

## MOLÉCULAS DE ÁCIDO NUCLEICO Y SUS USOS

### REFERENCIA A LA LISTA DE SECUENCIAS PRESENTADA

#### ELECTRÓNICAMENTE

- 5    **[0001]**El contenido de la lista de secuencias presentada electrónicamente en el archivo de texto ASCII (Nombre: 4159\_493PC01\_ST25; Tamaño: 434.561 bytes; y Fecha de creación: 9 de agosto de 2018) se incorpora en este documento por referencia en su totalidad.

#### ANTECEDENTES DE LA DESCRIPCIÓN

- 10    **[0002]**La terapia génica ofrece un medio duradero para tratar una variedad de enfermedades. Antes la terapia génica generalmente se basaba en el uso de virus. Existen numerosos agentes virales que podrían seleccionarse para este propósito, cada uno con propiedades distintas que los harían más o menos adecuados para la terapia génica. Zhou et al., Adv Drug Deliv Rev. 106(Pt A): 3-26, 2016. Sin embargo,
- 15    las propiedades no deseadas de algunos vectores virales, incluidos sus perfiles inmunogénicos o su propensión a causar cáncer, han dado lugar a problemas de seguridad clínica y, hasta hace poco, limitaban su uso real a ciertas aplicaciones clínicas, por ejemplo, vacunas y estrategias oncológicas. Cotter et al., Front Biosci. 10: 1098-105 (2005).
- 20    **[0003]**El virus adenoasociado (AAV) es uno de los vectores más investigados en terapia génica. El AAV está constituido por una cubierta proteica que rodea y protege a un genoma de ADN pequeño monocatenario de aproximadamente 4,8 kilobases (kb). Naso et al., BioDrugs, 31 (4): 317-334, 2017. El AAV pertenece a la familia de los parvovirus y depende de la infección conjunta con otros virus, principalmente
- 25    adenovirus, para poder replicarse. *Id.* Su genoma monocatenario contiene tres genes, Rep (Replicación), Cap (Cápside) y aap (Ensamblaje). *Id.* Estas secuencias codificantes están flanqueadas por repeticiones terminales invertidas (ITR) que se requieren para la replicación y empaquetamiento del genoma. *Id.* Las dos ITR de AAV que actúan en cis tienen aproximadamente 145 nucleótidos de longitud con
- 30    secuencias palindrómicas interrumpidas que pueden plegarse en estructuras de horquilla en forma de T que funcionan como cebadores durante el inicio de la replicación del ADN.
- 35    **[0004]**Sin embargo, el uso de AAV convencional como vector para la administración de genes ha dado lugar a algunos inconvenientes. Uno de los principales inconvenientes está asociado con la capacidad limitada de empaquetamiento viral del

AAV de aproximadamente 4,5 kb de ADN heterólogo. (Dong et al., Hum Gene Ther. 7(17): 2101-12, 1996). Además, la administración de vectores AAV puede inducir una respuesta inmunitaria en seres humanos. Aunque se ha demostrado que el AAV es menos inmunogénico que algunos otros virus (es decir, los adenovirus), las proteínas de la cápside pueden activar diversos componentes del sistema inmunitario humano. Véase Naso et al., 2017. El AAV es un virus común en la población humana, y la mayoría de las personas han estado expuestas al AAV, por lo tanto, la mayoría de las personas ya han desarrollado una respuesta inmunitaria contra las variantes particulares a las que habían estado expuestas anteriormente. Esta respuesta adaptativa preexistente puede incluir NAb y células T que podrían disminuir la eficacia clínica de reinfecciones posteriores con AAV y/o la eliminación de células que han sido transducidas, lo que excluye a los pacientes con inmunidad anti-AAV preexistente de los tratamientos en terapia génica basados en AAV. Ya sea administrado local o sistémicamente, el virus será visto como una proteína extraña, por lo tanto, el sistema inmune adaptativo intentará eliminarlo. Además, el anticuerpo neutralizante anti-AAV inducido por el tratamiento con AAV dificulta el tratamiento repetido por AAV cuando el nivel terapéutico de eficacia no alcanza el primer tratamiento con AAV. Además, las evidencias sugieren que los bucles de horquilla en forma de T de las ITR de AAV son susceptibles de inhibición por las proteínas de la célula huésped/complejos de proteínas que se unen a las estructuras de horquilla en forma de T de las ITR de AAV. Véase, por ejemplo, Zhou et al., *Scientific Reports* 7: 5432 (14 de julio de 2017).

**[0005]** Por lo tanto, existe la necesidad en la técnica de expresar de manera eficiente y persistente secuencias diana, por ejemplo, proteínas y/o ARNm terapéuticos, en entornos in vitro e in vivo, evitando algunas de las consecuencias y limitaciones no pretendidas de la tecnología de vectores AAV existente.

### COMPENDIO DE LA DESCRIPCIÓN

**[0006]** Un aspecto de la presente descripción se refiere a una molécula de ácido nucleico que comprende una primera repetición terminal invertida (ITR), una segunda ITR y un casete génico que codifica una proteína terapéutica; donde la primera ITR y/o la segunda ITR son una ITR de un virus no adenoasociado (VnoAA), donde el casete génico está situado entre la primera ITR y la segunda ITR, y donde la proteína terapéutica comprende un factor de coagulación. En algunas realizaciones, el VnoAA se selecciona del grupo que consiste en un miembro de la familia viral *Parvoviridae* y cualquiera de sus combinaciones. En algunas realizaciones, la primera ITR es una

ITR de un VnoAA y la segunda ITR es una ITR de un virus adenoasociado (AAV) o donde la primera ITR es una ITR de un AAV y la segunda ITR es una ITR de un VnoAA. En otras realizaciones, la primera ITR y la segunda ITR son ITR de un VnoAA. En algunas realizaciones, la primera ITR y la segunda ITR son idénticas. En algunas realizaciones, la primera ITR y/o la segunda ITR comprenden una secuencia palindrómica no interrumpida. En otras realizaciones, la primera ITR y/o la segunda ITR comprenden una secuencia palindrómica que se interrumpe.

**[0007]** En algunas realizaciones, el VnoAA en la molécula de ácido nucleico es un miembro de la familia viral *Parvoviridae*. En algunas realizaciones, el miembro de la familia viral *Parvoviridae* se selecciona del grupo que consiste en *Bocavirus*, *Dependovirus*, *Erythrovirus*, *Amdovirus*, *Parvovirus*, *Densovirus*, *Iteravirus*, *Contravirus*, *Aveparvovirus*, *Copiparvovirus*, *Protoparvovirus*, *Tetraparvovirus*, *Ambidensovirus*, *Brevidensovirus*, *Hepandensovirus*, *Penstyldensovirus* y sus combinaciones. En algunas realizaciones, el miembro de la familia viral *Parvoviridae* es el parvovirus y eritrovirus B19 (virus humano). En algunas realizaciones, el miembro de la familia viral *Parvoviridae* es una cepa de parvovirus de pato de Muscovy (MDPV). En algunas realizaciones, la cepa MDPV es FZ91-30 atenuada. En algunas realizaciones, la cepa MDPV es YY patógena. Todavía en algunas realizaciones, el *Dependoparvovirus* es una cepa de *Dependovirus* de ganso y parvovirus (GPV). En algunas realizaciones, la cepa GPV es 82-0321V atenuada. En algunas realizaciones, la cepa de GPV es B patógena. En algunas realizaciones, el miembro de la familia viral *Parvoviridae* se selecciona del grupo que consiste en parvovirus porcino (U44978), virus diminuto de ratones (U34256), parvovirus canino (M19296), virus de la enteritis de visón (D00765) y cualquiera de sus combinaciones.

**[0008]** En algunas realizaciones, la molécula de ácido nucleico comprende además un promotor. En algunas realizaciones, el promotor es un promotor específico de tejido. En algunas realizaciones, el promotor dirige la expresión de la proteína terapéutica en hepatocitos, células endoteliales, células musculares, células sinusoidales o cualquiera de sus combinaciones. En algunas realizaciones, el promotor está colocado 5' respecto a la secuencia de ácido nucleico que codifica el factor de coagulación. En algunas realizaciones, el promotor se selecciona del grupo que consiste en un promotor de tiretina de ratón (mTTR), un promotor endógeno humano del factor VIII (F8), un promotor de alfa-1-antitripsina humana (hAAT), un promotor mínimo de albúmina humana, un promotor de albúmina de ratón, un promotor de tristetraprolina (TTP), un promotor CASI, un promotor CAG, un promotor de

citomegalovirus (CMV),  $\alpha$ 1-antitripsina (AAT), creatina quinasa muscular (MCK), cadena pesada de miosina alfa ( $\alpha$ MHC), mioglobina (MB), desmina (DES), SPc5-12, 2R5Sc5-12, dMCK, tMCK, un promotor de fosfoglicerato quinasa (PGK) y cualquiera de sus combinaciones. En algunas realizaciones, el promotor comprende un promotor TTP.

**[0009]** En algunas realizaciones, la molécula de ácido nucleico comprende además una secuencia intrónica. En algunas realizaciones, la secuencia intrónica está colocada 5' respecto a la secuencia de ácido nucleico que codifica el factor de coagulación. En algunas realizaciones, la secuencia intrónica está colocada 3' respecto al promotor. En algunas realizaciones, la secuencia intrónica comprende una secuencia intrónica sintética. En algunas realizaciones, la secuencia intrónica comprende SEQ ID NO: 115.

**[0010]** En algunas realizaciones, la molécula de ácido nucleico comprende un elemento regulador postranscripcional. En algunas realizaciones, el elemento regulador postranscripcional está colocado 3' respecto a la secuencia de ácido nucleico que codifica el factor de coagulación. En algunas realizaciones, el elemento regulador post-transcripcional comprende un elemento regulador post-transcripcional del virus de la hepatitis de la marmota mutado (WPRE), un sitio de unión a microARN, una secuencia de direccionamiento nuclear de ADN, o cualquiera de sus combinaciones. En algunas realizaciones, el sitio de unión de microARN comprende un sitio de unión a miR142-3p.

**[0011]** En algunas realizaciones, la molécula de ácido nucleico comprende una secuencia de cola poli (A) 3'UTR. En algunas realizaciones, la secuencia de cola poli (A) 3'UTR se selecciona del grupo que consiste en bGH poli (A), actina poli (A), hemoglobina poli (A) y cualquiera de sus combinaciones. En algunas realizaciones, la secuencia de cola poli (A) 3'UTR comprende bGH poli (A).

**[0012]** En algunas realizaciones, la molécula de ácido nucleico comprende una secuencia potenciadora. En algunas realizaciones, la secuencia potenciadora está colocada entre la primera ITR y la segunda ITR.

**[0013]** En algunas realizaciones, la molécula de ácido nucleico comprende en este orden: la primera ITR, el casete génico y la segunda ITR; en donde el casete génico comprende una secuencia promotora específica de tejido, una secuencia intrónica, la secuencia de ácido nucleico que codifica una proteína terapéutica, por ejemplo, un factor de coagulación, o un ARNmi, un elemento regulador postranscripcional y una secuencia de cola poli (A) 3'UTR.

**[0014]** En algunas realizaciones, la molécula de ácido nucleico comprende en este orden: una secuencia promotora específica de tejido, una secuencia intrónica, la secuencia de ácido nucleico que codifica un polipéptido FVIII, un elemento regulador postranscripcional y una secuencia de cola poli (A) 3'UTR.

5 **[0015]** En una realización, la molécula de ácido nucleico comprende:

(a) la primera ITR es una ITR de un miembro de la familia de *Parvoviridae* VnoAA;

(b) la secuencia promotora específica de tejido, por ejemplo, un promotor TTP;

(c) el intrón, por ejemplo, un intrón sintético;

10 (d) la secuencia de nucleótidos que codifica una proteína terapéutica, por ejemplo, un factor de coagulación o un ARNm;

(e) el elemento regulador post-transcripcional, por ejemplo, WPRE;

(f) la secuencia de cola poli (A) 3'UTR, por ejemplo, bGHpA; y

(g) la segunda ITR es una ITR de un miembro de la familia de *Parvoviridae* VnoAA.

15

**[0016]** En algunas realizaciones, la molécula de ácido nucleico comprende un ácido nucleico monocatenario. En algunas realizaciones, el casete génico comprende un ácido nucleico bicatenario.

20

**[0017]** En algunas realizaciones, la molécula de ácido nucleico comprende un gen que codifica una proteína terapéutica, por ejemplo, un factor de coagulación, en el que el factor de coagulación se expresa por hepatocitos, células endoteliales, células musculares, células sinusoidales o cualquiera de sus combinaciones. En algunas realizaciones, el factor de coagulación comprende factor I (FI), factor II (FII), factor V (FV), factor VII (FVII), factor VIII (FVIII), factor IX (FIX), factor X (FX), factor XI (FXI),  
25 factor XII (FXII), factor XIII (FVIII), factor Von Willebrand (VWF), precalicreína, quinínogeno de alto peso molecular, fibronectina, antitrombina III, heparina cofactor II, proteína C, proteína S, proteína Z, inhibidor de la proteasa relacionada con la proteína Z (ZPI), plasminógeno, alfa 2-antiplasmina, activador de plasminógeno tisular (tPA), uroquinasa, inhibidor del activador de plasminógeno-1 (PAI-1), inhibidor del activador  
30 de plasminógeno-2 (PAI2) o cualquiera de sus combinaciones. En algunas realizaciones, el factor de coagulación es FVIII. En algunas realizaciones, el FVIII comprende FVIII maduro de longitud completa. En algunas realizaciones, el FVIII comprende una secuencia de aminoácidos al menos aproximadamente 70%, al menos aproximadamente 75%, al menos aproximadamente 80%, al menos  
35 aproximadamente 85%, al menos aproximadamente 90%, al menos aproximadamente

95%, al menos aproximadamente 96% o al menos aproximadamente 97%, al menos aproximadamente 98%, al menos aproximadamente 99%, o 100% idéntico a una secuencia de aminoácidos de SEQ ID NO: 106. En algunas realizaciones, el FVIII comprende dominio A1, dominio A2, dominio A3, dominio C1, dominio C2 y un dominio parcial o no B. En algunas realizaciones, el FVIII comprende una secuencia de aminoácidos de al menos aproximadamente 70%, al menos aproximadamente 75%, al menos aproximadamente 80%, al menos aproximadamente 85%, al menos aproximadamente 90%, al menos aproximadamente 95%, al menos aproximadamente 96%, al menos aproximadamente 97%, al menos aproximadamente 98%, al menos aproximadamente 99%, o 100% idéntico a la secuencia de aminoácidos de SEQ ID NO: 109.

**[0018]** En algunas realizaciones, el factor de coagulación comprende un resto heterólogo. En algunas realizaciones, el resto heterólogo se selecciona del grupo que consiste en albúmina o un fragmento de la misma, una región Fc de inmunoglobulina, el péptido C-terminal (CTP) de la subunidad  $\beta$  de la gonadotropina coriónica humana, una secuencia PAS, una secuencia HAP, una transferrina o un fragmento de la misma, un resto de unión a albúmina, un derivado de la misma y cualquiera de sus combinaciones. En algunas realizaciones, el resto heterólogo está unido al extremo N o al extremo C del FVIII o se inserta entre dos aminoácidos en el FVIII. En algunas realizaciones, el resto heterólogo se inserta entre dos aminoácidos en uno o más sitios de inserción seleccionados de los sitios de inserción enumerados en la Tabla 5. En algunas realizaciones, el FVIII comprende además dominio A1, dominio A2, dominio C1, dominio C2, un dominio B opcional y un resto heterólogo, en el que el resto heterólogo se inserta inmediatamente después del aminoácido 745 correspondiente al FVIII maduro (SEQ ID NO : 106).

**[0019]** En algunas realizaciones, el FVIII comprende además una pareja de unión a FcRn. En algunas realizaciones, la pareja de unión a FcRn comprende una región Fc de un dominio constante de inmunoglobulina. En algunas realizaciones, la secuencia de ácido nucleico que codifica el FVIII está optimizada con codón. En algunas realizaciones, la secuencia de ácido nucleico que codifica el FVIII es un codón optimizado para su expresión en un ser humano.

**[0020]** En algunas realizaciones, la secuencia de ácido nucleico que codifica el FVIII comprende una secuencia de nucleótidos al menos aproximadamente un 60%, al menos aproximadamente 65%, al menos aproximadamente 70%, al menos aproximadamente 75%, al menos aproximadamente 80%, al menos aproximadamente

85%, al menos aproximadamente 90%, al menos aproximadamente 95%, al menos aproximadamente 96%, al menos aproximadamente 97%, al menos aproximadamente 98%, al menos aproximadamente 99% o aproximadamente 100% idéntico a una secuencia de nucleótidos de SEQ ID NO: 107. En algunas realizaciones, la secuencia de ácido nucleico que codifica el FVIII comprende una secuencia de nucleótidos al menos aproximadamente 60%, al menos aproximadamente 65%, al menos aproximadamente 70%, al menos aproximadamente 75%, al menos aproximadamente 80%, al menos aproximadamente 85%, al menos aproximadamente 90%, al menos aproximadamente 95%, al menos aproximadamente 96%, al menos aproximadamente 97%, al menos aproximadamente 98%, al menos aproximadamente 99% o aproximadamente 100% idéntica con la secuencia de nucleótidos de SEQ ID NO: 71.

**[0021]** En algunas realizaciones, la molécula de ácido nucleico se formula con un agente de administración. En algunas realizaciones, el agente de administración comprende una o más nanopartículas lipídicas. En algunas realizaciones, el agente de administración se selecciona del grupo que consiste en liposomas, moléculas poliméricas no lipídicas, endosomas y cualquiera de sus combinaciones.

**[0022]** En algunas realizaciones, la molécula de ácido nucleico está formulada para la administración intravenosa, transdérmica, intradérmica, subcutánea, pulmonar u oral, o cualquiera de sus combinaciones. En algunas realizaciones, la molécula de ácido nucleico está formulada para la administración intravenosa.

**[0023]** En algunas realizaciones, se proporciona en este documento un vector que comprende la molécula de ácido nucleico como se describe a lo largo de la descripción. En algunas realizaciones, se proporciona en este documento un polipéptido codificado por la molécula de ácido nucleico como se describe a lo largo de la descripción. En algunas realizaciones, se proporciona en este documento una célula huésped que comprende la molécula de ácido nucleico como se describe a lo largo de la descripción.

**[0024]** En algunas realizaciones, se proporciona en el presente documento una composición farmacéutica que comprende (a) el ácido nucleico como se describe en el presente documento, el vector que comprende la molécula de ácido nucleico como se describe a lo largo de la descripción, el polipéptido codificado por la molécula de ácido nucleico como se describe a lo largo de la descripción o la célula huésped que comprende la molécula de ácido nucleico como se describe a lo largo de la descripción; (b) un LNP; y (c) un excipiente farmacéuticamente aceptable. En algunas realizaciones, en este documento se proporciona un kit que comprende la molécula

de ácido nucleico como se describe a lo largo de la descripción y las instrucciones para administrar la molécula de ácido nucleico a un sujeto que lo necesite. En algunas realizaciones, en este documento se proporciona un sistema de baculovirus para la producción de la molécula de ácido nucleico descrita en este documento. En algunas realizaciones, en el presente documento se proporciona un sistema de baculovirus, en el que la molécula de ácido nucleico descrita en el presente documento se produce en células de insecto.

**[0025]** En algunas realizaciones, en este documento se proporciona un sistema de suministro de nanopartículas para construcciones de expresión, en el que la construcción de expresión comprende la molécula de ácido nucleico descrita en este documento.

**[0026]** También se describe en el presente documento un método para producir un polipéptido con actividad de coagulación que comprende: cultivar una célula huésped descrita en el presente documento en condiciones adecuadas y recuperar el polipéptido con actividad de coagulación. En algunas realizaciones, se describe en el presente documento un método para expresar un factor de coagulación en un sujeto que lo necesita, que comprende administrar al sujeto la molécula de ácido nucleico descrita en el presente documento, el vector descrito en el presente documento, el polipéptido descrito en el presente documento o la composición farmacéutica descrita en el presente documento. En algunas realizaciones, se describe en el presente documento un método para tratar a un sujeto que tiene una deficiencia del factor de coagulación que comprende administrar al sujeto la molécula de ácido nucleico descrita en el presente documento, el vector descrito en el presente documento, el polipéptido descrito en el presente documento o la composición farmacéutica descrita en el presente documento. En algunas realizaciones, la molécula de ácido nucleico se administra por vía intravenosa, transdérmica, intradérmica, subcutánea, oral, pulmonar o cualquiera de sus combinaciones. En algunas realizaciones, la molécula de ácido nucleico se administra por vía intravenosa. En algunas realizaciones, el método comprende además administrar al sujeto un segundo agente. En algunas realizaciones, el sujeto es un mamífero. En algunas realizaciones, el sujeto es un ser humano.

**[0027]** En algunas realizaciones, la administración de la molécula de ácido nucleico al sujeto da como resultado una mayor actividad de FVIII, en relación con una actividad de FVIII en el sujeto antes de la administración, donde la actividad de FVIII aumenta al menos aproximadamente 2 veces, al menos aproximadamente 3 veces, al menos

aproximadamente 4 veces, al menos aproximadamente 5 veces, al menos  
aproximadamente 6 veces, al menos aproximadamente 7 veces, al menos  
aproximadamente 8 veces, al menos aproximadamente 9 veces, al menos  
aproximadamente 10 veces, al menos aproximadamente 11 veces, al menos  
5 aproximadamente 12 veces, al menos aproximadamente 13 veces, al menos  
aproximadamente 14 veces, al menos aproximadamente 15 veces, al menos  
aproximadamente 20 veces, al menos aproximadamente 25 veces, al menos  
aproximadamente 30 veces, al menos aproximadamente 35 veces, al menos  
aproximadamente 40 veces, al menos aproximadamente 50 veces, al menos  
10 aproximadamente 60 veces, al menos aproximadamente 70 veces, al menos  
aproximadamente 80 veces, al menos aproximadamente 90 veces, o al menos  
aproximadamente 100 veces.

**[0028]** En algunas realizaciones, el sujeto tiene un trastorno hemorrágico. En algunas  
realizaciones, el trastorno hemorrágico es una hemofilia. En algunas realizaciones, el  
15 trastorno hemorrágico es la hemofilia A.

### **BREVE DESCRIPCIÓN DE LAS FIGURAS**

**[0029]** FIG. 1A es una representación esquemática de un casete de expresión del  
factor de coagulación monocatenario (p. ej., FVIII). Se muestran las localizaciones de  
20 5' ITR de un VnoAA (con horquilla en el extremo de la estructura del ADN  
monocatenario), 3' ITR de un VnoAA (con horquilla), una secuencia promotora (por  
ejemplo, TTPp) y una secuencia transgén, por ejemplo, secuencia FVIIIco6XTEN con  
un XTEN144 insertadas dentro del dominio B. El casete de expresión ejemplar  
también muestra otros posibles elementos, por ejemplo, una secuencia intrónica, una  
25 secuencia WPREmut y una secuencia bGHpA.

**[0030]** Las FIG. 1B-1D son representaciones esquemáticas de plásmidos utilizados  
para preparar casetes de expresión del factor de coagulación monocatenario, tales  
como los casetes que se muestran en la FIG. 1A, donde las ITR del casete se derivan  
de AAV2 (FIG. 1B), B19 (FIG. 1C) o GPV (FIG. 1D). Una construcción plasmídica que  
30 comprende un casete de expresión de FVIII monocatenario como se muestra en este  
documento se digirió con PvuII (en sitios PvuII) (Figura 1B) o LglI (en sitios LglI)  
(Figuras 1C e 1D) para liberar el genoma viral. El ADN bicatenario se calentó a 95°C  
para producir ADN monocatenario y luego se incubó a 4°C para permitir la formación  
de la estructura ITR.

**[0031]** FIG. 2A es un árbol filogenético que ilustra las relaciones entre varios miembros de la familia del parvovirus. B19, AAV-2 y GPV están marcados con cuadros delineados.

**[0032]** FIG. 2B es un dibujo esquemático de los diversos casetes, incluidas las estructuras de horquilla.

**[0033]** Las FIG. 3A y 3B son alineamientos de las ITR de B19, GPV y AAV2 (Figura 3A) y B19 y GPV (Figura 3B). El sombreado gris muestra la homología.

**[0034]** Las figs. 4A-4C muestran la actividad plasmática de FVIII después de la administración de ADN desnudo de FVIII-AAV monocatenario (AAV-FVIII monocatenario; Figura 2 A), ADN monocatenario-B19 FVIII (Figura 2B) o ADN monocatenario-GPV FVIII (Figura 2C) por vía de inyección hidrodinámica (HDI) en ratones Hem A. La actividad de FVIII se midió (como porcentaje de un control) en muestras de plasma a las 24 horas, 3 días, 2 semanas, 3 semanas, 1 mes, 2 meses, 3 meses y 4 meses en ratones tratados con un solo HDI de ADN monocatenario en 50 µg/ratón (figura 2C), 20 µg/ratón (figuras 2A y 2B), 10 µg/ratón (figuras 2A y 2C) o 5 µg/ratón (figura 2A). Se administró un HDI de 5 µg/ratón de ADN plasmídico como control (Figuras 2A-2C).

#### **DESCRIPCIÓN DETALLADA DE LA DESCRIPCIÓN**

**[0035]** La presente descripción describe moléculas de ácido nucleico tipo plasmídicas que comprenden una primera repetición terminal invertida (ITR), una segunda ITR y un casete génico, por ejemplo, que codifica una proteína terapéutica o un ARNm, donde la primera ITR y/o la segunda ITR son una ITR de un virus no adenoasociado (por ejemplo, la primera ITR y/o la segunda ITR son de un VnoAA). En algunas realizaciones, el casete génico codifica una proteína terapéutica. En algunas realizaciones, la proteína terapéutica comprende una proteína seleccionada de un factor de coagulación, un factor de crecimiento, una hormona, una citocina, un anticuerpo, un fragmento de la misma o una sus combinaciones. En algunas realizaciones, el casete génico codifica distrofina ligada a X, MTMI (miotubularina), tirosina hidroxilasa, AADC, ciclohidrolasa, SMN1, FXN (frataxina), GUCY2D, RSI, CFH, HTRA, ARMS, CFB/CC2, CNGA/CNGB, Prf65, ARSA, PSAP, IDUA (MPS I), IDS (MPS II), PAH, GAA (alfa-glucosidasa ácida), o cualquiera de sus combinaciones.

**[0036]** En algunas realizaciones, la proteína terapéutica comprende un factor de coagulación. En una realización particular, la proteína terapéutica comprende una proteína FVIII o FIX.

[0037] En algunas realizaciones, el casete génico codifica un ARNmi. En ciertas realizaciones, el ARNmi regula a la baja la expresión de un gen diana seleccionado de SOD1, HTT, RHO o cualquiera de sus combinaciones.

[0038] En ciertas realizaciones, el VnoAA se selecciona del grupo que consiste en un miembro de la familia viral *Parvoviridae* y cualquiera de sus combinaciones. La presente descripción se dirige además a métodos para expresar una proteína terapéutica, por ejemplo, un factor de coagulación, por ejemplo, un FVIII, en un sujeto que lo necesita, que comprende administrar al sujeto una molécula de ácido nucleico que comprende una primera repetición terminal invertida (ITR), una segunda ITR y un casete génico, por ejemplo, que codifica una proteína terapéutica o un ARNmi, en el que la primera ITR y/o la segunda ITR son una ITR de un virus no adenoasociado (VnoAA). En ciertas realizaciones, la descripción describe una molécula de ácido nucleico aislada que comprende una secuencia de nucleótidos que tiene homología de secuencia con una secuencia de nucleótidos seleccionada de SEQ ID N°: 113 y 120.

[0039] Las construcciones ejemplares de la descripción se ilustran en las figuras adjuntas y el listado de secuencias. Para proporcionar una comprensión clara de la memoria descriptiva y las reivindicaciones, a continuación se proporcionan las siguientes definiciones.

## I. Definiciones

[0040] Debe observarse que el término "una" entidad se refiere a una entidad o más de una de esas entidades: por ejemplo, se entiende que "una secuencia de nucleótidos" representa una o más secuencias de nucleótidos. De manera similar, se entiende que "una proteína terapéutica" y "un ARNmi" representan una o más proteínas terapéuticas y uno o más ARNmi, respectivamente. Por lo tanto, los términos "una" (o "un/uno"), "uno/a o más" y "al menos uno/a" pueden usarse indistintamente en el presente documento.

[0041] El término "aproximadamente" se usa en el presente documento para dar a entender aproximadamente, cerca de, alrededor o en sus proximidades. Cuando el término "aproximadamente" se usa junto con un intervalo numérico, modifica ese intervalo extendiendo los límites por encima y por debajo de los valores numéricos establecidos. En general, el término "aproximadamente" se usa en el presente documento para modificar un valor numérico por encima y por debajo del valor establecido en una variación del 10 por ciento, hacia arriba o hacia abajo (mayor o menor).

**[0042]** También, como se usa en el presente documento, "y/o" se refiere y abarca todas y cada una de las combinaciones posibles de uno o más de los elementos enumerados asociados, así como la falta de combinaciones cuando se interpreta en la alternativa ("o").

5 **[0043]** "Ácidos nucleicos", "moléculas de ácido nucleico", "nucleótidos", "secuencia de nucleótidos" y "polinucleótido" se usan indistintamente y se refieren a la forma polimérica del éster fosfato de los ribonucleósidos (adenosina, guanosina, uridina o citidina; "moléculas de ARN") o desoxirribonucleósidos (desoxiadenosina, desoxiguanosina, desoxitimidina o desoxicitidina; "moléculas de ADN"), o cualquier  
10 análogo de fosfoéster de los mismos, como fosforotioatos y tioésteres, en forma monocatenaria o hélice bicatenaria. Las secuencias de ácido nucleico monocatenario se refieren a ADN monocatenario o ARN monocatenario. Las hélices de ADN-ADN, ADN-ARN y ARN-ARN de doble cadena son posibles. La expresión molécula de ácido nucleico y, en particular, molécula de ADN o ARN, se refiere solo a la estructura  
15 primaria y secundaria de la molécula, y no limita a ninguna forma terciaria particular. Por lo tanto, este término incluye ADN bicatenario encontrado, entre otros, en moléculas de ADN lineales o circulares (por ejemplo, fragmentos de restricción), plásmidos, ADN superenrollado y cromosomas. Al analizar la estructura de moléculas de ADN bicatenarias particulares, las secuencias se pueden describir en este  
20 documento de acuerdo con la convención normal de dar solo la secuencia en la dirección 5' a 3' a lo largo de la cadena de ADN no transcrita (es decir, la cadena que tiene una secuencia homóloga al ARNm). Una "molécula de ADN recombinante" es una molécula de ADN que ha sufrido una manipulación biológica molecular. El ADN incluye, entre otros, ADNc, ADN genómico, ADN plasmídico, ADN sintético y ADN  
25 semisintético. Una "composición de ácido nucleico" de la descripción comprende uno o más ácidos nucleicos como se describe en el presente documento.

**[0044]** Como se usa en el presente documento, una "repetición terminal invertida" (o "ITR") se refiere a una subsecuencia de ácido nucleico ubicada en el extremo 5' o 3' de una secuencia de ácido nucleico monocatenaria, que comprende un conjunto de  
30 nucleótidos (secuencia inicial) seguido después por su complemento inverso, es decir, una secuencia palindrómica. La secuencia intermedia de nucleótidos entre la secuencia inicial y el complemento inverso puede tener cualquier longitud, incluido el cero. En una realización, la ITR útil para la presente descripción comprende una o más secuencias palindrómicas. "Una ITR puede tener cualquier número de funciones.  
35 En algunas realizaciones, la ITR descrita en este documento forma una estructura de

horquilla. En algunas realizaciones, la ITR forma una estructura de horquilla en forma de T. En algunas realizaciones, la ITR forma una estructura de horquilla sin forma de T, por ejemplo, una estructura de horquilla en forma de U. En algunas realizaciones, la ITR promueve la supervivencia a largo plazo de la molécula de ácido nucleico en el núcleo de la célula. En algunas realizaciones, la ITR promueve la supervivencia permanente de la molécula de ácido nucleico en el núcleo de una célula (por ejemplo, durante toda la vida útil de la célula). En algunas realizaciones, la ITR promueve la estabilidad de la molécula de ácido nucleico en el núcleo de una célula. En algunas realizaciones, la ITR promueve la retención de la molécula de ácido nucleico en el núcleo de una célula. En algunas realizaciones, la ITR promueve la persistencia de la molécula de ácido nucleico en el núcleo de una célula. En algunas realizaciones, la ITR inhibe o previene la degradación de la molécula de ácido nucleico en el núcleo de la célula.

**[0045]** En una realización, la secuencia inicial y/o el complemento inverso comprenden

aproximadamente 2-600 nucleótidos, aproximadamente 2-550 nucleótidos, aproximadamente 2-500 nucleótidos, aproximadamente 2-450 nucleótidos, aproximadamente 2-400 nucleótidos, aproximadamente 2-350 nucleótidos, aproximadamente 2-300 nucleótidos, o aproximadamente 2-250 nucleótidos. En algunas realizaciones, la secuencia inicial y/o el complemento inverso comprenden aproximadamente 5-600 nucleótidos, aproximadamente 10-600 nucleótidos, aproximadamente 15-600 nucleótidos, aproximadamente 20-600 nucleótidos, aproximadamente 25-600 nucleótidos, aproximadamente 30-600 nucleótidos, aproximadamente 35-600 nucleótidos, aproximadamente 40-600 nucleótidos, aproximadamente 45-600 nucleótidos, aproximadamente 50-600 nucleótidos, aproximadamente 60-600 nucleótidos, aproximadamente 70-600 nucleótidos, aproximadamente 80-600 nucleótidos, aproximadamente 90-600 nucleótidos, aproximadamente 100-600 nucleótidos, aproximadamente 150-600 nucleótidos, aproximadamente 200-600 nucleótidos, aproximadamente 300-600 nucleótidos, aproximadamente 350-600 nucleótidos, aproximadamente 400-600 nucleótidos, aproximadamente 450-600 nucleótidos, aproximadamente 500-600 nucleótidos, o aproximadamente 550-600 nucleótidos. En algunas realizaciones, la secuencia inicial y/o el complemento inverso comprenden aproximadamente 5-550 nucleótidos, de aproximadamente 5 a 500 nucleótidos, aproximadamente 5-450 nucleótidos, de aproximadamente 5 a 400 nucleótidos, aproximadamente 5-350 nucleótidos, de aproximadamente 5 a 300 nucleótidos, o aproximadamente 5-250 nucleótidos. En

algunas realizaciones, la secuencia inicial y/o el complemento inverso comprenden aproximadamente 10-550 nucleótidos, aproximadamente 15-500 nucleótidos, aproximadamente 20-450 nucleótidos, aproximadamente 25-400 nucleótidos, aproximadamente 30-350 nucleótidos, aproximadamente 35-300 nucleótidos, o aproximadamente 40-250 nucleótidos. En ciertas realizaciones, la secuencia inicial y/o el complemento inverso comprenden aproximadamente 225 nucleótidos, aproximadamente 250 nucleótidos, aproximadamente 275 nucleótidos, aproximadamente 300 nucleótidos, aproximadamente 325 nucleótidos, aproximadamente 350 nucleótidos, aproximadamente 375 nucleótidos, aproximadamente 400 nucleótidos, aproximadamente 425 nucleótidos, aproximadamente 450 nucleótidos, aproximadamente 475 nucleótidos, aproximadamente 500 nucleótidos, aproximadamente 525 nucleótidos, aproximadamente 550 nucleótidos, aproximadamente 575 nucleótidos o aproximadamente 600 nucleótidos. En realizaciones particulares, la secuencia inicial y/o el complemento inverso comprenden aproximadamente 400 nucleótidos.

**[0046]** En otras realizaciones, la secuencia inicial y/o el complemento inverso comprenden aproximadamente 2-200 nucleótidos, aproximadamente 5-200 nucleótidos, aproximadamente 10-200 nucleótidos, aproximadamente 20-200 nucleótidos, aproximadamente 30-200 nucleótidos, aproximadamente 40-200 nucleótidos, aproximadamente 50-200 nucleótidos, aproximadamente 60-200 nucleótidos, aproximadamente 70-200 nucleótidos, aproximadamente 80-200 nucleótidos, aproximadamente 90-200 nucleótidos, aproximadamente 100-200 nucleótidos, aproximadamente 125-200 nucleótidos, aproximadamente 150-200 nucleótidos, o aproximadamente 175-200 nucleótidos. En otras realizaciones, la secuencia inicial y/o el complemento inverso comprenden aproximadamente 2-150 nucleótidos, aproximadamente 5-150 nucleótidos, aproximadamente 10-150 nucleótidos, aproximadamente 20-150 nucleótidos, aproximadamente 30-150 nucleótidos, aproximadamente 40-150 nucleótidos, aproximadamente 50-150 nucleótidos, aproximadamente 75-150 nucleótidos, aproximadamente 100-150 nucleótidos o aproximadamente 125-150 nucleótidos. En otras realizaciones, la secuencia inicial y/o el complemento inverso comprenden aproximadamente 2-100 nucleótidos, aproximadamente 5-100 nucleótidos, aproximadamente 10-100 nucleótidos, aproximadamente 20-100 nucleótidos, aproximadamente 30-100 nucleótidos, aproximadamente 40-100 nucleótidos, aproximadamente 50-100 nucleótidos, o aproximadamente 75-100 nucleótidos. En otras realizaciones, la

secuencia inicial y/o el complemento inverso comprenden aproximadamente 2-50 nucleótidos, aproximadamente 10-50 nucleótidos, aproximadamente 20-50 nucleótidos, aproximadamente 30-50 nucleótidos, aproximadamente 40-50 nucleótidos, aproximadamente 3-30 nucleótidos, aproximadamente 4-20 nucleótidos, o aproximadamente 5-10 nucleótidos. En otra realización, la secuencia inicial y/o el complemento inverso consisten en dos nucleótidos, tres nucleótidos, cuatro nucleótidos, cinco nucleótidos, seis nucleótidos, siete nucleótidos, ocho nucleótidos, nueve nucleótidos, diez nucleótidos, 11 nucleótidos, 12 nucleótidos, 13 nucleótidos, 14 nucleótidos, 15 nucleótidos, 16 nucleótidos, 17 nucleótidos, 18 nucleótidos, 19 nucleótidos o 20 nucleótidos. En otras realizaciones, un nucleótido intermedio entre la secuencia inicial y el complemento inverso tiene (por ejemplo, consta de) 0 nucleótidos, 1 nucleótido, dos nucleótidos, tres nucleótidos, cuatro nucleótidos, cinco nucleótidos, seis nucleótidos, siete nucleótidos, ocho nucleótidos, nueve nucleótidos, 10 nucleótidos, 11 nucleótidos, 12 nucleótidos, 13 nucleótidos, 14 nucleótidos, 15 nucleótidos, 16 nucleótidos, 17 nucleótidos, 18 nucleótidos, 19 nucleótidos o 20 nucleótidos.

**[0047]** Por lo tanto, una "ITR", como se usa en el presente documento, puede plegarse sobre sí misma y formar un segmento bicatenario. Por ejemplo, la secuencia GATCXXXGATC comprende una secuencia inicial de GATC y su complemento (3'CTAG5') cuando se pliega para formar una doble hélice. En algunas realizaciones, la ITR comprende una secuencia palindrómica continua (por ejemplo, GATCGATC) entre la secuencia inicial y el complemento inverso. En algunas realizaciones, la ITR comprende una secuencia palindrómica interrumpida (por ejemplo, GATCXXXGATC) entre la secuencia inicial y el complemento inverso. En algunas realizaciones, las secciones complementarias de la secuencia palindrómica continua o interrumpida interactúan entre sí para formar una estructura de "bucle de horquilla". Como se usa en el presente documento, se produce una estructura de "bucle de horquilla" cuando al menos dos secuencias complementarias en un par de bases de la molécula de nucleótido monocatenario forman una sección bicatenaria. En algunas realizaciones, solo una parte de la ITR forma un bucle de horquilla. En otras realizaciones, todo la ITR forma un bucle de horquilla.

**[0048]** En la presente descripción, al menos una ITR es una ITR de un virus no adenovirus asociado (VnoAA). En ciertas realizaciones, la ITR es una ITR de un miembro VnoAA de la familia viral *Parvoviridae*. En algunas realizaciones, la ITR es una ITR de un miembro VnoAA del género *Dependovirus* o el género *Erythrovirus*. En

realizaciones particulares, la ITR es una ITR de un parvovirus de ganso (GPV), un parvovirus de pato de Muscovy (MDPV) o un parvovirus de eritrovirus B19 (también conocido como parvovirus B19, eritroparvovirus de primates 1, virus B19 y eritrovirus). En ciertas realizaciones, cada ITR de dos ITR es una ITR de un AAV. En otras realizaciones, cada ITR de dos ITR en la construcción es una ITR de un serotipo AAV seleccionado entre los serotipos 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11 y cualquiera de sus combinaciones. En una realización particular, la ITR se deriva del serotipo 2 de AAV, por ejemplo, una ITR del serotipo 2 de AAV.

**[0049]** En ciertos aspectos de la presente descripción, la molécula de ácido nucleico comprende dos ITR, una 5' ITR y una 3' ITR, donde la 5' ITR está ubicada en el extremo 5' de la molécula de ácido nucleico, y la 3' ITR está ubicada en el extremo 3' de la molécula de ácido nucleico. 5' ITR y 3' ITR pueden derivarse del mismo virus o de virus diferentes. En ciertas realizaciones, la 5' ITR se deriva de un AAV y la 3' ITR no se deriva de ningún virus AAV (por ejemplo, un VnoAA). En algunas realizaciones, la ITR 3' se deriva de un AAV y la ITR 5' no se deriva de ningún virus AAV (por ejemplo, un VnoAA). En otras realizaciones, la 5' ITR no se deriva de ningún virus AAV (por ejemplo, un VnoAA), y la 3' ITR se deriva del mismo virus o un VnoAA diferente.

**[0050]** El término "parvovirus", como se usa en el presente documento, abarca a la familia *Parvoviridae*, que incluye entre otros parvovirus que se replican de forma autónoma y *Dependovirus*. Los parvovirus autónomos incluyen, por ejemplo, miembros de los géneros *Bocavirus*, *Dependovirus*, *Erythrovirus*, *Amdovirus*, *Parvovirus*, *Densovirus*, *Iteravirus*, *Contravirus*, *Aveparvovirus*, *Copiparvovirus*, *Protoparvovirus*, *Tetraparvovirus*, *Ambidensovirus*, *Brevidensovirus*, *Hepandensovirus*, *Hepandensovirus*, *Penstyldensovirus*.

**[0051]** Los parvovirus autónomos ejemplares incluyen, entre otros, parvovirus porcino, virus minucioso de ratones, parvovirus canino, virus de entertitus del visón, parvovirus bovino, parvovirus de pollo, virus de panleucopenia felina, parvovirus felino, parvovirus de ganso, parvovirus HI, parvovirus de pato muscovy, parvovirus de serpiente y virus B19. Los expertos en la técnica conocen otros parvovirus autónomos. Ver, por ejemplo, FIELDS et al. *VIROLOGY*, volumen 2, capítulo 69 (4ª ed., Lippincott-Raven Publishers).

**[0052]** La expresión "VnoAA", como se usa en el presente documento, abarca ácidos nucleicos, proteínas y virus de la familia *Parvoviridae*, excluyendo cualquier virus adenoasociado (AAV) de la familia *Parvoviridae*. "VnoAA" incluye, entre otros, miembros que se replican de manera autónoma de los géneros *Bocavirus*,

*Dependovirus, Erythrovirus, Amdovirus, Parvovirus, Densovirus, Iteravirus, Contravirus, Aveparvovirus, Copiparvovirus, Protoparvovirus, Tetraparvovirus, Ambidensovirus, Brevidensovirus, Hepadensovirus, y Penstyldensovirus.*

**[0053]** Tal como se usa en el presente documento, la expresión "virus adenoasociado"

5 (AAV) incluye, entre otros, AAV tipo 1, AAV tipo 2, AAV tipo 3 (incluidos los tipos 3 A y 3B), AAV tipo 4, AAV tipo 5, AAV tipo 6, AAV tipo 7, AAV tipo 8, AAV tipo 9, AAV tipo 10, AAV tipo 11, AAV tipo 12, AAV tipo 13, AAV de serpiente, AAV aviar, AAV bovino, AAV canino, AAV equino, AAV ovino, AAV de cabra, AAV de camarón, los serotipos y los clados de AAV descritos en Gao et al. ( J. Virol. 78: 6381 (2004)) y  
10 Moris et al. ( Virol. 33: 375 (2004)), y cualquier otro AAV conocido en la actualidad o descubierto posteriormente. Ver, por ejemplo, FIELDS et al. VIROLOGY, volumen 2, capítulo 69 (4ª ed., Lippincott-Raven Publishers).

**[0054]** La expresión "derivado de", como se usa en el presente documento, se refiere

a un componente que se aísla o se prepara usando una molécula u organismo  
15 específico, o información (por ejemplo, una secuencia de aminoácidos o ácido nucleico) de la molécula u organismo especificado. Por ejemplo, una secuencia de ácido nucleico (por ejemplo, ITR) que se deriva de una segunda secuencia de ácido nucleico (por ejemplo, ITR) puede incluir una secuencia de nucleótidos que sea idéntica o sustancialmente similar a la secuencia de nucleótidos de la segunda  
20 secuencia de ácido nucleico. En el caso de nucleótidos o polipéptidos, las especies derivadas pueden obtenerse, por ejemplo, mediante mutagénesis natural, mutagénesis dirigida artificial o mutagénesis aleatoria artificial. La mutagénesis utilizada para derivar nucleótidos o polipéptidos puede ser dirigida pretendida o intencionalmente al azar, o una mezcla de ambos. La mutagénesis de un nucleótido o  
25 polipéptido para crear un nucleótido o polipéptido diferente derivado del primero puede ser un evento aleatorio (p. ej., uno causado por la infidelidad de la polimerasa) y la identificación del nucleótido o polipéptido derivado puede realizarse mediante métodos de detección apropiados, p. ej., como se discute en este documento. La mutagénesis de un polipéptido típicamente implica la manipulación del polinucleótido  
30 que codifica el polipéptido. En algunas realizaciones, la secuencia de nucleótidos o aminoácidos que se deriva de una segunda secuencia de nucleótidos o aminoácidos tiene una identidad de secuencia de al menos 50%, al menos 51%, al menos 52%, al menos 53%, al menos 54%, al menos 55%, al menos 56%, al menos 57%, al menos 58%, al menos 59%, al menos 60%, al menos 61%, al menos 62%, al menos 63%, al  
35 menos 64%, al menos 65%, al menos 66%, al menos 67%, al menos 68%, al menos

69%, al menos 70%, al menos 71%, al menos 72%, al menos 73%, al menos 74 %, al menos 75%, al menos 76%, al menos 77%, al menos 78%, al menos 79%, al menos 80%, al menos 81%, al menos 82%, al menos 83%, al menos 84 %, al menos 85%, al menos 86%, al menos 87%, al menos 88%, al menos 89%, al menos 90%, al menos 91%, al menos 92%, al menos 93%, al menos 94%, al menos 95%, al menos 96%, al menos 97%, al menos 98%, al menos 99%, o 100% con respecto a la segunda secuencia de nucleótidos o aminoácidos, respectivamente, donde la primera secuencia de nucleótidos o aminoácidos retiene la actividad biológica de la segunda secuencia de nucleótidos o aminoácidos. En otras realizaciones, una ITR derivada de una ITR de un VnoAA (o AAV) es al menos un 90% idéntica con respecto a la ITR VnoAA (o ITR AAV, respectivamente), donde la ITR VnoAA (o AAV) retiene una propiedad funcional de la ITR de un VnoAA (o ITR AAV, respectivamente). En algunas realizaciones, una ITR derivada de una ITR de un VnoAA (o AAV) es al menos un 80% idéntica con respecto a la ITR VnoAA (o ITR AAV, respectivamente), donde la ITR VnoAA (o AAV) retiene una propiedad funcional de la ITR VnoAA (o ITR AAV, respectivamente). En algunas realizaciones, una ITR derivada de una ITR de un VnoAA (o AAV) es al menos un 70% idéntica con respecto a la ITR VnoAA (o ITR AAV, respectivamente), donde la ITR VnoAA (o AAV) retiene una propiedad funcional de la ITR VnoAA (o ITR AAV, respectivamente). En algunas realizaciones, la ITR derivada de una ITR de un VnoAA (o AAV) es al menos un 60% idéntico con respecto a la ITR VnoAA (o ITR AAV, respectivamente), donde la ITR VnoAA (o AAV) retiene una propiedad funcional de la ITR VnoAA (o ITR AAV, respectivamente). En algunas realizaciones, una ITR derivada de una ITR de un VnoAA (o AAV) es al menos 50% idéntico con respecto a la ITR VnoAA (o ITR AAV, respectivamente), donde la ITR VnoAA (o AAV) retiene una propiedad funcional de la ITR VnoAA (o ITR AAV, respectivamente).

**[0055]** En ciertas realizaciones, una ITR derivada de una ITR de un VnoAA (o AAV) comprende o consiste en un fragmento de la ITR del VnoAA (o AAV). En algunas realizaciones, la ITR derivada de una ITR de un VnoAA (o AAV) comprende o consiste en un fragmento de la ITR del VnoAA (o AAV), en el que el fragmento comprende al menos aproximadamente 5 nucleótidos, al menos aproximadamente 10 nucleótidos, al menos aproximadamente 15 nucleótidos, al menos aproximadamente 20 nucleótidos, al menos aproximadamente 25 nucleótidos, al menos aproximadamente 30 nucleótidos, al menos aproximadamente 35 nucleótidos, al menos aproximadamente 40 nucleótidos, al menos aproximadamente 45 nucleótidos, al

menos aproximadamente 50 nucleótidos, al menos aproximadamente 55 nucleótidos, al menos aproximadamente 60 nucleótidos, al menos aproximadamente 65 nucleótidos, al menos aproximadamente 70 nucleótidos, al menos aproximadamente 75 nucleótidos, al menos aproximadamente 80 nucleótidos, al menos aproximadamente 85 nucleótidos, al menos aproximadamente 90 nucleótidos, al menos aproximadamente 95 nucleótidos, al menos aproximadamente 100 nucleótidos, al menos aproximadamente 125 nucleótidos, al menos aproximadamente 150 nucleótidos, al menos aproximadamente 175 nucleótidos, al menos aproximadamente 200 nucleótidos, al menos aproximadamente 225 nucleótidos, al menos aproximadamente 250 nucleótidos, al menos aproximadamente 275 nucleótidos, al menos aproximadamente 300 nucleótidos, al menos aproximadamente 325 nucleótidos, al menos aproximadamente 350 nucleótidos, al menos aproximadamente 375 nucleótidos, al menos aproximadamente 400 nucleótidos, al menos aproximadamente 425 nucleótidos, al menos aproximadamente 450 nucleótidos, al menos aproximadamente 475 nucleótidos, al menos aproximadamente 500 nucleótidos, al menos aproximadamente 525 nucleótidos, al menos aproximadamente 550 nucleótidos, al menos aproximadamente 575 nucleótidos, o al menos aproximadamente 600 nucleótidos; donde la ITR derivada de una ITR de un VnoAA (o AAV) retiene una propiedad funcional de la ITR VnoAA (o AAV ITR, respectivamente). En ciertas realizaciones, la ITR derivada de una ITR de un VnoAA (o AAV) comprende o consiste en un fragmento de la ITR del VnoAA (o AAV), donde el fragmento comprende al menos aproximadamente 129 nucleótidos, y donde la ITR derivada de una ITR de un VnoAA (o AAV) retiene una propiedad funcional de la ITR VnoAA (o AAV ITR, respectivamente). En ciertas realizaciones, la ITR derivada de una ITR de un VnoAA (o AAV) comprende o consiste en un fragmento de la ITR del VnoAA (o AAV), donde el fragmento comprende al menos aproximadamente 102 nucleótidos, y donde la ITR derivada de una ITR de un VnoAA (o AAV) retiene una propiedad funcional de la ITR VnoAA (o AAV ITR, respectivamente).

**[0056]** En algunas realizaciones, la ITR derivada de una ITR de un VnoAA (o AAV) comprende o consiste en un fragmento de la ITR del VnoAA (o AAV), en donde el fragmento comprende al menos aproximadamente 5%, al menos aproximadamente 10%, al menos aproximadamente 15%, al menos aproximadamente 20%, al menos aproximadamente 25%, al menos aproximadamente 30%, al menos aproximadamente 35%, al menos aproximadamente 40%, al menos aproximadamente 45%, al menos aproximadamente 50%, al menos aproximadamente 55%, al menos aproximadamente

60%, al menos aproximadamente 65%, al menos aproximadamente 70%, al menos aproximadamente 75%, al menos aproximadamente 80%, al menos aproximadamente 85%, al menos aproximadamente 90%, al menos aproximadamente 95%, al menos aproximadamente 96%, al menos aproximadamente 97%, al menos aproximadamente 98%, o al menos aproximadamente 99% de la longitud de la ITR de los VnoAA (o AAV).

**[0057]** En ciertas realizaciones, una secuencia de nucleótidos o aminoácidos que se deriva de una segunda secuencia de nucleótidos o aminoácidos tiene una identidad de secuencia de al menos 50%, al menos 51%, al menos 52%, al menos 53%, al menos 54%, al menos 55%, al menos 56%, al menos 57%, al menos 58%, al menos 59%, al menos 60%, al menos 61%, al menos 62%, al menos 63%, al menos 64%, al menos 65%, al menos 66%, al menos 67%, al menos 68%, al menos 69%, al menos 70%, al menos 71%, al menos 72%, al menos 73%, al menos 74%, al menos 75%, al menos 76%, al menos 77%, al menos 78%, al menos 79%, al menos 80%, al menos 81%, al menos 82%, al menos 83%, al menos 84%, al menos 85%, al menos 86%, al menos 87%, al menos 88%, al menos 89%, al menos 90%, al menos 91%, al menos 92%, al menos 93%, al menos 94%, al menos 95%, al menos 96%, al menos 97%, al menos 98%, al menos 99% o 100% con respecto a una porción homóloga de la segunda secuencia de nucleótidos o aminoácidos, respectivamente, cuando está correctamente alineada, en donde el primer nucleótido o la secuencia de aminoácidos retiene la actividad biológica de la segunda secuencia de nucleótidos o aminoácidos. En otras realizaciones, la ITR derivada de una ITR de un VnoAA (o AAV) es al menos 90% idéntica con respecto a una porción homóloga de la ITR VnoAA (o AAV ITR, respectivamente), cuando está correctamente alineada, en donde la primera secuencia de nucleótidos o aminoácidos retiene la actividad biológica de la segunda secuencia de nucleótidos o aminoácidos. En algunas realizaciones, la ITR derivada de una ITR de un VnoAA (o AAV) es al menos 80% idéntica con respecto a una porción homóloga de la ITR VnoAA (o ITR AAV, respectivamente), cuando está correctamente alineada, en donde la primera secuencia de nucleótidos o aminoácidos retiene la actividad biológica de la segunda secuencia de nucleótidos o aminoácidos. En algunas realizaciones, la ITR derivada de una ITR de un VnoAA (o AAV) es al menos un 70% idéntica con respecto a una porción homóloga de la ITR VnoAA (o AAV ITR, respectivamente), cuando está correctamente alineada, en donde la primera secuencia de nucleótidos o aminoácidos retiene la actividad biológica de la segunda secuencia de nucleótidos o aminoácidos. En algunas realizaciones, una ITR derivada

de una ITR de un VnoAA (o AAV) es al menos 60% idéntica con respecto a una porción homóloga de la ITR VnoAA (o ITR AAV, respectivamente), cuando está correctamente alineada, en donde la primera secuencia de nucleótidos o aminoácidos retiene la actividad biológica de la segunda secuencia de nucleótidos o aminoácidos. En algunas realizaciones, la ITR derivada de una ITR de un VnoAA (o AAV) es al menos 50% idéntica con respecto a una porción homóloga de la ITR VnoAA (o ITR AAV, respectivamente), cuando está correctamente alineada, en donde la primera secuencia de nucleótidos o aminoácidos retiene la actividad biológica de la segunda secuencia de nucleótidos o aminoácidos.

**[0058]** Un vector o molécula de ácido nucleico "sin cápside" o "libre de cápside" se refiere a una construcción de vector libre de una cápside. En algunas realizaciones, el vector o la molécula de ácido nucleico sin cápside no contiene secuencias que codifiquen, por ejemplo, la proteína Rep AAV.

**[0059]** Como se usa en el presente documento, una "región codificante" o "secuencia codificante" es una porción de polinucleótido que consiste en codones traducibles en aminoácidos. Aunque un "codón de parada" (TAG, TGA o TAA) generalmente no se traduce en un aminoácido, puede considerarse parte de una región de codificación, pero cualesquiera secuencias de flaqueo, por ejemplo, promotores, sitios de unión a ribosomas, terminadores de la transcripción, intrones y similares, no son parte de la región de codificación. Los límites de una región de codificación se determinan típicamente mediante un codón de inicio en el terminal 5', que codifica el terminal amino del polipéptido resultante, y un codón de parada de la traducción en el terminal 3', que codifica el terminal carboxilo del polipéptido resultante. Dos o más regiones codificantes pueden estar presentes en una única construcción polinucleotídica, por ejemplo, en un solo vector, o en construcciones polinucleotídicas separadas, por ejemplo, en vectores separados (diferentes). Se deduce, entonces, que un solo vector puede contener una sola región de codificación, o comprender dos o más regiones de codificación.

**[0060]** Ciertas proteínas secretadas por las células de mamíferos se asocian con un péptido señal secretor que se escinde de la proteína madura una vez que se ha iniciado la exportación de la cadena proteica en crecimiento a través del retículo endoplásmico rugoso. Los expertos en la materia son conscientes de que los péptidos señal generalmente se fusionan con el extremo N del polipéptido y se escinden del polipéptido completo o de "longitud completa" para producir una forma secretada o "madura" del polipéptido. En ciertas realizaciones, el péptido señal nativo o un

derivado funcional de esa secuencia que retiene la capacidad de dirigir la secreción del polipéptido que está operativamente asociado con él. Alternativamente, se puede usar un péptido señal de mamífero heterólogo, por ejemplo, un activador de plasminógeno tisular humano (TPA) o un péptido señal de  $\beta$ -glucuronidasa de ratón B, o un derivado funcional del mismo.

**[0061]**El término "posterior" se refiere a una secuencia de nucleótidos que se encuentra 3' con respecto a una secuencia de nucleótidos de referencia. En ciertas realizaciones, las secuencias de nucleótidos posteriores se relacionan con las secuencias que siguen el punto de inicio de la transcripción. Por ejemplo, el codón de iniciación de la traducción de un gen se encuentra posterior al sitio de inicio de la transcripción.

**[0062]**El término "anterior" se refiere a una secuencia de nucleótidos que se encuentra 5' con respecto a una secuencia de nucleótidos de referencia. En ciertas realizaciones, las secuencias de nucleótidos anteriores se relacionan con secuencias que están ubicadas en el lado 5' de una región de codificación o punto de inicio de la transcripción. Por ejemplo, la mayoría de los promotores se encuentran anteriores al sitio de inicio de la transcripción.

**[0063]**Como se usa en el presente documento, la expresión "región reguladora de genes" o "región reguladora" se refiere a secuencias de nucleótidos ubicadas anterior (secuencias no codificantes 5'), en la misma o después (secuencias no codificantes 3') de una región codificante, y que influye en la transcripción, procesamiento de ARN, estabilidad o traducción de la región de codificación asociada. Las regiones reguladoras pueden incluir promotores, secuencias líderes de traducción, intrones, secuencias de reconocimiento de poliadenilación, sitios de procesamiento de ARN, sitios de unión a efectores o estructuras de bucle de tallo. Si una región de codificación está destinada a la expresión en una célula eucariota, la señal de poliadenilación y la secuencia de terminación de la transcripción generalmente se ubicarán en 3' de la secuencia de codificación.

**[0064]**Un polinucleótido que codifica un producto, por ejemplo, un ARNm o un producto génico (por ejemplo, un polipéptido tal como una proteína terapéutica), puede incluir un promotor y/u otros elementos del control de la expresión (por ejemplo, transcripción o traducción) asociados operativamente con una o más regiones de codificación. En una asociación operativa, una región codificante para un producto génico, por ejemplo, un polipéptido, está asociada con una o más regiones reguladoras de tal manera que la expresión del producto génico esté bajo la influencia

o el control de la región o las regiones reguladoras. Por ejemplo, una región de codificación y un promotor están "operativamente asociados" si la inducción de la función del promotor da como resultado la transcripción de un ARNm que codifica el producto génico codificado por la región de codificación, y si la naturaleza del enlace entre el promotor y la región de codificación lo hace no interferir con la capacidad del promotor para dirigir la expresión del producto génico o interferir con la capacidad de transcripción de la plantilla de ADN. Otros elementos de control de la expresión, además de un promotor, por ejemplo los potenciadores, operadores, represores y señales de terminación de la transcripción, también pueden asociarse operativamente con una región de codificación para dirigir la expresión del producto génico.

**[0065]** Las "secuencias de control transcripcional" se refieren a secuencias reguladoras de ADN, tales como promotores, potenciadores, terminadores y similares, que proporcionan la expresión de una secuencia codificante en una célula huésped. Los expertos en la materia conocen una variedad de regiones de control de la transcripción. Estos incluyen, sin limitación, regiones de control de la transcripción que funcionan en células de vertebrados, tales como, entre otros, segmentos promotores y potenciadores de citomegalovirus (el promotor temprano inmediato, junto con intrón-A), virus simio 40 (el promotor temprano) y retrovirus (como el virus del sarcoma de Rous). Otras regiones de control de la transcripción incluyen aquellas derivadas de genes de vertebrados tales como actina, proteína de choque térmico, hormona del crecimiento bovino y  $\beta$ -globina de conejo, así como otras secuencias capaces de controlar la expresión génica en células eucariotas. Las regiones de control de la transcripción adecuadas adicionales incluyen promotores y potenciadores específicos de tejido, así como promotores inducibles por linfocina (por ejemplo, promotores inducibles por interferones o interleucinas).

**[0066]** Del mismo modo, los expertos en la materia conocen una variedad de elementos de control de la traducción. Estos incluyen, entre otros, sitios de unión a ribosomas, codones de iniciación y terminación de la traducción, y elementos derivados de picornavirus (particularmente un sitio interno de entrada de ribosomas, o IRES, también denominado secuencia CITE).

**[0067]** El término "expresión", como se usa en el presente documento, se refiere a un proceso mediante el cual un polinucleótido produce un producto génico, por ejemplo, un ARN o un polipéptido. Incluye, sin limitación, la transcripción del polinucleótido en ARN mensajero (ARNm), ARN de transferencia (ARNt), ARN pequeño de horquilla (shRNA), ARN interferente pequeño (ARNip) o cualquier otro producto de ARN, y la

traducción de un ARNm en un polipéptido. La expresión produce un "producto genético". Como se usa en el presente documento, el producto génico puede ser un ácido nucleico, por ejemplo, un ARN mensajero producido por la transcripción de un gen, o un polipéptido que se traduzca desde una transcripción. Los productos génicos descritos en el presente documento incluyen además ácidos nucleicos con modificaciones posteriores a la transcripción, por ejemplo, poliadenilación o empalme, o polipéptidos con modificaciones posteriores a la traducción, por ejemplo, metilación, glucosilación, adición de lípidos, asociación con otras subunidades de proteínas o escisión proteolítica. El término "rendimiento", como se usa en el presente documento, se refiere a la cantidad de un polipéptido producido por la expresión de un gen.

**[0068]** Un "vector" se refiere a cualquier vehículo para la clonación y/o transferencia de un ácido nucleico en una célula huésped. Un vector puede ser un replicón al que se puede unir otro segmento de ácido nucleico para provocar la replicación del segmento unido. Un "replicón" se refiere a cualquier elemento genético (por ejemplo, un plásmido, fago, cósmido, cromosoma, virus) que funciona como una unidad autónoma de replicación in vivo, es decir, capaz de replicarse bajo su propio control. El término "vector" incluye vehículos para introducir el ácido nucleico en una célula in vitro, ex vivo o in vivo. Se conocen y se usan una gran cantidad de vectores en la técnica que incluyen, por ejemplo, plásmidos, virus eucariotas modificados o virus bacterianos modificados. La inserción de un polinucleótido en un vector adecuado se puede lograr ligando los fragmentos de polinucleótido apropiados en un vector elegido con terminales cohesivos complementarios.

**[0069]** Los vectores se pueden diseñar para codificar marcadores o reporteros seleccionables que proporcionen la selección o identificación de las células que han incorporado el vector. La expresión de marcadores o reporteros seleccionables permite la identificación y/o selección de células huésped que incorporan y expresan otras regiones de codificación contenidas en el vector. Los ejemplos de genes marcadores seleccionables conocidos y utilizados en la técnica incluyen: genes que proporcionan resistencia a la ampicilina, estreptomicina, gentamicina, kanamicina, higromicina, herbicida de bialafos, sulfonamida y similares; y genes que se usan como marcadores fenotípicos, es decir, genes reguladores de antocianinas, gen de isopentanol transferasa y similares. Los ejemplos de reporteros conocidos y utilizados en la técnica incluyen: luciferasa (Luc), proteína verde fluorescente (GFP), cloranfenicol acetiltransferasa (CAT),  $\beta$ -galactosidasa (LacZ),  $\beta$ -glucuronidasa (Gus) y similares. Los marcadores seleccionables también pueden considerarse reporteros.

**[0070]** La expresión "célula huésped", como se usa en el presente documento, se refiere, por ejemplo, a microorganismos, células de levadura, células de insecto y células de mamífero, que pueden usarse o han sido utilizadas como receptores de ADN monocatenario o vectores. El término incluye la progenie de la célula original que ha sido transducida. Por lo tanto, una "célula huésped", como se usa en el presente documento, generalmente se refiere a una célula que se ha transducido con una secuencia de ADN exógeno. Se entiende que la progenie de una sola célula parental puede no ser necesariamente completamente idéntica en morfología o en complemento de ADN genómico o total del progenitor original, debido a una mutación natural, accidental o deliberada. En algunas realizaciones, la célula huésped puede ser una célula huésped in vitro.

**[0071]** La expresión "marcador seleccionable" se refiere a un factor de identificación, generalmente un gen de resistencia a antibióticos o agentes químicos, que puede seleccionarse en función del efecto del gen marcador, es decir, resistencia a un antibiótico, resistencia a un herbicida, marcadores colorimétricos, enzimas, marcadores fluorescentes, y similares, en los que el efecto se usa para rastrear la herencia de un ácido nucleico de interés y/o para identificar una célula u organismo que ha heredado el ácido nucleico de interés. Los ejemplos de genes marcadores seleccionables conocidos y utilizados en la técnica incluyen: genes que proporcionan resistencia a la ampicilina, estreptomicina, gentamicina, kanamicina, higromicina, herbicida de bialafos, sulfonamida y similares; y genes que se usan como marcadores fenotípicos, es decir, genes reguladores de antocianinas, gen de isopentanol transferasa y similares.

**[0072]** La expresión "gen informador" se refiere a un ácido nucleico que codifica un factor de identificación que puede identificarse en función del efecto del gen informador, en el que el efecto se utiliza para rastrear la herencia de un ácido nucleico de interés, para identificar una célula u organismo que ha heredado el ácido nucleico de interés y/o para medir la inducción o transcripción de la expresión génica. Los ejemplos de genes informadores conocidos y utilizados en la técnica incluyen: luciferasa (Luc), proteína verde fluorescente (GFP), cloranfenicol acetiltransferasa (CAT),  $\beta$ -galactosidasa (LacZ),  $\beta$ -glucuronidasa (Gus) y similares. Los genes marcadores seleccionables también pueden considerarse genes informadores.

**[0073]** "Promotor" y "secuencia promotora" se usan indistintamente y se refieren a una secuencia de ADN capaz de controlar la expresión de una secuencia codificante o ARN funcional. En general, la secuencia de codificación se encuentra 3' con respecto

a una secuencia promotora. Los promotores pueden derivarse en su totalidad de un gen nativo, o estar compuestos de diferentes elementos derivados de diferentes promotores encontrados en la naturaleza, o incluso comprender segmentos de ADN sintéticos. Los expertos en la materia entienden que diferentes promotores pueden

5 dirigir la expresión de un gen en diferentes tejidos o tipos de células, o en diferentes etapas de desarrollo, o en respuesta a diferentes condiciones ambientales o fisiológicas. Los promotores que hacen que un gen se exprese en la mayoría de los tipos de células en la mayoría de los casos se denominan comúnmente "promotores constitutivos". Los promotores que hacen que un gen se exprese en un tipo celular

10 específico se denominan comúnmente "promotores específicos de células" o "promotores específicos de tejidos". "Los promotores que hacen que un gen se exprese en una etapa específica del desarrollo o de la diferenciación celular se denominan comúnmente "promotores específicos del desarrollo" o "promotores específicos de la diferenciación celular". Los promotores que son inducidos y hacen

15 que un gen se exprese después de la exposición o el tratamiento de la célula con un agente, molécula biológica, agente químico, ligando, luz o similar que induce al promotor se conocen comúnmente como "promotores inducibles" o "promotores regulables". Se reconoce además que, dado que en la mayoría de los casos los límites exactos de las secuencias reguladoras no se han definido completamente, los

20 fragmentos de ADN de diferentes longitudes pueden tener una actividad promotora idéntica.

**[0074]** La secuencia promotora está típicamente limitada en su terminal 3' por el sitio de inicio de la transcripción y se extiende antes (dirección 5') para incluir el número mínimo de bases o elementos necesarios para iniciar la transcripción a niveles

25 detectables por encima de la línea de fondo. Dentro de la secuencia promotora se encontrará un sitio de inicio de la transcripción (convenientemente definido, por ejemplo, mediante el mapeo con nucleasa S1), así como dominios de unión a proteínas (secuencias consenso) responsables de la unión de la ARN polimerasa.

**[0075]** En algunas realizaciones, la molécula de ácido nucleico comprende un

30 promotor específico de tejido. En ciertas realizaciones, el promotor específico de tejido conduce la expresión de la proteína terapéutica, por ejemplo, el factor de coagulación, en el hígado, por ejemplo, en hepatocitos y/o células endoteliales. En particular, en las realizaciones, el promotor se selecciona del grupo que consiste en un promotor de tiretina de ratón (mTTR), un promotor endógeno del factor VIII humano (F8), un

35 promotor de alfa-1 antitripsina humana (hAAT), un promotor mínimo de albúmina

humana, un promotor de albúmina de ratón, un promotor de tristetraprolina (TTP), un promotor CASI, un promotor CAG, un promotor de citomegalovirus (CMV), un promotor de fosfoglicerato quinasa (PGK) y cualquiera de sus combinaciones. En algunas realizaciones, el promotor se selecciona de un promotor específico del hígado (por ejemplo,  $\alpha$ 1-antitripsina (AAT)), un promotor específico del músculo (por ejemplo, creatina quinasa muscular (MCK), cadena alfa pesada de miosina ( $\alpha$ MHC), mioglobina (MB) y desmina (DES)), un promotor sintético (por ejemplo, SPc5-12, 2R5Sc5-12, dMCK y tMCK) y cualquiera de sus combinaciones. En una realización particular, el promotor comprende un promotor TTP.

**[0076]** Las expresiones "endonucleasa de restricción" y "enzima de restricción" se usan indistintamente y se refieren a una enzima que se une y corta dentro de una secuencia de nucleótidos específica dentro del ADN bicatenario.

**[0077]** El término "plásmido" se refiere a un elemento extracromosómico que a menudo lleva un gen que no forma parte del metabolismo central de la célula, y que generalmente está en forma de moléculas circulares de ADN bicatenario. Dichos elementos pueden ser secuencias que se replican de forma autónoma, secuencias integradoras del genoma, secuencias de fagos o nucleótidos, lineales, circulares o superenrolladas, de un ADN o ARN monocatenario o bicatenario, derivado de cualquier fuente, en donde se han producido varias secuencias de nucleótidos unidas o recombinadas en una construcción única capaz de introducir un fragmento de promotor y una secuencia de ADN para un producto genético seleccionado junto con la secuencia no traducida 3' apropiada en una célula.

**[0078]** Los vectores virales eucariotas que se pueden usar incluyen, entre otros, vectores de adenovirus, vectores de retrovirus, vectores de virus adenoasociados, virus de la viruela, por ejemplo, vectores del virus vaccinia, vectores de baculovirus o vectores del virus del herpes. Los vectores no virales incluyen plásmidos, liposomas, lípidos cargados eléctricamente (citofectinas), complejos de ADN-proteína y biopolímeros.

**[0079]** Un "vector de clonación" se refiere a un "replicón", que es una unidad de longitud de un ácido nucleico que se replica secuencialmente y que comprende un origen de replicación, como un plásmido, fago o cósmido, al que se puede unir otro segmento de ácido nucleico para lograr la replicación del segmento adjunto. Ciertos vectores de clonación son capaces de replicarse en un tipo de célula, por ejemplo, bacterias y expresión en otro, por ejemplo, células eucariotas. Los vectores de clonación comprenden típicamente una o más secuencias que pueden usarse para la

selección de células que comprenden el vector y/o uno o más sitios de clonación múltiples para la inserción de secuencias de ácido nucleico de interés.

**[0080]** La expresión "vector de expresión" se refiere a un vehículo diseñado para permitir la expresión de una secuencia de ácido nucleico insertada después de la inserción en una célula huésped. La secuencia de ácido nucleico insertada se coloca en asociación operativa con regiones reguladoras como se describió anteriormente.

**[0081]** Los vectores se introducen en las células huésped mediante métodos bien conocidos en la técnica, por ejemplo, transfección, electroporación, microinyección, transducción, fusión celular, DEAE dextrano, precipitación con fosfato de calcio, lipofección (fusión de lisosomas), uso de una pistola génica o un transportador de un vector de ADN. "Cultivo", "cultivar" y "cultivando", como se usan en el presente documento, significa incubar células en condiciones in vitro que permiten el crecimiento o la división celular o mantener células en un estado vivo. "Células cultivadas", como se usa en el presente documento, significa células que se propagan in vitro.

**[0082]** Como se usa en el presente documento, el término "polipéptido" pretende abarcar un "polipéptido" singular, así como "polipéptidos" plurales, y se refiere a una molécula compuesta de monómeros (aminoácidos) unidos linealmente por enlaces amida (también conocidos como enlaces peptídicos). El término "polipéptido" se refiere a cualquier cadena o cadenas de dos o más aminoácidos, y no se refiere a una longitud específica del producto. Por lo tanto, los péptidos, dipéptidos, tripéptidos, oligopéptidos, "proteína", "cadena de aminoácidos" o cualquier otro término utilizado para referirse a una cadena o cadenas de dos o más aminoácidos, se incluyen dentro de la definición de "polipéptido", y el término "polipéptido" puede usarse en lugar de, o de manera intercambiable con cualquiera de estos términos. El término "polipéptido" también se refiere a los productos de modificaciones posteriores a la expresión del polipéptido, que incluyen, sin limitación, glicosilación, acetilación, fosforilación, amidación, derivación por grupos protectores/bloqueadores conocidos, escisión proteolítica o modificación por aminoácidos presentes no naturales. Un polipéptido puede derivarse de una fuente biológica natural o producirse por tecnología recombinante, pero no se traduce necesariamente a partir de una secuencia de ácido nucleico designada. Se puede generar de cualquier manera, incluso mediante síntesis química.

