácido de origen natural

<220>

<221> misc_feature

5 <222> (7)..(8)

<223> Xaa puede ser cualquier aminoácido de origen natural

<400> 26

Xaa Gln Xaa Gly Gly Ser Xaa Xaa Thr Ile Gly Gly Tyr Ser Gly Ala
1 5 10 15

Ala Ser Ser Leu Gln Arg Thr Ile Gly Leu Leu Leu Ala Ser Leu Ala
20 25 30

Pro Leu Val Leu Ala Ser Leu Leu
35 40

<210> 27

10 <211> 83

<212> ADN

<213> Secuencia artificial

<220>

<223> cebador P2035

15 <220>

<221> misc_feature

<222> (11)..(12)

<223> n es a, c, g, o t

<220>

20 <221> misc_feature

<222> (17)..(18)

<223> n es a, c, g, o t

<220>

<221> misc_feature

25 <222> (29)..(30)

<223> n es a, c, g, o t

<220>

<221> misc_feature

<222> (32)..(33)

30 <223> n es a, c, g, o t

<400> 27

ctacaagctt nnkcaannkg gtggttctnn knnkactatt ggtggttatt ctggtgctgc 60

ttcttccttg cagagaacta ttg 83

	<210> 28		
	<211> 84		
	<212> ADN		
	<213> Secuencia artificial		
5	<220>		
	<223> Cebador P2036		
	<400> 28		
	gatgtctaga ttattataac aaagaagcta aaaccaatgg agctaaagaa gccataaca	60	
	aaccaatagt tctctgcaag gaag	84	
	<210> 29		
10	<211> 138		
	<212> ADN		
	<213> Secuencia artificial		
	<220>		
15	<223> Secuencia de ADN resultante de la digestión con enzimas de restricción de la reacción de alinear y prolongar oligos P2035 y P2036		
	<220>		
	<221> misc_feature		
	<222> (7)..(8)		
	<223> n es a, c, g, o t		
20	<220>		
	<221> misc_feature		
	<222> (13)..(14)		
	<223> n es a, c, g, o t		
	<220>		
25	<221> misc_feature		
	<222> (25)..(26)		
	<223> n es a, c, g, o t		
	<220>		
	<221> misc_feature		
30	<222> (28)..(29)		
	<223> n es a, c, g, o t		
	<400> 29		
	aagcttnnkc aannkgttg ttctnnknnk actattggtg gttattctgg tgctgcttct	60	
	tccttgcaaga gaactattgg ttgtttattg gcttcttttag ctccattggt tttagcttct	120	
	ttgttataat aatctaga	138	
	<210> 30		
35	<211> 42		

<212> PRT
 <213> Secuencia artificial
 <220>
 5 <223> Secuencia de aminoácidos de secuencia de ADN que resulta de la reacción de alinear y prolongar oligos P2035 y P2036
 <220>
 <221> misc_feature
 <222> (3)..(3)
 <223> Xaa puede ser cualquier aminoácido de origen natural
 10 <220>
 <221> misc_feature
 <222> (5)..(5)
 <223> Xaa puede ser cualquier aminoácido de origen natural
 <220>
 15 <221> misc_feature
 <222> (9)..(10)
 <223> Xaa puede ser cualquier aminoácido de origen natural
 <400> 30

Lys Leu Xaa Gln Xaa Gly Gly Ser Xaa Xaa Thr Ile Gly Gly Tyr Ser
1 5 10 15

Gly Ala Ala Ser Ser Leu Gln Arg Thr Ile Gly Leu Leu Leu Ala Ser
20 25 30

Leu Ala Pro Leu Val Leu Ala Ser Leu Leu
35 40

20 <210> 31
 <211> 24
 <212> ADN
 <213> Secuencia artificial
 <220>
 25 <223> Secuencia de anclaje de GPI consenso
 <220>
 <221> misc_feature
 <222> (1)..(2)
 <223> n es a, c, g, o t
 30 <220>
 <221> misc_feature
 <222> (7)..(8)
 <223> n es a, c, g, o t
 <220>

<221> misc_feature
 <222> (19)..(20)
 <223> n es a, c, g, o t
 <220>
 5 <221> misc_feature
 <222> (22)..(23)
 <223> n es a, c, g, o t
 <400> 31
 nnkcaannkg gtggttctnn knnk 24
 10 <210> 32
 <211> 24
 <212> ADN
 <213> Secuencia artificial
 <220>
 15 <223> Secuencia de anclaje de GPI de clon 23
 <400> 32
 tgtcaatagg gtggttctag gcct 24
 <210> 33
 <211> 24
 20 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> Secuencia de anclaje de GPI de clon 8
 <400> 33
 25 tgtcaaattg gtggttctta gtgt 24
 <210> 34
 <211> 6
 <212> PRT
 <213> Secuencia artificial
 30 <220>
 <223> Secuencia de anclaje de GPI de clon 8
 <400> 34

Cys Gln Ile Gly Gly Ser

1

5

35 <210> 35
 <211> 24
 <212> ADN
 <213> Secuencia artificial
 <220>

<223> Secuencia de anclaje de GPI de clon 15
 <400> 35
 cagcaatatg gtggttctgt ggat 24
 <210> 36
 5 <211> 8
 <212> PRT
 <213> Secuencia artificial
 <220>
 <223> Secuencia de anclaje de GPI de clon 15
 10 <400> 36
 Glu Gln Tyr Gly Gly Ser Val Asp
 1 5
 <210> 37
 <211> 24
 15 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> Secuencia de anclaje de GPI de clon 14
 <400> 37
 20 tctcaagttg gtggttctac ttgg 24
 <210> 38
 <211> 8
 <212> ADN
 <213> Secuencia artificial
 25 <220>
 <223> Secuencia de anclaje de GPI de clon 14
 <400> 38
 Ser Gln Val Gly Gly Ser Thr Trp
 1 5
 30 <210> 39
 <211> 24
 <212> ADN
 <213> Secuencia artificial
 <220>
 35 <223> Secuencia de anclaje de GPI de clon 5
 <220>
 <221> misc_feature
 <222> (1)..(2)
 <223> n es a, c, g, o t

<220>
 <221> misc_feature
 <222> (7)..(8)
 <223> n es a, c, g, o t
 5 <220>
 <221> misc_feature
 <222> (19)..(20)
 <223> n es a, c, g, o t
 <220>
 10 <221> misc_feature
 <222> (22)..(23)
 <223> n es a, c, g, o t
 <400> 39
 nnkcaannkg gtggttctnn knnk 24
 15 <210> 40
 <211> 24
 <212> ADN
 <213> Secuencia artificial
 <220>
 20 <223> Secuencia de anclaje de GPI de clon 2
 <400> 40
 catcaagggtg gtggttctat tcgg 24
 <210> 41
 <211> 8
 25 <212> PRT
 <213> Secuencia artificial
 <220>
 <223> Secuencia de anclaje de GPI de clon 2
 <400> 41
 30 **His Gln Gly Gly Gly Ser Ile Arg**
1 5
 <210> 42
 <211> 24
 <212> ADN
 35 <213> Secuencia artificial
 <220>
 <223> Secuencia de anclaje de GPI de clon 6
 <220>
 <221> misc_feature

<222> (1)..(2)
 <223> n es a, c, g, o t
 <220>
 <221> misc_feature
 5 <222> (7)..(8)
 <223> n es a, c, g, o t
 <220>
 <221> misc_feature
 <222> (19)..(20)
 10 <223> n es a, c, g, o t
 <220>
 <221> misc_feature
 <222> (22)..(23)
 <223> n es a, c, g, o t
 15 <400> 42
 nnkcaannkg gtggtctnn knnk 24
 <210> 43
 <211> 24
 <212> ADN
 20 <213> Secuencia artificial
 <220>
 <223> Secuencia de anclaje de GPI de clon 12
 <400> 43
 catcaattgg gtggtctgt tacg 24
 25 <210> 44
 <211> 8
 <212> PRT
 <213> Secuencia artificial
 <220>
 30 <223> Secuencia de anclaje de GPI de clon 12
 <400> 44

His Gln Leu Gly Gly Ser Val Thr

1

5

<210> 45
 35 <211> 24
 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> Secuencia de anclaje de GPI de clon 18

<400> 45
tatcaatcgg gtggttctgg gact 24
<210> 46
<211> 8
5 <212> ADN
<213> Secuencia artificial
<220>
<223> Secuencia de anclaje de GPI de clon 18
<400> 46
10 Tyr Gln Ser Gly Gly Ser Gly Thr
1 5
<210> 47
<211> 24
<212> ADN
15 <213> Secuencia artificial
<220>
<223> Secuencia de anclaje de GPI de clon 13
<400> 47
gggcaatatg gtggttctta gtgg 24
20 <210> 48
<211> 6
<212> PRT
<213> Secuencia artificial
<220>
25 <223> Secuencia de anclaje de GPI de clon 13
<400> 48
Gly Gln Tyr Gly Gly Ser
1 5
<210> 49
30 <211> 24
<212> ADN
<213> Secuencia artificial
<220>
<223> Secuencia de anclaje de GPI de clon 1
35 <400> 49
gtgcaagcgg gtggttctga ttag 24
<210> 50
<211> 7
<212> PRT

<210> 55
 <211> 24
 <212> ADN
 <213> Secuencia artificial
 5 <220>
 <223> Secuencia de anclaje de GPI de clon 7
 <400> 55
 ctgcaatgtg gtggttctta gtgg 24
 <210> 56
 10 <211> 6
 <212> PRT
 <213> Secuencia artificial
 <220>
 <223> Secuencia de anclaje de GPI de clon 7
 15 <400> 56
 Leu Gln Lys Gly Gly Ser
 1 5
 <210> 57
 <211> 24
 20 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> Secuencia de anclaje de GPI de clon 16
 <400> 57
 25 tagcaactgg gtggttcttt tggg 24
 <210> 58
 <211> 24
 <212> ADN
 <213> Secuencia artificial
 30 <220>
 <223> Secuencia de anclaje de GPI de clon 17
 <400> 58
 tagcaatatg gtggttctgt tcta 24
 <210> 59
 35 <211> 24
 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> Secuencia de anclaje de GPI de clon 19

<400> 59
 cttcaagtgg gtggttcttt gtag 24
 <210> 60
 <211> 7
 5 <212> PRT
 <213> Secuencia artificial
 <220>
 <223> Secuencia de anclaje de GPI de clon 19
 <400> 60
 10 Leu Gln Val Gly Gly Ser Leu
 1 5
 <210> 61
 <211> 24
 <212> ADN
 15 <213> Secuencia artificial
 <220>
 <223> Secuencia de anclaje de GPI de clon 24
 <400> 61
 tagcaatttg gtggttctca tgcg 24
 20 <210> 62
 <211> 24
 <212> ADN
 <213> Secuencia artificial
 <220>
 25 <223> Secuencia de anclaje de GPI de clon 9
 <400> 62
 cggcaacggg gtggttctaa gtgg 24
 <210> 63
 <211> 8
 30 <212> PRT
 <213> Secuencia artificial
 <220>
 <223> Secuencia de anclaje de GPI de clon 9
 <400> 63
 35 Arg Gln Arg Gly Gly Ser Lys Trp
 1 5
 <210> 64
 <211> 24
 <212> ADN

<213> Secuencia artificial

$\langle 220 \rangle$

<223> Secuencia de anclaje de GPI de clon 20

<400> 64

5 tcgcaaactg gtggttctgt tgct 24

<210> 65

 $\langle 211 \rangle$ 8

<212> PRT

<213> Secuencia artificial

10 $\langle 220 \rangle$

<223> Secuencia de anclaje de GPI de clon 20

<400> 65

Ser Gln Thr Gly Gly Ser Val Ala

1

5

15 <210> 66

<211> 2215

<212> ADN

<213> Homo sapiens

$\langle 220 \rangle$

20 <221> sig_péptido

<222> (40)..(93)

<220>

<221> CDS

<222> (40) (1866)

25 <220>

<221> mat_péptido

<222> (112)..(1866)

<400> 66

agctttttctc ttctgtcaac cccacacgcc tttggcaca atg aag tgg gta acc	54
Met Lys Trp Val Thr	
-20	
ttt att tcc ctt ctt ttt ctc ttt agc tcg gct tat tcc agg ggt gtg	102
Phe Ile Ser Leu Leu Phe Leu Phe Ser Ser Ala Tyr Ser Arg Gly Val	
-15 -10 -5	
ttt cgt cga gat gca cac aag agt gag gtt gct cat cgg ttt aaa gat	150
Phe Arg Arg Asp Ala His Lys Ser Glu Val Ala His Arg Phe Lys Asp	
-1 1 5 10	
ttg gga gaa gaa aat ttc aaa gcc ttg gtg ttg att gcc ttt gct cag	198
Leu Gly Glu Glu Asn Phe Lys Ala Leu Val Leu Ile Ala Phe Ala Gln	
15 20 25	
tat ctt cag cag tgt cca ttt gaa gat cat gta aaa tta gtg aat gaa	246
Tyr Leu Gln Gln Cys Pro Phe Glu Asp His Val Lys Leu Val Asn Glu	
30 35 40 45	
gta act gaa ttt gca aaa aca tgt gtt gct gat gag tca gct gaa aat	294
Val Thr Glu Phe Ala Lys Thr Cys Val Ala Asp Glu Ser Ala Glu Asn	
50 55 60	
tgt gac aaa tca ctt cat acc ctt ttt gga gac aaa tta tgc aca gtt	342
Cys Asp Lys Ser Leu His Thr Leu Phe Gly Asp Lys Leu Cys Thr Val	
65 70 75	
gca act ctt cgt gaa acc tat ggt gaa atg gct gac tgc tgt gca aaa	390
Ala Thr Leu Arg Glu Thr Tyr Gly Glu Met Ala Asp Cys Cys Ala Lys	
80 85 90	
caa gaa cct gag aga aat gaa tgc ttc ttg caa cac aaa gat gac aac	438
Gln Glu Pro Glu Arg Asn Glu Cys Phe Leu Gln His Lys Asp Asp Asn	
95 100 105	
cca aac ctc ccc cga ttg gtg aga cca gag gtt gat gtg atg tgc act	486
Pro Asn Leu Pro Arg Leu Val Arg Pro Glu Val Asp Val Met Cys Thr	
110 115 120 125	
gct ttt cat gac aat gaa gag aca ttt ttg aaa aaa tac tta tat gaa	534
Ala Phe His Asp Asn Glu Glu Thr Phe Leu Lys Lys Tyr Leu Tyr Glu	
130 135 140	
att gcc aga aga cat cct tac ttt tat gcc ccg gaa ctc ctt ttc ttt	582
Ile Ala Arg Arg His Pro Tyr Phe Tyr Ala Pro Glu Leu Leu Phe Phe	
145 150 155	