**[0083]** El término "aminoácido" incluye alanina (Ala o A); arginina (Arg o R); asparagina (Asn o N); ácido aspártico (Asp o D); cisteína (Cys o C); glutamina (Gln o Q); ácido

glutámico (Glu o E); glicina (Gly o G); histidina (His o H); isoleucina (Ile o I); leucina (Leu o L); lisina (Lys o K); metionina (Met o M); fenilalanina (Phe o F); prolina (Pro o P); serina (Ser o S); treonina (Thr o T); triptófano (Trp o W); tirosina (Tyr o Y); y valina (Val o V). Los aminoácidos no tradicionales también están dentro del alcance de la descripción e incluyen norleucina, ornitina, norvalina, homoserina y otros análogos de residuos de aminoácidos tales como los descritos en Ellman et al. Meth. Enzym 202:301-336 (1991). Para generar tales residuos de aminoácidos no naturales, pueden usarse los procedimientos de Noren et al. Science 244: 182 (1989) y Ellman et al., supra. Brevemente, estos procedimientos implican la activación química de un ARNt supresor con un residuo de aminoácido no natural seguido de la transcripción y traducción in vitro del ARN. La introducción del aminoácido no tradicional también se puede lograr usando químicas de péptidos conocidas en la técnica. Como se usa en este documento, la expresión "aminoácido polar" incluye aminoácidos que tienen carga neta cero, pero con cargas parciales distintas de cero en diferentes porciones de sus cadenas laterales (por ejemplo, M, F, W, S, Y, N, Q, C). Estos aminoácidos pueden participar en interacciones hidrofóbicas e interacciones electrostáticas. Como se usa en el presente documento, la expresión "aminoácido cargado" incluye aminoácidos que pueden tener una carga neta distinta de cero en sus cadenas laterales (por ejemplo, R, K, H, E, D). Estos aminoácidos pueden participar en interacciones hidrofóbicas e interacciones electrostáticas.

**[0084]** También se incluyen en la presente descripción fragmentos o variantes de polipéptidos, y cualquiera de sus combinaciones. El término "fragmento" o "variante", cuando se refiere a dominios de unión a polipéptidos o moléculas de unión de la presente descripción, incluye cualquier polipéptido que retenga al menos algunas de las propiedades (por ejemplo, la afinidad de unión a FcRn por un dominio de unión a FcRn o variante de Fc, la actividad de coagulación para una variante de FVIII, o la actividad de unión a FVIII para el fragmento de VWF) del polipéptido de referencia. Los fragmentos de polipéptidos incluyen fragmentos proteolíticos, así como fragmentos de delección, además de fragmentos de anticuerpos específicos discutidos en otra parte de este documento, pero no incluyen el polipéptido natural completo (o polipéptido maduro). Las variantes de dominios de unión a polipéptidos o moléculas de unión de la presente descripción incluyen fragmentos como se describe anteriormente y también polipéptidos con secuencias de aminoácidos modificadas debido a sustituciones, delecciones o inserciones de aminoácidos. Las variantes pueden ser naturales o no naturales. Se pueden producir variantes no naturales

utilizando técnicas de mutagénesis conocidas en la técnica. Los polipéptidos variantes pueden comprender sustituciones, deleciones o adiciones de aminoácidos conservativas o no conservativas.

**[0085]** Una "sustitución conservadora de aminoácidos" es aquella en donde el residuo de aminoácido se reemplaza con un residuo de aminoácido que tiene una cadena lateral similar. Las familias de residuos de aminoácidos que tienen cadenas laterales similares se han definido en la técnica, incluidas cadenas laterales básicas (por ejemplo, lisina, arginina, histidina), cadenas laterales ácidas (por ejemplo, ácido aspártico, ácido glutámico), cadenas laterales polares sin carga (por ejemplo, glicina, asparagina, glutamina, serina, treonina, tirosina, cisteína), cadenas laterales no polares (p. ej., alanina, valina, leucina, isoleucina, prolina, fenilalanina, metionina, triptófano), cadenas laterales ramificadas beta (p. ej., treonina, valina, isoleucina) y cadenas laterales aromáticas (p. ej., tirosina, fenilalanina, triptófano, histidina). Por lo tanto, si un aminoácido en un polipéptido se reemplaza con otro aminoácido de la misma familia de cadenas laterales, la sustitución se considera conservadora. En otra realización, una cadena de aminoácidos se puede reemplazar de forma conservadora por una cadena estructuralmente similar que difiera en orden y/o composición de los miembros de la familia de la cadena lateral.

**[0086]** La expresión "porcentaje de identidad" como se conoce en la técnica, es una relación entre dos o más secuencias de polipéptidos o dos o más secuencias de polinucleótidos, según se determina mediante la comparación de las secuencias. En la técnica, "identidad" también significa el grado de relación de secuencia entre secuencias de polipéptidos o polinucleótidos, según sea el caso, según lo determinado por la coincidencia entre las cadenas de tales secuencias. La "identidad" se puede calcular fácilmente mediante métodos conocidos, que incluyen, entre otros, los descritos en: Computational Molecular Biology (Lesk, A. M., ed.) Oxford University Press, Nueva York (1988); Biocomputing: Informatics and Genome Projects (Smith, D. W., ed.) Academic Press, New York (1993); Computer Analysis of Sequence Data, Part I (Griffin, A. M., and Griffin, H. G., eds.) Humana Press, New Jersey (1994); Sequence Analysis in Molecular Biology (von Heinje, G., ed.) Academic Press (1987); y Sequence Analysis Primer (Gribskov, M. and Devereux, J., eds.) Stockton Press, New York (1991). Los métodos preferidos para determinar la identidad están diseñados para dar la mejor coincidencia entre las secuencias probadas. Los métodos para determinar la identidad están codificados en programas informáticos disponibles al público. Las alineaciones de secuencias y los cálculos de porcentaje de identidad

se pueden realizar utilizando un software de análisis de secuencias como el programa Megalign del conjunto de computación bioinformática LASERGENE (DNASTAR Inc., Madison, WI), el conjunto de programas GCG (Wisconsin Package Version 9.0, Genetics Computer Group (GCG), Madison, WI), BLASTP, BLASTN, BLASTX (Altschul et al., J. Mol. Biol. 275: 403 (1990)) y DNASTAR (DNASTAR, Inc. 1228 S. Park St. Madison, WI 53715 EE. UU.). Dentro del contexto de esta aplicación, se entenderá que cuando el software de análisis de secuencia se utilice para el análisis, los resultados del análisis se basarán en los "valores predeterminados" del programa al que se hace referencia, a menos que se especifique lo contrario. Como se usa en este documento, "valores predeterminados" significará cualquier conjunto de valores o parámetros que originalmente se cargan con el software cuando este se inicia por primera vez. A los efectos de determinar el porcentaje de identidad entre una proteína terapéutica, por ejemplo, un factor de coagulación, la secuencia de la descripción y una secuencia de referencia, solo los nucleótidos en la secuencia de referencia correspondiente a los nucleótidos en la proteína terapéutica, por ejemplo, el factor de coagulación, la secuencia de la descripción se utiliza para calcular el porcentaje de identidad. Por ejemplo, cuando se compara una secuencia de nucleótidos de FVIII de longitud completa que contiene el dominio B con una secuencia de nucleótidos de FVIII eliminada del dominio B (BDD) optimizada de la descripción, la porción de la alineación que incluye el dominio A1, A2, A3, C1 y C2 será usado para calcular el porcentaje de identidad. Los nucleótidos en la porción de la secuencia de FVIII de longitud completa que codifica el dominio B (que dará como resultado un gran "espacio" en la alineación) no se contarán como una falta de coincidencia. Además, al determinar el porcentaje de identidad entre una secuencia BDD FVIII optimizada de la descripción, o una porción designada de la misma (por ejemplo, los nucleótidos 58-2277 y 2320-4374 de la SEQ ID NO: 3), y una secuencia de referencia, el porcentaje de identidad será calculado alineando y dividiendo el número de nucleótidos coincidentes por el número total de nucleótidos en la secuencia completa de la secuencia BDD-FVIII optimizada, o una porción designada de la misma, como se menciona en este documento.

**[0087]** Como se usa en el presente documento, los nucleótidos correspondientes a los nucleótidos en una secuencia particular de la descripción se identifican mediante la alineación de la secuencia de la descripción para maximizar la identidad de una secuencia de referencia. El número utilizado para identificar un aminoácido

equivalente en una secuencia de referencia se basa en el número utilizado para identificar el aminoácido correspondiente en la secuencia de la descripción.

**[0088]** Una proteína de "fusión" o "quimérica" comprende una primera secuencia de aminoácidos unida a una segunda secuencia de aminoácidos con la que no está unida de forma natural en la naturaleza. Las secuencias de aminoácidos que normalmente existen en proteínas separadas se pueden unir en el polipéptido de fusión, o las secuencias de aminoácidos que normalmente existen en la misma proteína se pueden colocar en una nueva disposición en el polipéptido de fusión, por ejemplo, la fusión de un dominio del Factor VIII de la descripción con un dominio Ig Fc. Se crea una proteína de fusión, por ejemplo, mediante síntesis química, o creando y traduciendo un polinucleótido en el que las regiones peptídicas se codifican en la relación deseada. Una proteína quimérica puede comprender además una segunda secuencia de aminoácidos asociada con la primera secuencia de aminoácidos mediante un enlace covalente, no peptídico o un enlace no covalente.

**[0089]** Como se usa en el presente documento, la expresión "sitio de inserción" se refiere a una posición en un polipéptido, o fragmento, variante o derivado del mismo, que está inmediatamente antes de la posición en donde se puede insertar un resto heterólogo. Un "sitio de inserción" se especifica como un número, siendo dicho número el número del aminoácido en una secuencia de referencia. Por ejemplo, un "sitio de inserción" en FVIII se refiere al número de la secuencia de aminoácidos en el FVIII nativo maduro (SEQ ID NO: 15) al que corresponde el sitio de inserción, que es inmediatamente N-terminal a la posición de la inserción. Por ejemplo, la frase "a3 comprende un resto heterólogo en un sitio de inserción que corresponde al aminoácido 1656 de la SEQ ID N°: 15" indica que el resto heterólogo está ubicado entre dos aminoácidos correspondientes al aminoácido 1656 y al aminoácido 1657 de la SEQ ID NO: 15.

**[0090]** La frase "inmediatamente después de un aminoácido", como se usa en el presente documento, se refiere a la posición justo al lado del grupo carboxilo terminal del aminoácido. De manera similar, la frase "inmediatamente antes de un aminoácido" se refiere a la posición justo al lado del grupo amino terminal del aminoácido.

**[0091]** Los términos "insertado", "se inserta", "insertado en" u otras expresiones gramaticalmente relacionadas, como se usan en el presente documento, se refieren a la posición de un resto heterólogo en un polipéptido, por ejemplo, un factor de coagulación, en relación con la posición análoga en el polipéptido parental. Por ejemplo, en cierta realización, "insertado" y similares se refieren a la posición de un

resto heterólogo en un polipéptido de FVIII recombinante, en relación con la posición análoga en un FVIII humano nativo maduro. Como se usa en el presente documento, los términos se refieren a las características del polipéptido y no indican, implican ni infieren ningún método o proceso por el cual se hizo el polipéptido.

5 **[0092]** Como se usa en el presente documento, el término "semivida" se refiere a una semivida biológica de un polipéptido particular in vivo. La semivida puede representarse por el tiempo requerido para que la mitad de la cantidad administrada a un sujeto se elimine de la circulación y/u otros tejidos en el animal. Cuando una curva de aclaramiento de un polipéptido dado se construye en función del tiempo, la curva  
10 suele ser bifásica con una fase  $\alpha$  rápida y una fase  $\beta$  más larga. La fase  $\alpha$  típicamente representa un equilibrio del polipéptido Fc administrado entre el espacio intravascular y extravascular y está, en parte, determinado por el tamaño del polipéptido. La fase  $\beta$  típicamente representa el catabolismo del polipéptido en el espacio intravascular. En algunas realizaciones, la proteína terapéutica, por ejemplo, el factor de coagulación,  
15 por ejemplo, FVIII, y las proteínas quiméricas que las comprenden son monofásicas y, por lo tanto, no tienen una fase alfa, sino solo la fase beta única. Por lo tanto, en ciertas realizaciones, el término semivida como se usa en el presente documento se refiere a la semivida del polipéptido en la fase  $\beta$ .

**[0093]** El término "unido", como se usa en el presente documento, se refiere a una  
20 primera secuencia de aminoácidos o secuencia de nucleótidos unida covalentemente o no covalentemente a una segunda secuencia de aminoácidos o secuencia de nucleótidos, respectivamente. La primera secuencia de aminoácidos o nucleótidos puede unirse directamente o yuxtaponerse a la segunda secuencia de aminoácidos o nucleótidos o, alternativamente, una secuencia intermedia puede unir covalentemente  
25 la primera secuencia a la segunda secuencia. El término "unido" significa no solo una fusión de una primera secuencia de aminoácidos con una segunda secuencia de aminoácidos en el C-terminal o el N-terminal, sino que también incluye la inserción de la primera secuencia completa de aminoácidos (o la segunda secuencia de aminoácidos) en cualquiera de los dos aminoácidos en la segunda secuencia de  
30 aminoácidos (o la primera secuencia de aminoácidos, respectivamente). En una realización, la primera secuencia de aminoácidos se puede unir a una segunda secuencia de aminoácidos mediante un enlace peptídico o un enlazante. La primera secuencia de nucleótidos puede unirse a una segunda secuencia de nucleótidos mediante un enlace fosfodiéster o un enlazante. El enlazante puede ser un péptido o  
35 un polipéptido (para cadenas de polipéptidos) o un nucleótido o una cadena de

nucleótidos (para cadenas de nucleótidos) o cualquier resto químico (tanto para cadenas de polipéptidos como de polinucleótidos). El término "unido" también se indica mediante un guión (-).

**[0094]** Hemostasia, como se usa en este documento, significa detener o ralentizar el sangrado o la hemorragia; o detener o ralentizar el flujo sanguíneo a través de un vaso sanguíneo o parte del cuerpo.

**[0095]** Un trastorno hemostático, como se usa en el presente documento, significa una afección genéticamente heredada o adquirida caracterizada por una tendencia a la hemorragia, ya sea espontáneamente o como resultado de un traumatismo, debido a una capacidad deteriorada o incapacidad para formar un coágulo de fibrina. Ejemplos de tales trastornos incluyen las hemofilias. Las tres formas principales son la hemofilia A (deficiencia del factor VIII), la hemofilia B (deficiencia del factor IX o "enfermedad de Navidad") y la hemofilia C (deficiencia del factor XI, tendencia leve a la hemorragia). Otros trastornos hemostáticos incluyen, por ejemplo, la enfermedad de von Willebrand, deficiencia del factor XI (deficiencia de PTA), deficiencia del factor XII, deficiencias o anomalías estructurales en fibrinógeno, protrombina, factor V, factor VII, factor X o factor XIII, síndrome de Bernard-Soulier, que es un defecto o una deficiencia en GPIb. GPIb, el receptor para vWF, puede ser defectuoso y conducir a la falta de formación de coágulos primarios (hemostasia primaria) y una mayor tendencia a la hemorragia), y trombastenia de Glanzman y Naegeli (trombastenia de Glanzmann). En la insuficiencia hepática (formas aguda y crónica), el hígado produce una insuficiente cantidad de factores de coagulación; esto puede aumentar el riesgo de sangrado.

**[0096]** Las moléculas de ácido nucleico aisladas, los polipéptidos aislados o los vectores que comprenden la molécula de ácido nucleico aislada de la descripción pueden usarse profilácticamente. Como se usa en el presente documento, la expresión "tratamiento profiláctico" se refiere a la administración de una molécula antes de un episodio de sangrado. En una realización, el sujeto que necesita un agente hemostático general está siendo sometido o está a punto de someterse a una cirugía. Se puede administrar un polinucleótido, polipéptido o vector de la descripción antes o después de la cirugía como profiláctico. El polinucleótido, polipéptido o vector de la descripción se puede administrar durante o después de la cirugía para controlar un episodio de sangrado agudo. La cirugía puede incluir, entre otros, trasplante de hígado, resección hepática, procedimientos dentales o trasplante de células madre.

**[0097]** Las moléculas de ácido nucleico aisladas, los polipéptidos aislados o los vectores de la descripción también se usan para el tratamiento a demanda. La expresión "tratamiento a demanda" se refiere a la administración de una molécula de ácido nucleico aislada, un polipéptido aislado o un vector en respuesta a los síntomas de un episodio de sangrado o antes de una actividad que pueda causar sangrado. En un aspecto, el tratamiento a demanda se puede administrar a un sujeto cuando comienza el sangrado, tal como después de una lesión, o cuando se espera sangrado, tal como antes de una cirugía. En otro aspecto, el tratamiento a demanda se puede administrar antes de actividades que aumenten el riesgo de sangrado, como los deportes de contacto.

**[0098]** Como se usa en el presente documento, la expresión "sangrado agudo" se refiere a un episodio de sangrado independientemente de la causa subyacente. Por ejemplo, un sujeto puede tener un trauma, uremia, un trastorno hemorrágico hereditario (por ejemplo, deficiencia del factor VII), un trastorno plaquetario o resistencia debida al desarrollo de anticuerpos contra los factores de coagulación.

**[0099]** Tratar, tratamiento, trato, tal como se usan en el presente documento, se refieren, por ejemplo, a la reducción en la gravedad de una enfermedad o afección; la reducción en la duración de un curso de la enfermedad; la mejora de uno o más síntomas asociados con una enfermedad o afección; la provisión de efectos beneficiosos para un sujeto con una enfermedad o afección, sin curar necesariamente la enfermedad o afección, o la profilaxis de uno o más síntomas asociados con una enfermedad o afección. En una realización, el término "tratar" o "tratamiento" significa mantener, por ejemplo, un nivel mínimo de FVIII de al menos aproximadamente 1 UI/dL, 2 UI/dL, 3 UI/dL, 4 UI/dL, 5 UI/dL, 6 UI/dL, 7 UI/dL, 8 UI/dL, 9 UI/dL, 10 UI/dL, 11 UI/dL, 12 UI/dL, 13 UI/dL, 14 UI/dL, 15 UI/dL, 16 UI/dL, 17 UI/dL, 18 UI/dL, 19 UI/dL, 20 UI/dL, 25 UI/dL, 30 UI/dL, 35 UI/dL, 40 UI/dL, 45 UI/dL, 50 UI/dL, 55 UI/dL, 60 UI/dL, 65 UI/dL, 70 UI/dL, 75 UI/dL, 80 UI/dL, 85 UI/dL, 90 UI/dL, 95 UI/dL, 100 IU/dL, 105 IU/dL, 110 IU/dL, 115 IU/dL, 120 IU/dL, 125 IU/dL, 130 IU/dL, 135 IU/dL, 140 IU/dL, 145 IU/dL, o 150 UI/dL en un sujeto mediante la administración de una molécula de ácido nucleico aislada, polipéptido aislado o vector de la descripción. En otra realización, tratar o tratamiento significa mantener un nivel mínimo de FVIII entre aproximadamente 1 y aproximadamente 150 UI/dL, aproximadamente 1 y aproximadamente 125 UI/dL, aproximadamente 1 y aproximadamente 100 UI/dL, aproximadamente 1 y aproximadamente 90 UI/dL, aproximadamente 1 y aproximadamente 85 UI/dL, aproximadamente 1 y aproximadamente 80 UI/dL,

aproximadamente 1 y aproximadamente 75 UI/dL, aproximadamente 1 y aproximadamente 70 UI/dL, aproximadamente 1 y aproximadamente 65 UI/dL, aproximadamente 1 y aproximadamente 60 UI/dL, aproximadamente 1 y aproximadamente 55 UI/dL, aproximadamente 1 y aproximadamente 50 UI/dL, aproximadamente 1 y aproximadamente 45 UI/dL, aproximadamente 1 y aproximadamente 40 UI/dL, aproximadamente 1 y aproximadamente 35 UI/dL, aproximadamente 1 y aproximadamente 30 UI/dL, aproximadamente 1 y aproximadamente 25 UI/dL, aproximadamente 25 y aproximadamente 125 UI/dL, aproximadamente 50 y aproximadamente 100 UI/dL, aproximadamente 50 y aproximadamente 75 UI/dL, aproximadamente 75 y aproximadamente 100 UI/dL, aproximadamente 1 y aproximadamente 20 UI/dL, aproximadamente 2 y aproximadamente 20 UI/dL, aproximadamente 3 y aproximadamente 20 UI/dL, aproximadamente 4 y aproximadamente 20 UI/dL, aproximadamente 5 y aproximadamente 20 UI/dL, aproximadamente 6 y aproximadamente 20 UI/dL, aproximadamente 7 y aproximadamente 20 UI/dL, aproximadamente 8 y aproximadamente 20 UI/dL, aproximadamente 9 y aproximadamente 20 UI/dL, o aproximadamente 10 y aproximadamente 20 UI/dL. El tratamiento o tratar una enfermedad o afección también puede incluir el mantenimiento de la actividad de FVIII en un sujeto a un nivel comparable al menos aproximadamente al 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9 %, 10%, 11%, 12%, 13%, 14%, 15%, 16%, 17%, 18%, 19%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 100%, 105%, 110%, 115%, 120%, 125%, 130% , 135%, 140%, 145% o 150% de la actividad de FVIII en un sujeto no hemofílico. El nivel mínimo requerido para el tratamiento puede medirse mediante uno o más métodos conocidos y puede ajustarse (aumentarse o disminuirse) para cada persona.

**[00100]** "Administrar", como se usa en el presente documento, significa dar una molécula de ácido nucleico farmacéuticamente aceptable, un polipéptido expresado a partir del mismo, o un vector que comprenda la molécula de ácido nucleico de la descripción a un sujeto a través de una ruta farmacéuticamente aceptable. Las vías de administración pueden ser intravenosas, por ejemplo, inyección intravenosa e infusión intravenosa. Las vías de administración adicionales incluyen, por ejemplo, administración subcutánea, intramuscular, oral, nasal y pulmonar. Las moléculas de ácido nucleico, polipéptidos y vectores se pueden administrar como parte de una composición farmacéutica que comprenda al menos un excipiente.

**[00101]** La expresión "farmacéuticamente aceptable", como se usa en el presente documento, se refiere a entidades moleculares y composiciones que son fisiológicamente tolerables y que típicamente no producen toxicidad o reacción alérgica alguna o similar, como malestar gástrico, mareos y similares, cuando se administra a un ser humano. Opcionalmente, como se usa en el presente documento, la expresión "farmacéuticamente aceptable" significa aprobado por alguna agencia reguladora del gobierno federal o estatal o incluido en la Farmacopea de EE. UU. u otra farmacopea generalmente reconocida para su uso en animales y, más particularmente, en seres humanos.

**[00102]** Como se usa en el presente documento, la frase "sujeto que lo necesita" incluye sujetos, tales como sujetos mamíferos, que se beneficiarían de la administración de una molécula de ácido nucleico, polipéptido o vector de la descripción, por ejemplo, para mejorar la hemostasia. En una realización, los sujetos incluyen, entre otros, individuos con hemofilia. En otra realización, los sujetos incluyen, entre otros, individuos que han desarrollado un inhibidor de la proteína terapéutica, por ejemplo, el factor de coagulación, por ejemplo, FVIII, y, por lo tanto, necesitan una terapia de derivación. El sujeto puede ser un adulto o un menor de edad (por ejemplo, menores de 12 años).

**[00103]** Como se usa en el presente documento, la expresión "proteína terapéutica" se refiere a cualquier polipéptido conocido en la técnica que pueda administrarse a un sujeto. En algunas realizaciones, la proteína terapéutica comprende una proteína seleccionada de un factor de coagulación, un factor de crecimiento, un anticuerpo, un fragmento funcional del mismo o una de sus combinaciones. Como se usa en el presente documento, la expresión "factor de coagulación" se refiere a moléculas, o análogos de las mismas, que se producen naturalmente o de forma recombinante y evitan o disminuyen la duración de algún episodio de sangrado en algún sujeto. En otras palabras, significa que las moléculas tienen actividad pro-coagulante, es decir, son responsables de la conversión de fibrinógeno en una malla de fibrina insoluble que hace que la sangre se coagule o condense. El "factor de coagulación", como se usa en el presente documento, incluye un factor de coagulación activado, su zimógeno o un factor de coagulación activable. Un "factor de coagulación activable" es un factor de coagulación en una forma inactiva (por ejemplo, en su forma de zimógeno) que puede convertirse en una forma activa. La expresión "factor de coagulación" incluye, entre otros, el factor I (FI), factor II (FII), factor V (FV), FVII, FVIII, FIX, factor X (FX), factor XI (FXI), factor XII (FXII), factor XIII

(FXIII), factor von Willebrand (VWF), precalicreína, quininógeno de alto peso molecular, fibronectina, antitrombina III, cofactor de heparina II, proteína C, proteína S, proteína Z, inhibidor de la proteasa relacionada con la proteína Z (ZPI), plasminógeno, alfa 2-antiplasmina, activador de plasminógeno tisular (tPA), uroquinasa, inhibidor del activador de plasminógeno-1 (PAI-1), inhibidor del activador de plasminógeno-2 (PAI2), zimógenos de los mismos, formas activadas de los mismos, o cualquiera de sus combinaciones.

**[00104]** La actividad de coagulación, como se usa en este documento, significa la capacidad de participar en una cascada de reacciones bioquímicas que culmina en la formación de un coágulo de fibrina y/o reduce la gravedad, duración o frecuencia de la hemorragia o el episodio de sangrado.

**[00105]** Un "factor de crecimiento", como se usa en el presente documento, incluye cualquier factor de crecimiento conocido en la técnica, incluidas las citocinas y las hormonas. En algunas realizaciones, el factor de crecimiento se selecciona de adrenomedulina (AM), angiopoyetina (Ang), factor de motilidad autocrina, una proteína morfogenética ósea (BMP) (p. ej. BMP2, BMP4, BMP5, BMP7), un miembro de la familia del factor neurotrófico ciliar (p. ej., Factor neurotrófico ciliar (CNTF), factor inhibidor de leucemia (LIF), interleucina-6 (IL-6)), un factor estimulante de colonias (p. ej., factor estimulante de colonias de macrófagos (m-CSF), factor estimulante de colonias de granulocitos (G-CSF), factor estimulante de colonias de macrófagos y granulocitos (GM-CSF), un factor de crecimiento epidérmico (EGF), una efrina (por ejemplo, efrina A1, efrina A2, efrina A3, efrina A4, efrina A5, efrina B1, efrina B2, efrina B3), eritropoyetina (EPO), un factor de crecimiento de fibroblastos (FGF) (p. ej., FGF1, FGF2, FGF3, FGF4, FGF5, FGF6, FGF7, FGF8, FGF9, FGF10, FGF11, FGF 12, FGF13, FGF14, FGF15, FGF16, FGF17, FGF18, FGF19, FGF20, FGF21, FGF22, FGF23), somatotrofina bovina fetal (FBS), un miembro de la familia GDNF (p. ej., factor neurotrófico derivado de la línea de células gliales (GDNF), neurturina, persefina, artemina), factor de diferenciación de crecimiento-9 (GDF9), factor de crecimiento de hepatocitos (HGF), factor de crecimiento derivado de hepatoma (HDGF), insulina, factores de crecimiento tipo insulina (p. ej., factor de crecimiento tipo insulina-1 (IGF-1) o IGF-2, una interleucina (IL) (por ejemplo, IL-1, IL-2, IL-3, IL-4, IL-5, IL-6, IL-7), factor de crecimiento de queratinocitos (KGF), factor estimulante de la migración (MSF), proteína estimuladora de macrófagos (MSP o proteína tipo factor de crecimiento de hepatocitos (HGFLP)), miostatina (GDF-8), una neuregulina (p. ej., neuregulina 1 (NRG1), NRG2, NRG3, NRG4), una neurotrofina (p. ej., factor

neurotrófico derivado del cerebro (BDNF), factor de crecimiento nervioso (NGF), neurotrofina-3 (NT-3), NT-4, factor de crecimiento placentario (PGF), factor de crecimiento derivado de plaquetas (PDGF), renalasa (RNLS), factor de crecimiento de células T (TCGF), trombopoyetina (TPO), un factor de crecimiento transformante (p. ej., factor de crecimiento transformante alfa (TGF- $\alpha$ ), TGF- $\beta$ , factor de necrosis tumoral alfa (TNF- $\alpha$ ) y factor de crecimiento endotelial vascular (VEGF).

**[00106]** En algunas realizaciones, la proteína terapéutica está codificada por un gen seleccionado de distrofina ligada a X, MTM1 (miotubularina), tirosina hidroxilasa, AADC, ciclohidrolasa, SMN1, FXN (frataxina), GUCY2D, RS1, CFH, HTRA, ARMS, CFB/CC2, CNGA/CNGB, Prf65, ARSA, PSAP, IDUA (MPS I), IDS (MPS II), PAH, GAA (alfa-glucosidasa ácida), o cualquiera de sus combinaciones.

**[00107]** Como se usa en el presente documento, los términos "heterólogo" o "exógeno" se refieren a tales moléculas que normalmente no se encuentran en un contexto dado, por ejemplo, en una célula o en un polipéptido. Por ejemplo, una molécula exógena o heteróloga puede introducirse en una célula y solo está presente después de la manipulación de la célula, por ejemplo, por transfección u otras formas de ingeniería genética o una secuencia de aminoácidos heteróloga puede estar presente en una proteína en donde no se encuentra de forma natural.

**[00108]** Como se usa en el presente documento, la expresión "secuencia de nucleótidos heterólogos" se refiere a una secuencia de nucleótidos que no se produce naturalmente con una secuencia de polinucleótidos dada. En una realización, la secuencia de nucleótidos heteróloga codifica un polipéptido capaz de extender la semivida de la proteína terapéutica, por ejemplo, el factor de coagulación, por ejemplo, FVIII. En otra realización, la secuencia de nucleótidos heteróloga codifica un polipéptido que aumenta el radio hidrodinámico de la proteína terapéutica, p. ej., el factor de coagulación, por ejemplo, FVIII. En otras realizaciones, la secuencia de nucleótidos heteróloga codifica un polipéptido que mejora una o más propiedades farmacocinéticas de la proteína terapéutica sin afectar significativamente su actividad o función biológica (por ejemplo, una actividad procoagulante). En algunas realizaciones, la proteína terapéutica está unida o conectada al polipéptido codificado por la secuencia de nucleótidos heterólogos por un enlazante. Los ejemplos no limitantes de restos de polipéptidos codificados por secuencias de nucleótidos heterólogos incluyen una región constante de inmunoglobulina o una porción de la misma, albúmina o un fragmento de la misma, un resto de unión a albúmina, una transferrina, los polipéptidos PAS de la Solicitud de Patente de los Estados Unidos

Núm. 20100292130, una secuencia HAP, transferrina o un fragmento de la misma, el péptido C-terminal (CTP) de la subunidad  $\beta$  de la gonadotropina coriónica humana, molécula pequeña de unión a albúmina, una secuencia XTEN, restos de unión a FcRn (por ejemplo, regiones Fc completas o porciones de las mismas que se unen a FcRn), regiones Fc monocatenarias (regiones ScFc, p. ej., como se describe en los documentos US 2008/0260738, WO 2008/012543 o WO 2008/1439545), enlazantes de poliglicina, enlazantes de poliserina, péptidos y polipéptidos cortos de 6-40 aminoácidos de dos tipos de aminoácidos seleccionados de glicina (G), alanina (A), serina (S), treonina (T), glutamato (E) y prolina (P) con diversos grados de estructura secundaria de menos del 50% a más de 50%, entre otros, o dos o más combinaciones de los mismos. En algunas realizaciones, el polipéptido codificado por la secuencia de nucleótidos heteróloga está unido a un resto no polipéptido. Los ejemplos no limitantes de los restos no polipeptídicos incluyen polietilenglicol (PEG), moléculas pequeñas que se unen a la albúmina, ácido polisiálico, hidroxietilalmidón (HES), un derivado del mismo o cualquiera de sus combinaciones.

**[00109]** Como se usa en el presente documento, la expresión "región Fc" se define como la porción de un polipéptido que corresponde a la región Fc de Ig nativa, es decir, según se forma por la asociación dimérica de los respectivos dominios Fc de sus dos cadenas pesadas. Una región Fc nativa forma un homodímero con otra región Fc. Por el contrario, la expresión "región Fc genéticamente fusionada" o "región Fc monocatenaria" (región scFc), como se usa en este documento, se refiere a una región Fc dimérica sintética compuesta de dominios Fc genéticamente unidos dentro de una cadena de polipéptidos única (es decir, codificada en una sola secuencia genética contigua).

**[00110]** En una realización, la "región Fc" se refiere a la porción de una sola cadena pesada de Ig que comienza en la región bisagra justo antes del sitio de escisión de papaína (es decir, el residuo 216 en IgG, tomando el primer residuo de la región constante de la cadena pesada como 114) y termina en el C-terminal del anticuerpo. Por consiguiente, un dominio Fc completo comprende al menos un dominio de bisagra, un dominio CH2 y un dominio CH3.

**[00111]** La región Fc de una región constante de Ig, dependiendo del isotipo Ig, puede incluir los dominios CH2, CH3 y CH4, así como la región bisagra. Las proteínas quiméricas que comprenden una región Fc de una Ig confieren varias propiedades deseables a una proteína quimérica, incluida una mayor estabilidad, una mayor semivida en suero (véase Capon et al, 1989, Nature 337: 525), así como la unión a

receptores Fc como el receptor neonatal de Fc (FcRn) (patentes de EE.UU. Nos. 6.086.875, 6.485.726, 6.030.613; y los documentos WO 03/077834; US2003-0235536A1), que se incorporan en este documento como referencia en su totalidad.

**[00112]** Una "secuencia de nucleótidos de referencia", cuando se usa en el presente documento como una comparación con una secuencia de nucleótidos de la descripción, es una secuencia de polinucleótidos esencialmente idéntica a la secuencia de nucleótidos de la descripción, excepto que la secuencia no está optimizada. Por ejemplo, la secuencia de nucleótidos de referencia para una molécula de ácido nucleico que consiste en el codón BDD FVIII optimizado de SEQ ID NO: 1 y una secuencia de nucleótidos heteróloga que codifica una región Fc de monocatenaria unida a SEQ ID NO: 1 en su extremo 3' es una molécula de ácido nucleico que consiste en el BDD FVIII original (o "progenitora") de SEQ ID NO: 16 y la secuencia de nucleótidos heteróloga idéntica que codifica una región Fc monocatenaria unida a SEQ ID NO: 16 en su extremo 3'.

**[00113]** Como se usa en el presente documento, el término "optimizado", con respecto a las secuencias de nucleótidos, se refiere a una secuencia de polinucleótidos que codifica un polipéptido, en el que la secuencia de polinucleótidos se ha mutado para potenciar una propiedad de esa secuencia de polinucleótidos. En algunas realizaciones, la optimización se realiza para aumentar los niveles de transcripción, aumentar los niveles de traducción, aumentar los niveles de ARNm en estado estacionario, aumentar o disminuir la unión de proteínas reguladoras tales como factores de transcripción generales, aumentar o disminuir el empalme o aumentar el rendimiento del polipéptido producido por la secuencia de polinucleótidos. Los ejemplos de cambios que se pueden hacer en una secuencia de polinucleótidos para optimizarla incluyen la optimización de codones, la optimización del contenido de G/C, la eliminación de secuencias repetidas, la eliminación de elementos ricos en AT, la eliminación de sitios de empalme críptico, la eliminación de elementos de acción cis que reprimen la transcripción o traducción, la adición o eliminación de secuencias de poli-T o poli-A, la adición de secuencias alrededor del sitio de inicio de la transcripción que mejoran la transcripción, como las secuencias de consenso de Kozak, la eliminación de secuencias que podrían formar estructuras de bucle de tallo, la eliminación de secuencias desestabilizadoras, y dos o más combinaciones de los mismos.

## **II. Moléculas de ácido nucleico**

**[00114]** La presente descripción está dirigida a una molécula de ácido nucleico sin cápsides, similar a un plásmido, que codifica una proteína terapéutica o un gen que puede modular la expresión de una proteína diana. Una cápside, la cubierta proteica de un virus, encierra el material genético del virus. Se sabe que las cápsides ayudan a las funciones del virión protegiendo el genoma viral, entregando el genoma a un huésped e interactuando con el huésped. No obstante, las cápsides virales pueden ser un factor para limitar la capacidad de empaquetamiento de los vectores y/o inducir respuestas inmunes, especialmente cuando se usan en terapia génica.

**[00115]** Los vectores AAV se han convertido en uno de los tipos más comunes de vectores de terapia génica. Sin embargo, la presencia de la cápside limita la utilidad de un vector AAV en terapia génica. En particular, la cápside en sí misma puede limitar el tamaño del transgén que se incluye en el vector en valores de menos de 4,5 kb. Diversas proteínas terapéuticas que pueden ser útiles en una terapia génica pueden superar fácilmente este tamaño incluso antes de que se agreguen elementos reguladores.

**[00116]** Además, las proteínas que forman la cápside pueden servir como antígenos que pueden ser atacados por el sistema inmunitario del sujeto. El AAV es muy común en la población general, y la mayoría de las personas han estado expuestas a un AAV durante toda su vida. Por lo tanto, la mayoría de los receptores potenciales de terapia génica probablemente ya hayan desarrollado una respuesta inmunitaria a un AAV y, por lo tanto, es más probable que rechacen la terapia.

**[00117]** Ciertos aspectos de la presente descripción tienen como objetivo superar estas deficiencias de vectores AAV. En particular, ciertos aspectos de la presente descripción están dirigidos a una molécula de ácido nucleico, que comprende un primer ITR, un segundo ITR y un casete génico, por ejemplo, que codifica una proteína terapéutica y/o un ARNmi. En algunas realizaciones, la molécula de ácido nucleico no comprende un gen que codifica una proteína de la cápside, una proteína de replicación y/o una proteína de ensamblaje. En algunas realizaciones, el casete génico codifica una proteína terapéutica. En algunas realizaciones, la proteína terapéutica comprende un factor de coagulación. En algunas realizaciones, el casete génico codifica un ARNmi. En ciertas realizaciones, el casete génico está colocado entre la primera ITR y la segunda ITR. En algunas realizaciones, la molécula de ácido nucleico comprende además una o más regiones no codificantes. En ciertas realizaciones, la una o más regiones no codificantes comprenden una secuencia

promotora, un intrón, un elemento regulador postranscripcional, una secuencia poli (A) 3'UTR o cualquiera de sus combinaciones.

**[00118]** En una realización, el casete génico es un ácido nucleico monocatenario. En otra realización, el casete génico es un ácido nucleico bicatenario.

5 **[00119]** En una realización, la molécula de ácido nucleico comprende:

(a) una primera ITR que es una ITR de un miembro de la familia de *Parvoviridae* VnoAA;

(b) una secuencia promotora específica de tejido, por ejemplo, un promotor TTP;

(c) un intrón, por ejemplo, un intrón sintético;

10 (d) un nucleótido que codifica un ARNmi o una proteína terapéutica, por ejemplo, un factor de coagulación;

(e) un elemento regulador postranscripcional, por ejemplo, WPRE;

(f) una secuencia de cola poli (A) 3'UTR, por ejemplo, bGHpA;

15 (g) una segunda ITR que es una ITR de un miembro de la familia *Parvoviridae* VnoAA

**[00120]** En una realización, la molécula de ácido nucleico comprende:

(a) una primera ITR que es una ITR de un miembro de la familia de *Parvoviridae* que no es AAV;

(b) una secuencia promotora específica de tejido, por ejemplo, un promotor TTP;

20 (c) un intrón, por ejemplo, un intrón sintético;

(d) un nucleótido que codifica un ARNmi, en donde el ARNmi regula la expresión de un gen diana seleccionado de SOD1, HTT, RHO y cualquiera de sus combinaciones;

(e) un elemento regulador postranscripcional, por ejemplo, WPRE;

25 (f) una secuencia de cola poli (A) 3'UTR, por ejemplo, bGHpA;

(g) una segunda ITR que es una ITR de un miembro de la familia *Parvoviridae* VnoAA

**[00121]** En una realización, la molécula de ácido nucleico comprende:

30 (a) una primera ITR que es una ITR de un miembro de la familia de *Parvoviridae* VnoAA;

(b) una secuencia promotora específica de tejido, por ejemplo, el promotor TTP;

(c) un intrón, por ejemplo, un intrón sintético;

35 (d) un nucleótido que codifica la distrofina ligada a X, MTM1 (miotubularina), tirosina hidroxilasa, AADC, ciclohidrolasa, SMN1, FXN (frataxina), GUCY2D, RSI, CFH, HTRA, ARMS, CFB/CC2, CNGA/CNGB, Prf65, ARSA, PSAP, IDUA

(MPS I), IDS (MPS II), PAH, GAA (alfa-glucosidasa ácida), o cualquiera de sus combinaciones;

(e) un elemento regulador postranscripcional, por ejemplo, WPRE;

(f) una secuencia de cola poli (A) 3'UTR, por ejemplo, bGHpA;

5 (g) una segunda ITR que es una ITR de un miembro de la familia *Parvoviridae* VnoAA

**[00122]** En una realización, la molécula de ácido nucleico comprende:

(a) una primera ITR que es una ITR de un AAV, por ejemplo, un genoma del serotipo 2 de AAV;

10 (b) una secuencia promotora específica de tejido, por ejemplo, el promotor TTP;

(c) un intrón, por ejemplo, un intrón sintético;

(d) un nucleótido que codifica FVIII; en donde el nucleótido tiene al menos 85%, al menos 90%, al menos 95%, al menos 96%, al menos 97%, al menos 98%, o al menos 99% de identidad de secuencia con una secuencia de nucleótidos seleccionada de SEQ ID NOs: 1-14 o SEQ ID NO: 71, en donde el FVIII codificado por el nucleótido retiene una actividad de FVIII;

15 (e) un elemento regulador postranscripcional, por ejemplo, WPRE;

(f) una secuencia de cola poli (A) 3'UTR, por ejemplo, bGHpA; y

(g) una segunda ITR que es una ITR de un AAV, por ejemplo, un genoma de serotipo 2 de AAV.

20

**[00123]** En una realización, la molécula de ácido nucleico comprende:

(a) una primera ITR que es una ITR de un AAV, por ejemplo, un genoma del serotipo 2 de AAV;

(b) una secuencia promotora específica de tejido, por ejemplo, el promotor TTP;

25 (c) un intrón, por ejemplo, un intrón sintético;

(d) un nucleótido que codifica un ARNmi, en el que el ARNmi regula la expresión de un gen diana, por ejemplo, SOD1, HTT, RHO y cualquiera de sus combinaciones;

(f) una secuencia de cola poli (A) 3'UTR, por ejemplo, bGHpA; y

30 (g) una segunda ITR que es una ITR de un AAV, por ejemplo, un genoma de serotipo 2 de AAV.

30

**[00124]** En una realización, la molécula de ácido nucleico comprende:

(a) una primera ITR que es una ITR de un AAV, por ejemplo, un genoma del serotipo 2 de AAV;

35 (b) una secuencia promotora específica de tejido, por ejemplo, el promotor TTP;

- (c) un intrón, por ejemplo, un intrón sintético;
- (d) un nucleótido que codifica la distrofina ligada a X, MTM1 (miotubularina), tirosina hidroxilasa, AADC, ciclohidrolasa, SMN1, FXN (frataxina), GUCY2D, RSI, CFH, HTRA, ARMS, CFB/CC2, CNGA/CNGB, Prf65, ARSA, PSAP, IDUA (MPS I), IDS (MPS II), PAH, GAA (alfa-glucosidasa ácida), o cualquiera de sus combinaciones;
- (f) una secuencia de cola poli (A) 3'UTR, por ejemplo, bGHpA; y
- (g) una segunda ITR que es una ITR de un AAV, por ejemplo, un genoma de serotipo 2 de AAV.

10 **[00125]** En otra realización, la molécula de ácido nucleico comprende:

- (a) una primera ITR;
- (b) una secuencia promotora específica de tejido, por ejemplo, el promotor TTP;
- (c) un intrón, por ejemplo, un intrón sintético;
- (d) un nucleótido que codifica un ARNmi o una proteína terapéutica, por ejemplo, un factor de coagulación;
- (e) un elemento regulador postranscripcional, por ejemplo, WPRE;
- (f) una secuencia de cola poli (A) 3'UTR, por ejemplo, bGHpA; y
- (g) una segunda ITR,

15 en donde una de la primera ITR o de la segunda ITR es una ITR de un miembro de la familia de *Parvoviridae* que no es AAV y la otra ITR es una ITR de un AAV, por ejemplo, un genoma del serotipo 2 de AAV.

**[00126]** En otra realización, la molécula de ácido nucleico comprende:

- (a) una primera ITR;
- (b) una secuencia promotora específica de tejido, el promotor TTP;
- (c) un intrón, por ejemplo, un intrón sintético;
- (d) un nucleótido que codifica un ARNmi o una proteína terapéutica, por ejemplo, un factor de coagulación;
- (e) un elemento regulador postranscripcional, por ejemplo, WPRE;
- (f) una secuencia de cola poli (A) 3'UTR, por ejemplo, bGHpA; y
- (g) una segunda ITR,

25 en el que la primera ITR es una ITR sintética, la segunda ITR es una ITR sintética, o tanto la primera ITR como la segunda ITR son ITR sintéticas.

#### **A. Repeticiones terminales invertidas**

**[00127]** Ciertos aspectos de la presente descripción están dirigidos a una molécula de ácido nucleico que comprende una primera ITR, por ejemplo, una 5'ITR,

35

y una segunda ITR, por ejemplo, una 3' ITR. Típicamente, las ITR están involucradas en la replicación y el rescate de ADN de parvovirus (p. ej., AAV), o escisión, de plásmidos procariotas (Samulski et al., 1983, 1987; Senapathy et al., 1984; Gottlieb y Muzyczka, 1988). Además, las ITR parecen ser las secuencias mínimas requeridas para la integración proviral de AAV y para el empaquetamiento del ADN de AAV en viriones (McLaughlin et al., 1988; Samulski et al., 1989). Estos elementos son esenciales para la multiplicación eficiente de un genoma de parvovirus. Se hipotetiza que los elementos de definición mínimos indispensables para la función ITR son un sitio de unión a Rep (p. ej., RBS; GCGCGCTCGCTCGCTC (SEQ ID NO: 104) para AAV2) y un sitio de resolución terminal (p. ej., TRS; AGTTGG (SEQ ID NO: 105) para AAV2) más una secuencia palindrómica variable que permite la formación de horquillas. Las regiones de nucleótidos palindrómicos normalmente funcionan juntas en cis como orígenes de la replicación del ADN y como señales de empaquetamiento del virus. Las secuencias complementarias en las ITR se pliegan en una estructura de horquilla durante la replicación del ADN. En algunas realizaciones, las ITR se pliegan en una estructura en forma de T de horquilla. En otras realizaciones, las ITR se pliegan en estructuras de horquilla sin forma de T, por ejemplo, en una estructura de horquilla en forma de U. Los datos sugieren que las estructuras de horquilla en forma de T de las ITR de AAV pueden inhibir la expresión de un transgén flanqueado por las ITR. Véase, por ejemplo, Zhou et al., Scientific Reports 7: 5432 (14 de julio de 2017). Al utilizar una ITR que no forme estructuras en forma de horquilla en forma de T, se puede evitar esta forma de inhibición. Por lo tanto, en ciertos aspectos, un polinucleótido que comprende una ITR VnoAA tiene una expresión transgénica mejorada en comparación con un polinucleótido que comprende una ITR AAV que forma una horquilla en forma de T.

**[00128]** En algunas realizaciones, la ITR comprende una ITR natural, p. ej. la ITR comprende todo o una parte de una ITR de parvovirus. En algunas realizaciones, la ITR comprende una secuencia sintética. En una realización, la primera ITR o la segunda ITR comprende una secuencia sintética. En otra realización, cada una de la primera ITR y la segunda ITR comprende una secuencia sintética. En algunas realizaciones, la primera ITR o la segunda ITR comprende una secuencia natural. En otra realización, cada una de la primera ITR y la segunda ITR comprende una secuencia natural.

**[00129]** En algunas realizaciones, la ITR comprende o consiste en una porción de una ITR natural, por ejemplo, una ITR truncada. En algunas realizaciones, la ITR

comprende o consiste en un fragmento de una ITR natural, en el que el fragmento comprende al menos aproximadamente 5 nucleótidos, al menos aproximadamente 10 nucleótidos, al menos aproximadamente 15 nucleótidos, al menos aproximadamente 20 nucleótidos, al menos aproximadamente 25 nucleótidos, al menos aproximadamente 30 nucleótidos, al menos aproximadamente 35 nucleótidos, al menos aproximadamente 40 nucleótidos, al menos aproximadamente 45 nucleótidos, al menos aproximadamente 50 nucleótidos, al menos aproximadamente 55 nucleótidos, al menos aproximadamente 60 nucleótidos, al menos aproximadamente 65 nucleótidos, al menos aproximadamente 70 nucleótidos, al menos aproximadamente 75 nucleótidos, al menos aproximadamente 80 nucleótidos, al menos aproximadamente 85 nucleótidos, al menos aproximadamente 90 nucleótidos, al menos aproximadamente 95 nucleótidos, al menos aproximadamente 100 nucleótidos, al menos aproximadamente 125 nucleótidos, al menos aproximadamente 150 nucleótidos, al menos aproximadamente 175 nucleótidos, al menos aproximadamente 200 nucleótidos, al menos aproximadamente 225 nucleótidos, al menos aproximadamente 250 nucleótidos, al menos aproximadamente 275 nucleótidos, al menos aproximadamente 300 nucleótidos, al menos aproximadamente 325 nucleótidos, al menos aproximadamente 350 nucleótidos, al menos aproximadamente 375 nucleótidos, al menos aproximadamente 400 nucleótidos, al menos aproximadamente 425 nucleótidos, al menos aproximadamente 450 nucleótidos, al menos aproximadamente 475 nucleótidos, al menos aproximadamente 500 nucleótidos, al menos aproximadamente 525 nucleótidos, al menos aproximadamente 550 nucleótidos, al menos aproximadamente 575 nucleótidos o al menos aproximadamente 600 nucleótidos; en donde la ITR retiene una propiedad funcional de la ITR natural. En ciertas realizaciones, la ITR comprende o consiste en un fragmento de una ITR natural, en el que el fragmento comprende al menos aproximadamente 129 nucleótidos; en donde la ITR retiene una propiedad funcional de la ITR natural. En ciertas realizaciones, la ITR comprende o consiste en un fragmento de una ITR natural, en el que el fragmento comprende al menos aproximadamente 102 nucleótidos; en donde la ITR retiene una propiedad funcional de la ITR natural.

**[00130]** En algunas realizaciones, la ITR comprende o consiste en una porción de una ITR natural, en donde el fragmento comprende al menos aproximadamente 5%, al menos aproximadamente 10%, al menos aproximadamente 15%, al menos aproximadamente 20%, al menos aproximadamente 25%, al menos aproximadamente

30%, al menos aproximadamente 35%, al menos aproximadamente 40%, al menos aproximadamente 45%, al menos aproximadamente 50%, al menos aproximadamente 55%, al menos aproximadamente 60%, al menos aproximadamente 65%, al menos aproximadamente 70%, al menos aproximadamente 75%, al menos aproximadamente 80%, al menos aproximadamente 85%, al menos aproximadamente 90%, al menos aproximadamente 95%, al menos aproximadamente 96%, al menos aproximadamente 97%, al menos aproximadamente 98%, o al menos aproximadamente 99% de la longitud de la ITR natural; en donde el fragmento retiene una propiedad funcional de la ITR natural.

- 10 **[00131]** En ciertas realizaciones, la ITR comprende o consiste en una secuencia que tiene una identidad de secuencia de al menos 50%, al menos 51%, al menos 52%, al menos 53%, al menos 54%, al menos 55%, al menos 56%, al menos 57%, al menos 58%, al menos 59%, al menos 60%, al menos 61%, al menos 62%, al menos 63%, al menos 64%, al menos 65%, al menos 66%, al menos 67%, al menos 68%, al menos 69%, al menos 70%, al menos 71%, al menos 72%, al menos 73%, al menos 74%, al menos 75%, al menos 76%, al menos 77%, al menos 78%, al menos 79%, al menos 80%, al menos 81%, al menos 82%, al menos 83%, al menos 84%, al menos 85%, al menos 86%, al menos 87%, al menos 88%, al menos 89%, al menos 90%, al menos 91%, al menos 92%, al menos 93%, al menos 94%, al menos 95%, al menos 96%, al menos 97%, al menos 98%, al menos 99% o 100% con respecto a una porción homóloga de una ITR natural, cuando está alineada adecuadamente; en donde la ITR retiene una propiedad funcional de la ITR natural. En otras realizaciones, la ITR comprende o consiste en una secuencia que tiene una identidad de secuencia de al menos 90% con respecto a una porción homóloga de una ITR natural, cuando está alineada adecuadamente; en donde la ITR retiene una propiedad funcional de la ITR natural. En algunas realizaciones, la ITR comprende o consiste en una secuencia que tiene una identidad de secuencia de al menos 80% con respecto a una porción homóloga de una ITR natural, cuando está alineada adecuadamente; en donde la ITR retiene una propiedad funcional de la ITR natural. En algunas realizaciones, la ITR comprende o consiste en una secuencia que tiene una identidad de secuencia de al menos 70% con respecto a una porción homóloga de una ITR natural, cuando está alineada adecuadamente; en donde la ITR retiene una propiedad funcional de la ITR natural. En algunas realizaciones, la ITR comprende o consiste en una secuencia que tiene una identidad de secuencia de al menos el 60% con una porción homóloga de una ITR natural, cuando está alineada adecuadamente; en donde la ITR retiene una

propiedad funcional de la ITR natural. En algunas realizaciones, la ITR comprende o consiste en una secuencia que tiene una identidad de secuencia de al menos 50% con respecto a una porción homóloga de una ITR natural, cuando está alineada adecuadamente; en donde la ITR retiene una propiedad funcional de la ITR natural.

5   **[00132]**        En algunas realizaciones, la ITR comprende una ITR de un genoma AAV. En algunas realizaciones, la ITR es una ITR de un genoma de AAV seleccionado de AAV1, AAV2, AAV3, AAV4, AAV5, AAV6, AAV7, AAV8, AAV9, AAV10 y AAV11, y cualquiera de sus combinaciones. En una realización particular, la ITR es una ITR del genoma AAV2. En otra realización, la ITR es una secuencia sintética diseñada  
10    genéticamente para incluir en sus extremos 5' y 3' ITR derivadas de uno o más de los genomas de AAV.

**[00133]**        En algunas realizaciones, la ITR no se deriva de un genoma de AAV. En algunas realizaciones, la ITR es una ITR de un VnoAA. En algunas realizaciones, la ITR es una ITR de un genoma VnoAA de la familia viral *Parvoviridae* seleccionado,  
15    entre otros, del grupo que consiste en *Bocavirus*, *Dependovirus*, *Erythrovirus*, *Amdovirus*, *Parvovirus*, *Densovirus*, *Iteravirus*, *Contravirus*, *Aveparvovirus*, *Copiparvovims*, *Protoparvovirus*, *Tetraparvovirus*, *Ambidensovirus*, *Brevidensovirus*, *Hepandensovirus*, *Penstyldensovirus* y cualquiera de sus combinaciones. En ciertas realizaciones, la ITR se deriva del eritrovirus parvovirus B19 (virus humano). En otra  
20    realización, la ITR se deriva de una cepa de parvovirus de pato Muscovy (MDPV). En ciertas realizaciones, la cepa de MDPV está atenuada, por ejemplo, la cepa de MDPV FZ91-30. En otras realizaciones, la cepa MDPV es patógena, por ejemplo, la cepa MDPV YY. En algunas realizaciones, la ITR se deriva de un parvovirus porcino, por ejemplo, el parvovirus porcino U44978. En algunas realizaciones, la ITR se deriva de  
25    un virus diminuto de ratones, por ejemplo, el virus diminuto de ratones U34256. En algunas realizaciones, la ITR se deriva de un parvovirus canino, por ejemplo, el parvovirus canino M19296. En algunas realizaciones, la ITR se deriva de un virus de la enteritis de visón, por ejemplo, el virus de la enteritis de visón D00765. En algunas realizaciones, la ITR se deriva de un *Dependoparvovirus*. En una realización, el  
30    *Dependoparvovirus* es una cepa de parvovirus *Dependovirus* Goose (GPV). En una realización específica, la cepa GPV está atenuada, por ejemplo, la cepa GPV 82-0321V. En otra realización específica, la cepa de GPV es patógena, por ejemplo, la cepa B de GPV.

**[00134]**        La primera ITR y la segunda ITR de la molécula de ácido nucleico  
35    pueden derivarse del mismo genoma, por ejemplo, del genoma del mismo virus, o de

genomas diferentes, por ejemplo, de los genomas de dos o más genomas de virus diferentes. En ciertas realizaciones, la primera ITR y la segunda ITR se derivan del mismo genoma de AAV. En una realización específica, las dos ITR presentes en la molécula de ácido nucleico de la invención son iguales y, en particular, pueden ser ITR AAV2. En otras realizaciones, la primera ITR se deriva de un genoma de AAV y la segunda ITR no se deriva de un genoma de AAV (por ejemplo, un genoma VnoAA). En otras realizaciones, la primera ITR no se deriva de un genoma de AAV (por ejemplo, un genoma VnoAA) y la segunda ITR se deriva de un genoma de AAV. En otras realizaciones más, tanto la primera ITR como la segunda ITR no se derivan de un genoma AAV (por ejemplo, un genoma VnoAA). En una realización particular, la primera ITR y la segunda ITR son idénticas.

**[00135]** En algunas realizaciones, la primera ITR se deriva de un genoma AAV, y la segunda ITR se deriva de un genoma seleccionado del grupo que consiste en *Bocavirus*, *Dependovirus*, *Erythrovirus*, *Amdovirus*, *Parvovirus*, *Densovirus*, *Iteravirus*, *Contravirus*, *Aveparvovirus*, *Copiparvovirus*, *Protoparvovirus*, *Tetraparvovirus*, *Ambidensovirus*, *Brevidensovirus*, *Hepandensovirus*, *Penstyldensovirus* y cualquiera de sus combinaciones. En otras realizaciones, la segunda ITR se deriva de un genoma de AAV, y la primera ITR se deriva de un genoma seleccionado del grupo que consiste en *Bocavirus*, *Dependovirus*, *Erythrovirus*, *Amdovirus*, *Parvovirus*, *Densovirus*, *Iteravirus*, *Contravirus*, *Aveparvovirus*, *Copiparvovirus*, *Protoparvovirus*, *Tetraparvovirus*, *Ambidensovirus*, *Brevidensovirus*, *Hepandensovirus*, *Penstyldensovirus*, y cualquiera de sus combinaciones. En otras realizaciones, la primera ITR y la segunda ITR se derivan de un genoma seleccionado del grupo que consiste en *Bocavirus*, *Dependovirus*, *Erythrovirus*, *Amdovirus*, *Parvovirus*, *Densovirus*, *Iteravirus*, *Contravirus*, *Aveparvovirus*, *Copiparvovirus*, *Protoparvovirus*, *Tetraparvovirus*, *Ambidensovirus*, *Brevidensovirus*, *Hepandensovirus*, *Penstyldensovirus*, y cualquiera de sus combinaciones, en donde la primera ITR y la segunda ITR se derivan del mismo genoma. En otras realizaciones, la primera ITR y la segunda ITR se derivan de un genoma seleccionado del grupo que consiste en *Bocavirus*, *Dependovirus*, *Erythrovirus*, *Amdovirus*, *Parvovirus*, *Densovirus*, *Iteravirus*, *Contravirus*, *Aveparvovirus*, *Copiparvovirus*, *Protoparvovirus*, *Tetraparvovirus*, *Ambidensovirus*, *Brevidensovirus*, *Hepandensovirus*, *Penstyldensovirus*, y cualquiera de sus combinaciones, en donde la primera ITR y la segunda ITR se derivan de los diferentes genomas.

**[00136]** En algunas realizaciones, la primera ITR se deriva de un genoma AAV, y la segunda ITR se deriva del parvovirus y eritrovirus B19 (virus humano). En otras realizaciones, la segunda ITR se deriva de un genoma de AAV, y la primera ITR se deriva del parvovirus y eritrovirus B19 (virus humano).

5 **[00137]** En ciertas realizaciones, la primera ITR y/o la segunda ITR comprenden o consisten en todo o una parte de una ITR derivada de B19. En algunas realizaciones, la primera ITR y/o la segunda ITR comprenden o consisten en una secuencia de nucleótidos de al menos aproximadamente 50%, al menos aproximadamente 55%, al menos aproximadamente 60%, al menos aproximadamente 65%, al menos aproximadamente 65%, al menos aproximadamente 70%, al menos aproximadamente 75%, al menos aproximadamente 80%, al menos aproximadamente 85%, al menos aproximadamente 90%, al menos aproximadamente 95%, al menos aproximadamente 96%, al menos aproximadamente 97%, al menos aproximadamente 98%, al menos aproximadamente 99%, o 100% idéntica a una secuencia de nucleótidos seleccionada de SEQ ID NOs: 167, 168, 169, 170 y 171, en donde la primera ITR y/o la segunda ITR conserva una propiedad funcional del ITR B19 de la que se deriva. En algunas realizaciones, la primera ITR y/o la segunda ITR comprenden o consisten en una secuencia de nucleótidos al menos aproximadamente 50%, al menos aproximadamente 55%, al menos aproximadamente 60%, al menos aproximadamente 65%, al menos aproximadamente 65%, al menos aproximadamente 70%, al menos aproximadamente 75%, al menos aproximadamente 80%, al menos aproximadamente 85%, al menos aproximadamente 90%, al menos aproximadamente 95%, al menos aproximadamente 96%, al menos aproximadamente 97%, al menos aproximadamente 98%, al menos aproximadamente 99% o, o 100% idéntica a una secuencia de nucleótidos seleccionada de SEQ ID NOs: 167, 168, 169, 170 y 171, en donde la primera ITR y/o la segunda ITR es capaz de formar una estructura de horquilla. En ciertas realizaciones, la estructura de horquilla no comprende una horquilla en forma de T.

**[00138]** En algunas realizaciones, la primera ITR y/o la segunda ITR comprenden o consisten en una secuencia de nucleótidos seleccionada de SEQ ID NO: 167, 168, 169, 170 y 171. En algunas realizaciones, la primera ITR y/o la segunda ITR comprenden o consisten en la secuencia de nucleótidos establecida en SEQ ID NO: 167. En algunas realizaciones, la primera ITR y/o la segunda ITR comprenden o consisten en la secuencia de nucleótidos establecida en SEQ ID NO: 168. En algunas realizaciones, la primera ITR y/o la segunda ITR comprenden o consisten en la

secuencia de nucleótidos establecida en SEQ ID NO: 169. En algunas realizaciones, la primera ITR y/o la segunda ITR comprenden o consisten en la secuencia de nucleótidos establecida en SEQ ID NO: 170. En algunas realizaciones, la primera ITR y/o la segunda ITR comprenden o consisten en la secuencia de nucleótidos establecida en SEQ ID NO: 171.

5

Tabla 1. Muestras de secuencias de parvovirus ITR.

| Parvo-virus | ITR ID | Descripción                                        | Longitud (nt) | Secuencia                                                                                                                                                                                                                                                                                                                                                                                                                                                    |
|-------------|--------|----------------------------------------------------|---------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| B19         | wt     | Gene Bank: KY940273.1                              | 383           | CCAAATCAGATGCCGCCGGTCGCCGCCGG<br>TAGGCGGGACTTCCGGTACAAGATGGCGG<br>ACAATTACGTCATTTCTGTGACGTCATTTCTGTGACGTCACCTTCCGGTGGGCGGGACT<br>TCCGGAATTAGGGTTGGCTCTGGGCCAGC<br>TTGCTTGGGGTTGCCTTGACACTAAGACAA<br>GCGGCGCGCCGCTTGATCTTAGTGGCACG<br>TCAACCCCAAGCGCTGGCCCAGAGCCAAC<br>CCTAATTCCGGAAGTCCCGCCACCGGAA<br>GTGACGTCACAGGAAATGACGTCACAGGAA<br>ATGACGTAATTGTCCGCCATCTTGTACCGG<br>AAGTCCCGCCTACCGGCGGCGACCGGCGG<br>CATCTGATTTGGTGTCTTCTTTAAATTTT<br>(SEQ ID NO: 167) |
|             | d135   | excluye los primeros 135 nucleótidos               | 248           | CTCTGGGCCAGCTTGCTTGGGGTTGCCTTG<br>ACACTAAGACAAGCGGCGCGCCGCTTGAT<br>CTTAGTGGCACGTCAACCCCAAGCGCTGG<br>CCCAGAGCCAACCCTAATTCCGGAAGTCCC<br>GCCACCGGAAGTGACGTCACAGGAAATG<br>ACGTCACAGGAAATGACGTAATTGTCCGCC<br>ATCTTGTACCGGAAGTCCCGCCTACCGGC<br>GGCGACCGGCGGCATCTGATTTGGTGTCTT<br>CTTTTAAATTTT (SEQ ID NO: 168)                                                                                                                                                     |
|             | v1     | secuencia mínima basada en la comparación con AAV2 | 129           | CGGCGCGCCGCTTGATCTTAGTGGCACGT<br>CAACCAGCGCTGGCCCAGAGCCAACCCTA<br>ATTCCGGAAGTCTCAGTCCGCCATCTTGC<br>CCGCCTACCGGCGGCGACCGGCGGCATCA<br>TTTGGTGTCTT (SEQ ID NO: 169)                                                                                                                                                                                                                                                                                             |
|             | v2     | excluye los primeros 135 nucleótidos y             | 113           | CTCTGGGCCAGCTTGCTTGGGGTTGCCTTG<br>ACACTAAGACAAGCGGCGCGCCGCTTGAT<br>CTTAGTGGCACGTCAACCCCAAGCGCTGG                                                                                                                                                                                                                                                                                                                                                             |

|     |      |                                                                    |     |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |
|-----|------|--------------------------------------------------------------------|-----|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|     |      | los 135 nucleótidos correspondientes complementarios en palíndromo |     | CCCAGAGTGTCTTCTTTTAAATTTT (SEQ ID NO: 170)                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |
|     | v3   | secuencia mínima basada en la comparación con GPV                  | 340 | CAAATCAGATGCCGCCGGTCGCCGCCGGT<br>AGGCGGGACTTCCGGTACAAGATGGCGGA<br>CAATTACGTCATTTCTGTGACGTATTTCTT<br>GTGACGTACTTCCGGTGGCGGGACTTCCG<br>GAATTTTGGCTCTGGGCCAGCTTGCTTGGG<br>GTTGCCTTGACCAAGCGCGCGCCGCTTGA<br>TCACCCCAAGCGCTGGCCCAGAGCCACCT<br>AACCGGAAGTCCCCCACCAGGAAGTGACG<br>TCACAGGAAAGACGTCACAGGAAGTAATTG<br>TCCGCCATCTTGACCGGAAGTCCCGCACC<br>GGCGGCGACCGGCGGCATCTGATTTGGTG<br>TCTTCTTTTAAATTTT (SEQ ID NO: 171)                                                                                                                         |
| GPV | wt   | Gene Bank: U25749.1                                                | 444 | CTCATTGGAGGGTTCGTTTCGTTTCAACCAG<br>CCAATCAGGGGAGGGGGAAGTGACGCAAG<br>TTCCGGTCACATGCTTCCGGTGACGCACAT<br>CCGGTGACGTAGTTCCGGTCACGTGCTTCC<br>TGTCACGTGTTTCCGGTCACGTGACTTCCG<br>GTCATGTGACTTCCGGTGACGTGTTTCCGG<br>CTGTTAGGTTGACCACGCGCATGCCGCGC<br>GGTCAGCCCAATAGTTAAGCCGGAACACG<br>TCACCGGAAGTCACATGACCGGAAGTCAC<br>GTGACCGGAACACGTGACAGGAAGCACG<br>TGACCGGAACTACGTACCGGATGTGCGT<br>CACCGGAAGCATGTGACCGGAACTTGCGT<br>CACTTCCCCCTCCCCTGATTGGCTGGTTCCG<br>AACGAACGAACCCTCCAATGAGACTCAAGG<br>ACAAGAGGATATTTTGCGCGCCAGGAAGTG<br>(SEQ ID NO: 172) |
|     | d162 | excluye los primeros 162 nucleótidos                               | 282 | CGGTGACGTGTTTCCGGCTGTTAGGTTGAC<br>CACGCGCATGCCGCGCGGTGACCCCAATA<br>GTTAAGCCGGAACACGTACCGGAAGTCA<br>CATGACCGGAAGTCACGTGACCGGAAACA<br>CGTGACAGGAAGCACGTGACCGGAACTAC<br>GTCACCGGATGTGCGTCACCGGAAGCATG                                                                                                                                                                                                                                                                                                                                      |

|      |     |                                                                                                                                        |     |                                                                                                                                                                                              |
|------|-----|----------------------------------------------------------------------------------------------------------------------------------------|-----|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|      |     |                                                                                                                                        |     | TGACCGGAACTTGCGTCACTTCCCCCTCCC<br>CTGATTGGCTGGTTCGAACGAACGAACCCT<br>CCAATGAGACTCAAGGACAAGAGGATATTT<br>TGCGCGCCAGGAAGTG <b>(SEQ ID NO: 173)</b>                                               |
|      | v1  | secuencia<br>mínima<br>basada en la<br>comparación<br>con AAV2                                                                         | 145 | TTGACCACGCGCATGCCGCGCGGTGAGCC<br>CAATAGTTAAGCCGGGTGACCACACGTGAC<br>AGGAAGCACGGGATGTGCGTCACCGGAAG<br>CAGTGACCGGGCTGGTTCGAACGAACGAA<br>CCCTCCAACTCAAGGACAAGAGGATATT<br><b>(SEQ ID NO: 174)</b> |
|      | v2  | excluye los<br>primeros 162<br>nucleótidos y<br>los 162<br>nucleótidos<br>correspondien<br>tes<br>complementar<br>ios en<br>palíndromo | 120 | CGGTGACGTGTTTCCGGCTGTTAGGTTGAC<br>CACGCGCATGCCGCGCGGTGAGCCCAATA<br>GTTAAGCCGGAACACGTACCGACTCAAG<br>GACAAGAGGATATTTTGCGCGCCAGGAAGT<br><b>G (SEQ ID NO: 175)</b>                               |
|      | v3  | secuencia<br>mínima<br>basada en la<br>comparación<br>con B19                                                                          | 102 | GGGAACAATCAGGGGAAGTGACCGGTGAC<br>GTCATGTAAGTTGCGTCACTTCCCGTTCTGA<br>ACGAACGAACGAGACTCAAGGACAAGAGG<br>CGCGCCAGGAAGTG <b>(SEQ ID NO: 176)</b>                                                  |
| AAV2 | wt  | Gene Bank:<br>NC_001401.2                                                                                                              | 145 | TTGGCCACTCCCTCTCTGCGGCTCGCTCG<br>CTCACTGAGGCCGGGCGACCAAAGGTCGC<br>CCGACGCCCCGGGCTTTGCCCGGGCGGCCT<br>CAGTGAGCGAGCGAGCGCGCAGAGAGGG<br>AGTGGCCAACTCCATCACTAGGGGTTCTT<br><b>(SEQ ID NO: 177)</b> |
|      | GTx | usado en los<br>vectores GTx                                                                                                           | 130 | CTGCGCGCTCGCTCGCTCACTGAGGCCGC<br>CCGGGCAAAGCCCGGGCGTCGGGCGACC<br>TTTGGTCGCCCCGGCCTCAGTGAGCGAGCG<br>AGCGCGCAGAGAGGGAGTGGCCAACTCCA<br>TCACTAGGGGTTCTT <b>(SEQ ID NO: 178)</b>                  |

**[00139]** En ciertas realizaciones, la primera ITR y/o la segunda ITR comprende una secuencia de nucleótidos, en donde la secuencia de nucleótidos comprende la secuencia de nucleótidos mínima establecida en SEQ ID NO: 169, y en donde la secuencia de nucleótidos es al menos aproximadamente 50%, al menos

aproximadamente 55%, al menos aproximadamente 60%, al menos aproximadamente 65%, al menos aproximadamente 65%, al menos aproximadamente 70%, al menos aproximadamente 75%, al menos aproximadamente 80%, al menos aproximadamente 85%, al menos aproximadamente 90%, al menos aproximadamente 95%, al menos aproximadamente 96%, al menos aproximadamente 97%, al menos aproximadamente 98%, al menos aproximadamente 99% o 100% idéntica a la secuencia de nucleótidos establecida en SEQ ID NO: 167, conserva una propiedad funcional de la ITR B19 de la que deriva. En algunas realizaciones, la primera ITR y/o la segunda ITR comprenden una secuencia de nucleótidos, en donde la secuencia de nucleótidos comprende la secuencia de nucleótidos mínima establecida en SEQ ID NO: 169, y en donde la secuencia de nucleótidos es al menos aproximadamente 50%, a al menos aproximadamente 55%, al menos aproximadamente 60%, al menos aproximadamente 65%, al menos aproximadamente 65%, al menos aproximadamente 70%, al menos aproximadamente 75%, al menos aproximadamente 80%, al menos aproximadamente 85%, al menos aproximadamente 90%, al menos aproximadamente 95%, al menos aproximadamente 96%, al menos aproximadamente 97%, al menos aproximadamente 98%, al menos aproximadamente 99% o 100% idéntica con la secuencia de nucleótidos establecida en SEQ ID NO: 167, en donde la primera ITR y/o la segunda ITR es capaz de formar una estructura de horquilla. En ciertas realizaciones, la estructura de horquilla no comprende una horquilla en forma de T.

**[00140]** En algunas realizaciones, la primera ITR se deriva de un genoma de AAV, y la segunda ITR se deriva de GPV. En otras realizaciones, la segunda ITR se deriva de un genoma de AAV, y la primera ITR se deriva de GPV.

**[00141]** En ciertas realizaciones, la primera ITR y/o la segunda ITR comprenden o consisten en todo o una parte de una ITR derivada de GPV. En algunas realizaciones, la primera ITR y/o la segunda ITR comprenden o consisten en una secuencia de nucleótidos al menos aproximadamente 50%, al menos aproximadamente 55%, al menos aproximadamente 60%, al menos aproximadamente 65%, al menos aproximadamente 65%, al menos aproximadamente 70%, al menos aproximadamente 75%, al menos aproximadamente 80%, al menos aproximadamente 85%, al menos aproximadamente 90%, al menos aproximadamente 95%, al menos aproximadamente 96%, al menos aproximadamente 97%, al menos aproximadamente 98%, al menos aproximadamente 99%, o 100% idéntica a una secuencia de nucleótidos seleccionada de SEQ ID NOS: 172, 173, 174, 175 y 176, en donde la primera ITR y/o la segunda ITR retiene una propiedad funcional de la ITR GPV de la

que se deriva. En algunas realizaciones, la primera ITR y/o la segunda ITR comprenden o consisten en todo o una parte de una ITR derivada de GPV. En algunas realizaciones, la primera ITR y/o la segunda ITR comprenden o consisten en una secuencia de nucleótidos al menos aproximadamente 50%, al menos aproximadamente 55%, al menos aproximadamente 60%, al menos aproximadamente 65%, al menos aproximadamente 65%, al menos aproximadamente 70%, al menos aproximadamente 75%, al menos aproximadamente 80%, al menos aproximadamente 85%, al menos aproximadamente 90%, al menos aproximadamente 95%, al menos aproximadamente 96%, al menos aproximadamente 97%, al menos aproximadamente 98%, al menos aproximadamente 99%, o 100% idéntica a una secuencia de nucleótidos seleccionada de SEQ ID NOs: 172, 173, 174, 175 y 176, en donde la primera ITR y/o la segunda ITR es capaz de formar una estructura de horquilla. En ciertas realizaciones, la estructura de horquilla no comprende una horquilla en forma de T. En algunas realizaciones, la primera ITR y/o la segunda ITR comprenden o consisten en una secuencia de nucleótidos seleccionada de SEQ ID NO: 172, 173, 174, 175 y 176. En algunas realizaciones, la primera ITR y/o la segunda ITR comprenden o consisten en la secuencia de nucleótidos establecida en SEQ ID NO: 172. En algunas realizaciones, la primera ITR y/o la segunda ITR comprenden o consisten en la secuencia de nucleótidos establecida en SEQ ID NO: 173. En algunas realizaciones, la primera ITR y/o la segunda ITR comprenden o consisten en la secuencia de nucleótidos establecida en SEQ ID NO: 174. En algunas realizaciones, la primera ITR y/o la segunda ITR comprenden o consisten en la secuencia de nucleótidos establecida en SEQ ID NO: 175. En algunas realizaciones, la primera ITR y/o la segunda ITR comprenden o consisten en la secuencia de nucleótidos establecida en SEQ ID NO: 176.