gct aaa agg tat aaa gct gct ttt aca gaa tgt tgc caa gct gct gat Ala Lys Arg Tyr Lys Ala Ala Phe Thr Glu Cys Cys Gln Ala Ala Asp 160 165 170	630
aaa gct gcc tgc ctg ttg cca aag ctc gat gaa ctt cgg gat gaa ggg Lys Ala Ala Cys Leu Leu Pro Lys Leu Asp Glu Leu Arg Asp Glu Gly 175 180 185	678
aag gct tcg tct gcc aaa cag aga ctc aag tgt gcc agt ctc caa aaa Lys Ala Ser Ser Ala Lys Gln Arg Leu Lys Cys Ala Ser Leu Gln Lys 190 195 200 205	726
ttt gga gaa aga gct ttc aaa gca tgg gca gta gct cgc ctg agc cag Phe Gly Glu Arg Ala Phe Lys Ala Trp Ala Val Ala Arg Leu Ser Gln 210 215 220	774
aga ttt ccc aaa gct gag ttt gca gaa gtt tcc aag tta gtg aca gat Arg Phe Pro Lys Ala Glu Phe Ala Glu Val Ser Lys Leu Val Thr Asp 225 230 235	822
ctt acc aaa gtc cac acg gaa tgc tgc cat gga gat ctg ctt gaa tgt Leu Thr Lys Val His Thr Glu Cys Cys His Gly Asp Leu Leu Glu Cys 240 245 250	870
gct gat gac agg gcg gac ctt gcc aag tat atc tgt gaa aat caa gat Ala Asp Asp Arg Ala Asp Leu Ala Lys Tyr Ile Cys Glu Asn Gln Asp 255 260 265	918
tcg atc tcc agt aaa ctg aag gaa tgc tgt gaa aaa cct ctg ttg gaa Ser Ile Ser Ser Lys Leu Lys Glu Cys Cys Glu Lys Pro Leu Leu Glu 270 275 280 285	966
aaa tcc cac tgc att gcc gaa gtg gaa aat gat gag atg cct gct gac Lys Ser His Cys Ile Ala Glu Val Glu Asn Asp Glu Met Pro Ala Asp 290 295 300	1014
ttg cct tca tta gct gct gat ttt gtt gaa agt aag gat gtt tgc aaa Leu Pro Ser Leu Ala Ala Asp Phe Val Glu Ser Lys Asp Val Cys Lys 305 310 315	1062
aac tat gct gag gca aag gat gtc ttc ctg ggc atg ttt ttg tat gaa Asn Tyr Ala Glu Ala Lys Asp Val Phe Leu Gly Met Phe Leu Tyr Glu 320 325 330	1110
tat gca aga agg cat cct gat tac tct gtc gtg ctg ctg ctg aga ctt Tyr Ala Arg Arg His Pro Asp Tyr Ser Val Val Leu Leu Leu Arg Leu 335 340 345	1158
gcc aag aca tat gaa acc act cta gag aag tgc tgt gcc gct gca gat Ala Lys Thr Tyr Glu Thr Thr Leu Glu Lys Cys Cys Ala Ala Ala Asp 350 355 360 365	1206
cct cat gaa tgc tat gcc aaa gtg ttc gat gaa ttt aaa cct ctt gtg Pro His Glu Cys Tyr Ala Lys Val Phe Asp Glu Phe Lys Pro Leu Val 370 375 380	1254
gaa gag cct cag aat tta atc aaa caa aat tgt gag ctt ttt gag cag Glu Glu Pro Gln Asn Leu Ile Lys Gln Asn Cys Glu Leu Phe Glu Gln 385 390 395	1302
ctt gga gag tac aaa ttc cag aat gcg cta tta gtt cgt tac acc aag Leu Gly Glu Tyr Lys Phe Gln Asn Ala Leu Leu Val Arg Tyr Thr Lys 400 405 410	1350
aaa gta ccc caa gtg tca act cca act ctt gta gag gtc tca aga aac	1398

Lys	Val	Pro	Gln	Val	Ser	Thr	Pro	Thr	Leu	Val	Glu	Val	Ser	Arg	Asn	
415						420					425					
cta	gga	aaa	gtg	ggc	agc	aaa	tgt	tgt	aaa	cat	cct	gaa	gca	aaa	aga	1446
Leu	Gly	Lys	Val	Gly	Ser	Lys	Cys	Cys	Lys	His	Pro	Glu	Ala	Lys	Arg	
430					435					440					445	
atg	ccc	tgt	gca	gaa	gac	tat	cta	tcc	gtg	gtc	ctg	aac	cag	tta	tgt	1494
Met	Pro	Cys	Ala	Glu	Asp	Tyr	Leu	Ser	Val	Val	Leu	Asn	Gln	Leu	Cys	
				450					455					460		
gtg	ttg	cat	gag	aaa	acg	cca	gta	agt	gac	aga	gtc	acc	aaa	tgc	tgc	1542
Val	Leu	His	Glu	Lys	Thr	Pro	Val	Ser	Asp	Arg	Val	Thr	Lys	Cys	Cys	
			465					470					475			
aca	gaa	tcc	ttg	gtg	aac	agg	cga	cca	tgc	ttt	tca	gct	ctg	gaa	gtc	1590
Thr	Glu	Ser	Leu	Val	Asn	Arg	Arg	Pro	Cys	Phe	Ser	Ala	Leu	Glu	Val	
		480					485					490				
gat	gaa	aca	tac	gtt	ccc	aaa	gag	ttt	aat	gct	gaa	aca	ttc	acc	ttc	1638
Asp	Glu	Thr	Tyr	Val	Pro	Lys	Glu	Phe	Asn	Ala	Glu	Thr	Phe	Thr	Phe	
	495					500					505					
cat	gca	gat	ata	tgc	aca	ctt	tct	gag	aag	gag	aga	caa	atc	aag	aaa	1686
His	Ala	Asp	Ile	Cys	Thr	Leu	Ser	Glu	Lys	Glu	Arg	Gln	Ile	Lys	Lys	
510					515					520					525	
caa	act	gca	ctt	gtt	gag	ctc	gtg	aaa	cac	aag	ccc	aag	gca	aca	aaa	1734
Gln	Thr	Ala	Leu	Val	Glu	Leu	Val	Lys	His	Lys	Pro	Lys	Ala	Thr	Lys	
				530					535					540		
gag	caa	ctg	aaa	gct	gtt	atg	gat	gat	ttc	gca	gct	ttt	gta	gag	aag	1782
Glu	Gln	Leu	Lys	Ala	Val	Met	Asp	Asp	Phe	Ala	Ala	Phe	Val	Glu	Lys	
			545					550					555			
tgc	tgc	aag	gct	gac	gat	aag	gag	acc	tgc	ttt	gcc	gag	gag	ggt	aaa	1830
Cys	Cys	Lys	Ala	Asp	Asp	Lys	Glu	Thr	Cys	Phe	Ala	Glu	Glu	Gly	Lys	
		560					565				570					
aaa	ctt	gtt	gct	gca	agt	caa	gct	gcc	tta	ggc	tta	taacatctac				1876
Lys	Leu	Val	Ala	Ala	Ser	Gln	Ala	Ala	Leu	Gly	Leu					
	575					580				585						
atttaaaagc	atctcagcct	accatgagaa	taagagaaag	aaaatgaaga	tcaaaagctt											1936
attcatctgt	tttctttttc	gttggtgtaa	agccaacacc	ctgtctaaaa	aacataaatt											1996
tctttaatca	ttttgcctct	tttctctgtg	cttcaattaa	taaaaaatgg	aaagaatcta											2056
atagagtggg	acagcactgt	tattttttcaa	agatgtgttg	ctatcctgaa	aattctgtag											2116
gttctgtgga	agttccagtg	ttctctctta	ttccacttcg	gtagaggatt	tctagtttct											2176
gtgggcta	taaataaatc	actaatactc	ttctaagtt													2215

<210> 67

<211> 609

<212> PRT

5 <213> Homo sapiens

<400> 67

Met Lys Trp Val Thr Phe Ile Ser Leu Leu Phe Leu Phe Ser Ser Ala
-20 -15 -10

Tyr Ser Arg Gly Val Phe Arg Arg Asp Ala His Lys Ser Glu Val Ala
-5 -1 1 5

His Arg Phe Lys Asp Leu Gly Glu Glu Asn Phe Lys Ala Leu Val Leu
10 15 20

Ile Ala Phe Ala Gln Tyr Leu Gln Gln Cys Pro Phe Glu Asp His Val
25 30 35 40

Lys Leu Val Asn Glu Val Thr Glu Phe Ala Lys Thr Cys Val Ala Asp
45 50 55

Glu Ser Ala Glu Asn Cys Asp Lys Ser Leu His Thr Leu Phe Gly Asp
60 65 70

Lys Leu Cys Thr Val Ala Thr Leu Arg Glu Thr Tyr Gly Glu Met Ala
75 80 85

Asp Cys Cys Ala Lys Gln Glu Pro Glu Arg Asn Glu Cys Phe Leu Gln
90 95 100

His Lys Asp Asp Asn Pro Asn Leu Pro Arg Leu Val Arg Pro Glu Val
105 110 115 120

Asp Val Met Cys Thr Ala Phe His Asp Asn Glu Glu Thr Phe Leu Lys
125 130 135

Lys Tyr Leu Tyr Glu Ile Ala Arg Arg His Pro Tyr Phe Tyr Ala Pro
140 145 150

Glu Leu Leu Phe Phe Ala Lys Arg Tyr Lys Ala Ala Phe Thr Glu Cys
155 160 165

Cys Gln Ala Ala Asp Lys Ala Ala Cys Leu Leu Pro Lys Leu Asp Glu
170 175 180

Leu Arg Asp Glu Gly Lys Ala Ser Ser Ala Lys Gln Arg Leu Lys Cys
185 190 195 200

Ala Ser Leu Gln Lys Phe Gly Glu Arg Ala Phe Lys Ala Trp Ala Val
205 210 215

Ala Arg Leu Ser Gln Arg Phe Pro Lys Ala Glu Phe Ala Glu Val Ser
220 225 230

Lys Leu Val Thr Asp Leu Thr Lys Val His Thr Glu Cys Cys His Gly
235 240 245

ES 2 535 358 T3

Asp Leu Leu Glu Cys Ala Asp Asp Arg Ala Asp Leu Ala Lys Tyr Ile
 250 255 260
 Cys Glu Asn Gln Asp Ser Ile Ser Ser Lys Leu Lys Glu Cys Cys Glu
 265 270 275 280
 Lys Pro Leu Leu Glu Lys Ser His Cys Ile Ala Glu Val Glu Asn Asp
 285 290 295
 Glu Met Pro Ala Asp Leu Pro Ser Leu Ala Ala Asp Phe Val Glu Ser
 300 305 310
 Lys Asp Val Cys Lys Asn Tyr Ala Glu Ala Lys Asp Val Phe Leu Gly
 315 320 325
 Met Phe Leu Tyr Glu Tyr Ala Arg Arg His Pro Asp Tyr Ser Val Val
 330 335 340
 Leu Leu Leu Arg Leu Ala Lys Thr Tyr Glu Thr Thr Leu Glu Lys Cys
 345 350 355 360
 Cys Ala Ala Ala Asp Pro His Glu Cys Tyr Ala Lys Val Phe Asp Glu
 365 370 375
 Phe Lys Pro Leu Val Glu Glu Pro Gln Asn Leu Ile Lys Gln Asn Cys
 380 385 390
 Glu Leu Phe Glu Gln Leu Gly Glu Tyr Lys Phe Gln Asn Ala Leu Leu
 395 400 405
 Val Arg Tyr Thr Lys Lys Val Pro Gln Val Ser Thr Pro Thr Leu Val
 410 415 420
 Glu Val Ser Arg Asn Leu Gly Lys Val Gly Ser Lys Cys Cys Lys His
 425 430 435 440
 Pro Glu Ala Lys Arg Met Pro Cys Ala Glu Asp Tyr Leu Ser Val Val
 445 450 455
 Leu Asn Gln Leu Cys Val Leu His Glu Lys Thr Pro Val Ser Asp Arg
 460 465 470
 Val Thr Lys Cys Cys Thr Glu Ser Leu Val Asn Arg Arg Pro Cys Phe
 475 480 485
 Ser Ala Leu Glu Val Asp Glu Thr Tyr Val Pro Lys Glu Phe Asn Ala
 490 495 500