**[00142]** En ciertas realizaciones, la primera ITR y/o la segunda ITR comprenden una secuencia de nucleótidos, en donde la secuencia de nucleótidos comprende la secuencia de nucleótidos mínima establecida en SEQ ID NO: 174, y en donde la secuencia de nucleótidos es al menos aproximadamente 50%, al menos aproximadamente 55%, al menos aproximadamente 60%, al menos aproximadamente 65%, al menos aproximadamente 65%, al menos aproximadamente 70%, al menos aproximadamente 75%, al menos aproximadamente 80%, al menos aproximadamente 85%, al menos aproximadamente 90%, al menos aproximadamente 95%, al menos aproximadamente 96%, al menos aproximadamente 97%, al menos aproximadamente 98%, al menos aproximadamente 99% o 100% idéntica a la secuencia de nucleótidos

establecida en SEQ ID NO: 172, en donde la primera ITR y/o la segunda ITR retiene una propiedad funcional de la ITR GPV de la cual se deriva. En algunas realizaciones, la primera ITR y/o la segunda ITR comprenden una secuencia de nucleótidos, en donde la secuencia de nucleótidos comprende la secuencia de nucleótidos mínima establecida en SEQ ID NO: 174, y en donde la secuencia de nucleótidos es al menos aproximadamente 50%, al menos aproximadamente 55%, al menos aproximadamente 60%, al menos aproximadamente 65%, al menos aproximadamente 65%, al menos aproximadamente 70%, al menos aproximadamente 75%, al menos aproximadamente 80%, al menos aproximadamente 85%, al menos aproximadamente 90%, al menos aproximadamente 95%, al menos aproximadamente 96%, al menos aproximadamente 97%, al menos aproximadamente 98%, al menos aproximadamente 99% o 100% idéntica a la secuencia de nucleótidos establecida en la SEQ ID NO : 172, en donde la primera ITR y/o la segunda ITR es capaz de formar una estructura de horquilla. En ciertas realizaciones, la estructura de horquilla no comprende una horquilla en forma de T.

**[00143]** En ciertas realizaciones, la primera ITR y/o la segunda ITR comprenden una secuencia de nucleótidos, en donde la secuencia de nucleótidos comprende la secuencia de nucleótidos mínima establecida en SEQ ID NO: 176, y en donde la secuencia de nucleótidos es al menos aproximadamente 50%, al menos aproximadamente 55%, al menos aproximadamente 60%, al menos aproximadamente 65%, al menos aproximadamente 65%, al menos aproximadamente 70%, al menos aproximadamente 75%, al menos aproximadamente 80%, al menos aproximadamente 85%, al menos aproximadamente 90%, al menos aproximadamente 95%, al menos aproximadamente 96%, al menos aproximadamente 97%, al menos aproximadamente 98%, al menos aproximadamente 99% o 100% idéntica a la secuencia de nucleótidos establecida en SEQ ID NO: 172, en donde la primera ITR y/o la segunda ITR retiene una propiedad funcional de la ITR GPV de la cual se deriva. En algunas realizaciones, la primera ITR y/o la segunda ITR comprenden una secuencia de nucleótidos, en donde la secuencia de nucleótidos comprende la secuencia de nucleótidos mínima establecida en SEQ ID NO: 176, y en donde la secuencia de nucleótidos es al menos aproximadamente 50%, al menos aproximadamente 55%, al menos aproximadamente 60%, al menos aproximadamente 65%, al menos aproximadamente 65%, al menos aproximadamente 70%, al menos aproximadamente 75%, al menos aproximadamente 80%, al menos aproximadamente 85%, al menos aproximadamente 90%, al menos aproximadamente 95%, al menos aproximadamente 96%, al menos aproximadamente

97%, al menos aproximadamente 98%, al menos aproximadamente 99% o 100% idéntica a la secuencia de nucleótidos establecida en la SEQ ID NO : 172, en donde la primera ITR y/o la segunda ITR es capaz de formar una estructura de horquilla. En ciertas realizaciones, la estructura de horquilla no comprende una horquilla en forma de T.

**[00144]** En ciertas realizaciones, una de las primeras ITR o las segundas ITR comprende o consiste en todo o una parte de una ITR derivada de AAV2. En algunas realizaciones, la primera ITR o la segunda ITR comprende o consiste en una secuencia de nucleótidos al menos aproximadamente 50%, al menos aproximadamente 55%, al menos aproximadamente 60%, al menos aproximadamente 65%, al menos aproximadamente 65%, al menos aproximadamente 70%, al menos aproximadamente 75%, al menos aproximadamente 80%, al menos aproximadamente 85%, al menos aproximadamente 90%, al menos aproximadamente 95%, al menos aproximadamente 96%, al menos aproximadamente 97%, al menos aproximadamente 98%, al menos aproximadamente 99%, o 100% idéntica a una secuencia de nucleótidos establecida en SEQ ID NOs: 177 o 178, en donde la primera ITR y/o la segunda ITR retiene una propiedad funcional de la ITR de AAV2 de la cual se deriva. En algunas realizaciones, la primera ITR o la segunda ITR comprende o consiste en una secuencia de nucleótidos al menos aproximadamente 50%, al menos aproximadamente 55%, al menos aproximadamente 60%, al menos aproximadamente 65%, al menos aproximadamente 65%, al menos aproximadamente 70%, al menos aproximadamente 75%, al menos aproximadamente 80%, al menos aproximadamente 85%, al menos aproximadamente 90%, al menos aproximadamente 95%, al menos aproximadamente 96%, al menos aproximadamente 97%, al menos aproximadamente 98%, al menos aproximadamente 99%, o 100% idéntica a una secuencia de nucleótidos establecida en SEQ ID NOs: 177 o 178, en donde la primera ITR y/o la segunda ITR es capaz de formar una estructura de horquilla. En ciertas realizaciones, la estructura de horquilla no comprende una horquilla en forma de T. En algunas realizaciones, la primera ITR y/o la segunda ITR comprenden o consisten en una secuencia de nucleótidos establecida en SEQ ID NOs: 177 o 178. En algunas realizaciones, la primera ITR y/o la segunda ITR comprenden o consisten en la secuencia de nucleótidos establecida en SEQ ID NO: 177. En algunas realizaciones, la primera ITR y/o la segunda ITR comprenden o consisten en la secuencia de nucleótidos establecida en la SEQ ID NO: 178.

**[00145]** En algunas realizaciones, la primera ITR se deriva de un genoma de AAV, y la segunda ITR se deriva de una cepa de parvovirus de pato de Muscovy (MDPV). En otras realizaciones, la segunda ITR se deriva de un genoma de AAV, y la primera ITR se deriva de una cepa de parvovirus de pato de Muscovy (MDPV). En ciertas realizaciones, la cepa de MDPV está atenuada, por ejemplo, la cepa de MDPV FZ91-30. En otras realizaciones, la cepa MDPV es patógena, por ejemplo, la cepa MDPV YY.

**[00146]** En algunas realizaciones, la primera ITR se deriva de un genoma AAV, y la segunda ITR se deriva de un *Dependoparvovirus*. En algunas realizaciones, la segunda ITR se deriva de un genoma AAV, y la primera ITR se deriva de un *Dependoparvovirus*. En otras realizaciones, la primera ITR se deriva de un genoma de AAV, y la segunda ITR se deriva de una cepa de parvovirus de ganso de *Dependovirus* (GPV). En otras realizaciones, la segunda ITR se deriva de un genoma de AAV, y la primera ITR se deriva de una cepa de GPV de *Dependovirus*. En ciertas realizaciones, la cepa GPV está atenuada, por ejemplo, la cepa GPV 82-0321V. En otras realizaciones, la cepa de GPV es patogénica, por ejemplo, la cepa B de GPV.

**[00147]** En ciertas realizaciones, la primera ITR se deriva de un genoma de AAV, y la segunda ITR se deriva de un genoma seleccionado del grupo que consiste en parvovirus porcino, por ejemplo, la cepa de parvovirus porcino U44978; virus minucioso de ratones, por ejemplo, la cepa de virus diminuto de ratones U34256; parvovirus canino, por ejemplo, la cepa de parvovirus canino M19296; el virus de la enteritis del visón, p. ej., cepa del virus de la enteritis del visón D00765; y cualquiera de sus combinaciones. En otras realizaciones, la segunda ITR se deriva de un genoma de AAV, y la primera ITR se deriva de un genoma seleccionado del grupo que consiste en parvovirus porcino, por ejemplo, la cepa de parvovirus porcino U44978; virus minucioso de ratones, por ejemplo, cepa de virus diminuto de ratones U34256; parvovirus canino, por ejemplo, la cepa de parvovirus canino M19296; virus de la enteritis del visón, p. ej., la cepa del virus de la enteritis del visón D00765; y cualquiera de sus combinaciones.

**[00148]** En otra realización particular, la ITR es una secuencia sintética modificada genéticamente para incluir en sus extremos 5' y 3' ITR no derivados de un genoma de AAV. En otra realización particular, la ITR es una secuencia sintética diseñada genéticamente para incluir en sus extremos 5' y 3' ITR derivados de uno o más de los genomas VnoAA. Las dos ITR presentes en la molécula de ácido nucleico de la invención pueden ser los mismos o diferentes genomas VnoAA. En particular,

las ITR pueden derivarse del mismo genoma VnoAA. En una realización específica, las dos ITR presentes en la molécula de ácido nucleico de la invención son iguales y, en particular, pueden ser ITR AAV2.

**[00149]** En algunas realizaciones, la secuencia ITR comprende una o más secuencias palindrómicas. Una secuencia palindrómica de una ITR descrita en el presente documento incluye, entre otras, secuencias palindrómicas nativas (es decir, secuencias que se encuentran en la naturaleza), secuencias sintéticas (es decir, secuencias que no se encuentran en la naturaleza), tales como secuencias pseudo palindrómicas, y combinaciones o formas modificadas de las mismas. Una "secuencia pseudo palindrómica" es una secuencia de ADN palindrómica, que incluye una secuencia palindrómica imperfecta, que comparte menos del 80%, incluyendo menos del 70%, 60%, 50%, 40%, 30%, 20%, 10% o 5 %, o no, identidad de secuencia de ácido nucleico con secuencias en una secuencia palindrómica AAV nativa o VnoAA que forma una estructura secundaria. Las secuencias palindrómicas nativas se pueden obtener o derivar de cualquier genoma descrito en este documento. La secuencia palindrómica sintética puede basarse en cualquier genoma descrito en este documento.

**[00150]** La secuencia palindrómica puede ser continua o interrumpida. En algunas realizaciones, la secuencia palindrómica se interrumpe, en donde la secuencia palindrómica comprende una inserción de una segunda secuencia. En algunas realizaciones, la segunda secuencia comprende un promotor, un potenciador, un sitio de integración para una integrasa (por ejemplo, sitios para la recombinasa Cre o Flp), un marco de lectura abierto para un producto génico, o una de sus combinaciones.

**[00151]** En algunas realizaciones, las ITR forman estructuras de bucle de horquilla. En una realización, la primera ITR forma una estructura de horquilla. En otra realización, la segunda ITR forma una estructura de horquilla. Todavía en otra realización, tanto la primera ITR como la segunda ITR forman estructuras de horquilla. En algunas realizaciones, la primera ITR y/o la segunda ITR no forman una estructura de horquilla en forma de T. En ciertas realizaciones, la primera ITR y/o la segunda ITR forman una estructura de horquilla sin forma de T. En algunas realizaciones, la estructura de horquilla sin forma de T comprende una estructura de horquilla en forma de U.

**[00152]** En algunas realizaciones, una ITR en una molécula de ácido nucleico descrita en el presente documento puede ser una ITR activada por transcripción. Una

ITR activada por transcripción puede comprender todo o una parte de una ITR de tipo salvaje que se ha activado por transcripción mediante la inclusión de al menos un elemento transcripcionalmente activo. Varios tipos de elementos transcripcionalmente activos son adecuados para su uso en este contexto. En algunas realizaciones, el elemento transcripcionalmente activo es un elemento constitutivo transcripcionalmente activo. Los elementos constitutivos transcripcionalmente activos proporcionan un nivel continuo de la transcripción génica, y se prefieren cuando se desea que el transgén se exprese de forma continua. En otras realizaciones, el elemento transcripcionalmente activo es un elemento inducible transcripcionalmente activo. Los elementos transcripcionalmente activos inducibles generalmente exhiben una baja actividad en ausencia de un inductor (o condición inductora), y están sobre-regulados por el aumento en presencia del inductor (o cambian a una condición inductora). Los elementos inducibles transcripcionalmente activos pueden preferirse cuando la expresión se desea solo en ciertos momentos o en ciertos lugares, o cuando es deseable valorar el nivel de expresión usando un agente inductor. Los elementos transcripcionalmente activos también pueden ser específicos de tejido; es decir, exhiben actividad solo en ciertos tejidos o tipos de células.

**[00153]** Los elementos transcripcionalmente activos se pueden incorporar a una ITR de varias maneras. En algunas realizaciones, un elemento transcripcionalmente activo se incorpora 5' a cualquier porción de una ITR o 3' a cualquier porción de una ITR. En otras realizaciones, un elemento transcripcionalmente activo de una ITR activado transcripcionalmente se encuentra entre dos secuencias de ITR. Si el elemento transcripcionalmente activo comprende dos o más elementos que deben estar separados, esos elementos pueden alternarse con partes de la ITR. En algunas realizaciones, una estructura de horquilla de una ITR se elimina y se reemplaza con repeticiones invertidas de un elemento transcripcional. Esta última disposición crearía una horquilla que imita la porción eliminada en la estructura. Múltiples elementos activos transcripcionalmente en tándem también pueden estar presentes en una ITR activada transcripcionalmente, y estos pueden estar adyacentes o separados. Además, los sitios de unión a proteínas (p. ej., los sitios de unión a Rep) pueden introducirse en elementos transcripcionalmente activos de las ITR activadas transcripcionalmente. Un elemento transcripcionalmente activo puede comprender cualquier secuencia que permita la transcripción controlada de ADN por la ARN polimerasa para formar ARN, y puede comprender, por ejemplo, un elemento transcripcionalmente activo, como se define a continuación.

**[00154]** Las ITR activadas transcripcionalmente proporcionan tanto la activación transcripcional como las funciones de ITR a la molécula de ácido nucleico en una longitud de secuencia de nucleótidos relativamente limitada que maximiza de forma efectiva la longitud de un transgén que puede transportarse y expresarse desde la molécula de ácido nucleico. La incorporación de un elemento transcripcionalmente activo en una ITR se puede lograr de varias maneras. Una comparación de la secuencia ITR y los requisitos de la secuencia del elemento transcripcionalmente activo pueden proporcionar información sobre formas de codificar el elemento dentro de una ITR. Por ejemplo, la actividad transcripcional se puede agregar a una ITR mediante la introducción de cambios específicos en la secuencia de ITR que replica los elementos funcionales del elemento transcripcionalmente activo. Existen varias técnicas en la técnica para agregar, eliminar y/o cambiar secuencias de nucleótidos particulares en sitios específicos (véase, por ejemplo, Deng y Nickoloff (1992) Anal. Biochem. 200:81-88). Otra forma de crear ITR activadas por transcripción implica la introducción de un sitio de restricción en una ubicación deseada en la ITR. Además, se pueden incorporar múltiples elementos activados transcripcionalmente en una ITR activada transcripcionalmente, usando métodos conocidos en la técnica.

**[00155]** A modo de ilustración, las ITR activadas transcripcionalmente se pueden generar mediante la inclusión de uno o más elementos transcripcionalmente activos tales como: caja TATA, caja GC, caja CCAAT, sitio Spl, región Inr, sitio CRE (elemento regulador cAMP), sitio ATF-1/CRE, caja APB $\beta$ , caja APB $\alpha$ , caja CArG, caja CCAC o cualquier otro elemento involucrado en la transcripción como se conoce en la técnica.

## **B. Proteínas terapéuticas**

**[00156]** Ciertos aspectos de la presente descripción están dirigidos a una molécula de ácido nucleico que comprende una primera ITR, una segunda ITR y un casete génico que codifica una proteína terapéutica. En algunas realizaciones, el casete génico codifica una proteína terapéutica. En algunas realizaciones, el casete génico codifica más de una proteína terapéutica. En algunas realizaciones, el casete génico codifica dos o más copias de la misma proteína terapéutica. En algunas realizaciones, el casete génico codifica dos o más variantes de la misma proteína terapéutica. En algunas realizaciones, el casete génico codifica dos o más proteínas terapéuticas diferentes.

**[00157]** Ciertas realizaciones de la presente descripción están dirigidas a una molécula de ácido nucleico que comprende una primera ITR, una segunda ITR y un casete génico que codifica una proteína terapéutica, en donde la proteína terapéutica

comprende un factor de coagulación. En algunas realizaciones, el factor de coagulación se selecciona del grupo que consiste en FI, FII, FIII, FIV, FV, FVI, FVII, FVIII, FIX, FX, FXI, FXII, FXIII), VWF, precalicreína, quinínogeno de alto peso molecular, fibronectina, antitrombina III, cofactor de heparina II, proteína C, proteína S, proteína Z, inhibidor de la proteasa relacionada con la proteína Z (ZPI), plasminógeno, alfa 2-antiplasmina, activador del plasminógeno tisular (tPA), uroquinasa, inhibidor del activador del plasminógeno-1 (PAI-1), inhibidor del activador del plasminógeno-2 (PAI2), cualquier zimógeno del mismo, cualquier forma activa del mismo y cualquiera de sus combinaciones. En una realización, el factor de coagulación comprende FVIII o una variante o fragmento del mismo. En otra realización, el factor de coagulación comprende FIX o una variante o fragmento del mismo. En otra realización, el factor de coagulación comprende FVII o una variante o fragmento del mismo. En otra realización, el factor de coagulación comprende VWF o una variante o fragmento del mismo.

## **1. Factores de coagulación**

**[00158]** En algunas realizaciones, la molécula de ácido nucleico comprende una primera ITR, una segunda ITR y un casete génico que codifica una proteína terapéutica, en donde la proteína terapéutica comprende un polipéptido del factor VIII. "Factor VIII", abreviado a lo largo de la presente solicitud como "FVIII", según se usa en este documento, significa un polipéptido de FVIII funcional en su papel normal en la coagulación, a menos que se especifique lo contrario. Por lo tanto, el término FVIII incluye polipéptidos variantes que son funcionales. "Una proteína FVIII" se usa indistintamente con el polipéptido (o la proteína) FVIII o FVIII. Los ejemplos de las funciones del FVIII incluyen, entre otros, la capacidad de activar la coagulación, la capacidad de actuar como cofactor del factor IX o la capacidad de formar un complejo de tenasa con el factor IX en presencia de  $\text{Ca}^{2+}$  y fosfolípidos, que luego convierten el Factor X en la forma activada Xa. La proteína FVIII puede ser la proteína FVIII humana, porcina, canina, de rata o murina. Además, las comparaciones entre el FVIII de humanos y otras especies han identificado residuos conservados que probablemente sean necesarios para la función (Cameron et al, Thromb. Haemost 79:317-22 (1998); documento US 6.251.632). Se conocen las secuencias de polipéptidos y polinucleótidos de longitud completa, al igual que muchos fragmentos funcionales, mutantes y versiones modificadas. Varias secuencias de aminoácidos y nucleótidos del FVIII se describen, por ejemplo, en las publicaciones de EE. UU. 2015/0158929 AI, 2014/0308280 AI, y 2014/0370035 AI y la Publicación Internacional

No. WO 2015/106052 AI. Los polipéptidos de FVIII incluyen, por ejemplo, FVIII de longitud completa, FVIII de longitud completa menos Met en el extremo N, FVIII maduro (menos la secuencia de señal), FVIII maduro con un Met adicional en el extremo N, y/o FVIII con un eliminación total o parcial del dominio B. Las variantes de FVIII incluyen eliminaciones de dominio B, ya sean eliminaciones parciales o completas.

**a. Secuencias de FVIII y polinucleótidos que codifican la proteína FVIII**

**[00159]** En algunas realizaciones, la molécula de ácido nucleico comprende una primera ITR, una segunda ITR y un casete génico que codifica una proteína terapéutica, en donde la proteína terapéutica comprende un polipéptido del factor VIII. "Factor VIII", abreviado a lo largo de la presente solicitud como "FVIII", según se usa en este documento, significa el polipéptido de FVIII funcional en su papel normal en la coagulación, a menos que se especifique lo contrario. Por lo tanto, el término FVIII incluye polipéptidos variantes que son funcionales. "Una proteína FVIII" se usa indistintamente con el polipéptido (o proteína) FVIII o FVIII. Los ejemplos de las funciones del FVIII incluyen, entre otros, la capacidad de activar la coagulación, la capacidad de actuar como cofactor del factor IX o la capacidad de formar un complejo de tenasa con el factor IX en presencia de  $\text{Ca}^{2+}$  y fosfolípidos, que luego convierten el Factor X en la forma activada Xa. La proteína FVIII puede ser la proteína FVIII humana, porcina, canina, de rata o murina. Además, las comparaciones entre el FVIII de humanos y otras especies han identificado residuos conservados que probablemente sean necesarios para la función (Cameron et al, Thromb. Haemost 79:317-22 (1998); documento US 6.251.632). Las secuencias de polipéptidos y polinucleótidos de longitud completa son conocidas, al igual que muchos fragmentos funcionales, mutantes y versiones modificadas. Varias secuencias de aminoácidos y nucleótidos del FVIII se describen, por ejemplo, en las publicaciones de EE. UU. 2015/0158929 AI, 2014/0308280 AI, y 2014/0370035 AI y Publicación Internacional No. WO 2015/106052 AI. Los polipéptidos de FVIII incluyen, por ejemplo, FVIII de longitud completa, FVIII de longitud completa menos Met en el extremo N, FVIII maduro (menos la secuencia de señal), FVIII maduro con un Met adicional en el extremo N, y/o FVIII con un eliminación total o parcial del dominio B. Las variantes de FVIII incluyen eliminaciones de dominio B, ya sean eliminaciones parciales o completas.

**[00160]** La porción de FVIII en la proteína quimérica usada en este documento tiene actividad de FVIII. La actividad de FVIII se puede medir por cualquier método

conocido en la técnica. Hay varias pruebas disponibles para evaluar la función del sistema de coagulación: prueba del tiempo de tromboplastina parcial activada (aPTT), prueba cromogénica, prueba ROTEM, prueba de tiempo de protrombina (PT) (también utilizada para determinar INR), prueba de fibrinógeno (a menudo por el Método de Clauss), recuento de plaquetas, pruebas de la función plaquetaria (a menudo por PFA-100), TCT, tiempo de sangrado, prueba de mezcla (si una anomalía corrige si el plasma del paciente se mezcla con plasma normal), ensayos del factor de coagulación, anticuerpos antifosfolípidos, D-dímero, pruebas genéticas (p. ej., factor V Leiden, mutación de protrombina G20210A), tiempo diluido de veneno de víbora de Russell (dRVVT), diversas pruebas de función plaquetaria, tromboelastografía (TEG o Sonoclot), tromboelastometría (TEM®, p. ej., ROTEM®) o tiempo de lisis de euglobulina (ELT).

**[00161]** La prueba aPTT es un indicador del rendimiento que mide la eficacia tanto de la "ruta intrínseca" (también referida a la vía de activación del contacto) como de las vías de coagulación comunes. Esta prueba se usa comúnmente para medir la actividad de coagulación de factores de coagulación recombinantes disponibles comercialmente, por ejemplo, FVIII. Se utiliza junto con el tiempo de protrombina (PT), que mide la vía extrínseca.

**[00162]** El análisis ROTEM proporciona información sobre la cinética completa de la hemostasia: tiempo de coagulación, formación de coágulos, estabilidad del coágulo y lisis. Los diferentes parámetros en la tromboelastometría dependen de la actividad del sistema de coagulación plasmática, la función plaquetaria, la fibrinólisis o muchos factores que influyen en estas interacciones. Este ensayo puede proporcionar una vista completa de la hemostasia secundaria.

**[00163]** El mecanismo de ensayo cromogénico se basa en los principios de la cascada de coagulación de la sangre, donde el FVIII activado acelera la conversión del Factor X en Factor Xa en presencia de Factor IX activado, fosfolípidos e iones de calcio. La actividad del Factor Xa se evalúa por hidrólisis de un sustrato de p-nitroanilida (pNA) específico para el Factor Xa. La tasa inicial de liberación de p-nitroanilina medida a 405 nM es directamente proporcional a la actividad del Factor Xa y, por lo tanto, a la actividad del FVIII en la muestra.

**[00164]** El ensayo cromogénico es recomendado por el Subcomité FVIII y Factor IX del Comité Científico y de Normalización (SSC) de la Sociedad Internacional de Trombosis y Hemostasis (ISTH). Desde 1994, el ensayo cromogénico también ha sido el método de referencia de la Farmacopea Europea para la asignación de la

potencia del concentrado de FVIII. Por lo tanto, en una realización, el polipéptido quimérico que comprende FVIII tiene una actividad FVIII comparable a un polipéptido quimérico que comprende FVIII maduro o un FVIII BDD (por ejemplo, ADVATE®, REFACTO®, o ELOCTATE®).

5 **[00165]** En otra realización, la proteína quimérica que comprende FVIII de esta descripción tiene una tasa de generación de Factor Xa comparable a una proteína quimérica que comprende FVIII maduro o un FVIII BDD (por ejemplo, ADVATE®, REFACTO® o ELOCTATE®).

**[00166]** Para activar el Factor X al Factor Xa, el Factor IX activado (Factor IXa) hidroliza un enlace arginina-isoleucina en el Factor X para formar el Factor Xa en presencia de  $\text{Ca}^{2+}$ , fosfolípidos de membrana y un cofactor FVIII. Por lo tanto, la interacción del FVIII con el Factor IX es crítica en la vía de coagulación. En ciertas realizaciones, el polipéptido quimérico que comprende FVIII puede interactuar con el Factor IXa a una velocidad comparable a un polipéptido quimérico que comprende una secuencia de FVIII madura o un FVIII BDD (por ejemplo, ADVATE®, REFACTO®, o ELOCTATE®).

**[00167]** Además, el FVIII está unido al factor von Willebrand mientras está inactivo en circulación. El FVIII se degrada rápidamente cuando no se une al VWF y se libera del VWF por la acción de la trombina. En algunas realizaciones, el polipéptido quimérico que comprende FVIII se une al Factor von Willebrand a un nivel comparable a un polipéptido quimérico que comprende una secuencia de FVIII madura o un FVIII BDD (por ejemplo, ADVATE®, REFACTO® o ELOCTATE®).

**[00168]** El FVIII puede ser inactivado por la proteína C activada en presencia de calcio y fosfolípidos. La proteína C activada escinde la cadena pesada de FVIII después de la Arginina 336 en el dominio A1, que interrumpe un sitio de interacción del sustrato de Factor X, y escinde después de la Arginina 562 en el dominio A2, lo que mejora la disociación del dominio A2, así como también interrumpe un sitio de interacción con el Factor IXa. Esta división también divide el dominio A2 (43 kDa) y genera dominios A2-N (18 kDa) y A2-C (25 kDa). Por lo tanto, la proteína C activada puede catalizar múltiples sitios de escisión en la cadena pesada. En una realización, el polipéptido quimérico que comprende FVIII es inactivado por la proteína C activada a un nivel comparable a un polipéptido quimérico que comprende una secuencia de FVIII madura o un FVIII BDD (p. ej., ADVATE®, REFACTO®, o ELOCTATE®).

**[00169]** En otras realizaciones, la proteína quimérica que comprende FVIII tiene actividad de FVIII in vivo comparable a un polipéptido quimérico que comprende una

secuencia de FVIII madura o un FVIII BDD (por ejemplo, ADVATE®, REFACTO® o ELOCTATE®). En una realización particular, el polipéptido quimérico que comprende FVIII es capaz de proteger un ratón HemA a un nivel comparable a un polipéptido quimérico que comprende una secuencia FVIII madura o un FVIII BDD (por ejemplo, ADVATE®, REFACTO® o ELOCTATE®) en un modelo de transección de la vena de la cola de ratón HemA.

**[00170]** Un "dominio B" de FVIII, como se usa en el presente documento, es el mismo que el dominio B conocido en la técnica que se define por la identidad de secuencia de aminoácidos interna y sitios de escisión proteolítica por trombina, por ejemplo, residuos Ser741-Arg1648 de FVIII humano maduro. Los otros dominios humanos de FVIII están definidos por los siguientes residuos de aminoácidos, en relación con el FVIII humano maduro: A1, residuos Ala1-Arg372; A2, residuos Ser373-Arg740; A3, residuos Ser1690-Ile2032; C1, residuos Arg2033-Asn2172; C2, residuos Ser2173-Tyr2332 de FVIII maduro. Los números de residuos de secuencia utilizados en este documento sin hacer referencia a ningún número de SEQ ID corresponden a la secuencia de FVIII sin la secuencia de péptido señal (19 aminoácidos) a menos que se indique lo contrario. La secuencia A3-C1-C2, también conocida como la cadena pesada del FVIII, incluye los residuos Ser1690-Tyr2332. La secuencia restante, los residuos Glu1649-Arg1689, generalmente se denomina péptido de activación de la cadena ligera del FVIII. Las ubicaciones de los límites para todos los dominios, incluidos los dominios B, para FVIII porcino, de ratón y canino también se conocen en la técnica. En una realización, el dominio B del FVIII se elimina ("FVIII con dominio B eliminado" o "FVIII BDD"). Un ejemplo de un FVIII BDD es REFACTO® (FVIII BDD recombinante). En una realización particular, la variante de FVIII eliminada del dominio B comprende una delección de los residuos de aminoácidos 746 a 1648 de FVIII maduro.

**[00171]** Un "FVIII con dominio B eliminado" puede tener las delecciones totales o parciales descritas en las patentes de EE.UU. Nos. 6.316.226, 6.346.513, 7.041.635, 5.789.203, 6.060.447, 5.595.886, 6.228.620, 5.972.885, 6.048.720, 5.543.502, 5.610.278, 5.171.844, 5.112.950, 4.868.112 y 6.458.563 y la publicación internacional No. WO 2015106052 A1 (documento PCT/US2015/010738). En algunas realizaciones, una secuencia de FVIII con dominio B eliminado utilizada en los métodos de la presente descripción comprende cualquiera de las delecciones descritas en la columna 4, línea 4 a col. 5, línea 28 y ejemplos 1-5 de la patente de EE.UU. Nº 6.316.226 (también en el documento US 6.346.513). En otra realización, un Factor VIII con

dominio B eliminado es el Factor VIII con el dominio B S743/Q1638 (FVIII SQ BDD) (por ejemplo, Factor VIII que tiene una delección del aminoácido 744 al aminoácido 1637, por ejemplo, Factor VIII que tiene los aminoácidos 1-743 y aminoácidos 1638-2332 de FVIII maduro). En algunas realizaciones, un FVIII con dominio B eliminado  
5 utilizado en los métodos de la presente descripción tiene una delección descrita en la col. 2, líneas 26-51 y ejemplos 5-8 de la Patente de Estados Unidos Nº 5.789.203 (también en los documentos US 6.060.447, US 5.595.886 y US 6.228.620). En algunas realizaciones, un Factor VIII con dominio B eliminado tiene una delección descrita en la col. 1, líneas 25 a col. 2, línea 40 de la Patente de Estados Unidos Nº  
10 5.972.885; columna 6, líneas 1-22 y ejemplo 1 de la patente de EE. UU. No. 6.048.720; columna 2, líneas 17-46 de la Patente de Estados Unidos Nº 5.543.502; columna 4, línea 22 a col. 5, línea 36 de la patente de EE. UU. No. 5.171.844; columna 2, líneas 55-68, figura 2 y ejemplo 1 de la Patente de Estados Unidos Nº 5.112.950; columna 2, línea 2 a col. 19, línea 21 y tabla 2 de la Patente de Estados Unidos Nº 4.868.112;  
15 columna 2, línea 1 a col. 3, línea 19, col. 3, línea 40 a col. 4, línea 67, col. 7, línea 43 a col. 8, línea 26 y col. 11, línea 5 a col. 13, línea 39 de la patente de EE. UU. No. 7.041.635; o col. 4, líneas 25-53 de la Patente de Estados Unidos Nº 6.458.563. En algunas realizaciones, un FVIII con dominio B eliminado tiene una delección de la mayoría del dominio B, pero aún contiene secuencias amino-terminales del dominio B  
20 que son esenciales para el procesamiento proteolítico in vivo del producto de traducción primario en dos cadenas de polipéptidos, como se describe en el documento WO 91/09122. En algunas realizaciones, un FVIII con dominio B eliminado se construye con una eliminación de los aminoácidos 747-1638, es decir, prácticamente una eliminación completa del dominio B. Hoebe R.C., et al. J. Biol.  
25 Chem 265 (13): 7318-7323 (1990). Un Factor VIII con dominio B eliminado también puede contener una delección de los aminoácidos 771-1666 o los aminoácidos 868-1562 del FVIII. Meulien P. et al. Protein Eng. 2(4): 301-6 (1988). Las otras delecciones del dominio B que son parte de la invención incluyen: la delección de los aminoácidos 982 a 1562 o de 760 a 1639 (Toole et al, Proc. Natl. Acad. Sci. U.S.A. (1986) 83, 5939-  
30 5942)), de 797 a 1562 (Eaton, et al. Biochemistry (1986) 25: 8343-8347)), de 741 a 1646 (Kaufman (solicitud publicada PCT No. WO 87/04187)) , 747-1560 (Sarver, et al, DNA (1987) 6: 553-564)), de 741 a 1648 (Pasek (solicitud PCT No.88/00831)), o de 816 a 1598 o de 741 a 1648 (Lagner (Behring Inst. Mitt. (1988) No 82: 16-25, documento EP 295597)). En una realización particular, el FVIII con dominio B  
35 eliminado comprende una delección de los residuos de aminoácidos 746 a 1648 del

FVIII maduro. En otra realización, el FVIII con dominio B eliminado comprende una delección de los residuos de aminoácidos 745 a 1648 del FVIII maduro. En algunas realizaciones, el FVIII BDD comprende FVIII monocatenario que contiene una delección en los aminoácidos 765 a 1652 correspondientes al FVIII de longitud completa madura (también conocido como rVIII-SingleChain y AFSTYLA®). Ver la Patente de Estados Unidos N° 7.041.635.

**[00172]** En otras realizaciones, FVIII BDD incluye un polipéptido FVIII que contiene fragmentos del dominio B que retienen uno o más sitios de glicosilación unidos a N, por ejemplo, residuos 757, 784, 828, 900, 963 u opcionalmente 943, que corresponden a la secuencia de aminoácidos de la secuencia de FVIII de longitud completa. Los ejemplos de los fragmentos del dominio B incluyen 226 aminoácidos o 163 aminoácidos del dominio B como se describe en Miao, HZ, et al, Blood 103 (a): 3412-3419 (2004), Kasuda, A, et al, J. Thromb. Haemost 6: 1352-1359 (2008), y Pipe, S.W., et al., J. Thromb. Haemost 9: 2235-2242 (2011) (es decir, se retienen los primeros 226 aminoácidos o 163 aminoácidos del dominio B). En otras realizaciones más, FVIII BDD comprende además una mutación puntual en el residuo 309 (de Phe a Ser) para mejorar la expresión de la proteína FVIII BDD. Ver Miao, H.Z., et al., Blood 103 (a): 3412-3419 (2004). En otras realizaciones más, el FVIII BDD incluye un polipéptido FVIII que contiene una porción del dominio B, pero que no contiene uno o más sitios de escisión de furina (por ejemplo, Arg1313 y Arg 1648). Ver Pipe, S.W., et al., J. Thromb. Haemost 9: 2235-2242 (2011). En algunas realizaciones, el FVIII BDD comprende FVIII monocatenario que contiene una delección en los aminoácidos 765 a 1652 correspondientes al FVIII maduro completo (también conocido como rVIII-SingleChain y AFSTYLA®). Ver la Patente de Estados Unidos N° 7.041.635. Cada una de las eliminaciones anteriores se puede realizar en cualquier secuencia de FVIII.

**[00173]** Se conocen muchas variantes funcionales de FVIII, como se discute antes y más abajo. Además, se han identificado cientos de mutaciones no funcionales en el FVIII en pacientes con hemofilia, y se ha determinado que el efecto de estas mutaciones en la función del FVIII se debe más al lugar donde se encuentran dentro de la estructura tridimensional del FVIII que a la naturaleza de la sustitución (Cutler et al, Hum. Mutat 79: 274-8 (2002)), incorporado en este documento como referencia en su totalidad. Además, las comparaciones entre el FVIII de humanos y otras especies han identificado residuos conservados que probablemente sean necesarios para la función (Cameron et al, Thromb. Haemost 79:317-22 (1998); documento US 6.251.632), incorporado en este documento como referencia en su totalidad.

**[00174]** En algunas realizaciones, el polipéptido FVIII comprende una variante de FVIII o un fragmento del mismo, en donde la variante de FVIII o el fragmento del mismo tiene una actividad de FVIII. En algunas realizaciones, el casete génico codifica un polipéptido FVIII de longitud completa. En otras realizaciones, el casete génico codifica un polipéptido FVIII con dominio B eliminado (BDD), en el que se elimina todo o una parte del dominio B de FVIII. En una realización particular, el casete génico codifica un polipéptido que comprende una secuencia de aminoácidos que tiene al menos aproximadamente 80%, al menos aproximadamente 85%, al menos aproximadamente 86%, al menos aproximadamente 87%, al menos aproximadamente 88%, al menos aproximadamente 89%, al menos aproximadamente 90%, al menos aproximadamente 95%, al menos aproximadamente 96%, al menos aproximadamente 97%, al menos aproximadamente 98% o al menos aproximadamente 99% de identidad de secuencia con las SEQ ID NO: 106, 107, 109, 110, 111 o 112. En algunas realizaciones, el casete génico codifica un polipéptido que tiene la secuencia de aminoácidos de SEQ ID NO: 17 o un fragmento del mismo. En algunas realizaciones, el casete génico codifica un polipéptido que tiene la secuencia de aminoácidos de SEQ ID NO: 106 o un fragmento del mismo. En algunas realizaciones, el casete génico comprende una secuencia de nucleótidos que tiene al menos 85%, al menos 90%, al menos 95%, al menos 96%, al menos 97%, al menos 98% o al menos 99% de identidad de secuencia con respecto a la SEQ ID NO: 107. En algunas realizaciones, el casete génico codifica un polipéptido que tiene la secuencia de aminoácidos de SEQ ID NO: 109 o un fragmento del mismo. En algunas realizaciones, el casete génico comprende una secuencia de nucleótidos que tiene al menos 85%, al menos 90%, al menos 95%, al menos 96%, al menos 97%, al menos 98% o al menos 99% de identidad de secuencia con la SEQ ID NO: 16. En algunas realizaciones, el casete génico comprende una secuencia de nucleótidos que tiene al menos 85%, al menos 90%, al menos 95%, al menos 96%, al menos 97%, al menos 98% o al menos 99% de identidad de secuencia con la SEQ ID NO: 109.

**[00175]** En algunas realizaciones, el casete génico de la descripción codifica un polipéptido FVIII que comprende un péptido señal o un fragmento del mismo. En otras realizaciones, el casete génico codifica un polipéptido FVIII que carece de un péptido señal. En algunas realizaciones, el péptido señal comprende los aminoácidos 1-19 de la SEQ ID NO: 17.

**[00176]** En algunas realizaciones, el casete génico comprende una secuencia de nucleótidos que codifica un polipéptido FVIII, en el que la secuencia de nucleótidos

está optimizada por codones. En ciertas realizaciones, el casete génico comprende una secuencia de nucleótidos que se describe en la Solicitud Internacional N° PCT/US2017/015879, que se incorpora por referencia en su totalidad. En algunas realizaciones, el casete génico comprende una secuencia de nucleótidos que codifica un polipéptido FVIII, en el que la secuencia de nucleótidos está optimizada por codones. En ciertas realizaciones, el casete génico comprende una secuencia de nucleótidos que tiene al menos un 85%, al menos un 90%, al menos un 95%, al menos un 96%, al menos un 97%, al menos un 98% o al menos un 99% de identidad de secuencia con una secuencia de nucleótidos seleccionada de SEQ ID NOs: 1-14. En algunas realizaciones, el casete génico comprende una secuencia de nucleótidos que tiene al menos 85%, al menos 90%, al menos 95%, al menos 96%, al menos 97%, al menos 98% o al menos 99% de identidad de secuencia con la SEQ ID NO: 71. En algunas realizaciones, el casete génico comprende una secuencia de nucleótidos que tiene al menos 85%, al menos 90%, al menos 95%, al menos 96%, al menos 97%, al menos 98% o al menos 99% de identidad de secuencia con la SEQ ID NO: 19.

**i. Secuencias de nucleótidos optimizadas con codones que codifican polipéptidos FVIII**

**[00177]** En algunas realizaciones, una molécula de ácido nucleico de la presente descripción comprende una primera ITR, una segunda ITR y un casete génico que codifica una proteína terapéutica, en donde la primera ITR y la segunda ITR se derivan de un genoma de AAV, y en donde el casete génico comprende una secuencia de nucleótidos optimizada por codón que codifica un polipéptido FVIII. En algunas realizaciones, la secuencia de nucleótidos optimizada con codón codifica un polipéptido FVIII de longitud completa. En otras realizaciones, la secuencia de nucleótidos optimizada con codón codifica un polipéptido FVIII con un dominio B eliminado (BDD), en el que se elimina todo o una parte del dominio B del FVIII. En una realización particular, la secuencia de nucleótidos optimizada por codón codifica un polipéptido que comprende una secuencia de aminoácidos que tiene al menos aproximadamente 80%, al menos aproximadamente 85%, al menos aproximadamente 86%, al menos aproximadamente 87%, al menos aproximadamente 88%, al menos aproximadamente 89%, al menos aproximadamente 90%, al menos aproximadamente 95%, al menos aproximadamente 96%, al menos aproximadamente 97%, al menos aproximadamente 98%, o al menos aproximadamente el 99% de identidad de secuencia con la SEQ ID NO: 17 o un fragmento de la misma. En una realización, la

secuencia de nucleótidos optimizada por codón codifica un polipéptido que tiene la secuencia de aminoácidos de SEQ ID NO: 17 o un fragmento de la misma.

**[00178]** En algunas realizaciones, la secuencia de nucleótidos optimizada con codón codifica un polipéptido FVIII que comprende un péptido señal o un fragmento del mismo. En otras realizaciones, la secuencia optimizada con codón codifica un polipéptido FVIII que carece de un péptido señal. En algunas realizaciones, el péptido señal comprende los aminoácidos 1-19 de la SEQ ID NO: 17.

**[00179]** En algunas realizaciones, la secuencia de nucleótidos optimizada por codón que codifica un polipéptido FVIII comprende una secuencia de nucleótidos que comprende una primera secuencia de ácido nucleico que codifica una porción N-terminal de un polipéptido FVIII y una segunda secuencia de ácido nucleico que codifica una porción C-terminal de un polipéptido FVIII; en donde la primera secuencia de ácido nucleico tiene al menos 85%, al menos 90%, al menos 95%, al menos 96%, al menos 97%, al menos 98% o al menos 99% de identidad de secuencia con (i) los nucleótidos 58-1791 de SEQ ID NO: 3 o (ii) los nucleótidos 58-1791 de SEQ ID NO: 4; y en donde la porción N-terminal y la porción C-terminal juntas tienen una actividad del polipéptido FVIII. En una realización particular, la primera secuencia de ácido nucleico tiene al menos 85%, al menos 90%, al menos 95%, al menos 96%, al menos 97%, al menos 98% o al menos 99% de identidad de secuencia con los nucleótidos 58-1791 de SEQ ID NO: 3. En otra realización, la primera secuencia de ácido nucleico tiene al menos un 85%, al menos un 90%, al menos un 95%, al menos un 96%, al menos un 97%, al menos un 98% o al menos un 99% de identidad de secuencia con los nucleótidos 58-1791 de SEQ ID NO: 4. En otras realizaciones, la primera secuencia de nucleótidos comprende los nucleótidos 58-1791 de SEQ ID NO: 3 o los nucleótidos 58-1791 de SEQ ID NO: 4.

**[00180]** En otras realizaciones, la secuencia de nucleótidos optimizada por codón que codifica un polipéptido FVIII comprende una secuencia de nucleótidos que comprende una primera secuencia de ácido nucleico que codifica una porción N-terminal de un polipéptido FVIII y una segunda secuencia de ácido nucleico que codifica una porción C-terminal de un polipéptido FVIII; en donde la primera secuencia de ácido nucleico tiene al menos 85%, al menos 90%, al menos 95%, al menos 96%, al menos 97%, al menos 98% o al menos 99% de identidad de secuencia con (i) los nucleótidos 1-1791 de SEQ ID NO: 3 o (ii) los nucleótidos 1-1791 de SEQ ID NO: 4; y en donde la porción N-terminal y la porción C-terminal juntas tienen una actividad de polipéptido FVIII. En una realización, la primera secuencia de nucleótidos comprende

los nucleótidos 1-1791 de SEQ ID NO: 3 o los nucleótidos 1-1791 de SEQ ID NO: 4. En otra realización, la segunda secuencia de nucleótidos tiene al menos 60%, al menos 70%, al menos 80%, al menos 90%, al menos 95%, al menos 96%, al menos 97%, al menos 98%, o al menos 99% de identidad de secuencia con los nucleótidos 1792-4374 de la SEQ ID N°: 3 o 1792-4374 de la SEQ ID N°: 4. En una realización particular, la segunda secuencia de nucleótidos comprende los nucleótidos 1792-4374 de SEQ ID NO: 3 o 1792-4374 de SEQ ID NO: 4. En otra realización más, la segunda secuencia de nucleótidos tiene al menos 60%, al menos 70%, al menos 80%, al menos 90%, al menos 95%, al menos 96%, al menos 97%, al menos 98%, o al menos 99% de identidad de secuencia con los nucleótidos 1792-2277 y 2320-4374 de SEQ ID NO: 3 o 1792-277 y 2320-4374 de SEQ ID NO: 4 (es decir, los nucleótidos 1792-4374 de SEQ ID NO: 3 o 1792- 4374 de SEQ ID NO: 4 sin los nucleótidos que codifican el dominio B o el fragmento del dominio B). En una realización particular, la segunda secuencia de nucleótidos comprende los nucleótidos 1792-2277 y 2320-4374 de SEQ ID NO: 3 o 1792-2277 y 2320-4374 de SEQ ID NO: 4 (es decir, los nucleótidos 1792-4374 de SEQ ID NO: 3 o 1792-4374 de SEQ ID NO: 4 sin los nucleótidos que codifican el dominio B o el fragmento del dominio B).

**[00181]** En algunas realizaciones, la secuencia de nucleótidos optimizada por codón que codifica un polipéptido FVIII comprende una secuencia de nucleótidos que comprende una primera secuencia de ácido nucleico que codifica una porción N-terminal de un polipéptido FVIII y una segunda secuencia de ácido nucleico que codifica una porción C-terminal de un polipéptido FVIII; en donde la segunda secuencia de ácido nucleico tiene al menos 85%, al menos 90%, al menos 95%, al menos 96%, al menos 97%, al menos 98% o al menos 99% de identidad de secuencia con (i) los nucleótidos 1792-4374 de SEQ ID NO: 5 o (ii) 1792-4374 de SEQ ID NO: 6; y en donde la porción N-terminal y la porción C-terminal juntas tienen una actividad de polipéptido FVIII. En ciertas realizaciones, la segunda secuencia de ácido nucleico tiene al menos 85%, al menos 90%, al menos 95%, al menos 96%, al menos 97%, al menos 98% o al menos 99% de identidad de secuencia con los nucleótidos 1792-4374 de SEQ ID NO: 5. En otras realizaciones, la segunda secuencia de ácido nucleico tiene al menos 85%, al menos 90%, al menos 95%, al menos 96%, al menos 97%, al menos 98% o al menos 99% de identidad de secuencia con los nucleótidos 1792-4374 de SEQ ID NO: 6. En una realización particular, la segunda secuencia de ácido nucleico comprende los nucleótidos 1792-4374 de SEQ ID NO: 5 o 1792-4374 de SEQ ID NO: 6. En algunas realizaciones, la primera secuencia de ácido nucleico unida a la

segunda secuencia de ácido nucleico enumerada anteriormente tiene al menos 60%, al menos 70%, al menos 80%, al menos 90%, al menos 95%, al menos 96%, al menos 97%, al menos 98%, o al menos 99% de identidad de secuencia con los nucleótidos 58-1791 de la SEQ ID NO: 5 o los nucleótidos 58-1791 de la SEQ ID NO: 6. En otras realizaciones, la primera secuencia de ácido nucleico unida a la segunda secuencia de ácido nucleico enumerada anteriormente tiene al menos 60%, al menos 70%, al menos 80%, al menos 90%, al menos 95%, al menos 96%, al menos 97%, al menos 98%, o al menos 99% de identidad de secuencia con los nucleótidos 1-1791 de SEQ ID NO: 5 o los nucleótidos 1-1791 de SEQ ID NO: 6.

**[00182]** En otras realizaciones, la secuencia de nucleótidos optimizada por codón que codifica un polipéptido FVIII comprende una secuencia de nucleótidos que comprende una primera secuencia de ácido nucleico que codifica una porción N-terminal de un polipéptido FVIII y una segunda secuencia de ácido nucleico que codifica una porción C-terminal de un polipéptido FVIII; en donde la segunda secuencia de ácido nucleico tiene al menos 85%, al menos 90%, al menos 95%, al menos 96%, al menos 97%, al menos 98% o al menos 99% de identidad de secuencia con (i) los nucleótidos 1792-2277 y 2320-4374 de SEQ ID NO: 5 (es decir, los nucleótidos 1792-4374 de SEQ ID NO: 5 sin los nucleótidos que codifican el dominio B o el fragmento del dominio B) o (ii) 1792-2277 y 2320-4374 de SEQ ID NO: 6 (es decir, los nucleótidos 1792-4374 de SEQ ID NO: 6 sin los nucleótidos que codifican el dominio B o el fragmento del dominio B); y en donde la porción N-terminal y la porción C-terminal juntas tienen una actividad de polipéptido FVIII. En ciertas realizaciones, la segunda secuencia de ácido nucleico tiene al menos 85%, al menos 90%, al menos 95%, al menos 96%, al menos 97%, al menos 98% o al menos 99% de identidad de secuencia con los nucleótidos 1792-2277 y 2320-4374 de SEQ ID NO: 5 (es decir, los nucleótidos 1792-4374 de SEQ ID NO: 5 sin los nucleótidos que codifican el dominio B o el fragmento del dominio B). En otras realizaciones, la segunda secuencia de ácido nucleico tiene al menos 85%, al menos 90%, al menos 95%, al menos 96%, al menos 97%, al menos 98% o al menos 99% de identidad de secuencia con los nucleótidos 1792-2277 y 2320-4374 de SEQ ID NO: 6 (es decir, los nucleótidos 1792-4374 de SEQ ID NO: 6 sin los nucleótidos que codifican el dominio B o el fragmento del dominio B). En una realización particular, la segunda secuencia de ácido nucleico comprende los nucleótidos 1792-2277 y 2320-4374 de SEQ ID NO: 5 o 1792-2277 y 2320-4374 de SEQ ID NO: 6 (es decir, los nucleótidos 1792-4374 de SEQ ID NO: 5 o 1792-4374 de SEQ ID NO: 6 sin los nucleótidos que codifican el dominio B o el fragmento del

dominio B). En algunas realizaciones, la primera secuencia de ácido nucleico unida a la segunda secuencia de ácido nucleico enumerada anteriormente tiene al menos 60%, al menos 70%, al menos 80%, al menos 90%, al menos 95%, al menos 96%, al menos 97%, al menos 98% o al menos 99% de identidad de secuencia con los nucleótidos 58-1791 de la SEQ ID NO: 5 o los nucleótidos 58-1791 de la SEQ ID NO: 6. En otras realizaciones, la primera secuencia de ácido nucleico unida a la segunda secuencia de ácido nucleico enumerada anteriormente tiene al menos 60%, al menos 70%, al menos 80%, al menos 90%, al menos 95%, al menos 96%, al menos 97%, al menos 98% o al menos 99% de identidad de secuencia con los nucleótidos 1-1791 de SEQ ID NO: 5 o los nucleótidos 1-1791 de SEQ ID NO: 6.

**[00183]** En algunas realizaciones, la secuencia de nucleótidos optimizada por codón que codifica un polipéptido FVIII comprende una secuencia de nucleótidos que comprende una primera secuencia de ácido nucleico que codifica una porción N-terminal de un polipéptido FVIII y una segunda secuencia de ácido nucleico que codifica una porción C-terminal de un polipéptido FVIII; en donde la primera secuencia de ácido nucleico tiene al menos 90%, al menos 95%, al menos 96%, al menos 97%, al menos 98%, o al menos 99% de identidad de secuencia con (i) los nucleótidos 58-1791 de la SEQ ID NO: 1, (ii) nucleótidos 58-1791 de SEQ ID NO: 2, (iii) nucleótidos 58-1791 de SEQ ID NO: 70, o (iv) nucleótidos 58-1791 de SEQ ID NO: 71; y en donde la porción N-terminal y la porción C-terminal juntas tienen una actividad de polipéptido FVIII. En otras realizaciones, la primera secuencia de nucleótidos comprende los nucleótidos 58-1791 de SEQ ID NO: 1, los nucleótidos 58-1791 de SEQ ID NO: 2, (iii) los nucleótidos 58-1791 de SEQ ID NO: 70, o (iv) los nucleótidos 58-1791 de SEQ ID NO: 71.

**[00184]** En otras realizaciones, la secuencia de nucleótidos optimizada por codón que codifica un polipéptido FVIII comprende una secuencia de nucleótidos que comprende una primera secuencia de ácido nucleico que codifica una porción N-terminal de un polipéptido FVIII y una segunda secuencia de ácido nucleico que codifica una porción C-terminal de un polipéptido FVIII; en donde la primera secuencia de ácido nucleico tiene al menos 90%, al menos 95%, al menos 96%, al menos 97%, al menos 98% o al menos 99% de identidad de secuencia con (i) los nucleótidos 1-1791 de la SEQ ID NO: 1, (ii) nucleótidos 1-1791 de SEQ ID NO: 2, (iii) nucleótidos 1-1791 de SEQ ID NO: 70, o (iv) nucleótidos 1-1791 de SEQ ID NO: 71; y en donde la porción N-terminal y la porción C-terminal juntas tienen una actividad de polipéptido FVIII. En una realización, la primera secuencia de nucleótidos comprende los

nucleótidos 1-1791 de SEQ ID NO: 1, los nucleótidos 1-1791 de SEQ ID NO: 2, (iii) los nucleótidos 1-1791 de SEQ ID NO: 70, o (iv) los nucleótidos 1-1791 de SEQ ID NO: 71. En otra realización, la segunda secuencia de nucleótidos unida a la primera secuencia de nucleótidos tiene al menos 60%, al menos 70%, al menos 80%, al menos 90%, al menos 95%, al menos 96%, al menos 97%, al menos 98%, o al menos 99% de identidad de secuencia con los nucleótidos 1792-4374 de la SEQ ID NO: 1, 1792-4374 de la SEQ ID NO: 2, (iii) los nucleótidos 1792-4374 de la SEQ ID NO: 70, o (iv) los nucleótidos 1792-4374 de SEQ ID NO: 71. En una realización particular, la segunda secuencia de nucleótidos unida a la primera secuencia de nucleótidos comprende (i) los nucleótidos 1792-4374 de SEQ ID NO: 1, (ii) los nucleótidos 1792-4374 de SEQ ID NO: 2, (iii) los nucleótidos 1792-4374 de SEQ ID NO: 70, o (iv) los nucleótidos 1792-4374 de SEQ ID NO: 71. En otras realizaciones, la segunda secuencia de nucleótidos unida a la primera secuencia de nucleótidos tiene al menos 60%, al menos 70%, al menos 80%, al menos 90%, al menos 95%, al menos 96%, al menos 97%, al menos 98%, o al menos 99% de identidad de secuencia con (i) los nucleótidos 1792-2777 y 2320-4374 de SEQ ID NO: 1, (ii) los nucleótidos 1792-2277 y 2320-4374 de SEQ ID NO: 2, (iii) los nucleótidos 1792-2277 y 2320-4374 de SEQ ID NO: 70, o (iv) los nucleótidos 1792-277 y 2320-4374 de SEQ ID NO: 71. En una realización, la segunda secuencia de nucleótidos comprende (i) los nucleótidos 1792-2277 y 2320-4374 de SEQ ID NO: 1, (ii) los nucleótidos 1792-2277 y 2320-4374 de SEQ ID NO: 2, (iii) los nucleótidos 1792-2277 y 2320-4374 de SEQ ID NO: 70, o (iv) los nucleótidos 1792-2277 y 2320-4374 de SEQ ID NO: 71.

**[00185]** En otra realización, la secuencia de nucleótidos optimizada por codón que codifica un polipéptido FVIII comprende una secuencia de nucleótidos que comprende una primera secuencia de ácido nucleico que codifica una porción N-terminal de un polipéptido FVIII y una segunda secuencia de ácido nucleico que codifica una porción C-terminal de un polipéptido FVIII; en donde la segunda secuencia de ácido nucleico tiene al menos 90%, al menos 95%, al menos 96%, al menos 97%, al menos 98% o al menos 99% de identidad de secuencia con (i) los nucleótidos 1792-4374 de SEQ ID NO: 1, (ii) los nucleótidos 1792-4374 de SEQ ID NO: 2, (iii) los nucleótidos 1792-4374 de SEQ ID NO: 70, o (iv) los nucleótidos 1792-4374 de SEQ ID NO: 71; y en donde la porción N-terminal y la porción C-terminal juntas tienen una actividad de polipéptido FVIII. En una realización particular, la segunda secuencia de ácido nucleico comprende (i) los nucleótidos 1792-4374 de SEQ ID NO: 1, (ii) los nucleótidos 1792-4374 de SEQ ID NO: 2, (iii) los nucleótidos

1792-4374 de SEQ ID NO: 70, o (iv) los nucleótidos 1792-4374 de SEQ ID NO: 71. En algunas realizaciones, la secuencia optimizada de codón que codifica un polipéptido FVIII comprende una secuencia de nucleótidos que comprende una primera secuencia de ácido nucleico que codifica una porción N-terminal de un polipéptido FVIII y una  
5 segunda secuencia de ácido nucleico que codifica una porción C-terminal de un polipéptido FVIII; en donde la segunda secuencia de ácido nucleico tiene al menos 90%, al menos 95%, al menos 96%, al menos 97%, al menos 98% o al menos 99% de identidad de secuencia con (i) los nucleótidos 1792-277 y 2320- 4374 de SEQ ID NO: 1, (ii) los nucleótidos 1792-2277 y 2320-4374 de SEQ ID NO: 2, (iii) los nucleótidos  
10 1792-2277 y 2320-4374 de SEQ ID NO: 70, o (iv) los nucleótidos 1792-2277 y 2320-4374 de SEQ ID NO: 71 (es decir, los nucleótidos 1792-4374 de SEQ ID NO: 1, los nucleótidos 1792-4374 de SEQ ID NO: 2, los nucleótidos 1792-4374 de SEQ ID NO: 70, o los nucleótidos 1792-4374 de SEQ ID NO: 71 sin los nucleótidos que codifican el dominio B o el fragmento del dominio B); y en donde la porción N-terminal y la  
15 porción C-terminal juntas tienen una actividad de polipéptido FVIII. En una realización, la segunda secuencia de ácido nucleico comprende (i) los nucleótidos 1792-2277 y 2320-4374 de SEQ ID NO: 1, (ii) los nucleótidos 1792-2277 y 2320-4374 de SEQ ID NO: 2, (iii) los nucleótidos 1792-2277 y 2320-4374 de SEQ ID NO: 70, o (iv) los nucleótidos 1792-2277 y 2320-4374 de SEQ ID NO: 71 (es decir, los nucleótidos 1792-  
20 4374 de SEQ ID NO: 1, los nucleótidos 1792-4374 de SEQ ID NO: 2, los nucleótidos 1792-4374 de SEQ ID NO: 70, o los nucleótidos 1792-4374 de SEQ ID NO: 71 sin los nucleótidos que codifican el dominio B o el fragmento de dominio B).

**[00186]** En algunas realizaciones, la secuencia de nucleótidos con codón optimizado que codifica un polipéptido FVIII comprende una secuencia de nucleótidos  
25 que codifica un polipéptido con actividad de FVIII, en donde la secuencia de nucleótidos comprende una secuencia de ácido nucleico que tiene al menos 90%, al menos 91%, al menos 92%, al menos 93%, al menos 94%, al menos 95%, al menos 96%, al menos 97%, al menos 98% o al menos 99% de identidad de secuencia con los nucleótidos 58 a 4374 de la SEQ ID N°: 1. En otras realizaciones, la secuencia de  
30 nucleótidos comprende una secuencia de ácido nucleico que tiene al menos 90%, al menos 91%, al menos 92%, al menos 93%, al menos 94%, al menos 95%, al menos 96%, al menos 97%, al menos 98%, o al menos 99% de identidad de secuencia con los nucleótidos 58-2277 y 2320-4374 de SEQ ID NO: 1 (es decir, los nucleótidos 58-4374 de SEQ ID NO: 1 sin los nucleótidos que codifican el dominio B o fragmento de  
35 dominio B). En otras realizaciones, la secuencia de ácido nucleico tiene al menos 90%,

al menos 91%, al menos 92%, al menos 93%, al menos 94%, al menos 95%, al menos 96%, al menos 97%, al menos 98%, o al menos el 99% de identidad de secuencia con la SEQ ID NO: 1. En otras realizaciones, la secuencia de nucleótidos comprende los nucleótidos 58-2277 y 2320-4374 de SEQ ID NO: 1 (es decir, los nucleótidos 58-4374 de SEQ ID NO: 1 sin los nucleótidos que codifican el dominio B o el fragmento de dominio B) o los nucleótidos 58 a 4374 de SEQ ID NO: 1. En otras realizaciones más, la secuencia de nucleótidos comprende los nucleótidos 1-2277 y 2320-4374 de SEQ ID NO: 1 (es decir, los nucleótidos 1-4374 de SEQ ID NO: 1 sin los nucleótidos que codifican el dominio B o el fragmento de dominio B) o los nucleótidos 1 a 4374 de SEQ ID NO: 1.

**[00187]** En algunas realizaciones, la secuencia de nucleótidos con codón optimizado que codifica un polipéptido FVIII comprende una secuencia de nucleótidos que codifica un polipéptido con actividad de FVIII, en donde la secuencia de nucleótidos comprende una secuencia de ácido nucleico que tiene al menos 94%, al menos 95%, al menos 96%, al menos 97%, al menos 98% o al menos 99% de identidad de secuencia para los nucleótidos 58 a 4374 de la SEQ ID N°: 2. En otras realizaciones, la secuencia de nucleótidos comprende una secuencia de ácido nucleico que tiene al menos 94%, al menos 95%, al menos 96%, al menos 97%, al menos 98% o al menos 99% de identidad de secuencia con los nucleótidos 58-2277 y 2320-4374 de SEQ ID NO: 2. En otras realizaciones, la secuencia de ácido nucleico tiene al menos un 94%, al menos un 95%, al menos un 96%, al menos un 97%, al menos un 98% o al menos un 99% de identidad de secuencia con la SEQ ID NO: 2. En otras realizaciones, la secuencia de nucleótidos comprende los nucleótidos 58-2277 y 2320-4374 de SEQ ID NO: 2 (es decir, los nucleótidos 58-4374 de SEQ ID NO: 2 sin los nucleótidos que codifican el dominio B o el fragmento de dominio B) o los nucleótidos 58 a 4374 de SEQ ID NO: 2. En otras realizaciones más, la secuencia de nucleótidos comprende los nucleótidos 1-2277 y 2320-4374 de SEQ ID NO: 2 (es decir, los nucleótidos 1-4374 de SEQ ID NO: 2 sin los nucleótidos que codifican el dominio B o el fragmento de dominio B) o los nucleótidos 1 a 4374 de SEQ ID NO: 2.

**[00188]** En algunas realizaciones, la secuencia de nucleótidos con codón optimizado que codifica un polipéptido FVIII comprende una secuencia de nucleótidos que codifica un polipéptido con actividad de FVIII, en donde la secuencia de nucleótidos comprende una secuencia de ácido nucleico que tiene al menos 85%, al menos 86%, al menos 87%, al menos 88%, al menos 89%, al menos 90%, al menos 91%, al menos 95%, al menos 96%, al menos 97%, al menos 98%, o al menos 99%

de identidad de secuencia con los nucleótidos 58 a 4374 de la SEQ ID NO: 70. En otras realizaciones, la secuencia de nucleótidos comprende una secuencia de ácido nucleico que tiene al menos 85%, al menos 86%, al menos 87%, al menos 88%, al menos 89%, al menos 90%, al menos 91%, al menos 95%, al menos 96%, al menos 97%, al menos 98% o al menos 99% de identidad de secuencia a los nucleótidos 58-2277 y 2320-4374 de SEQ ID NO: 70 (es decir, los nucleótidos 58-4374 de SEQ ID NO: 70 sin los nucleótidos que codifican el dominio B o el fragmento del dominio B). En otras realizaciones, la secuencia de ácido nucleico tiene al menos 85%, al menos 86%, al menos 87%, al menos 88%, al menos 89%, al menos 90%, al menos 91%, al menos 95%, al menos 96%, al menos 97%, al menos 98%, o al menos 99% de identidad de secuencia con la SEQ ID NO: 70. En otras realizaciones, la secuencia de nucleótidos comprende los nucleótidos 58-2277 y 2320-4374 de SEQ ID NO: 70 (es decir, los nucleótidos 58-4374 de SEQ ID NO: 70 sin los nucleótidos que codifican el dominio B o el fragmento de dominio B) o los nucleótidos 58 a 4374 de SEQ ID NO: 70. En otras realizaciones más, la secuencia de nucleótidos comprende los nucleótidos 1-2277 y 2320-4374 de SEQ ID NO: 70 (es decir, los nucleótidos 1-4374 de SEQ ID NO: 70 sin los nucleótidos que codifican el dominio B o el fragmento de dominio B) o los nucleótidos 1 a 4374 de la SEQ ID NO: 70.

**[00189]** En algunas realizaciones, la secuencia de nucleótidos con codón optimizado que codifica un polipéptido FVIII comprende una secuencia de nucleótidos que codifica un polipéptido con actividad de FVIII, en donde la secuencia de nucleótidos comprende una secuencia de ácido nucleico que tiene al menos 85%, al menos 86%, al menos 87%, al menos 88%, al menos 89%, al menos 90%, al menos 91%, al menos 95%, al menos 96%, al menos 97%, al menos 98% o al menos 99% de identidad de secuencia con los nucleótidos 58 a 4374 de la SEQ ID NO: 71. En otras realizaciones, la secuencia de nucleótidos comprende una secuencia de ácido nucleico que tiene al menos 85%, al menos 86%, al menos 87%, al menos 88%, al menos 89%, al menos 90%, al menos 91%, al menos 95%, al menos 96%, al menos 97%, al menos 98%, o al menos 99% de identidad de secuencia con los nucleótidos 58-2277 y 2320-4374 de SEQ ID NO: 71 (es decir, los nucleótidos 58-4374 de SEQ ID NO: 71 sin los nucleótidos que codifican el dominio B o el fragmento del dominio B). En otras realizaciones, la secuencia de ácido nucleico tiene al menos 85%, al menos 86%, al menos 87%, al menos 88%, al menos 89%, al menos 90%, al menos 91%, al menos 95%, al menos 96%, al menos 97%, al menos 98%, o al menos 99% de identidad de secuencia con la SEQ ID NO: 71. En otras realizaciones, la secuencia

de nucleótidos comprende los nucleótidos 58-2277 y 2320-4374 de SEQ ID NO: 71 (es decir, los nucleótidos 58-4374 de SEQ ID NO: 71 sin los nucleótidos que codifican el dominio B o el fragmento del dominio B) o los nucleótidos 58 a 4374 de la SEQ ID NO: 71. En otras realizaciones más, la secuencia de nucleótidos comprende los nucleótidos 1-2277 y 2320-4374 de SEQ ID NO: 71 (es decir, los nucleótidos 1-4374 de SEQ ID NO: 71 sin los nucleótidos que codifican el dominio B o el fragmento de dominio B) o los nucleótidos 1 a 4374 de la SEQ ID NO: 71.

**[00190]** En algunas realizaciones, la secuencia de nucleótidos con codón optimizado que codifica un polipéptido FVIII comprende una secuencia de nucleótidos que codifica un polipéptido con actividad FVIII, en donde la secuencia de nucleótidos comprende una secuencia de ácido nucleico que tiene al menos 92%, al menos 93%, al menos 94%, al menos 95%, al menos 96%, al menos 97%, al menos 98% o al menos 99% de identidad de secuencia con los nucleótidos 58 a 4374 de la SEQ ID NO: 3. En otras realizaciones, la secuencia de nucleótidos comprende una secuencia de ácido nucleico que tiene al menos 92%, al menos 93%, al menos 94%, al menos 95%, al menos 96%, al menos 97%, al menos 98% o al menos 99% de identidad de secuencia con los nucleótidos 58-2277 y 2320-4374 de SEQ ID NO: 3 (es decir, los nucleótidos 58-4374 de SEQ ID NO: 3 sin los nucleótidos que codifican el dominio B o el fragmento de dominio B). En ciertas realizaciones, la secuencia de ácido nucleico tiene al menos 92%, al menos 93%, al menos 94%, al menos 95%, al menos 96%, al menos 97%, al menos 98% o al menos 99% de identidad de secuencia con la SEQ ID NO: 3. En algunas realizaciones, la secuencia de nucleótidos comprende los nucleótidos 58-2277 y 2320-4374 de SEQ ID NO: 3 (es decir, los nucleótidos 58-4374 de SEQ ID NO: 3 sin los nucleótidos que codifican el dominio B o el fragmento de dominio B) o los nucleótidos 58 a 4374 de SEQ ID NO: 3. En otras realizaciones más, la secuencia de nucleótidos comprende los nucleótidos 58-2277 y 2320-4374 de SEQ ID NO: 3 (es decir, los nucleótidos 1-4374 de SEQ ID NO: 3 sin los nucleótidos que codifican el dominio B o el fragmento de dominio B) o los nucleótidos 1 a 4374 de SEQ ID NO: 3.

**[00191]** En algunas realizaciones, la secuencia de nucleótidos con codón optimizado que codifica un polipéptido FVIII comprende una secuencia de nucleótidos que codifica un polipéptido con actividad de FVIII, en donde la secuencia de nucleótidos comprende una secuencia de ácido nucleico que tiene al menos 91%, al menos 92%, al menos 93%, al menos 94%, al menos 95%, al menos 96%, al menos 97%, al menos 98% o al menos 99% de identidad de secuencia con los nucleótidos

58 a 4374 de la SEQ ID N°: 4. En otras realizaciones, la secuencia de nucleótidos comprende una secuencia de ácido nucleico que tiene al menos 91%, al menos 92%, al menos 93%, al menos 94%, al menos 95%, al menos 96%, al menos 97%, al menos 98%, o al menos 99% de identidad de secuencia con los nucleótidos 58-2277 y 2320-4374 de SEQ ID NO: 4 (es decir, los nucleótidos 58-4374 de SEQ ID NO: 4 sin los nucleótidos que codifican el dominio B o el fragmento del dominio B). En otras realizaciones, la secuencia de ácido nucleico tiene al menos 91%, al menos 92%, al menos 93%, al menos 94%, al menos 95%, al menos 96%, al menos 97%, al menos 98%, o al menos 99% de identidad de secuencia con la SEQ ID NO: 4. En otras realizaciones, la secuencia de nucleótidos comprende los nucleótidos 58-2277 y 2320-4374 de SEQ ID NO: 4 (es decir, los nucleótidos 58-4374 de SEQ ID NO: 4 sin los nucleótidos que codifican el dominio B o el fragmento del dominio B) o los nucleótidos 58 a 4374 de la SEQ ID NO: 4. En otras realizaciones más, la secuencia de nucleótidos comprende los nucleótidos 1-2277 y 2320-4374 de SEQ ID NO: 4 (es decir, los nucleótidos 1-4374 de SEQ ID NO: 4 sin los nucleótidos que codifican el dominio B o el fragmento del dominio B) o los nucleótidos 1 a 4374 de SEQ ID NO: 4.

**[00192]** En algunas realizaciones, la secuencia de nucleótidos con codón optimizado que codifica un polipéptido FVIII comprende una secuencia de nucleótidos que codifica un polipéptido con actividad de FVIII, en donde la secuencia de nucleótidos comprende una secuencia de ácido nucleico que tiene al menos 89%, al menos 90%, al menos 91%, al menos 92%, al menos 93%, al menos 94%, al menos 95%, al menos 96%, al menos 97%, al menos 98% o al menos 99% de identidad de secuencia con los nucleótidos 58 a 4374 de SEQ ID NO: 5. En otras realizaciones, la secuencia de nucleótidos comprende una secuencia de ácido nucleico que tiene al menos 89%, al menos 90%, al menos 91%, al menos 92%, al menos 93%, al menos 94%, al menos 95%, al menos 96%, al menos 97%, al menos 98% o al menos 99% de identidad de secuencia con los nucleótidos 58-2277 y 2320-4374 de SEQ ID NO: 5 (es decir, los nucleótidos 58-4374 de SEQ ID NO: 5 sin los nucleótidos que codifican el dominio B o el fragmento del dominio B). En ciertas realizaciones, la secuencia de ácido nucleico tiene al menos 89%, al menos 90%, al menos 91%, al menos 92%, al menos 93%, al menos 94%, al menos 95%, al menos 96%, al menos 97%, al menos 98%, o al menos 99% de identidad de secuencia con la SEQ ID NO: 5. En algunas realizaciones, la secuencia de nucleótidos comprende los nucleótidos 58-2277 y 2320-4374 de la SEQ ID NO: 5 (es decir, los nucleótidos 58-4374 de la SEQ ID NO: 5 sin los nucleótidos que codifican el dominio B o el fragmento del dominio B) o los

nucleótidos 58 a 4374 de la SEQ ID NO: 5. En otras realizaciones más, la secuencia de nucleótidos comprende los nucleótidos 1-2277 y 2320-4374 de SEQ ID NO: 5 (es decir, los nucleótidos 1-4374 de SEQ ID NO: 5 sin los nucleótidos que codifican el dominio B o el fragmento de dominio B) o los nucleótidos 1 a 4374 de SEQ ID NO: 5.

5 **[00193]** En algunas realizaciones, la secuencia de nucleótidos con codón optimizado que codifica un polipéptido FVIII comprende una secuencia de nucleótidos que codifica un polipéptido con actividad de FVIII, en donde la secuencia de nucleótidos comprende una secuencia de ácido nucleico que tiene al menos 89%, al menos 90%, al menos 91%, al menos 92%, al menos 93%, al menos 94%, al menos 95%, al menos 96%, al menos 97%, al menos 98% o al menos 99% de identidad de  
10 secuencia con los nucleótidos 58 a 4374 de SEQ ID NO: 6. En otras realizaciones, la secuencia de nucleótidos comprende una secuencia de ácido nucleico que tiene al menos 89%, al menos 90%, al menos 91%, al menos 92%, al menos 93%, al menos 94%, al menos 95%, al menos 96%, al menos 97%, al menos 98% o al menos 99%  
15 de identidad de secuencia con los nucleótidos 58-2277 y 2320-4374 de SEQ ID NO: 6 (es decir, los nucleótidos 58-4374 de SEQ ID NO: 6 sin los nucleótidos que codifican el dominio B o el fragmento del dominio B). En ciertas realizaciones, la secuencia de ácido nucleico tiene al menos 89%, al menos 90%, al menos 91%, al menos 92%, al menos 93%, al menos 94%, al menos 95%, al menos 96%, al menos 97%, al menos  
20 98% o al menos 99% de identidad de secuencia con la SEQ ID NO: 6. En algunas realizaciones, la secuencia de nucleótidos comprende los nucleótidos 58-2277 y 2320-4374 de la SEQ ID NO: 6 (es decir, los nucleótidos 58-4374 de la SEQ ID NO: 6 sin los nucleótidos que codifican el dominio B o el fragmento del dominio B) o los nucleótidos 58 a 4374 de SEQ ID NO: 6. En otras realizaciones más, la secuencia de  
25 nucleótidos comprende los nucleótidos 1-2277 y 2320-4374 de SEQ ID NO: 6 (es decir, los nucleótidos 1-4374 de SEQ ID NO: 6 sin los nucleótidos que codifican el dominio B o el fragmento de dominio B) o los nucleótidos 1 a 4374 de la SEQ ID NO: 6.

**[00194]** En algunas realizaciones, la secuencia de nucleótidos optimizada por  
30 codón que codifica un polipéptido FVIII comprende una secuencia de ácido nucleico que codifica un péptido señal. En ciertas realizaciones, el péptido señal es un péptido señal FVIII. En algunas realizaciones, la secuencia de ácido nucleico que codifica un péptido señal está optimizada por codón. En una realización particular, la secuencia de ácido nucleico que codifica un péptido señal tiene al menos 60%, al menos 70%,  
35 al menos 80%, al menos 90%, al menos 95%, al menos 96%, al menos 97%, al menos

98%, al menos 99% o 100% de identidad de secuencia para (i) los nucleótidos 1 a 57 de la SEQ ID NO: 1; (ii) los nucleótidos 1 a 57 de la SEQ ID NO: 2; (iii) los nucleótidos 1 a 57 de SEQ ID NO: 3; (iv) los nucleótidos 1 a 57 de SEQ ID NO: 4; (v) los nucleótidos 1 a 57 de SEQ ID NO: 5; (vi) los nucleótidos 1 a 57 de la SEQ ID NO: 6; (vii) los nucleótidos 1 a 57 de SEQ ID NO: 70; (viii) los nucleótidos 1 a 57 de SEQ ID NO: 71; o (ix) los nucleótidos 1 a 57 de la SEQ ID NO: 68.

**[00195]** SEQ ID NO: 1-6, 70 y 71 son versiones optimizadas de SEQ ID NO: 16, la secuencia de nucleótidos de FVIII inicial o "progenitora" o "natural". La SEQ ID NO: 16 codifica un FVIII humano con dominio B eliminado. Mientras que las SEQ ID N°: 1-6, 70 y 71 se derivan de una forma específica de FVIII con dominio B eliminado (SEQ ID N°: 16), debe entenderse que la presente descripción también incluye versiones optimizadas de ácidos nucleicos que codifican otras versiones de FVIII. Por ejemplo, otras versiones de FVIII pueden incluir FVIII de longitud completa, otras deleciones de FVIII en el dominio B (descritas en este documento) u otros fragmentos de FVIII que retienen la actividad de FVIII.

**[00196]** En una realización, el casete génico comprende una construcción FVIII, que incluye una secuencia de polinucleótidos (6526 nucleótidos) como se describe en las Tablas 2A-2C. En una realización, el casete génico comprende una construcción FVIII, que incluye una secuencia de polinucleótidos (6526 nucleótidos) como se describe en la Tabla 2B.

**[00197]** En ciertas realizaciones, la molécula de ácido nucleico aislada comprende una secuencia de nucleótidos que tiene al menos aproximadamente 60%, al menos aproximadamente 65%, al menos aproximadamente 70%, al menos aproximadamente 75%, al menos aproximadamente 80%, al menos aproximadamente 85%, al menos aproximadamente 90%, al menos aproximadamente 95%, al menos aproximadamente 96%, al menos aproximadamente 97%, al menos aproximadamente 98%, al menos aproximadamente 99% o aproximadamente 100% de identidad de secuencia con la secuencia de nucleótidos de SEQ ID NO: 179 o 182, en donde la molécula de ácido nucleico aislada retiene la capacidad de expresar una proteína FVIII funcional.

Tabla 2A: Ejemplo de construcción FVIII (nucleótidos 1-6526; SEQ ID NO: 110)

| Descripción | Secuencia |
|-------------|-----------|
|-------------|-----------|

|                                                                                                                                                              |                                                                                                                                                                                                                                                                                                                                                                                                                                                            |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 5'ITR (5'-extremo repetición terminal invertida AAV2) (SEQ ID NO:111)                                                                                        | 1--<br>CTGCGCGCTCGCTCGCTCACTGAGGCCGCCCGGGCAAAGCCCCGGGCGT<br>CGGGCGACCTTT<br>GGTCGCCCCGGCCTCAGTGAGCGAGCGAGCGCGCAGAGAGGGAGTGGC<br>CAACTCCATCACTAGGGGTTCT -- 130                                                                                                                                                                                                                                                                                              |
| Secuencia estructural plasmídica (PBS)- 1 (SEQ ID NO:112)                                                                                                    | 131--<br>GCGGCAATTCAGTCGATAACTATAACGGTCCTAAGGTAGCGATTTAAATA<br>CGCGCTCT<br>CTTAAGGTAGCCCCGGGACGCGTCAATTGAGATCTGGATCCGGTACCGA<br>ATTCGCGGCCGCTCGACGACTAGCGTTTAATTAA -- 272                                                                                                                                                                                                                                                                                  |
| TTPp (promotor específico del hígado) (SEQ ID NO:113)                                                                                                        | 273 --<br>ACGCGTGTCTGTCTGCACATTTCTAGAGCGAGTGTTCCGATACTCTAAT<br>CTCCCTAG<br>GCAAGGTTTCATATTTGTGTAGGTTACTTATTCTCCTTTTGTGACTAAGTC<br>AATAATCAGAATCAGCAGGTTTGGAGTCAGCTTGGCAGGGATCAGCAGC<br>CTGGGTTGGAAGGAGGGGGTATAAAAGCCCCTTCACCAGGAGAAGCCGT<br>CACACAGATCCACAAGCTCCTG -- 501                                                                                                                                                                                  |
| PBS-2 (SEQ ID NO:114)                                                                                                                                        | 502 -- AG -- 503                                                                                                                                                                                                                                                                                                                                                                                                                                           |
| Intrón sintético (SEQ ID NO:115)                                                                                                                             | 504 --<br>GTAAGTGCCGTGTGTGGTTCCCGCGGGCCTGGCCTCTTTACGGGTTATG<br>GCCCTTGCG<br>TGCCTTGAATTACTGACACTGACATCCACTTTTTCTTTTTCTCCACAG --<br>609                                                                                                                                                                                                                                                                                                                     |
| PBS-3 (SEQ ID NO:116)                                                                                                                                        | 610 -- CTAGCGCCACC -- 620                                                                                                                                                                                                                                                                                                                                                                                                                                  |
| FVIIIco6XTEN (SEQ ID NO:117) (marco de lectura abierto para la versión 6 del FVIII optimizado con codón que contiene XTEN144; la secuencia XTEN se marca con | 621 --<br>ATGCAGATTGAGCTGTCCACTTGTTTCTTCCTGTGCCTCCTGCGCTTCTG<br>TTTCTCCG<br>CCACTCGCCGGTACTACCTTGGAGCCGTGGAGCTTTCATGGGACTACAT<br>GCAGAGCGACCTGGGCGAACTCCCCGTGGATGCCAGATCCCCCCCCCG<br>CGTGCCAAAGTCCTTCCCCTTTAACACCTCCGTGGTGTACAAGAAAACCC<br>TCTTTGTCGAGTTCACTGACCACCTGTTCAACATCGCCAAGCCGCGCCCA<br>CCTTGGATGGGCCTCCTGGGACCGACCATTCAAGCTGAAGTGTACGACA<br>CCGTGGTGATCACCCTGAAGAACATGGCGTCCCACCCCGTGTCCCTGCA<br>TGCGGTCGGAGTGTCTACTGGAAGGCCTCCGAAGGAGCTGAGTACGA |

|                                                    |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |
|----------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <p><u>doble subrayado</u><br/>(SEQ ID NO:118))</p> | <p>CGACCAGACTAGCCAGCGGGAAAAGGAGGACGATAAAGTGTTCCCGGG<br/>CGGCTCGCATACTTACGTGTGGCAAGTCCTGAAGGAAAACGGACCTATG<br/>GCATCCGATCCTCTGTGCCTGACTTACTCCTACCTTTCCCATGTGGACCT<br/>CGTGAAGGACCTGAACAGCGGGCTGATTGGTGCATTCTCGTGTGCCGC<br/>GAAGGTTGCTCGCTAAGGAAAAGACCCAGACCCTCCATAAGTTCATCCT<br/>TTTGTTGCTGTGTTGATGAAGGAAAGTCATGGCATTCCGAAACTAAGA<br/>ACTCGCTGATGCAGGACCGGGATGCCGCCTCAGCCCGCGCCTGGCCTA<br/>AAATGCATACAGTCAACGGATACGTGAATCGGTCACTGCCCAGGGCTCAT<br/>CGGTTGTCACAGAAAGTCCGTGTACTGGCACGTCATCGGCATGGGCACT<br/>ACGCCTGAAGTGCCTCATCTTCTGGAAGGGCACACCTTCTCTGTCG<br/>GCAACCACCGCCAGGCCTCTCTGGAATCTCCCCGATTACCTTTCTGACC<br/>GCCCAGACTCTGCTCATGGACCTGGGGCAGTTCTTCTCTTCTGCCACAT<br/>CTCCAGCCATCAGCACGACGGAATGGAGGCCTACGTGAAGGTGGACTCA<br/>TGCCCGGAAGAACCTCAGTTGCGGATGAAGAACAACGAGGAGGCCGAG<br/>GACTATGACGACGATTTGACTGACTCCGAGATGGACGTCGTGCGGTTG<br/>ATGACGACAACAGCCCCAGCTTCATCCAGATTCGCAGCGTGGCCAAGAA<br/>GCACCCCAAACCTGGGTGCACTACATCGCGGCCGAGGAAGAAGATTGG<br/>GACTACGCCCCGTTGGTGCTGGCACCCGATGACCGGTCGTACAAGTCCC<br/>AGTATCTGAACAATGGTCCGCAGCGGATTGGCAGAAAGTACAAGAAAGT<br/>GCGGTTTCATGGCGTACACTGACGAAACGTTTAAGACCCGGGAGGCCATT<br/>CAACATGAGAGCGGCATTCTGGGACCACTGCTGTACGGAGAGGTGCGC<br/>GATACCCTGCTCATCATCTTCAAAAACCAGGCCTCCCGGCCTTACAACAT<br/>CTACCCTCACGGAATCACCGACGTGCGGCCACTCTACTCGCGGCGCCTG<br/>CCGAAGGGCGTCAAGCACCTGAAAGACTTCCCTATCCTGCCGGGCGAAA<br/>TCTTCAAGTATAAGTGGACCGTCACCGTGGAGGACGGGCCCACCAAGAG<br/>CGATCCTAGGTGTCTGACTCGGTACTACTCCAGCTTCGTGAACATGGAAC<br/>GGGACCTGGCATCGGGACTCATTGGACCGCTGCTGATCTGCTACAAAGA<br/>GTCGGTGGATCAACGCGGCAACCAGATCATGTCCGACAAGCGCAACGTG<br/>ATCCTGTTCTCCGTGTTTGATGAAAACAGATCCTGGTACCTCACTGAAAA<br/>CATCCAGAGGTTCTCCCAAACCCCGCAGGAGTGCAACTGGAGGACCCT<br/>GAGTTTCAGGCCTCGAATATCATGCACTCGATTAACGGTTACGTGTTTCA<br/>CTCGCTGCAACTGAGCGTGTGCCTCCATGAAGTCGCTTACTGGTACATTC<br/>TGTCCATCGGCGCCCAGACTGACTTCCTGAGCGTGTCTTTTCCGGTTAC<br/>ACCTTTAAGCACAAGATGGTGTACGAAGATACCCTGACCCTGTTCCCTTT<br/>CTCCGGCGAAACGGTGTTCATGTGATGGAGAACC CGGGTCTGTGGATT<br/>CTGGGATGCCACAACAGCGACTTTCGGAACCGCGGAATGACTGCCCTGC<br/>TGAAGGTGTCCTCATGCGACAAGAACACCGGAGACTACTACGAGGACTC<br/>CTACGAGGATATCTCAGCCTACCTCCTGTCCAAGAACAACGCGATCGAG<br/>CCGCGCAGCTTCAGCCAGAACGGCGCGCCAACATCAGAGAGCGCCACC<br/><u>CCTGAAAGTGGTCCCGGGAGCGAGCCAGCCACATCTGGGTGCGAAACG</u><br/><u>CCAGGCACAAGTGAGTCTGCAACTCCCGAGTCCGGACCTGGCTCCGAG</u></p> |
|----------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

CCTGCCACTAGCGGCTCCGAGACTCCGGGAACTTCCGAGAGCGCTACAC  
CAGAAAGCGGACCCGGAACCAGTACCGAACCTAGCGAGGGCTCTGCTC  
CGGGCAGCCCAGCCGGCTCTCCTACATCCACGGAGGAGGGCACTTCCG  
AATCCGCCACCCCGGAGTCAGGGCCAGGATCTGAACCCGCTACCTCAG  
GCAGTGAGACGCCAGGAACGAGCGAGTCCGCTACACCGGAGAGTGGGC  
CAGGGAGCCCTGCTGGATCTCCTACGTCCACTGAGGAAGGGTCACCAGC  
GGGCTCGCCCACCAGCACTGAAGAAGGTGCCTCGAGCCCGCCTGTGCT  
GAAGAGGCACCAGCGAGAAATTACCCGGACCACCCTCCAATCGGATCAG  
GAGGAAATCGACTACGACGACACCATCTCGGTGGAAATGAAGAAGGAAG  
ATTTGATATCTACGACGAGGACGAAAATCAGTCCCCTCGCTCATTCAA  
AAGAAAAGTAGACACTACTTTATCGCCGCGGTGGAAAGACTGTGGGACTA  
TGGAATGTCATCCAGCCCTCACGTCTTCGGAACCGGGCCAGAGCGGA  
TCGGTGCCTCAGTTCAAGAAAGTGGTGTTCAGGAGTTCACCGACGGCA  
GCTTCACCCAGCCGCTGTACCGGGGAGAACTGAACGAACACCTGGGCCT  
GCTCGGTCCCTACATCCGCGCGGAAGTGGAGGATAACATCATGGTGACC  
TTCCGTAACCAAGCATCCAGACCTTACTCCTTCTATTCTCCCTGATCTCA  
TACGAGGAGGACCAGCGCCAAGGCGCCGAGCCCCGCAAGAACTTCGTC  
AAGCCCAACGAGACTAAGACCTACTTCTGGAAGGTCCAACACCATATGG  
CCCCGACCAAGGATGAGTTTGACTGCAAGGCCTGGGCCTACTTCTCCGA  
CGTGGACCTTGAGAAGGATGTCCATTCCGGCCTGATCGGGCCGCTGCTC  
GTGTGTCACACCAACACCCTGAACCCAGCGCATGGACGCCAGGTCACCG  
TCCAGGAGTTTGCTCTGTTCTTCACCATTTTTGACGAACTAAGTCCTGGT  
ACTTCACCGAGAATATGGAGCGAACTGTAGAGCGCCCTGCAATATCCA  
GATGGAAGATCCGACTTTCAAGGAGAACTATAGATTCCACGCCATCAACG  
GGTACATCATGGATACTCTGCCGGGGCTGGTCATGGCCCAGGATCAGAG  
GATTCGGTGGTACTTGCTGTCAATGGGATCGAACGAAAACATTCACTCCA  
TTCATTCTCCGGTCACGTGTTCACTGTGCGCAAGAAGGAGGAGTACAA  
GATGGCGCTGTACAATCTGTACCCCGGGGTGTTGAACTGTGGAGATG  
CTGCCGTCCAAGGCCGGCATCTGGAGAGTGGAGTGCCTGATCGGAGAG  
CACCTCCACGCGGGGATGTCCACCCTCTTCCTGGTGTACTCGAATAAGT  
GCCAGACCCCGCTGGGCATGGCCTCGGGCCACATCAGAGACTTCCAGA  
TCACAGCAAGCGGACAATACGGCCAATGGGCGCCGAAGCTGGCCCGCT  
TGCACTACTCCGGATCGATCAACGCATGGTCCACCAAGGAACCGTTCTC  
GTGGATTAAGGTGGACCTCCTGGCCCCTATGATTATCCACGGAATTAAGA  
CCCAGGGCGCCAGGCAGAAGTTCTCCTCCCTGTACATCTCGCAATTCAT  
CATCATGTACAGCCTGGACGGGAAGAAGTGGCAGACTTACAGGGGAAAC  
TCCACCGGCACCCTGATGGTCTTTTTTCGGCAACGTGGATTCTCCGGCA  
TTAAGCACAACATCTTCAACCCACCGATCATAGCCAGATATATTAGGCTC  
CACCCCACTCACTACTCAATCCGCTCAACTCTTCGGATGGAACCATGGG  
GTGCGACCTGAACTCCTGCTCCATGCCGTTGGGGATGGAATCAAAGGCT  
ATTAGCGACGCCAGATCACCGCGAGCTCCTACTTCACTAACATGTTTCG

|                                                                                                            |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |
|------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|                                                                                                            | CACCTGGAGCCCCTCCAAGGCCAGGCTGCACTTGCAGGGACGGTCAAAT<br>GCCTGGCGGCCGCAAGTGAACAATCCGAAGGAATGGCTTCAAGTGGATT<br>TCCAAAAGACCATGAAAGTGACCGGAGTCACCACCCAGGGAGTGAAGTC<br>CCTTCTGACCTCGATGTATGTGAAGGAGTTCCTGATTAGCAGCAGCCAG<br>GACGGGCACCAAGTGGACCCTGTTCTTCCAAAACGGAAAGGTCAAGGTGT<br>TCCAGGGGAACCAGGACTCGTTCACACCCGTGGTGAAGTCCCTGGACCC<br>CCCACTGCTGACGCGGTACTTGAGGATTCATCCTCAGTCCTGGGTCCAT<br>CAGATTGCATTGCGAATGGAAGTCCTGGGCTGCGAGGCCCAGGACCTGT<br>ACTGA -- 5444                                                                                                                                                                                                                                 |
| PBS-4 (SEQ ID NO:119)                                                                                      | 5445 -- ATCAGCCTGAGCTCGCTGA -- 5463                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |
| WPRe (elemento regulador post-transcripcional del virus de la hepatitis de marmota mutado) (SEQ ID NO:120) | 5464 --<br>TCATAATCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGGTATTCT<br>TAACTA<br><br>TGTGCTCCTTTTACGCTATGTGGATACGCTGCTTTAATGCCTTTGTATCA<br>TGCTATTGCTTCCCGTATGGCTTTCAATTTCTCCTCCTTGATAAATCCTG<br>GTTGCTGTCTCTTTATGAGGAGTTGTGGCCCGTTGTCAGGCAACGTGGC<br>GTGGTGTGCACTGTGTTTGCTGACGCAACCCCCACTGGTTGGGGCATTG<br>CCACCACCTGTCAGTCTCTTTCCGGGACTTTTCGCTTTCCCCCTCCCTATT<br>GCCACGGCGGAACATCATCGCCGCTGCCTTGCCCGCTGCTGGACAGGG<br>GCTCGGCTGTTGGGCACTGACAATTCCGTGGTGTGTCGGGGAAATCAT<br>CGTCCTTTCTTGGCTGCTCGCCTGTGTTGCCACCTGGATTCTGCGCGG<br>GACGTCCTTCTGCTACGTCCCTTCGGCCCTCAATCCAGCGGACCTTCCTT<br>CCCGCGGCCTGCTGCCGGCTCTGCGGCCTCTTCCGCGTCTTCGCCTTC<br>GCCCTCAGACGAGTCGGATCTCCCTTTGGGCCGCTCCCCGCTG -- 6058 |
| PBS-5 (SEQ ID NO:121)                                                                                      | 6059 -- ATCAGCCT -- 6066                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |
| bGHpA (señal de poliadenilación de la hormona del crecimiento de bovina) (SEQ ID NO:122)                   | 6067 --<br>CGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCTCCCCCGTG<br>CCTTCCT<br><br>TGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAATGAGGAA<br>ATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGT<br>GGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGC<br>TGGGGA --6277                                                                                                                                                                                                                                                                                                                                                                                                                           |
| PBS-6 (SEQ ID NO:123)                                                                                      | 6278 --<br>TGCGGTGGGCTCTATGGCTTCTGAGGCGGAAAGAACGGGCTCGAGAAG<br>CTTCTAGAT                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |

|                                                                                |                                                                                                                                                                    |
|--------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|                                                                                | ATCCTCTCTTAAGGTAGCATCGAGATTTAAATTAGGGATAACAGGGTAAT<br>GGCGCGGGCCGC -- 6396                                                                                         |
| 3'ITR (repetición<br>terminal invertido<br>extremo 3' AAV2)<br>(SEQ ID NO:124) | 6397 --<br>AGGAACCCCTAGTGATGGAGTTGGCCACTCCCTCTCTGCGCGCTCGCTC<br>GCTCACTG<br>AGGCCGGGCGACCAAAGGTCGCCCCGACGCCCGGGCTTTGCCCGGGCG<br>GCCTCAGTGAGCGAGCGAGCGCGCAG -- 6526 |

Tabla 2B: Construcción del ejemplo B19-FVIII (nucleótidos 1-6762; SEQ ID NO: 179)

| Descripción                                                                                                                                                                            | Secuencia                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 5'ITR (SEQ ID<br>NO: 180)                                                                                                                                                              | 1 –<br>CTCTGGGCCAGCTTGCTTGGGGTTGCCTTGACACTAAGACAAGCGGCGC<br>GCCGCTTGATCTTAGTGGCACGTCAACCCCAAGCGCTGGCCCAGAGCCA<br>ACCCTAATTCCGGAAGTCCCGCCACCGGAAGTGACGTACAGGAAATG<br>ACGTACAGGAAATGACGTAATTGTCCGCCATCTTGTACCGGAAGT<br>CCCGCCTACCGGCGGCGACCGGCGGCATCTGATTTGGTGTCTTCTTTTA<br>A ATTTT -- 248                                                                                                                                                                                                                                                                                                                     |
| TTPp (promotor<br>específico del<br>hígado) (SEQ ID<br>NO:113)                                                                                                                         | 391 –<br>ACGCGTGTCTGTCTGCACATTTCTAGAGCGAGTGTTCCGATACTCTAAT<br>CTCCCTAG<br>GCAAGGTTTCATATTTGTGTAGGTTACTTATTCTCCTTTTGTTGACTAAGTC<br>AATAATCAGAATCAGCAGGTTTGGAGTCAGCTTGGCAGGGATCAGCAGC<br>CTGGGTTGGAAGGAGGGGGTATAAAAGCCCCTTCACCAGG<br>AGAAGCCGTCACACAGATCCACAAGCTCCTG -- 619                                                                                                                                                                                                                                                                                                                                    |
| Intrón sintético<br>(SEQ ID NO:115)                                                                                                                                                    | 622 –<br>GTAAGTGCCGTGTGTGGTTCCCGCGGGCCTGGCCTCTTTACGGGTTA<br>TGGCCCTTGCGTGCCTTGAATTACTGACACTGACATCCACTTTTCTTTTT<br>CTCCACAG --727                                                                                                                                                                                                                                                                                                                                                                                                                                                                             |
| FVIIIco6XTEN<br>(SEQ ID NO:117)<br>(marco de lectura<br>abierto para la<br>versión 6 FVIII<br>optimizada con<br>codón que<br>contiene<br>XTEN144; la<br>secuencia XTEN<br>se marca con | 739 –<br>ATGCAGATTGAGCTGTCCACTTGTTTCTTCCTGTGCCTCCTGCGCTTC<br>TGTTTCTCCGCACTCGCCGGTACTACCTTGGAGCCGTGGAGCTTTCATG<br>GGA CTACATGCAGAGCGACCTGGGCGAACTCCCCGTGGATGCCAGATTC<br>CCCCCCCCGCGTGCCAAAGTCCTTCCCCTTTAACACCTCCGTGGTGTACAA<br>GAAAACCCTCTTTGTGAGTTCACTGACCACCTGTTCAACATCGCCAAGC<br>CGCGCCACCTTGGATGGGCCTCCTGGGACCGACCATTCAAGCTGAAGT<br>GTACGACACCGTGGTGATCACCTGAAGAACATGGCGTCCCACCCCGTG<br>TCCCTGCATGCGGTGCGAGTGTCTACTGGAAGGCCTCCGAAGGAGCTG<br>AGTACGACGACCAGACTAGCCAGCGGGAAAAGGAGGACGATAAAGTGTT<br>CCCGGGCGGCTCGCATACTTACGTGTGGCAAGTCCTGAAGGAAAACGGA<br>CCTATGGCATCCGATCCTCTGTGCCTGACTTACTCCTACCTTTCCCATGT |

|                                           |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |
|-------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <u>doble subrayado</u><br>(SEQ ID NO:118) | GGACCTCGTGAAGGACCTGAACAGCGGGCTGATTGGTGCACTTCTCGTG<br>TGCCGCGAAGGTTTCGCTCGCTAAGGAAAAGACCCAGACCCTCCATAAGT<br>TCATCCTTTTGTTCGCTGTGTTTCGATGAAGGAAAGTCATGGCATTCCGAA<br>ACTAAGAACTCGCTGATGCAGGACCGGGATGCCGCCTCAGCCCCGCGCCT<br>GGCCTAAAATGCATACAGTCAACGGATACGTGAATCGGTCACTGCCCGG<br>GCTCATCGGTTGTACAGAAAGTCCGTGTACTGGCACGTATCGGCATG<br>GGCCTACGCCTGAAGTGCCTCATCTTCTGGAAGGGCACACCTTCC<br>TCGTGCGCAACCACCGCCAGGCCTCTCTGGAATCTCCCCGATTACCTTT<br>CTGACCGCCCAGACTCTGCTCATGGACCTGGGGCAGTTCCTTCTCTTCT<br>GCCACATCTCCAGCCATCAGCACGACGGAATGGAGGCCTACGTGAAGGT<br>GGACTCATGCCCCGAAGAACCTCAGTTGCGGATGAAGAACAACGAGGAG<br>GCCGAGGACTATGACGACGATTTGACTGACTCCGAGATGGACGTCTGTC<br>GGTTTCGATGACGACAACAGCCCCAGCTTCATCCAGATTGCGAGCGTGGC<br>CAAGAAGCACCCCCAAAACCTGGGTGCACTACATCGCGGCCGAGGAAGAA<br>GATTGGGACTACGCCCCGTTGGTGCTGGCACCCGATGACCGGTCTGACA<br>AGTCCCAGTATCTGAACAATGGTCCGCAGCGGATTGGCAGAAAGTACAA<br>GAAAGTGCGTTTCATGGCGTACACTGACGAAACGTTTAAGACCCGGGAG<br>GCCATTCAACATGAGAGCGGCATTCTGGGACCACTGCTGTACGGAGAGG<br>TCGGCGATACCCTGCTCATCATCTTCAAAAACAGGCCTCCCGGCCTTAC<br>AACATCTACCCTCACGGAATCACCGACGTGCGGCCACTCTACTCGCGGC<br>GCCTGCCGAAGGGCGTCAAGCACCTGAAAGACTTCCCTATCCTGCCGGG<br>CGAAATCTTCAAGTATAAGTGGACCGTCACCGTGGAGGACGGGCCACC<br>AAGAGCGATCCTAGGTGTCTGACTCGGTACTACTCCAGCTTCGTGAACAT<br>GGAACGGGACCTGGCATCGGGACTCATTGGACCGCTGCTGATCTGCTAC<br>AAAGAGTCGGTGGATCAACGCGGCAACCAGATCATGTCCGACAAGCGCA<br>ACGTGATCCTGTTCTCCGTGTTTGATGAAAACAGATCCTGGTACCTCACT<br>GAAAACATCCAGAGGTTCTCCCAAACCCCGCAGGAGTGCAACTGGAGG<br>ACCCTGAGTTTCAGGCCTCGAATATCATGCACTCGATTAACGGTTACGTG<br>TTCGACTCGCTGCAACTGAGCGTGTGCCTCCATGAAGTCGCTTACTGGTA<br>CATTCTGTCCATCGGCGCCCAGACTGACTTCCTGAGCGTGTTCTTTTCCG<br>GTTACACCTTTAAGCACAAAGATGGTGTACGAAGATACCCTGACCCTGTTC<br>CCTTTCTCCGGCGAAACGGTGTTTCATGTCGATGGAGAACCCGGGTCTGT<br>GGATTCTGGGATGCCACAACAGCGACTTTCGGAACCGCGGAATGACTGC<br>CCTGCTGAAGGTGTCCTCATGCGACAAGAACACCGGAGACTACTACGAG<br>GACTCCTACGAGGATATCTCAGCCTACCTCCTGTCCAAGAACAACGCGAT<br>CGAGCCGCGCAGCTTCAGCCAGAACGGCGCGCCAACATCAGAGAGCGC<br>CACCCCTGAAAGTGGTCCCGGGAGCGAGCCAGCCACATCTGGGTCCGA<br>AACGCCAGGCACAAGTGAGTCTGCAACTCCCGAGTCCGGACCTGGCTCC<br>GAGCCTGCCACTAGCGGCTCCGAGACTCCGGGAACCTCCGAGAGCGCT<br>ACACCAGAAAGCGGACCCGGAACAGTACCGAACCTAGCGAGGGCTCT<br>GCTCCGGGCAGCCCAGCCGGCTCTCCTACATCCACGGAGGAGGGCACT |
|-------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

|  |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |
|--|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|  | <p>TCCGAATCCGCCACCCCGGAGTCAGGGCCAGGATCTGAACCCGCTACCT<br/>CAGGCAGTGAGACGCCAGGAACGAGCGAGTCCGCTACACCGGAGAGTG<br/>GGCCAGGGAGCCCTGCTGGATCTCCTACGTCCACTGAGGAAGGGTCAC<br/>CAGCGGGCTCGCCCACCAGCACTGAAGAAGGTGCCTCGAGCCCGCCTG<br/>TGCTGAAGAGGCACCAGCGAGAAATTACCCGACCACCCTCCAATCGGA<br/>TCAGGAGGAAATCGACTACGACGACACCATCTCGGTGGAATGAAGAAG<br/>GAAGATTTGATATCTACGACGAGGACGAAAATCAGTCCCCTCGCTCATT<br/>CCAAAAGAAAAGTAGACACTACTTTATCGCCGCGGTGGAAGACTGTGG<br/>GACTATGGAATGTCATCCAGCCCTCACGTCCTTCGGAACCGGGCCCAGA<br/>GCGGATCGGTGCCTCAGTTCAAGAAAGTGGTGTTCCAGGAGTTCACCGA<br/>CGGCAGCTTCACCCAGCCGCTGTACCGGGGAGAACTGAACGAACACCTG<br/>GGCCTGCTCGGTCCCTACATCCGCGCGGAAGTGAGGATAACATCATGG<br/>TGACCTTCCGTAACCAAGCATCCAGACCTTACTCCTTCTATTCTCCCTGA<br/>TCTCATACGAGGAGGACCAGCGCCAAGGCGCCGAGCCCCGCAAGAAGT<br/>TCGTCAAGCCCAACGAGACTAAGACCTACTTCTGGAAGGTCCAACACCAT<br/>ATGGCCCCGACCAAGGATGAGTTTGACTGCAAGGCCTGGGCCTACTTCT<br/>CCGACGTGGACCTTGAGAAGGATGTCCATTCCGGCCTGATCGGGCCGCT<br/>GCTCGTGTGTCACACCAACACCCTGAACCCAGCGCATGGACGCCAGGTC<br/>ACCGTCCAGGAGTTTGCTCTGTTCTTCACCATTTTTTGACGAACTAAGTCC<br/>TGGTACTTCACCGAGAATATGGAGCGAACTGTAGAGCGCCCTGCAATAT<br/>CCAGATGGAAGATCCGACTTTCAAGGAGAACTATAGATTCCACGCCATCA<br/>ACGGGTACATCATGGATACTCTGCCGGGGCTGGTCATGGCCCAGGATCA<br/>GAGGATTCGGTGGTACTTGCTGTCAATGGGATCGAACGAAAACATTCACT<br/>CCATTCACTTCTCCGGTCACGTGTTCACTGTGCGCAAGAAGGAGGAGTA<br/>CAAGATGGCGCTGTACAATCTGTACCCCGGGGTGTTGAAACTGTGGAG<br/>ATGCTGCCGTCCAAGGCCGCATCTGGAGAGTGGAGTGCCTGATCGGA<br/>GAGCACCTCCACGCGGGGATGTCCACCCTCTTCCTGGTGTACTCGAATA<br/>AGTGCCAGACCCCGCTGGGCATGGCCTCGGGCCACATCAGAGACTTCCA<br/>GATCACAGCAAGCGGACAATACGGCCAATGGGCGCCGAAGCTGGCCCG<br/>CTTGCACTACTCCGGATCGATCAACGCATGGTCCACCAAGGAACCGTTCT<br/>CGTGGAATTAAGGTGGACCTCCTGGCCCCTATGATTATCCACGGAATTAAG<br/>ACCCAGGGCGCCAGGCAGAAGTTCTCCTCCCTGTACATCTCGCAATTCAT<br/>CATCATGTACAGCCTGGACGGGAAGAAGTGGCAGACTTACAGGGGAAAC<br/>TCCACCGGCACCCTGATGGTCTTTTTCGGCAACGTGGATTCTCCGGCAT<br/>TAAGCACAACATCTTCAACCCACCGATCATAGCCAGATATATTAGGCTCC<br/>ACCCCACTCACTACTCAATCCGCTCAACTCTTCGGATGGAAGTCAATGGGG<br/>TGCGACCTGAACTCCTGCTCCATGCCGTTGGGGATGGAATCAAAGGCTA<br/>TTAGCGACGCCAGATCACCGCGAGCTCCTACTTCACTAACATGTTCCGC<br/>ACCTGGAGCCCCTCCAAGGCCAGGCTGCACTTGACAGGGACGGTCAAATG<br/>CCTGGCGGCCGCAAGTGAACAATCCGAAGGAATGGCTTCAAGTGGATTT<br/>CCAAAAGACCATGAAAGTGACCGGAGTCACCACCCAGGGAGTGAAGTCC</p> |
|--|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       | CTTCTGACCTCGATGTATGTGAAGGAGTTCCTGATTAGCAGCAGCCAGGA<br>CGGGCACCAGTGGACCCTGTTCTTCCAAAACGGAAAGGTCAAGGTGTTT<br>CAGGGGAACCAGGACTCGTTCACACCCGTGGTGAACCTCCCTGGACCCCC<br>CACTGCTGACGCGGTACTTGAGGATTCATCCTCAGTCCTGGGTCCATCA<br>GATTGCATTGCGAATGGAAGTCCTGGGCTGCGAGGCCAGGACCTGTAC<br>TGA -- 5562                                                                                                                                                                                                                                                                                                                                                                                           |
| WPRe (elemento<br>regulador post-<br>transcrip-cional<br>del virus de la<br>hepatitis de<br>marmota mutado)<br>(SEQ ID NO:120)                                                                                                                                                                                                                                                                                                                                                                                                                        | 5582 – TCATAATCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGGT<br>ATTCTTAACATATGTTGCTCCTTTTACGCTATGTGGATACGCTGCTTTAATG<br>CCTTTGTATCATGCTATTGCTTCCCGTATGGCTTTCATTTTCTCCTCCTTG<br>TATAAATCCTGGTTGCTGTCTCTTTATGAGGAGTTGTGGCCCGTTGTCAG<br>GCAACGTGGCGTGGTGTGCACTGTGTTTGTGACGCAACCCCCACTGGT<br>TGGGGCATTGCCACCACCTGTCAGCTCCTTTCCGGGACTTTCGCTTTCCC<br>CCTCCCTATTGCCACGGCGGAACTCATCGCCGCCTGCCTTGCCCGCTGC<br>TGGACAGGGGCTCGGCTGTTGGGCACTGACAATTCCGTGGTGTGTCGG<br>GGAAATCATCGTCCTTTCTTGCTGCTCGCCTGTGTTGCCACCTGGATT<br>CTGCGCGGGACGTCCTTCTGCTACGTCCCTTCGGCCCTCAATCCAGCGG<br>ACCTTCCTTCCCGCGGCCTGCTGCCGGCTCTGCGGCCTCTTCCGCGTCT<br>TCGCCTTCGCCCTCAGACGAGTCGGATCTCCCTTTGGGCCGCCTCCCCG<br>CTG -- 6176 |
| bGHpA (señal de<br>poliadenilación de<br>la hormona del<br>crecimiento de<br>bovina) (SEQ ID<br>NO:122)                                                                                                                                                                                                                                                                                                                                                                                                                                               | 6185 –<br>CGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCTCCCCC<br>GTGCCTTCCTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATA<br>AAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGG<br>GGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATA<br>GCAGGCATGCTGGGGA -- 6395                                                                                                                                                                                                                                                                                                                                                                                                                              |
| repetición<br>terminal invertida<br>3' ITR (SEQ ID<br>NO: 181)                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        | 6515 –<br>CTCTGGGCCAGCTTGCTTGGGGTTGCCTTGACACTAAGACAAGCGG<br>CGCGCCGCTTGATCTTAGTGGCACGTCAACCCCAAGCGCTGGCCCAGAG<br>CCAACCCTAATTCCGGAAGTCCCGCCCACCGGAAGTGACGTCACAGGAA<br>ATGACGTCACAGGAAATGACGTAATTGTCCGCCATCTTGACCGGAAGTC<br>CCGCCTACCGGCGGCGACCGGCGGCATCTGATTTGGTGTCTTCTTTTAA<br>ATTTT -- 6762                                                                                                                                                                                                                                                                                                                                                                                   |
| <b>Secuencia de longitud completa (SEQ ID NO: 179)</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |
| CTCTGGGCCAGCTTGCTTGGGGTTGCCTTGACACTAAGACAAGCGGCGCGCCGCTTGATCTT<br>AGTGGCACGTCAACCCCAAGCGCTGGCCCAGAGCCAACCCTAATTCCGGAAGTCCCGCCCAC<br>CGGAAGTGACGTCACAGGAAATGACGTCACAGGAAATGACGTAATTGTCCGCCATCTTGATACC<br>GGAAGTCCCGCCTACCGGCGGCGACCGGCGGCATCTGATTTGGTGTCTTCTTTTAAATTTTGC<br>GGCAATTCAGTCGATAACTATAACGGTCCTAAGGTAGCGATTTAAATACGCGCTCTCTTAAGGT<br>AGCCCCGGGACGCGTCAATTGAGATCTGGATCCGGTACCGAATTCGCGGCCGCCTCGACGAC<br>TAGCGTTTAATTAAACGCGTGTCTGTCTGCACATTTCTAGAGCGAGTGTTCCGATACTCTAAT<br>CTCCCTAGGCAAGGTTTCATATTTGTGTAGGTTACTTATTCTCCTTTTGTGACTAAGTCAATAAT |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |

CAGAATCAGCAGGTTTGGAGTCAGCTTGGCAGGGATCAGCAGCCTGGGTTGGAAGGAGGGG  
GTATAAAAGCCCCCTTACCAGGAGAAGCCGTCACACAGATCCACAAGCTCCTGAGGTAAGTGC  
CGTGTGTGGTTCCCGCGGGCCTGGCCTCTTTACGGGTTATGGCCCTTGC GTGCCTTGAATTAC  
TGACACTGACATCCACTTTTTCTTTTTCTCCACAGCTAGCGCCACCATGCAGATTGAGCTGTCC  
ACTTGTTTCTTCCTGTGCCTCCTGCGCTTCTGTTTCTCCGCCACTCGCCGGTACTACCTTGAG  
CCGTGGAGCTTTCATGGGACTACATGCAGAGCGACCTGGGCGAACTCCCCGTGGATGCCAGA  
TTCCCCCCCCCGCTGCCAAAGTCCTTCCCCTTTAACACCTCCGTGGTGTACAAGAAAACCTC  
TTTGTGAGTTCACTGACCACCTGTTCAACATCGCCAAGCCGCGCCACCTTGATGGGCCTC  
CTGGGACCGACCATTCAAGCTGAAGTGTACGACACCGTGGTGATCACCTGAAGAACATGGC  
GTCCACCCCCGTGTCCCTGCATGCGGTGCGAGTGTCTACTGGAAGGCCTCCGAAGGAGCTG  
AGTACGACGACCAGACTAGCCAGCGGGAAAAGGAGGACGATAAAGTGTTCCCGGGCGGCTC  
GCATACTTACGTGTGGCAAGTCCTGAAGGAAAACGGACCTATGGCATCCGATCCTCTGTGCCT  
GACTTACTCCTACCTTTCCCATGTGGACCTCGTGAAGGACCTGAACAGCGGGCTGATTGGTGC  
ACTTCTCGTGTGCCGCGAAGGTTGCTCGCTAAGGAAAAGACCCAGACCCTCATAAGTTCAT  
CCTTTTGTTGCTGTGTTGATGAAGGAAAGTCATGGCATTCCGAAACTAAGAACTCGCTGATG  
CAGGACCGGGATGCCGCCTCAGCCCGCGCCTGGCCTAAAATGCATACAGTCAACGGATACGT  
GAATCGGTCACTGCCCCGGGCTCATCGGTTGTACAGAAAGTCCGTGTACTGGCACGTATCG  
GCATGGGCACTACGCCTGAAGTGA CTCTTCCTGGAAGGGCACACCTTCCTCGTGCGC  
AACCACCGCCAGGCCTCTCTGGAATCTCCCCGATTACCTTTCTGACCGCCCAGACTCTGCTC  
ATGGACCTGGGGCAGTTCCTTCTCTTCTGCCACATCTCCAGCCATCAGCACGACGGAATGGAG  
GCCTACGTGAAGGTGGACTCATGCCCGGAAGAACCTCAGTTGCGGATGAAGAACAACGAGGA  
GGCCGAGGACTATGACGACGATTTGACTGACTCCGAGATGGACGTCGTGCGGTTGATGACG  
ACAACAGCCCCAGCTTCATCCAGATTCGCAGCGTGGCCAAGAAGCACCCCAAAACCTGGGTG  
CACTACATCGCGGCCGAGGAAGAAGATTGGGACTACGCCCCGTTGGTGCTGGCACCCGATGA  
CCGGTCGTACAAGTCCCAGTATCTGAACAATGGTCCGCAGCGGATTGGCAGAAAGTACAAGAA  
AGTGCGGTTTCATGGCGTACACTGACGAAACGTTTAAGACCCGGGAGGCCATTCAACATGAGA  
GCGGCATTCTGGGACCACTGCTGTACGGAGAGGTGCGCGATACCCTGCTCATCATCTTCAAAA  
ACCAGGCCTCCCGGCCCTTACAACATCTACCCTCACGGAATCACCGACGTGCGGCCACTCTACT  
CGCGGCGCCTGCCGAAGGGCGTCAAGCACCTGAAAGACTTCCCTATCCTGCCGGGCGAAATC  
TTCAAGTATAAGTGGACCGTCACCGTGGAGGACGGGCCCACCAAGAGCGATCCTAGGTGTCT  
GACTCGGTACTACTCCAGCTTCGTGAACATGGAACGGGACCTGGCATCGGGACTCATTGGAC  
CGCTGCTGATCTGCTACAAAGAGTCGGTGGATCAACGCGGCAACCAGATCATGTCCGACAAG  
CGCAACGTGATCCTGTTCTCCGTGTTTGATGAAAACAGATCCTGGTACCTCACTGAAAACATCC  
AGAGGTTCCCTCCCAAACCCCGCAGGAGTGCAACTGGAGGACCCTGAGTTTCAGGCCTCGAAT  
ATCATGCACTCGATTAACGGTTACGTGTTGACTCGCTGCAACTGAGCGTGTGCCTCCATGAA  
GTCGCTTACTGGTACATTCTGTCCATCGGCGCCCAGACTGACTTCCTGAGCGTGTTCTTTTCC  
GGTTACACCTTTAAGCACAAGATGGTGTACGAAGATACCCTGACCCTGTTCCCTTTCTCCGGC  
GAAACGGTGTTTCATGTGATGGAGAACCCGGGTCTGTGGATTCTGGGATGCCACAACAGCGA  
CTTTCGGAACCGCGGAATGACTGCCCTGCTGAAGGTGTCTCATGCGACAAGAACACCGGAG  
ACTACTACGAGGACTCCTACGAGGATATCTCAGCCTACCTCCTGTCCAAGAACAACGCGATCG  
AGCCGCGCAGCTTCAGCCAGAACGGCGCGCCAACATCAGAGAGCGCCACCCCTGAAAGTGG

TCCCGGGAGCGAGCCAGCCACATCTGGGTCGGAAACGCCAGGCACAAGTGAGTCTGCAACTC  
CCGAGTCCGGACCTGGCTCCGAGCCTGCCACTAGCGGCTCCGAGACTCCGGGAACTTCCGA  
GAGCGCTACACCAGAAAGCGGACCCGGAACCAGTACCGAACCTAGCGAGGGCTCTGCTCCG  
GGCAGCCCAGCCGGCTCTCCTACATCCACGGAGGAGGGCACTTCCGAATCCGCCACCCCGG  
AGTCAGGGCCAGGATCTGAACCCGCTACCTCAGGCAGTGAGACGCCAGGAACGAGCGAGTC  
CGCTACACCGGAGAGTGGGCCAGGGAGCCCTGCTGGATCTCCTACGTCCACTGAGGAAGGG  
TCACCAGCGGGCTCGCCCACCAGCACTGAAGAAGGTGCCTCGAGCCCGCCTGTGCTGAAGA  
GGCACCAGCGAGAAATTACCCGGACCACCCTCCAATCGGATCAGGAGGAAATCGACTACGAC  
GACACCATCTCGGTGGAAATGAAGAAGGAAGATTTGATATCTACGACGAGGACGAAAATCAG  
TCCCCTCGCTCATTCCAAAAGAAAAGTAGACACTACTTTATCGCCGCGGTGGAAGACTGTGG  
GACTATGGAATGTCATCCAGCCCTCACGTCCTTCGGAACCGGGCCCAGAGCGGATCGGTGCC  
TCAGTTCAAGAAAGTGGTGTTCAGGAGTTACCCGACGGCAGCTTACCCAGCCGCTGTACC  
GGGGAGAACTGAACGAACACCTGGGCCTGCTCGGTCCCTACATCCGCGCGGAAGTGGAGGA  
TAACATCATGGTGACCTTCGTAACCAAGCATCCAGACCTTACTCCTTCTATTCTCCCTGATC  
TCATACGAGGAGGACCAGCGCCAAGGCGCCGAGCCCCGCAAGAACTTCGTCAAGCCCAACG  
AGACTAAGACCTACTTCTGGAAGGTCCAACACCATATGGCCCCGACCAAGGATGAGTTTGACT  
GCAAGGCCTGGGCCTACTTCTCCGACGTGGACCTTGAGAAGGATGTCCATTCCGGCCTGATC  
GGGCCGCTGCTCGTGTGTACACCAACACCCTGAACCCAGCGCATGGACGCCAGGTCACCGT  
CCAGGAGTTTGCTCTGTTCTTACCATTTTTGACGAACTAAGTCCTGGTACTTCACCGAGAAT  
ATGGAGCGAACTGTAGAGCGCCCTGCAATATCCAGATGGAAGATCCGACTTTCAAGGAGAAC  
TATAGATTCCACGCCATCAACGGGTACATCATGGATACTCTGCCGGGGCTGGTCATGGCCCAG  
GATCAGAGGATTGCGTGGTACTTGCTGTCAATGGGATCGAACGAAAACATTCACTCCATTAC  
TTCTCCGGTCACGTGTTCACTGTGCGCAAGAAGGAGGAGTACAAGATGGCGCTGTACAATCTG  
TACCCCGGGGTGTTGAACTGTGGAGATGCTGCCGTCCAAGGCCGGCATCTGGAGAGTGGA  
GTGCCTGATCGGAGAGCACCTCCACGCGGGGATGTCCACCCTCTTCTGGTGTACTCGAATA  
AGTGCCAGACCCCGCTGGGCATGGCCTCGGGCCACATCAGAGACTTCCAGATCACAGCAAGC  
GGACAATACGGCCAATGGGCGCCGAAGCTGGCCCGCTTGCACTACTCCGGATCGATCAACGC  
ATGGTCCACCAAGGAACCGTTCTCGTGGATTAAGGTGGACCTCCTGGCCCCTATGATTATCCA  
CGGAATTAAGACCCAGGGCGCCAGGCAGAAGTTCTCCTCCCTGTACATCTCGCAATTCATCAT  
CATGTACAGCCTGGACGGGAAGAAGTGGCAGACTTACAGGGGAACTCCACCGGCACCCTGA  
TGGTCTTTTTTCGGCAACGTGGATTCTCCGGCATTAAGCACAACATCTTCAACCCACCGATCAT  
AGCCAGATATATTAGGCTCCACCCCACTCACTACTCAATCCGCTCAACTCTTCGGATGGAACTC  
ATGGGGTGCGACCTGAACTCCTGCTCCATGCCGTTGGGGATGGAATCAAAGGCTATTAGCGA  
CGCCCAGATCACCGCGAGCTCCTACTTCACTAACATGTTGCCACCTGGAGCCCCTCCAAGG  
CCAGGCTGCACTTGCAAGGACGGTCAAATGCCTGGCGGCCGCAAGTGAACAATCCGAAGGAA  
TGGCTTCAAGTGGATTTCCAAAAGACCATGAAAGTGACCGGAGTACCACCCAGGGAGTGAA  
GTCCCTTCTGACCTCGATGTATGTGAAGGAGTTCCTGATTAGCAGCAGCCAGGACGGGCACC  
AGTGGACCCTGTTCTTCCAAAACGGAAAGGTCAAGGTGTTCCAGGGGAACCAGGACTCGTTCA  
CACCCGTGGTGAACTCCCTGGACCCCCCACTGCTGACGCGGTACTTGAGGATTCATCCTCAG  
TCCTGGGTCCATCAGATTGCATTGCGAATGGAAGTCCTGGGCTGCGAGGCCAGGACCTGTA  
CTGAATCAGCCTGAGCTCGCTGATCATAATCAACCTCTGGATTACAAAATTTGTGAAAGATTGA

CTGGTATTCTTAACTATGTTGCTCCTTTTACGCTATGTGGATACGCTGCTTTAATGCCTTTGTAT  
CATGCTATTGCTTCCCGTATGGCTTTCATTTTCTCCTCCTTGTATAAATCCTGGTTGCTGTCTCT  
TTATGAGGAGTTGTGGCCCGTTGTCAGGCAACGTGGCGTGGTGTGCACTGTGTTTGCTGACG  
CAACCCCCACTGGTTGGGGCATTGCCACCACCTGTCAGCTCCTTTCCGGGACTTTCGCTTTCC  
CCCTCCCTATTGCCACGGCGGAACATCGCCGCCTGCCTTGCCCGCTGCTGGACAGGGGCT  
CGGCTGTTGGGCACTGACAATTCCGTGGTGTGTCGGGGAAATCATCGTCCTTTCTTGCTG  
CTCGCTGTGTTGCCACCTGGATTCTGCGCGGGACGTCCTTCTGCTACGTCCCTTCGGCCCT  
CAATCCAGCGGACCTTCCTTCCCGCGGCCTGCTGCCGGCTCTGCGGCCTCTTCGCGCTCTTC  
GCCTTCGCCCTCAGACGAGTCGGATCTCCCTTTGGGCCGCCTCCCCGCTGATCAGCCTCGAC  
TGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCTCCCCCGTGCCTTCCTTGACCCTGGA  
AGGTGCCACTCCCCTGTCCTTTCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGG  
TGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACA  
ATAGCAGGCATGCTGGGGATGCGGTGGGCTCTATGGCTTCTGAGGCGGAAAGAACGGGCTC  
GAGAAGCTTCTAGATATCCTCTCTTAAGGTAGCATCGAGATTTAAATTAGGGATAACAGGGTAA  
TGCGCGGGCCCGCAAAATTTAAAAGAAGACACCAAATCAGATGCCGCCGGTCGCCGCCGGTA  
GGCGGGACTTCCGGTACAAGATGGCGGACAATTACGTCATTTCTGTGACGTCATTTCTGTG  
ACGTCATTCCGGTGGGCGGGACTTCCGGAATTAGGGTTGGCTCTGGGCCAGCGCTTGGGGT  
TGACGTGCCACTAAGATCAAGCGGCGCGCCGCTTGTCTTAGTGTCAAGGCAACCCCAAGCAA  
GCTGGCCCAGAG

Tabla 2C: Ejemplo de construcción GPV-FVIII (nucleótidos 1-6830; SEQ ID NO: 182)

| Descripción                                           | Secuencia                                                                                                                                                                                                                                                                                                                       |
|-------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 5' ITR (SEQ ID NO: 183)                               | 1<br>CGGTGACGTGTTTCCGGCTGTTAGGTTGACCACGCGCATGCCGCGCGGT<br>CAGCCCAATA<br>GTTAAGCCGGAAACACGTCACCGGAAGTCACATGACCGGAAGTCACGTGA<br>CCGGAAACACGTGACAGGAAGCACGTGACCGGAACTACGTCACCGGATG<br>TGCGTCACCGGAAGCATGTGACCGGAACTTGCGTCACTTCCCCCTCCCCT<br>GATTGGCTGGTTCGAACGAACGAACCCTCCAATGAGACTCAAGGACAAGA<br>GGATATTTTGC GCGCCAGGAAGTG -- 282 |
| TTPp (promotor específico del hígado) (SEQ ID NO:113) | 425<br>ACGCGTGTCTGTCTGCACATTTCTAGAGCGAGTGTTCCGATACTCTAATC<br>TCCCTAG<br>GCAAGGTTTCATATTTGTGTAGGTTACTTATTCTCCTTTTGTTGACTAAGTCA<br>ATAATCAGAATCAGCAGGTTTGGAGTCAGCTTGGCAGGGATCAGCAGCCT<br>GGGTTGGAAGGAGGGGGTATAAAAGCCCCCTTCACCAGGAGAAGCCGTCA<br>CACAGATCCACAAGCTCCTG -- 653                                                        |
| Intrón sintético (SEQ ID NO:115)                      | 656<br>GTAAGTGCCGTGTGTGGTTCCCGCGGGCCTGGCCTCTTTACGGGTTATGG<br>CCCTTGCG<br>TGCCTTGAATTACTGACACTGACATCCACTTTTTCTTTTCTCCACAG -- 761                                                                                                                                                                                                 |

|                                                                                                                                                                                                                                                                           |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <p>FVIIIco6XTEN<br/>(SEQ ID NO:117)<br/>(marco de lectura<br/>abierto para la<br/>versión 6 FVIII<br/>optimizada con<br/>codón que<br/>contiene<br/>XTEN144; la<br/>secuencia XTEN<br/>se marca con<br/><u>doble subrayado</u><br/><u>(SEQ ID</u><br/><u>NO:118))</u></p> | <p>773<br/>—</p> <p>ATGCAGATTGAGCTGTCCACTTGTTTCTTCCTGTGCCTCCTGCGCTTCTGT<br/>TTCTCCG</p> <p>CCACTCGCCGGTACTACCTTGGAGCCGTGGAGCTTTCATGGGACTACATG<br/>CAGAGCGACCTGGGCGAACTCCCCGTGGATGCCAGATTCCCCCCCCGCG<br/>TGCCAAAGTCCTTCCCCTTTAACACCTCCGTGGTGTACAAGAAAACCCTCT<br/>TTGTTCGAGTTCAGTGACCACCTGTTCAACATCGCCAAGCCGCGCCACCT<br/>TGGATGGGCCTCCTGGGACCGACCATTCAGCTGAAGTGTACGACACCGT<br/>GGTGATCACCTGAAGAACATGGCGTCCCACCCCGTGTCCCTGCATGCGG<br/>TCGGAGTGTCTACTGGAAGGCCTCCGAAGGAGCTGAGTACGACGACCA<br/>GACTAGCCAGCGGGAAAAGGAGGACGATAAAGTGTTCGCGGGCGGCTCG<br/>CATACTTACGTGTGGCAAGTCCTGAAGGAAAACGGACCTATGGCATCCGA<br/>TCCTCTGTGCCTGACTTACTCCTACCTTTCCCATGTGGACCTCGTGAAGGA<br/>CCTGAACAGCGGGCTGATTGGTGCATTCTCGTGTGCCGCGAAGGTTTCG<br/>TCGCTAAGGAAAAGACCCAGACCCTCCATAAGTTCATCCTTTTGTTCGCTG<br/>TGTTTCGATGAAGGAAAAGTCATGGCATTCCGAACTAAGAACTCGCTGATGC<br/>AGGACCGGGATGCCGCCTCAGCCCGCGCCTGGCCTAAAATGCATACAGT<br/>CAACGGATACGTGAATCGGTCACTGCCCCGGGCTCATCGGTTGTACAGAA<br/>AGTCCGTGTACTGGCACGTATCGGCATGGGCACTACGCCTGAAGTGCAC<br/>TCCATCTTCCTGGAAGGGCACACCTTCCTCGTGCGCAACCACCGCCAGGC<br/>CTCTCTGGAAATCTCCCCGATTACCTTTCTGACCGCCCAGACTCTGCTCAT<br/>GGACCTGGGGCAGTTCCTTCTTCTGCCACATCTCCAGCCATCAGCACG<br/>ACGGAATGGAGGCCTACGTGAAGGTGGACTCATGCCCGGAAGAACCTCA<br/>GTTGCGGATGAAGAACAACGAGGAGGCCGAGGACTATGACGACGATTTGA<br/>CTGACTCCGAGATGGACGTCGTGCGGTTTCGATGACGACAACAGCCCCAG<br/>CTTCATCCAGATTTCGACGCGTGCCCAAGAAGCACCCCAAAACCTGGGTGC<br/>ACTACATCGCGGCCGAGGAAGAAGATTGGGACTACGCCCCGTTGGTGCT<br/>GGCACCCGATGACCGGTCTGACAAAGTCCCAGTATCTGAACAATGGTCCGC<br/>AGCGGATTGGCAGAAAGTACAAGAAAGTGCGGTTTCATGGCGTACACTGAC<br/>GAAACGTTTAAGACCCGGGAGGCCATTCAACATGAGAGCGGCATTCTGGG<br/>ACCACTGCTGTACGGAGAGGTGCGCGATACCCTGCTCATCATCTTCAAAA<br/>ACCAGGCCTCCCGGCCTTACAACATCTACCCTCACGGAATCACCGACGTG<br/>CGGCCACTCTACTCGCGGCGCCTGCCGAAGGGCGTCAAGCACCTGAAAG<br/>ACTTCCCTATCCTGCCGGGCGAAATCTTCAAGTATAAGTGGACCGTCACC<br/>GTGGAGGACGGGCCCCACCAAGAGCGATCCTAGGTGTCTGACTCGGTACT<br/>ACTCCAGCTTCGTGAACATGGAACGGGACCTGGCATCGGGACTCATTGGA<br/>CCGCTGCTGATCTGCTACAAAGAGTCGGTGGATCAACGCGGCAACCAGAT<br/>CATGTCCGACAAGCGCAACGTGATCCTGTTCTCCGTGTTTGATGAAAACAG<br/>ATCCTGGTACCTCACTGAAAACATCCAGAGGTTCTCCCAAACCCCGCAG<br/>GAGTGCAACTGGAGGACCCTGAGTTTCAGGCCTCGAATATCATGCACTCG<br/>ATTAACGGTTACGTGTTCTGACTCGCTGCAACTGAGCGTGTGCCTCCATGAA</p> |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

|  |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |
|--|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|  | <p>GTCGCTTACTGGTACATTCTGTCCATCGGCGCCCAGACTGACTTCCTGAG<br/>CGTGTTCTTTTCCGGTTACACCTTTAAGCACAAAGATGGTGTACGAAGATAC<br/>CCTGACCCTGTTCCCTTTCTCCGGCGAAACGGTGTTTCATGTTCGATGGAGA<br/>ACCCGGGTCTGTGGATTCTGGGATGCCACAACAGCGACTTTTCGGAACCGC<br/>GGAATGACTGCCCTGCTGAAGGTGTCCTCATGCGACAAGAACACCGGAGA<br/>CTACTACGAGGACTCCTACGAGGATATCTCAGCCTACCTCCTGTCCAAGAA<br/>CAACGCGATCGAGCCGCGCAGCTTCAGCCAGAACGGCGCGCCAACATCA<br/>GAGAGCGCCACCCCTGAAAGTGGTCCCGGGAGCGAGCCAGCCACATCTG<br/>GGTCGGAAACGCCAGGCACAAGTGAGTCTGCAACTCCCGAGTCCGGACC<br/>TGGCTCCGAGCCTGCCACTAGCGGCTCCGAGACTCCGGGAACCTCCGAG<br/>AGCGCTACACCAGAAAGCGGACCCGGAACCAGTACCGAACCTAGCGAGG<br/>GCTCTGCTCCGGGCAGCCCAGCCGGCTCTCCTACATCCACGGAGGAGGG<br/>CACTTCCGAATCCGCCACCCCGGAGTCAGGGCCAGGATCTGAACCCGCTA<br/>CCTCAGGCAGTGAGACGCCAGGAACGAGCGAGTCCGCTACACCGGAGAG<br/>TGGGCCAGGGAGCCCTGCTGGATCTCCTACGTCCACTGAGGAAGGGTCA<br/>CCAGCGGGCTCGCCCACCAGCACTGAAGAAGGTGCCTCGAGCCCGCCTG<br/>TGCTGAAGAGGCACCAGCGAGAAATTACCCGGACCACCCTCCAATCGGAT<br/>CAGGAGGAAATCGACTACGACGACACCATCTCGGTGGAATGAAGAAGGA<br/>AGATTTTCGATATCTACGACGAGGACGAAAATCAGTCCCCTCGCTCATTCCA<br/>AAAGAAAAGTAGACACTACTTTATCGCCGCGGTGGAAGACTGTGGGACT<br/>ATGGAATGTCATCCAGCCCTCACGTCCTTCGGAACCGGGGCCAGAGCGGA<br/>TCGGTGCCTCAGTTCAAGAAAGTGGTGTTCAGGAGTTCACCGACGGCAG<br/>CTTCACCCAGCCGCTGTACCGGGGAGAACTGAACGAACACCTGGGCCTG<br/>CTCGGTCCCTACATCCGCGCGGAAGTGGAGGATAACATCATGGTGACCTT<br/>CCGTAACCAAGCATCCAGACCTTACTCCTTCTATTCTCCTGATCTCATA<br/>CGAGGAGGACCAGCGCCAAGGCGCCGAGCCCCGCAAGAACTTCGTCAAG<br/>CCCAACGAGACTAAGACCTACTTCTGGAAGGTCCAACACCATATGGCCCC<br/>GACCAAGGATGAGTTTGACTGCAAGGCCTGGGCCTACTTCTCCGACGTGG<br/>ACCTTGAGAAGGATGTCCATTCCGGCCTGATCGGGCCGCTGCTCGTGTGT<br/>CACACCAACACCCTGAACCCAGCGCATGGACGCCAGGTCACCGTCCAGG<br/>AGTTTGCTCTGTTCTTCACCATTTTTGACGAAACTAAGTCCTGGTACTTCAC<br/>CGAGAATATGGAGCGAAACTGTAGAGCGCCCTGCAATATCCAGATGGAAG<br/>ATCCGACTTTCAAGGAGAACTATAGATTCCACGCCATCAACGGGTACATCA<br/>TGGATACTCTGCCGGGGCTGGTCATGGCCCAGGATCAGAGGATTCGGTG<br/>GTACTTGCTGTCAATGGGATCGAACGAAAACATTCACTCCATTCACTTCTC<br/>CGGTACGTGTTCACTGTGCGCAAGAAGGAGGAGTACAAGATGGCGCTGT<br/>ACAATCTGTACCCCGGGGTGTTGAAACTGTGGAGATGCTGCCGTCCAAG<br/>GCCGGCATCTGGAGAGTGGAGTGCCTGATCGGAGAGCACCTCCACGCGG<br/>GGATGTCCACCCTCTTCCTGGTGTACTCGAATAAGTGCCAGACCCCGCTG<br/>GGCATGGCCTCGGGCCACATCAGAGACTTCCAGATCACAGCAAGCGGAC<br/>AATACGGCCAATGGGCGCCGAAGCTGGCCCGCTTGCACTACTCCGGATC</p> |
|--|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

|                                                                                                                                            |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |
|--------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|                                                                                                                                            | <p>GATCAACGCATGGTCCACCAAGGAACCGTTCTCGTGGATTAAGGTGGACC<br/> TCCTGGCCCCTATGATTATCCACGGAATTAAGACCCAGGGCGCCAGGCAG<br/> AAGTTCTCCTCCCTGTACATCTCGCAATTCATCATCATGTACAGCCTGGAC<br/> GGGAAGAAGTGGCAGACTTACAGGGGAAACTCCACCGGCACCCTGATGG<br/> TCTTTTTTCGGCAACGTGGATTCTCCGGCATTAAAGCACAACATCTTCAACC<br/> CACCGATCATAGCCAGATATATTAGGCTCCACCCCACTCACTACTCAATCC<br/> GCTCAACTCTTCGGATGGAATCATGGGGTGCGACCTGAACTCCTGCTCC<br/> ATGCCGTTGGGGATGGAATCAAAGGCTATTAGCGACGCCAGATCACCGC<br/> GAGCTCCTACTTCACTAACATGTTCCGCACCTGGAGCCCCTCCAAGGCCA<br/> GGCTGCACTTGACAGGGACGGTCAAATGCCTGGCGGCCGCAAGTGAACAA<br/> TCCGAAGGAATGGCTTCAAGTGGATTTCAAAAGACCATGAAAGTGACCG<br/> GAGTCACCACCCAGGGAGTGAAGTCCCTTCTGACCTCGATGTATGTGAAG<br/> GAGTTCCTGATTAGCAGCAGCCAGGACGGGCACCAGTGGACCCTGTTCTT<br/> CCAAAACGGAAAGGTCAAGGTGTTCCAGGGGAACCAGGACTCGTTACAC<br/> CCGTGGTGAATCCCTGGACCCCCCACTGCTGACGCGGTACTTGAGGATT<br/> CATCCTCAGTCCTGGGTCCATCAGATTGCATTGCGAATGGAAGTCCTGGG<br/> CTGCGAGGCCCAGGACCTGTACTGA --5596</p> |
| <p>WPRE (elemento<br/>regulador post-<br/>transcripcional<br/>del virus de la<br/>hepatitis de<br/>marmota mutado)<br/>(SEQ ID NO:120)</p> | <p>5616<br/> TCATAATCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGGTATTCTT<br/> AACT<br/> ATGTTGCTCCTTTTACGCTATGTGGATACGCTGCTTTAATGCCTTTGTATCA<br/> TGCTATTGCTTCCCGTATGGCTTTCATTTTCTCCTCCTTGATAAATCCTGG<br/> TTGCTGTCTCTTTATGAGGAGTTGTGGCCCGTTGTCAGGCAACGTGGCGT<br/> GGTGTGCACTGTGTTTGCTGACGCAACCCCCCACTGGTTGGGGCATTGCCA<br/> CCACCTGTCAGCTCCTTTCCGGGACTTTCGCTTTCCCCCTCCCTATTGCCA<br/> CGGCGGAACTCATCGCCGCTGCCTTGCCCGCTGCTGGACAGGGGCTCG<br/> GCTGTTGGGCACTGACAATTCCGTGGTGTGTCGGGGAAATCATCGTCCT<br/> TTCCTTGGCTGCTCGCCTGTGTTGCCACCTGGATTCTGCGCGGGACGTCC<br/> TTCTGCTACGTCCCTTCGGCCCTCAATCCAGCGGACCTTCCTTCCCGCGG<br/> CCTGCTGCCGGCTCTGCGGCCTCTTCGCGCTTCGCTTCGCCCTCAGA<br/> CGAGTCGGATCTCCCTTTGGGCCGCTCCCCGCTG -- 6210</p>                                                                                                                                                                                                                                                        |
| <p>bGHpA (señal de<br/>poliadenilación<br/>de la hormona<br/>del crecimiento<br/>de bovina) (SEQ<br/>ID NO:122)</p>                        | <p>6219<br/> CGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTG<br/> CCTTCCTT<br/> GACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAATGAGGAAAT<br/> TGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGG<br/> GGCAGGACAGCAAGGGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGG<br/> GGA -- 6429</p>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           |
| <p>repetición<br/>terminal invertida</p>                                                                                                   | <p>6549<br/> CGGTGACGTGTTTCCGGCTGTTAGGTTGACCACGCGCATGCCGCGCGGT<br/> CAGCCCA</p>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |

|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     |                                                                                                                                                                                                                                                         |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 3' ITR (SEQ ID NO: 184)                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             | ATAGTTAAGCCGGAACACGTACCGGAAGTCACATGACCGGAAGTCACG<br>TGACCGGAACACGTGACAGGAAGCACGTGACCGGAACTACGTACCGG<br>ATGTGCGTCACCGGAAGCATGTGACCGGAACTTGCGTCACTTCCCCCTCC<br>CCTGATTGGCTGGTTCGAACGAACGAACCCTCCAATGAGACTCAAGGACA<br>AGAGGATATTTTGC GCGCCAGGAAGTG -- 6830 |
| <b>Secuencia completa (SEQ ID NO: 182)</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |                                                                                                                                                                                                                                                         |
| CGGTGACGTGTTTCCGGCTGTTAGGTTGACCACGCGCATGCCGCGCGGTGAGCCCAATAGTT<br>AAGCCGGAACACGTACCGGAAGTCACATGACCGGAAGTCACGTGACCGGAACACGTGAC<br>AGGAAGCACGTGACCGGAACTACGTACCGGATGTGCGTCACCGGAAGCATGTGACCGGAAC<br>TTGCGTCACTTCCCCCTCCCCTGATTGGCTGGTTCGAACGAACGAACCCTCCAATGAGACTCA<br>AGGACAAGAGGATATTTTGC GCGCCAGGAAGTGCGGCAATTCAGTCGATAACTATAACGGTC<br>CTAAGGTAGCGATTTAAATACGCGCTCTCTTAAGGTAGCCCCGGGACGCGTCAATTGAGATCT<br>GGATCCGGTACCGAATTCGCGGCCGCCTCGACGACTAGCGTTTAATTAAACGCGTGTCTGTCT<br>GCACATTTCTAGAGCGAGTGTTCCGATACTCTAATCTCCCTAGGCAAGGTTTCATATTTGTGTA<br>GGTTACTTATTCTCCTTTTGTGACTAAGTCAATAATCAGAATCAGCAGGTTTGGAGTCAGCTTG<br>GCAGGGATCAGCAGCTGGGTTGGAAGGAGGGGGTATAAAAGCCCCCTTACCAGGAGAAGC<br>CGTCACACAGATCCACAAGCTCCTGAGGTAAGTGCCGTGTGTGGTTCCCGCGGGCCTGGCCT<br>CTTTACGGGTTATGGCCCTTGCGTGCCTTGAATTACTGACACTGACATCCACTTTTTCTTTTTCT<br>CCACAGCTAGCGCCACCATGCAGATTGAGCTGTCCACTTGTTTCTTCTGTGCCTCCTGCGCT<br>TCTGTTTCTCCGCCACTCGCCGGTACTACCTTGAGCCGTGGAGCTTTCATGGGACTACATGC<br>AGAGCGACCTGGGCGAACTCCCCGTGGATGCCAGATTCCCCCCCCGCGTGCCAAAGTCCTTC<br>CCCTTTAACACCTCCGTGGTGTACAAGAAAACCCTCTTTGTGAGTTCACTGACCACCTGTTCA<br>ACATCGCCAAGCCGCGCCACCTTGGATGGGCCTCCTGGGACCGACCATTAAGCTGAAGTG<br>TACGACACCGTGTTGATCACCTGAAGAACATGGCGTCCCACCCCGTGTCCCTGCATGCGGT<br>CGGAGTGTCTACTGGAAGGCCTCCGAAGGAGCTGAGTACGACGACCAGACTAGCCAGCGG<br>GAAAAGGAGGACGATAAAGTGTTCCCGGGCGGCTCGCATACTTACGTGTGGCAAGTCCTGAA<br>GGAAAACGGACCTATGGCATCCGATCCTCTGTGCCTGACTTACTCCTACCTTTCCCATGTGGA<br>CCTCGTGAAGGACCTGAACAGCGGGCTGATTGGTGCATTCTCGTGTGCCGGAAGGTTTCGC<br>TCGCTAAGGAAAAGACCCAGACCCTCCATAAGTTCATCCTTTTGTTCGCTGTGTTTCGATGAAGG<br>AAAGTCATGGCATTCCGAAACTAAGAACTCGCTGATGCAGGACCGGGATGCCGCCTCAGCCC<br>GCGCCTGGCCTAAAATGCATACAGTCAACGGATACGTGAATCGGTCACTGCCCGGGCTCATC<br>GGTTGTCACAGAAAGTCCGTGTACTGGCACGTATCGGCATGGGCACTACGCCTGAAGTGCA<br>CTCCATCTTCTGGAAGGGCACACCTTCCTCGTGCGCAACCACCGCCAGGCCTCTCTGGAAAT<br>CTCCCCGATTACCTTTCTGACCGCCCAGACTCTGCTCATGGACCTGGGGCAGTTCCTTCTCTT<br>CTGCCACATCTCCAGCCATCAGCACGACGAATGGAGGCCTACGTGAAGGTGGACTCATGCC<br>CGGAAGAACCTCAGTTGCGGATGAAGAACAACGAGGAGGCCGAGGACTATGACGACGATTG<br>ACTGACTCCGAGATGGACGTGTCGGTTCGATGACGACAACAGCCCCAGCTTCATCCAGATT<br>CGCAGCGTGGCCAAGAAGCACCCCAAACCTGGGTGCACTACATCGCGGCCGAGGAAGAAG<br>ATTGGGACTACGCCCCGTTGGTGCTGGCACCCGATGACCGGTCGTACAAGTCCCAGTATCTG<br>AACAAATGGTCCGCAGCGGATTGGCAGAAAGTACAAGAAAGTGCGGTTTCATGGCGTACACTGA<br>CGAAACGTTTAAGACCCGGGAGGCCATTCAACATGAGAGCGGCATTCTGGGACCACTGCTGT |                                                                                                                                                                                                                                                         |