ES 2 535 358 T3

Glu Thr Phe Thr Phe His Ala Asp Ile Cys Thr Leu Ser Glu Lys Glu
505 510 515 520

Arg Gln Ile Lys Lys Gln Thr Ala Leu Val Glu Leu Val Lys His Lys
525 530 535

Pro Lys Ala Thr Lys Glu Gln Leu Lys Ala Val Met Asp Asp Phe Ala
540 545 550

Ala Phe Val Glu Lys Cys Cys Lys Ala Asp Asp Lys Glu Thr Cys Phe
555 560 565

Ala Glu Glu Gly Lys Lys Leu Val Ala Ala Ser Gln Ala Ala Leu Gly
570 575 580

Leu
585

<210> 68

<211> 1507

<212> ADN

5 <213> Homo sapiens

<400> 68

tctactagta	ctcccagcta	cactacctca	atcaccacca	ccgagacccc	ctcacacagt	60
actcccagct	acactacctc	aatcaccacc	accgagaccc	catcacacag	tactcccagc	120
ttcacttctt	caatcaccac	caccgagacc	acatcccaca	gtactcccag	cttcacttct	180
tcaatcagga	ccaccgagac	cacatcctac	agtactccca	gcttcacttc	ttcaaatacc	240
atcactgaga	ccacctcaca	cagtactccc	agctacatta	cctcaatcac	caccaccgag	300
acccctcaa	gcagtactcc	cagcttcagt	tcttcgatca	ccaccactga	gaccacatcc	360
cacagtactc	ccggcttcac	ttcttcaatc	accaccactg	agactacatc	ccacagtact	420
cccagcttca	cttcttcgat	caccaccact	gagaccacct	cacatgatac	tcccagcttc	480
acttcttcaa	tcaccaccag	tgagaccccc	tcacacagta	ctcccagctc	cacttcttta	540
atcaccacca	ccaagaccac	ctcacacagt	actcccagct	tcacttcttc	gatcaccacc	600
accgagacca	cctcacacag	tgctcgcagc	ttcacttctt	cgatcaccac	caccgagacc	660
acctcacaca	atactcggag	cttcacttct	tcgatcacca	ccaccgagac	caactctcac	720
agtactacca	gcttcacttc	ttcgatcacc	accaccgaga	ccacctcaca	cagtactccc	780
agcttcagtt	cttcaatcac	caccactgag	accccttac	acagtactcc	tggcctacct	840
tcgtgggtca	ccaccaccaa	gaccacctca	cacattactc	ctggcctcac	ttcttcaatc	900
accaccactg	agactacctc	acacagtact	cccggcttca	cttcttcaat	caccaccact	960
gagaccacct	cagagagtac	tcccagcctc	agttcttcaa	ccatctactc	cacagtcagc	1020
acatccacaa	ctgccatcac	ctcacatttt	actacctcag	agactgcggt	gactcccaca	1080
cctgtaaccc	catcttctct	gagtacagac	atcccgacca	caagcctacg	aactctcacc	1140
ccttcgtctg	tgggcaccag	cacttcattg	actacaacca	cagactttcc	ctctataccc	1200
actgatatca	gtaccttacc	aactcgaaca	cacatcattt	catcttctcc	ctccatccaa	1260
agtacagaaa	cctcatccct	tgtgggcacc	acctctccca	ccatgtccac	tgtgagaatg	1320
accctcagaa	ttactgagaa	cacccaatc	agttccttta	gcacaagtat	tgttggtata	1380
cctgaaaccc	caacacagac	ccctcctgta	ctgacgtcag	ccactgggac	ccaaacatct	1440
cctgcaccta	ctactgtcac	ctttggaagt	acggattcct	ccacgtccac	tcttcataag	1500
cttctct						1507

<210> 69

<211> 502

<212> PRT

5 <213> Homo sapiens

<400> 69

ES 2 535 358 T3

Ser Thr Ser Thr Pro Ser Tyr Thr Thr Ser Ile Thr Thr Thr Glu Thr
1 5 10 15

Pro Ser His Ser Thr Pro Ser Tyr Thr Thr Ser Ile Thr Thr Thr Glu
20 25 30

Thr Pro Ser His Ser Thr Pro Ser Phe Thr Ser Ser Ile Thr Thr Thr
35 40 45

Glu Thr Thr Ser His Ser Thr Pro Ser Phe Thr Ser Ser Ile Arg Thr
50 55 60

Thr Glu Thr Thr Ser Tyr Ser Thr Pro Ser Phe Thr Ser Ser Asn Thr
65 70 75 80

Ile Thr Glu Thr Thr Ser His Ser Thr Pro Ser Tyr Ile Thr Ser Ile
85 90 95

Thr Thr Thr Glu Thr Pro Ser Ser Ser Thr Pro Ser Phe Ser Ser Ser
100 105 110

Ile Thr Thr Thr Glu Thr Thr Ser His Ser Thr Pro Gly Phe Thr Ser
115 120 125

Ser Ile Thr Thr Thr Glu Thr Thr Ser His Ser Thr Pro Ser Phe Thr
130 135 140

Ser Ser Ile Thr Thr Thr Glu Thr Thr Ser His Asp Thr Pro Ser Phe
145 150 155 160

ES 2 535 358 T3

Thr Ser Ser Ile Thr Thr Ser Glu Thr Pro Ser His Ser Thr Pro Ser
 165 170 175

Ser Thr Ser Leu Ile Thr Thr Thr Lys Thr Thr Ser His Ser Thr Pro
 180 185 190

Ser Phe Thr Ser Ser Ile Thr Thr Thr Glu Thr Thr Ser His Ser Ala
 195 200 205

Arg Ser Phe Thr Ser Ser Ile Thr Thr Thr Glu Thr Thr Ser His Asn
 210 215 220

Thr Arg Ser Phe Thr Ser Ser Ile Thr Thr Thr Glu Thr Asn Ser His
 225 230 235 240

Ser Thr Thr Ser Phe Thr Ser Ser Ile Thr Thr Thr Glu Thr Thr Ser
 245 250 255

His Ser Thr Pro Ser Phe Ser Ser Ser Ile Thr Thr Thr Glu Thr Pro
 260 265 270

Leu His Ser Thr Pro Gly Leu Pro Ser Trp Val Thr Thr Thr Lys Thr
 275 280 285

Thr Ser His Ile Thr Pro Gly Leu Thr Ser Ser Ile Thr Thr Thr Glu
 290 295 300

Thr Thr Ser His Ser Thr Pro Gly Phe Thr Ser Ser Ile Thr Thr Thr
 305 310 315 320

Glu Thr Thr Ser Glu Ser Thr Pro Ser Leu Ser Ser Ser Thr Ile Tyr
 325 330 335

Ser Thr Val Ser Thr Ser Thr Thr Ala Ile Thr Ser His Phe Thr Thr
 340 345 350

Ser Glu Thr Ala Val Thr Pro Thr Pro Val Thr Pro Ser Ser Leu Ser
 355 360 365

Thr Asp Ile Pro Thr Thr Ser Leu Arg Thr Leu Thr Pro Ser Ser Val
 370 375 380

Gly Thr Ser Thr Ser Leu Thr Thr Thr Thr Asp Phe Pro Ser Ile Pro
 385 390 395 400

Thr Asp Ile Ser Thr Leu Pro Thr Arg Thr His Ile Ile Ser Ser Ser
 405 410 415

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

420	425	430
Pro Thr Met Ser Thr Val Arg Met Thr Leu Arg Ile Thr Glu Asn Thr		
435	440	445
Pro Ile Ser Ser Phe Ser Thr Ser Ile Val Val Ile Pro Glu Thr Pro		
450	455	460
Thr Gln Thr Pro Pro Val Leu Thr Ser Ala Thr Gly Thr Gln Thr Ser		
465	470	475
Pro Ala Pro Thr Thr Val Thr Phe Gly Ser Thr Asp Ser Ser Thr Ser		
485	490	495
Thr Leu His Lys Leu Leu		
500		

<210> 70

<211> 738

<212> ADN

5 <213> Homo sapiens

<400> 70

actagtacag gttctgggtca tgcaagctct accccaggtg gagaaaagga gacttcggct	60
accagagaa gttcagtgcc cagctctact gagaagaatg ctgtgagtat gaccagcagc	120
gtactctcca gccacagccc cggttcaggc tcctccacca ctcagggaca ggatgtcact	180
ctggccccgg ccacggaacc agcttcaggt tcagctgcc cctggggaca ggatgtcacc	240
tcgggtccag tcaccaggcc agccctgggc tccaccacc cgccagccca cgatgtcacc	300
tcagccccgg acaacaagcc agccccgggc tccaccgccc cccagccca cgggtgtcacc	360
tcggccccgg acaacaggcc cgccttggcg tccaccgccc ctccagtcca caatgtcacc	420
tcggcctcag gctctgcac aggtcagct tctactctgg tgcacaacgg cacctctgcc	480
agggtacca caacccagc cagcaagagc actccattct caattcccag ccaccactct	540
gatactccta ccacccttgc cagccatagc accaagactg atgccagtag cactcaccat	600
agcacggtac ctctctcac ctctccaat cacagcactt ctcccagtt gtctactggg	660
gtctctttct ttttctgtc ttttcacatt tcaaacctcc agtttaattc ctctctggaa	720
gatcccagca ccgactac	738

<210> 71

<211> 246

10 <212> PRT

<213> Homo sapiens

ES 2 535 358 T3

<400> 71

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

<210> 72

<211> 1632

<212> ADN

<213> Saccharomyces sp.

5 <400> 72

```

agaacaaccg ctatcagctc attatccgaa gtaggaacta caaccgtggg atcatccagc      60
gccattgaac catcaagtgc ctctataatc tcacctgtca cctctacact ttcgagtaca      120
acatcgtcca atccaactac tacctcccta agttcgacat ctacatctcc aagctctaca      180
tctacatctc caagctctac atctacctca tcaagttcga catctacctc atcaagttcg      240
acatctacct catcaagttc gacatctaca tctccaagtt cgacatccac atcttcaagt      300
ttgacatcca catcttcaag ttctacatct acatcccaaa gttctacatc tacctcatca      360
agttcgacat ctacatctcc aagctctaca tctacctcat caagttcaac atctacatct      420
ccaagttcta aatctacttc tgcaagctcc acttccactt cttcatattc aacatctaca      480
tccccaagtt tgacttcttc atctccaact ttgggttcca cttctccaag ttcaacatct      540
attagctcta cttttactga ttcaacttca tcccttggct cctctatagc atcttcatca      600
acgtctgtgt cattatacag cccatccaca cctgtttact ccgtcccttc gacttcgtca      660
aatgttgcaa ctctttctat gacttcttca actgttgaaa caactgtagg ttcacaaagt      720
tcgtctgaat atatcaccaa atcctcaatt tctactacta tcccatcatt ttccatgtct      780
acatatttca ccactgtagg tggagtcact acaatgtata cgacatgggt tccttatagc      840
tctgaatctg agactagcac attaaccagt atgcatgaaa cggttacaac agacgctaca      900
gtctgcactc acgagtcttg catgccctcg cagacaacaa gtttgattac atcttctata      960
aaaatgtcca ctaaaaacgt cgcaacttct gtaagcacct caacggttga atcctcatat     1020
gcatgctcca catgtgctga aacgtcacac tcgtattctt ccgtgcaaac agcttcatca     1080
agttctgtaa cacagcagac cacatccaca aagagttggg taagttcaat gacaacttcg     1140
gatgaagatt tcaataagca cgctaccggg aagtatcatg taacatcttc aggtacctca     1200
accatttcga ctagtgtgaa tgaagccacg agtacatcaa gcattgactc agaatctcaa     1260
gaacaatcat cacacttatt atcgacatcg gtcctttcat cctcctcctt gtctgctaca     1320
ttatcctctg acagtactat tttgctattc agttctgtat catcactaag tgtcgaacag     1380
tcaccagtta ccacacttca aattttcttca acatcagaga ttttacaacc cacttcttcc     1440
acagctattg ctacaatatc tgccctctaca tcatcacttt ccgcaacatc tatctctaca     1500
ccatctacct ctgtggaatc gactattgaa tcttcatcat tgactccgac ggtatcttct     1560
atcttctctc catcatcatc tgctccctct tctctacaaa catctgttac cactacagaa     1620
gtttccacta ct                                     1632

```

<210> 73

<211> 544

<212> PRT

5 <213> Saccharomyces sp.

ES 2 535 358 T3

<400> 73

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

ES 2 535 358 T3

Phe Ser Met Ser Thr Tyr Phe Thr Thr Val Ser Gly Val Thr Thr Met
 260 265 270
 Tyr Thr Thr Trp Cys Pro Tyr Ser Ser Glu Ser Glu Thr Ser Thr Leu
 275 280 285
 Thr Ser Met His Glu Thr Val Thr Thr Asp Ala Thr Val Cys Thr His
 290 295 300
 Glu Ser Cys Met Pro Ser Gln Thr Thr Ser Leu Ile Thr Ser Ser Ile
 305 310 315 320
 Lys Met Ser Thr Lys Asn Val Ala Thr Ser Val Ser Thr Ser Thr Val
 325 330 335
 Glu Ser Ser Tyr Ala Cys Ser Thr Cys Ala Glu Thr Ser His Ser Tyr
 340 345 350
 Ser Ser Val Gln Thr Ala Ser Ser Ser Ser Val Thr Gln Gln Thr Thr
 355 360 365
 Ser Thr Lys Ser Trp Val Ser Ser Met Thr Thr Ser Asp Glu Asp Phe
 370 375 380
 Asn Lys His Ala Thr Gly Lys Tyr His Val Thr Ser Ser Gly Thr Ser
 385 390 395 400
 Thr Ile Ser Thr Ser Val Ser Glu Ala Thr Ser Thr Ser Ser Ile Asp
 405 410 415
 Ser Glu Ser Gln Glu Gln Ser Ser His Leu Leu Ser Thr Ser Val Leu
 420 425 430
 Ser Ser Ser Ser Leu Ser Ala Thr Leu Ser Ser Asp Ser Thr Ile Leu
 435 440 445
 Leu Phe Ser Ser Val Ser Ser Leu Ser Val Glu Gln Ser Pro Val Thr
 450 455 460
 Thr Leu Gln Ile Ser Ser Thr Ser Glu Ile Leu Gln Pro Thr Ser Ser
 465 470 475 480
 Thr Ala Ile Ala Thr Ile Ser Ala Ser Thr Ser Ser Leu Ser Ala Thr
 485 490 495
 Ser Ile Ser Thr Pro Ser Thr Ser Val Glu Ser Thr Ile Glu Ser Ser
 500 505 510

ES 2 535 358 T3

Ser Leu Thr Pro Thr Val Ser Ser Ile Phe Leu Ser Ser Ser Ser Ala
515 520 525

Pro Ser Ser Leu Gln Thr Ser Val Thr Thr Thr Glu Val Ser Thr Thr
530 535 540

<210> 74

<211> 1629

<212> ADN

5 <213> Saccharomyces sp.

<400> 74

acaaccgcta tcagctcatt atccgaagta ggaactacaa ccgtggtatc atccagcgcc	60
attgaaccat caagtgcctc tataatctca cctgtcacct ctacactttc gagtacaaca	120
tcgtccaatc caactactac ctccctaagt tcgacatcta catctccaag ctctacatct	180
acatctccaa gctctacatc tacctcatca agttcgacat ctacctcatc aagttcgaca	240
tctacctcat caagttcgac atctacatct ccaagttcga catccacatc ttcaagtttg	300
acatccacat cttcaagttc tacatctaca tcccaaagtt ctacatctac ctcatcaagt	360
tcgacatcta catctccaag ctctacatct acctcatcaa gttcaacatc tacatctcca	420
agttctaaat ctactttctgc aagctccact tccactttct catattcaac atctacatcc	480
ccaagtttga cttcttcac cccaactttg gcttccactt ctccaagttc aacatctatt	540
agctctactt ttactgattc aacttcatcc cttggctcct ctatagcatc ttcatcaacg	600
tctgtgtcat tatacagccc atccacacct gtttactccg tcccttcgac ttcgtcaaat	660
gttgcaactc cttctatgac ttcttcaact gttgaaacaa ctggttagttc acaaagttcg	720
tctgaatata tcaccaaate ctcaatttct actactatcc catcattttc catgtctaca	780
tatttcacca ctggttagtgg agtcaactaca atgtatacga catggtgtcc ttatagctct	840
gaatctgaga ctagcacatt aaccagtatg catgaaacgg ttacaacaga cgctacagtc	900
tgacttcacg agtcttgcac gccctcgcag acaacaagtt tgattacatc ttctataaaa	960
atgtccacta aaaacgtcgc aacttctgta agcacctcaa cggttgaatc ctcatatgca	1020
tgctccacat gtgctgaaac gtcacactcg tattcttccg tgcaaacagc ttcatcaagt	1080
tctgtaacac agcagaccac atccacaaag agttgggtaa gttcaatgac aacttcggat	1140
gaagattttca ataagcacgc taccggtaag tatcatgtaa catcttcagg tacctcaacc	1200
atttcgacta gtgtaagtga agccacgagt acatcaagca ttgactcaga atctcaagaa	1260
caatcatcac acttattatc gacatcggtc ctttcatcct cctccttgtc tgctacatta	1320
tcctctgaca gtactatttt gctattcagt tctgtatcat cactaagtgt cgaacagtca	1380
ccagttacca cacttcaa atcttcaaca tcagagattt tacaaccac ttcttccaca	1440
gctattgcta caatatctgc ctctacatca tcactttccg caacatctat ctctacacca	1500
tctacctctg tggaatcgac tattgaatct tcatcattga ctccgacggg atcttctatt	1560
ttcctctcat catcatctgc tccctcttct ctacaaacat ctgttaccac tacagaagtt	1620
tccactact	1629

<210> 75

<211> 543

<212> PRT

5 <213> Saccharomyces sp.

ES 2 535 358 T3

<400> 75

Thr Thr Ala Ile Ser Ser Leu Ser Glu Val Gly Thr Thr Thr Val Val
1 5 10 15

Ser Ser Ser Ala Ile Glu Pro Ser Ser Ala Ser Ile Ile Ser Pro Val
20 25 30

Thr Ser Thr Leu Ser Ser Thr Thr Ser Ser Asn Pro Thr Thr Thr Ser
35 40 45

Leu Ser Ser Thr Ser Thr Ser Pro Ser Ser Thr Ser Thr Ser Pro Ser
50 55 60

Ser Thr Ser Thr Ser Ser Ser Ser Thr Ser Thr Ser Ser Ser Ser Thr
65 70 75 80

Ser Thr Ser Ser Ser Ser Thr Ser Thr Ser Pro Ser Ser Thr Ser Thr
85 90 95

Ser Ser Ser Leu Thr Ser Thr Ser Ser Ser Ser Thr Ser Thr Ser Gln
100 105 110

Ser Ser Thr Ser Thr Ser Ser Ser Ser Thr Ser Thr Ser Pro Ser Ser
115 120 125

Thr Ser Thr Ser Ser Ser Ser Thr Ser Thr Ser Pro Ser Ser Lys Ser
130 135 140

Thr Ser Ala Ser Ser Thr Ser Thr Ser Ser Tyr Ser Thr Ser Thr Ser
145 150 155 160

Pro Ser Leu Thr Ser Ser Ser Pro Thr Leu Ala Ser Thr Ser Pro Ser
165 170 175

Ser Thr Ser Ile Ser Ser Thr Phe Thr Asp Ser Thr Ser Ser Leu Gly
180 185 190

Ser Ser Ile Ala Ser Ser Ser Thr Ser Val Ser Leu Tyr Ser Pro Ser
195 200 205

ES 2 535 358 T3

Thr Pro Val Tyr Ser Val Pro Ser Thr Ser Ser Asn Val Ala Thr Pro
 210 215 220
 Ser Met Thr Ser Ser Thr Val Glu Thr Thr Val Ser Ser Gln Ser Ser
 225 230 235 240
 Ser Glu Tyr Ile Thr Lys Ser Ser Ile Ser Thr Thr Ile Pro Ser Phe
 245 250 255
 Ser Met Ser Thr Tyr Phe Thr Thr Val Ser Gly Val Thr Thr Met Tyr
 260 265 270
 Thr Thr Trp Cys Pro Tyr Ser Ser Glu Ser Glu Thr Ser Thr Leu Thr
 275 280 285
 Ser Met His Glu Thr Val Thr Thr Asp Ala Thr Val Cys Thr His Glu
 290 295 300
 Ser Cys Met Pro Ser Gln Thr Thr Ser Leu Ile Thr Ser Ser Ile Lys
 305 310 315 320
 Met Ser Thr Lys Asn Val Ala Thr Ser Val Ser Thr Ser Thr Val Glu
 325 330 335
 Ser Ser Tyr Ala Cys Ser Thr Cys Ala Glu Thr Ser His Ser Tyr Ser
 340 345 350
 Ser Val Gln Thr Ala Ser Ser Ser Ser Val Thr Gln Gln Thr Thr Ser
 355 360 365
 Thr Lys Ser Trp Val Ser Ser Met Thr Thr Ser Asp Glu Asp Phe Asn
 370 375 380
 Lys His Ala Thr Gly Lys Tyr His Val Thr Ser Ser Gly Thr Ser Thr
 385 390 395 400
 Ile Ser Thr Ser Val Ser Glu Ala Thr Ser Thr Ser Ser Ile Asp Ser
 405 410 415
 Glu Ser Gln Glu Gln Ser Ser His Leu Leu Ser Thr Ser Val Leu Ser
 420 425 430
 Ser Ser Ser Leu Ser Ala Thr Leu Ser Ser Asp Ser Thr Ile Leu Leu
 435 440 445
 Phe Ser Ser Val Ser Ser Leu Ser Val Glu Gln Ser Pro Val Thr Thr
 450 455 460
 Leu Gln Ile Ser Ser Thr Ser Glu Ile Leu Gln Pro Thr Ser Ser Thr

ES 2 535 358 T3

465		470		475		480									
Ala	Ile	Ala	Thr	Ile	Ser	Ala	Ser	Thr	Ser	Ser	Leu	Ser	Ala	Thr	Ser
				485				490						495	
Ile	Ser	Thr	Pro	Ser	Thr	Ser	Val	Glu	Ser	Thr	Ile	Glu	Ser	Ser	Ser
			500					505					510		
Leu	Thr	Pro	Thr	Val	Ser	Ser	Ile	Phe	Leu	Ser	Ser	Ser	Ser	Ala	Pro
		515					520					525			
Ser	Ser	Leu	Gln	Thr	Ser	Val	Thr	Thr	Thr	Glu	Val	Ser	Thr	Thr	
	530					535					540				

<210> 75

<211> 1506

<212> ADN

5 <213> Secuencia artificial

<220>

<213> Gen de la MUC3 humana modificado

<400> 76

tctactagta caccaagcta tactacatca attactagca ctgagacacc atctcattca	60
actccttcta gttcgacttc aattacaact acagagacac ccagtcacag caccatctct	120
tacactagct ctgtctccac atccgagact acatcacatt ctactccatc agaaacaagc	180
tccagtagaa caacagaaaag cacctcttat agttcaccta gctccacttc atctaacaca	240
attactgaga caagctcaca ctccactcct agtactgcta cttctatttc ctcgaccgaa	300
acacctagtt caagtacacc atctgtatca tcgtccatta ctgttactga gagttcatct	360
catagcactc ctggagctac ttccactttg acatcgagtg aaacttctac ttggtcaaca	420
ccatctagca caagttctat tatgtcaagc tcctacactt cagctgacac tccatctgaa	480
acatcagttt atacttccag cgaaacccca tcgtcctcaa gcccaactag cacatctttg	540
atttctagtt cgaagtcaac atcgaccagt acaccttcgt ttacttcttc gattactagc	600
actgagacct cctcatattc tgctagttcc tatacacctt cagttagtag cacagcaagt	660
tctagcaaga acacaacgag ttccactgct tctataagca gtacagagac tgttagttca	720
tcgactagct ctgtctctag tactattcct tcttctcaat ccacaagtta ttctacacca	780
tcattctcca gttcggcaac aagcagtgtt actccattgc attcaacacc atctctacca	840
tcttgggtta ctacaagtaa gaccacatca catattacac caggtctgac ttcgtccatg	900
tcttcgagcg agacctatag ccatagtact ccaggtttta caagctctat tacttcgaca	960
gaatcgacaa gtgagtcaac tccatcattg tccagttcta caatttatag tactgtttca	1020
acatctacta cagctattac ttcacatttt acaacttctg agacagctgt tactccaaca	1080
ccagttacac cctctagctt gagtacagat atcccaacta caagtttgag aactttgact	1140
ccttcctctg ttggtacctc gacttccttg acaactacaa ctgactttcc atcaattcca	1200
acagacatta gcactttgcc aacaagaact catattattt caagctcacc atctattcaa	1260
tcaacagaga ctagcagttt ggttggtact acatctccaa ctatgtcaac agttagaatg	1320
actttgagaa ttacagagaa cactccaatt tcttcattca gtacttccat tgttggtatt	1380
ccagagacac caactcaaac accaccagtt ttgacttctg ctacaggtac tcaaacatca	1440
ccagctccaa ctacagttac ttttggttct actgactcat ctacatcaac tttgcataag	1500
cttttg	1506

REIVINDICACIONES

1. Una proteína de fusión que comprende una proteína MUC3 humana modificada codificada por el ácido nucleico que comprende la SEC ID NO: 76.
2. La proteína de fusión de la reivindicación 1, que comprende adicionalmente un resto de anclaje.
- 5 3. La proteína de fusión de la reivindicación 2, en la que la proteína de anclaje es un glucosil-fosfatidil-inositol (GPI).
4. Una célula huésped que comprende el ácido nucleico que comprende la SEC ID NO: 76.
5. La célula huésped de la reivindicación 4, en la que la célula huésped es una célula huésped de levadura.
6. Una biblioteca que comprende una pluralidad de proteínas de fusión de acuerdo con una cualquiera de las reivindicaciones 1 a 3.

Presentación de levaduras y selección de transferrinas de unión a dianas

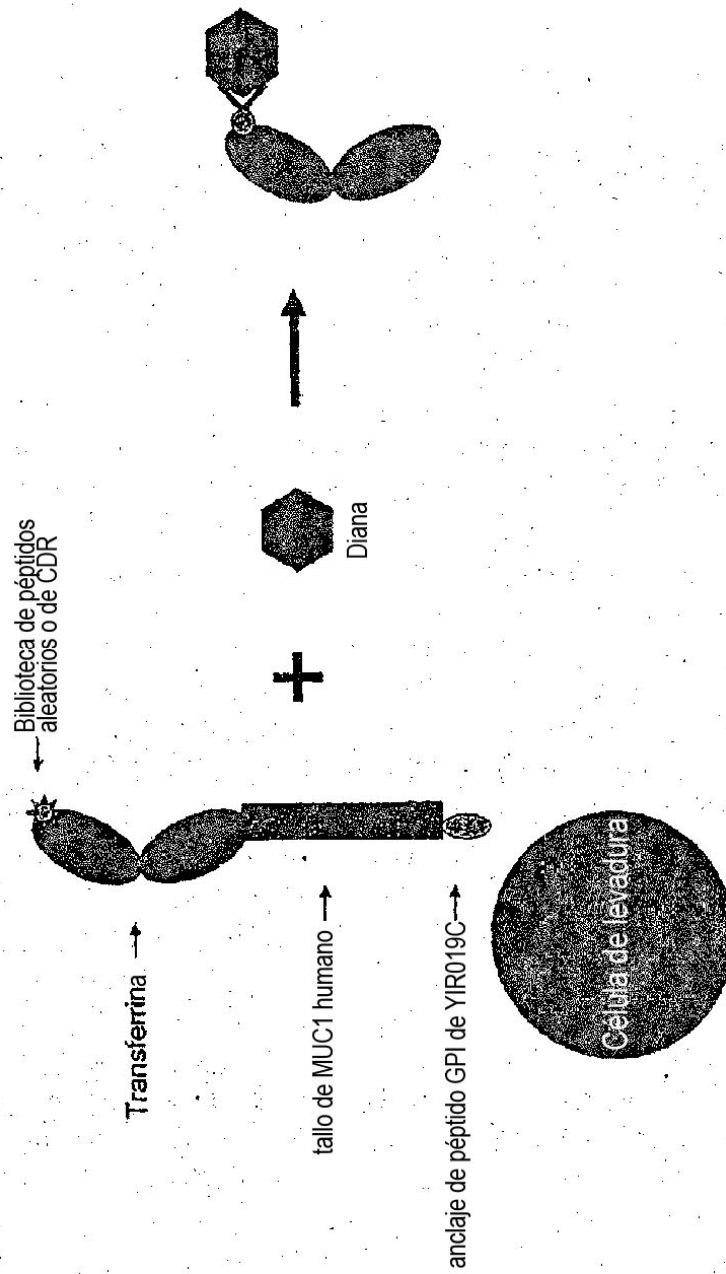
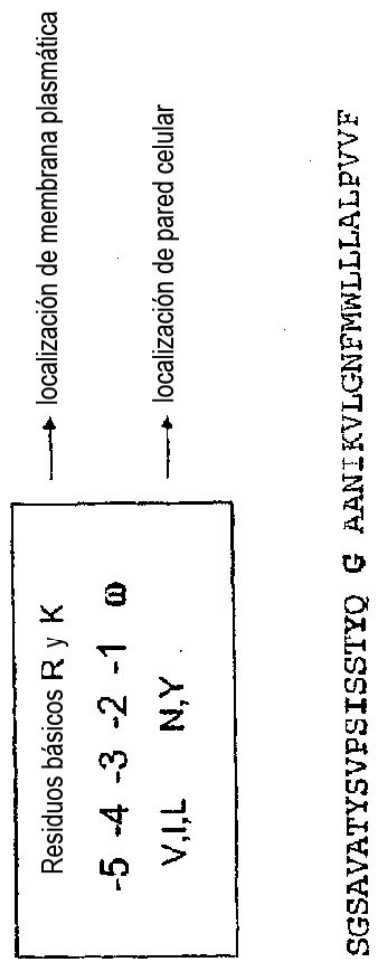