ACGGAGAGGTCGGCGATACCCTGCTCATCATCTTCAAAAACCAGGCCTCCCGGCCTTACAACA  
TCTACCCTCACGGAATCACCGACGTGCGGCCACTCTACTCGCGGCGCCTGCCGAAGGGCGTC  
AAGCACCTGAAAGACTTCCCTATCCTGCCGGGCGAAATCTTCAAGTATAAGTGGACCGTCACC  
GTGGAGGACGGGCCCACCAAGAGCGATCCTAGGTGTCTGACTCGGTACTACTCCAGCTTCGT  
GAACATGGAACGGGACCTGGCATCGGGACTCATTGGACCGCTGCTGATCTGCTACAAAGAGT  
CGGTGGATCAACGCGGCAACCAGATCATGTCCGACAAGCGCAACGTGATCCTGTTCTCCGTG  
TTTGATGAAAACAGATCCTGGTACCTCACTGAAAACATCCAGAGGTTCCCTCCCAAACCCCGCA  
GGAGTGCAACTGGAGGACCCTGAGTTTCAGGCCTCGAATATCATGCACTCGATTAACGGTTAC  
GTGTTGCACTCGCTGCAACTGAGCGTGTGCCTCCATGAAGTCGCTTACTGGTACATTCTGTCC  
ATCGGCGCCCAGACTGACTTCCTGAGCGTGTTCTTTTCCGGTTACACCTTTAAGCACAAGATG  
GTGTACGAAGATACCCTGACCCTGTTCCCTTTCTCCGGCGAAACGGTGTTTCATGTGATGGAG  
AACCCGGGTCTGTGGATTCTGGGATGCCACAACAGCGACTTTCGGAACCGCGGAATGACTGC  
CCTGCTGAAGGTGTCTCATGCGACAAGAACACCGGAGACTACTACGAGGACTCCTACGAGG  
ATATCTCAGCCTACCTCCTGTCCAAGAACAACGCGATCGAGCCGCGCAGCTTCAGCCAGAACG  
GCGCGCCAACATCAGAGAGCGCCACCCCTGAAAGTGGTCCCGGGAGCGAGCCAGCCACATC  
TGGGTGCGAAACGCCAGGCACAAGTGAGTCTGCAACTCCCGAGTCCGGACCTGGCTCCGAGC  
CTGCCACTAGCGGCTCCGAGACTCCGGGAACTTCCGAGAGCGCTACACCAGAAAGCGGACCC  
GGAACCAGTACCGAACCTAGCGAGGGCTCTGCTCCGGGCAGCCCAGCCGGCTCTCCTACATC  
CACGGAGGAGGGCACTTCCGAATCCGCCACCCCGGAGTCAGGGCCAGGATCTGAACCCGCT  
ACCTCAGGCAGTGAGACGCCAGGAACGAGCGAGTCCGCTACACCGGAGAGTGGGCCAGGGA  
GCCCTGCTGGATCTCCTACGTCCACTGAGGAAGGGTCAACCAGCGGGCTCGCCCACCAGCACT  
GAAGAAGGTGCCTCGAGCCCGCCTGTGCTGAAGAGGCACCAGCGAGAAATTACCCGGACCAC  
CCTCCAATCGGATCAGGAGGAAATCGACTACGACGACACCATCTCGGTGGAAATGAAGAAGG  
AAGATTTGATATCTACGACGAGGACGAAAATCAGTCCCCTCGCTCATTCCAAAAGAAAAGTAG  
ACACTACTTTATCGCCGCGGTGGAAAGACTGTGGGACTATGGAATGTCATCCAGCCCTCACGT  
CCTTCGGAACCGGGGCCAGAGCGGATCGGTGCCTCAGTTCAAGAAAGTGGTGTTCCAGGAGT  
TCACCGACGGCAGCTTCACCCAGCCGCTGTACCGGGGAGAACTGAACGAACACCTGGGCCTG  
CTCGGTCCCTACATCCGCGCGGAAGTGGAGGATAACATCATGGTGACCTTCCGTAACCAAGCA  
TCCAGACCTTACTCCTTCTATTCTCCTGATCTCATACGAGGAGGACCAGCGCCAAGGCGCC  
GAGCCCCGCAAGAACTTCGTCAAGCCCAACGAGACTAAGACCTACTTCTGGAAGGTCCAACAC  
CATATGGCCCCGACCAAGGATGAGTTTGACTGCAAGGCCTGGGCCTACTTCTCCGACGTGGA  
CCTTGAGAAGGATGTCCATTCCGGCCTGATCGGGCCGCTGCTCGTGTGTACACCAACACCC  
TGAACCCAGCGCATGGACGCCAGGTCACCGTCCAGGAGTTTGCTCTGTTCTTACCATTTTTG  
ACGAAACTAAGTCCTGGTACTTCACCGAGAATATGGAGCGAACTGTAGAGCGCCCTGCAATA  
TCCAGATGGAAGATCCGACTTTCAAGGAGAACTATAGATTCCACGCCATCAACGGGTACATCA  
TGGATACTCTGCCGGGGCTGGTCATGGCCCAGGATCAGAGGATTCCGGTGGTACTTGCTGTCA  
ATGGGATCGAACGAAAACATTCACTCCATTCACTTCTCCGGTCACGTGTTCACTGTGCGCAAG  
AAGGAGGAGTACAAGATGGCGCTGTACAATCTGTACCCCGGGGTGTTGCAAACTGTGGAGAT  
GCTGCCGTCCAAGGCCGGCATCTGGAGAGTGGAGTGCCTGATCGGAGAGCACCTCCACGCG  
GGGATGTCCACCCTCTTCTGGTGTACTCGAATAAGTGCCAGACCCCGCTGGGCATGGCCTC  
GGGCCACATCAGAGACTTCCAGATCACAGCAAGCGGACAATACGGCCAATGGGCGCCGAAGC

```

TGGCCCGCTTGCACTACTCCGGATCGATCAACGCATGGTCCACCAAGGAACCGTTCTCGTGG
ATTAAGGTGGACCTCCTGGCCCCCTATGATTATCCACGGAATTAAGACCCAGGGCGCCAGGCA
GAAGTTCTCCTCCCTGTACATCTCGCAATTCATCATCATGTACAGCCTGGACGGGAAGAAGTG
GCAGACTTACAGGGGAAACTCCACCGGCACCCTGATGGTCTTTTTTCGGCAACGTGGATTCTCTC
CGGCATTAAGCACAAACATCTTCAACCCACCGATCATAGCCAGATATATTAGGCTCCACCCCACT
CACTACTCAATCCGCTCAACTCTTCGGATGGAACCTCATGGGGTGCGACCTGAACTCCTGCTCC
ATGCCGTTGGGGATGGAATCAAAGGCTATTAGCGACGCCAGATCACCGCGAGCTCCTACTTC
ACTAACATGTTTCGCCACCTGGAGCCCCTCCAAGGCCAGGCTGCACTTGAGGGACGGTCAAA
TGCCTGGCGGCCGCAAGTGAACAATCCGAAGGAATGGCTTCAAGTGGATTTCAAAAGACCAT
GAAAGTGACCGGAGTCACCACCCAGGGAGTGAAGTCCCTTCTGACCTCGATGTATGTGAAGG
AGTTCCTGATTAGCAGCAGCCAGGACGGGCACCACTGGACCCTGTTCTTCCAAAACGGAAAG
GTCAAGGTGTTCCAGGGGAACCAAGGACTCGTTTCACACCCGTGGTGAACCTCCCTGGACCCCC
ACTGCTGACGCGGTACTTGAGGATTCATCCTCAGTCCCTGGGTCCATCAGATTGCATTGCGAAT
GGAAGTCTGGGCTGCGAGGCCAGGACCTGTACTGAATCAGCCTGAGCTCGCTGATCATAA
TCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGGTATTCTTAACCTATGTTGCTCCTTTTA
CGCTATGTGGATACGCTGCTTTAATGCCTTTGTATCATGCTATTGCTTCCCGTATGGCTTTCATT
TTCTCCTCCTTGATAAATCCTGGTTGCTGTCTCTTTATGAGGAGTTGTGGCCCGTTGTCAGGC
AACGTGGCGTGGTGTGCACTGTGTTTGCTGACGCAACCCCCACTGGTTGGGGCATTGCCACC
ACCTGTCAGTCCCTTCCGGGACTTTTCGCTTTCCCCCTCCCTATTGCCACGGCGGAACTCATC
GCCGCTGCCTTGCCCGCTGCTGGACAGGGGCTCGGCTGTTGGGCACTGACAATTCCGTGGT
GTTGTGCGGGAAATCATCGTCCCTTTCCTTGGCTGCTCGCCTGTGTTGCCACCTGGATTCTGCG
CGGGACGTCCTTCTGCTACGTCCCTTCGGCCCTCAATCCAGCGGACCTTCCTTCCCGCGGCC
TGCTGCCGGCTCTGCGGCCTCTTCCGCGTCTTCGCCTTCGCCCTCAGACGAGTCGGATCTCC
CTTTGGGCCGCTCCCCGCTGATCAGCCTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGT
TTGCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATA
AAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCAATTCTATTCTGGGGGGTGGGGTGGG
GCAGGACAGCAAGGGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGATGCGGTGGGC
TCTATGGCTTCTGAGGCGGAAAGAACGGGCTCGAGAAGCTTCTAGATATCCTCTCTTAAGGTA
GCATCGAGATTTAAATTAGGGATAACAGGGTAATGGCGCGGGCCGCCACTTCCTGGCGCGCA
AAATATCCTCTTGTCCTTGAGTCTCATTGGAGGGTTCGTTTCGTTTGAACCAGCCAATCAGGGG
AGGGGGAAGTGACGCAAGTTCGGGTCACATGCTTCCGGTGACGCACATCCGGTGACGTAGTT
CCGGTCACGTGCTTCTGTACGTGTTTCCGGTCACGTGACTTCCGGTCATGTGACTTCCGGT
GACGTGTTTCCGGCTTAACCTATTGGGCTGACCGCGCGGCATGCGCGTGGTCAACCTAACAGC
CGGAAACACGTCACCG

```

### ***A. Índice de adaptación de codones***

**[00198]** En una realización, el casete génico comprende una secuencia de nucleótidos con codón optimizada que codifica un polipéptido FVIII, en el que el índice de adaptación del codón humano de la secuencia de nucleótidos con codón optimizado aumenta en relación con la SEQ ID NO: 16. Por ejemplo, la secuencia de

nucleótidos optimizada con codón puede tener un índice de adaptación de codón humano que sea al menos aproximadamente 0,75 (75%, al menos aproximadamente 0,76 (76%, al menos aproximadamente 0,77 (77%, al menos aproximadamente 0,78 (78%, al menos aproximadamente 0,79 (79%, al menos aproximadamente 0,80 (80%, al menos aproximadamente 0,81 (81%, al menos aproximadamente 0,82 (82%, al menos aproximadamente 0,83 (83%, al menos aproximadamente 0,84 (84%, al menos aproximadamente 0,85 (85%, al menos aproximadamente 0,86 (86%, al menos aproximadamente 0,87 (87%, al menos aproximadamente 0,88 (88%, al menos aproximadamente 0,89 (89%, al menos aproximadamente 0,90 (90%, al menos aproximadamente 0,91 (91%, al menos aproximadamente 0,92 (92%, al menos aproximadamente 0,93 (93%, al menos aproximadamente 0,94 (94%, al menos aproximadamente 0,95 (95%, al menos aproximadamente 0,96 (96%, al menos aproximadamente 0,97 (97%, al menos aproximadamente 0,98 (98%, o al menos aproximadamente 0,99 (99%). En algunas realizaciones, la secuencia de nucleótidos optimizada con codón tiene un índice de adaptación de codón humano que es al menos aproximadamente 0,88 (88%). En otras realizaciones, la secuencia de nucleótidos optimizada con codón tiene un índice de adaptación de codón humano que es al menos aproximadamente 0,91 (91%). En otras realizaciones, la secuencia de nucleótidos con codón optimizado tiene un índice de adaptación de codón humano que es al menos aproximadamente 0,91 (97%).

**[00199]** En una realización particular, la secuencia de nucleótidos optimizada por codón que codifica un polipéptido FVIII comprende una secuencia de nucleótidos que comprende una primera secuencia de ácido nucleico que codifica una porción N-terminal de un polipéptido FVIII y una segunda secuencia de ácido nucleico que codifica una porción C-terminal de un polipéptido FVIII; en donde la primera secuencia de ácido nucleico tiene al menos aproximadamente 80%, al menos aproximadamente 85%, al menos aproximadamente 86%, al menos aproximadamente 87%, al menos aproximadamente 88%, al menos aproximadamente 89%, al menos aproximadamente 90%, al menos aproximadamente 91%, al menos aproximadamente 92%, al menos aproximadamente 93%, al menos aproximadamente 94%, al menos aproximadamente 95%, al menos aproximadamente 96%, al menos aproximadamente 97%, al menos aproximadamente 98% o al menos aproximadamente 99% de identidad de secuencia con (i) los nucleótidos 58-1791 de la SEQ ID N°: 3; (ii) los nucleótidos 1-1791 de SEQ ID NO: 3; (iii) los nucleótidos 58-1791 de SEQ ID NO: 4; o (iv) los nucleótidos 1-1791 de SEQ ID NO: 4; en donde la porción N-terminal y la porción C-terminal juntas tienen

una actividad de polipéptido FVIII; y en donde el índice de adaptación del codón humano de la secuencia de nucleótidos se incrementa con relación a la SEQ ID NO: 16. En algunas realizaciones, la secuencia de nucleótidos tiene un índice de adaptación del codón humano que es al menos aproximadamente 0,75 (75%, al menos aproximadamente 0,76 (76%, al menos aproximadamente 0,77 (77%, al menos aproximadamente 0,78 (78%), al menos aproximadamente 0,79 (79%, al menos aproximadamente 0,80 (80%, al menos aproximadamente 0,81 (81%, al menos aproximadamente 0,82 (82%, al menos aproximadamente 0,83 (83%, al menos aproximadamente 0,84 (84%, al menos aproximadamente 0,85 (85%, al menos aproximadamente 0,86 (86%, al menos aproximadamente 0,87 (87%, al menos aproximadamente 0,88 (88%, al menos aproximadamente 0,89 (89%, al menos aproximadamente 0,90 (90%), o al menos aproximadamente 0,91 (91%). En una realización particular, la secuencia de nucleótidos tiene un índice de adaptación del codón humano que es al menos aproximadamente 0,88 (88%). En otra realización, la secuencia de nucleótidos tiene un índice de adaptación del codón humano que es al menos aproximadamente 0,91 (91%).

**[00200]** En otra realización, la secuencia de nucleótidos optimizada por codón que codifica un polipéptido FVIII comprende una secuencia de nucleótidos que comprende una primera secuencia de ácido nucleico que codifica una porción N-terminal de un polipéptido FVIII y una segunda secuencia de ácido nucleico que codifica una porción C-terminal de un polipéptido FVIII; en donde la segunda secuencia de ácido nucleico tiene al menos aproximadamente 80%, al menos aproximadamente 85%, al menos aproximadamente 86%, al menos aproximadamente 87%, al menos aproximadamente 88%, al menos aproximadamente 89%, al menos aproximadamente 90%, al menos aproximadamente 91%, al menos aproximadamente 92%, al menos aproximadamente 93%, al menos aproximadamente 94%, al menos aproximadamente 95%, al menos aproximadamente 96%, al menos aproximadamente 97%, al menos aproximadamente 98%, o al menos aproximadamente 99% de identidad de secuencia con (i) los nucleótidos 1792-2277 y 2320-4374 de SEQ ID NO: 5 o (ii) 1792-2277 y 2320-4374 de SEQ ID NO: 6; en donde la porción N-terminal y la porción C-terminal juntas tienen una actividad de polipéptido FVIII; y en donde el índice de adaptación del codón humano de la secuencia de nucleótidos se incrementa con relación a la SEQ ID NO: 16. En algunas realizaciones, la secuencia de nucleótidos tiene un índice de adaptación del codón humano que es al menos aproximadamente 0,75 (75%, al menos aproximadamente 0,76 (76%, al menos aproximadamente 0,77

(77%, al menos aproximadamente 0,78 (78%), al menos aproximadamente 0,79 (79%, al menos aproximadamente 0,80 (80%, al menos aproximadamente 0,81 (81%, al menos aproximadamente 0,82 (82%, al menos aproximadamente 0,83 (83%, al menos aproximadamente 0,84 (84%, al menos aproximadamente 0,85 (85%, al menos aproximadamente 0,86 (86%, al menos aproximadamente 0,87 (87%) o al menos aproximadamente 0,88 (88%). En una realización particular, la secuencia de nucleótidos tiene un índice de adaptación de codones humanos que es al menos aproximadamente 0,83 (83%)). En otra realización, la secuencia de nucleótidos tiene un índice de adaptación del codón humano que es al menos aproximadamente 0,88 (88%).

**[00201]** En otras realizaciones, la secuencia de nucleótidos optimizada por codón que codifica un polipéptido FVIII comprende una secuencia de nucleótidos que codifica un polipéptido con actividad de FVIII, en donde la secuencia de nucleótidos comprende una secuencia de ácido nucleico que tiene al menos aproximadamente 80%, al menos aproximadamente 85%, al menos aproximadamente 89%, al menos aproximadamente 90%, al menos aproximadamente 91%, al menos aproximadamente 92%, al menos aproximadamente 93%, al menos aproximadamente 94%, al menos aproximadamente 95%, al menos aproximadamente 96%, al menos aproximadamente 97%, al menos aproximadamente 98%, o al menos aproximadamente 99% de identidad de secuencia con los nucleótidos 58-2277 y 2320-4374 de una secuencia de aminoácidos seleccionada de SEQ ID NO: 1, 2, 3, 4, 5, 6, 70 y 71 (es decir, los nucleótidos 58-4374 de SEQ ID NO: 1, 2, 3, 4, 5, 6, 70 o 71 sin los nucleótidos que codifican el dominio B o el fragmento del dominio B); y en donde el índice de adaptación del codón humano de la secuencia de nucleótidos se incrementa con relación a la SEQ ID NO: 16. En algunas realizaciones, la secuencia de nucleótidos tiene un índice de adaptación del codón humano que es al menos aproximadamente 0,75 (75%, al menos aproximadamente 0,76 (76%, al menos aproximadamente 0,77 (77%, al menos aproximadamente 0,78 (78%), al menos aproximadamente 0,79 (79%, al menos aproximadamente 0,80 (80%, al menos aproximadamente 0,81 (81%, al menos aproximadamente 0,82 (82%, al menos aproximadamente 0,83 (83%, al menos aproximadamente 0,84 (84%, al menos aproximadamente 0,85 (85%, al menos aproximadamente 0,86 (86%, al menos aproximadamente 0,87 (87%) o al menos aproximadamente 0,88 (88%). En una realización particular, la secuencia de nucleótidos tiene un índice de adaptación del codón humano que es al menos aproximadamente 0,75 (75%). En otra realización, la secuencia de nucleótidos tiene

un índice de adaptación del codón humano que es al menos aproximadamente 0,83 (83%). En otra realización, la secuencia de nucleótidos tiene un índice de adaptación del codón humano que es al menos aproximadamente 0,88 (88%). En otra realización, la secuencia de nucleótidos tiene un índice de adaptación del codón humano que es al menos aproximadamente 0,91 (91%). En otra realización, la secuencia de nucleótidos tiene un índice de adaptación del codón humano que es al menos aproximadamente 0,97 (97%).

**[00202]** En algunas realizaciones, la secuencia de nucleótidos optimizada por codón que codifica un polipéptido FVIII de la presente descripción tiene una frecuencia aumentada de codones óptimos (FOP) con respecto a la SEQ ID NO: 16. En ciertas realizaciones, la FOP de la secuencia de nucleótidos optimizada por codón que codifica un polipéptido FVIII es al menos aproximadamente 40, al menos aproximadamente 45, al menos aproximadamente 50, al menos aproximadamente 55, al menos aproximadamente 60, al menos aproximadamente 64, al menos aproximadamente 65, al menos aproximadamente 70, al menos aproximadamente 75, al menos aproximadamente 79, al menos aproximadamente 80, al menos aproximadamente 85 o al menos aproximadamente 90.

**[00203]** En otras realizaciones, la secuencia de nucleótidos optimizada por codón que codifica un polipéptido FVIII de la presente descripción tiene un uso de codón sinónimo relativo aumentado (RCSU) en relación con la SEQ ID NO: 16. En algunas realizaciones, la RCSU de la molécula de ácido nucleico aislada es mayor que 1,5. En otras realizaciones, la RCSU de la molécula de ácido nucleico aislada es mayor que 2,0. En ciertas realizaciones, la RCSU de la molécula de ácido nucleico aislada es al menos aproximadamente 1,5, al menos aproximadamente 1,6, al menos aproximadamente 1,7, al menos aproximadamente 1,8, al menos aproximadamente 1,9, al menos aproximadamente 2,0, al menos aproximadamente 2,1, al menos aproximadamente 2,2, al menos aproximadamente 2,3, al menos aproximadamente 2,4, al menos aproximadamente 2,5, al menos aproximadamente 2,6, o al menos aproximadamente 2,7.

**[00204]** En otras realizaciones más, la secuencia de nucleótidos optimizada por codón que codifica un polipéptido FVIII de la presente descripción tiene un número eficaz disminuido de codones en relación con la SEQ ID NO: 16. En algunas realizaciones, la secuencia de nucleótidos optimizada por codón que codifica un polipéptido FVIII tiene un número efectivo de codones de menos de aproximadamente 50, menos de aproximadamente 45, menos de aproximadamente 40, menos de

aproximadamente 35, menos de aproximadamente 30 o menos de aproximadamente 25. En una realización particular, la molécula de ácido nucleico aislada tiene un número efectivo de codones de aproximadamente 40, aproximadamente 35, aproximadamente 30, aproximadamente 25 o aproximadamente 20.

## 5 **B. Optimización de contenido G/C**

**[00205]** En algunas realizaciones, el casete génico comprende una secuencia de nucleótidos con codón optimizado que codifica un polipéptido FVIII, en donde la secuencia de nucleótidos con codón optimizado contiene un mayor porcentaje de nucleótidos G/C en comparación con el porcentaje de nucleótidos G/C en la SEQ ID  
10 NO: 16. En otras realizaciones, la secuencia de nucleótidos optimizada por codón que codifica un polipéptido FVIII tiene un contenido de G/C que es al menos aproximadamente 45%, al menos aproximadamente 46%, al menos aproximadamente 47%, al menos aproximadamente 48%, al menos aproximadamente 49%, al menos aproximadamente 50%, al menos aproximadamente 51%, al menos aproximadamente  
15 52%, al menos aproximadamente 53%, al menos aproximadamente 54%, al menos aproximadamente 55%, al menos aproximadamente 56%, al menos aproximadamente 57%, al menos aproximadamente 58%, al menos aproximadamente 59%, o al menos aproximadamente 60%.

**[00206]** En una realización particular, la secuencia de nucleótidos optimizada por  
20 codón que codifica un polipéptido FVIII comprende una primera secuencia de ácido nucleico que codifica una porción N-terminal de un polipéptido FVIII y una segunda secuencia de ácido nucleico que codifica una porción C-terminal de un polipéptido FVIII; en donde la primera secuencia de ácido nucleico tiene al menos aproximadamente 80%, al menos aproximadamente 85%, al menos aproximadamente  
25 86%, al menos aproximadamente 87%, al menos aproximadamente 88%, al menos aproximadamente 89%, al menos aproximadamente 90%, al menos aproximadamente 91%, al menos aproximadamente 92%, al menos aproximadamente 93%, al menos aproximadamente 94%, al menos aproximadamente 95%, al menos aproximadamente 96%, al menos aproximadamente 97%, al menos aproximadamente 98% o al menos  
30 aproximadamente 99% de identidad de secuencia con (i) los nucleótidos 58-1791 de la SEQ ID N°: 3; (ii) los nucleótidos 1-1791 de SEQ ID NO: 3; (iii) los nucleótidos 58-1791 de SEQ ID NO: 4; o (iv) los nucleótidos 1-1791 de SEQ ID NO: 4; en donde la porción N-terminal y la porción C-terminal juntas tienen una actividad de polipéptido FVIII; y en el que la secuencia de nucleótidos contiene un mayor porcentaje de  
35 nucleótidos G/C en comparación con el porcentaje de nucleótidos G/C en SEQ ID NO:

16. En algunas realizaciones, la secuencia de nucleótidos optimizada por codón que codifica un polipéptido FVIII comprende que tiene un contenido de G/C que es al menos aproximadamente 45%, al menos aproximadamente 46%, al menos aproximadamente 47%, al menos aproximadamente 48%, al menos aproximadamente 49%, al menos aproximadamente 50%, al menos aproximadamente 51%, al menos aproximadamente 52%, al menos aproximadamente 53%, al menos aproximadamente 54%, al menos aproximadamente 55%, al menos aproximadamente 56%, al menos aproximadamente 57%, o al menos aproximadamente el 58%. En una realización particular, la secuencia de nucleótidos que codifica un polipéptido con actividad de FVIII tiene un contenido de G/C que es al menos aproximadamente 58%.

**[00207]** En otra realización, la secuencia de nucleótidos optimizada por codón que codifica un polipéptido FVIII comprende una primera secuencia de ácido nucleico que codifica una porción N-terminal de un polipéptido FVIII y una segunda secuencia de ácido nucleico que codifica una porción C-terminal de un polipéptido FVIII; en donde la segunda secuencia de ácido nucleico tiene al menos aproximadamente 80%, al menos aproximadamente 85%, al menos aproximadamente 86%, al menos aproximadamente 87%, al menos aproximadamente 88%, al menos aproximadamente 89%, al menos aproximadamente 90%, al menos aproximadamente 91%, al menos aproximadamente 92%, al menos aproximadamente 93%, al menos aproximadamente 94%, al menos aproximadamente 95%, al menos aproximadamente 96%, al menos aproximadamente 97%, al menos aproximadamente 98%, o al menos aproximadamente 99% de identidad de secuencia con (i) los nucleótidos 1792-4374 de SEQ ID NO: 5; (ii) los nucleótidos 1792-4374 de SEQ ID NO: 6; (iii) los nucleótidos 1792-2277 y 2320-4374 de SEQ ID NO: 5 (es decir, los nucleótidos 1792-4374 de SEQ ID NO: 5 sin los nucleótidos que codifican el dominio B o el fragmento de dominio B), o (iv) los 1792-2277 y 2320-4374 de SEQ ID NO: 6 (es decir, los nucleótidos 1792-4374 de SEQ ID NO: 6 sin los nucleótidos que codifican el dominio B o el fragmento de dominio B); en donde la porción N-terminal y la porción C-terminal juntas tienen una actividad de polipéptido FVIII; y en el que la secuencia de nucleótidos optimizada con codón contiene un mayor porcentaje de nucleótidos G/C en comparación con el porcentaje de nucleótidos G/C en SEQ ID NO: 16. En otras realizaciones, la secuencia de nucleótidos optimizada por codón que codifica un polipéptido FVIII tiene un contenido de G/C que es al menos aproximadamente 45%, al menos aproximadamente 46%, al menos aproximadamente 47%, al menos aproximadamente 48%, al menos aproximadamente 49%, al menos aproximadamente 50%, al menos

aproximadamente 51%, al menos aproximadamente 52%, al menos aproximadamente 53%, al menos aproximadamente 54%, al menos aproximadamente 55%, al menos aproximadamente 56% o al menos aproximadamente 57%. En una realización particular, la secuencia de nucleótidos optimizada por codón que codifica un polipéptido FVIII tiene un contenido de G/C que es al menos aproximadamente 52%. En otra realización, la secuencia de nucleótidos optimizada por codón que codifica un polipéptido FVIII tiene un contenido de G/C que es al menos aproximadamente 55%. En otra realización, la secuencia de nucleótidos optimizada por codón que codifica un polipéptido FVIII tiene un contenido de G/C que es al menos aproximadamente 57%.

**[00208]** En otras realizaciones, el casete génico comprende una secuencia de nucleótidos con codón optimizado que codifica un polipéptido FVIII, en donde la secuencia de nucleótidos con codón optimizado comprende una secuencia de ácido nucleico que tiene al menos aproximadamente 80%, al menos aproximadamente 85%, al menos aproximadamente 89%, al menos aproximadamente 90%, al menos aproximadamente 91%, al menos aproximadamente 92%, al menos aproximadamente 93%, al menos aproximadamente 94%, al menos aproximadamente 95%, al menos aproximadamente 96%, al menos aproximadamente 97%, al menos aproximadamente 98%, o al menos aproximadamente 99% de identidad de secuencia con (i) los nucleótidos 58-4374 o (ii) los nucleótidos 58-2277 y 2320-4374 de una secuencia de aminoácidos seleccionada de SEQ ID NOs: 1, 2, 3, 4, 5, 6, 70 y 71 (es decir, los nucleótidos 58-4374 de SEQ ID NO: 1, 2, 3, 4, 5, 6, 70 o 71 sin los nucleótidos que codifican el dominio B o el fragmento del dominio B); y en el que la secuencia de nucleótidos contiene un mayor porcentaje de nucleótidos G/C en comparación con el porcentaje de nucleótidos G/C en SEQ ID NO: 16. En otras realizaciones, la secuencia de nucleótidos optimizada por codón que codifica un polipéptido FVIII tiene un contenido de G/C que es al menos aproximadamente 45%. En una realización particular, la secuencia de nucleótidos optimizada por codón que codifica un polipéptido FVIII tiene un contenido de G/C que es al menos aproximadamente 52%. En otra realización, la secuencia de nucleótidos optimizada por codón que codifica un polipéptido FVIII tiene un contenido de G/C que es al menos aproximadamente 55%. En otra realización, la secuencia de nucleótidos optimizada por codón que codifica un polipéptido FVIII tiene un contenido de G/C que es al menos aproximadamente 57%. En otra realización, la secuencia de nucleótidos optimizada por codón que codifica un polipéptido FVIII tiene un contenido de G/C que es al menos aproximadamente 58%. En otra realización más, la secuencia de nucleótidos optimizada con codón n que

codifica un polipéptido FVIII tiene un contenido de G/C que es al menos aproximadamente 60%.

**[00209]** "Contenido de G/C" (o contenido de guanina-citosina), o "porcentaje de nucleótidos de G/C", se refiere al porcentaje de bases nitrogenadas en una molécula de ADN que son guanina o citosina. El contenido G/C se puede calcular usando la siguiente fórmula:

$$\frac{G + C}{A + T + G + C} \times 100$$

**[00210]** Los genes humanos son altamente heterogéneos en su contenido de G/C, con algunos genes que tienen un contenido de G/C tan bajo como 20%, y otros genes que tienen un contenido de G/C tan alto como 95%. En general, los genes ricos en G/C se expresan más altamente. De hecho, se ha demostrado que aumentar el contenido de G/C de un gen puede conducir a una mayor expresión del gen, debido principalmente a un aumento en la transcripción y niveles más altos de ARNm en estado estacionario. Ver Kudla et al., PLoS Biol., 4 (6): el 80 (2006).

### **C. Secuencias de tipo región de unión a matriz**

**[00211]** En algunas realizaciones, el casete génico comprende una secuencia de nucleótidos con codón optimizado que codifica un polipéptido FVIII, en el que la secuencia de nucleótidos con codón optimizado contiene menos secuencias MARS/ARS en relación con la SEQ ID NO: 16. En otras realizaciones, la secuencia de nucleótidos optimizada por codón que codifica un polipéptido FVIII contiene como máximo 6, como máximo 5, como máximo 4, como máximo 3, o como máximo 2 secuencias MARS/ARS. En otras realizaciones, la secuencia de nucleótidos optimizada por codón que codifica un polipéptido FVIII contiene como máximo 1 secuencia MARS/ARS. En otras realizaciones más, la secuencia de nucleótidos optimizada por codón que codifica un polipéptido FVIII no contiene ninguna secuencia MARS/ARS.

**[00212]** En una realización particular, la secuencia de nucleótidos optimizada por codón que codifica un polipéptido FVIII comprende una primera secuencia de ácido nucleico que codifica una porción N-terminal de un polipéptido FVIII y una segunda secuencia de ácido nucleico que codifica una porción C-terminal de un polipéptido FVIII; en donde la primera secuencia de ácido nucleico tiene al menos aproximadamente 80%, al menos aproximadamente 85%, al menos aproximadamente 86%, al menos aproximadamente 87%, al menos aproximadamente 88%, al menos aproximadamente 89%, al menos aproximadamente 90%, al menos aproximadamente

91%, al menos aproximadamente 92%, al menos aproximadamente 93%, al menos aproximadamente 94%, al menos aproximadamente 95%, al menos aproximadamente 96%, al menos aproximadamente 97%, al menos aproximadamente 98%, o al menos aproximadamente 99% de identidad de secuencia con (i) los nucleótidos 58-1791 de la SEQ ID N°: 3; (ii) los nucleótidos 1-1791 de SEQ ID NO: 3; (iii) los nucleótidos 58-1791 de SEQ ID NO: 4; o (iv) los nucleótidos 1-1791 de SEQ ID NO: 4; en donde la porción N-terminal y la porción C-terminal juntas tienen una actividad del polipéptido FVIII; y en el que la secuencia de nucleótidos optimizada con codón contiene menos secuencias MARS/ARS en relación con la SEQ ID NO: 16. En otras realizaciones, la secuencia de nucleótidos que codifica un polipéptido con actividad de FVIII contiene como máximo 6, como máximo 5, como máximo 4, como máximo 3, o como máximo 2 secuencias MARS/ARS. En otras realizaciones, la secuencia de nucleótidos optimizada por codón que codifica un polipéptido FVIII contiene como máximo 1 secuencia MARS/ARS. En otras realizaciones más, la secuencia de nucleótidos optimizada por codón que codifica un polipéptido FVIII no contiene ninguna secuencia MARS/ARS.

**[00213]** En otra realización, la secuencia de nucleótidos optimizada por codón que codifica un polipéptido FVIII comprende una primera secuencia de ácido nucleico que codifica una porción N-terminal de un polipéptido FVIII y una segunda secuencia de ácido nucleico que codifica una porción C-terminal de un polipéptido FVIII; en donde la segunda secuencia de ácido nucleico tiene al menos aproximadamente 80%, al menos aproximadamente 85%, al menos aproximadamente 86%, al menos aproximadamente 87%, al menos aproximadamente 88%, al menos aproximadamente 89%, al menos aproximadamente 90%, al menos aproximadamente 91%, al menos aproximadamente 92%, al menos aproximadamente 93%, al menos aproximadamente 94%, al menos aproximadamente 95%, al menos aproximadamente 96%, al menos aproximadamente 97%, al menos aproximadamente 98%, o al menos aproximadamente 99% de identidad de secuencia con (i) los nucleótidos 1792-4374 de la SEQ ID N°: 5; (ii) los nucleótidos 1792-4374 de SEQ ID NO: 6; (iii) los nucleótidos 1792-2277 y 2320-4374 de SEQ ID NO: 5 (es decir, los nucleótidos 1792-4374 de SEQ ID NO: 5 sin los nucleótidos que codifican el dominio B o el fragmento de dominio B); o (iv) los nucleótidos 1792-2277 y 2320-4374 de la SEQ ID NO: 6 (es decir, los nucleótidos 1792-4374 de la SEQ ID NO: 6 sin los nucleótidos que codifican el dominio B o el fragmento del dominio B); en donde la porción N-terminal y la porción C-terminal juntas tienen una actividad de polipéptido FVIII; y en donde la secuencia de

nucleótidos contiene menos secuencias MARS/ARS en relación con la SEQ ID NO: 16. En otras realizaciones, la secuencia de nucleótidos optimizada por codón que codifica un polipéptido FVIII contiene como máximo 6, como máximo 5, como máximo 4, como máximo 3, o como máximo 2 secuencias MARS/ARS. En otras realizaciones, la secuencia de nucleótidos optimizada por codón que codifica un polipéptido FVIII contiene como máximo 1 secuencia MARS/ARS. En otras realizaciones más, la secuencia de nucleótidos optimizada por codón que codifica un polipéptido FVIII no contiene una secuencia MARS/ARS.

**[00214]** En otras realizaciones, el casete génico comprende una secuencia de nucleótidos con codón optimizado que codifica un polipéptido FVIII, en donde la secuencia de nucleótidos con codón optimizado comprende una secuencia de ácido nucleico que tiene al menos aproximadamente 80%, al menos aproximadamente 85%, al menos aproximadamente 89%, al menos aproximadamente 90%, al menos aproximadamente 91%, al menos aproximadamente 92%, al menos aproximadamente 93%, al menos aproximadamente 94%, al menos aproximadamente 95%, al menos aproximadamente 96%, al menos aproximadamente 97%, al menos aproximadamente 98%, o al menos aproximadamente 99% de identidad de secuencia con (i) los nucleótidos 58-4374 de SEQ ID N°: 1, 2, 3, 4, 5, 6, 70 o 71 o (ii) los nucleótidos 58-2277 y 2320-4374 de SEQ ID NO: 1, 2, 3, 4, 5, 6, 70 o 71 (es decir, los nucleótidos 58-4374 de SEQ ID NO: 1, 2, 3, 4, 5, 6, 70, o 71 sin los nucleótidos que codifican el dominio B o el fragmento del dominio B); y en el que la secuencia de nucleótidos optimizada con codón contiene menos secuencias MARS/ARS en relación con la SEQ ID NO: 16. En otras realizaciones, la secuencia de nucleótidos optimizada por codón que codifica un polipéptido FVIII contiene como máximo 6, como máximo 5, como máximo 4, como máximo 3, o como máximo 2 secuencias MARS/ARS. En otras realizaciones, la secuencia de nucleótidos optimizada por codón que codifica un polipéptido FVIII contiene como máximo 1 secuencia MARS/ARS. En otras realizaciones más, la secuencia de nucleótidos optimizada por codón que codifica un polipéptido FVIII no contiene una secuencia MARS/ARS.

**[00215]** Se han identificado elementos ricos en AT en la secuencia de nucleótidos del FVIII humana que comparten similitud de secuencia con las secuencias de replicación autónoma de *Saccharomyces cerevisiae* (ARS) y las regiones de unión a la matriz nuclear (MAR). (Fallux et al., Mol. Cell. Biol. 16:4264-4272 (1996). Se ha demostrado que uno de estos elementos se une a factores nucleares in vitro y reprime la expresión de un gen informador de cloranfenicol

acetiltransferasa (CAT). Id. Se ha planteado la hipótesis de que estas secuencias pueden contribuir a la represión transcripcional del gen FVIII humano. Por lo tanto, en una realización, todas las secuencias MAR/ARS se eliminan en la secuencia de nucleótidos optimizada por codón que codifica un polipéptido FVIII de la presente descripción. Hay cuatro secuencias MAR/ARS ATATTT (SEQ ID NO: 21) y tres secuencias MAR/ARS AAATAT (SEQ ID NO: 22) en la secuencia progenitora FVIII (SEQ ID NO: 16). Todos estos sitios fueron mutados para destruir las secuencias MAR/ARS en las secuencias optimizadas de FVIII (SEQ ID NOs: 1-6). La ubicación de cada uno de estos elementos y la secuencia de nucleótidos correspondientes en las secuencias optimizadas se muestran en la Tabla 3, a continuación.

Tabla 3: Resumen de cambios en elementos represivos

| Localiza-<br>ción del<br>elemento      | Secuencia<br>de partida<br>BDD FVIII<br>(SEQ ID<br>NO: 16) | Secuencia BDD FVIII optimizada |                 |                 |                 |                 |                 |                  |                  |
|----------------------------------------|------------------------------------------------------------|--------------------------------|-----------------|-----------------|-----------------|-----------------|-----------------|------------------|------------------|
|                                        |                                                            | SEQ ID<br>NO: 1                | SEQ ID<br>NO: 2 | SEQ ID<br>NO: 3 | SEQ ID<br>NO: 4 | SEQ ID<br>NO: 5 | SEQ ID<br>NO: 6 | SEQ ID<br>NO: 70 | SEQ ID<br>NO: 71 |
| Secuencias desestabilizantes           |                                                            |                                |                 |                 |                 |                 |                 |                  |                  |
| 639                                    | ATTTA                                                      | GTTTA                          | GTTCA           | GTTCA           | GTTCA           | GTTCA           | GTTCA           | GTTCA            | GTTCA            |
| 1338                                   | ATTTA                                                      | GTTTA                          | GTTCA           | CTTCA           | GTTCA           | GTTCA           | GTTCA           | CTTCA            | GTTCA            |
| 1449                                   | ATTTA                                                      | CTTTA                          | CTTCA           | CTTCA           | CTTCA           | CTTCA           | CTTCA           | CTTCA            | CTTCA            |
| 1590                                   | TAAAT                                                      | TAAAT                          | CAAGT           | CAAGT           | TAAGT           | CAAGT           | CAAGT           | CAAGT            | TAAGT            |
| 1623                                   | TAAAT                                                      | CAAAA                          | GAAGA           | CTAAG           | CAAGA           | CAAGA           | CAAGA           | TAAGT            | CAAGA            |
| 2410                                   | ATTTA                                                      | ATCTA                          | ATCTA           | ATCTA           | ATCTA           | ATCTA           | ATCTA           | ATCTA            | ATCTA            |
| 2586                                   | ATTTA                                                      | GTTTA                          | GTTCA           | GTTCA           | GTTCA           | GTTCA           | GTTCA           | GTTCA            | GTTCA            |
| 2630                                   | TAAAT                                                      | TGAAT                          | TGAAC           | TGAAC           | TGAAC           | TCAAT           | TGAAC           | TCAAT            | TGAAC            |
| 3884                                   | ATTTA                                                      | ATCTG                          | ACCTG           | ACCTG           | ACCTG           | ATCTG           | ACCTG           | ATCTG            | ACCTG            |
| 3887                                   | TAAAT                                                      | TGAAC                          | TGAAC           | TGAAC           | TGAAC           | TGAAC           | TGAAC           | TGAAC            | TGAAC            |
| Sitios de unión del promotor potencial |                                                            |                                |                 |                 |                 |                 |                 |                  |                  |
| 641                                    | TTATA                                                      | TTATC                          | TCATC           | TCATT           | TCATC           | TCATC           | TCATC           | TCATT            | TCATC            |
| 1275                                   | TATAA                                                      | CTATA                          | TTACA           | CTACA           | GTACA           | CTACA           | CTACA           | CTACA            | GTACA            |
| 1276                                   | TTATA                                                      | TATAA                          | TACAA           | TACAA           | TACAA           | TACAA           | TACAA           | TACAA            | TACAA            |
| 1445                                   | TTATA                                                      | TCATC                          | TCATC           | TTATC           | TCATC           | TCATC           | TCATC           | TTATC            | TCATC            |
| 1474                                   | TATAA                                                      | TATAA                          | TACAA           | TACAA           | TACAA           | TACAA           | TACAA           | TACAA            | TACAA            |
| 1588                                   | TATAA                                                      | TATAA                          | TACAA           | TACAA           | TATAA           | TACAA           | TACAA           | TACAA            | TATAA            |
| 2614                                   | TTATA                                                      | CTGTA                          | CTGTA           | CTGTA           | CTGTA           | TTGTA           | CTGTA           | TTGTA            | CTGTA            |
| 2661                                   | TATAA                                                      | CATCA                          | CATCA           | CATCA           | CATCA           | CATCA           | CATCC           | CATCA            | CATCC            |

|                                                     |          |          |          |          |          |          |          |          |          |
|-----------------------------------------------------|----------|----------|----------|----------|----------|----------|----------|----------|----------|
| 3286                                                | TATAA    | TATAA    | TACAA    | TACAA    | TACAA    | TACAA    | TACAA    | TACAA    | TACAA    |
| 3840                                                | TTATA    | TTATA    | TTACT    | CTACA    | CTACA    | CTACA    | CTACT    | CTACA    | CTACT    |
| <b>Secuencias tipo unión a la matriz (MARS/ARS)</b> |          |          |          |          |          |          |          |          |          |
| 1287                                                | ATATTT   | GTATCT   | GTACCT   | GTACCT   | GTATCT   | GTACCT   | GTACCT   | GTACCT   | GTATCT   |
| 1447                                                | ATATTT   | ATCTTT   | ATCTTC   | ATCTTC   | ATCTTC   | ATCTTC   | ATCTTC   | ATCTTC   | ATCTTC   |
| 1577                                                | AAATAT   | AAATCT   | AGATCT   | AAATCT   | AAATCT   | AGATCT   | AGATCT   | AAATCT   | AAATCT   |
| 1585                                                | AAATAT   | AAGTAT   | AAGTAC   | AAGTAC   | AAGTAT   | AAGTAC   | AAGTAC   | AAGTAC   | AAGTAT   |
| 2231                                                | ATATTT   | ACATCA   | ATATCA   | ACATCA   | ACATCA   | ACATCT   | ATATCT   | ACATCT   | ATATCT   |
| 3054                                                | AAATAT   | AAACAT   | GAATAT   | GAACAT   | GAACAT   | GAACAT   | GAATAT   | GAACAT   | GAATAT   |
| 3788                                                | ATATTT   | ATATCT   | ATATCT   | ACATCT   | ACATCT   | ACATCT   | ACATCT   | ACATCT   | ACATCT   |
| <b>Elementos de secuencia ricos en AU (AREs)</b>    |          |          |          |          |          |          |          |          |          |
| 2468                                                | ATTTTATT | ACTTCATC | ACTTCATC | ACTTCATT | ACTTCATT | ACTTTATT | ACTTTATC | ACTTTATT | ACTTTATC |
| 3790                                                | ATTTTTAA | ATCTTTAA | ATCTTCAA | ATCTTCAA | ATCTTCAA | ATCTTCAA | ATCTTCAA | ATCTTCAA | ATCTTCAA |
| <b>Secuencias Poly A/Poly T</b>                     |          |          |          |          |          |          |          |          |          |
| 3273                                                | AAAAAAA  | GAAAAAA  | GAAGAAG  | GAAGAAG  | GAAGAAG  | GAAGAAG  | CAAGAAG  | GAAGAAG  | CAAGAAG  |
| 4195                                                | TTTTTT   | TTCTTT   | TTCTTC   | TTCTTC   | TTCTTC   | TTCTTC   | TTCTTC   | TTCTTC   | TTCTTC   |
| <b>Sitios de división</b>                           |          |          |          |          |          |          |          |          |          |
| 2203                                                | GGTGAT   | GGGGAC   | GGCGAC   | GGGGAC   | GGGGAC   | GGAGAC   | GGAGAC   | GGAGAC   | GGAGAC   |

#### D. Secuencias desestabilizadoras

- [00216]** En algunas realizaciones, el casete génico comprende una secuencia de nucleótidos optimizada con codón que codifica un polipéptido FVIII, en donde la secuencia de nucleótidos optimizada con codón contiene menos elementos desestabilizadores con respecto a la SEQ ID NO: 16. En otras realizaciones, la secuencia de nucleótidos optimizada por codón que codifica un polipéptido FVIII contiene como máximo 9, como máximo 8, como máximo 7, como máximo 6 o como máximo 5 elementos desestabilizadores. En otras realizaciones, la secuencia de nucleótidos optimizada por codón que codifica un polipéptido FVIII contiene como máximo 4, como máximo 3, como máximo 2, o como máximo 1 elementos desestabilizadores. En otras realizaciones más, la secuencia de nucleótidos optimizada por codón que codifica un polipéptido FVIII no contiene ningún elemento desestabilizador.
- [00217]** En una realización particular, la secuencia de nucleótidos optimizada por codón que codifica un polipéptido FVIII comprende una primera secuencia de ácido nucleico que codifica una porción N-terminal de un polipéptido FVIII y una segunda secuencia de ácido nucleico que codifica una porción C-terminal de un polipéptido FVIII; en donde la primera secuencia de ácido nucleico tiene al menos

aproximadamente 80%, al menos aproximadamente 85%, al menos aproximadamente 86%, al menos aproximadamente 87%, al menos aproximadamente 88%, al menos aproximadamente 89%, al menos aproximadamente 90%, al menos aproximadamente 91%, al menos aproximadamente 92%, al menos aproximadamente 93%, al menos aproximadamente 94%, al menos aproximadamente 95%, al menos aproximadamente 96%, al menos aproximadamente 97%, al menos aproximadamente 98%, o al menos aproximadamente 99% de identidad de secuencia con (i) los nucleótidos 58-1791 de la SEQ ID N°: 3; (ii) los nucleótidos 1-1791 de SEQ ID NO: 3; (iii) los nucleótidos 58-1791 de SEQ ID NO: 4; o (iv) los nucleótidos 1-1791 de SEQ ID NO: 4; en donde la porción N-terminal y la porción C-terminal juntas tienen una actividad de polipéptido FVIII; y en donde la secuencia de nucleótidos optimizada con codón contiene menos elementos desestabilizadores con respecto a la SEQ ID NO: 16. En otras realizaciones, la secuencia de nucleótidos que codifica un polipéptido con actividad de FVIII contiene como máximo 9, como máximo 8, como máximo 7, como máximo 6 o como máximo 5 elementos desestabilizadores. En otras realizaciones, la secuencia de nucleótidos optimizada por codón que codifica un polipéptido FVIII contiene como máximo 4, como máximo 3, como máximo 2, o como máximo 1 elemento desestabilizador. En otras realizaciones más, la secuencia de nucleótidos optimizada por codón que codifica un polipéptido FVIII no contiene ningún elemento desestabilizador.

**[00218]** En otra realización, la secuencia de nucleótidos optimizada por codón que codifica un polipéptido FVIII comprende una primera secuencia de ácido nucleico que codifica una porción N-terminal de un polipéptido FVIII y una segunda secuencia de ácido nucleico que codifica una porción C-terminal de un polipéptido FVIII; en donde la segunda secuencia de ácido nucleico tiene al menos aproximadamente 80%, al menos aproximadamente 85%, al menos aproximadamente 86%, al menos aproximadamente 87%, al menos aproximadamente 88%, al menos aproximadamente 89%, al menos aproximadamente 90%, al menos aproximadamente 91%, al menos aproximadamente 92%, al menos aproximadamente 93%, al menos aproximadamente 94%, al menos aproximadamente 95%, al menos aproximadamente 96%, al menos aproximadamente 97%, al menos aproximadamente 98%, o al menos aproximadamente el 99% de identidad de secuencia con (i) los nucleótidos 1792-4374 de SEQ ID NO: 5; (ii) los nucleótidos 1792-4374 de SEQ ID NO: 6; (iii) los nucleótidos 1792-2277 y 2320-4374 de SEQ ID NO: 5 (es decir, los nucleótidos 1792-4374 de SEQ ID NO: 5 sin los nucleótidos que codifican el dominio B o el fragmento de dominio

B); o (iv) los nucleótidos 1792-2277 y 2320-4374 de la SEQ ID NO: 6 (es decir, los nucleótidos 1792-4374 de la SEQ ID NO: 6 sin los nucleótidos que codifican el dominio B o el fragmento del dominio B); en donde la porción N-terminal y la porción C-terminal juntas tienen una actividad de polipéptido FVIII; y en donde la secuencia de nucleótidos optimizada con codón contiene menos elementos desestabilizadores con respecto a la SEQ ID NO: 16. En otras realizaciones, la secuencia de nucleótidos que codifica un polipéptido con actividad de FVIII contiene como máximo 9, como máximo 8, como máximo 7, como máximo 6 o como máximo 5 elementos desestabilizadores. En otras realizaciones, la secuencia de nucleótidos optimizada por codón que codifica un polipéptido FVIII contiene como máximo 4, como máximo 3, como máximo 2, o como máximo 1 elemento desestabilizador. En otras realizaciones más, la secuencia de nucleótidos optimizada por codón que codifica un polipéptido FVIII no contiene ningún elemento desestabilizador.

**[00219]** En otras realizaciones, el casete génico comprende una secuencia de nucleótidos con codón optimizado que codifica un polipéptido FVIII, en donde la secuencia de nucleótidos con codón optimizado comprende una secuencia de ácido nucleico que tiene al menos aproximadamente 80%, al menos aproximadamente 85%, al menos aproximadamente 89%, al menos aproximadamente 90%, al menos aproximadamente 91%, al menos aproximadamente 92%, al menos aproximadamente 93%, al menos aproximadamente 94%, al menos aproximadamente 95%, al menos aproximadamente 96%, al menos aproximadamente 97%, al menos aproximadamente 98%, o al menos aproximadamente 99% de identidad de secuencia con (i) los nucleótidos 58-4374 de una secuencia de aminoácidos seleccionada de SEQ ID NO: 1, 2, 3, 4, 5, 6, 70 y 71 o (ii) los nucleótidos 58-2277 y 2320-4374 de una secuencia de aminoácidos seleccionada de SEQ ID NO: 1, 2, 3, 4, 5, 6, 70 y 71 (es decir, los nucleótidos 58-4374 de SEQ ID NO: 1, 2, 3, 4, 5, 6, 70 o 71 sin los nucleótidos que codifican el dominio B o el fragmento del dominio B); y en donde la secuencia de nucleótidos optimizada con codón contiene menos elementos desestabilizadores con respecto a la SEQ ID NO: 16. En otras realizaciones, la secuencia de nucleótidos optimizada por codón que codifica un polipéptido FVIII contiene como máximo 9, como máximo 8, como máximo 7, como máximo 6 o como máximo 5 elementos desestabilizadores. En otras realizaciones, la secuencia de nucleótidos optimizada por codón que codifica un polipéptido FVIII contiene como máximo 4, como máximo 3, como máximo 2, o como máximo 1 elemento desestabilizador. En otras realizaciones

más, la secuencia de nucleótidos optimizada por codón que codifica un polipéptido FVIII no contiene ningún elemento desestabilizador.

**[00220]** Hay diez elementos desestabilizadores en la secuencia progenitora del FVIII (SEQ ID NO: 16): seis secuencias ATTTA (SEQ ID NO: 23) y cuatro secuencias TAAAT (SEQ ID NO: 24). En una realización, las secuencias de estos sitios se mutaron para destruir los elementos desestabilizadores en las SEQ ID NOs de FVIII optimizadas: 1-6, 70 y 71. La ubicación de cada uno de estos elementos y la secuencia de los nucleótidos correspondientes en las secuencias optimizadas se muestran en la Tabla 3.

## **E. Posibles sitios de unión de promotores**

**[00221]** En algunas realizaciones, el casete génico comprende una secuencia de nucleótidos optimizada por codón que codifica un polipéptido FVIII, en donde la secuencia de nucleótidos contiene menos sitios potenciales de unión al promotor en relación con la SEQ ID NO: 16. En otras realizaciones, la secuencia de nucleótidos optimizada por codón que codifica un polipéptido FVIII contiene como máximo 9, como máximo 8, como máximo 7, como máximo 6 o como máximo 5 sitios de unión a promotores potenciales. En otras realizaciones, la secuencia de nucleótidos optimizada por codón que codifica un polipéptido FVIII contiene como máximo 4, como máximo 3, como máximo 2, o como máximo 1 sitio de unión a promotores potenciales. En otras realizaciones más, la secuencia de nucleótidos optimizada por codón que codifica un polipéptido FVIII no contiene ningún sitio potencial de unión al promotor.

**[00222]** En una realización particular, la secuencia de nucleótidos optimizada por codón que codifica un polipéptido FVIII comprende una primera secuencia de ácido nucleico que codifica una porción N-terminal de un polipéptido FVIII y una segunda secuencia de ácido nucleico que codifica una porción C-terminal de un polipéptido FVIII; en donde la primera secuencia de ácido nucleico tiene al menos aproximadamente 80%, al menos aproximadamente 85%, al menos aproximadamente 86%, al menos aproximadamente 87%, al menos aproximadamente 88%, al menos aproximadamente 89%, al menos aproximadamente 90%, al menos aproximadamente 91%, al menos aproximadamente 92%, al menos aproximadamente 93%, al menos aproximadamente 94%, al menos aproximadamente 95%, al menos aproximadamente 96%, al menos aproximadamente 97%, al menos aproximadamente 98% o al menos aproximadamente 99% de identidad de secuencia con (i) los nucleótidos 58-1791 de la SEQ ID N°: 3; (ii) los nucleótidos 1-1791 de SEQ ID NO: 3; (iii) los nucleótidos 58-1791 de SEQ ID NO: 4; o (iv) los nucleótidos 1-1791 de SEQ ID NO: 4; en donde la

porción N-terminal y la porción C-terminal juntas tienen una actividad de polipéptido FVIII; y en donde la secuencia de nucleótidos optimizada con codón contiene menos sitios potenciales de unión al promotor con relación a la SEQ ID NO: 16. En otras realizaciones, la secuencia de nucleótidos optimizada por codón que codifica un polipéptido FVIII contiene como máximo 9, como máximo 8, como máximo 7, como máximo 6 o como máximo 5 sitios de unión a promotores potenciales. En otras realizaciones, la secuencia de nucleótidos optimizada por codón que codifica un polipéptido FVIII contiene como máximo 4, como máximo 3, como máximo 2, o como máximo 1 sitio de unión a promotores potenciales. En otras realizaciones más, la secuencia de nucleótidos optimizada por codón que codifica un polipéptido FVIII no contiene ningún sitio potencial de unión al promotor.

**[00223]** En otra realización, la secuencia de nucleótidos optimizada por codón que codifica un polipéptido FVIII comprende una primera secuencia de ácido nucleico que codifica una porción N-terminal de un polipéptido FVIII y una segunda secuencia de ácido nucleico que codifica una porción C-terminal de un polipéptido FVIII; en donde la segunda secuencia de ácido nucleico tiene al menos aproximadamente 80%, al menos aproximadamente 85%, al menos aproximadamente 86%, al menos aproximadamente 87%, al menos aproximadamente 88%, al menos aproximadamente 89%, al menos aproximadamente 90%, al menos aproximadamente 91%, al menos aproximadamente 92%, al menos aproximadamente 93%, al menos aproximadamente 94%, al menos aproximadamente 95%, al menos aproximadamente 96%, al menos aproximadamente 97%, al menos aproximadamente 98%, o al menos aproximadamente 99% de identidad de secuencia con (i) los nucleótidos 1792-4374 de la SEQ ID N°: 5; (ii) los nucleótidos 1792-4374 de SEQ ID NO: 6; (iii) los nucleótidos 1792-2277 y 2320-4374 de SEQ ID NO: 5 (es decir, los nucleótidos 1792-4374 de SEQ ID NO: 5 sin los nucleótidos que codifican el dominio B o el fragmento de dominio B); o (iv) los nucleótidos 1792-2277 y 2320-4374 de la SEQ ID NO: 6 (es decir, los nucleótidos 1792-4374 de la SEQ ID NO: 6 sin los nucleótidos que codifican el dominio B o el fragmento del dominio B); en donde la porción N-terminal y la porción C-terminal juntas tienen una actividad de polipéptido FVIII; y en donde la secuencia de nucleótidos optimizada con codón contiene menos sitios potenciales de unión al promotor con relación a la SEQ ID NO: 16. En otras realizaciones, la secuencia de nucleótidos optimizada por codón que codifica un polipéptido FVIII contiene como máximo 9, como máximo 8, como máximo 7, como máximo 6 o como máximo 5 sitios de unión a promotores potenciales. En otras realizaciones, la secuencia de

nucleótidos optimizada por codón que codifica un polipéptido FVIII contiene como máximo 4, como máximo 3, como máximo 2, o como máximo 1 sitio de unión a promotores potenciales. En otras realizaciones más, la secuencia de nucleótidos optimizada por codón que codifica un polipéptido FVIII no contiene ningún sitio potencial de unión al promotor.

**[00224]** En otras realizaciones, el casete génico comprende una secuencia de nucleótidos optimizada por codón que codifica un polipéptido FVIII, en donde la secuencia de nucleótidos comprende una secuencia de ácido nucleico que tiene al menos aproximadamente 80%, al menos aproximadamente 85%, al menos aproximadamente 89%, al menos aproximadamente 90%, al menos aproximadamente 91%, al menos aproximadamente 92%, al menos aproximadamente 93%, al menos aproximadamente 94%, al menos aproximadamente 95%, al menos aproximadamente 96%, al menos aproximadamente 97%, al menos aproximadamente 98%, o al menos aproximadamente 99% de identidad de secuencia para (i) los nucleótidos 58-4374 de una secuencia de aminoácidos seleccionada de SEQ ID NO: 1, 2, 3, 4, 5, 6, 70 y 71 o (ii) los nucleótidos 58-2277 y 2320-4374 de una secuencia de aminoácidos seleccionada de SEQ ID NO: 1, 2, 3, 4, 5, 6, 70 y 71 (es decir, los nucleótidos 58-4374 de SEQ ID NO: 1, 2, 3, 4, 5, 6, 70 o 71 sin los nucleótidos que codifican el dominio B o el fragmento del dominio B); y en donde la secuencia de nucleótidos optimizada con codón contiene menos sitios potenciales de unión al promotor con relación a la SEQ ID NO: 16. En otras realizaciones, la secuencia de nucleótidos optimizada por codón que codifica un polipéptido FVIII contiene como máximo 9, como máximo 8, como máximo 7, como máximo 6 o como máximo 5 sitios de unión de promotores potenciales. En otras realizaciones, la secuencia de nucleótidos optimizada por codón que codifica un polipéptido FVIII contiene como máximo 4, como máximo 3, como máximo 2, o como máximo 1 sitio de unión de promotores potenciales. En otras realizaciones más, la secuencia de nucleótidos optimizada por codón que codifica un polipéptido FVIII no contiene ningún sitio potencial de unión al promotor.

**[00225]** Las cajas TATA son secuencias reguladoras que a menudo se encuentran en las regiones promotoras de las células eucariotas. Sirven como el sitio de unión de la proteína de unión TATA (TBP), un factor de transcripción general. Las cajas TATA generalmente comprenden la secuencia TATAA (SEQ ID NO: 28) o una variante cercana. Sin embargo, las cajas TATA dentro de una secuencia de codificación pueden inhibir la traducción de la proteína de longitud completa. Hay diez secuencias potenciales de unión al promotor en la secuencia FVIII BDD natural (SEQ

ID NO: 16): cinco secuencias TATAA (SEQ ID NO: 28) y cinco secuencias TTATA (SEQ ID NO: 29). En algunas realizaciones, al menos 1, al menos 2, al menos 3, o al menos 4 de los sitios de unión al promotor están eliminados en los genes FVIII de la presente descripción. En algunas realizaciones, al menos 5 de los sitios de unión al promotor están eliminados en los genes FVIII de la presente descripción. En otras realizaciones, al menos 6, al menos 7, o al menos 8 de los sitios de unión al promotor están eliminados en los genes FVIII de la presente descripción. En una realización, al menos 9 de los sitios de unión al promotor están eliminados en los genes FVIII de la presente descripción. En una realización particular, todos los sitios de unión al promotor se eliminan en los genes FVIII de la presente descripción. La ubicación de cada sitio potencial de unión al promotor y la secuencia de los nucleótidos correspondientes en las secuencias optimizadas se muestran en la Tabla 3.

#### **F. Otros elementos reguladores negativos que actúan en Cis**

**[00226]** Además de las secuencias MAR/ARS, los elementos desestabilizadores y los posibles sitios promotores descritos anteriormente, se pueden identificar varias secuencias potencialmente inhibitorias adicionales en la secuencia FVIII BDD natural (SEQ ID NO: 16). Se pueden identificar dos elementos de secuencia rica en AU (ARE) (ATTTTATT (SEQ ID NO: 30); y ATTTTTAA (SEQ ID NO: 31), junto con un sitio poli-A (AAAAAAA; SEQ ID NO: 26), un sitio poli-T (TTTTTT; SEQ ID NO: 25), y un sitio de empalme (GGTGAT; SEQ ID NO: 27) en la secuencia FVIII BDD no optimizada. Uno o más de estos elementos se pueden eliminar de las secuencias de FVIII optimizadas. La ubicación de cada uno de estos sitios y la secuencia de los nucleótidos correspondientes en las secuencias optimizadas se muestran en la Tabla 3.

**[00227]** En ciertas realizaciones, la secuencia de nucleótidos optimizada por codón que codifica un polipéptido FVIII comprende una primera secuencia de ácido nucleico que codifica una porción N-terminal de un polipéptido FVIII y una segunda secuencia de ácido nucleico que codifica una porción C-terminal de un polipéptido FVIII; en donde la primera secuencia de ácido nucleico tiene al menos aproximadamente 80%, al menos aproximadamente 85%, al menos aproximadamente 86%, al menos aproximadamente 87%, al menos aproximadamente 88%, al menos aproximadamente 89%, al menos aproximadamente 90%, al menos aproximadamente 91%, al menos aproximadamente 92%, al menos aproximadamente 93%, al menos aproximadamente 94%, al menos aproximadamente 95%, al menos aproximadamente 96%, al menos aproximadamente 97%, al menos aproximadamente 98%, o al menos aproximadamente 99% de identidad de secuencia con (i) los nucleótidos 58-1791 de

la SEQ ID N°: 3; (ii) los nucleótidos 1-1791 de SEQ ID NO: 3; (iii) los nucleótidos 58-1791 de SEQ ID NO: 4; o (iv) los nucleótidos 1-1791 de SEQ ID NO: 4; en donde la porción N-terminal y la porción C-terminal juntas tienen una actividad de polipéptido FVIII; y en donde la secuencia de nucleótidos optimizada por codón no contiene ni uno ni otros más elementos reguladores negativos que actúen en cis, por ejemplo, un sitio de empalme, una secuencia poli-T, una secuencia poli-A, una secuencia ARE, o cualquiera de sus combinaciones.

**[00228]** En otra realización, la secuencia de nucleótidos optimizada por codón que codifica un polipéptido FVIII comprende una primera secuencia de ácido nucleico que codifica una porción N-terminal de un polipéptido FVIII y una segunda secuencia de ácido nucleico que codifica una porción C-terminal de un polipéptido FVIII; en donde la segunda secuencia de ácido nucleico tiene al menos aproximadamente 80%, al menos aproximadamente 85%, al menos aproximadamente 86%, al menos aproximadamente 87%, al menos aproximadamente 88%, al menos aproximadamente 89%, al menos aproximadamente 90%, al menos aproximadamente 91%, al menos aproximadamente 92%, al menos aproximadamente 93%, al menos aproximadamente 94%, al menos aproximadamente 95%, al menos aproximadamente 96%, al menos aproximadamente 97%, al menos aproximadamente 98%, o al menos aproximadamente 99% de identidad de secuencia con (i) los nucleótidos 1792-4374 de la SEQ ID N°: 5; (ii) los nucleótidos 1792-4374 de SEQ ID NO: 6; (iii) los nucleótidos 1792-2277 y 2320-4374 de SEQ ID NO: 5 (es decir, los nucleótidos 1792-4374 de SEQ ID NO: 5 sin los nucleótidos que codifican el dominio B o el fragmento de dominio B); o (iv) los nucleótidos 1792-2277 y 2320-4374 de SEQ ID NO: 6 (es decir, los nucleótidos 1792-4374 de SEQ ID NO: 6 sin los nucleótidos que codifican el dominio B o el fragmento de dominio B); en donde la porción N-terminal y la porción C-terminal juntas tienen una actividad de polipéptido FVIII; y en donde la secuencia de nucleótidos optimizada por codón no contiene ni uno ni otros más elementos reguladores negativos que actúen en cis, por ejemplo, un sitio de empalme, una secuencia poli-T, una secuencia poli-A, una secuencia ARE, o cualquiera de sus combinaciones.

**[00229]** En otras realizaciones, el casete génico comprende una secuencia de nucleótidos optimizada por codón que codifica un polipéptido FVIII, en donde la secuencia de nucleótidos comprende una secuencia de ácido nucleico que tiene al menos aproximadamente 80%, al menos aproximadamente 85%, al menos aproximadamente 89%, al menos aproximadamente 90%, al menos aproximadamente

91%, al menos aproximadamente 92%, al menos aproximadamente 93%, al menos aproximadamente 94%, al menos aproximadamente 95%, al menos aproximadamente 96%, al menos aproximadamente 97%, al menos aproximadamente 98%, o al menos aproximadamente 99% de identidad de secuencia para (i) los nucleótidos 58-4374 de una secuencia de aminoácidos seleccionada de SEQ ID NO: 1, 2, 3, 4, 5, 6, 70 y 71 o (ii) los nucleótidos 58-2277 y 2320-4374 de una secuencia de aminoácidos seleccionada de SEQ ID NO: 1, 2, 3, 4, 5, 6, 70 y 71 (es decir, los nucleótidos 58-4374 de SEQ ID NO: 1, 2, 3, 4, 5, 6, 70 o 71 sin los nucleótidos que codifican el dominio B o el fragmento del dominio B); y en donde la secuencia de nucleótidos optimizada por codón no contiene ni uno ni otros más elementos reguladores negativos que actúen en cis, por ejemplo, un sitio de empalme, una secuencia poli-T, una secuencia poli-A, una secuencia ARE, o cualquiera de sus combinaciones.

**[00230]** En algunas realizaciones, la secuencia de nucleótidos optimizada por codón que codifica un polipéptido FVIII comprende una primera secuencia de ácido nucleico que codifica una porción N-terminal de un polipéptido FVIII y una segunda secuencia de ácido nucleico que codifica una porción C-terminal de un polipéptido FVIII; en donde la primera secuencia de ácido nucleico tiene al menos aproximadamente 80%, al menos aproximadamente 85%, al menos aproximadamente 86%, al menos aproximadamente 87%, al menos aproximadamente 88%, al menos aproximadamente 89%, al menos aproximadamente 90%, al menos aproximadamente 91%, al menos aproximadamente 92%, al menos aproximadamente 93%, al menos aproximadamente 94%, al menos aproximadamente 95%, al menos aproximadamente 96%, al menos aproximadamente 97%, al menos aproximadamente 98% o al menos aproximadamente 99% de identidad de secuencia con (i) los nucleótidos 58-1791 de la SEQ ID N°: 3; (ii) los nucleótidos 1-1791 de la SEQ ID N°: 3; (iii) los nucleótidos 58-1791 de SEQ ID NO: 4; o (iv) los nucleótidos 1-1791 de SEQ ID NO: 4; en donde la porción N-terminal y la porción C-terminal juntas tienen una actividad de polipéptido FVIII; y en donde la secuencia de nucleótidos optimizada por codón no contiene el sitio de empalme GGTGAT (SEQ ID N°: 27). En algunas realizaciones, la secuencia de nucleótidos optimizada por codón que codifica un polipéptido FVIII comprende una primera secuencia de ácido nucleico que codifica una porción N-terminal de un polipéptido FVIII y una segunda secuencia de ácido nucleico que codifica una porción C-terminal de un polipéptido FVIII; en donde la primera secuencia de ácido nucleico tiene al menos aproximadamente 80%, al menos aproximadamente 85%, al menos aproximadamente 86%, al menos aproximadamente 87%, al menos aproximadamente 88%, al menos aproximadamente 89%, al menos aproximadamente 90%, al menos aproximadamente 91%, al menos aproximadamente 92%, al menos aproximadamente 93%, al menos aproximadamente 94%, al menos aproximadamente 95%, al menos aproximadamente 96%, al menos aproximadamente 97%, al menos aproximadamente 98% o al menos aproximadamente 99% de identidad de secuencia con (i) los nucleótidos 58-1791 de la SEQ ID N°: 3; (ii) los nucleótidos 1-1791 de la SEQ ID N°: 3; (iii) los nucleótidos 58-1791 de SEQ ID NO: 4; o (iv) los nucleótidos 1-1791 de SEQ ID NO: 4; en donde la porción N-terminal y la porción C-terminal juntas tienen una actividad de polipéptido FVIII; y en donde la secuencia de nucleótidos optimizada por codón no contiene el sitio de empalme GGTGAT (SEQ ID N°: 27).

88%, al menos aproximadamente 89%, al menos aproximadamente 90%, al menos  
aproximadamente 91%, al menos aproximadamente 92%, al menos aproximadamente  
93%, al menos aproximadamente 94%, al menos aproximadamente 95%, al menos  
aproximadamente 96%, al menos aproximadamente 97%, al menos aproximadamente  
5 98% o al menos aproximadamente 99% de identidad de secuencia con (i) los  
nucleótidos 58-1791 de la SEQ ID N°: 3; (ii) los nucleótidos 1-1791 de SEQ ID NO: 3;  
(iii) los nucleótidos 58-1791 de SEQ ID NO: 4; o (iv) los nucleótidos 1-1791 de SEQ ID  
NO: 4; en donde la porción N-terminal y la porción C-terminal juntas tienen una  
actividad de polipéptido FVIII; y en donde la secuencia de nucleótidos optimizada por  
10 codón no contiene ninguna secuencia de poli-T (SEQ ID NO: 25). En algunas  
realizaciones, la secuencia de nucleótidos optimizada por codón que codifica un  
polipéptido FVIII comprende una primera secuencia de ácido nucleico que codifica una  
porción N-terminal de un polipéptido FVIII y una segunda secuencia de ácido nucleico  
que codifica una porción C-terminal de un polipéptido FVIII; en donde la primera  
15 secuencia de ácido nucleico tiene al menos aproximadamente 80%, al menos  
aproximadamente 85%, al menos aproximadamente 86%, al menos aproximadamente  
87%, al menos aproximadamente 88%, al menos aproximadamente 89%, al menos  
aproximadamente 90%, al menos aproximadamente 91%, al menos aproximadamente  
92%, al menos aproximadamente 93%, al menos aproximadamente 94%, al menos  
20 aproximadamente 95%, al menos aproximadamente 96%, al menos aproximadamente  
97%, al menos aproximadamente 98% o al menos aproximadamente 99% de  
identidad de secuencia con (i) los nucleótidos 58-1791 de la SEQ ID N°: 3; (ii) los  
nucleótidos 1-1791 de SEQ ID NO: 3; (iii) los nucleótidos 58-1791 de SEQ ID NO: 4;  
o (iv) los nucleótidos 1-1791 de SEQ ID NO: 4; en donde la porción N-terminal y la  
25 porción C-terminal juntas tienen una actividad de polipéptido FVIII; y en donde la  
secuencia de nucleótidos optimizada por codón no contiene ninguna secuencia de  
poli-A (SEQ ID NO: 26). En algunas realizaciones, la secuencia de nucleótidos  
optimizada por codón que codifica un polipéptido FVIII comprende una primera  
secuencia de ácido nucleico que codifica una porción N-terminal de un polipéptido  
30 FVIII y una segunda secuencia de ácido nucleico que codifica una porción C-terminal  
de un polipéptido FVIII; en donde la primera secuencia de ácido nucleico tiene al  
menos aproximadamente 80%, al menos aproximadamente 85%, al menos  
aproximadamente 86%, al menos aproximadamente 87%, al menos aproximadamente  
88%, al menos aproximadamente 89%, al menos aproximadamente 90%, al menos  
35 aproximadamente 91%, al menos aproximadamente 92%, al menos aproximadamente

93%, al menos aproximadamente 94%, al menos aproximadamente 95%, al menos aproximadamente 96%, al menos aproximadamente 97%, al menos aproximadamente 98% o al menos aproximadamente 99% de identidad de secuencia con (i) los nucleótidos 58-1791 de la SEQ ID N°: 3; (ii) los nucleótidos 1-1791 de SEQ ID NO: 3; (iii) los nucleótidos 58-1791 de SEQ ID NO: 4; o (iv) los nucleótidos 1-1791 de SEQ ID NO: 4; en donde la porción N-terminal y la porción C-terminal juntas tienen una actividad de polipéptido FVIII; y en donde la secuencia de nucleótidos optimizada por codón no contiene ningún elemento ARE (SEQ ID NO: 30 o SEQ ID NO: 31).