Figura 1

anclaje de GPI



Levadura YIR019C
GPI

Figura 2

FIG. 3

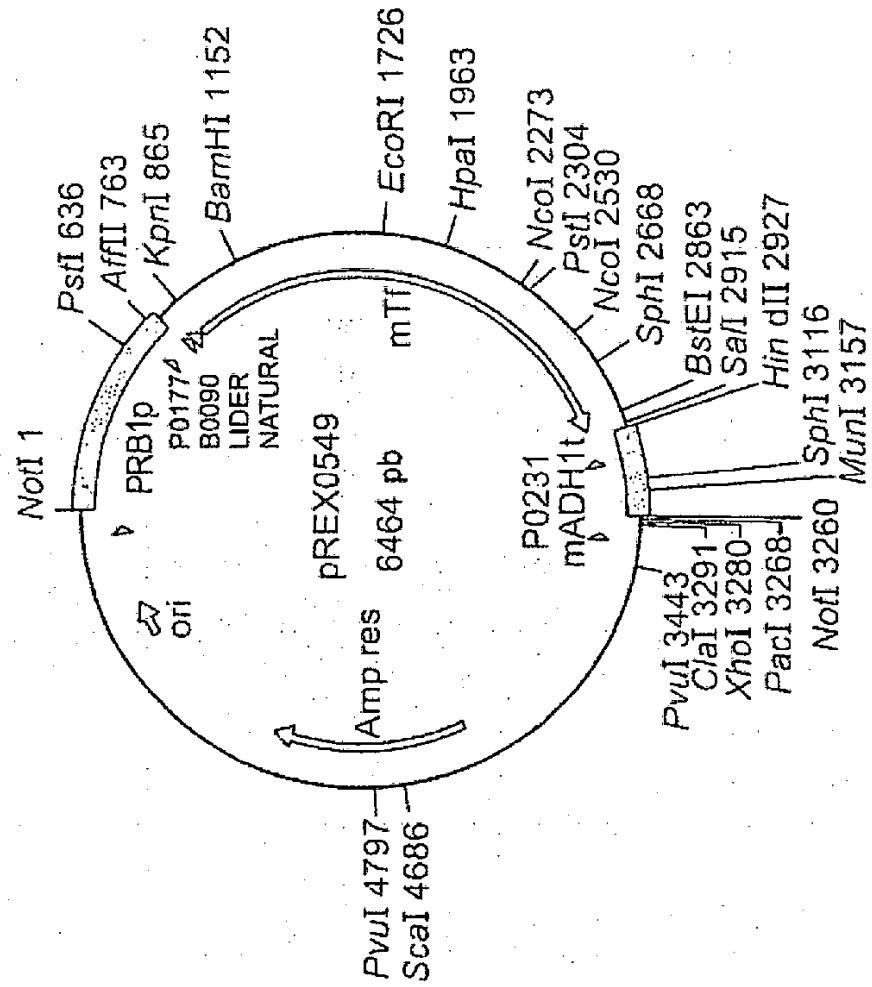


FIG. 4

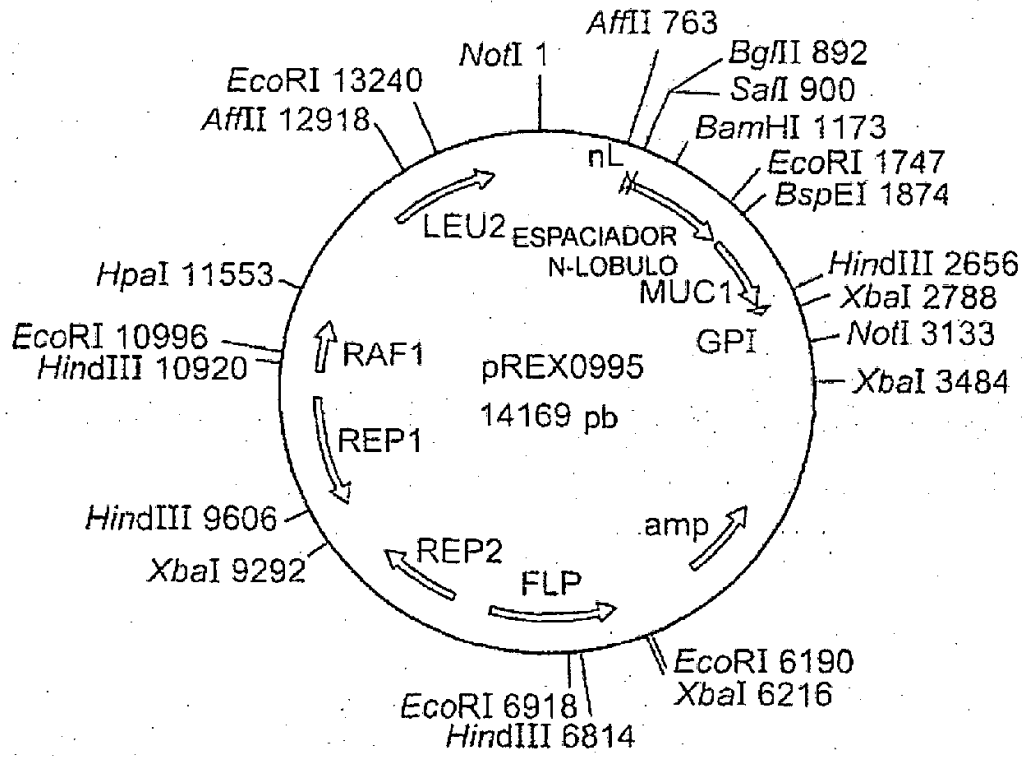


FIG. 5

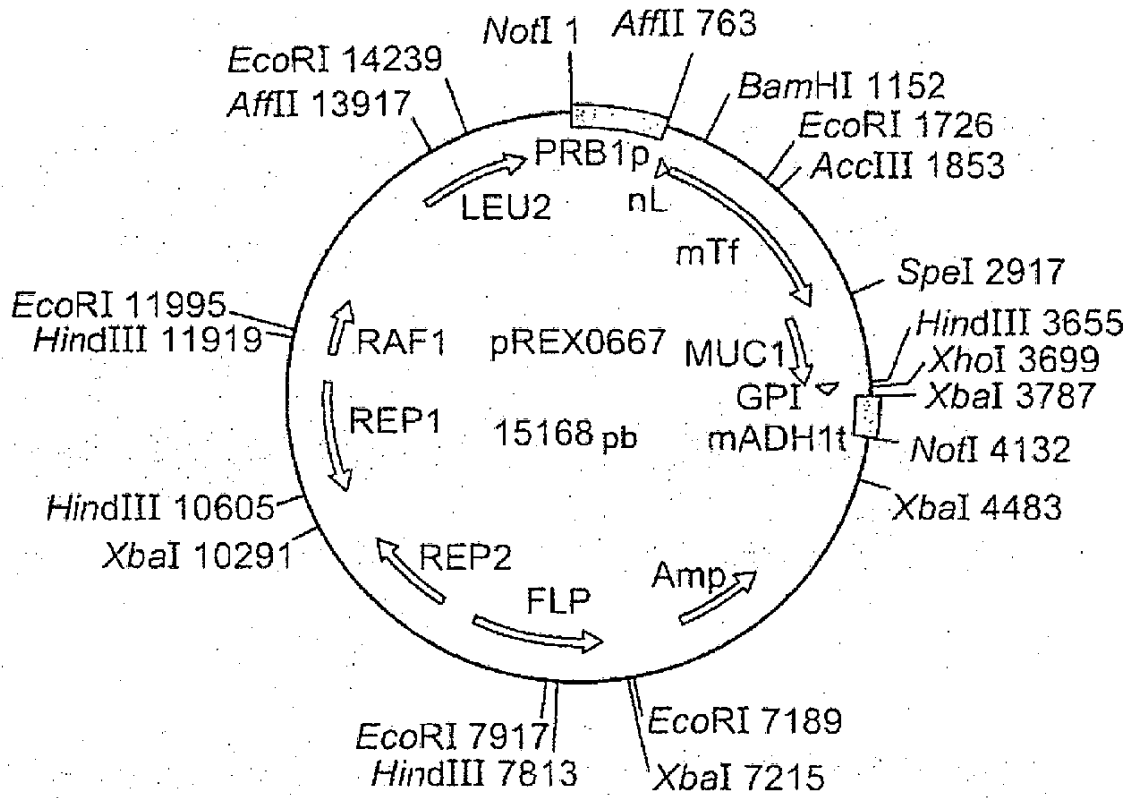


FIG. 6

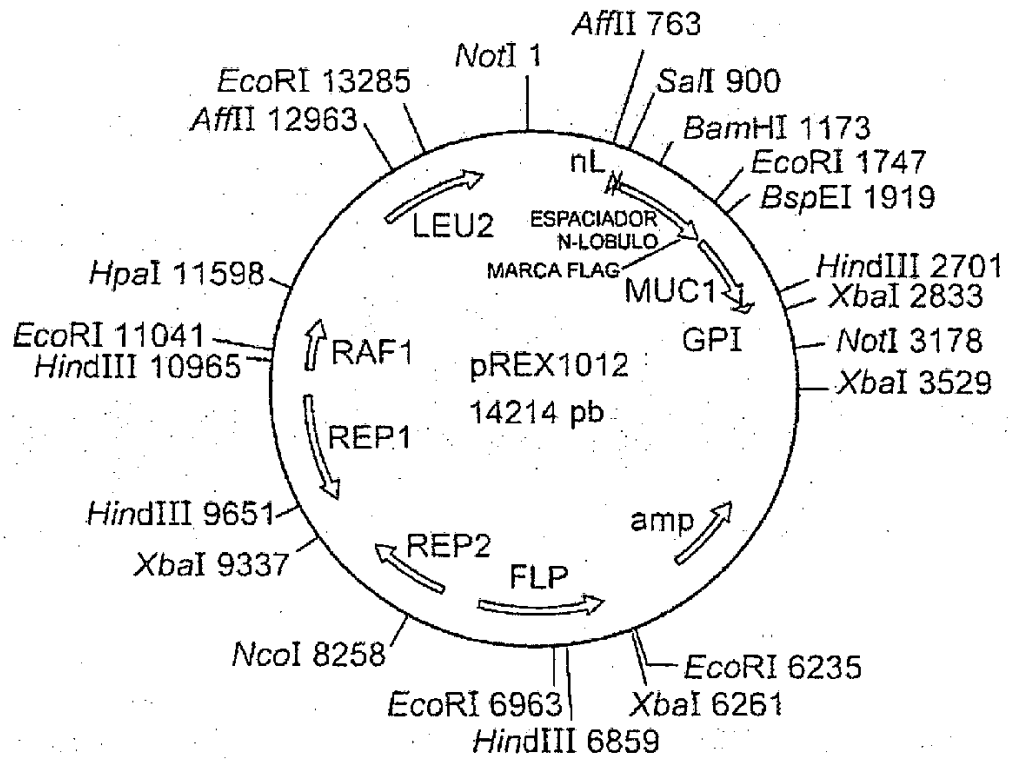
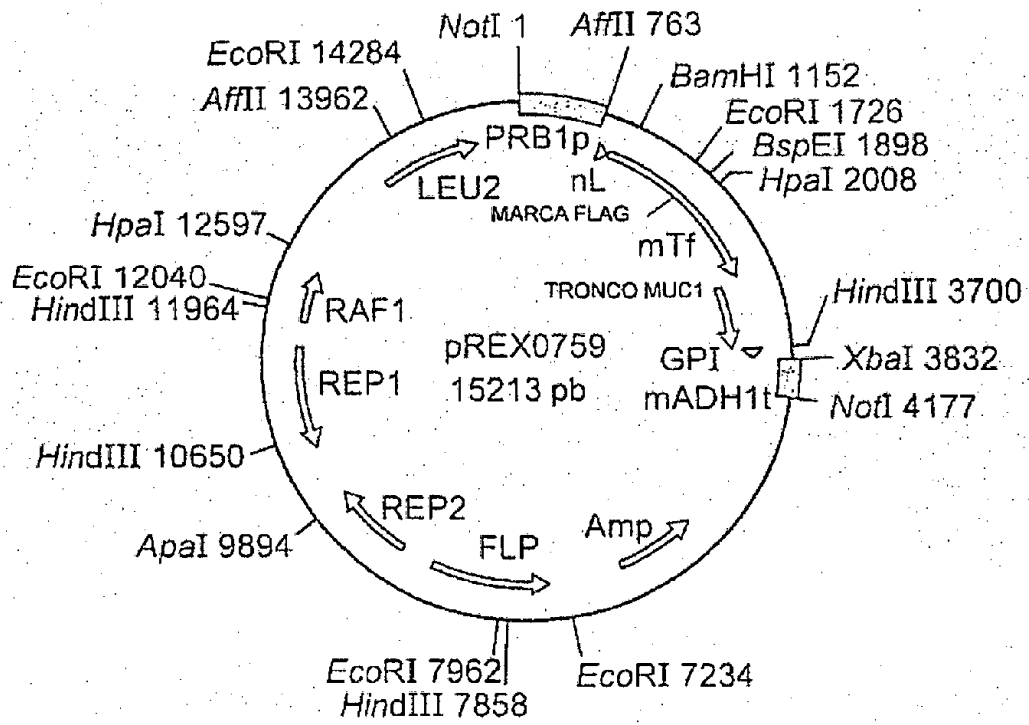