**[00231]**

En algunas realizaciones, el casete génico comprende una secuencia de nucleótidos con codón optimizado que codifica un polipéptido FVIII, en donde la secuencia de nucleótidos con codón optimizado comprende una secuencia de ácido nucleico que tiene al menos aproximadamente 80%, al menos aproximadamente 85%, al menos aproximadamente 89%, al menos aproximadamente 90%, al menos aproximadamente 91%, al menos aproximadamente 92%, al menos aproximadamente 93%, al menos aproximadamente 94%, al menos aproximadamente 95%, al menos aproximadamente 96%, al menos aproximadamente 97%, al menos aproximadamente 98%, o al menos aproximadamente 99% de identidad de secuencia con (i) los nucleótidos 58-4374 de una secuencia de aminoácidos seleccionada de SEQ ID NO: 1, 2, 3, 4, 5, 6, 70 y 71 o (ii) los nucleótidos 58-2277 y 2320-4374 de una secuencia de aminoácidos seleccionada de SEQ ID NO: 1, 2, 3, 4, 5, 6, 70 y 71 (es decir, los nucleótidos 58-4374 de SEQ ID NO: 1, 2, 3, 4, 5, 6, 70 o 71 sin los nucleótidos que codifican el dominio B o el fragmento del dominio B); y en donde la secuencia de nucleótidos optimizada por codón no contiene el sitio de empalme GGTGAT (SEQ ID NO: 27). En algunas realizaciones, el casete génico comprende una secuencia de nucleótidos con codón optimizado que codifica un polipéptido FVIII, en donde la secuencia de nucleótidos con codón optimizado comprende una secuencia de ácido nucleico que tiene al menos aproximadamente 80%, al menos aproximadamente 85%, al menos aproximadamente 89%, al menos aproximadamente 90%, al menos aproximadamente 91%, al menos aproximadamente 92%, al menos aproximadamente 93%, al menos aproximadamente 94%, al menos aproximadamente 95%, al menos aproximadamente 96%, al menos aproximadamente 97%, al menos aproximadamente 98%, o al menos aproximadamente 99% de identidad de secuencia para (i) los nucleótidos 58- 4374 de una secuencia de aminoácidos seleccionada de SEQ ID NO: 1, 2, 3, 4, 5, 6, 70 y 71 o (ii) los nucleótidos 58-2277 y 2320-4374 de una secuencia de aminoácidos seleccionada de SEQ ID NO: 1, 2, 3, 4, 5, 6, 70 y 71 (es decir, los

nucleótidos 58-4374 de SEQ ID NO: 1, 2, 3, 4, 5, 6, 70 o 71 sin los nucleótidos que codifican el dominio B o el fragmento del dominio B); y en donde la secuencia de nucleótidos optimizada por codón no contiene ninguna secuencia de poli-T (SEQ ID NO: 25). En algunas realizaciones, el casete génico comprende una secuencia de nucleótidos con codón optimizado que codifica un polipéptido FVIII, en donde la secuencia de nucleótidos con codón optimizado comprende una secuencia de ácido nucleico que tiene al menos aproximadamente 80%, al menos aproximadamente 85%, al menos aproximadamente 89%, al menos aproximadamente 90%, al menos aproximadamente 91%, al menos aproximadamente 92%, al menos aproximadamente 93%, al menos aproximadamente 94%, al menos aproximadamente 95%, al menos aproximadamente 96%, al menos aproximadamente 97%, al menos aproximadamente 98%, o al menos aproximadamente 99% de identidad de secuencia para (i) los nucleótidos 58-4374 de una secuencia de aminoácidos seleccionada de SEQ ID NO: 1, 2, 3, 4, 5, 6, 70 y 71 o (ii) los nucleótidos 58-2277 y 2320-4374 de una secuencia de aminoácidos seleccionada de SEQ ID NO: 1, 2, 3, 4, 5, 6, 70 y 71 (es decir, los nucleótidos 58-4374 de SEQ ID NO: 1, 2, 3, 4, 5, 6, 70 o 71 sin los nucleótidos que codifican el dominio B o el fragmento del dominio B); y en donde la secuencia de nucleótidos optimizada por codón no contiene una secuencia de poli-A (SEQ ID NO: 26). En algunas realizaciones, el casete génico comprende una secuencia de nucleótidos con codón optimizado que codifica un polipéptido FVIII, en donde la secuencia de nucleótidos con codón optimizada comprende una secuencia de ácido nucleico que tiene al menos aproximadamente 80%, al menos aproximadamente 85%, al menos aproximadamente 89%, al menos aproximadamente 90%, al menos aproximadamente 91%, al menos aproximadamente 92%, al menos aproximadamente 93%, al menos aproximadamente 94%, al menos aproximadamente 95%, al menos aproximadamente 96%, al menos aproximadamente 97%, al menos aproximadamente 98%, o al menos aproximadamente 99% de identidad de secuencia para (i) los nucleótidos 58-4374 de una secuencia de aminoácidos seleccionada de SEQ ID NO: 1, 2, 3, 4, 5, 6, 70 y 71 o (ii) los nucleótidos 58-2277 y 2320-4374 de una secuencia de aminoácidos seleccionada de SEQ ID NO: 1, 2, 3, 4, 5, 6, 70 y 71 (es decir, los nucleótidos 58-4374 de SEQ ID NO: 1, 2, 3, 4, 5, 6, 70 o 71 sin los nucleótidos que codifican el dominio B o el fragmento del dominio B); y en donde la secuencia de nucleótidos optimizada por codón no contiene ningún elemento ARE (SEQ ID NO: 30 o SEQ ID NO: 31).

**[00232]** En otras realizaciones, una secuencia de FVIII optimizada de la descripción no comprende ni uno ni otros más motivos antivirales, estructuras de bucle y secuencias repetidas.

**[00233]** En otras realizaciones más, los nucleótidos que rodean el sitio de inicio de la transcripción se cambian a una secuencia consenso de kozak (GCCGCCACCATGC (SEQ ID NO: 32), en donde los nucleótidos subrayados son el codón de inicio). En otras realizaciones, se pueden agregar o eliminar sitios de restricción para facilitar el proceso de clonación.

#### **b. FIX y secuencias de polinucleótidos que codifican la proteína FIX**

**[00234]** En algunas realizaciones, la molécula de ácido nucleico comprende una primera ITR, una segunda ITR, y un casete génico que codifica una proteína terapéutica, en donde la proteína terapéutica comprende un polipéptido FIX. En algunas realizaciones, el polipéptido FIX comprende FIX o una variante o fragmento del mismo, en donde el FIX o la variante o fragmento del mismo tiene una actividad FIX.

**[00235]** El FIX humano es una serina proteasa que es un componente importante de la vía intrínseca de la cascada de coagulación sanguínea. El "factor IX" o "FIX", según se usa en el presente documento, se refiere a una proteína del factor de coagulación y especies y variantes de secuencia de la misma, e incluye, entre otros, la secuencia de aminoácidos monocatenaria 461 del polipéptido precursor FIX humano ("prepro"), la secuencia de aminoácidos 415 monocatenaria de FIX humano maduro (SEQ ID NO: 125) y la variante R338L FIX (Padua) (SEQ ID NO: 126). FIX incluye cualquier forma de molécula FIX con las características típicas de la coagulación sanguínea FIX. Como se usa en el presente documento, tanto "Factor IX" como "FIX" pretenden abarcar polipéptidos que comprenden los dominios Gla (región que contiene residuos de ácido  $\gamma$ -carboxiglutámico), EGF1 y EGF2 (regiones que contienen secuencias homólogas respecto al factor de crecimiento epidérmico humano), péptido de activación ("AP", formado por los residuos R136-R180 del FIX maduro), y el dominio de proteasa C-terminal ("Pro"), o sinónimos de estos dominios conocidos en la técnica, o puede ser un fragmento truncado o una variante de secuencia que retenga al menos una parte de la actividad biológica de la proteína nativa. Se han clonado FIX o variantes de secuencia, como se describe en las patentes de los Estados Unidos Nos. 4.770.999 y 7.700.734, y se ha aislado, caracterizado y clonado el ADNc que codifica FIX humano en vectores de expresión (véase, por ejemplo, Choo et al., Nature 299: 178-180 (1982); Fair et al., Blood 64:

194-204 (1984); y Kurachi et al., Proc. Natl. Acad. Sci., U.S.A. 79: 6461-6464 (1982)). Una variante particular de FIX, la variante R338L FIX (Padua) (SEQ ID NO: 2), caracterizada por Simioni et al, 2009, comprende una mutación de ganancia de función, que se correlaciona con un aumento de casi 8 veces en la actividad de la variante de Padua en relación con FIX nativo (Tabla 4). Las variantes de FIX también pueden incluir cualquier polipéptido FIX que tenga una o más sustituciones conservadoras de aminoácidos, que no afecten a la actividad FIX del polipéptido FIX. En algunas realizaciones, la variante FIX comprende rFIX-albúmina fusionada por un enlazante escindible, por ejemplo, IDELVION®. Ver el documento US 7.939.632, incorporado en este documento como referencia en su totalidad.

Tabla 4: Ejemplos de secuencias FIX

|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b><u>SEQ ID NO: 125 (polipéptido FIX maduro)</u></b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                             |
| 1:YNSGKLEEFV QGNLERECME EKCSFEEARE VFENTERTE FWKQYVDGDQ<br>CESNPCLNGG<br>61:SCKDDINSYE CWCPFGFEGK NCELDVTCNI KNGRCEQFCK NSADNKVVCS<br>CTEGYRLAEN 121:QKSCEPAVPF PCGRVSVSQT SKLTRAETVF PDVDYVNSTE<br>AETILDNITQ STQSFNDFTR<br>181:VVGGEDAKPG QFPWQVVLNG KVDAFCGGS VNEKWIVTAA HCVETGVKIT<br>VVAGEHNIEE<br>241:TEHTEQKRN IRIIPHHNYN AAINKYNHDI ALLELDEPLV LNSYVTPICI<br>ADKEYTNIFL<br>301:KFGSGYVSGW GRVFKHGRSA LVLQYLRVPL VDRATCLRST KFTIYNNMFC<br>AGFHEGGRDS<br>361:CQGDSSGPHV TEVEGTSFLT GIISWGEECA MKGKYGIYTK VSRVYNWIK<br>KTKLT |
| <b><u>SEQ ID NO: 126 (Polipéptido Padua(R338L)FIX maduro)</u></b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |
| 1:YNSGKLEEFV QGNLERECME EKCSFEEARE VFENTERTE FWKQYVDGDQ<br>CESNPCLNGG<br>61:SCKDDINSYE CWCPFGFEGK NCELDVTCNI KNGRCEQFCK NSADNKVVCS<br>CTEGYRLAEN 121:QKSCEPAVPF PCGRVSVSQT SKLTRAETVF PDVDYVNSTE<br>AETILDNITQ STQSFNDFTR<br>181:VVGGEDAKPG QFPWQVVLNG KVDAFCGGS VNEKWIVTAA HCVETGVKIT<br>VVAGEHNIEE                                                                                                                                                                                                                              |

241:TEHTEQKRN V IRIIPHHN YN AAINKYNHDI ALLELDEPLV LNSYVTPICI  
ADKEYTNIFL  
301:KFGSGYVSGW GRV FHKGRSA LVLQYLRVPL VDRATCLLST KFTIYNNMFC  
AGFHEGGRDS  
361:CQGDSSGPHV TEVEGTSFLT GIISWGEECA MKGKYGIYTK VSRYVNWIKE  
KTKLT

**SEQ ID NO: 127 (Polipéptido Señal FIX y Propéptido)**

1: MQRVNMIMAE SPGLITICLL GYLLSAECTV *FLDHENANKI LNRPKR*

**SEQ ID NO: 160 (FIX-Enlazante-Albúmina)**

YNSGKLEEFV QGNLERECME EKCSFEARE VFENTERTE FWKQYVDGDQ 50  
CESNPCLNGG SCKDDINSYE CWCPFGFEGK NCELDVTCNI KNGRCEQFCK 100  
NSADNKVVCS CTEGYRLAEN QKSCEPAVPF PCGRVSVSQT SKLTRAETVF 150  
PDVDYVNSTE AETILDNITQ STQSFNDFTR VVGGEDAKPG QFPWQVVLNG 200  
KVDAFCGCSI VNEKWIVTAA HCVETGVKIT VVAGEHNIEE TEHTEQKRN V 250  
IRIIPHHN YN AAINKYNHDI ALLELDEPLV LNSYVTPICI ADKEYTNIFL 300  
KFGSGYVSGW GRV FHKGRSA LVLQYLRVPL VDRATCLRST KFTIYNNMFC 350  
AGFHEGGRDS CQGDSSGPHV TEVEGTSFLT GIISWGEECA MKGKYGIYTK 400  
VSRYVNWIKE KTKLTPVSQT SKLTRAETVF PDVDAH KSEV AHRFKDLGEE 450  
NFKALVLIAF AQYLQQCPFE DHVKLVNEVT EFAKTCVADE SAENCDKSLH 500  
TLFGDKLCTV ATLRETYGEM ADCCAKQEPE RNECFLQHKD DNP NL PRLVR 550  
PEVDVMCTAF HDNEETFLKK YLYEIARRHP YFYAPELLFF AKRYKAA FTE 600  
CCQAADKAAC LLPKLDEL RD EGKASSAKQR LK CASLQKFG ERAFKAWAVA 650  
RLSQRFPKAE FAEVSKLVTD LTKVHTECCH GDLLECADDR ADLAKYICEN 700  
QDSISSKLKE CCEKPLLEKS HCIAEVENDE MPADLPSLAA DFVESKDVCK 750  
NYAEAKDVFL GMFLY EYARR HPDYSVLLLL RLA KTYETTL EKCCAAADPH 800  
ECYAKVFDEF KPLVEEPQNL IKQNC ELF EQ LGEYKFQNAL LVRYTKKVPQ 850  
VSTPTLVEVS RNLGKVGSKC CKHPEAKRMP CAEDYLSVVL NQLCVLHEKT 900  
PVSDRVTKCC TESLVNRRPC FSALEVDETY VPKEFNAETF TFHADICTLS 950  
EKERQIKKQT ALVELVKHKP KATKEQLKAV MDDFAAFVEK CCKADDKETC 1000  
FAEEGKKLVA ASQAALGL 1018

**SEQ ID NO: 161 (FIX)**

YNSGKLEEFVQGNLERECMEEEKCSFEEAREVFENTERTEFWKQYVDGDQCESNP  
CLNGGSCCKDDINSYECWCPFGFEGKNCELDVTCNIKNRCEQFCKNSADNKVVCSC  
TEGYRLAENQKSCEPAVPFPCGRVSVSQTSLTRAETVFPDVDYVNSTEAETILDNIT  
QSTQSFNDFTRVVGGEDAKPGQFPWQVVLNGKVDAFCGGSIVNEKWIVTAAHCVET  
GVKITVVAGEHNIEETEHETEQKRNVIRIIPHHNYNAAINKYNHDIALLELDEPLVLNSYVT  
PICIADKEYTNIFLKFGSGYVSGWGRVFBKGRSALVLQYLRVPLVDRATCLRSTKFTIY  
NNMFCAGFHEGGRDSCQGDSSGPHVTEVEGTSFLTGIISWGEECAMKGKYGIYTKV  
SRYVNWIEKTKLTPVSQTSKLT

**SEQ ID NO: 162 (Enlazante)**

RAETVFPDV

**SEQ ID NO: 163 (Albúmina)**

DAHKSEVAHRFKDLGEENFKALVLIAFAQYLQQCPFEDHVKLNVNEVTEFAKTCVADES  
AENCCKSLHTLFGDKLCTVATLRETYGEMADCCAKQEPERNECFLQHKDDNPNLPR  
LVRPEVDVMCTAFHDNEETFLKKYLYEIARRHPYFYAPELLFFAKRYKAAFTECCQAA  
DKAACLLPKLDELDEGKASSAKQRLKCASLQKFGERAFAKAWAVARLSQRFPAEFA  
EVSKLVTDLTKEVHTECCHGDLLECADDRADLAKYICENQDSISSKLKECCEKPLLEKS  
HCIAEVENDEMPADLPSLAADFVESKDVCKNYAEAKDVFLGMFLYEYARRHPDYSVV  
LLLRLAKTYETTLEKCCAAADPHECYAKVFDEFKPLVEEPQNLKQNCLEFEQLGEYK  
FQNALLVRYTKKVPQVSTPTLVEVSRNLGKVGSKCKHPEAKRMPCAEDYLSVVLNQ  
LCVLHEKTPVSDRVTKCCTESLVNRRPCFSALEVDETYVPKEFNAETFTFHADICTLS  
EKERQIKKQTALVELVKHKPKATKEQLKAVMDDFAAFVEKCKADDKETCFAEEGKK  
LVAASQAALGL

**SEQ ID NO: 164 (FIX(XTEN)-Fc)\***

MQRVNMIMAESPLITICLLGYLLSAECTVFLDHENANKILNRPKRYNSGKLEEFVQG  
NLERECMEEEKCSFEEAREVFENTERTEFWKQYVDGDQCESNPCLNGGSCCKDDINS  
YECWCPFGFEGKNCELDVTCNIKNRCEQFCKNSADNKVVCSCTEGYRLAENQKSC  
EPAVPFPCGRVSVSQTSLTRAETVFPDVDYVNSTEAETILDGPSPGSPTSTEEGTSE  
SATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAP  
GASSNITQSTQSFNDFTRVVGGEDAKPGQFPWQVVLNGKVDAFCGGSIVNEKWIVT  
AAHCVETGVKITVVAGEHNIEETEHETEQKRNVIRIIPHHNYNAAINKYNHDIALLELDEP  
LVLNSYVTPICIADKEYTNIFLKFGSGYVSGWGRVFBKGRSALVLQYLRVPLVDRATC  
LLSTKFTIYNNMFCAGFHEGGRDSCQGDSSGPHVTEVEGTSFLTGIISWGEECAMKG  
KYGIYTKVSRYVNWIEKTKLTDKTHTCPPCPAPELLGGPSVFLFPPKPKDTLMISRT

PEVTCVVVDVSHEDPEVKFNWYVDGVEVHNAKTKPREEQYNSTYRVVSVLTVLHQ  
DWLNGKEYKCKVSNKALPAPIEKTISKAKGQPREPQVYTLPPSRDELTKNQVSLTCL  
LVKGFYPSDIAVEWESNGQPENNYKTTPPVLDSDGSFFLYSKLTVDKSRWQQGNVFC  
SCSVMHEALHNHYTQKSLSLSPG

**SEQ ID NO: 165 (FIX-FXla-AE288)\***

YNSGKLEEFVQGNLERECMEEKCSFEEAREVFENTERTEFWKQYVDGDQCESNP  
CLNGGSCCKDDINSYECWCPFGFEGKNCELDVTCTNIKNGRCEQFCKNSADNKVVCSC  
TEGYRLAENQKSCEPAVPFPCGRVSVSQTSLTRAETVFPDVDYVNSTEAETILDNIT  
QSTQSFNDFTRVVGGEDAKPGQFPWQVVLNGKVD AFCGGSIVNEKWIVTAAHCVET  
GVKITVVAGEHNIEETEHEQKRN VIRIIPHHNYNAAINKYNHDIALLELDEPLVLNSYVT  
PICIADKEYTNIFLKFGSGYVSGWGRV FHKGRSALVLQYLRVPLVDRATCLRSTKFTIY  
NNMFCAGFHEGG RDSCQGD SGGPHVTEVEGTSFLTGIISWGEECAMKGKYG IYTKV  
SRYVNWIKEKTKLTGKLTRAETGGTSESATPESGPGSEPATSGSETPGTSESATPES  
GPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESA  
TPESGPGSEPATSGSETPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGT  
STEPSEGSAPGTSESATPESGPGTSESATPESGPGTSESATPESGPGSEPATSGSE  
TPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGSEPATSG  
GSETPGTSESATPESGPGTSTEPSEGSAP

**SEQ ID NO: 166 (FIX-Fc-Fc)\*\***

MQRVNMIMAESPLITICLLGYLLSAECTVFLDHENANKILNRPKRYNSGKLEEFVQG  
NLERECMEEKCSFEEAREVFENTERTEFWKQYVDGDQCESNPCLNGGSCCKDDINS  
YECWCPFGFEGKNCELDVTCTNIKNGRCEQFCKNSADNKVVCSCTEGYRLAENQKSC  
EPAVPFPCGRVSVSQTSLTRAETVFPDVDYVNSTEAETILDNITQSTQSFNDFTRV  
GGEDAKPGQFPWQVVLNGKVD AFCGGSIVNEKWIVTAAHCVETGVKITVVAGEHNIE  
ETEHEQKRN VIRIIPHHNYNAAINKYNHDIALLELDEPLVLNSYVTPICIADKEYTNIFLK  
FGSGYVSGWGRV FHKGRSALVLQYLRVPLVDRATCLRSTKFTIYNMFCAGFHEGG  
RDSCQGD SGGPHVTEVEGTSFLTGIISWGEECAMKGKYG IYTKVSRYVNWIKEKTKL  
TDKTHTCPPCPAPELLGGPSVFLFPPKPKDTLMISRTPEVTCVVVDVSHEDPEVKFN  
WYVDGVEVHNAKTKPREEQYNSTYRVVSVLTVLHQDWLNGKEYKCKVSNKALPAPI  
EKTISKAKGQPREPQVYTLPPSRDELTKNQVSLTCLVKGFYPSDIAVEWESNGQPEN  
NYKTTPPVLDSDGSFFLYSKLTVDKSRWQQGNVFCSCSVMHEALHNHYTQKSLSLSP  
GKRRRRSGGGGSGGGGSGGGGSGGGGSRRRRDKTHTCPPCPAPELLGGPSVFLF  
PPKPKDTLMISRTPEVTCVVVDVSHEDPEVKFNWYVDGVEVHNAKTKPREEQYNST  
YRVVSVLTVLHQDWLNGKEYKCKVSNKALPAPIEKTISKAKGQPREPQVYTLPPSRD

|                                                                                                |
|------------------------------------------------------------------------------------------------|
| ELTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTTPVLDSGDSFFLYSKLTVD<br>KSRWQQGNVFSCSVMHEALHNHYTQKSLSLSPGK |
|------------------------------------------------------------------------------------------------|

\* Sombreado gris = péptido señal; subrayado = secuencia XTEN; negrita = Fc.

\*\* SEQ ID NO: 67 de la Patente de los Estados Unidos Núm. 9.856.468, que se incorpora por referencia en este documento en su totalidad.

5   **[00236]**     El polipéptido FIX tiene 55 kDa, sintetizado como una cadena prepropolipéptida (ID. SEQ. NO: 125) compuesto por tres regiones: un péptido señal de 28 aminoácidos (aminoácidos 1 a 28 de la SEQ ID NO: 127), un propéptido de 18 aminoácidos (aminoácidos 29 a 46), que se requiere para la gamma-carboxilación de residuos de ácido glutámico, y un Factor IX maduro de 415 aminoácidos (SEQ ID NO: 10   125 o 126). El propéptido es una secuencia de residuos de 18 aminoácidos N-terminal con respecto al dominio de gamma-carboxiglutamato. El propéptido se une a la gamma carboxilasa dependiente de la vitamina K y luego se escinde del polipéptido precursor de FIX por una proteasa endógena, muy probablemente PACE (enzima de corte de aminoácidos básica combinada), también conocida como furina o PCSK3. 15   Sin la gamma carboxilación, el dominio Gla no puede unir el calcio para asumir la conformación correcta necesaria para anclar la proteína a las superficies de fosfolípidos con carga negativa, lo que hace que el Factor IX no sea funcional. Incluso si está carboxilado, el dominio Gla también depende de la escisión del propéptido para su correcto funcionamiento, ya que el propéptido retenido interfiere con los cambios 20   conformacionales del dominio Gla necesarios para una unión óptima al calcio y al fosfolípido. En seres humanos, el factor IX maduro resultante es secretado por las células hepáticas en el torrente sanguíneo como un zimógeno inactivo, una proteína monocatenaria de 415 residuos de aminoácidos que contiene aproximadamente el 17% de carbohidratos en peso (Schmidt, A. E., et al. (2003) Trends Cardiovasc Med, 25   13: 39).

**[00237]**     El FIX maduro está compuesto por varios dominios que en una configuración de terminal N a C son: un dominio GLA, un dominio EGF1, un dominio EGF2, un dominio de péptido de activación (AP) y un dominio de proteasa (o catalítico). Un enlazante corto conecta el dominio EGF2 con el dominio AP. FIX 30   contiene dos péptidos de activación formados por R145-A146 y R180-V181, respectivamente. Después de la activación, el FIX monocatenario se convierte en una molécula de 2 cadenas, en donde las dos cadenas están unidas por un enlace disulfuro. Los factores de coagulación pueden diseñarse reemplazando sus péptidos

de activación, lo que da como resultado una especificidad de activación alterada. En mamíferos, el FIX maduro debe activarse mediante el Factor XI activado para producir el Factor IXa. El dominio de proteasa proporciona, tras la activación de FIX a FIXa, la actividad catalítica de FIX. El factor VIII activado (FVIIIa) es el cofactor específico para la expresión completa de la actividad de FIXa.

**[00238]** En ciertas realizaciones, un polipéptido FIX comprende una forma alélica Thr148 de FIX derivado de plasma y tiene características estructurales y funcionales similares al FIX endógeno.

**[00239]** Muchas variantes FIX funcionales son conocidas en la técnica. La publicación internacional número WO 02/040544 A3 describe mutantes que exhiben una mayor resistencia a la inhibición por heparina en la página 4, líneas 9-30 y página 15, líneas 6-31. La publicación internacional número WO 03/020764 A2 describe mutantes FIX con inmunogenicidad reducida de células T en las Tablas 2 y 3 (en las páginas 14-24), y en la página 12, líneas 1-27. La publicación internacional número WO 2007/149406 A2 describe moléculas funcionales FIX mutantes que exhiben una mayor estabilidad de la proteína, una mayor semivida in vivo e in vitro, y una mayor resistencia a las proteasas en la página 4, línea 1 a la página 19, línea 11. El documento WO 2007/149406 A2 también describe moléculas FIX quiméricas y otras variantes de FIX en la página 19, línea 12 a la página 20, línea 9. La publicación internacional número WO 08/118507 A2 describe mutantes FIX que exhiben una actividad de coagulación incrementada en la página 5, línea 14 a la página 6, línea 5. El número de publicación internacional WO 09/051717 A2 describe mutantes FIX que tienen un mayor número de sitios de glicosilación unidos a N y/o unidos a O, lo que da como resultado una semivida y/o recuperación aumentada en la página 9, línea 11 a página 20, línea 2. La publicación internacional número WO 09/137254 A2 también describe mutantes del Factor IX con un mayor número de sitios de glicosilación en la página 2, párrafo [006] a la página 5, párrafo [011] y página 16, párrafo [044] a la página 24, párrafo [057]. La publicación internacional número WO 09/130198 A2 describe moléculas FIX mutantes funcionales que tienen un mayor número de sitios de glicosilación, que dan como resultado una semivida aumentada, en la página 4, línea 26 a la página 12, línea 6. El número de publicación internacional WO 09/140015 A2 describe mutantes FIX funcionales que aumentan el número de residuos Cys, que pueden usarse para la conjugación de polímeros (p. ej., PEG), en la página 11, párrafo [0043] a la página 13, párrafo [0053]. Los polipéptidos FIX descritos en la solicitud internacional nº PCT/US2011/043569 presentada el 11 de julio de 2011 y publicada

como WO 2012/006624 el 12 de enero de 2012 también se incorporan en este documento como referencia en su totalidad. En algunas realizaciones, el polipéptido FIX comprende un polipéptido FIX fusionado a una albúmina, por ejemplo, albúmina FIX. En ciertas realizaciones, el polipéptido FIX es IDELVION® o rIX-FP.

5 **[00240]** Además, se han identificado cientos de mutaciones no funcionales en FIX en sujetos con hemofilia, muchas de las cuales se describen en la Tabla 6, en las páginas 11-14 de la publicación internacional número WO 09/137254 A2. Dichas mutaciones no funcionales no se incluyen en la invención, pero proporcionan una guía adicional para las mutaciones que tienen más o menos probabilidades de dar como  
10 resultado un polipéptido FIX funcional.

**[00241]** En una realización, el polipéptido FIX (o la porción del Factor IX de un polipéptido de fusión) comprende una secuencia de aminoácidos al menos 70%, al menos 80%, al menos 85%, al menos 90%, al menos 95%, al menos 96%, al menos 97%, al menos 98%, al menos 99% o 100% idénticos a la secuencia establecida en  
15 SEQ ID NO: 1 o 2 (aminoácidos 1 a 415 de SEQ ID NO: 125 o 126), o alternativamente, con una secuencia de propéptido, o con una secuencia de propéptido y señal (FIX de longitud completa). En otra realización, el polipéptido FIX comprende una secuencia de aminoácidos al menos 70%, al menos 80%, al menos 85%, al menos 90%, al menos 95%, al menos 96%, al menos 97%, al menos 98 %, al menos 99% o 100% idéntica a la secuencia establecida en SEQ ID NO: 2.  
20

**[00242]** La actividad coagulante FIX se expresa como Unidad(es) Internacional(es). Una UI de actividad FIX corresponde aproximadamente a la cantidad de FIX en un mililitro de plasma humano normal. Hay varios ensayos disponibles para medir la actividad de FIX, incluido el ensayo de coagulación de una  
25 etapa (tiempo de tromboplastina parcial activada; aPTT), tiempo de generación de trombina (TGA) y tromboelastometría rotacional (ROTEM®). La invención contempla secuencias que tienen homología con secuencias FIX, fragmentos de secuencia que son naturales, tales como humanos, primates no humanos, mamíferos (incluidos animales domésticos) y variantes de secuencia no naturales que retienen al menos  
30 una parte de la actividad biológica o la función biológica de FIX y/o que son útiles para prevenir, tratar, mediar o mejorar una enfermedad, deficiencia, trastorno o afección relacionada con el factor de coagulación (p. ej., episodios de sangrado relacionados con traumatismos, cirugía o deficiencia de algún factor de coagulación). Se pueden encontrar secuencias con homología con FIX humano mediante técnicas de búsqueda  
35 de homología estándar, como NCBI BLAST.

**[00243]** En ciertas realizaciones, la secuencia FIX está optimizada para codones. Los ejemplos de secuencias FIX optimizadas para codones incluyen, entre otros, las SEQ ID N°: 1 y 54-58 de la Publicación internacional N° WO 2016/004113 AI, que se incorpora en este documento como referencia en su totalidad.

5 **C. Secuencias de FVII y polinucleótidos que codifican la proteína FVII**

**[00244]** En algunas realizaciones, la molécula de ácido nucleico comprende una primera ITR, una segunda ITR, y un casete génico que codifica una proteína terapéutica, en donde la proteína terapéutica comprende un polipéptido del Factor VII. En algunas realizaciones, el polipéptido FVII comprende FVII o una variante o  
10 fragmento del mismo, en donde la variante o fragmento del mismo tiene una actividad FVII.

**[00245]** El "Factor VII" ("FVII" o "F7;" también denominado Factor 7, factor de coagulación VII, factor de suero VII, acelerador de la conversión de protrombina en suero, SPCA, proconvertina y eptacog alfa) es una serina proteasa que forma parte  
15 de la cascada de coagulación. En una realización, el factor de coagulación en el ácido nucleico descrito en este documento es FVII. El factor VII activado recombinante ("FVII") se ha utilizado ampliamente para el tratamiento de hemorragias graves, como la que ocurre en pacientes con hemofilia A o B, deficiencia del factor de coagulación XI, FVII, función plaquetaria defectuosa, trombocitopenia o enfermedad von  
20 Willebrand.

**[00246]** El FVII activado recombinante (rFVIIa; NOVOSEVEN®) se usa para tratar episodios de sangrado en (i) pacientes con hemofilia con anticuerpos neutralizantes contra FVIII o FIX (inhibidores), (ii) pacientes con deficiencia de FVII o (iii) pacientes con hemofilia A o B con inhibidores sometidos a procedimientos  
25 quirúrgicos. Sin embargo, NOVOSEVEN® muestra poca eficacia. A menudo se requieren dosis repetidas de FVIIa a alta concentración para controlar un sangrado, debido a su baja afinidad por las plaquetas activadas, semivida corta y poca actividad enzimática en ausencia de factor tisular. En consecuencia, existe una necesidad médica insatisfecha de mejores opciones de tratamiento y prevención para pacientes  
30 con hemofilia con inhibidores de FVIII y FIX y/o con deficiencia de FVII.

**[00247]** En una realización, el casete génico codifica una forma madura de FVII o una variante del mismo. FVII incluye un dominio Gla, dos dominios EGF (EGF-1 y EGF-2) y un dominio de serina proteasa (o dominio de peptidasa SI) que está altamente conservado entre todos los miembros de la familia de serina proteasas de  
35 peptidasa SI, como por ejemplo con quimotripsina. El FVII aparece como un zimógeno

monocatenario (es decir, FVII activable) y una forma de dos cadenas completamente activada.

### **C. Factores de crecimiento**

**[00248]** En algunas realizaciones, la molécula de ácido nucleico comprende una primera ITR, una segunda ITR, y un casete génico que codifica una proteína terapéutica, en donde la proteína terapéutica comprende un factor de crecimiento. El factor de crecimiento se puede seleccionar de cualquier factor de crecimiento conocido en la técnica. En algunas realizaciones, el factor de crecimiento es una hormona. En otras realizaciones, el factor de crecimiento es una citocina. En algunas realizaciones, el factor de crecimiento es una quimiocina.

**[00249]** En algunas realizaciones, el factor de crecimiento es adrenomedulina (AM). En algunas realizaciones, el factor de crecimiento es angiopoyetina (Ang). En algunas realizaciones, el factor de crecimiento es el factor de motilidad autocrina. En algunas realizaciones, el factor de crecimiento es una proteína morfogenética ósea (BMP). En algunas realizaciones, el BMP se selecciona de BMP2, BMP4, BMP5 y BMP7. En algunas realizaciones, el factor de crecimiento es un miembro de la familia del factor neurotrófico ciliar. En algunas realizaciones, el miembro de la familia del factor neurotrófico ciliar se selecciona de factor neurotrófico ciliar (CNTF), factor inhibidor de leucemia (LIF), interleucina-6 (IL-6). En algunas realizaciones, el factor de crecimiento es un factor estimulante de colonias. En algunas realizaciones, el factor estimulante de colonias se selecciona de factor estimulante de colonias de macrófagos (m-CSF), factor estimulante de colonias de granulocitos (G-CSF) y factor estimulante de colonias de macrófagos y granulocitos (GM-CSF). En algunas realizaciones, el factor de crecimiento es un factor de crecimiento epidérmico (EGF). En algunas realizaciones, el factor de crecimiento es una efrina. En algunas realizaciones, la efrina se selecciona de efrina A1, efrina A2, efrina A3, efrina A4, efrina A5, efrina B1, efrina B2 y efrina B3. En algunas realizaciones, el factor de crecimiento es la eritropoyetina (EPO). En algunas realizaciones, el factor de crecimiento es un factor de crecimiento de fibroblastos (FGF). En algunas realizaciones, el FGF se selecciona de FGF1, FGF2, FGF3, FGF4, FGF5, FGF6, FGF7, FGF8, FGF9, FGF10, FGF11, FGF12, FGF13, FGF14, FGF15, FGF16, FGF17, FGF18, FGF19, FGF20, FGF21, FGF22 y FGF23. En algunas realizaciones, el factor de crecimiento es la somatotropina bovina fetal (FBS). En algunas realizaciones, el factor de crecimiento es un miembro de la familia GDNF. En algunas realizaciones, el miembro de la familia GDNF se selecciona del factor neurotrófico derivado de la línea de células gliales

(GDNF), neurturina, persefina y artemina. En algunas realizaciones, el factor de crecimiento es el factor de diferenciación de crecimiento-9 (GDF9). En algunas realizaciones, el factor de crecimiento es el factor de crecimiento de hepatocitos (HGF). En algunas realizaciones, el factor de crecimiento es el factor de crecimiento derivado de hepatoma (HDGF). En algunas realizaciones, el factor de crecimiento es la insulina. En algunas realizaciones, el factor de crecimiento es un factor de crecimiento similar a la insulina. En algunas realizaciones, el factor de crecimiento similar a la insulina es el factor de crecimiento similar a la insulina-1 (IGF-1) o IGF-2. En algunas realizaciones, el factor de crecimiento es una interleucina (IL). En algunas realizaciones, la IL se selecciona de IL-1, IL-2, IL-3, IL-4, IL-5, IL-6 e IL-7. En algunas realizaciones, el factor de crecimiento es el factor de crecimiento de queratinocitos (KGF). En algunas realizaciones, el factor de crecimiento es el factor estimulante de la migración (MSF). En algunas realizaciones, el factor de crecimiento es la proteína estimulante de macrófagos (MSP o proteína similar al factor de crecimiento de hepatocitos (HGFLP)). En algunas realizaciones, el factor de crecimiento es la miostatina (GDF-8). En algunas realizaciones, el factor de crecimiento es una neuregulina. En algunas realizaciones, la neuregulina se selecciona de neuregulina 1 (NRG1), NRG2, NRG3 y NRG4. En algunas realizaciones, el factor de crecimiento es una neurotrofina. En algunas realizaciones, el factor de crecimiento es el factor neurotrófico derivado del cerebro (BDNF). En algunas realizaciones, el factor de crecimiento es el factor de crecimiento nervioso (NGF). En algunas realizaciones, el NGF es neurotrofina-3 (NT-3) o NT-4. En algunas realizaciones, el factor de crecimiento es el factor de crecimiento placentario (PGF). En algunas realizaciones, el factor de crecimiento es el factor de crecimiento derivado de plaquetas (PDGF). En algunas realizaciones, el factor de crecimiento es la renalasa (RNLS). En algunas realizaciones, el factor de crecimiento es el factor de crecimiento de células T (TCGF). En algunas realizaciones, el factor de crecimiento es la trombopoyetina (TPO). En algunas realizaciones, el factor de crecimiento es un factor de crecimiento transformante. En algunas realizaciones, el factor de crecimiento transformante es el factor de crecimiento transformante alfa (TGF- $\alpha$ ) o TGF- $\beta$ . En algunas realizaciones, el factor de crecimiento es el factor de necrosis tumoral alfa (TNF- $\alpha$ ). En algunas realizaciones, el factor de crecimiento es el factor de crecimiento endotelial vascular (VEGF).

#### **D. Micro ARN (ARNmi)**

**[00250]** Los microARN (ARNmi) son pequeñas moléculas de ARN no codificantes (aproximadamente 18-22 nucleótidos) que regulan negativamente la expresión génica al inhibir la traducción o inducir la degradación del ARN mensajero (ARNm). Desde su descubrimiento, los ARNmi han estado implicados en varios procesos celulares, incluida la apoptosis, la diferenciación y la proliferación celular, y han demostrado desempeñar un papel clave en la carcinogénesis. La capacidad de los ARNmi para regular la expresión génica hace que la expresión de ARNmi in vivo sea una herramienta valiosa en la terapia génica.

**[00251]** Ciertos aspectos de la presente descripción están dirigidos a moléculas de ácido nucleico de tipo plásmido que comprenden una primera ITR, una segunda ITR y un casete génico que codifica un ARNmi, en donde la primera ITR y/o la segunda ITR son una ITR de un virus no adeno-asociado (por ejemplo, la primera ITR y/o la segunda ITR son de un VnoAA). El ARNmi puede ser cualquier ARNmi conocido en la técnica. En algunas realizaciones, el ARNmi regula a la baja la expresión de un gen diana. En ciertas realizaciones, el gen diana se selecciona de SOD1, HTT, RHO o cualquiera de sus combinaciones.

**[00252]** En algunas realizaciones, el casete génico codifica un ARNmi. En algunas realizaciones, el casete génico codifica más de un ARNmi. En algunas realizaciones, el casete génico codifica dos o más ARNmi diferentes. En algunas realizaciones, el casete génico codifica dos o más copias del mismo ARNmi. En algunas realizaciones, el casete génico codifica dos o más variantes de la misma proteína terapéutica. En ciertas realizaciones, el casete génico codifica uno o más ARNmi y una o más proteínas terapéuticas.

**[00253]** En algunas realizaciones, el ARNmi es un ARNmi natural. En algunas realizaciones, el ARNmi es un ARNmi diseñado. En algunas realizaciones, el ARNmi es un ARNmi artificial. En ciertas realizaciones, el ARNmi comprende el ARNmi diseñado por miHTT descrito por Evers et al., Molecular Therapy 26 (9): 1 - 15 (epub antes de la impresión de junio de 2018). En ciertas realizaciones, el ARNmi comprende el ARNmi artificial miR SOD1 descrito por Dirren et al., Annals of Clinical and Translational Neurology 2 (2): 167-84 (febrero de 2015). En ciertas realizaciones, el ARNmi comprende miR-708, que se dirige a RHO (ver Behrman et al., JCB 192 (6): 919-27 (2011)).

**[00254]** En algunas formas de realización, el ARNmi regula al alza la expresión de un gen regulando negativamente la expresión de un inhibidor del gen. En algunas realizaciones, el inhibidor es un inhibidor natural, por ejemplo, natural. En algunas

realizaciones, el inhibidor es el resultado de un gen mutado, heterólogo y/o poco expresado.

**E. Restos heterólogos**

**[00255]** En algunas realizaciones, la molécula de ácido nucleico comprende una primera ITR, una segunda ITR, y un casete génico que codifica una proteína terapéutica, en el que la proteína terapéutica comprende al menos un resto heterólogo. En algunas realizaciones, el resto heterólogo se fusiona con el extremo N o el extremo C de la proteína terapéutica. En otras realizaciones, el resto heterólogo se inserta entre dos aminoácidos dentro de la proteína terapéutica.

**[00256]** En algunas realizaciones, la proteína terapéutica comprende un polipéptido FVIII y un resto heterólogo, que se inserta entre dos aminoácidos dentro del polipéptido FVIII. En algunas realizaciones, el resto heterólogo se inserta dentro del polipéptido FVIII en uno o más sitios de inserción seleccionados de la Tabla 5. En algunas realizaciones, la secuencia de aminoácidos heteróloga se puede insertar dentro del polipéptido del factor de coagulación codificado por la molécula de ácido nucleico de la descripción en cualquier sitio descrito en la Publicación Internacional No. WO 2013/123457 A1, WO 2015/106052 A1 o la Publicación de los Estados Unidos No. 2015/0158929 A1, que se incorporan en este documento como referencia en su totalidad. En una realización particular, la proteína terapéutica comprende un FVIII y un resto heterólogo, en el que el resto heterólogo se inserta dentro del FVIII inmediatamente después del aminoácido 745 con respecto al FVIII maduro. En una realización particular, la proteína terapéutica comprende un FVIII y un XTEN en el que el XTEN se inserta dentro del FVIII inmediatamente después del aminoácido 745 con respecto al FVIII maduro. En una realización particular, el FVIII comprende una delección de los aminoácidos 746-1646, correspondientes al FVIII humano maduro (SEQ ID N°: 15), y el resto heterólogo se inserta inmediatamente después del aminoácido 745, correspondiente al FVIII humano maduro (SEQ. ID NO: 15).

Tabla 5: Sitios de inserción de restos heterólogos del FVIII

| Sitio de inserción | Dominio | Sitio de inserción | Dominio | Sitio de inserción | Dominio |
|--------------------|---------|--------------------|---------|--------------------|---------|
| 3                  | A1      | 375                | A2      | 1749               | A3      |
| 18                 | A1      | 378                | A2      | 1796               | A3      |
| 22                 | A1      | 399                | A2      | 1802               | A3      |
| 26                 | A1      | 403                | A2      | 1827               | A3      |

|     |    |      |           |      |    |
|-----|----|------|-----------|------|----|
| 40  | A1 | 409  | A2        | 1861 | A3 |
| 60  | A1 | 416  | A2        | 1896 | A3 |
| 65  | A1 | 442  | A2        | 1900 | A3 |
| 81  | A1 | 487  | A2        | 1904 | A3 |
| 116 | A1 | 490  | A2        | 1905 | A3 |
| 119 | A1 | 494  | A2        | 1910 | A3 |
| 130 | A1 | 500  | A2        | 1937 | A3 |
| 188 | A1 | 518  | A2        | 2019 | A3 |
| 211 | A1 | 599  | A2        | 2068 | C1 |
| 216 | A1 | 603  | A2        | 2111 | C1 |
| 220 | A1 | 713  | A2        | 2120 | C1 |
| 224 | A1 | 745  | B         | 2171 | C2 |
| 230 | A1 | 1656 | región a3 | 2188 | C2 |
| 333 | A1 | 1711 | A3        | 2227 | C2 |
| 336 | A1 | 1720 | A3        | 2332 | CT |
| 339 | A1 | 1725 | A3        |      |    |

**[00257]** En algunas realizaciones, la proteína terapéutica comprende un polipéptido FIX y un resto heterólogo, que se inserta entre dos aminoácidos dentro del polipéptido FIX. En algunas realizaciones, el resto heterólogo se inserta dentro del polipéptido FIX en uno o más sitios de inserción seleccionados de la Tabla 5. En algunas realizaciones, la secuencia de aminoácidos heteróloga se puede insertar dentro del polipéptido del factor de coagulación codificado por la molécula de ácido nucleico de la descripción en cualquier sitio descrito en la Solicitud Internacional N° PCT/US2017/015879, que se incorpora en este documento como referencia en su totalidad. En una realización particular, la proteína terapéutica comprende un polipéptido FIX y un resto heterólogo, en el que el resto heterólogo se inserta dentro del polipéptido FIX inmediatamente después del aminoácido 166 en relación con el FIX maduro. En una realización particular, la proteína terapéutica comprende un polipéptido FIX y un XTEN, en el que el XTEN se inserta dentro del FIX inmediatamente después del aminoácido 166 en relación con el FVIII maduro.

Tabla 6: Sitios de inserción de restos heterólogos FIX

| Sitio de inserción | Dominio | Sitio de inserción | Dominio | Sitio de inserción | Dominio |
|--------------------|---------|--------------------|---------|--------------------|---------|
|--------------------|---------|--------------------|---------|--------------------|---------|

|     |        |     |            |     |            |
|-----|--------|-----|------------|-----|------------|
| 52  | EGF1   | 149 | AP         | 257 | Catalítico |
| 59  | EGF1   | 162 | AP         | 265 | Catalítico |
| 66  | EGF1   | 166 | AP         | 277 | Catalítico |
| 80  | EGF1   | 174 | AP         | 283 | Catalítico |
| 85  | EGF2   | 188 | Catalítico | 292 | Catalítico |
| 89  | EGF2   | 202 | Catalítico | 316 | Catalítico |
| 103 | EGF2   | 224 | Catalítico | 341 | Catalítico |
| 105 | EGF2   | 226 | Catalítico | 354 | Catalítico |
| 113 | EGF2   | 228 | Catalítico | 392 | Catalítico |
| 129 | Linker | 230 | Catalítico | 403 | Catalítico |
| 142 | Linker | 240 | Catalítico | 413 | Catalítico |

**[00258]** En otras realizaciones, las proteínas terapéuticas de la descripción comprenden además dos, tres, cuatro, cinco, seis, siete u ocho secuencias de nucleótidos heterólogos. En algunas realizaciones, todos los restos heterólogos son idénticos. En algunas realizaciones, al menos un resto heterólogo es diferente de los otros restos heterólogos. En algunas realizaciones, la descripción puede comprender dos, tres, cuatro, cinco, seis o más de siete restos heterólogos en tándem.

**[00259]** En algunas realizaciones, el resto heterólogo aumenta la semivida (es un "extensor de semivida") de la proteína terapéutica.

**[00260]** En algunas realizaciones, el resto heterólogo es un péptido o un polipéptido con características no estructuradas o estructuradas que están asociadas con la prolongación de la semivida in vivo cuando se incorpora en una proteína de la descripción. Los ejemplos no limitantes incluyen albúmina, fragmentos de albúmina, fragmentos Fc de inmunoglobulinas, el péptido C-terminal (CTP) de la subunidad  $\beta$  de la gonadotropina coriónica humana, una secuencia HAP, una secuencia XTEN, una transferrina o un fragmento de la misma, un polipéptido PAS, enlazantes de poliglicina, enlazantes de poliserina, restos de unión a albúmina, o cualesquiera fragmentos, derivados, variantes o combinaciones de estos polipéptidos. En una realización particular, la secuencia de aminoácidos heteróloga es una región constante de inmunoglobulina o una porción de la misma, transferrina, albúmina o una secuencia PAS. En algunos aspectos, un resto heterólogo incluye el factor von Willebrand o un fragmento del mismo. En otros aspectos relacionados, un resto heterólogo puede incluir un sitio de unión (por ejemplo, un aminoácido cisteína) para un resto no

polipeptídico, tal como polietilenglicol (PEG), hidroxietilalmidón (HES), ácido poli  
siálico o cualesquiera derivados, variantes o combinaciones de estos elementos. En  
algunos aspectos, un resto heterólogo comprende un aminoácido cisteína que  
funciona como un sitio de unión para un resto no polipéptido tal como polietilenglicol  
5 (PEG), hidroxietilalmidón (HES), ácido poli siálico o cualesquiera derivados, variantes  
o combinaciones de estos elementos

**[00261]** En una realización específica, un primer resto heterólogo es una  
molécula que extiende la semivida que se conoce en la técnica, y un segundo resto  
heterólogo es una molécula que extiende la semivida que se conoce en la técnica. En  
10 ciertas realizaciones, el primer resto heterólogo (por ejemplo, un primer resto Fc) y el  
segundo resto heterólogo (por ejemplo, un segundo resto Fc) se asocian entre sí para  
formar un dímero. En una realización, el segundo resto heterólogo es un segundo  
resto Fc, en el que el segundo resto Fc está unido o asociado con el primer resto  
heterólogo, por ejemplo, el primer resto Fc. Por ejemplo, el segundo resto heterólogo  
15 (por ejemplo, el segundo resto Fc) puede estar unido al primer resto heterólogo (por  
ejemplo, el primer resto Fc) por un enlazante o asociado con el primer resto heterólogo  
por un enlace covalente o no covalente.

**[00262]** En algunas realizaciones, el resto heterólogo es un polipéptido que  
comprende, que consiste esencialmente, o que consiste sin más en al menos  
20 aproximadamente 10, al menos aproximadamente 100, al menos aproximadamente  
200, al menos aproximadamente 300, al menos aproximadamente 400, al menos  
aproximadamente 500, al menos aproximadamente 600, al menos aproximadamente  
700, al menos aproximadamente 800, al menos aproximadamente 900, al menos  
aproximadamente 1000, al menos aproximadamente 1100, al menos  
25 aproximadamente 1200, al menos aproximadamente 1300, al menos  
aproximadamente 1400, al menos aproximadamente 1500, al menos  
aproximadamente 1600, al menos aproximadamente 1700, al menos  
aproximadamente 1800, al menos aproximadamente 1900, al menos  
aproximadamente 2000, al menos aproximadamente 2500, al menos  
30 aproximadamente 3000 o al menos aproximadamente 4000 aminoácidos. En otras  
realizaciones, el resto heterólogo es un polipéptido que comprende, que consiste  
esencialmente, o que consiste sin más entre aproximadamente 100 y  
aproximadamente 200 aminoácidos, aproximadamente 200 y aproximadamente 300  
aminoácidos, aproximadamente 300 y aproximadamente 400 aminoácidos,  
35 aproximadamente 400 y aproximadamente 500 aminoácidos, aproximadamente 500

y aproximadamente 600 aminoácidos, aproximadamente 600 y aproximadamente 700 aminoácidos, aproximadamente 700 y aproximadamente 800 aminoácidos, aproximadamente 800 y aproximadamente 900 aminoácidos, o aproximadamente 900 y aproximadamente 1000 aminoácidos.

- 5 **[00263]** En ciertas realizaciones, un resto heterólogo mejora una o más propiedades farmacocinéticas de la proteína terapéutica sin afectar significativamente su actividad o función biológica.

**[00264]** En ciertas realizaciones, un resto heterólogo aumenta la semivida in vivo y/o in vitro de la proteína terapéutica de la descripción. En otras realizaciones, un resto heterólogo facilita la visualización o localización de la proteína terapéutica de la descripción o un fragmento de la misma (por ejemplo, un fragmento que comprende un resto heterólogo después de la escisión proteolítica de la proteína FVIII). La visualización y/o ubicación de la proteína terapéutica de la descripción o un fragmento de la misma puede ser in vivo, in vitro, ex vivo o combinaciones de las mismas.

- 15 **[00265]** En otras realizaciones, un resto heterólogo aumenta la estabilidad de la proteína terapéutica de la descripción o un fragmento de la misma (por ejemplo, un fragmento que comprende un resto heterólogo después de la escisión proteolítica de la proteína terapéutica, por ejemplo, un factor de coagulación). Como se usa en el presente documento, el término "estabilidad" se refiere a una medida reconocida en la técnica del mantenimiento de una o más propiedades físicas de la proteína terapéutica en respuesta a una condición ambiental (por ejemplo, una temperatura elevada o baja). En ciertos aspectos, la propiedad física puede ser el mantenimiento de la estructura covalente de la proteína terapéutica (por ejemplo, la ausencia de escisión proteolítica, oxidación o desamidación no deseada). En otros aspectos, la propiedad física también puede ser la presencia de la proteína terapéutica en un estado plegado adecuadamente (por ejemplo, la ausencia de agregados o precipitados solubles o insolubles). En un aspecto, la estabilidad de la proteína terapéutica se mide analizando una propiedad biofísica de la proteína terapéutica, por ejemplo, estabilidad térmica, perfil de desnaturalización frente al pH, eliminación estable de la glucosilación, solubilidad, función bioquímica (por ejemplo, capacidad para unirse a una proteína, receptor o ligando), etc., y/o combinaciones de los mismos. En otro aspecto, la función bioquímica se demuestra por la afinidad de unión de la interacción. En un aspecto, una medida de la estabilidad de la proteína es la estabilidad térmica, es decir, la resistencia a la exposición al calor. La estabilidad se puede medir utilizando métodos conocidos en la técnica, como HPLC (cromatografía

líquida de alto resolución), SEC (cromatografía de exclusión por tamaños), DLS (dispersión dinámica de la luz), etc. Los métodos para medir la estabilidad térmica incluyen, entre otros, calorimetría de barrido diferencial (DSC), fluorimetría de barrido diferencial (DSF), dicroísmo circular (CD) y ensayo por exposición al calor.

5 **[00266]** En ciertos aspectos, una proteína terapéutica codificada por la molécula de ácido nucleico de la descripción comprende al menos un extensor de semivida, es decir, un resto heterólogo que aumenta la semivida in vivo de la proteína terapéutica con respecto a la semivida in vivo de la proteína terapéutica correspondiente que carece de tal resto heterólogo. La semivida in vivo de una proteína terapéutica puede  
10 determinarse mediante cualquier método conocido por los expertos en la técnica, por ejemplo, ensayos de actividad (por ejemplo, ensayo cromogénico o ensayo de aPTT de coagulación en una etapa en el que la proteína terapéutica comprende un polipéptido FVIII), ELISA, ROTEM®, etc.

**[00267]** En algunas realizaciones, la presencia de uno o más extensores de  
15 semivida da como resultado que se incremente la semivida de la proteína terapéutica en comparación con la semivida de la proteína correspondiente que carece de uno o más extensores de semivida. La semivida de la proteína terapéutica que comprende un extensor de semivida es al menos aproximadamente 1,5 veces, al menos aproximadamente 2 veces, al menos aproximadamente 2,5 veces, al menos  
20 aproximadamente 3 veces, al menos aproximadamente 4 veces, al menos aproximadamente 5 veces, al menos aproximadamente 6 veces, al menos aproximadamente 7 veces, al menos aproximadamente 8 veces, al menos aproximadamente 9 veces, al menos aproximadamente 10 veces, al menos aproximadamente 11 veces o al menos aproximadamente 12 veces más que la  
25 semivida in vivo de la proteína terapéutica correspondiente que carece de dicho extensor de semivida.

**[00268]** En una realización, la semivida de la proteína terapéutica que comprende un extensor de semivida es de aproximadamente 1,5 veces a aproximadamente 20 veces, de aproximadamente 1,5 veces a aproximadamente 15  
30 veces, o de aproximadamente 1,5 veces a aproximadamente 10 veces más larga que la semivida in vivo de la proteína correspondiente que carece de dicho extensor de semivida. En otra realización, la semivida de la proteína terapéutica que comprende un extensor de semivida se extiende aproximadamente de 2 veces a aproximadamente 10 veces, aproximadamente 2 veces a aproximadamente 9 veces,  
35 aproximadamente 2 veces a aproximadamente 8 veces, aproximadamente 2 veces a

aproximadamente 7 veces, aproximadamente 2 veces a aproximadamente 6 veces, aproximadamente 2 veces a aproximadamente 5 veces, aproximadamente 2 veces a aproximadamente 4 veces, aproximadamente 2 veces a aproximadamente 3 veces, aproximadamente 2,5 veces a aproximadamente 10 veces, aproximadamente 2,5 veces a aproximadamente 9 veces, aproximadamente 2,5 veces a aproximadamente 8 veces, aproximadamente 2,5 veces a aproximadamente 7 veces, aproximadamente 2,5 veces a aproximadamente 6 veces, aproximadamente 2,5 veces a aproximadamente 5 veces, aproximadamente 2,5 veces a aproximadamente 4 veces, aproximadamente 2,5 veces a aproximadamente 3 veces, aproximadamente 3 veces a aproximadamente 10 veces, aproximadamente 3 veces a aproximadamente 9 veces, aproximadamente 3 veces a aproximadamente 8 veces, aproximadamente 3 veces a aproximadamente 7 veces, aproximadamente 3 veces a aproximadamente 6 veces, aproximadamente 3 veces a aproximadamente 5 veces, aproximadamente 3 veces a aproximadamente 4 veces, aproximadamente 4 veces a aproximadamente 6 veces, aproximadamente 5 veces a aproximadamente 7 veces, o aproximadamente 6 veces a aproximadamente 8 veces en comparación con la semivida in vivo de la proteína correspondiente que carece de dicho extensor de semivida.

**[00269]** En otras realizaciones, la semivida de la proteína terapéutica que comprende un extensor de semivida es de al menos aproximadamente 17 horas, al menos aproximadamente 18 horas, al menos aproximadamente 19 horas, al menos aproximadamente 20 horas, al menos aproximadamente 21 horas, al menos aproximadamente 22 horas, al menos aproximadamente 23 horas, al menos aproximadamente 24 horas, al menos aproximadamente 25 horas, al menos aproximadamente 26 horas, al menos aproximadamente 27 horas, al menos aproximadamente 28 horas, al menos aproximadamente 29 horas, al menos aproximadamente 30 horas, al menos aproximadamente 31 horas, al menos aproximadamente 32 horas, al menos aproximadamente 33 horas, al menos aproximadamente 34 horas, al menos aproximadamente 35 horas, al menos aproximadamente 36 horas, al menos aproximadamente 48 horas, al menos aproximadamente 60 horas, al menos aproximadamente 72 horas, al menos aproximadamente 84 horas, al menos aproximadamente 96 horas o al menos aproximadamente 108 horas.

**[00270]** En otras realizaciones más, la semivida de la proteína terapéutica que comprende un extensor de semivida es de aproximadamente 15 horas a aproximadamente dos semanas, de aproximadamente 16 horas a aproximadamente

una semana, de aproximadamente 17 horas a aproximadamente una semana, de aproximadamente 18 horas a aproximadamente una semana, aproximadamente 19 horas a aproximadamente una semana, aproximadamente 20 horas a aproximadamente una semana, aproximadamente 21 horas a aproximadamente una semana, aproximadamente 22 horas a aproximadamente una semana, aproximadamente 23 horas a aproximadamente una semana, aproximadamente 24 horas a aproximadamente una semana, aproximadamente 36 horas a aproximadamente una semana, aproximadamente 48 horas a aproximadamente una semana, aproximadamente 60 horas a aproximadamente una semana, aproximadamente 24 horas a aproximadamente seis días, aproximadamente 24 horas a aproximadamente cinco días, aproximadamente 24 horas a aproximadamente cuatro días, aproximadamente 24 horas a aproximadamente tres días, o aproximadamente 24 horas a aproximadamente dos días.

**[00271]** En algunas realizaciones, la semivida promedio por sujeto de la proteína terapéutica que comprende un extensor de semivida es de aproximadamente 15 horas, aproximadamente 16 horas, aproximadamente 17 horas, aproximadamente 18 horas, aproximadamente 19 horas, aproximadamente 20 horas, aproximadamente 21 horas, aproximadamente 22 horas, aproximadamente 23 horas, aproximadamente 24 horas (1 día), aproximadamente 25 horas, aproximadamente 26 horas, aproximadamente 27 horas, aproximadamente 28 horas, aproximadamente 29 horas, aproximadamente 30 horas, aproximadamente 31 horas, aproximadamente 32 horas, aproximadamente 33 horas, aproximadamente 34 horas, aproximadamente 35 horas, aproximadamente 36 horas, aproximadamente 40 horas, aproximadamente 44 horas, aproximadamente 48 horas (2 días), aproximadamente 54 horas, aproximadamente 60 horas, aproximadamente 72 horas (3 días), aproximadamente 84 horas, aproximadamente 96 horas (4 días), aproximadamente 108 horas, aproximadamente 120 horas (5 días), aproximadamente seis días, aproximadamente siete días (una semana), aproximadamente ocho días, aproximadamente nueve días, aproximadamente 10 días, aproximadamente 11 días, aproximadamente 12 días, aproximadamente 13 días o aproximadamente 14 días.

**[00272]** Uno o más extensores de semivida pueden fusionarse con el extremo C o el extremo N de la proteína terapéutica o insertarse dentro de la proteína terapéutica.

### **1. Una región constante de inmunoglobulina o una porción de la misma**

**[00273]** En otro aspecto, un resto heterólogo comprende una o más regiones constantes de inmunoglobulina o porciones de las mismas (por ejemplo, una región

Fc). En una realización, una molécula de ácido nucleico aislada de la descripción comprende además una secuencia de ácido nucleico heteróloga que codifica una región constante de inmunoglobulina o una porción de la misma. En algunas realizaciones, la región constante de inmunoglobulina o una porción de la misma es una región Fc.

**[00274]** Una región constante de inmunoglobulina está compuesta por los dominios denominados dominios CH (constantes pesados) (CH1, CH2, etc.). Dependiendo del isotipo, (es decir, IgG, IgM, IgA IgD o IgE), la región constante puede estar compuesta por tres o cuatro dominios CH. Algunas regiones constantes de isotipos (p. ej. IgG) también contienen una región bisagra. Ver Janeway et al. 2001, Immunobiology, Garland Publishing, N.Y., N.Y.

**[00275]** Se puede obtener una región constante de inmunoglobulina o una porción de la misma de la presente descripción a partir de varias fuentes diferentes. En una realización, una región constante de inmunoglobulina o una porción de la misma se deriva de una inmunoglobulina humana. Sin embargo, se entiende que la región constante de inmunoglobulina o una porción de la misma puede derivarse de una inmunoglobulina de otra especie de mamífero, que incluya, por ejemplo, un roedor (por ejemplo, un ratón, rata, conejo, conejillo de indias) o un primate no humano (p. ej., chimpancés, macacos). Además, la región constante de inmunoglobulina o una porción de la misma puede derivarse de cualquier clase de inmunoglobulina, incluyendo IgM, IgG, IgD, IgA e IgE, y cualquier isotipo de inmunoglobulina, incluyendo IgG1, IgG2, IgG3 e IgG4. En una realización, se usa el isotipo humano IgG1.

**[00276]** Una variedad de secuencias de genes de la región constante de inmunoglobulina (por ejemplo, secuencias de genes de la región constante humana) están disponibles en forma de depósitos con acceso público. La secuencia de dominios de la región constante puede seleccionarse teniendo una función efectora particular (o careciendo de una función efectora particular) o con una modificación particular para reducir la inmunogenicidad. Se han publicado muchas secuencias de anticuerpos y genes que codifican anticuerpos y se pueden derivar secuencias adecuadas de la región constante de Ig (por ejemplo, secuencias de bisagra, CH2 y/o CH3, o porciones de las mismas) a partir de estas secuencias usando técnicas reconocidas en la técnica. El material genético obtenido usando cualquiera de los métodos anteriores se puede modificar o sintetizar para obtener los polipéptidos de la presente descripción. Se apreciará además que el alcance de esta descripción abarca alelos, variantes y mutaciones de secuencias de ADN de la región constante.

**[00277]** Las secuencias de la región constante de inmunoglobulina o una porción de la misma se pueden clonar, por ejemplo, usando la reacción en cadena de la polimerasa y los cebadores que se seleccionan para amplificar el dominio de interés. Para clonar una secuencia de la región constante de inmunoglobulina o una porción de la misma a partir de un anticuerpo, el ARNm puede aislarse de las células de hibridoma, bazo o linfa, transcribirse de forma inversa en el ADN y los genes del anticuerpo amplificarse por PCR. Los métodos de amplificación por PCR se describen en detalle en las patentes de EE.UU. Nos. 4.683.195; 4.683.202; 4.800.159; 4.965.188; y en, por ejemplo, "PCR Protocols: A Guide to Methods and Applications" Innis et al. eds., Academic Press, San Diego, CA (1990); Ho et al. 1989. Gene 77:51; Horton et al. 1993. Methods Enzymol. 217:270). La PCR puede iniciarse mediante cebadores de la región constante de consenso o por cebadores más específicos basados en las secuencias de ADN y aminoácidos de cadena pesada y ligera publicadas. La PCR también se puede usar para aislar clones de ADN que codifican las cadenas ligera y pesada del anticuerpo. En este caso, las bibliotecas pueden seleccionarse mediante cebadores de consenso o sondas homólogas más grandes, como sondas de la región constante de ratón. Se conocen numerosos conjuntos de cebadores adecuados para la amplificación de genes de anticuerpos en la técnica (por ejemplo, cebadores 5' basados en la secuencia N-terminal de anticuerpos purificados (Benhar y Pastan. 1994. Protein Engineering 7: 1509); amplificación rápida de los extremos de ADNc (Ruberti, F. et al. 1994. J. Immunol. Methods 173: 33); secuencias líder de anticuerpos (Larrick et al. 1989 Biochem. Biophys. Res. Commun. 160: 1250). La clonación de secuencias de anticuerpos se describe además en Newman et al., Patente de EE.UU. No. 5.658.570, presentada el 25 de enero de 1995, que se incorpora en este documento como referencia.

**[00278]** Una región constante de inmunoglobulina usada en el presente documento puede incluir todos los dominios y la región bisagra o porciones de los mismos. En una realización, la región constante de inmunoglobulina o una porción de la misma comprende el dominio CH2, el dominio CH3 y una región bisagra, es decir, una región Fc o una pareja de unión a FcRn.

**[00279]** Como se usa en el presente documento, la expresión "región Fc" se define como la porción de un polipéptido que corresponde a la región Fc de Ig nativa, es decir, como se forma por la asociación dimérica de los respectivos dominios Fc de sus dos cadenas pesadas. Una región Fc nativa forma un homodímero con otra región Fc. Por el contrario, la expresión "región Fc genéticamente fusionada" o "región Fc

monocatenaria" (región scFc), como se usa en el presente documento, se refiere a una región Fc dimérica sintética compuesta de dominios Fc genéticamente unidos dentro de una cadena de polipéptidos única (es decir, codificada en una sola secuencia genética contigua). Véase la Publicación internacional No. WO 2012/006635, incorporada en este documento como referencia en su totalidad.

**[00280]** En una realización, la "región Fc" se refiere a la porción de una sola cadena pesada de Ig que comienza en la región bisagra justo arriba del sitio de escisión de la papaína (es decir, el residuo 216 en IgG, tomando el primer residuo de la región constante de la cadena pesada 114) y terminando en el C-terminal del anticuerpo. En consecuencia, una región Fc completa comprende al menos un dominio de bisagra, un dominio CH2 y un dominio CH3.

**[00281]** Una región constante de inmunoglobulina o una porción de la misma puede ser una pareja de unión a FcRn. FcRn es activo en tejidos epiteliales adultos y se expresa en la luz de los intestinos, vías respiratorias pulmonares, superficies nasales, superficies vaginales, colon y superficies rectales (Patente de Estados Unidos No. 6.485.726). Una pareja de unión a FcRn es una porción de una inmunoglobulina que se une a FcRn.

**[00282]** El receptor FcRn se ha aislado de varias especies de mamíferos, incluidos los humanos. Se conocen las secuencias de FcRn humano, FcRn de mono, FcRn de rata y FcRn de ratón (Story et al. 1994, J. Exp. Med. 180:2377). El receptor FcRn se une a la IgG (pero no a otras clases de inmunoglobulinas como IgA, IgM, IgD e IgE) a un pH relativamente bajo, transporta activamente la IgG transcelularmente en una dirección luminal a serosa, y luego libera la IgG a un pH relativamente más alto encontrado en los fluidos intersticiales. Se expresa en tejido epitelial adulto (patentes de EE.UU. Nos. 6.485.726, 6.030.613, 6.086.875; WO 03/077834; US2003-0235536A1) incluyendo epitelio pulmonar e intestinal (Israel et al. 1997, Immunology 92:69), epitelio tubular proximal renal (Kobayashi et al. 2002, Am. J. Physiol. Renal Physiol. 282:F358), así como el epitelio nasal, las superficies vaginales y las superficies de los árboles biliares.

**[00283]** Las parejas de unión a FcRn útiles en la presente descripción abarcan moléculas que pueden unirse específicamente por el receptor de FcRn incluyendo IgG completa, el fragmento Fc de IgG y otros fragmentos que incluyen la región de unión completa del receptor de FcRn. La región de la porción Fc de IgG que se une al receptor FcRn se ha descrito en base a la cristalografía de rayos X (Burmeister et al. 1994, Nature 372: 379). El área principal de contacto de Fc con FcRn está cerca de la

unión de los dominios CH2 y CH3. Los contactos Fc-FcRn están todos dentro de una sola cadena pesada de Ig. Las parejas de unión a FcRn incluyen IgG completa, el fragmento Fc de IgG y otros fragmentos de IgG que incluyen la región de unión completa de FcRn. Los principales sitios de contacto incluyen los residuos de aminoácidos 248, 250-257, 272, 285, 288, 290-291, 308-311 y 314 del dominio CH2 y los residuos de aminoácidos 385-387, 428 y 433-436 del dominio CH3. Las referencias hechas a la numeración de aminoácidos de inmunoglobulinas o fragmentos de inmunoglobulinas, o regiones, se basan en Kabat et al. 1991, Sequences of Proteins of Immunological Interest, Departamento de Salud Pública de EE. UU., Bethesda, Maryland.

**[00284]** Las regiones Fc o las parejas de unión a FcRn unidas a FcRn pueden ser transportadas de manera efectiva a través de barreras epiteliales por FcRn, proporcionando así un medio no invasivo para administrar sistémicamente una molécula terapéutica deseada. Además, las proteínas de fusión que comprenden una región Fc o una pareja de unión a FcRn son endocitosadas por células que expresan el FcRn. Pero en lugar de estar marcadas por la degradación, estas proteínas de fusión se reciclan nuevamente en circulación, lo que aumenta la semivida in vivo de estas proteínas. En ciertas realizaciones, las porciones de las regiones constantes de inmunoglobulina son una región Fc o una pareja de unión a FcRn que típicamente se asocia, mediante enlaces disulfuro y otras interacciones no específicas, con otra región Fc u otra pareja de unión a FcRn para formar dímeros y multímeros de orden superior.

**[00285]** Dos receptores de FcRn pueden unirse a una sola molécula de Fc. Los datos cristalográficos sugieren que cada molécula de FcRn se une a un único polipéptido del homodímero de Fc. En una realización, la unión de la pareja de unión a FcRn, por ejemplo, un fragmento Fc de una IgG, a una molécula biológicamente activa proporciona un medio para administrar la molécula biológicamente activa por vía oral, bucal, sublingual, rectal, vaginal, como un aerosol administrado por vía nasal o vía una ruta pulmonar o por vía ocular. En otra realización, la proteína del factor de coagulación se puede administrar de forma invasiva, por ejemplo, por vía subcutánea o intravenosa.

**[00286]** Una región de pareja de unión a FcRn es una molécula o porción de la misma que puede unirse específicamente por el receptor FcRn con el consiguiente transporte activo por el receptor FcRn de la región Fc. Específicamente unido se refiere a dos moléculas que forman un complejo que es relativamente estable en

condiciones fisiológicas. La unión específica se caracteriza por una alta afinidad y una capacidad de baja a moderada, a diferencia de la unión no específica, que generalmente tiene una baja afinidad con una capacidad de moderada a alta. Típicamente, la unión se considera específica cuando la constante de afinidad  $K_A$  es superior a  $10^6 \text{ M}^{-1}$ , o superior a  $10^8 \text{ M}^{-1}$ . Si es necesario, la unión no específica se puede reducir sin afectar sustancialmente la unión específica variando las condiciones de unión. Cualquier experto en la técnica puede optimizar las condiciones de unión apropiadas usando técnicas rutinarias, tales como la concentración de las moléculas, la fuerza iónica de la solución, la temperatura, el tiempo permitido para la unión, la concentración de un agente bloqueante (por ejemplo, albúmina sérica, caseína de leche), etc.

**[00287]** En ciertas realizaciones, una proteína terapéutica codificada por la molécula de ácido nucleico de la descripción comprende una o más regiones Fc truncadas que, sin embargo, son suficientes para conferir propiedades de unión al receptor Fc (FcR) a la región Fc. Por ejemplo, la porción de una región Fc que se une a FcRn (es decir, la porción de unión a FcRn) comprende de aproximadamente los aminoácidos 282-438 de IgG1, numeración EU (siendo los sitios de contacto primarios los aminoácidos 248, 250-257, 272, 285, 288, 290-291, 308-311 y 314 del dominio CH2 y los residuos de aminoácidos 385-387, 428 y 433-436 del dominio CH3). Por lo tanto, una región Fc de la descripción puede comprender o consistir en una porción de unión a FcRn. Las porciones de unión a FcRn pueden derivarse de cadenas pesadas de cualquier isotipo, que incluyen IgG1, IgG2, IgG3 e IgG4. En una realización, se usa una porción de unión a FcRn de un anticuerpo del isotipo humano IgG1. En otra realización, se usa una porción de unión a FcRn de un anticuerpo del isotipo humano IgG4.

**[00288]** La región Fc se puede obtener de varias fuentes diferentes. En una realización, una región Fc del polipéptido se deriva de una inmunoglobulina humana. Sin embargo, se entiende que un resto Fc puede derivarse de una inmunoglobulina de otra especie de mamífero, que incluye, por ejemplo, un roedor (por ejemplo, un ratón, rata, conejo, conejillo de Indias) o un primate no humano (por ejemplo, chimpancé, macacos). Además, el polipéptido de los dominios Fc o porciones del mismo pueden derivarse de cualquier clase de inmunoglobulina, incluyendo IgM, IgG, IgD, IgA e IgE, y cualquier isotipo de inmunoglobulina, incluyendo IgG1, IgG2, IgG3 e IgG4. En otra realización, se usa el isotipo humano IgG1.

**[00289]** En ciertas realizaciones, la variante Fc confiere un cambio en al menos una función efectora impartida por un resto Fc que comprende dicho dominio Fc natural (p. ej., una mejora o reducción en la capacidad de la región Fc de unirse a los receptores Fc (p. ej. FcγRI, FcγRII o FcγRIII) o proteínas del complemento (por ejemplo, C1q), o para desencadenar una citotoxicidad dependiente de anticuerpos (ADCC), fagocitosis o citotoxicidad dependiente del complemento (CDCC)). En otras realizaciones, la variante Fc proporciona un residuo de cisteína modificado.

**[00290]** La región Fc de la descripción puede emplear variantes de Fc reconocidas en la técnica que se sabe que imparten un cambio (por ejemplo, una mejora o reducción) en la función efectora y/o la unión de FcR o FcRn. Específicamente, una región Fc de la descripción puede incluir, por ejemplo, un cambio (por ejemplo, una sustitución) en una o más de las posiciones de aminoácidos descritas en las publicaciones internacionales PCT WO88/07089A1, WO96/14339A1, WO98/05787A1, WO98/23289A1, WO99/51642A1, W099/58572A1, WO00/09560A2, WO00/32767A1, WO00/42072A2, WO02/44215A2, WO02/02/919A2, WO03/0745692, WO04/02A2, 02/02/02/02/02/2014 WO04/074455A2, WO04/099249A2, WO05/040217A2, WO04/044859, WO05/070963A1, WO05/077981A2, WO05/092925A2, WO05/123780A2, WO06/019447A1, WO06/047350A2, y WO06/085967A2 y las publicaciones de patentes de EE.UU. US2007/0231329, US2007/0231329, US2007/0237765, US2007/0237766, US2007/0237767, US2007/0243188, US20070248603, US20070286859, US20080057056; o las Patentes de Estados Unidos 5.648.260; 5.739.277; 5.834.250; 5.869.046; 6.096.871; 6.121.022; 6.194.551; 6.242.195; 6.277.375; 6.528.624; 6.538.124; 6.737.056; 6.821.505; 6.998.253; 7.083.784; 7.404.956 y 7.317.091, cada una de las cuales se incorpora en este documento como referencia. En una realización, el cambio específico (por ejemplo, la sustitución específica de uno o más aminoácidos descritos en la técnica) puede realizarse en una o más de las posiciones de aminoácidos descritas. En otra realización, se puede hacer un cambio diferente en una o más de las posiciones de aminoácidos descritas (por ejemplo, la sustitución diferente de una o más posiciones de aminoácidos descritas en la técnica).

**[00291]** La región Fc o la pareja de unión a FcRn de IgG pueden modificarse de acuerdo con procedimientos bien reconocidos tales como la mutagénesis dirigida al sitio y similares para producir fragmentos modificados de IgG o Fc o porciones de los mismos que se unirán por FcRn. Dichas modificaciones incluyen modificaciones alejadas de los sitios de contacto de FcRn, así como modificaciones dentro de los

sitios de contacto que preservan o incluso mejoran la unión al FcRn. Por ejemplo, los siguientes residuos de aminoácidos únicos en IgG1 Fc humana (Fcγ1) se pueden sustituir sin pérdida significativa de afinidad de unión a Fc por FcRn: P238A, S239A, K246A, K248A, D249A, M252A, T256A, E258A, T260A, D265A, S267A, H268A, E269A, D270A, E272A, L274A, N276A, Y278A, D280A, V282A, E283A, H285A, N286A, T289A, K290A, R292A, E293A, E294A, Q295A, Y296F, N297A, S298A, Y300F, R301A, V303A, V305A, T307A, L309A, Q311A, D312A, N315A, K317A, E318A, K320A, K322A, S324A, K326A, A327Q, P329A, A330Q, P331A, E333A, K334A, T335A, S337A, K338A, K340A, Q342A, R344A, E345A, Q347A, R355A, E356A, M358A, T359A, K360A, N361A, Q362A, Y373A, S375A, D376A, A378Q, E380A, E382A, S383A, N384A, Q386A, E388A, N389A, N390A, Y391F, K392A, L398A, S400A, D401A, D413A, K414A, R416A, Q418A, Q419A, N421A, V422A, S424A, E430A, N434A, T437A, Q438A, K439A, S440A, S444A y K447A, donde, por ejemplo, P238A representa la posición de prolina natural sustituida por alanina en la posición número 238. Como ejemplo, una realización específica incorpora la mutación N297A, eliminando un sitio de N-glucosilación altamente conservado. Además de la alanina, se pueden sustituir otros aminoácidos por los aminoácidos naturales en las posiciones especificadas anteriormente. Las mutaciones se pueden introducir individualmente en Fc dando lugar a más de cien regiones de Fc distintas de la Fc nativa. Además, las combinaciones de dos, tres o más de estas mutaciones individuales se pueden introducir juntas, dando lugar a cientos de regiones Fc más.

**[00292]** Algunas de las mutaciones anteriores pueden conferir una nueva funcionalidad a la región Fc o pareja de unión de FcRn. Por ejemplo, una realización incorpora N297A, eliminando un sitio de N-glucosilación altamente conservado. El efecto de esta mutación es reducir la inmunogenicidad, mejorando así la semivida circulante de la región Fc, y hacer que la región Fc sea incapaz de unirse a FcγRI, FcγRIIA, FcγRIIB y FcγRIIIA, sin comprometer la afinidad por FcRn (Routledge et al. 1995, Transplantation 60: 847; Friend et al. 1999, Transplantation 68: 1632; Shields et al. 1995, J. Biol. Chem 276:6591). Como otro ejemplo de la nueva funcionalidad que surge de las mutaciones descritas anteriormente, la afinidad por FcRn puede incrementarse más allá de la natural en algunos casos. Este aumento de la afinidad puede reflejar un aumento de la tasa de "activación", una disminución de la tasa de "desactivación" o tanto un aumento de la tasa de "activación" como una disminución de la tasa de "desactivación". Los ejemplos de mutaciones que se cree que imparten

una mayor afinidad por FcRn incluyen, entre otros, T256A, T307A, E380A y N434A (Shields et al. 2001, J. Biol. Chem 276:6591).

**[00293]** Además, al menos tres receptores gamma Fc humanos parecen reconocer un sitio de unión en IgG dentro de la región bisagra inferior, generalmente los aminoácidos 234-237. Por lo tanto, otro ejemplo de nueva funcionalidad y potencial inmunogenicidad disminuida puede surgir de mutaciones de esta región, como, por ejemplo, reemplazando los aminoácidos 233-236 de IgG1 humano "ELLG" (SEQ ID NO: 45) con la secuencia correspondiente de IgG2 "PVA" (con una delección de aminoácidos). Se ha demostrado que FcγRI, FcγRII y FcγRIII, que median diversas funciones efectoras, no se unirán a IgG1 cuando se hayan introducido tales mutaciones. Ward y Ghetie 1995, Therapeutic Immunology 2:77 y Armour et al. 1999, Eur. J. Immunol. 29:2613.

**[00294]** En otra realización, la región constante de inmunoglobulina o una porción de la misma comprende una secuencia de aminoácidos en la región bisagra o una porción de la misma que forma uno o más enlaces disulfuro con una segunda región constante de inmunoglobulina o una porción de la misma. La segunda región constante de inmunoglobulina o una porción de la misma se puede unir a un segundo polipéptido, uniendo la proteína terapéutica y el segundo polipéptido. En algunas realizaciones, el segundo polipéptido es un resto potenciador. Como se usa en el presente documento, la expresión "resto potenciador" se refiere a una molécula, fragmento de la misma o un componente de un polipéptido que es capaz de potenciar la actividad de la proteína terapéutica. El resto potenciador puede ser un cofactor, tal como uno en el que la proteína terapéutica sea un factor de coagulación, un factor de tejido soluble (sTF) o un péptido procoagulante. Por lo tanto, tras la activación del factor de coagulación, el resto potenciador estará disponible para mejorar la actividad del factor de coagulación.

**[00295]** En ciertas realizaciones, una proteína terapéutica codificada por una molécula de ácido nucleico de la descripción comprende una sustitución de aminoácidos en una región constante de inmunoglobulina o una porción de la misma (por ejemplo, variantes Fc), que modifica las funciones efectoras independientes del antígeno de la región constante de Ig, en particular la semivida circulante de la proteína.

## **2. Regiones scFc**

**[00296]** En otro aspecto, un resto heterólogo comprende una región scFc (Fc monocatenario). En una realización, una molécula de ácido nucleico aislada de la

descripción comprende además una secuencia de ácido nucleico heterólogo que codifica una región scFc. La región scFc comprende al menos dos regiones constantes de inmunoglobulina o porciones de las mismas (por ejemplo, restos Fc o dominios (por ejemplo, 2, 3, 4, 5, 6 o más restos o dominios Fc)) dentro de la misma cadena de polipéptidos lineal que son capaces de plegamiento (por ejemplo, plegamiento intramolecular o intermolecular) para formar una región scFc funcional que está unida por un enlazante peptídico Fc. Por ejemplo, en una realización, un polipéptido de la descripción es capaz de unirse, a través de su región scFc, a al menos un receptor Fc (por ejemplo, un FcRn, un receptor FcγR (por ejemplo, FcγRIII) o una proteína del complemento (por ejemplo, C1q)) para mejorar la semivida o desencadenar una función efectora inmune (por ejemplo, citotoxicidad dependiente de anticuerpos (ADCC), fagocitosis o citotoxicidad dependiente del complemento (CDCC) y/o mejorar la capacidad de fabricación).

### 3. CTP

**[00297]** En otro aspecto, un resto heterólogo comprende un péptido C-terminal (CTP) de la subunidad β de la gonadotropina coriónica humana o fragmento, variante o derivado de la misma. Se sabe que uno o más péptidos CTP insertados en una proteína recombinante aumentan la semivida in vivo de esa proteína. Véase, por ejemplo, la Patente de los Estados Unidos Núm. 5.712.122, incorporada en este documento como referencia en su totalidad.

**[00298]** Los péptidos CTP ejemplares incluyen DPRFQDSSSSKAPPSLPSPSRLPGPSDTPIL (SEQ ID NO: 33) o SSSSKAPPSLPSPSRLPGPSDTPILPQ (SEQ ID NO: 34). Véase, por ejemplo, la Publicación de Solicitud de Patente de los Estados Unidos Núm. US 2009/0087411 A1, incorporada como referencia.

### 4. Secuencia XTEN

**[00299]** En algunas realizaciones, un resto heterólogo comprende una o más secuencias XTEN, fragmentos, variantes o derivados de las mismas. Tal como se usa en este documento, la "secuencia XTEN" se refiere a polipéptidos de longitud extendida con secuencias no repetitivas que se producen de manera no natural que están compuestas principalmente de aminoácidos hidrófilos pequeños, con la secuencia que tiene un bajo grado o ninguna estructura secundaria o terciaria en condiciones fisiológicas. Como un resto heterólogo, los XTEN pueden servir como un resto de extensión de semivida. Además, XTEN puede proporcionar propiedades

deseables que incluyen, entre otras, parámetros farmacocinéticos mejorados y características de solubilidad.

**[00300]** La incorporación de un resto heterólogo que comprende una secuencia XTEN en una proteína de la descripción puede conferir a la proteína una o más de las siguientes propiedades ventajosas: flexibilidad conformacional, solubilidad acuosa mejorada, alto grado de resistencia a la proteasa, baja inmunogenicidad, baja unión a mamíferos receptores o aumento de radios hidrodinámicos (o de Stokes).

**[00301]** En ciertos aspectos, una secuencia XTEN puede aumentar las propiedades farmacocinéticas, como una semivida in vivo más larga o un área aumentada bajo la curva (AUC), de modo que alguna proteína de la descripción permanezca in vivo y tenga actividad procoagulante durante un período de tiempo mayor en comparación con alguna proteína con lo mismo pero sin el resto heterólogo XTEN.

**[00302]** En algunas realizaciones, la secuencia XTEN útil para la descripción es un péptido o un polipéptido que tiene más de aproximadamente 20, 30, 40, 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1200, 1400, 1600, 1800 o 2000 residuos de aminoácidos. En ciertas realizaciones, XTEN es un péptido o un polipéptido que tiene más de aproximadamente 20 a aproximadamente 3000 residuos de aminoácidos, más de 30 a aproximadamente 2500 residuos, más de 40 a aproximadamente 2000 residuos, más de 50 a aproximadamente 1500 residuos, más de 60 a aproximadamente 1000 residuos, más de 70 a aproximadamente 900 residuos, más de 80 a aproximadamente 800 residuos, más de 90 a aproximadamente 700 residuos, más de 100 a aproximadamente 600 residuos, más de 110 a aproximadamente 500 residuos o más de 120 a aproximadamente 400 residuos. En una realización particular, el XTEN comprende una secuencia de aminoácidos de más de 42 aminoácidos y de menos de 144 aminoácidos de longitud.

**[00303]** La secuencia XTEN de la descripción puede comprender uno o más motivos de secuencia de 5 a 14 (por ejemplo, 9 a 14) residuos de aminoácidos o una secuencia de aminoácidos al menos 80%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% o 99% idéntica al motivo de secuencia, en donde el motivo comprende, consiste esencialmente o consiste en sí mismo entre 4 y 6 tipos de aminoácidos (por ejemplo, 5 aminoácidos) seleccionados del grupo que consiste en glicina (G), alanina (A), serina (S), treonina (T), glutamato (E) y prolina (P). Ver el documento US 2010-0239554 A1.

**[00304]** En algunas realizaciones, el XTEN comprende motivos de secuencia no superpuestos en los que aproximadamente 80%, o al menos aproximadamente 85%, o al menos aproximadamente 90%, o aproximadamente 91%, o aproximadamente 92%, o aproximadamente 93%, o aproximadamente 94 %, o aproximadamente 95%, o aproximadamente 96%, o aproximadamente 97%, o aproximadamente 98%, o aproximadamente 99% o aproximadamente 100% de la secuencia consiste en múltiples unidades de secuencias no superpuestas seleccionadas de una única familia de motivos seleccionada de la Tabla 7, que resulta en una secuencia familiar. Como se usa en el presente documento, "familia" significa que el XTEN tiene motivos seleccionados solo de una categoría de motivos única de la Tabla 7; es decir, AD, AE, AF, AG, AM, AQ, BC o BD XTEN, y que cualquier otro aminoácido en XTEN que no sea de un motivo familiar se selecciona para lograr una propiedad necesaria, como para permitir la incorporación de un sitio de restricción por los nucleótidos codificadores, incorporación de una secuencia de escisión, o para lograr un mejor enlace con la proteína terapéutica. En algunas realizaciones de familias XTEN, una secuencia XTEN comprende múltiples unidades de motivos de secuencia no superpuestos de la familia de motivos AD, o de la familia de motivos AE, o de la familia de motivos AF, o de la familia de motivos AG, o de la familia de motivos AM, o de la familia de motivos AQ, o de la familia de BC, o de la familia de BD, exhibiendo la XTEN resultante el intervalo de homología descrito anteriormente. En otras realizaciones, el XTEN comprende múltiples unidades de secuencias de motivos de dos o más de las familias de motivos de la Tabla 7. Estas secuencias se pueden seleccionar para lograr las características físicas/químicas deseadas, incluidas propiedades tales como carga neta, hidrofilia, falta de estructura secundaria o falta de repetitividad que confiere la composición de aminoácidos de los motivos, que se describe más detalladamente a continuación. En las realizaciones descritas anteriormente en este párrafo, los motivos incorporados en el XTEN pueden seleccionarse y ensamblarse usando los métodos descritos en este documento para lograr un XTEN de aproximadamente 36 a aproximadamente 3000 residuos de aminoácidos.

Tabla 7. Motivos de secuencia XTEN de 12 aminoácidos y familias de motivos

| <b>Familia<br/>Motivo*</b> | <b>SECUENCIA MOTIVO</b> | <b>SEQ ID NO:</b> |
|----------------------------|-------------------------|-------------------|
| AD                         | GESPGGSSGSES            | 73                |
| AD                         | GSEGSSGPGESE            | 74                |

| <b>Familia<br/>Motivo*</b> | <b>SECUENCIA MOTIVO</b> | <b>SEQ ID NO:</b> |
|----------------------------|-------------------------|-------------------|
| AD                         | GSSESGSSEGGP            | 75                |
| AD                         | GSGGEPSESGSS            | 76                |
| AE, AM                     | GSPAGSPTSTEE            | 77                |
| AE, AM, AQ                 | GSEPATSGSETP            | 78                |
| AE, AM, AQ                 | GTSESATPESGP            | 79                |
| AE, AM, AQ                 | GTSTEPSEGSAP            | 80                |
| AF, AM                     | GSTSESPSGTAP            | 81                |
| AF, AM                     | GTSTPESGSASP            | 82                |
| AF, AM                     | GTSPSGESSTAP            | 83                |
| AF, AM                     | GSTSSTAESPGP            | 84                |
| AG, AM                     | GTPGSGTASSSP            | 85                |
| AG, AM                     | GSSTPSGATGSP            | 86                |
| AG, AM                     | GSSPSASTGTGP            | 87                |
| AG, AM                     | GASPGTSSTGSP            | 88                |
| AQ                         | GEPAGSPTSTSE            | 89                |
| AQ                         | GTGEPSSTPASE            | 90                |
| AQ                         | GSGPSTESAPTE            | 91                |
| AQ                         | GSETPSGPSETA            | 92                |
| AQ                         | GPSETSTSEPGA            | 93                |
| AQ                         | GSPSEPTEGTSA            | 94                |
| BC                         | GSGASEPTSTEP            | 95                |
| BC                         | GSEPATSGTEPS            | 96                |
| BC                         | GTSEPSTSEPGA            | 97                |
| BC                         | GTSTEPSEPGSA            | 98                |
| BD                         | GSTAGSETSTE A           | 99                |
| BD                         | GSETATSGSETA            | 100               |
| BD                         | GTSESATSESGA            | 101               |
| BD                         | GTSTEASEGSAS            | 102               |

\* Denota secuencias de motivos individuales que, cuando se usan juntas en diversas permutaciones, dan como resultado una "secuencia familiar"

**[00305]** Se describen ejemplos de secuencias XTEN que se pueden usar como restos heterólogos en las proteínas terapéuticas de la descripción, por ejemplo, en las Publicaciones de patentes de EE.UU. Nos. 2010/0239554 AI, 2010/0323956 AI, 2011/0046060 AI, 2011/0046061 AI, 2011/0077199 AI, o 2011/0172146 AI, o las publicaciones internacionales de patentes núms. WO 2010091122 AI, WO 2010144502 A2, WO 2010144508 AI, WO 2011028228 AI, WO 2011028229 AI o WO 2011028344 A2, cada una de los cuales se incorpora en este documento como referencia en su totalidad.

**[00306]** XTEN puede tener diferentes longitudes para su inserción o unión a una proteína terapéutica. En una realización, la longitud de la(s) secuencia(s) XTEN se elige en función de la propiedad o función a alcanzar en la proteína de fusión. Dependiendo de la propiedad o función prevista, XTEN puede ser una secuencia de longitud corta o intermedia o una secuencia más larga que puede servir como portadores. En ciertas realizaciones, el XTEN incluye segmentos cortos de aproximadamente 6 a aproximadamente 99 residuos de aminoácidos, longitudes intermedias de aproximadamente 100 a aproximadamente 399 residuos de aminoácidos y longitudes más largas de aproximadamente 400 a aproximadamente 1000 y hasta aproximadamente 3000 residuos de aminoácidos. Por lo tanto, el XTEN insertado o unido a una proteína terapéutica puede tener longitudes de aproximadamente 6, aproximadamente 12, aproximadamente 36, aproximadamente 40, aproximadamente 42, aproximadamente 72, aproximadamente 96, aproximadamente 144, aproximadamente 288, aproximadamente 400, aproximadamente 500, aproximadamente 576, aproximadamente 600, aproximadamente 700, aproximadamente 800, aproximadamente 864, aproximadamente 900, aproximadamente 1000, aproximadamente 1500, aproximadamente 2000, aproximadamente 2500, o hasta aproximadamente 3000 residuos de aminoácidos de longitud. En otras realizaciones, las secuencias XTEN son de aproximadamente 6 a aproximadamente 50, aproximadamente 50 a aproximadamente 100, aproximadamente 100 a 150, aproximadamente 150 a 250, aproximadamente 250 a 400, aproximadamente 400 a aproximadamente 500, aproximadamente 500 a aproximadamente 900, aproximadamente 900 a 1500, aproximadamente 1500 a 2000, o aproximadamente 2000 a aproximadamente 3000 residuos de aminoácidos de longitud. La longitud precisa de una XTEN insertada o unida a una proteína terapéutica puede variar sin afectar negativamente a la actividad de la proteína terapéutica. En una realización, uno o más de los XTEN usados en este

documento tienen 42 aminoácidos, 72 aminoácidos, 144 aminoácidos, 288 aminoácidos, 576 aminoácidos u 864 aminoácidos de longitud y pueden seleccionarse de uno o más de las secuencias familiares XTEN; es decir, AD, AE, AF, AG, AM, AQ, BC o BD.

- 5 **[00307]** En algunas realizaciones, la proteína terapéutica comprende un polipéptido FVIII y un XTEN, en donde el XTEN comprende 288 aminoácidos. En una realización, la proteína terapéutica comprende un polipéptido FVIII y un XTEN, en donde el XTEN comprende 288 aminoácidos, y el XTEN se inserta dentro del dominio B del polipéptido FVIII. En una realización particular, la proteína terapéutica
- 10 comprende un polipéptido FVIII y un XTEN que comprende la SEQ ID N°: 109, y el XTEN se inserta dentro del dominio B del polipéptido FVIII. En una realización particular, la proteína terapéutica comprende un polipéptido FVIII y un XTEN que comprende la SEQ ID N°: 109, y el XTEN se inserta dentro del polipéptido FVIII inmediatamente después del aminoácido 745 del FVIII maduro.
- 15 **[00308]** En algunas realizaciones, la proteína terapéutica comprende un polipéptido FIX y un XTEN, en donde el XTEN comprende 72 aminoácidos. En una realización, la proteína terapéutica comprende un polipéptido FIX y un XTEN, en donde el XTEN comprende 72 aminoácidos, y el XTEN se inserta dentro del polipéptido FIX inmediatamente después del aminoácido 166 del FIX maduro.
- 20 **[00309]** En algunas realizaciones, la secuencia XTEN usada en la descripción es al menos un 60%, 70%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% o 100% idéntica con una secuencia seleccionada del grupo que consiste en AE42, AG42, AE48, AM48, AE72, AG72, AE108, AG108, AE144, AF144, AG144, AE180, AG180, AE216, AG216, AE252, AG252, AE288, AG288, AE324, AG324,
- 25 AE360, AG360, AE396, AG396, AE432, AG432, AE468, AG468, AE504, AG504, AF504, AE540, AG540, AF540, AD576, AE576, AF576, AG576, AE612, AG612, AE624, AE648, AG648, AG684, AE720, AG720, AE756, AG756, AE792, AG792, AE828, AG828, AD836, AE864, AF864, AG864, AM875, AE912, AM923, AM1318, BC864, BD864, AE948, AE1044, AE1140, AE1236, AE1332, AE1428, AE1524,
- 30 AE1620, AE1716, AE1812, AE1908, AE2004A, AG948, AG1044, AG1140, AG1236, AG1332, AG1428, AG1524, AG1620, AG1716, AG1812, AG1908, AG2004, y cualquiera de sus combinaciones. Ver US 2010-0239554 A1. En una realización particular, el XTEN comprende AE42, AE72, AE144, AE288, AE576, AE864, AG 42, AG72, AG144, AG288, AG576, AG864, o cualquiera de sus combinaciones.

**[00310]** Las secuencias XTEN ejemplares que se pueden usar como restos heterólogos en la proteína terapéutica de la descripción incluyen XTEN AE42-4 (SEQ ID NO: 46, codificado por SEQ ID NO: 47), XTEN AE144-2A (SEQ ID NO: 48, codificado por SEQ ID NO: 49), XTEN AE144-3B (SEQ ID NO: 50, codificado por SEQ ID NO: 51), XTEN AE144-4A (SEQ ID NO: 52, codificado por SEQ ID NO: 53), XTEN AE144-5A (SEQ ID NO: 54, codificado por SEQ ID NO: 55), XTEN AE144-6B (SEQ ID NO: 56, codificado por SEQ ID NO: 57), XTEN AG144-1 (SEQ ID NO: 58, codificado por SEQ ID NO: 59), XTEN AG144-A (SEQ ID NO: 60, codificado por SEQ ID NO: 61), XTEN AG144-B (SEQ ID NO: 62, codificado por SEQ ID NO: 63), XTEN AG144-C (SEQ ID NO: 64, codificado por SEQ ID NO: 65), y XTEN AG144-F (SEQ ID NO: 66, codificado por SEQ ID NO: 67). En una realización particular, el XTEN está codificado por la SEQ ID NO: 18.

**[00311]** En otra realización, la secuencia XTEN se selecciona del grupo que consiste en AE36 (SEQ ID NO: 130), AE42 (SEQ ID NO: 131), AE72 (SEQ ID NO: 132), AE78 (SEQ ID NO: 133), AE144 (SEQ ID NO: 134), AE144\_2A (SEQ ID NO: 48), AE144\_3B (SEQ ID NO: 50), AE144\_4A (SEQ ID NO: 52), AE144\_5A (SEQ ID NO: 54), AE144\_6B (SEQ ID NO: 135), AG144 (SEQ ID NO: 136), AG144\_A (SEQ ID NO: 137), AG144\_B (SEQ ID NO: 62), AG144\_C (SEQ ID NO: 64), AG144\_F (SEQ ID NO: 66), AE288 (SEQ ID NO: 138), AE288\_2 (SEQ ID NO: 139), AG288 (SEQ ID NO: 140), AE576 (SEQ ID NO: 141), AG576 (SEQ ID NO: 142), AE864 (SEQ ID NO: 143), AG864 (SEQ ID NO: 144), XTEN\_AE72\_2A\_1 (SEQ ID NO: 145), XTEN\_AE72\_2A\_2 (SEQ ID NO: 146), XTEN\_AE72\_3B\_1 (SEQ ID NO: 147), XTEN\_AE72\_3B\_2 (SEQ ID NO: 148), XTEN\_AE72\_4A\_2 (SEQ ID NO: 149), XTEN\_AE72\_5A\_2 (SEQ ID NO: 150), XTEN\_AE72\_6B\_1 (SEQ ID NO: 151), XTEN\_AE72\_6B\_2 (SEQ ID NO: 152), XTEN\_AE72\_1A\_1 (SEQ ID NO: 153), XTEN\_AE72\_1A\_2 (SEQ ID NO: 154), XTEN\_AE144\_1A (SEQ ID NO: 155), AE150 (SEQ ID NO: 156), AG150 (SEQ ID NO: 157), AE294 (SEQ ID NO: 158), AG294 (SEQ ID NO: 159), y cualquiera de sus combinaciones. En una realización específica, la secuencia XTEN se selecciona del grupo que consiste en AE72, AE144 y AE288. Las secuencias de aminoácidos para ciertas secuencias XTEN de la invención se muestran en la Tabla 8.

**Tabla 8. Secuencias XTEN**

| XTEN | Secuencia de aminoacidos |
|------|--------------------------|
|------|--------------------------|

|                          |                                                                                                                                                   |
|--------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------|
| AE42-4 (SEQ ID NO: 46)   | GAPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPASS                                                                                                        |
| AE144-2A (SEQ ID NO: 48) | TSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPG                          |
| A144-3B (SEQ ID NO: 50)  | SPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPG              |
| AE144-4A (SEQ ID NO: 52) | TSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPG  |
| AE144-5A (SEQ ID NO: 54) | TSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSSESATPESGPGSEPATSGSETPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEG |
| AE144-6B (SEQ ID NO: 56) | TSTEPSEGSAPGTSESATPESGPGTSESATPESGPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSSTEPSEGSAPGTSTEPSEGSAPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPG |
| AG144-1 (SEQ ID NO: 58)  | PGSSPSASTGTGPGSSPSASTGTGPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSS  |
| AG144-A (SEQ ID NO: 60)  | GASPGTSSTGSPGSSPSASTGTGPGSSPSASTGTGPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSP  |

|                              |                                                                                                                                                              |
|------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------|
| AG144-B (SEQ ID NO: 62)      | GTPGSGTASSSPGSSTPSGATGSPGASPGTSSTGSPGTPGSGT<br>ASSSPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGS<br>STPSGATGSPGSSTPSGATGSPGASPGTSSTGSPGASPGTSST<br>GSPGASPGTSSTGSP |
| AG144-C (SEQ ID NO: 64)      | GTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGASPGTS<br>STGSPGSSPSASTGTGPGTPGSGTASSSPGASPGTSSTGSPGA<br>SPGTSSTGSPGASPGTSSTGSPGSSTPSGATGSPGSSTPSGAT<br>GSPGASPGTSSTGSP |
| XTEN AG144-F (SEQ ID NO: 66) | GSSPSASTGTGPGSSPSASTGTGPGASPGTSSTGSPGASPGTS<br>STGSPGSSTPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGS<br>SPSASTGTGPGTPGSGTASSSPGSSTPSGATGSPGSSTPSGAT<br>GSPGASPGTSSTGSP |
| AE36 (SEQ ID NO: 130)        | GSPAGSPTSTEEGTSESATPESGPGSEPATSGSETP                                                                                                                         |
| AE42 (SEQ ID NO: 131)        | GAPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPASS                                                                                                                   |
| AE72 (SEQ ID NO: 132)        | GAPTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPA<br>TSGSETPGTSESATPESGPGTSTEPSEGSAPGASS                                                                           |
| AE78 (SEQ ID NO: 133)        | GAPTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPA<br>TSGSETPGTSESATPESGPGTSTEPSEGSAPGASS                                                                           |
| AE144 (SEQ ID NO: 134)       | GSEPATSGSETPGTSESATPESGPGSEPATSGSETPGSPAGSP<br>TSTEEGTSTEPSEGSAPGSEPATSGSETPGSEPATSGSETPGS<br>EPATSGSETPGTSTEPSEGSAPGTSESAPESGPGSEPATSGSE<br>TPGTSTEPSEGSAP  |

|                                     |                                                                                                                                                                                                                                                                                                                          |
|-------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| AE144_6B<br><br>(SEQ ID NO:<br>135) | TSTEPSEGSAPGTSESATPESGPGTSESATPESGPGTSESATP<br>ESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTS<br>TEPSEGSAPGTSTEPSEGSAPGSEPATSGSETPGTSESATPES<br>GPGTSTEPSEGSAPG                                                                                                                                                             |
| AG144<br><br>(SEQ ID<br>NO:136)     | GTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGSSPSAS<br>TGTGPGASPGTSSTGSPGASPGTSSTGSPGSSTPSGATGSPGS<br>SPSASTGTGPGASPGTSSTGSPGSSPSASTGTGPGTPGSGTAS<br>SSPGSSTPSGATGSP                                                                                                                                                             |
| AG144_A<br><br>(SEQ ID NO:<br>137)  | GASPGTSSTGSPGSSPSASTGTGPGSSPSASTGTGPGTPGSGT<br>ASSSPGSSTPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGT<br>PGSGTASSSPGSSTPSGATGSPGTPGSGTASSSPGASPGTSST<br>GSPGASPGTSSTGSP                                                                                                                                                             |
| AE288<br><br>(SEQ ID NO:<br>138)    | GTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPAT<br>GSETPGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGT<br>SESATPESGPGSEPATSGSETPGTSESATPESGPGSPAGSPTS<br>TEEGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSE<br>SATPESGPGTSESATPESGPGSEPATSGSETPGSEPATSGSET<br>PGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGSEPAT<br>SGSETPGTSESATPESGPGTSTEPSEGSAP  |
| AE288_2<br><br>(SEQ ID NO:<br>139)  | GSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESAT<br>PESGPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGT<br>STEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTS<br>TEEGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSE<br>SATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSA<br>PGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGSPAGS<br>PTSTEEGTSESATPESGPGTSTEPSEGSAP |
| AG288<br><br>(SEQ ID NO:<br>140)    | PGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPS<br>GATGSPGTPGSGTASSSPGSSTPSGATGSPGTPGSGTASSSPG<br>SSTPSGATGSPGSSTPSGATGSPGSSPSASTGTGPGSSPSAST<br>GTGPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGSS<br>PSASTGTGPGSSPSASTGTGPGASPGTSSTGSPGASPGTSSTG                                                                                  |

|                                  |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      |
|----------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|                                  | SPGSSTPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGSSPS<br>ASTGTGPGTPGSGTASSSPGSSTPSGATGS                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |
| AE576<br><br>(SEQ ID NO:<br>141) | GSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSP<br>TSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGS<br>EPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSESATPE<br>SGPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTST<br>EPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSTEPSEGS<br>PGTSESATPESGPGSEPATSGSETPGTSTEPSEGSAPGTSTEP<br>SEGSAPGTSESATPESGPGTSESATPESGPGSPAGSPTSTEEG<br>TSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSE<br>GSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTS<br>TEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGS<br>APGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPA<br>TSGSETPGTSESATPESGPGTSTEPSEGSAPGTSESATPESGP<br>GSPAGSPTSTEEGSPAGSPTSTEEGSPAGSPTSTEEGTSESAT<br>PESGPGTSTEPSEGSAP  |
| AG576<br><br>(SEQ ID NO:<br>142) | PGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGSSPSA<br>STGTGPGSSTPSGATGSPGSSTPSGATGSPGASPGTSSTGSPG<br>ASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGASPGTSS<br>TGSPGASPGTSSTGSPGASPGTSSTGSPGSSPSASTGTGPGTP<br>GSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTG<br>SPGSSTPSGATGSPGSSTPSGATGSPGASPGTSSTGSPGTPGS<br>GTASSSPGSSTPSGATGSPGSSTPSGATGSPGSSTPSGATGSP<br>GSSPSASTGTGPGASPGTSSTGSPGASPGTSSTGSPGTPGSGT<br>ASSSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGA<br>SPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGTPGSGTAS<br>SSPGSSTPSGATGSPGTPGSGTASSSPGSSTPSGATGSPGSST<br>PSGATGSPGSSPSASTGTGPGSSPSASTGTGPGASPGTSSTGS<br>PGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGSSPSA<br>STGTGPGASPGTSSTGS |
| AE864                            | GSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSP<br>TSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGS<br>EPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSESATPE                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |

|                                            |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             |
|--------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <p>(SEQ ID NO:<br/>143)</p>                | <p>SGPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTST<br/>EPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSTEPSEGSA<br/>PGTSESATPESGPGSEPATSGSETPGTSTEPSEGSAPGTSTEP<br/>SEGSAPGTSESATPESGPGTSESATPESGPGSPAGSPTSTEEG<br/>TSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSE<br/>GSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTS<br/>TEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGS<br/>APGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPA<br/>TSGSETPGTSESATPESGPGTSTEPSEGSAPGTSESATPESGP<br/>GSPAGSPTSTEEGSPAGSPTSTEEGSPAGSPTSTEEGTSESAT<br/>PESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGT<br/>SESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEG<br/>SAPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSE<br/>SATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSA<br/>PGTSESATPESGPGTSESATPESGPGTSESATPESGPGSEPAT<br/>SGSETPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPG<br/>TSTEPSEGSAPGSEPATSGSETPGTSESATPESGPGTSTEPSE<br/>GSAP</p> |
| <p>AG864<br/><br/>(SEQ ID NO:<br/>144)</p> | <p>GASPGTSSTGSPGSSPSASTGTGPGSSPSASTGTGPGTPGSGT<br/>ASSSPGSSTPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGT<br/>PGSGTASSSPGSSTPSGATGSPGTPGSGTASSSPGASPGTSST<br/>GSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGASP<br/>GTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTG<br/>PGSSPSASTGTGPGSSTPSGATGSPGSSTPSGATGSPGASPGT<br/>SSTGSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPG<br/>ASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGSSPSAST<br/>GTGPGTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGAS<br/>PGTSSTGSPGSSTPSGATGSPGSSTPSGATGSPGASPGTSSTG<br/>SPGTPGSGTASSSPGSSTPSGATGSPGSSTPSGATGSPGSSTP<br/>SGATGSPGSSPSASTGTGPGASPGTSSTGSPGASPGTSSTGSP<br/>GTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGASPGTS<br/>STGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGT<br/>PGSGTASSSPGSSTPSGATGSPGTPGSGTASSSPGSSTPSGAT<br/>GSPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGASP</p>                                                          |

|                                                       |                                                                                                                                                                                                                          |
|-------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|                                                       | <p>GTSSSTGSPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTG</p> <p>PGSSPSASTGTGPGASPGTSSTGSPGASPGTSSTGSPGSSTPS</p> <p>GATGSPGSSPSASTGTGPGASPGTSSTGSPGSSPSASTGTGPG</p> <p>TPGSGTASSSPGSSTPSGATGSPGSSTPSGATGSPGASPGTSS</p> <p>TGSP</p> |
| <p>XTEN_AE72_2A</p> <p>_1</p> <p>(SEQ ID NO: 145)</p> | <p>TSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSE</p> <p>GSAPGTSESATPESGPGTSTEPSEGSAPG</p>                                                                                                                                  |
| <p>XTEN_AE72_2A</p> <p>_2</p> <p>(SEQ ID NO: 146)</p> | <p>TSESATPESGPGSEPATSGSETPGTSTEPSEGSAPGTSTEPSE</p> <p>GSAPGTSESATPESGPGTSESATPESGPG</p>                                                                                                                                  |
| <p>XTEN_AE72_3B</p> <p>_1</p> <p>(SEQ ID NO:147)</p>  | <p>SPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATP</p> <p>ESGPGTSTEPSEGSAPGTSTEPSEGSAPG</p>                                                                                                                                  |
| <p>XTEN_AE72_3B</p> <p>_2</p> <p>(SEQ ID NO: 148)</p> | <p>TSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSE</p> <p>GSAPGSPAGSPTSTEEGTSTEPSEGSAPG</p>                                                                                                                                  |
| <p>XTEN_AE72_4A</p> <p>_2</p> <p>(SEQ ID NO: 149)</p> | <p>TSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGSPAGSPT</p> <p>STEEGTSESATPESGPGTSTEPSEGSAPG</p>                                                                                                                                  |
| <p>XTEN_AE72_5A</p> <p>_2</p> <p>(SEQ ID NO: 150)</p> | <p>SPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATP</p> <p>ESGPGSPAGSPTSTEEGSPAGSPTSTEEG</p>                                                                                                                                  |

|                                               |                                                                                                                                                                    |
|-----------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| XTEN_AE72_6B<br>_1 (SEQ ID NO:<br>151)        | TSTEPSEGSAPGTSESATPESGPGTSESATPESGPGTSESATP<br>ESGPGSEPATSGSETPGSEPATSGSETPG                                                                                       |
| XTEN_AE72_6B<br>_2<br><br>(SEQ ID NO:<br>152) | SPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGSEPATSG<br>SETPGTSESATPESGPGTSTEPSEGSAPG                                                                                       |
| XTEN_AE72_1A<br>_1<br><br>(SEQ ID NO:<br>153) | SPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPT<br>STEEGTSTEPSEGSAPGTSTEPSEGSAPG                                                                                       |
| XTEN_AE72_1A<br>_2<br><br>(SEQ ID NO:<br>154) | TSESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPT<br>STEEGTSESATPESGPGTSTEPSEGSAPG                                                                                       |
| XTEN_AE144_1<br>A<br><br>(SEQ ID NO:<br>155)  | SPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPT<br>STEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGSE<br>PATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSESATPES<br>GPGTSTEPSEGSAPG       |
| AE150<br><br>(SEQ ID NO:<br>156)              | GAPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGSPA<br>GSPTSTEEGTSTEPSEGSAPGSEPATSGSETPGSEPATSGSET<br>PGSEPATSGSETPGTSTEPSEGSAPGTSESATPESGPGSEPAT<br>SGSETPGTSTEPSEGSAPASS |
| G150<br><br>(SEQ ID NO:<br>157)               | GAPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGSSP<br>SASTGTGPGASPGTSSTGSPGASPGTSSTGSPGSSTPSGATGS<br>PGSSPSASTGTGPGASPGTSSTGSPGSSPSASTGTGPGTPGSG<br>TASSSPGSSTPSGATGSPASS |
| AE294                                         | GAPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEP<br>ATSGSETPGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTE                                                                         |

|                                  |                                                                                                                                                                                                                                                                                                                                |
|----------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| (SEQ ID NO:<br>158)              | EGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSPAGS<br>PTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPG<br>TSESATPESGPGTSESATPESGPGSEPATSGSETPGSEPATSG<br>SETPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGSE<br>PATSGSETPGTSESATPESGPGTSTEPSEGSAPASS                                                                                               |
| AG294<br><br>(SEQ ID NO:<br>159) | GAPPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGSS<br>TPSGATGSPGTPGSGTASSSPGSSTPSGATGSPGTPGSGTASS<br>SPGSSTPSGATGSPGSSTPSGATGSPGSSPSASTGTGPGSSPS<br>ASTGTGPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSP<br>GSSPSASTGTGPGSSPSASTGTGPGASPGTSSTGSPGASPGTS<br>STGSPGSSTPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGS<br>SPSASTGTGPGTPGSGTASSSPGSSTPSGATGSASS |

**[00312]** En algunas realizaciones, menos del 100% de los aminoácidos de un XTEN se seleccionan de glicina (G), alanina (A), serina (S), treonina (T), glutamato (E) y prolina (P), o menos del 100% de la secuencia consiste en los motivos de secuencia de la Tabla 7 o una secuencia XTEN proporcionada en este documento. En tales realizaciones, los restos de aminoácidos restantes de XTEN se seleccionan de cualquiera de los otros 14 L-aminoácidos naturales, pero pueden seleccionarse preferentemente de aminoácidos hidrófilos de modo que la secuencia XTEN contenga al menos aproximadamente 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, o al menos aproximadamente 99% de aminoácidos hidrofílicos. El contenido de aminoácidos hidrófobos en el XTEN utilizado en las construcciones de conjugación puede ser inferior al 5%, o inferior al 2%, o inferior al 1% del contenido de aminoácidos hidrófobos. Los residuos hidrofóbicos menos favorecidos en la construcción de XTEN incluyen triptófano, fenilalanina, tirosina, leucina, isoleucina, valina y metionina. Además, las secuencias XTEN pueden contener menos del 5% o menos del 4% o menos del 3% o menos del 2% o menos del 1% o ninguno de los siguientes aminoácidos: metionina (por ejemplo, para evitar la oxidación) o asparagina y glutamina (para evitar la desamidación).

**[00313]** La una o más secuencias XTEN pueden insertarse en el extremo C o en el extremo N de la proteína terapéutica o insertarse entre dos aminoácidos en la secuencia de aminoácidos de la proteína terapéutica. Por ejemplo, cuando la proteína terapéutica comprende un polipéptido FVIII, el XTEN se puede insertar entre dos

aminoácidos en uno o más sitios de inserción seleccionados de la Tabla 5. Cuando la proteína terapéutica comprende un polipéptido FIX, el XTEN se puede insertar entre dos aminoácidos en uno o más sitios de inserción seleccionados de la Tabla 5.

**[00314]** Ejemplos adicionales de secuencias XTEN que se pueden usar de acuerdo con la presente invención se describen en las publicaciones de patentes de EE. UU. 2010/0239554 AI, 2010/0323956 AI, 2011/0046060 AI, 2011/0046061 AI, 2011/0077199 AI, o 2011/0172146 AI, o las publicaciones internacionales de patentes núms. WO 2010091122 AI, WO 2010144502 A2, WO 2010144508 AI, WO 2011028228 AI, WO 2011028229 AI, WO 2011028344 A2, WO 2014/011819 A2 o WO 2015/023891.

## **5. Albúmina o fragmento, derivado o variante del mismo**

**[00315]** En algunas realizaciones, el resto heterólogo comprende albúmina o algún fragmento funcional de la misma. La albúmina sérica humana (HSA o HA), una proteína de 609 aminoácidos en su forma completa, es responsable de una proporción significativa de la presión osmótica del suero y también funciona como portadora de ligandos endógenos y exógenos. El término "albúmina" como se usa en el presente documento incluye albúmina de longitud completa o un fragmento funcional, variante, derivado o análogo de la misma. Los ejemplos de albúmina o sus fragmentos o variantes se describen en las Publ. de Patentes de EE.UU. Nos. 2008/0194481A1, 2008/0004206 AI, 2008/0161243 AI, 2008/0261877 AI, o 2008/0153751 AI o Publ. de solicitud PCT Nos. 2008/033413 A2, 2009/058322 AI, o 2007/021494 A2, que se incorporan en este documento como referencia en su totalidad.

**[00316]** En una realización, la proteína terapéutica de la descripción comprende albúmina, un fragmento o una variante del mismo que está además unida a un segundo resto heterólogo seleccionado del grupo que consiste en una región constante de inmunoglobulina o una porción de la misma (por ejemplo, una región Fc), una secuencia PAS, HES, PEG y cualquiera de sus combinaciones.

## **6. Resto de unión a la albúmina**

**[00317]** En ciertas realizaciones, el resto heterólogo es un resto de unión a albúmina que comprende un péptido de unión a albúmina, un dominio de unión a albúmina bacteriano, un fragmento de anticuerpo de unión a albúmina, o cualquiera de sus combinaciones.

**[00318]** Por ejemplo, la proteína de unión a la albúmina puede ser una proteína de unión a la albúmina bacteriana, un anticuerpo o un fragmento de anticuerpo que incluya anticuerpos de dominio (véase la Patente de Estados Unidos N° 6.696.245).

Una proteína de unión a la albúmina, por ejemplo, puede ser un dominio de unión a la albúmina bacteriana, como el de la proteína estreptocócica G (Konig, T. y Skerra, A. (1998) J. Immunol. Methods 218, 73-83). Otros ejemplos de péptidos que se unen a la albúmina que pueden usarse como pareja de conjugación son, por ejemplo, aquellos que tienen una secuencia consenso Cys-Xaa<sub>1</sub>-Xaa<sub>2</sub>-Xaa<sub>3</sub>-Xaa<sub>4</sub>-Cys, donde Xaa<sub>1</sub> es Asp, Asn, Ser, Thr, o Trp; Xaa<sub>2</sub> es Asn, Gln, His, Ile, Leu, o Lys; Xaa<sub>3</sub> es Ala, Asp, Phe, Trp, o Tyr; y Xaa<sub>4</sub> es Asp, Gly, Leu, Phe, Ser, o Thr como se describe en la solicitud de patente de EE. UU. 2003/0069395 o Dennis et al. (Dennis et al. (2002) J. Biol. Chem 277, 35035-35043).

**[00319]** El dominio 3 de la proteína estreptocócica G, según lo descrito por Kraulis et al, FEBS Lett. 378: 190-194 (1996) y Linhult et al, Protein Sci. 11: 206-213 (2002) es un ejemplo de un dominio de unión a albúmina bacteriana. Los ejemplos de péptidos de unión a albúmina incluyen una serie de péptidos que tienen la secuencia central DICLPRWGCLW (SEQ ID NO: 35). Ver, por ejemplo, Dennis et al., J. Biol. Chem 2002, 277: 35035-35043 (2002). Ejemplos de fragmentos de anticuerpos que se unen a la albúmina se describen en Muller y Kontermann, Curr. Opin. Mol. Ther. 9:319-326 (2007); Roovers et al., Cancer Immunol. Immunother. 56: 303-317 (2007), y Holt et al., Prot. Eng. Design Sci., 21: 283-288 (2008), que se incorporan en este documento como referencia en su totalidad. Un ejemplo de tal resto de unión a la albúmina es 2-(3-maleimidopropanamido)-6-(4-(4-yodofenil)butanamido)hexanoato (marcador "Albu") como lo describen Trussel et al, Bioconjugate Chem. 20:2286-2292 (2009).

**[00320]** Los ácidos grasos, en particular los ácidos grasos de cadena larga (LCFA) y los compuestos de unión a albúmina similares a los ácidos grasos de cadena larga se pueden usar para extender la semivida in vivo de las proteínas del factor de coagulación de la descripción. Un ejemplo de un compuesto de unión a la albúmina similar a LCFA es el ácido 16-(1-(3-(9-(((2,5-dioxopirrolidin-1-iloxi) carboniloxi)-metil)-7-sulfo-9H-fluoren-2-ilamino)-3-oxopropil)-2,5-dioxopirrolidin-3-iltio)-hexadecanoico (véase, por ejemplo, el documento WO 2010/140148).

## **7. Secuencia PAS**

**[00321]** En otras realizaciones, el resto heterólogo es una secuencia PAS. Una secuencia PAS, como se usa en el presente documento, significa una secuencia de aminoácidos que comprende principalmente residuos de alanina y serina o que comprende principalmente residuos de alanina, serina y prolina, formando la secuencia de aminoácidos una conformación enrollada al azar en condiciones

fisiológicas. En consecuencia, la secuencia PAS es un bloque de construcción, un polímero de aminoácidos o un casete de secuencia que comprende, que consiste esencialmente, o que consiste en sí mismo en alanina, serina y prolina que se puede usar como parte del resto heterólogo en la proteína quimérica. Sin embargo, cualquier experto en la materia sabe que un polímero de aminoácidos también puede formar una conformación enrollada aleatoria cuando se agregan residuos distintos de alanina, serina y prolina como un constituyente menor en la secuencia PAS. La expresión "constituyente menor", como se usa en el presente documento, significa que se pueden agregar aminoácidos distintos de alanina, serina y prolina en la secuencia de PAS en cierto grado, por ejemplo, hasta aproximadamente 12%, es decir, de aproximadamente 12 de 100 aminoácidos de la secuencia PAS, hasta aproximadamente el 10%, es decir, de aproximadamente 10 de cada 100 aminoácidos de la secuencia PAS, hasta aproximadamente el 9%, es decir, aproximadamente 9 de cada 100 aminoácidos, hasta aproximadamente el 8%, es decir, aproximadamente 8 de cada 100 aminoácidos, aproximadamente 6%, es decir, aproximadamente 6 de cada 100 aminoácidos, aproximadamente 5%, es decir, aproximadamente 5 de cada 100 aminoácidos, aproximadamente 4%, es decir, aproximadamente 4 de cada 100 aminoácidos, aproximadamente 3%, es decir, aproximadamente 3 de cada 100 aminoácidos, aproximadamente 2%, es decir, aproximadamente 2 de cada 100 aminoácidos, aproximadamente 1%, es decir, aproximadamente 1 de cada 100 de los aminoácidos. Los aminoácidos diferentes de alanina, serina y prolina pueden seleccionarse del grupo que consiste en Arg, Asn, Asp, Cys, Gin, Glu, Gly, His, He, Leu, Lys, Met, Phe, Thr, Tip, Tyr, y Val.

**[00322]** En condiciones fisiológicas, la extensión de la secuencia PAS forma una conformación enrollada aleatoria y, por lo tanto, puede mediar una mayor estabilidad in vivo y/o in vitro a la proteína del factor de coagulación. Dado que el dominio enrollado aleatorio no adopta una estructura o función estable por sí misma, la actividad biológica mediada por la proteína del factor de coagulación se conserva esencialmente. En otras realizaciones, las secuencias de PAS que forman un dominio enrollado aleatorio son biológicamente inertes, especialmente con respecto a la proteólisis en plasma sanguíneo, inmunogenicidad, punto isoeléctrico/comportamiento electrostático, unión a receptores de superficie celular o internalización, pero aún son biodegradables, lo que proporciona claras ventajas respecto a los polímeros sintéticos como el PEG.

**[00323]** Los ejemplos no limitantes de las secuencias PAS que forman una conformación enrollada aleatoria comprenden una secuencia de aminoácidos seleccionada del grupo que consiste en ASPAAPAPASPAAPAPSAPA (SEQ ID NO: 36), AAPASPAPAAAPSAPAPAAPS (SEQ ID NO: 37), APSSPSPSAPSSPSPASPSS (SEQ ID NO: 38), APSSPSPSAPSSPSPASPS (SEQ ID NO: 39), SSPSAPSPSSPASPSPPSPA (SEQ ID NO: 40), AASPAAPSAPPAAASPAAPSAPPA (SEQ ID NO: 41), ASAAAPAAASAAASAPSAAA (SEQ ID NO: 42) y cualquiera de sus combinaciones. Se conocen ejemplos adicionales de secuencias PAS de, por ejemplo, la Publ. de patente de EE.UU. No. 2010/0292130 A1 y la Publ. de solicitud PCT No. WO 2008/155134 A1.

## **8. Secuencia HAP**

**[00324]** En ciertas realizaciones, el resto heterólogo es un polímero de hom-aminoácido (HAP) rico en glicina. La secuencia HAP puede comprender una secuencia repetitiva de glicina, que tiene al menos 50 aminoácidos, al menos 100 aminoácidos, 120 aminoácidos, 140 aminoácidos, 160 aminoácidos, 180 aminoácidos, 200 aminoácidos, 250 aminoácidos, 300 aminoácidos, 350 aminoácidos, 400 aminoácidos, 450 aminoácidos o 500 aminoácidos de longitud. En una realización, la secuencia HAP es capaz de extender la semivida de un resto fusionado o unido a la secuencia HAP. Los ejemplos no limitantes de la secuencia HAP incluyen, entre otros (Gly)<sub>n</sub>, (Gly<sub>4</sub>Ser)<sub>n</sub> o S(Gly<sub>4</sub>Ser)<sub>n</sub>, donde n es 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 o 20. En una realización, n es 20, 21, 22, 23, 24, 25, 26, 26, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39 o 40. En otra realización, n es 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190 o 200.

## **9. Transferrina o un fragmento de la misma**

**[00325]** En ciertas realizaciones, el resto heterólogo es transferrina o un fragmento de la misma. Se puede usar cualquier transferrina para fabricar las proteínas del factor de coagulación de la descripción. Como ejemplo, el TF humano (TF) natural es una proteína de 679 aminoácidos, de aproximadamente 75 KDa (sin tener en cuenta la glucosilación), con dos dominios principales, N (aproximadamente 330 aminoácidos) y C (aproximadamente 340 aminoácidos), que parecen originarse de una duplicación de genes. Consúltense los números de entrada del GenBank NM001063, XM002793, M12530, XM039845, XM039847 y S95936 ([www.ncbi.nlm.nih.gov/](http://www.ncbi.nlm.nih.gov/)), todos los cuales se incorporan en este documento como referencia en su totalidad. La transferrina comprende dos dominios, el dominio N y el

dominio C. El dominio N comprende dos subdominios, el dominio N1 y el dominio N2, y el dominio C comprende dos subdominios, el dominio C1 y el dominio C2.

**[00326]** En una realización, el resto heterólogo de transferrina incluye una variante de empalme de transferrina. En un ejemplo, una variante de empalme de transferrina puede ser una variante de empalme de transferrina humana, por ejemplo, nº de entrada del Genbank AAA61140. En otra realización, la porción de transferrina de la proteína quimérica incluye uno o más dominios de la secuencia de transferrina, por ejemplo, el dominio N, dominio C, dominio N1, dominio N2, dominio C1, dominio C2 o cualquiera de sus combinaciones.

## **10. Receptores de aclaramiento**

**[00327]** En ciertas realizaciones, el resto heterólogo es un receptor de aclaramiento, fragmento, variante o derivado del mismo. LRP1 es una proteína de membrana integral de 600 kDa que está implicada en el aclaramiento mediado por el receptor de una variedad de proteínas, como el Factor X. Véase, por ejemplo, Narita et al., Blood 91: 555-560 (1998).

## **11. Factor von Willebrand o fragmentos del mismo**

**[00328]** En ciertas realizaciones, el resto heterólogo es el Factor von Willebrand (VWF) o uno o más fragmentos del mismo.

**[00329]** El VWF (también conocido como F8VWF) es una glucoproteína multimérica grande presente en el plasma sanguíneo y producida constitutivamente en el endotelio (en los cuerpos de Weibel-Palade), megacariocitos ( $\alpha$ -gránulos de plaquetas) y tejido conectivo subendotelial. El monómero básico de VWF es una proteína de 2813 aminoácidos. Cada monómero contiene varios dominios específicos con una función específica, los dominios D' y D3 (que se unen al Factor VIII), el dominio A1 (que se une al receptor de plaquetas GPIb, heparina y/o posiblemente colágeno), Dominio A3 (que se une al colágeno), el dominio C1 (en el que el dominio RGD se une a la integrina plaquetaria  $\alpha$ Ib $\beta$ 3 cuando está activado) y el dominio "nudo de cisteína" en el extremo C-terminal de la proteína (cuyo VWF comparte con el factor de crecimiento derivado de plaquetas (PDGF), el factor de crecimiento transformante  $\beta$  (TGFP) y la gonadotropina coriónica humana  $\beta$  (PHCG)).

**[00330]** La secuencia de aminoácidos del monómero 2813 para el VWF humano aparece con el número de entrada NP000543.2 en el Genbank. La secuencia de nucleótidos que codifica el VWF humano aparece con el número de entrada NM000552.3 en el Genbank. La SEQ ID NO: 129 es la secuencia de aminoácidos codificada por la SEQ ID NO: 128. El dominio D' incluye los aminoácidos 764 a 866

de la SEQ ID NO: 129. El dominio D3 incluye los aminoácidos 867 a 1240 de la SEQ ID NO: 44.

**[00331]** En plasma, el 95-98% del FVIII circula en un complejo no covalente apretado con VWF de longitud completa. La formación de este complejo es importante para el mantenimiento de niveles plasmáticos apropiados de FVIII in vivo. Lenting et al., Blood. 92(11): 3983-96 (1998); Lenting et al., J. Thromb. Haemost 5(7): 1353-60 (2007). Cuando el FVIII se activa debido a la proteólisis en las posiciones 372 y 740 en la cadena pesada y en la posición 1689 en la cadena ligera, el VWF unido al FVIII se elimina del FVIII activado.

**[00332]** En ciertas realizaciones, el resto heterólogo es el Factor von Willebrand de longitud completa. En otras realizaciones, el resto heterólogo es un fragmento del factor von Willebrand. Como se usa en este documento, la expresión "fragmento de VWF" o "fragmentos de VWF", usadas en este documento, significan cualquier fragmento de VWF que interactúa con FVIII y retiene al menos una o más propiedades que normalmente proporciona FVIII al VWF de longitud completa, por ejemplo, para prevenir la activación prematura a FVIII, evitando la proteólisis prematura, evitando la asociación con membranas de fosfolípidos que podrían conducir a un aclaramiento prematuro, evitando la unión a receptores de aclaramiento del FVIII que pueden unirse al FVIII desnudo pero no al FVIII unido al VWF, y/o estabilizar las interacciones de la cadena pesada y la cadena ligera del FVIII. En una realización específica, el resto heterólogo es un fragmento (VWF) que comprende un dominio D' y un dominio D3 de VWF. El fragmento VWF que comprende el dominio D' y el dominio D3 puede comprender además un dominio VWF seleccionado del grupo que consiste en un dominio A1, un dominio A2, un dominio A3, un dominio D1, un dominio D2, un dominio D4, un dominio B1, un dominio B2, un dominio B3, un dominio C1, un dominio C2, un dominio CK, uno o más fragmentos del mismo, y cualquiera de sus combinaciones. Ejemplos adicionales del polipéptido que tiene actividad FVIII fusionada con el fragmento VWF se describen en la solicitud de patente provisional de EE. UU. No. 61/667.901, presentada el 3 de julio de 2012, y la publicación de EE. UU. No. 2015/0023959 A1, que se incorporan a la presente por referencia en su totalidad.

## **12. Restos de enlazante**

**[00333]** En ciertas realizaciones, el resto heterólogo es un enlazante peptídico.

**[00334]** Como se usa en el presente documento, las expresiones "enlazantes de péptidos" o "restos enlazantes" se refieren a una secuencia de péptidos o polipéptidos

(por ejemplo, una secuencia de péptidos o polipéptidos sintéticos) que conecta dos dominios en una secuencia lineal de aminoácidos de una cadena de polipéptidos.

**[00335]** En algunas realizaciones, los enlazantes peptídicos pueden insertarse entre la proteína terapéutica de la descripción y un resto heterólogo descrito anteriormente, tal como la albúmina. Los enlazantes de péptidos pueden proporcionar flexibilidad a la molécula de polipéptido quimérica. Los enlazantes no suelen escindirse normalmente, sin embargo, dicha escisión puede ser deseable. En una realización, estos enlazantes no se eliminan durante el procesamiento.

**[00336]** Un tipo de enlazante que puede estar presente en una proteína quimérica de la descripción es un enlazante escindible por proteasa que comprende un sitio de escisión (es decir, un sustrato del sitio de escisión de proteasa, por ejemplo, un factor XIa, Xa o un sitio de escisión de trombina) y que puede incluir enlazantes adicionales en el terminal N o el terminal C o en ambos lados del sitio de escisión. Estos enlazantes escindibles, cuando se incorporan a una construcción de la descripción, dan como resultado una molécula quimérica que tiene un sitio de escisión heterólogo.

**[00337]** En una realización, una proteína terapéutica codificada por una molécula de ácido nucleico de la presente descripción comprende dos o más dominios o restos Fc unidos a través de un enlazante cscFc para formar una región Fc comprendida en una única cadena de polipéptidos. El enlazante cscFc está flanqueado por al menos un sitio de procesamiento intracelular, es decir, un sitio escindido por una enzima intracelular. La escisión del polipéptido en al menos un sitio de procesamiento intracelular da como resultado un polipéptido que comprende al menos dos cadenas de polipéptidos.

**[00338]** Otros enlazantes peptídicos pueden usarse opcionalmente en una construcción de la descripción, por ejemplo, para conectar una proteína de factor de coagulación a una región Fc. Algunos enlazantes ejemplares que pueden usarse en conexión con la descripción incluyen, por ejemplo, polipéptidos que comprenden aminoácidos GlySer descritos con más detalle a continuación.

**[00339]** En una realización, el enlazante peptídico es sintético, es decir, de origen no natural. En una realización, un enlazante peptídico incluye péptidos (o polipéptidos) (que pueden o no pueden darse naturalmente) que comprenden una secuencia de aminoácidos que une o fusiona genéticamente una primera secuencia lineal de aminoácidos a una segunda secuencia lineal de aminoácidos a la que no está naturalmente unida o genéticamente fusionada en la naturaleza. Por ejemplo, en

una realización, el enlazante peptídico puede comprender polipéptidos de origen no natural que son formas modificadas de polipéptidos de origen natural (por ejemplo, que comprenden una mutación tal como una adición, sustitución o delección). En otra realización, el enlazante peptídico puede comprender aminoácidos no naturales. En otra realización, el enlazante peptídico puede comprender aminoácidos naturales que se producen en una secuencia lineal que no se produce en la naturaleza. En otra realización más, el enlazante peptídico puede comprender una secuencia de polipéptidos de origen natural.

**[00340]** Por ejemplo, en ciertas realizaciones, se puede usar un enlazante peptídico para fusionar restos Fc idénticos, formando así una región scFc homodimérica. En otras realizaciones, se puede usar un enlazante peptídico para fusionar diferentes restos Fc (por ejemplo, un resto Fc natural y una variante del resto Fc), formando así una región scFc heterodimérica.

**[00341]** En otra realización, un enlazante peptídico comprende o consiste en un enlazante gly-ser. En una realización, un enlazante scFc o cscFc comprende al menos una porción de una bisagra de inmunoglobulina y un enlazante de glicerina. Como se usa en este documento, la expresión "enlazante de gly-ser" se refiere a un péptido que consiste en residuos de glicina y serina. En ciertas realizaciones, dicho enlazante de gly-ser se puede insertar entre otras dos secuencias del enlazante de péptido. En otras realizaciones, un enlazante gly-ser se une a uno o ambos extremos de otra secuencia del enlazante peptídico. En otras realizaciones más, dos o más enlazantes de gly-ser se incorporan en serie en un enlazante peptídico. En una realización, un enlazante peptídico de la descripción comprende al menos una porción de una región de bisagra superior (por ejemplo, derivada de una molécula de IgG1, IgG2, IgG3 o IgG4), al menos una porción de una región de bisagra media (por ejemplo, derivada de una molécula de IgG1, IgG2, IgG3 o IgG4) y una serie de residuos de aminoácidos gly/ser.

**[00342]** Los enlazantes peptídicos de la descripción tienen al menos un aminoácido de longitud y pueden tener longitudes variables. En una realización, un enlazante peptídico de la descripción tiene una longitud de aproximadamente 1 a aproximadamente 50 aminoácidos. Como se usa en este contexto, el término "aproximadamente" indica +/- dos residuos de aminoácidos. Dado que la longitud del enlazante debe ser un número entero positivo, la longitud de aproximadamente 1 a aproximadamente 50 aminoácidos de longitud significa una longitud de 1-3 a 48-52 aminoácidos de longitud. En otra realización, un enlazante peptídico de la descripción

tiene una longitud de aproximadamente 10 a aproximadamente 20 aminoácidos. En otra realización, un enlazante peptídico de la descripción tiene una longitud de aproximadamente 15 a aproximadamente 50 aminoácidos. En otra realización, un enlazante peptídico de la descripción tiene una longitud de aproximadamente 20 a aproximadamente 45 aminoácidos. En otra realización, un enlazante peptídico de la descripción tiene una longitud de aproximadamente 15 a aproximadamente 35 o aproximadamente 20 a aproximadamente 30 aminoácidos. En otra realización, un enlazante peptídico de la descripción tiene aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 40, 50, 60, 70, 80, 90, 100, 500, 1000 o 2000 aminoácidos de longitud. En una realización, un enlazante peptídico de la descripción tiene una longitud de 20 o 30 aminoácidos.

**[00343]** En algunas realizaciones, el enlazante peptídico puede comprender al menos dos, al menos tres, al menos cuatro, al menos cinco, al menos 10, al menos 20, al menos 30, al menos 40, al menos 50, al menos 60, al menos 70, al menos 80, al menos 90 o al menos 100 aminoácidos. En otras realizaciones, el enlazante peptídico puede comprender al menos 200, al menos 300, al menos 400, al menos 500, al menos 600, al menos 700, al menos 800, al menos 900, o al menos 1.000 aminoácidos. En algunas realizaciones, el enlazante peptídico puede comprender al menos aproximadamente 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 150, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1100, 1200, 1300, 1400, 1500, 1600, 1700, 1800, 1900 o 2000 aminoácidos. El enlazante peptídico puede comprender 1-5 aminoácidos, 1-10 aminoácidos, 1-20 aminoácidos, 10-50 aminoácidos, 50-100 aminoácidos, 100-200 aminoácidos, 200-300 aminoácidos, 300-400 aminoácidos, 400-500 aminoácidos, 500-600 aminoácidos, 600-700 aminoácidos, 700-800 aminoácidos, 800-900 aminoácidos o 900-1000 aminoácidos.

**[00344]** Los enlazantes de péptidos pueden introducirse en secuencias de polipéptidos usando técnicas conocidas en la técnica. Las modificaciones pueden confirmarse mediante análisis de secuencia de ADN. El ADN plasmídico puede usarse para transformar células huésped para la producción estable de los polipéptidos producidos.

### **13. Híbridos de monómero-dímero**

**[00345]** En algunas realizaciones, la proteína terapéutica de la descripción comprende una molécula híbrida monómero-dímero que comprende un factor de coagulación.

**[00346]** La expresión "híbrido monómero-dímero" usada en el presente documento se refiere a una proteína quimérica que comprende una primera cadena de polipéptidos y una segunda cadena de polipéptidos, que están asociadas entre sí por un enlace disulfuro, en el que la primera cadena comprende un factor de coagulación, por ejemplo, FVIII, y una primera región Fc y la segunda cadena comprende, consiste esencialmente, o consiste en sí misma en una segunda región Fc sin el factor de coagulación. La construcción híbrida monómero-dímero, por lo tanto, es un híbrido que comprende un aspecto monómero que tiene solo un factor de coagulación y un aspecto dímero que tiene dos regiones Fc.

#### **14. Secuencias de control de la expresión**

**[00347]** En algunas realizaciones, la molécula de ácido nucleico de la descripción comprende además al menos una secuencia de control de la expresión. Una secuencia de control de la expresión, como se usa en el presente documento, es cualquier secuencia de nucleótidos reguladora, tal como una secuencia promotora o combinación de promotor-potenciador, que facilita la transcripción y traducción eficientes del ácido nucleico codificante al que está operativamente unido. Por ejemplo, la molécula de ácido nucleico de la descripción puede estar operativamente unida a al menos una secuencia de control de la transcripción. La secuencia de control de la expresión génica puede ser, por ejemplo, un promotor viral o de mamífero, tal como un promotor constitutivo o inducible.

**[00348]** Los promotores constitutivos de mamíferos incluyen, entre otros, los promotores de los siguientes genes: hipoxantina fosforibosil transferasa (HPRT), adenosina desaminasa, piruvato quinasa, promotor de beta-actina y otros promotores constitutivos. Los promotores virales ejemplares que funcionan constitutivamente en células eucariotas incluyen, por ejemplo, promotores del citomegalovirus (CMV), virus simio (por ejemplo, SV40), virus del papiloma, adenovirus, virus de la inmunodeficiencia humana (VIH), virus del sarcoma de Rous, citomegalovirus, repeticiones terminales largas (LTR) del virus de la leucemia de Moloney y otros retrovirus, y el promotor de timidina quinasa del virus del herpes simple. Los expertos en la materia conocen otros promotores constitutivos. Los promotores útiles como secuencias de expresión génica de la descripción también incluyen promotores inducibles. Los promotores inducibles se expresan en presencia de un agente inductor. Por ejemplo, el promotor de la metalotioneína se induce para promover la transcripción y traducción en presencia de ciertos iones metálicos. Los expertos habituales en la técnica conocen otros promotores inducibles.

**[00349]** En una realización, la descripción incluye la expresión de un transgén bajo el control de un promotor y/o potenciador específico de tejido. En otra realización, el promotor u otra secuencia de control de la expresión mejora selectivamente la expresión del transgén en células hepáticas. En ciertas realizaciones, el promotor u otra secuencia de control de la expresión selectiva potencia la expresión del transgén en hepatocitos, células sinusoidales y/o células endoteliales. En una realización particular, el promotor u otra secuencia de control de la expresión selectiva potencia la expresión del transgén en células endoteliales. En ciertas realizaciones, el promotor u otra secuencia de control de la expresión selectiva potencia la expresión del transgén en las células musculares, el sistema nervioso central, el ojo, el hígado, el corazón o cualquiera de sus combinaciones. Los ejemplos de promotores específicos del hígado incluyen, entre otros, un promotor de tiretina de ratón (mTTR), un promotor endógeno humano del factor VIII (F8), promotor de alfa-1-antitripsina humana (hAAT), promotor mínimo de albúmina humana y promotor de albúmina de ratón. En una realización particular, el promotor comprende un promotor mTTR. El promotor mTTR se describe en R. H. Costa et al., 1986, Mol. Cell. Biol. 6:4697. El promotor F8 se describe en Figueiredo y Brownlee, 1995, J. Biol. Chem 270: 11828-11838. En algunas realizaciones, el promotor se selecciona de un promotor específico del hígado (por ejemplo,  $\alpha$ 1-antitripsina (AAT)), un promotor específico del músculo (por ejemplo, creatina quinasa muscular (MCK), alfa de la cadena pesada de miosina ( $\alpha$ MHC), mioglobina (MB) y desmina (DES)), un promotor sintético (por ejemplo, SPc5-12, 2R5Sc5-12, dMCK y tMCK), y cualquiera de sus combinaciones.

**[00350]** En una realización, el promotor se selecciona del grupo que consiste en un promotor de tiretina de ratón (mTTR), un promotor endógeno del factor VIII humano (F8), promotor de alfa-1 antitripsina humana (hAAT), promotor mínimo de albúmina humana, promotor de albúmina de ratón, TTPp, un promotor CASI, un promotor CAG, un promotor de citomegalovirus (CMV),  $\alpha$ 1-antitripsina (AAT), creatina quinasa muscular (MCK), cadena pesada alfa de miosina ( $\alpha$ MHC), mioglobina (MB), desmina (DES), SPc5-12, 2R5Sc5-12, dMCK y tMCK, un promotor de fosfoglicerato quinasa (PGK) y cualquiera de sus combinaciones.

**[00351]** Los niveles de expresión se pueden mejorar aún más para lograr la eficacia terapéutica utilizando uno o más potenciadores. Se pueden proporcionar uno o más potenciadores solos o junto con uno o más elementos promotores. Típicamente, la secuencia de control de la expresión comprende una pluralidad de elementos potenciadores y un promotor específico de tejido. En una realización, un potenciador

comprende una o más copias del potenciador de  $\alpha$ -1-microglobulina/bikunina (Rouet et al., 1992, J. Biol. Chem 267:20765-20773; Rouet et al., 1995, Nucleic Acids Res. 23: 395-404; Rouet et al., 1998, Biochem. J. 334: 577-584; Ill et al., 1997, Blood Coagulation Fibrinolysis 8: S23-S30). En otra realización, un potenciador se deriva de  
5 sitios de unión del factor de transcripción específico del hígado, tales como EBP, DBP, HNF1, HNF3, HNF4, HNF6, con Enh1, que comprende HNF1, (sentido)-HNF3, (sentido)-HNF4, (antisentido)-HNF1, (antisentido)-HNF6, (sentido)-EBP, (antisentido)-HNF4 (antisentido).

**[00352]** En un ejemplo particular, un promotor útil para la descripción comprende  
10 SEQ ID NO: 69 (es decir, promotor ET), que también se conoce como GenBank No. AY661265. Ver también Vigna et al., Molecular Therapy 11 (5). 763 (2005). Se describen ejemplos de otros vectores adecuados y elementos reguladores de genes en los documentos WO 02/092134, EP1395293, o las patentes de los Estados Unidos Nos. 6.808.905, 7.745.179, o 7.179.903, que se incorporan en este documento por  
15 referencia en su totalidad.

**[00353]** En una realización, las moléculas de ácido nucleico de la presente descripción además comprenden una secuencia intrónica. En algunas realizaciones, la secuencia intrónica está colocada 5' respecto a la secuencia de ácido nucleico que codifica el polipéptido FVIII. En algunas realizaciones, la secuencia intrónica es una  
20 secuencia intrónica natural. En algunas realizaciones, la secuencia intrónica es una secuencia sintética. En algunas realizaciones, la secuencia intrónica se deriva de una secuencia intrónica natural. En ciertas realizaciones, la secuencia intrónica comprende el intrón T pequeño SV40. En una realización, la secuencia intrónica comprende SEQ ID NO: 115.

**[00354]** En algunas realizaciones, la molécula de ácido nucleico comprende además un elemento regulador postranscripcional. En ciertas realizaciones, el elemento regulador post-transcripcional comprende un elemento regulador post-transcripcional del virus de la hepatitis de marmota mutado (WPRE). En una  
25 realización particular, el elemento regulador postranscripcional comprende la SEQ ID NO: 120.

**[00355]** En algunas realizaciones, la molécula de ácido nucleico comprende un sitio de unión de microARN (ARNmi). En una realización, el sitio de unión de ARNmi es un sitio de unión de ARNmi para miR-142-3p. En otras realizaciones, el sitio de unión de ARNmi se selecciona de un sitio de unión de ARNmi descrito por Rennie et  
35 al., RNA Biol. 13(6): 554-560 (2016), y STarMirDB, disponible en

<http://sfold.wadsworth.org/starmirDB.php>, que se incorpora en este documento por referencia en su totalidad.

**[00356]** En algunas realizaciones, la molécula de ácido nucleico comprende una o más secuencias de direccionamiento nuclear de ADN (DTS). Un DTS promueve la translocación de moléculas de ADN que contienen tales secuencias en el núcleo. En ciertas realizaciones, el DTS comprende una secuencia potenciadora de SV40. En ciertas realizaciones, el DTS comprende una secuencia potenciadora de c-Myc. En algunas realizaciones, los DTS están entre la primera ITR y la segunda ITR. En algunas realizaciones, el DTS es 3' para la primera ITR y 5' para la proteína terapéutica. En otras realizaciones, el DTS es 3' para la proteína terapéutica y 5' para la segunda ITR.

**[00357]** En algunas realizaciones, la molécula de ácido nucleico comprende además una secuencia de cola poli (A) 3'UTR. En una realización, la secuencia de cola poli (A) 3'UTR comprende bGH poli (A). En una realización, la cola poli (A) 3'UTR comprende un sitio de actina poli (A). En una realización, la cola poli (A) 3'UTR comprende un sitio de hemoglobina poli (A).