FIG. 7



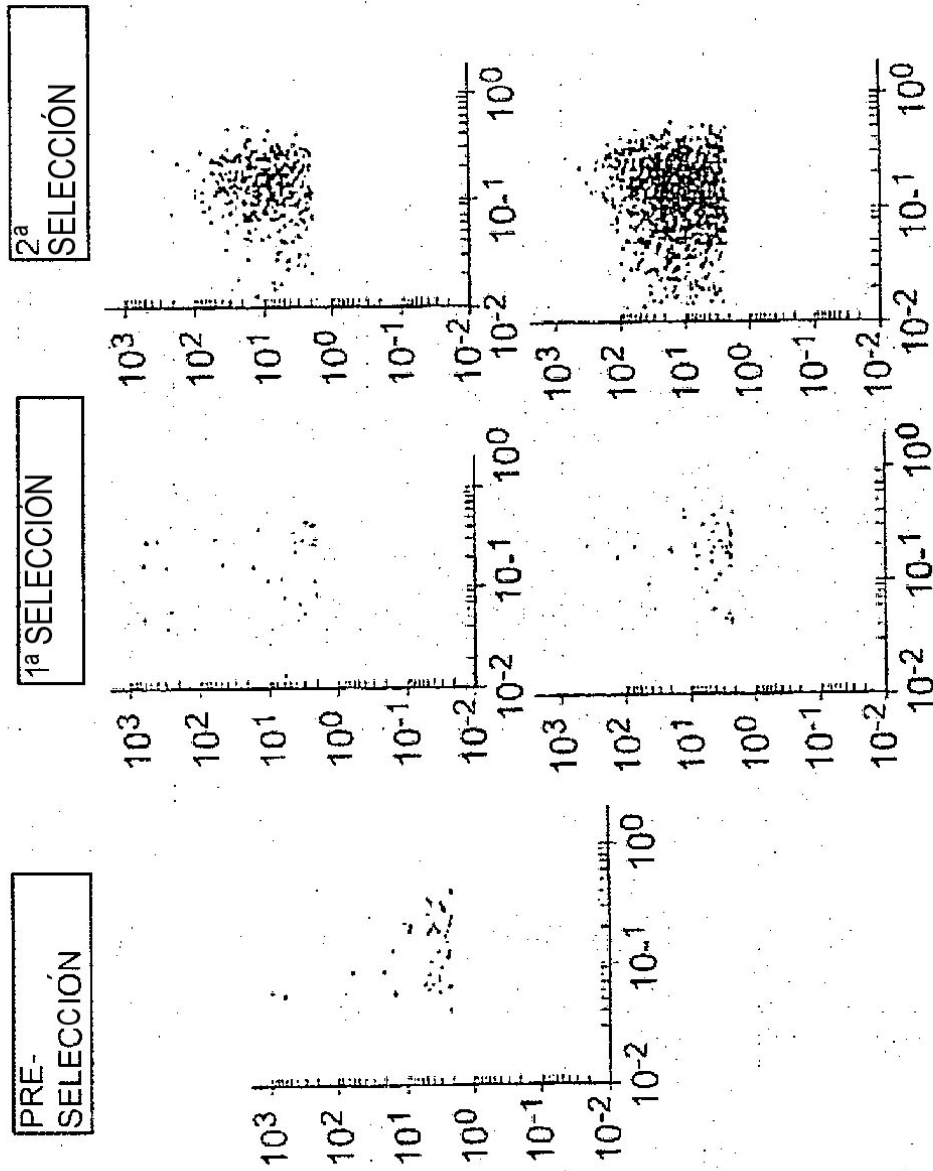


FIG. 9

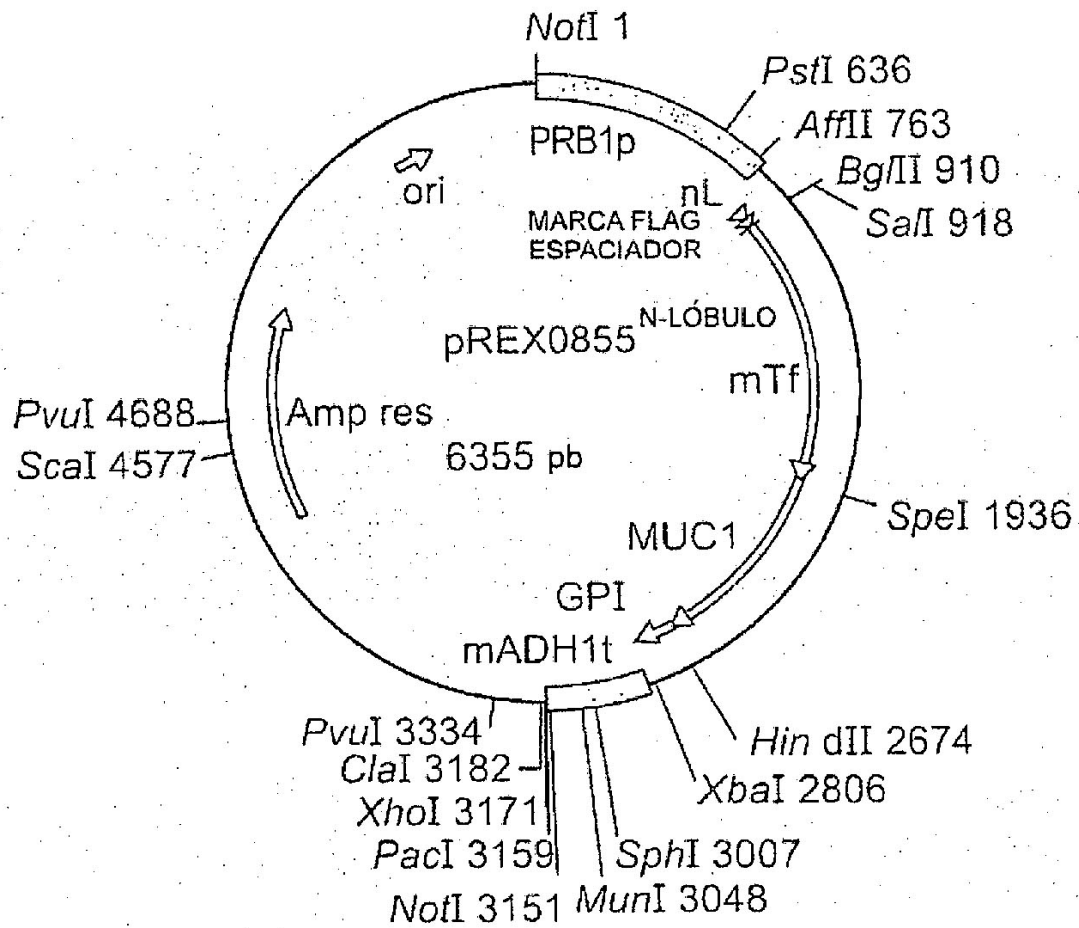


FIG. 10

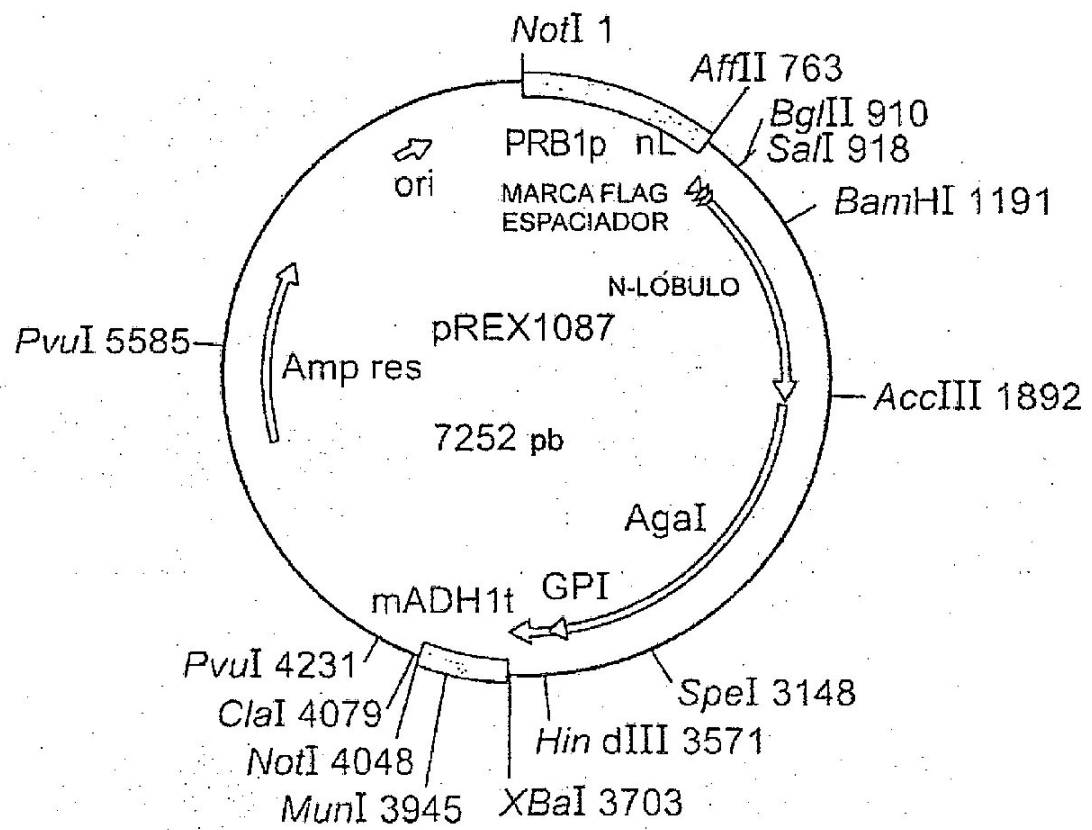
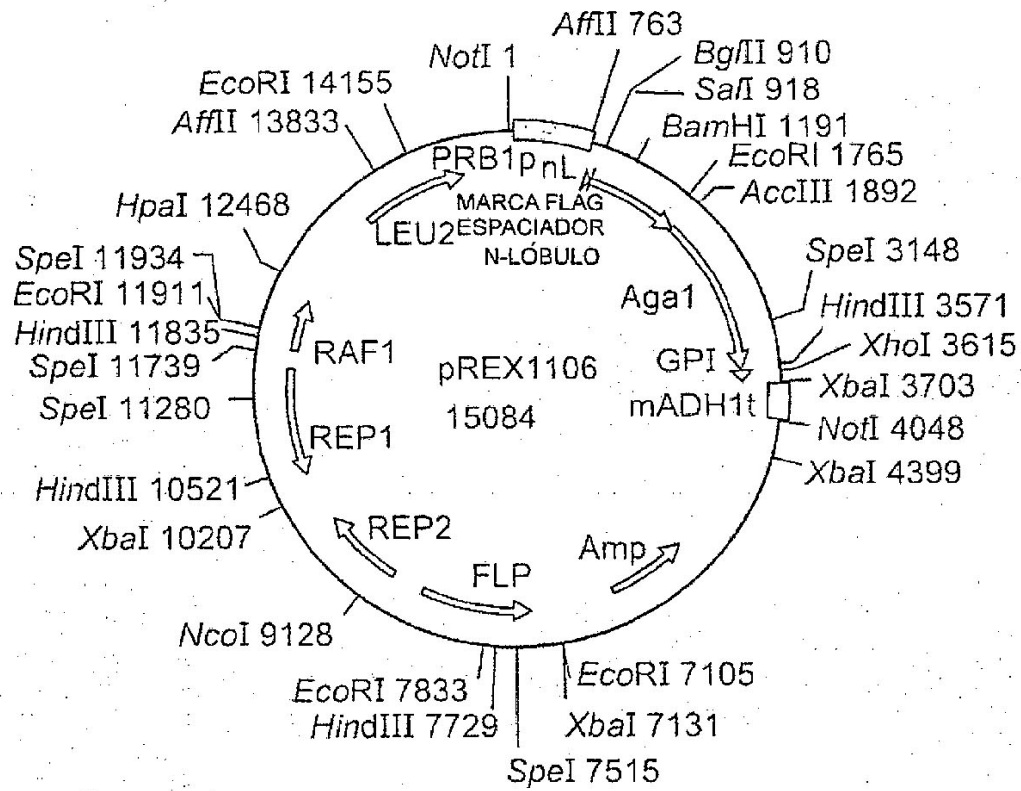
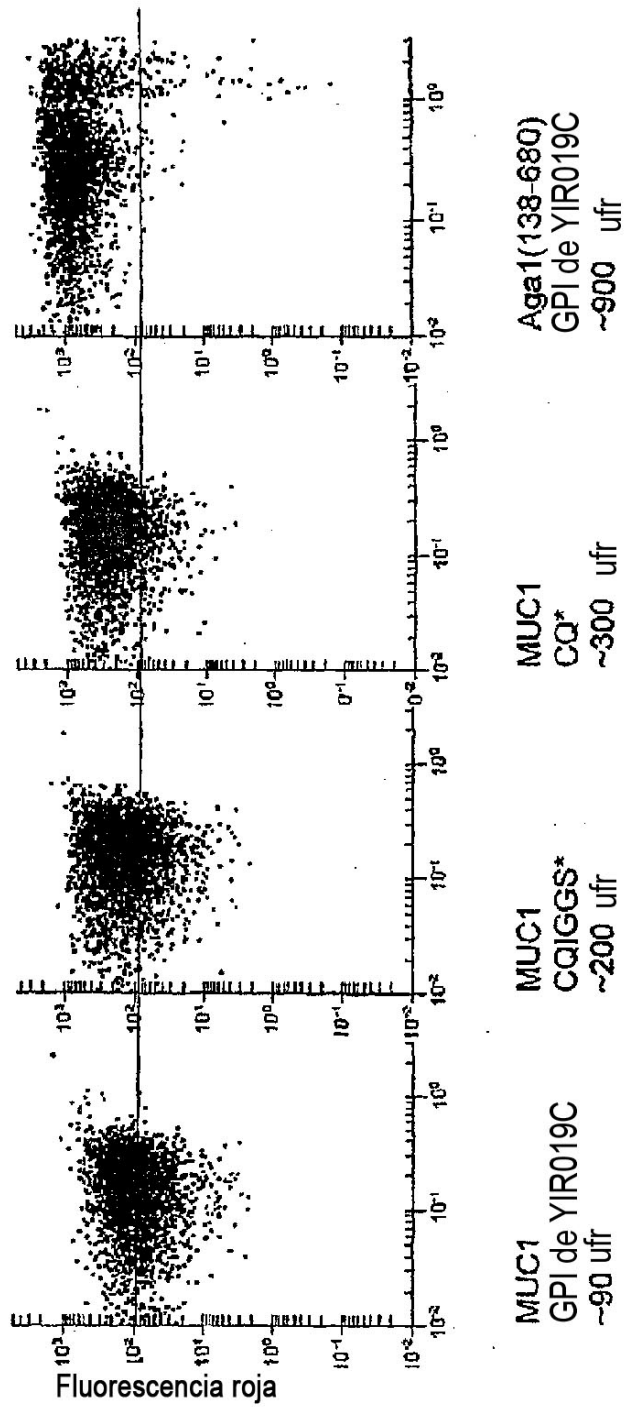


FIG. 11





El tallo MUC1 corresponde a residuos 22-325 de huMUC1
Esta versión es de varias repeticiones menos y es de 243 aa más que de 303 aa.

Figura 12

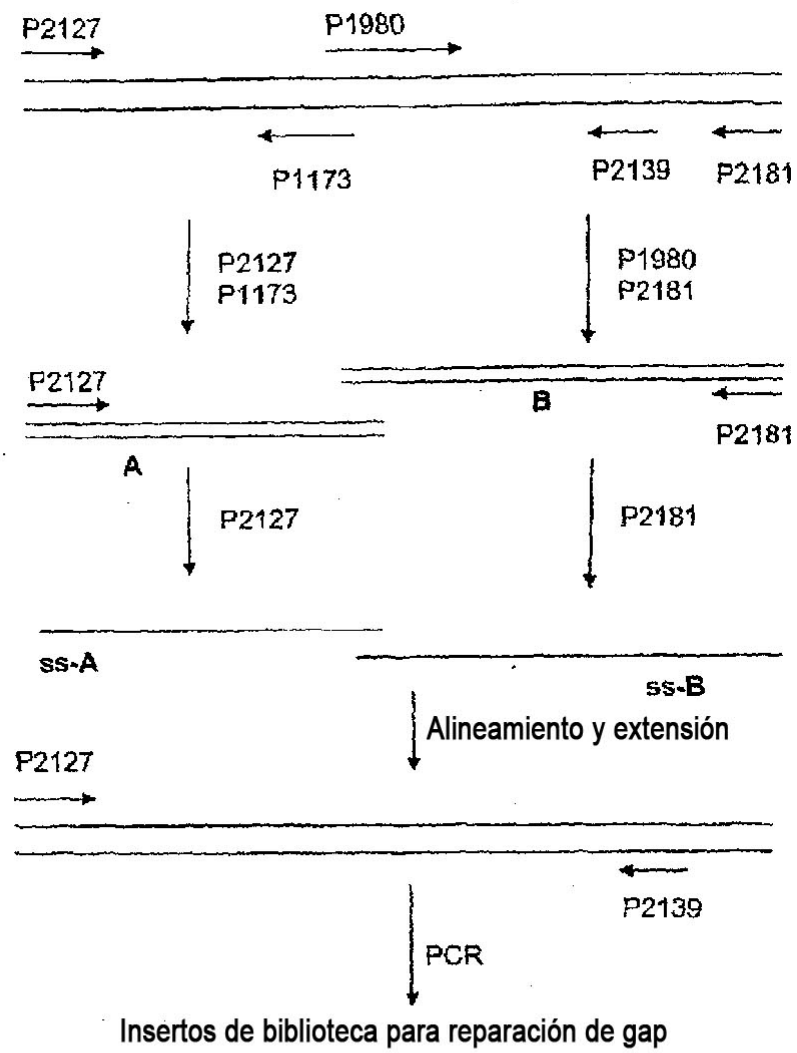
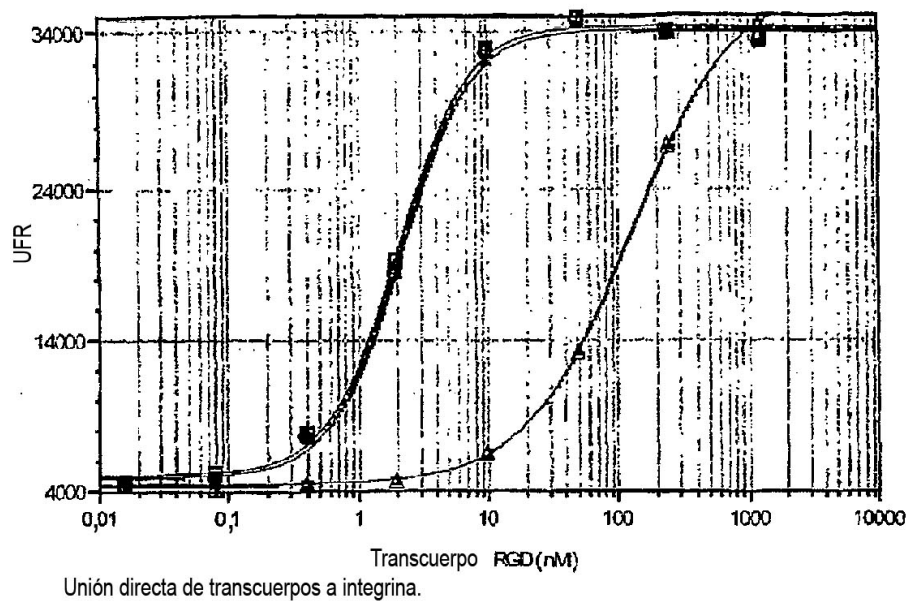


Figura 13

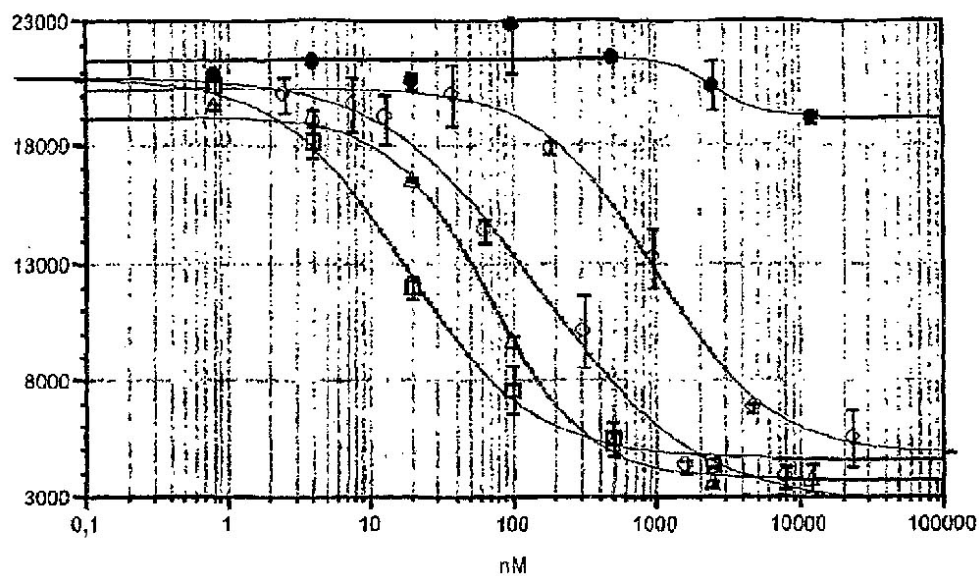


n.º: 14

n.º: 41

F531

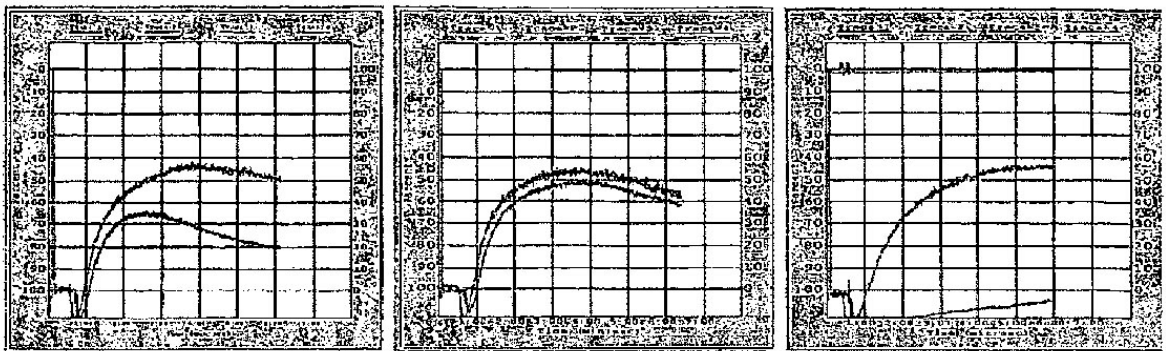
Figura 14



Análisis de competición de unión de *Trans-cuerpo*.

- péptido GRGDSP
- n.º 32
- △ n.º 14
- ◇ F531
- Tf

Figura 15



El clon de trans-cuerpo de RGD n.º: 32 inhibe agregación plaquetaria en ratones

- A.** 25 μ M de péptido GRGDSD + 10 μ M de ADP (Azul) frente a ADP solo: inhibición del 37 %
- B.** 25 μ M de Tf + ADP (Azul) frente a 10 μ M de ADP: sin ninguna inhibición
- C.** 25 μ M de n.º: 32 + ADP (Azul) frente a ADP solo: inhibición del 100 %

Figura 16

FIG. 17

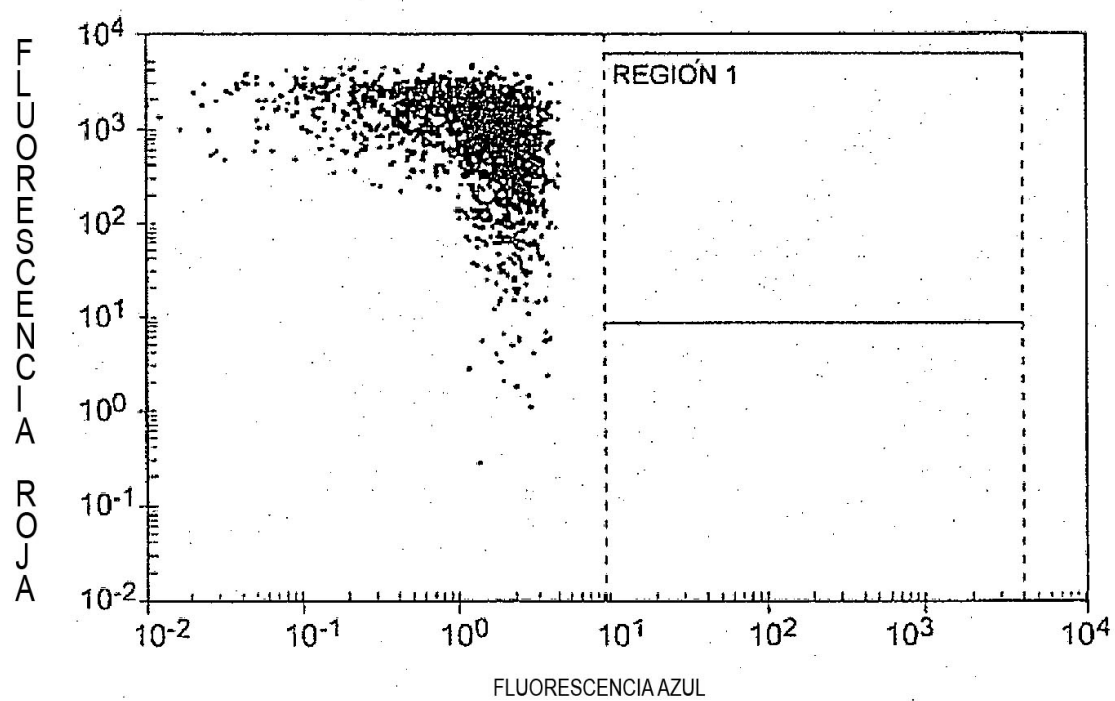


FIG. 18

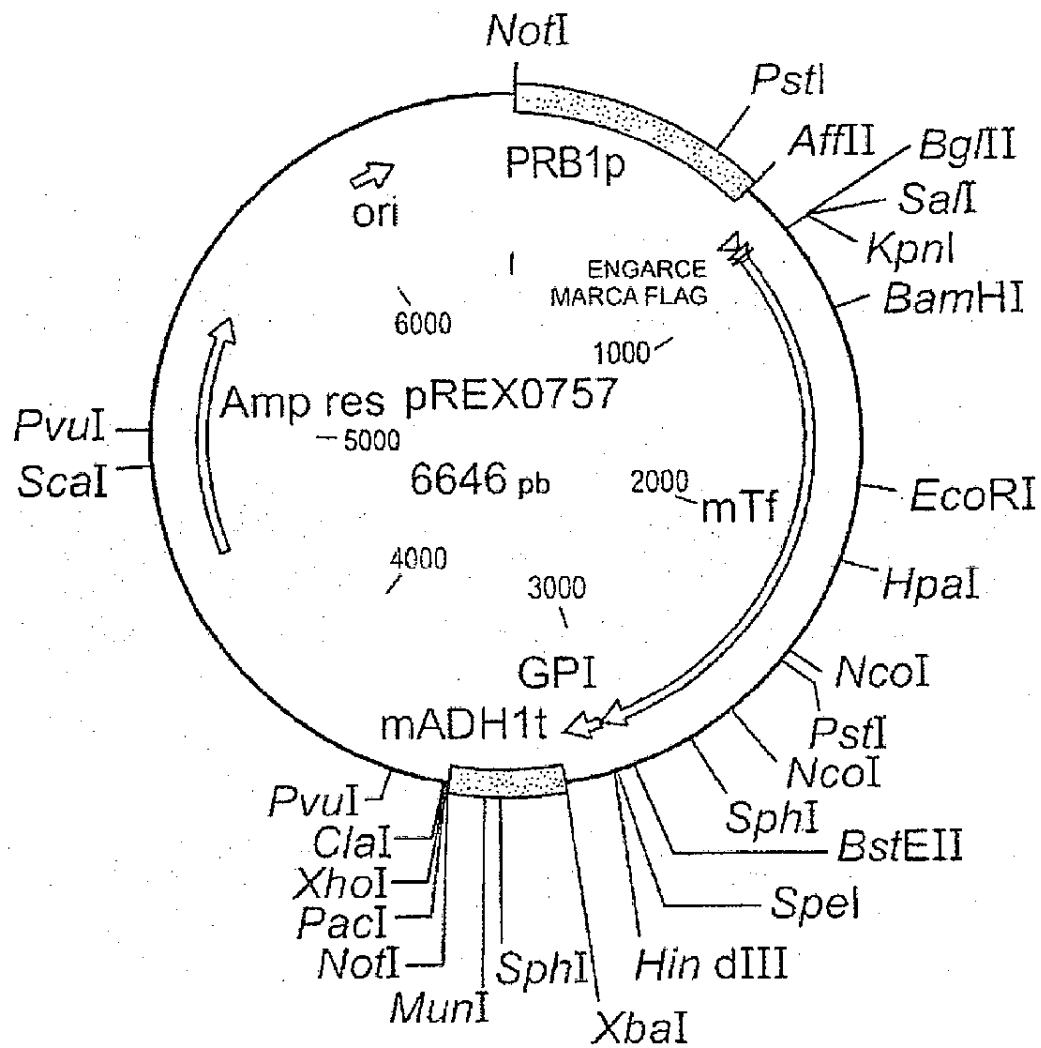


FIG. 19

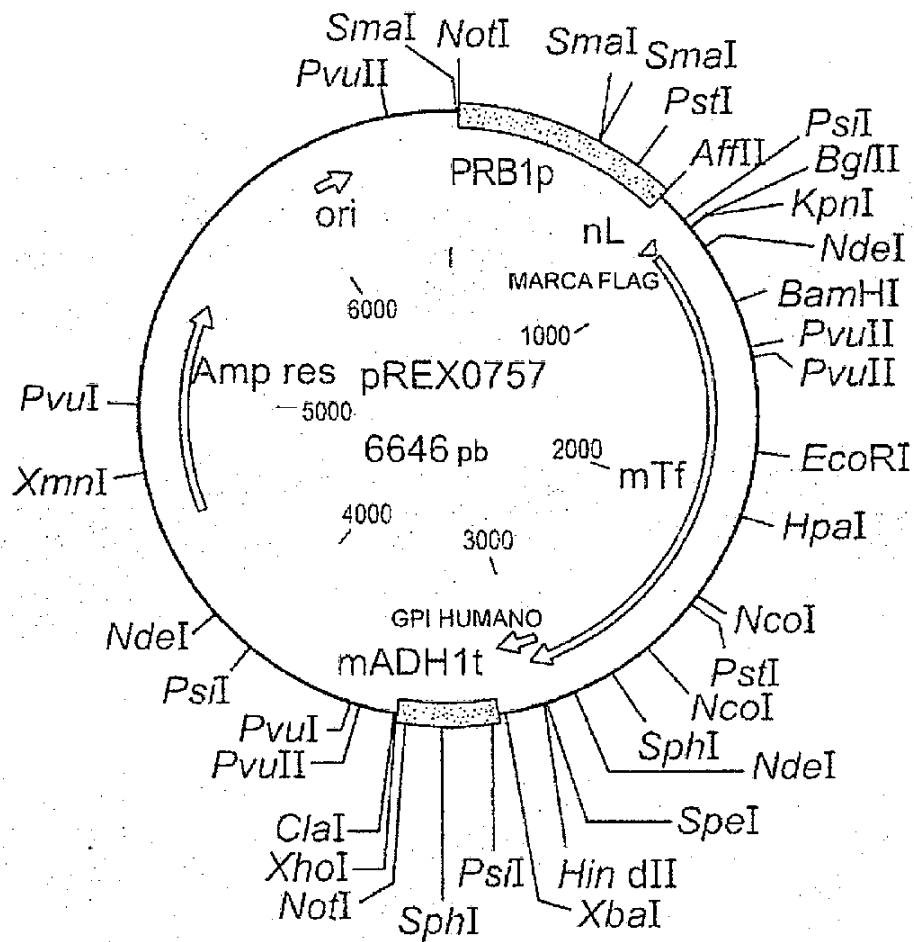


FIG. 20

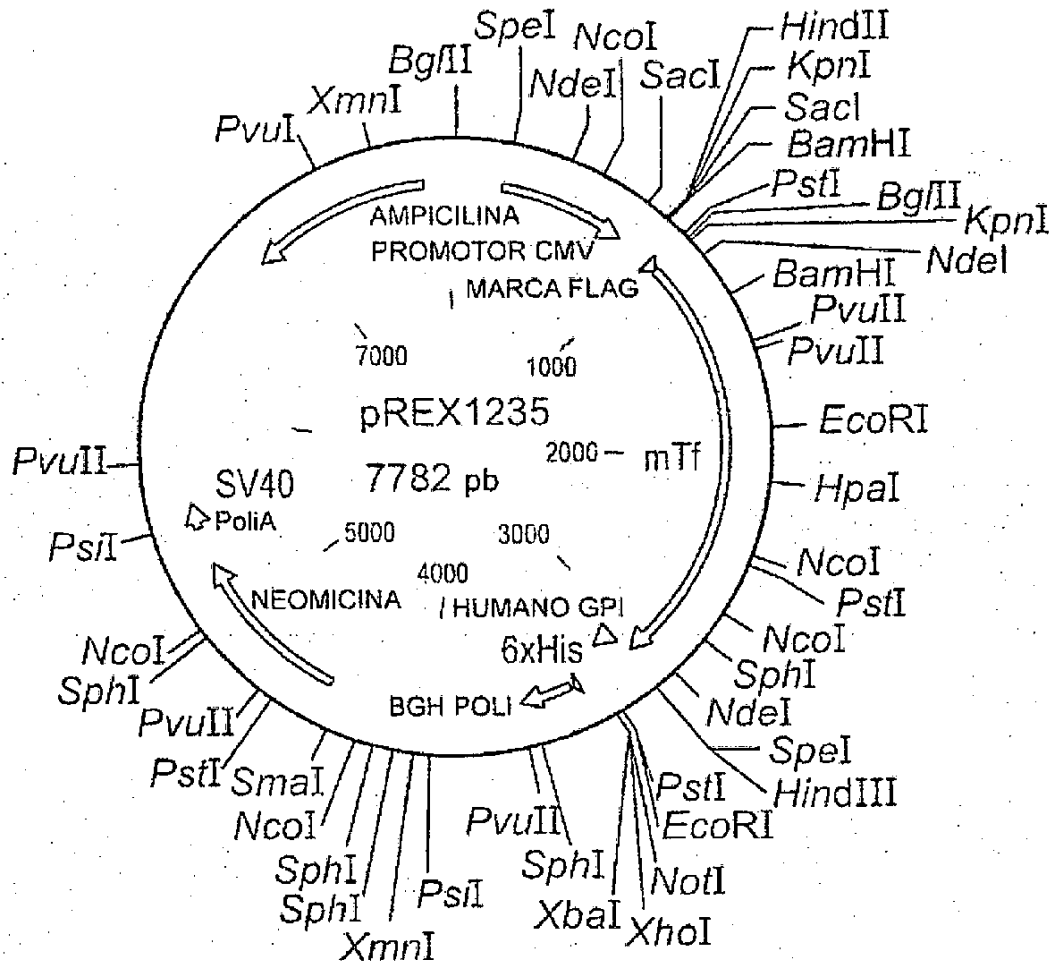


FIG. 21

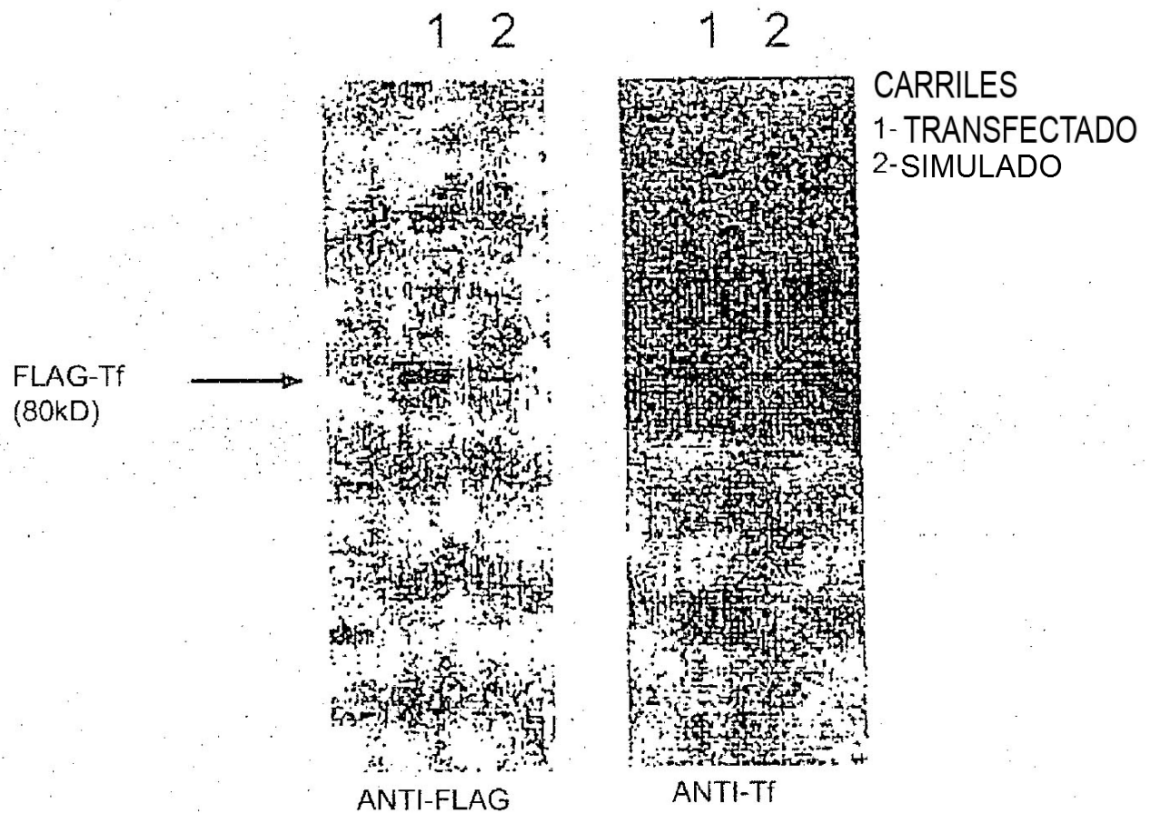


FIG. 22

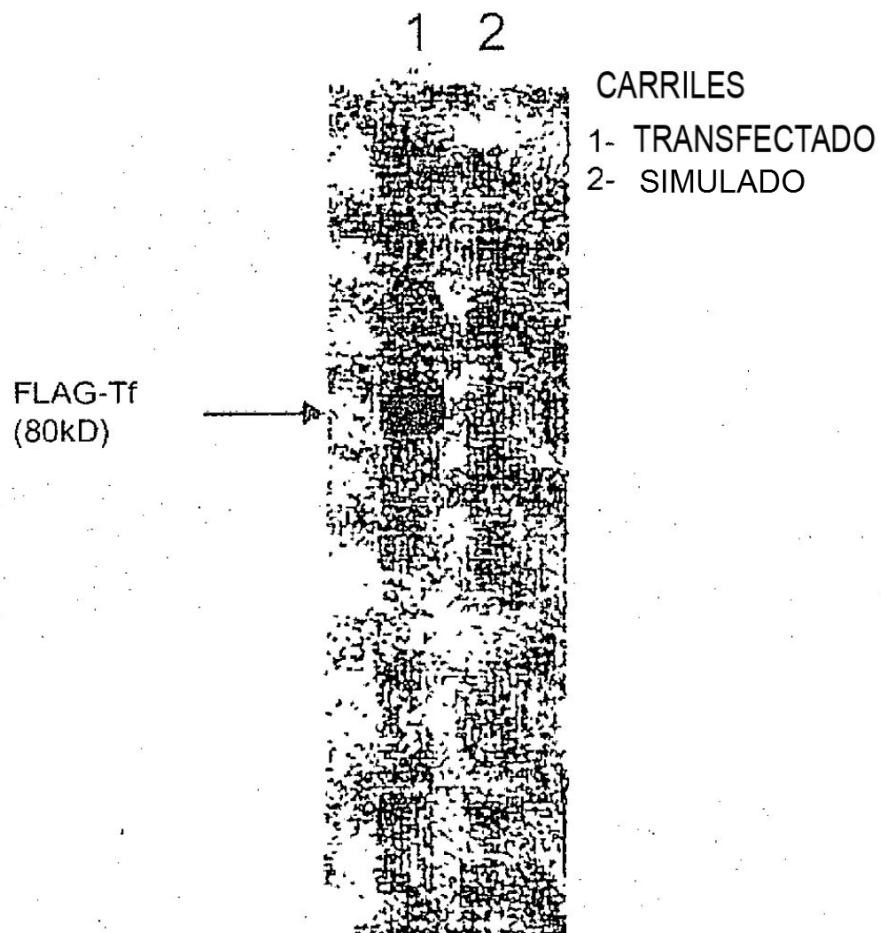


FIG. 23