**[00358]** En una realización particular, la secuencia de cola poli (A) 3'UTR comprende a la SEQ ID NO: 122.

### **III. Expresión específica de tejido**

**[00359]** En ciertas realizaciones, será útil incluir dentro del vector una o más secuencias diana de ARNmi que, por ejemplo, estén operativamente unidas al transgén del factor de coagulación. Por lo tanto, la descripción también proporciona al menos un objetivo de secuencia de ARNmi unido operativamente a la secuencia de nucleótidos del factor de coagulación o insertado de otro modo dentro de un vector. Más de una copia de una secuencia objetivo de ARNmi incluida en el vector puede aumentar la efectividad del sistema. También se incluyen diferentes secuencias diana de ARNmi. Por ejemplo, los vectores que expresan más de un transgén pueden tener el transgén bajo control de más de una secuencia objetivo de ARNmi, que puede ser igual o diferente. Las secuencias diana de ARNmi pueden estar en tándem, pero también se incluyen otras disposiciones. El casete de expresión transgénica, que contiene secuencias diana de ARNmi, también se puede insertar dentro del vector en orientación antisentido. La orientación antisentido puede ser útil en la producción de partículas virales para evitar la expresión de productos génicos que de otro modo pueden ser tóxicos para las células productoras. En otras realizaciones, el vector comprende 1, 2, 3, 4, 5, 6, 7 u 8 copias de la misma o diferente secuencia diana de

ARNmi. Sin embargo, en ciertas otras realizaciones, el vector no incluirá ninguna secuencia diana de ARNmi. La elección de incluir o no una secuencia objetivo de ARNmi (y cuántas) se guiará por parámetros conocidos como el objetivo de tejido deseado, el nivel de expresión requerido, etc.

5   **[00360]**        En una realización, la secuencia objetivo es un objetivo miR-223 que se ha informado que bloquea la expresión más eficazmente en progenitores comprometidos con mieloides y al menos parcialmente en el HSPC más primitivo. El objetivo miR-223 puede bloquear la expresión en células mieloides diferenciadas, incluidos granulocitos, monocitos, macrófagos y células dendríticas mieloides. El  
10   objetivo miR-223 también puede ser adecuado para aplicaciones de terapia génica que dependan de una expresión transgénica fuerte en el linaje linfóide o eritroide. El objetivo miR-223 también puede bloquear la expresión de manera muy efectiva en HSC humano.

15   **[00361]**        En otra realización, la secuencia objetivo es un objetivo miR142 (tccataaagt aggaaacact aca (SEQ ID NO: 43)). En una realización, el vector comprende 4 copias de secuencias diana miR-142. En ciertas realizaciones, la secuencia complementaria de microARN específicos hematopoyéticos, como miR-142 (142T), se incorpora en la región no traducida 3' de un vector, por ejemplo, vectores lentivirales (LV), haciendo que el transcrito que codifica el transgén sea  
20   susceptible a la regulación descendente mediada por ARNmi. Mediante este método, se puede prevenir la expresión transgénica en las células presentadoras de antígeno (APC) de linaje hematopoyético, mientras se mantiene en células no hematopoyéticas (Brown et al., Nat Med 2006). Esta estrategia puede imponer un estricto control postranscripcional en la expresión transgénica y, por lo tanto, permite la entrega  
25   estable y la expresión a largo plazo de transgenes. En algunas realizaciones, la regulación miR-142 evita el aclaramiento mediado por el sistema inmune de las células transducidas y/o induce células T reguladoras (T regs) específicas de antígeno y media una tolerancia inmunológica fuerte respecto al antígeno codificado por el transgén.

30   **[00362]**        En algunas realizaciones, la secuencia objetivo es un objetivo miR181. Chen C-Z y Lodish H, Seminars in Immunology (2005) 17 (2): 155-165 describe miR-181, un ARNmi expresado específicamente en células B dentro de la médula ósea de ratón (Chen y Lodish, 2005). También describe que algunos ARNmi humanos están relacionados con leucemias.

**[00363]** La secuencia objetivo puede ser total o parcialmente complementaria al ARNm. La expresión "completamente complementaria" significa que la secuencia diana tiene una secuencia de ácido nucleico que es 100% complementaria con la secuencia del ARNm que la reconoce. La expresión "parcialmente complementaria" significa que la secuencia diana es solo en parte complementaria a la secuencia del ARNm que la reconoce, por lo que la secuencia parcialmente complementaria todavía es reconocida por el ARNm. En otras palabras, una secuencia diana parcialmente complementaria en el contexto de la presente descripción es eficaz para reconocer el ARNm correspondiente y efectuar la prevención o reducción de la expresión transgénica en células que expresan ese ARNm. Se describen ejemplos de las secuencias diana de ARNm en los documentos WO2007/000668, WO2004/094642, WO2010/055413 o WO2010/125471, que se incorporan en este documento como referencia en su totalidad.

**[00364]** En algunas realizaciones, la expresión transgénica está dirigida al hígado. En ciertas realizaciones, la expresión transgénica se dirige a los hepatocitos. En otra realización, la expresión transgénica está dirigida a las células endoteliales. En una realización particular, la expresión transgénica se dirige a cualquier tejido que exprese naturalmente FVIII endógeno.

**[00365]** En algunas realizaciones, la expresión transgénica está dirigida al sistema nervioso central. En ciertas realizaciones, la expresión transgénica está dirigida a las neuronas. En algunas realizaciones, la expresión transgénica está dirigida a neuronas aferentes. En algunas realizaciones, la expresión transgénica está dirigida a neuronas eferentes. En algunas realizaciones, la expresión transgénica está dirigida a interneuronas. En algunas realizaciones, la expresión transgénica está dirigida a las células gliales. En algunas realizaciones, la expresión transgénica se dirige a los astrocitos. En algunas realizaciones, la expresión transgénica está dirigida a los oligodendrocitos. En algunas realizaciones, la expresión transgénica está dirigida a la microglía. En algunas realizaciones, la expresión transgénica está dirigida a células ependimarias. En algunas realizaciones, la expresión transgénica está dirigida a las células de Schwann. En algunas realizaciones, la expresión transgénica se dirige a las células satélite.

**[00366]** En algunas realizaciones, la expresión transgénica está dirigida al tejido muscular. En algunas realizaciones, la expresión transgénica está dirigida al músculo liso. En algunas realizaciones, la expresión transgénica está dirigida al músculo

cardíaco. En algunas realizaciones, la expresión transgénica está dirigida al músculo esquelético.

**[00367]** En algunas realizaciones, la expresión transgénica está dirigida al ojo. En algunas realizaciones, la expresión transgénica se dirige a una célula fotorreceptora. En algunas realizaciones, la expresión transgénica se dirige a la célula ganglionar retiniana.

#### **IV. Células huésped**

**[00368]** La descripción también proporciona una célula huésped que comprende una molécula de ácido nucleico o un vector de la descripción. Como se usa en el presente documento, el término "transformación" se usará en un sentido amplio para referirse a la introducción de ADN en una célula huésped receptora que cambia el genotipo y, en consecuencia, da como resultado un cambio en la célula receptora.

**[00369]** "Células huésped" se refiere a células que se han transformado con vectores contruidos usando técnicas de ADN recombinante y que codifican al menos un gen heterólogo. Las células huésped de la presente descripción son preferiblemente de origen mamífero; más preferiblemente de origen humano o de ratón. A los expertos en la materia se les acredita la capacidad de determinar preferentemente líneas celulares huésped particulares que se adapten mejor a su propósito. Las líneas de células huésped ejemplares incluyen, entre otras, CHO, DG44 y DUXB 11 (líneas de ovario de hámster chino, DHFR minus), HELA (carcinoma cervical humano), CVI (línea de riñón de mono), COS (un derivado de CVI con Antígeno T SV40), R1610 (fibroblastos de hámster chino) BALBC/3T3 (fibroblastos de ratón), HAK (línea de riñón de hámster), SP2/O (mieloma de ratón), P3.times.63-Ag3.653 (mieloma de ratón), BFA-lclBPT (células endoteliales bovinas), RAJI (linfocito humano), PER.C6®, NSO, CAP, BHK21 y HEK 293 (riñón humano). En una realización particular, la célula huésped se selecciona del grupo que consiste en: una célula CHO, una célula HEK293, una célula BHK21, una célula PER.C6®, una célula NSO, una célula CAP y cualquiera de sus combinaciones. En algunas realizaciones, las células huésped de la presente descripción son de origen insecto. En una realización particular, las células huésped son células SF9. Las líneas celulares del huésped están típicamente disponibles en los servicios comerciales, la American Tissue Culture Collection o en la literatura publicada.

**[00370]** La introducción de las moléculas de ácido nucleico o los vectores de la descripción en la célula huésped se puede lograr mediante diversas técnicas bien conocidas por los expertos en la materia. Estos incluyen, entre otros, transfección

(incluyendo electroforesis y electroporación), fusión de protoplastos, precipitación con fosfato de calcio, fusión celular con ADN envuelto, microinyección e infección con virus intacto. Ver, Ridgway, A. A. G. "*Mammalian Expression Vectors*" Capítulo 24.2, pp. 470-472 Vectors, Rodríguez and Denhardt, Eds. (Butterworths, Boston, Massachusetts, 1988). Lo más preferible es que la introducción del plásmido en el huésped sea por electroporación. Las células transformadas se cultivan en condiciones apropiadas para la producción de cadenas ligeras y cadenas pesadas, y se analizan para realizar la síntesis de proteínas de cadena pesada y/o ligera. Las técnicas de ensayo ejemplares incluyen el ensayo inmunosorbente ligado a enzimas (ELISA), radioinmunoensayo (RIA) o análisis de clasificación de células activadas por fluorescencia (FACS), inmunohistoquímica y similares.

**[00371]** Las células huésped que comprenden las moléculas de ácido nucleico aisladas o los vectores de la descripción se cultivan en un medio de crecimiento apropiado. Como se usa en el presente documento, la expresión "medio de crecimiento apropiado" significa un medio que contiene nutrientes necesarios para el crecimiento de las células. Los nutrientes necesarios para el crecimiento celular pueden incluir una fuente de carbono, una fuente de nitrógeno, aminoácidos esenciales, vitaminas, minerales y factores de crecimiento. Opcionalmente, los medios pueden contener uno o más factores de selección. Opcionalmente, los medios pueden contener suero de ternera bovina o suero de ternera fetal (FCS). En una realización, el medio no contiene sustancialmente ninguna IgG. El medio de crecimiento generalmente seleccionará las células que contienen la construcción de ADN mediante, por ejemplo, la selección del fármaco o la deficiencia en un nutriente esencial que se complementa con el marcador seleccionable en la construcción de ADN o se transfecta junto con la construcción de ADN. Las células de mamífero cultivadas generalmente se cultivan en medios que contienen suero o están libres de suero disponibles comercialmente (por ejemplo, MEM, DMEM, DMEM/F12). En una realización, el medio es CDoptiCHO (Invitrogen, Carlsbad, CA.). En otra realización, el medio es CD 17 (Invitrogen, Carlsbad, CA.). La selección de un medio apropiado para la línea celular particular utilizada está dentro del nivel de los expertos ordinarios en la técnica.

## **V. Preparación de polipéptidos.**

**[00372]** La descripción también proporciona un polipéptido codificado por una molécula de ácido nucleico de la descripción. En otras realizaciones, el polipéptido de la descripción está codificado por un vector que comprende las moléculas de ácido

nucleico de la descripción. En otras realizaciones más, el polipéptido de la descripción es producido por una célula huésped que comprende las moléculas de ácido nucleico de la descripción.

**[00373]** En otras realizaciones, la descripción también proporciona un método para producir un polipéptido con actividad de factor de coagulación, por ejemplo, FVIII, que comprende cultivar una célula huésped de la descripción en condiciones en las que se produce un polipéptido con factor de coagulación, por ejemplo, FVIII, y recuperar el polipéptido con la actividad del factor de coagulación, por ejemplo, de FVIII. En algunas realizaciones, la expresión del polipéptido con la actividad del factor de coagulación, por ejemplo, FVIII, aumenta en relación con una célula huésped cultivada en las mismas condiciones pero que comprende una secuencia de nucleótidos de referencia (por ejemplo, SEQ ID NO: 16, la secuencia del gen FVIII progenitora)

**[00374]** En otras realizaciones, la descripción proporciona un método para aumentar la expresión de un polipéptido con la actividad del factor de coagulación, por ejemplo, FVIII, que comprende cultivar una célula huésped de la descripción en condiciones por las cuales un polipéptido con la actividad de factor de coagulación, por ejemplo, FVIII, se expresa por la molécula de ácido nucleico, en donde la expresión del polipéptido con la actividad de factor de coagulación, por ejemplo, FVIII, aumenta en relación con una célula huésped cultivada en las mismas condiciones que comprende una molécula de ácido nucleico de referencia (por ejemplo, SEQ ID NO: 16, secuencia de genes de FVIII progenitora).

**[00375]** En otras realizaciones, la descripción proporciona un método para mejorar el rendimiento de un polipéptido con la actividad del factor de coagulación, por ejemplo, FVIII, que comprende cultivar una célula huésped en condiciones en las que la molécula de ácido nucleico produce una actividad de polipéptido con factor de coagulación, por ejemplo, FVIII dada a conocer en el presente documento, en el que el rendimiento del polipéptido con la actividad del factor de coagulación, por ejemplo, FVIII, aumenta en relación con una célula huésped cultivada en las mismas condiciones que comprende una secuencia de ácido nucleico de referencia (por ejemplo, SEQ ID NO: 16, la secuencia del gen FVIII progenitora).

**[00376]** La proteína terapéutica, p. ej. el factor de coagulación de la descripción puede sintetizarse en un animal transgénico, como un roedor, cabra, oveja, cerdo o vaca. La expresión "animales transgénicos" se refiere a animales no humanos que han incorporado un gen extraño en su genoma. Debido a que este gen está presente

en los tejidos de la línea germinal, se transmite de padres a hijos. Se introducen genes exógenos en embriones unicelulares (Brinster et al. 1985, Proc. Natl. Acad. Sci. USA 82: 4438). Los métodos para producir animales transgénicos son conocidos en la técnica, incluidos los transgénicos que producen moléculas de inmunoglobulina (Wagner et al. 1981, Proc. Natl. Acad. Sci. USA 78: 6376; McKnight et al. 1983, Cell 34: 335; Brinster et al. 1983, Nature 306: 332; Ritchie et al. 1984, Nature 312: 517; Baldassarre et al. 2003, Theriogenology 59: 831; Robl et al. 2003, Theriogenology 59: 107; Malassagne et al. 2003, Xenotransplantation 10 (3): 267).

## **VII. Composición farmacéutica**

**[00377]** Las composiciones que contienen una molécula de ácido nucleico, un polipéptido codificado por la molécula de ácido nucleico, un vector o una célula huésped de la presente descripción pueden contener un vehículo farmacéuticamente aceptable adecuado. Por ejemplo, pueden contener excipientes y/o agentes auxiliares que faciliten el procesamiento de los compuestos activos en preparaciones diseñadas para la entrega al sitio de acción.

**[00378]** En una realización, la presente descripción se dirige a una composición farmacéutica que comprende (a) una molécula de ácido nucleico, un vector, un polipéptido o una célula huésped descrita en el presente documento; y (b) un excipiente farmacéuticamente aceptable.

**[00379]** En algunas realizaciones, la composición farmacéutica comprende además un agente de administración. En ciertas realizaciones, el agente de administración comprende una nanopartícula lipídica (LNP). En otras realizaciones, la composición farmacéutica comprende además liposomas, otras moléculas poliméricas y exosomas.

**[00380]** La composición farmacéutica puede formularse para su administración parenteral (es decir, intravenosa, subcutánea o intramuscular) mediante inyección en bolo. Las formulaciones para inyección pueden presentarse en forma de dosificación unitaria, por ejemplo, en ampollas o en envases multidosis con un conservante añadido. Las composiciones pueden tomar formas tales como suspensiones, soluciones o emulsiones en vehículos oleosos o acuosos, y contienen agentes de formulación tales como agentes de suspensión, estabilizantes y/o dispersantes. Alternativamente, el ingrediente activo puede estar en forma de polvo para constituir con un vehículo adecuado, por ejemplo, agua libre de pirógenos.

**[00381]** Las formulaciones adecuadas para la administración parenteral también incluyen soluciones acuosas de los compuestos activos en forma soluble en agua, por

ejemplo, sales solubles en agua. Además, se pueden administrar suspensiones de los compuestos activos como suspensiones de inyección oleosas apropiadas. Los solventes o vehículos lipofílicos adecuados incluyen aceites grasos, por ejemplo, aceite de sésamo, o ésteres de ácidos grasos sintéticos, por ejemplo, oleato de etilo o triglicéridos. Las suspensiones de inyección acuosa pueden contener sustancias que aumentan la viscosidad de la suspensión, que incluyen, por ejemplo, carboximetilcelulosa de sodio, sorbitol y dextrano. Opcionalmente, la suspensión también puede contener estabilizadores. Los liposomas también se pueden usar para encapsular las moléculas de la descripción para el suministro a las células o espacios intersticiales. Los ejemplos de vehículos farmacéuticamente aceptables son disolventes fisiológicamente compatibles, medios de dispersión, recubrimientos, agentes antibacterianos y antifúngicos, agentes isotónicos y retardadores de la absorción, agua, solución salina, solución salina tamponada con fosfato, dextrosa, glicerol, etanol y similares. En algunas realizaciones, la composición comprende agentes isotónicos, por ejemplo, azúcares, polialcoholes tales como manitol, sorbitol o cloruro de sodio. En otras realizaciones, las composiciones comprenden sustancias farmacéuticamente aceptables tales como agentes humectantes o cantidades menores de sustancias auxiliares tales como agentes humectantes o emulsionantes, conservantes o tampones, que mejoran la vida útil o la efectividad de los ingredientes activos.

**[00382]** Las composiciones de la descripción pueden estar en una variedad de formas, que incluyen, por ejemplo, líquidos (por ejemplo, soluciones inyectables e infusibles), dispersiones, suspensiones, formas de dosificación semisólidas y sólidas. La forma preferida depende del modo de administración y la aplicación terapéutica.

**[00383]** La composición puede formularse como una solución, microemulsión, dispersión, liposoma u otra estructura ordenada adecuada para una alta concentración de fármaco. Las soluciones inyectables estériles se pueden preparar incorporando el ingrediente activo en la cantidad requerida en un solvente apropiado con uno o una combinación de ingredientes mencionados anteriormente, según sea necesario, seguido de su esterilización filtrada. Generalmente, las dispersiones se preparan incorporando el ingrediente activo en un vehículo estéril que contiene un medio de dispersión básico y los otros ingredientes requeridos de los mencionados anteriormente. En el caso de polvos estériles para la preparación de soluciones inyectables estériles, los métodos preferidos de preparación son el secado al vacío y la liofilización que produce un polvo del ingrediente activo más cualquier ingrediente

adicional deseado de una solución filtrada previamente estéril. La fluidez adecuada de una solución se puede mantener, por ejemplo, mediante el uso de un recubrimiento como la lecitina, mediante el mantenimiento del tamaño de partícula requerido en el caso de una dispersión y mediante el uso de tensioactivos. La absorción prolongada de las composiciones inyectables puede lograrse incluyendo en la composición un agente que retrase la absorción, por ejemplo, sales de monoestearato y gelatina.

**[00384]** El ingrediente activo puede formularse con una formulación o dispositivo de liberación controlada. Los ejemplos de tales formulaciones y dispositivos incluyen implantes, parches transdérmicos y sistemas de administración microencapsulados.

Se pueden usar polímeros biodegradables y biocompatibles, por ejemplo, acetato de etileno vinilo, polianhídridos, ácido poliglicólico, colágeno, poliortoésteres y ácido poliláctico. Los métodos para la preparación de tales formulaciones y dispositivos son conocidos en la técnica. Ver, por ejemplo, Sustained and Controlled Release Drug Delivery Systems, J. R. Robinson, ed., Marcel Dekker, Inc., Nueva York, 1978.

**[00385]** Las formulaciones de depósito inyectables pueden prepararse formando matrices microencapsuladas del fármaco en polímeros biodegradables tales como polilactida-poliglicólido. Dependiendo de la relación de fármaco a polímero, y la naturaleza del polímero empleado, se puede controlar la velocidad de liberación del fármaco. Otros polímeros biodegradables ejemplares son los poliortoésteres y los polianhídridos. Las formulaciones inyectables de depósito también se pueden preparar atrapando el medicamento en liposomas o microemulsiones.

**[00386]** Se pueden incorporar compuestos activos suplementarios en las composiciones. En una realización, la molécula de ácido nucleico de la descripción se formula con un factor de coagulación, o una variante, fragmento, análogo o derivado del mismo. Por ejemplo, el factor de coagulación incluye, entre otros, factor V, factor VII, factor VIII, factor IX, factor X, factor XI, factor XII, factor XIII, protrombina, fibrinógeno, factor von Willebrand o factor de tejido soluble recombinante (rsTF) o formas activadas de cualquiera de los anteriores. El factor de coagulación del agente hemostático también puede incluir fármacos antifibrinolíticos, por ejemplo, ácido épsilon-aminocaproico, ácido tranexámico.

**[00387]** Los regímenes de dosificación se pueden ajustar para proporcionar la respuesta óptima deseada. Por ejemplo, se puede administrar un solo bolo, se pueden administrar varias dosis divididas con el tiempo, o la dosis se puede reducir o aumentar proporcionalmente según lo indiquen las exigencias de la situación terapéutica. Es ventajoso formular composiciones parenterales en formas de unidad de dosificación

para facilitar la administración y la uniformidad de la dosificación. Ver, por ejemplo, Remington's Pharmaceutical Sciences (Mack Pub. Co., Easton, Pa. 1980).

**[00388]** Además del compuesto activo, la forma de dosificación líquida puede contener ingredientes inertes como agua, alcohol etílico, carbonato de etilo, acetato de etilo, alcohol bencílico, benzoato de bencilo, propilenglicol, 1,3-butilenglicol, dimetilformamida, aceites, glicerol, alcohol tetrahidrofurfurílico, polietilenglicoles y ésteres de sorbitán de ácido graso.

**[00389]** Ejemplos no limitantes de vehículos farmacéuticos adecuados también se describen en Remington's Pharmaceutical Sciences por E. W. Martin. Algunos ejemplos de excipientes incluyen almidón, glucosa, lactosa, sacarosa, gelatina, malta, arroz, harina, yeso, gel de sílice, estearato de sodio, monoestearato de glicerol, talco, cloruro de sodio, leche descremada en polvo, glicerol, propilenglicol, agua, etanol, y similares. La composición también puede contener reactivos tamponadores del pH y agentes humectantes o emulsionantes.

**[00390]** Para la administración oral, la composición farmacéutica puede tomar la forma de tabletas o cápsulas preparadas por medios convencionales. La composición también se puede preparar como un líquido, por ejemplo, un jarabe o una suspensión. El líquido puede incluir agentes de suspensión (por ejemplo, jarabe de sorbitol, derivados de celulosa o grasas comestibles hidrogenadas), agentes emulsionantes (lecitina o acacia), vehículos no acuosos (por ejemplo, aceite de almendras, ésteres oleosos, alcohol etílico o aceites vegetales fraccionados), y conservantes (por ejemplo, metil o propil-p-hidroxibenzoatos o ácido sórbico). Las preparaciones también pueden incluir agentes saborizantes, colorantes y edulcorantes. Alternativamente, la composición puede presentarse como un producto seco para su reconstitución en agua u otro vehículo adecuado.

**[00391]** Para la administración bucal, la composición puede tomar la forma de tabletas o pastillas de acuerdo con los protocolos convencionales.

**[00392]** Para la administración por inhalación, los compuestos para su uso de acuerdo con la presente descripción se administran convenientemente en forma de aerosol nebulizado con o sin excipientes o en forma de aerosol de un paquete o nebulizador presurizado, opcionalmente, con un propelente, por ejemplo, diclorodifluorometano, triclorofluorometano, diclorotetrafluorometano, dióxido de carbono u otros gases adecuados. En el caso de un aerosol presurizado, la unidad de dosificación se puede determinar proporcionando una válvula para administrar una cantidad medida. Se pueden formular cápsulas y cartuchos de, por ejemplo, gelatina

para usar en un inhalador o insuflador que contenga una mezcla en polvo del compuesto y una base en polvo adecuada tal como lactosa o almidón.

[00393] La composición farmacéutica también puede formularse para administración rectal como un supositorio o enema de retención, por ejemplo, que  
5 contiene bases de supositorio convencionales tales como manteca de cacao u otros glicéridos.

[00394] En algunas realizaciones, la composición se administra por una ruta seleccionada del grupo que consiste en administración tópica, administración intraocular, administración parenteral, administración intratecal, administración  
10 subdural y administración oral. La administración parenteral puede ser intravenosa o subcutánea.

### **VIII. Métodos de tratamiento**

[00395] En algunas realizaciones, la molécula de ácido nucleico comprende una primera ITR, una segunda ITR, y un casete génico, en donde el casete génico  
15 comprende un factor de coagulación, y en donde la molécula de ácido nucleico se usa para tratar una enfermedad o afección hemorrágica en un sujeto que lo necesita. La enfermedad o afección hemorrágica se selecciona del grupo que consiste en un trastorno de coagulación hemorrágica, hemartrosis, sangrado muscular, sangrado oral, hemorragia, hemorragia en los músculos, hemorragia oral, traumatismo,  
20 traumatismo capitis, hemorragia gastrointestinal, hemorragia intracraneal, hemorragia intraabdominal, hemorragia intratorácica, fractura ósea, sangrado del sistema nervioso central, sangrado en el espacio retrofaríngeo, sangrado en el espacio retroperitoneal, sangrado en la vaina del iliopsoas y cualquiera de sus combinaciones. En otras realizaciones más, el sujeto está programado para someterse a una cirugía.  
25 En otras realizaciones más, el tratamiento es profiláctico o bajo demanda.

[00396] La descripción proporciona un método para tratar un trastorno hemorrágico que comprende administrar a un sujeto que lo necesite una molécula de ácido nucleico, un vector o un polipéptido de la descripción. En algunas realizaciones, el trastorno hemorrágico se caracteriza por una deficiencia en un factor de  
30 coagulación, por ejemplo, FVIII. En algunas realizaciones, el trastorno hemorrágico es la hemofilia. En algunas realizaciones, el trastorno hemorrágico es la hemofilia A. En algunas realizaciones del método para tratar un trastorno hemorrágico, la actividad plasmática de un factor de coagulación, por ejemplo, FVIII, a las 24 horas después de la administración aumenta en relación con un sujeto al que se le administró una  
35 molécula de ácido nucleico de referencia (p. ej., SEQ ID NO: 16, la secuencia del gen

FVIII progenitora), un vector que comprende una molécula de ácido nucleico de referencia o un polipéptido codificado por una molécula de ácido nucleico de referencia.

**[00397]** La descripción también se refiere a un método para tratar, mejorar o prevenir un trastorno hemostático en un sujeto que comprende administrar una cantidad terapéuticamente efectiva de una molécula de ácido nucleico aislada de la descripción o un polipéptido que tiene una actividad de factor de coagulación, por ejemplo, FVIII, codificada por la molécula de ácido nucleico de la descripción. El tratamiento, mejora y prevención mediante la molécula de ácido nucleico aislada o el polipéptido codificado puede ser una terapia de derivación. El sujeto que recibe la terapia de derivación puede haber desarrollado un inhibidor de un factor de coagulación, por ejemplo, FVIII, o puede generar un inhibidor del factor de coagulación.

**[00398]** Las moléculas de ácido nucleico, los vectores o los polipéptidos de la descripción tratan o previenen un trastorno hemostático promoviendo la formación de un coágulo de fibrina. El polipéptido que tiene actividad del factor de coagulación, por ejemplo, FVIII, codificado por la molécula de ácido nucleico de la descripción puede activar un miembro de una cascada de coagulación. El factor de coagulación puede ser un participante en la vía extrínseca, la vía intrínseca o ambas.

**[00399]** Las moléculas de ácido nucleico, los vectores o los polipéptidos de la descripción pueden usarse para tratar trastornos hemostáticos que se sabe que son tratables con un factor de coagulación. Los trastornos hemostáticos que pueden tratarse utilizando los métodos de la descripción incluyen, entre otros, hemofilia A, hemofilia B, enfermedad von Willebrand, deficiencia del Factor XI (deficiencia de PTA), deficiencia del Factor XII, así como deficiencias o anomalías estructurales en fibrinógeno, protrombina, Factor V, Factor VII, Factor X o Factor XIII, hemartrosis, sangrado muscular, sangrado oral, hemorragia, hemorragia en los músculos, hemorragia oral, traumatismo, traumatismo capitis, hemorragia gastrointestinal, hemorragia intracraneal, hemorragia intraabdominal, hemorragia intratorácica, fractura ósea, sangrado del sistema nervioso central, sangrado en el espacio retrofaríngeo, sangrado en el espacio retroperitoneal y sangrado en la vaina del iliopsoas.

**[00400]** En algunas realizaciones, el trastorno hemostático es un trastorno hereditario. En una realización, el sujeto tiene hemofilia A. En otras realizaciones, el trastorno hemostático es el resultado de una deficiencia en un factor de coagulación.

En otras realizaciones, el trastorno hemostático es el resultado de una deficiencia en FVIII. En otras realizaciones, el trastorno hemostático puede ser el resultado de un factor de coagulación de FVIII defectuoso.

**[00401]** En otra realización, el trastorno hemostático puede ser un trastorno adquirido. El trastorno adquirido puede ser el resultado de una enfermedad o afección secundaria subyacente. La condición no relacionada puede ser, por ejemplo, pero no como una limitación, cáncer, una enfermedad autoinmune o embarazo. El trastorno adquirido puede ser el resultado de la vejez o de medicamentos para tratar un trastorno secundario subyacente (por ejemplo, quimioterapia contra el cáncer).

**[00402]** La descripción también se refiere a métodos para tratar a un sujeto que no tiene un trastorno hemostático o una enfermedad o afección secundaria que dé como resultado la adquisición de un trastorno hemostático. Por lo tanto, la descripción se refiere a un método para tratar a un sujeto que necesita un agente hemostático general que comprende administrar una cantidad terapéuticamente efectiva de la molécula, vector o polipéptido de ácido nucleico aislado de la descripción. Por ejemplo, en una realización, el sujeto que necesita un agente hemostático general está siendo sometido o está a punto de someterse a una cirugía. La molécula, el vector o el polipéptido de ácido nucleico aislado de la descripción se puede administrar antes o después de la cirugía como profiláctico. La molécula, el vector o el polipéptido de ácido nucleico aislado de la descripción se puede administrar durante o después de la cirugía para controlar un episodio de sangrado agudo. La cirugía puede incluir, entre otros, trasplante de hígado, resección hepática o trasplante de células madre.

**[00403]** En otra realización, la molécula, el vector o el polipéptido de ácido nucleico aislado de la descripción se puede usar para tratar a un sujeto que tiene un episodio de sangrado agudo que no tiene un trastorno hemostático. El episodio de sangrado agudo puede ser el resultado de un traumatismo grave, por ejemplo, cirugía, un accidente automovilístico, heridas, un disparo con arma de laceración o cualquier otro evento traumático que provoque un sangrado incontrolado.

**[00404]** La molécula, el vector o la proteína de ácido nucleico aislados se pueden usar para tratar profilácticamente a un sujeto con un trastorno hemostático. La molécula, el vector o la proteína de ácido nucleico aislados se pueden usar para tratar un episodio de sangrado agudo en un sujeto con un trastorno hemostático.

**[00405]** En otra realización, la expresión de la proteína del factor de coagulación mediante la administración de la molécula de ácido nucleico aislada o el vector de la descripción no induce ninguna respuesta inmunitaria en el sujeto. En algunas

realizaciones, la respuesta inmunitaria comprende el desarrollo de anticuerpos contra un factor de coagulación. En una realización, la respuesta inmunitaria comprende el desarrollo de anticuerpos contra FVIII. En algunas realizaciones, la respuesta inmunitaria comprende la secreción de citoquinas. En algunas realizaciones, la respuesta inmunitaria comprende la activación de células B, células T, o tanto células B como células T. En algunas realizaciones, la respuesta inmunitaria es una respuesta inmunitaria inhibitoria, en donde la respuesta inmunitaria en el sujeto reduce la actividad de una proteína del factor de coagulación en relación con la actividad del factor de coagulación en un sujeto que no ha desarrollado ninguna respuesta inmunitaria. En ciertas realizaciones, la expresión de una proteína del factor de coagulación mediante la administración de la molécula de ácido nucleico o el vector aislado de la descripción evita el surgimiento de una respuesta inmunitaria inhibitoria contra la proteína del factor de coagulación o la proteína del factor de coagulación expresada a partir de la molécula de ácido nucleico aislada o el vector.

**[00406]** En algunas realizaciones, una composición aislada de molécula, vector o proteína de ácido nucleico de la descripción se administra en combinación con al menos otro agente que promueve la hemostasia. Dicho otro agente promueve la hemostasia en una terapéutica con actividad de coagulación demostrada. Como ejemplo, pero no como una limitación, el agente hemostático puede incluir FV, FVII, FIX, FX, FXI, FXII, FXIII, protrombina o fibrinógeno o formas activadas de cualquiera de los anteriores. El factor de coagulación o el agente hemostático también pueden incluir fármacos antifibrinolíticos, por ejemplo, ácido épsilon-aminocaproico, ácido tranexámico.

**[00407]** En una realización de la descripción, la composición (por ejemplo, la molécula, el vector o el polipéptido de ácido nucleico aislado) es aquella en donde el factor de coagulación está presente en forma activable cuando se administra a un sujeto. Dicha molécula activable se puede activar in vivo en el sitio de coagulación después de la administración a un sujeto.

**[00408]** La molécula, vector o polipéptido de ácido nucleico aislado puede administrarse por vía intravenosa, subcutánea, intramuscular o a través de cualquier superficie de la mucosa, por ejemplo, por vía oral, sublingual, bucal, sublingual, nasal, rectal, vaginal o vía pulmonar. La proteína del factor de coagulación puede implantarse o unirse a un soporte sólido de biopolímero que permita la liberación lenta de la proteína quimérica en el sitio deseado.

**[00409]** Para la administración oral, la composición farmacéutica puede tomar la forma de tabletas o cápsulas preparadas por medios convencionales. La composición también se puede preparar como un líquido, por ejemplo, un jarabe o una suspensión. El líquido puede incluir agentes de suspensión (por ejemplo, jarabe de sorbitol, derivados de celulosa o grasas comestibles hidrogenadas), agentes emulsionantes (lecitina o acacia), vehículos no acuosos (por ejemplo, aceite de almendras, ésteres oleosos, alcohol etílico o aceites vegetales fraccionados) y conservantes (por ejemplo, metil- o propil-p-hidroxibenzoatos o ácido sórbico). Las preparaciones también pueden incluir agentes saborizantes, colorantes y edulcorantes. Alternativamente, la composición puede presentarse como un producto seco para su reconstitución con agua u otro vehículo adecuado.

**[00410]** Para la administración bucal y sublingual, la composición puede tomar la forma de tabletas, pastillas o películas de disolución rápida de acuerdo con los protocolos convencionales.

**[00411]** Para la administración por inhalación, el polipéptido que tiene actividad de factor de coagulación para su uso de acuerdo con la presente descripción se administra convenientemente en forma de aerosol de un paquete presurizado o nebulizador (por ejemplo, en PBS), con un propelente adecuado, por ejemplo, diclorodifluorometano, triclorofluorometano, diclorotetrafluorometano, dióxido de carbono u otro gas adecuado. En el caso de un aerosol presurizado, la unidad de dosificación se puede determinar proporcionando una válvula para administrar una cantidad medida. Se pueden formular cápsulas y cartuchos de, por ejemplo, gelatina para usar en un inhalador o insuflador que contenga una mezcla en polvo del compuesto y una base en polvo adecuada tal como lactosa o almidón.

**[00412]** En una realización, la ruta de administración de la molécula, vector o polipéptido de ácido nucleico aislado es parenteral. El término parenteral como se usa en el presente documento incluye la administración intravenosa, intraarterial, intraperitoneal, intramuscular, subcutánea, rectal o vaginal. Se prefiere la forma intravenosa de la administración parenteral. Si bien todas estas formas de administración se contemplan claramente dentro del alcance de la descripción, una forma de administración sería una solución inyectable, en particular para la inyección intravenosa o intraarterial o el goteo. Normalmente, una composición farmacéutica adecuada para su inyección puede comprender un tampón (por ejemplo, tampón de acetato, fosfato o citrato), un tensioactivo (por ejemplo, polisorbato), opcionalmente un agente estabilizador (por ejemplo, albúmina humana), etc. Sin embargo, en otros

métodos compatibles con las enseñanzas de la presente memoria, la molécula, el vector o el polipéptido de ácido nucleico aislado se puede administrar directamente al sitio de la población celular adversa aumentando así la exposición del tejido enfermo al agente terapéutico.

5 **[00413]** Las preparaciones para la administración parenteral incluyen soluciones, suspensiones y emulsiones acuosas o no acuosas estériles. Ejemplos de disolventes no acuosos son propilenglicol, polietilenglicol, aceites vegetales como el aceite de oliva y ésteres orgánicos inyectables como el oleato de etilo. Los portadores acuosos incluyen agua, soluciones alcohólicas/acuosas, emulsiones o suspensiones,  
10 que incluyen solución salina y medios tamponados. En la descripción del tema, los vehículos farmacéuticamente aceptables incluyen, entre otros, 0,01-0,1 M y preferiblemente tampón fosfato 0,05 M o solución salina al 0,8%. Otros vehículos parenterales comunes incluyen soluciones de fosfato de sodio, dextrosa de Ringer, dextrosa y cloruro de sodio, Ringer de lactato o aceites fijos. Los vehículos  
15 intravenosos incluyen reabastecedores de fluidos y nutrientes, reabastecedores de electrolitos, como los basados en la dextrosa de Ringer, y similares. También pueden estar presentes conservantes y otros aditivos tales como, por ejemplo, antimicrobianos, antioxidantes, agentes quelantes y gases inertes y similares.

**[00414]** Más particularmente, las composiciones farmacéuticas adecuadas para  
20 un uso inyectable incluyen soluciones acuosas estériles (cuando son solubles en agua) o dispersiones y polvos estériles para la preparación extemporánea de soluciones o dispersiones inyectables estériles. En tales casos, la composición debe ser estéril y debe ser fluida en la medida en que exista una fácil capacidad de inyección. Debe ser estable en las condiciones de fabricación y almacenamiento y  
25 preferiblemente se preservará contra la acción contaminante de microorganismos, como bacterias y hongos. El vehículo puede ser un disolvente o medio de dispersión que contenga, por ejemplo, agua, etanol, poliol (por ejemplo, glicerol, propilenglicol y polietilenglicol líquido, y similares), y mezclas adecuadas de los mismos. Se puede mantener la fluidez adecuada, por ejemplo, mediante el uso de un recubrimiento como  
30 la lecitina, mediante el mantenimiento del tamaño de partícula requerido en el caso de dispersión y mediante el uso de tensioactivos.

**[00415]** La composición farmacéutica también puede formularse para administración rectal como un supositorio o enema de retención, por ejemplo, con bases de supositorio convencionales tales como manteca de cacao u otros glicéridos.

**[00416]** Las dosis efectivas de las composiciones de la presente descripción para el tratamiento de afecciones varían dependiendo de muchos factores diferentes, que incluyen medios de administración, sitio objetivo, estado fisiológico del paciente, si el paciente es humano o animal, otros medicamentos administrados y si el tratamiento es profiláctico o terapéutico. Por lo general, el paciente es un ser humano, pero también se pueden tratar mamíferos no humanos, incluidos los mamíferos transgénicos. Las dosis de tratamiento pueden ajustarse utilizando métodos de rutina conocidos por los expertos en la materia para optimizar la seguridad y la eficacia.

**[00417]** La molécula de ácido nucleico, el vector o los polipéptidos de la descripción pueden administrarse opcionalmente en combinación con otros agentes que son eficaces en el tratamiento del trastorno o afección que necesita tratamiento (por ejemplo, profiláctico o terapéutico).

**[00418]** Como se usa en el presente documento, la administración de moléculas de ácido nucleico aisladas, vectores o polipéptidos de la descripción en combinación o junto con una terapia complementaria significa la administración o aplicación secuencial, simultánea, coextensiva, concurrente, concomitante o contemporánea de la terapia y los polipéptidos descritos. Los expertos en la materia apreciarán que la administración o aplicación de los diversos componentes del régimen terapéutico combinado puede programarse para mejorar la eficacia general del tratamiento. Cualquier experto en la materia (p. ej., un médico) podría discernir fácilmente regímenes terapéuticos combinados efectivos sin una experimentación excesiva basada en la terapia adjunta seleccionada y las enseñanzas de la presente memoria descriptiva.

**[00419]** Se apreciará además que la molécula, el vector o el polipéptido de ácido nucleico aislado de la presente descripción se puede usar en conjunción o combinación con un agente o varios agentes (por ejemplo, para proporcionar un régimen terapéutico combinado). Los agentes ejemplares con los que se puede combinar un polipéptido o polinucleótido de la descripción incluyen agentes que representan el estándar actual de atención para un trastorno particular que se está tratando. Dichos agentes pueden ser de naturaleza química o biológica. El término "biológico" o "agente biológico" se refiere a cualquier agente farmacéuticamente activo hecho de organismos vivos y/o sus productos destinados para ser utilizados como agentes terapéuticos.

**[00420]** La cantidad de agente que se utilizará en combinación con los polinucleótidos o polipéptidos de la presente descripción puede variar según el sujeto

o puede administrarse de acuerdo con lo que se conoce en la técnica. Véase, por ejemplo, Bruce A Chabner et al., Antineoplastic Agents, en GOODMAN & GILMAN'S THE FARMACOLOGICAL BASIS OF THERAPEUTICS 1233-1287 ((Joel G. Hardman et al., Eds., 9ª ed. 1996). En otra realización, se administra una cantidad de dicho agente consistente con los estándares de las buenas prácticas en el cuidado.

**[00421]** En una realización, también se describe en el presente documento un kit que comprende la molécula de ácido nucleico descrita en el presente documento y las instrucciones para administrar la molécula de ácido nucleico a un sujeto que lo necesite. En otra realización, se describe en el presente documento un sistema de baculovirus para la producción de la molécula de ácido nucleico proporcionada en el presente documento. La molécula de ácido nucleico se produce en células de insectos. En otra realización, se proporciona un sistema de suministro de nanopartículas para construcciones de expresión. La construcción de expresión comprende la molécula de ácido nucleico descrita en este documento.

## **IX. Terapia de genes**

**[00422]** Ciertos aspectos de la presente descripción proporcionan un método para expresar una construcción genética en un sujeto, que comprende administrar la molécula de ácido nucleico aislada de la descripción a un sujeto que lo necesite. En algunos aspectos, la descripción proporciona un método para aumentar la expresión de un polipéptido en un sujeto que comprende administrar la molécula de ácido nucleico aislada de la descripción a un sujeto que lo necesita. En otros aspectos, la descripción proporciona un método de modulación de la expresión de un polipéptido en un sujeto que lo necesita que comprende administrar una molécula de ácido nucleico aislada de la descripción, por ejemplo, una secuencia de ácido nucleico que comprende un ARNmi, al sujeto. En algunos aspectos, la descripción proporciona un método de regulación a la baja de la expresión de un gen diana en un sujeto que lo necesita que comprende administrar una molécula de ácido nucleico aislada de la descripción, por ejemplo, una secuencia de ácido nucleico que comprende un ARNmi al sujeto.

**[00423]** La terapia génica somática se ha explorado como un posible tratamiento para una variedad de afecciones, que incluyen, entre otras, la hemofilia A. La terapia génica es un tratamiento particularmente atractivo para la hemofilia debido a su potencial para curar enfermedades mediante la producción endógena continua de un factor de coagulación, por ejemplo, FVIII, después de una única administración del vector. La hemofilia A es muy adecuada para un enfoque de reemplazo de genes

porque sus manifestaciones clínicas son completamente atribuibles a la falta de un solo producto génico (por ejemplo, FVIII) que circula en pequeñas cantidades (200 ng/ml) en el plasma.

**[00424]** Una proteína del factor de coagulación de la descripción puede producirse in vivo en un mamífero, por ejemplo, un paciente humano, usando un enfoque de terapia génica para el tratamiento de una enfermedad o trastorno hemorrágico seleccionado del grupo que consiste en un trastorno de coagulación hemorrágica, hemartrosis, sangrado muscular, sangrado oral, hemorragia, hemorragia en los músculos, hemorragia oral, traumatismo, traumatismo capitis, sangrado gastrointestinal, hemorragia intracraneal, hemorragia intraabdominal, hemorragia intratorácica, fractura ósea, sangrado del sistema nervioso central, sangrado en el espacio retrofaríngeo, sangrado en el espacio retroperitoneal, y el sangrado en la vaina del iliopsoas sería terapéuticamente beneficioso. En una realización, la enfermedad o trastorno hemorrágico es la hemofilia. En otra realización, la enfermedad o trastorno hemorrágico es la hemofilia A.

**[00425]** Otras condiciones también son adecuadas para el tratamiento usando las moléculas de ácido nucleico descritas en este documento. En ciertas realizaciones, los métodos descritos en este documento se usan para tratar una enfermedad o afección que afecta a un órgano diana seleccionado del músculo, sistema nervioso central (SNC), ocular, hígado, corazón, riñón, páncreas, pulmones, piel, vejiga, tracto urinario, o cualquiera de sus combinaciones. En ciertas realizaciones, los métodos descritos en este documento se usan para tratar una enfermedad o afección seleccionada de DMD (distrofia muscular de Duchenne), XLMTM (miopatía miotubular ligada al cromosoma X), Parkinson, SMA (atrofia muscular espinal), ataxia de Friedreich, GUCY2D-LCA ( Amaurosis congénita de Leber), XLRS (retinosquiasis ligada al cromosoma X), AMD (degeneración macular relacionada con la edad), ACHM (acromatopsia), IRD mediada por RPF65 (Tabla 9).

**Tabla 9. Enfermedades y trastornos tratables por los métodos descritos en este documento.**

| Enfermedad                           | Órgano diana | Gen defectuoso                   | Terapia génica       |
|--------------------------------------|--------------|----------------------------------|----------------------|
| DMD (distrofia muscular de Duchenne) | Músculo      | Distrofina ligada al cromosoma X | Introducción del gen |

|                                                                           |                               |                                            |                      |
|---------------------------------------------------------------------------|-------------------------------|--------------------------------------------|----------------------|
| XLMTM (miopatía miotubular ligada al cromosoma X)                         | Músculo                       | MTM1 (miotubularin)                        | Introducción del gen |
| Parkinson                                                                 | CNS                           | Tirosina hidroxilasa, AADC, ciclohidrolasa | introducción del gen |
| SMA (atrofia muscular espinal)                                            | CNS                           | SMN1                                       | introducción del gen |
| Ataxia de Friedreich                                                      | CNS                           | FXN (Frataxina)                            | introducción del gen |
| GUCY2D-LCA Amaurosis Congénita Leber                                      | Ocular                        | GUCY2D                                     | introducción del gen |
| XLRS Retinosquiasis ligada a X                                            | Ocular                        | RS1                                        | introducción del gen |
| AMD Degeneración macular relacionada con la edad                          | Ocular                        | CFH HTRA ARMS CFB/CC2                      | introducción del gen |
| ACHM acromatopsia                                                         | Ocular                        | CNGA/CNGB                                  | introducción del gen |
| IRD mediada por RPF65                                                     | Ocular                        | Prf65                                      | introducción del gen |
| Trastornos de almacenamiento lisosómico                                   | Órgano diana                  | Gen defectuoso                             | Terapia génica       |
| MLD leucodistrofia metacromática (Trastorno de almacenamiento lisosómico) | CNS                           | ARSA PSAP                                  | introducción del gen |
| MPS Mucopolisacaridosis (Trastorno de almacenamiento lisosómico)          | Hígado                        | IDUA (MPS I) IDS (MPS II)                  | Introducción del gen |
| PKU Fenilcetonuria (Trastorno de almacenamiento lisosómico)               | Hígado                        | PAH                                        | Introducción del gen |
| Pompe Enfermedad de almacenamiento de glucógeno tipo II                   | Corazón, hígado, músculo, CNS | GAA (ácida alfa-glucosidasa)               | Introducción de gen  |

| Terapias de micro ARN                           | Órgano diana | Gen defectuoso               | Terapia génica |
|-------------------------------------------------|--------------|------------------------------|----------------|
| ALS Esclerosis lateral amiotrófica              | CNS          | SOD1 <sup>1</sup>            | miRNA          |
| Enfermedad de Huntington                        | CNS          | HTT <sup>2</sup>             | miRNA          |
| AdRP Retinitis Pigmentaria Autosómica Dominante | Ocular       | RHO <sup>2</sup> (Rodopsina) | ARNmi          |

1: La mutación del gen SOD1 representó el 20% del caso ALS heredado. SOD1 natural ha demostrado propiedades antiapoptóticas en cultivos neurales, mientras que se ha observado que la SOD1 mutante promueve la apoptosis en las mitocondrias de la médula espinal, pero no en las mitocondrias hepáticas, aunque se expresa igualmente en ambos. La expresión de SOD1 mutada con regulación negativa podría inhibir la degeneración de la neurona motora en la ALS.

2: HD es uno de los diversos trastornos repetidos de trinucleótidos que están causados por la longitud de una sección repetida de un gen que excede un rango normal. HTT contiene una secuencia de tres bases de ADN: citosina-adenina-guanina (CAG), repetidas varias veces (es decir ... CAGCAGCAG ...), conocido como una repetición de trinucleótidos. CAG es el código genético de 3 letras (codón) para el aminoácido glutamina, por lo que una serie de ellos da como resultado la producción de una cadena de glutamina conocida como tracto de poliglutamina (o tracto polyQ), y la parte repetida del gen, la región PolyQ.

En general, las personas tienen menos de 36 glutaminas repetidas en la región polyQ, lo que produce la producción de la proteína citoplasmática Huntingtin. Sin embargo, una secuencia de 36 o más glutaminas da como resultado la producción de una proteína que tiene características diferentes. Esta forma alterada, llamada huntingtina mutante (mHTT), aumenta la tasa de descomposición de ciertos tipos de neuronas.

En general, el número de repeticiones de CAG está relacionado con cuánto se ve afectado este proceso y representa aproximadamente el 60% de la variación de la edad de aparición de los síntomas. La variación restante se atribuye al ambiente y a otros genes que modifican el mecanismo de la HD. 36-39 repeticiones dan lugar a una forma de penetración reducida de la enfermedad, con un inicio mucho más tardío y una progresión más lenta de los síntomas. En algunos casos, el inicio puede ser tan tardío que los síntomas nunca se notan. Con recuentos de repetición muy grandes, la HD tiene una penetrancia total y puede ocurrir antes de los 20 años, que entonces se conoce como HD juvenil, HD rígida acinética o variante de Westphal. Esto representa

aproximadamente el 7% de los vehículos de HD. 3: La mayoría de las mutaciones del gen RHO responsables de la retinitis pigmentosa alteran el plegamiento o el transporte de la proteína rodopsina. Algunas mutaciones hacen que la rodopsina se active constitutivamente en lugar de activarse en respuesta a la luz. Los estudios sugieren que las versiones alteradas de la rodopsina interfieren con las funciones celulares esenciales, causando que los bastones se autodestruyan (sufran apoptosis). Debido a que los bastones son esenciales para la visión en condiciones de poca luz, la pérdida de estas células conduciría a una ceguera nocturna progresiva en personas con retinitis pigmentosa.

**[00426]** En algunas realizaciones, los métodos descritos en este documento se usan para tratar un trastorno de almacenamiento lisosómico. En algunas realizaciones, el trastorno de almacenamiento lisosómico se selecciona de MLD (leucodistrofia metacromática), MPS (mucopolisacaridosis), PKU (fenilcetonuria), enfermedad de almacenamiento de glucógeno pompe tipo II, o cualquiera de sus combinaciones.

**[00427]** En algunas realizaciones, los métodos descritos en este documento se usan en una terapia microARN (ARNmi). En algunas realizaciones, el ARNmi trata la afección causada por la sobreexpresión de un gen o una proteína. En algunas realizaciones, el ARNmi trata una afección causada por la acumulación de una proteína. En algunas realizaciones, el ARNmi trata una afección causada por la eliminación de la expresión de un gen o proteína. En algunas realizaciones, el ARNmi trata una afección causada por la expresión de un gen mutante. En algunas realizaciones, el ARNmi trata una afección causada por la expresión de un gen heterólogo. En ciertas realizaciones, la terapia con ARNmi trata una afección seleccionada de ALS (esclerosis lateral amiotrófica), enfermedad de Huntington, AdRP (retinitis pigmentosa autosómica dominante) y cualquiera de sus combinaciones. En ciertas realizaciones, los métodos de la presente descripción comprenden el tratamiento dirigido de ALS mediante la administración de una molécula de ácido nucleico descrita en este documento, en donde la molécula de ácido nucleico comprende un casete génico que codifica un ARNmi, en donde el ARNmi se dirige a la expresión de SOD1. En ciertas realizaciones, el ARNmi comprende el ARNmi artificial miR SOD1 descrito por Dirren et al., Annals of Clinical and Translational Neurology 2 (2): 167- 84 (febrero de 2015). La mutación del gen SOD1 representa el 20% de los casos de ALS heredados. SOD1 natural ha demostrado propiedades antiapoptóticas en cultivos neurales, mientras que se ha observado que

la SOD1 mutante promueve la apoptosis en las mitocondrias de la médula espinal, pero no en las mitocondrias hepáticas, aunque se expresa igualmente en ambos. La regulación negativa de la expresión de SOD1 mutada podría inhibir la degeneración de la neurona motora en ALS.

5 **[00428]** En ciertas realizaciones, los métodos de la presente descripción comprenden el tratamiento dirigido de la enfermedad de Huntington mediante la administración de una molécula de ácido nucleico descrita en el presente documento, en donde la molécula de ácido nucleico comprende un casete génico que codifica un ARNmi, donde el ARNmi se dirige a la expresión de HTT. En ciertas realizaciones, el

10 ARNmi comprende el ARNmi diseñado por miHTT descrito por Evers et al., Molecular Therapy 26 (9): 1-15 (epub antes de la impresión de junio de 2018). La enfermedad de Huntington es uno de varios trastornos repetidos por trinucleótidos que son causados por la longitud de una sección repetida de un gen que excede un rango normal. HTT contiene una secuencia de tres bases de ADN: citosina-adenina-guanina

15 (CAG), repetidas varias veces (es decir ... CAGCAGCAG ...), que se conoce como una repetición de trinucleótidos. CAG es el código genético de 3 letras (codón) para el aminoácido glutamina, por lo que una serie de estas repeticiones da como resultado la producción de una cadena de glutamina conocida como tracto de poliglutamina (o tracto polyQ), y la parte repetida del gen, la región PolyQ. En general, las personas

20 tienen menos de 36 glutaminas repetidas en la región polyQ, lo que resulta en la producción de la proteína citoplasmática huntingtina. Sin embargo, una secuencia de 36 o más glutaminas da como resultado la producción de una proteína que tiene características diferentes. Esta forma alterada, llamada huntingtina mutante (mHTT), aumenta la tasa de descomposición de ciertos tipos de neuronas. En general, el

25 número de repeticiones de CAG está relacionado con cuánto se ve afectado este proceso y representa aproximadamente el 60% de la variación de la edad de aparición de los síntomas. La variación restante se atribuye al medio ambiente y a otros genes que modifican el mecanismo de la enfermedad de Huntington. 36-39 repeticiones resultan en una forma de penetración reducida de la enfermedad, con un inicio mucho

30 más tardío y una progresión más lenta de los síntomas. En algunos casos, el inicio puede ser tan tardío que los síntomas nunca se noten. Con recuentos de repetición muy grandes, la enfermedad de Huntington tiene una penetrancia total y puede ocurrir con menos de 20 años de edad, que es cuando se la conoce como enfermedad de Huntington juvenil, enfermedad de Huntington con rigidez acinética o variante de

Westphal. Esto representa aproximadamente el 7% de los portadores de la enfermedad de Huntington.

[00429] En ciertas realizaciones, los métodos de la presente descripción comprenden el tratamiento dirigido de la retinitis pigmentosa autosómica dominante (AdRP) mediante la administración de una molécula de ácido nucleico descrita en el presente documento, donde la molécula de ácido nucleico comprende un casete génico que codifica un ARNmi, donde el ARNmi se dirige a la expresión de RHO (Rodopsina). En ciertas realizaciones, el ARNmi comprende miR-708 (véase Behrman et al., JCB 192 (6): 919-27 (2011)). La mayoría de las mutaciones del gen RHO responsables de la retinitis pigmentosa alteran el plegamiento o el transporte de la proteína rodopsina. Algunas mutaciones hacen que la rodopsina se active constitutivamente en lugar de activarse en respuesta a la luz. Los estudios sugieren que las versiones alteradas de la rodopsina interfieren con las funciones celulares esenciales, causando que los bastones se autodestruyan (sufran apoptosis). Debido a que los bastones son esenciales para la visión en condiciones de poca luz, la pérdida de estas células conduce a la ceguera nocturna progresiva en personas con retinitis pigmentosa.

[00430] Todos los diversos aspectos, realizaciones y opciones descritos en este documento pueden combinarse en cualquiera y en todas las variaciones.

[00431] Todas las publicaciones, patentes y solicitudes de patentes mencionadas en esta memoria descriptiva se incorporan en este documento por referencia en la misma medida que si cada publicación individual, patente o solicitud de patente se indicara específica e individualmente para ser incorporada por referencia.

[00432] Habiendo descrito de manera general esta descripción, se puede obtener otra comprensión por referencia a los ejemplos proporcionados en este documento. Estos ejemplos son solo para fines ilustrativos y no pretenden ser limitantes.

## EJEMPLOS

### **Ejemplo 1. Generación de construcciones de la expresión de FVIII que llevan ITR parvovirales de AAV y VnoAA**

*Ejemplo 1a. Clonación del gen FVIII optimizado con codón y regiones de repetición terminal invertida (ITR) de AAV en casetes genéticos.*

**[00433]** El casete génico FVIII se generó en base al genoma del serotipo 2 de AAV. Sin embargo, las regiones ITR originadas de cualquier serotipo (incluido el sintético) pueden usarse en este enfoque (Fig. 1A).

**[00434]** El plásmido de expresión AAV2-FVIIIco6XTEN que codifica una secuencia codificante de FVIII optimizada con codón bajo la regulación de un promotor específico de hígado (TTPp, Figuras 1A e 1B) flanqueado por regiones de repetición terminal invertida (ITR) de AAV (AAV-FVIII) fue diseñado para su expresión in vitro e in vivo, como se muestra en la FIG. 1B. El casete génico también contiene elementos WPRE y bGHpA para la expresión óptima del transgén (Figuras 1A e 1B). La secuencia de FVIII optimizada por codón flanqueada por ITR se clonó en una estructura plasmídica que comprendía un origen de replicación ColEI y un casete de expresión para beta-lactamasa, que confería resistencia frente a la ampicilina (Figura 1B). Los sitios de reconocimiento para la endonucleasa de restricción PvuII que flanqueaban el casete de expresión se diseñaron para permitir la escisión precisa de la construcción AAV-FVIII tras la digestión con PvuII (Figura 1B).

*Ejemplo 1b. Clonación del gen FVIII optimizado con codón y regiones de repetición terminal invertida (ITR) de parvovirus VnoAA en casetes genéticos.*

**[00435]** Basado en la relación filogénica entre los miembros de la familia viral *Parvoviridae*, al que pertenece el AAV (figura 2A), se planteó la hipótesis de que otros miembros no pertenecientes al AAV del género *Dependovirus* y los miembros del género *Erythrovirus* (y posiblemente todos/muchos otros miembros de la familia viral) utilizaban mecanismos celulares similares para el mantenimiento del ciclo de vida viral y, lo más importante, el establecimiento de una infección persistente y latente. Por lo tanto, las regiones ITR originadas en los genomas de estos virus podrían utilizarse para desarrollar casetes de expresión genética de tipo AAV (pero basados en VnoAA). Los siguientes parvovirus se probaron para determinar la idoneidad de sus regiones ITR para el desarrollo de construcciones genéticas para aplicaciones de terapia génica: *Dependovirus*, cepa de parvovirus de ganso (GPV) y parvovirus de eritrovirus B19 (Fig. 2A).

**[00436]** La inestabilidad de las regiones ITR parvovirales durante la propagación de vectores plasmídicos en células bacterianas es un obstáculo común para la generación y manipulación de construcciones genéticas. Aunque las construcciones genéticas que contienen las ITR AAV2 de longitud completa (145 nt) han sido generadas con éxito por varios grupos, incluido el de los inventores, estas construcciones son altamente inestables y la mayoría de los plásmidos basados en

ITR AAV2 contienen una versión truncada de 130 nt de la región ITR (ejemplificada en la Tabla 1). De manera similar, las construcciones de plásmidos que llevan secuencias de longitud completa de ITR B19 y GPV que se generaron exhibían un alto grado de inestabilidad en el huésped bacteriano (datos no mostrados), lo que limitaba significativamente la utilidad de estos ITR para el desarrollo de vectores genéticos en aplicaciones de terapia génica.

**[00437]** Anteriormente, se desarrolló un sistema de genética inversa para el rescate del virus B19 recombinante con una versión truncada de la ITR (Manaresi et al., 2017, Virology, doi: 10.1016/j.virol.2017.05.006) (Tabla 2B, ITR ID: B 19dl35). Por lo tanto, se utilizó B19dl35 ITR para generar un plásmido de expresión de FVIII genéticamente estable B 19-FVIIIco6XTEN (Figura 1C). Para utilizar aún más este enfoque para la síntesis de la construcción basada en ITR de GPV, se compararon las secuencias naturales de longitud completa de ITR de B19, GPV y AAV2 (Figura 3 A). Según la homología con los primeros 135 y 15 nucleótidos de las secuencias ITR B19 y AAV2, respectivamente, que son prescindibles para la función ITR, se planteó la hipótesis de que los primeros 162 nucleótidos del ITR GPV se podían eliminar para sintetizar construcciones genéticas estables con ITR completamente funcionales (figura 3A, secuencias sombreadas en gris). Por lo tanto, de manera similar a las construcciones AAV2-FVIIIco6XTEN y B19-FVIIIco6XTEN que tienen versiones truncadas de sus ITR correspondientes, se usó GPVdl62 (Tabla 2C) para generar una construcción plasmídica de expresión de FVIII estable GPV-FVIIIco6XTEN (FIG. 1D). Notablemente, tanto las ITR B19 como GPV de longitud completa son mucho más largas que las ITR AAV2 de longitud completa (Tabla 1) y no forman la estructura distintiva en forma de horquilla y en forma de T de las ITR AAV2 (Figura 2B).

**[00438]** Los plásmidos B19-FVIIIco6XTEN (FIG. 1C; Tabla 2B) y GPV-FVIIIco6XTEN (Figura 1D; Tabla 2C) que contiene casetes de expresión de FVIII flanqueados por ITR parvovirales VnoAA (B19dl35-FVIII y GPVdl62-FVIII) se generaron como se describe en el Ejemplo 1a. Se usaron sitios de reconocimiento para la endonucleasa de restricción Lgl para flanquear ambos casetes de expresión de FVIII (Figuras 1C e 1D).

*Ejemplo 1c. Preparación de fragmentos de ADN monocatenarios que contienen casetes de expresión de FVIII flanqueados con ITR parvovirales AAV y VnoAA.*

**[00439]** Se planteó la hipótesis de que la formación de estructuras de horquilla dentro de las regiones ITR que flanqueaban el casete de expresión FVIII conduciría a una transducción persistente de las células diana. Para estudios de prueba de

concepto, el plásmido basado en ITR AAV2-FVIIIco6XTEN y los plásmidos B19-FVIIIco6XTEN y GPV-FVIIIco6XTEN a base de ITR VnoAA se digirieron con PvuII y Lgl, respectivamente. Se generaron fragmentos AAV-FVIII, B19-FVIII o GPV-FVIII monocatenarios (ss) con estructuras de horquilla ITR formadas desnaturalizando los productos de fragmentos de ADN bicatenarios (casete de expresión FVIII y estructura de plásmido) de digestión con PvuII o Lgl a 95°C y luego se enfrió a 4°C para permitir que las secuencias ITR palindrómicas se plegaran (Figura 1A). El AAV-FVIII monocatenario, B19-FVIII monocatenario o GPV-FVIII monocatenario resultante se probó en el modelo de ratón HemA (hemofilia A) para determinar la capacidad de establecer la transducción persistente de hepatocitos.

*Ejemplo 1d. Uso de un sistema de expresión de baculovirus para generar construcciones de expresión de FVIII*

**[00440]** Se utilizará un sistema de expresión de baculovirus descrito en Li et al., PLoS ONE 8 (8): e69879 (2013) para la producción de construcciones AAV-FVIII, B19-FVIII y GPV-FVIII en forma de moléculas de ADN de extremo cerrado (ADNec) en células de insecto. Se ha demostrado que el suministro sistémico de casetes de expresión de ADNec establece una transducción persistente de hepatocitos y conduce a una expresión transgénica estable a largo plazo en el hígado.

**Ejemplo 2. La inyección sistémica de construcciones genéticas que comprenden casetes de expresión de FVIII flanqueados por ITR parvovirales AAV y VnoAA da como resultado la expresión de FVIII a largo plazo en ratones HemA.**

*Ejemplo 2a. Evaluación in vivo de la expresión de FVIII mediada por AAV-FVIII monocatenario.*

**[00441]** Para validar la capacidad de AAV-FVIII monocatenario con regiones ITR AAV para mediar la expresión transgénica persistente in vivo, el casete de expresión genética se administró por vía sistémica mediante inyección hidrodinámica (HDI) en ratones Hem A de 5-12 semanas de edad (4 animales/grupo) a 5 µg, 10 µg, o 20 µg del casete de expresión genética de ADN monocatenario (AAV-FVIII monocatenario) (Figura 4A). El HDI da como resultado la entrega primaria del material inyectado en el hígado de animales experimentales. Los animales de control inyectados con 5 µg/ratón del plásmido de expresión progenitor mostraron altos niveles de actividad plasmática de FVIII poco después de la administración. Sin embargo, el nivel de FVIII circulante disminuyó rápidamente y se volvió indetectable 15 días después de la inyección (i.p.). Se recogieron muestras de plasma de animales experimentales a las

18 horas, 3 días, 2 semanas, 3 semanas, 1 mes, 2 meses, 3 meses y 4 meses después de una única inyección hidrodinámica de AAV-FVIII monocatenario. La actividad plasmática de FVIII en sangre se analizó mediante un ensayo de actividad de FVIII cromogénico. Los animales experimentales inyectados con 5, 10 y 20 µg/ratón de AAV-FVIII monocatenario desarrollaron una expresión a largo plazo del transgén con niveles estables de FVIII circulante aproximadamente 8, 16 y 32% del nivel normal de FVIII, respectivamente (Fig. 4A). Cabe destacar que se observó una fuerte respuesta frente a la dosis, lo que sugiere un alto grado de correlación entre la dosis inyectada y el resultado del tratamiento.

*Ejemplo 2b. Evaluación in vivo de la expresión de FVIII mediada por B19-FVIII monocatenario y GPV-FVIII monocatenario.*

**[00442]** Para evaluar la expresión in vivo de FVIII a partir de B19-FVIII monocatenario y GPV-FVIII monocatenario que tienen regiones ITR parvovirales VnoAA B19dl35 y GPVdl62, respectivamente, 10 o 20 µg/ratón de B19-FVIII monocatenario, y 10 o 50 µg/ratón del casete de expresión genética GPV-FVIII monocatenario se administró por vía sistémica a través de HDI en ratones con hemofilia A (HemA) de 5-12 semanas de edad. Se recogieron muestras de sangre a los 1, 3, 7, 14, 21 y 28 días i.p. y la actividad de FVIII en sangre se analizó mediante el ensayo de actividad cromogénica de FVIII. Como se observó anteriormente en los experimentos con la construcción AAV-FVIII, los animales de control inyectados con 5 µg/ratón de plásmido de expresión de FVIII progenitor mostraron altos niveles de actividad plasmática de FVIII a las 24 horas i.p. que disminuyó rápidamente y se volvió indetectable a los 14 días i.p. Los animales experimentales inyectados con B19-FVIII monocatenario mostraron una actividad plasmática máxima de FVIII a las 24 horas i.p. que luego disminuyó gradualmente durante el período de 21 días y se estabilizó alrededor de 28 días i.p. (FIG. 4B). Los ratones HemA inyectados con GPV-FVIII monocatenario, por otro lado, desarrollaron rápidamente niveles estables de actividad plasmática de FVIII alrededor del día 7 que se mantuvieron durante todo el período de observación de 28 días (Fig. 4C). En particular, los animales inyectados con 10 µg/ratón de AAV-FVIII monocatenario (Figura 4 A) o GPV-FVIII monocatenario (Figura 4C) desarrollaron niveles estables muy similares de actividad plasmática de FVIII, lo que sugiere que tanto las regiones ITR AAV2 como GPV comprenden los factores genéticos requeridos para un establecimiento eficiente de la transducción persistente de células diana.

### **Ejemplo 3. Generación y evaluación in vivo de construcciones de expresión de FVIII que llevan los derivados ITR parvovirales VnoAA de B19dl35 y GPVdl62.**

#### *Ejemplo 3a. Determinación de secuencias esenciales mínimas de B19 y GPV ITR.*

En base a la comparación entre las secuencias ITR de *Dependovirus* AAV2 y GPV, y el eritrovirus B19 (números de entrada del banco de genes NC\_001401.2, U25749.1 y KY940273.1, respectivamente), se diseñaron secuencias mínimas de ITR de GPV y B19 de parvovirus que serían necesarios con o sin secuencias adicionales (espaciadores, inserciones, inversiones, adiciones y/o recombinación con secuencias naturales de otras ITR parvovirales) para la transducción persistente de células eucariotas con construcciones genéticas que llevaban tales ITR (Figuras 3A y 3B). Por lo tanto, la alineación de secuencias de ITR AAV2, GPV y B19 reveló las regiones conservadas B19vl y GPVvl entre las tres especies virales que se presentan en las Tablas 2A-2C como secuencias continuas sin regiones espaciadoras de secuencia variable. Del mismo modo, las variantes de secuencia esencial mínima B19v3 y GPVv3 se diseñaron en base a la comparación de secuencias entre ITR B19 y GPV. Dado que las construcciones de expresión de FVIII que llevan ITR de GPVdl62 se comportaron mejor en experimentos in vivo que las construcciones genéticas que llevaban ITR de B19dl35, se asumió que todos los factores genéticos requeridos para la transducción persistente estaban presentes en la secuencia de ITR de GPV; sin embargo, la secuencia ITR B19 carecía de algunos de estos factores genéticos. Por lo tanto, la secuencia B19v3 comprende regiones de secuencia ITR B19 mínimas que se conservan entre las secuencias ITR B19 y GPV, y la secuencia GPVv3 comprende regiones mínimas de secuencia ITR GPV que están presentes en la secuencia ITR GPV y carecen de la secuencia ITR B19 (Tablas 2B y 2C). Las secuencias B19v2 y GPVv2 se generaron excluyendo los primeros 135 y 162 nucleótidos y los correspondientes nucleótidos complementarios 135 y 162 en las regiones palindrómicas ITR de las secuencias ITR B19 y GPV, respectivamente (Tablas 2B y 2C).

#### *Ejemplo 3b. Orientación de las regiones palindrómicas de ITR B19 y GPV y sus derivados en construcciones genéticas funcionales.*

**[00443]** Parte de las ITR parvovirales consiste en una región palindrómica autocomplementaria. Anteriormente se demostró que los parvovirus B19 infecciosos recombinantes que rescataron virus con regiones palindrómicas en orientación directa e inversa exhibían propiedades de crecimiento similares (Manaresi et al., 2017, Virology, doi: 10.1016/j.virol.2017.05.006). Por lo tanto, se propone que las

construcciones de expresión genética que llevan ITR B19 y GPV y sus derivados permanezcan funcionales independientemente de si las regiones palindrómicas de tales ITR están en combinación directa, inversa o cualquier combinación posible de ITR 5' y 3' con respecto al casete de expresión genética. *Ejemplo 3c. Inyección*

5 *sistémica de construcciones genéticas con los derivados de ITR parvovirales VnoAA B19dl35 y GPVdl62 en ratones Hema.*

**[00444]** Para evaluar la expresión in vivo de FVIII a partir de construcciones de ADN monocatenario que llevaban derivados de ITR parvovirales VnoAA B19dl35 y GPVdl62, 5, 10, 20 o 50 µg/ratón de cada casete de expresión genética de ADN

10 monocatenario se entregará sistémicamente a través de HDI en ratones Hema de 5-12 semanas de edad. Las muestras de sangre se recolectarán a los 1, 3, 7, 14, 21 y 28 días i.p., y luego una vez al mes durante un período de 4 meses. La actividad de FVIII en sangre se analizará mediante el ensayo de actividad de FVIII cromogénico.

**Ejemplo 4. Producción y evaluación in vivo de construcciones de expresión de**

15 **ADNec que tienen derivados de ITR parvovirales VnoAA B19dl35 y GPVdl62 en células de insecto**

*Ejemplo 4a. Uso de un sistema de expresión de baculovirus para generar construcciones de expresión de ADNec que llevan derivados de ITR de B19dl35 y GPVdl62*

20 **[00445]** De manera similar a las construcciones AAV-FVIII, B19-FVIII y GP V-FVIII descritas en el ejemplo 1d, se usarán el sistema de expresión de baculovirus para la producción de la expresión de FVIII para construcciones genéticas que lleven derivados de ITR parvovirales VnoAA BTR19dl35 y GPVdl62 en forma de ADNec en células de insecto.

25 *Ejemplo 4b. Inyección sistémica de construcciones de expresión de ADNec que llevan los derivados de ITR parvovirales VnoAA B19dl35 y GPVdl62 en ratones Hema.*

**[00446]** Para evaluar la expresión in vivo de FVIII a partir de construcciones de ADNec que llevan los derivados de ITR parvovirales VnoAA B19dl35 y GPVdl62, se administrarán 5, 10, 20 o 50 µg/ratón de cada casete de expresión genética de ADNec

30 sistémicamente a través de HDI en ratones Hema de 5-12 semana de edad. Las muestras de sangre se recolectarán a los 1, 3, 7, 14, 21 y 28 días i.p., y luego una vez al mes durante un período de 4 meses. La actividad de FVIII en sangre se analizará mediante el ensayo de actividad de FVIII cromogénico.

**Ejemplo 5. Generación de formulaciones de nanopartículas lipídicas de**

35 **construcciones de expresión de ADN monocatenario y ADNec FVIII**

**[00447]** Después de que cada ADN monocatenario o ADNec se produzca como se describe en los Ejemplos 1 y 4, cada construcción genética se formulará en nanopartículas lipídicas (LNP) utilizando composiciones lipídicas apropiadas mediante mezcla microfluídica (LNP-ADN monocatenario y LNP-ADNec). Los plásmidos progenitores que codifican casetes de expresión de FVIII flanqueados por ITR parvovirales AAV o VnoAA formulados en LNP se utilizarán como controles para la eficiencia de la transducción.

**Ejemplo 6. Evaluación in vitro e in vivo de LNP-ADN monocatenario y LNP-ADNec**

*Ejemplo 6a. Evaluación in vitro de la expresión de FVIII mediada por ADN monocatenario y ADNec en hepatocitos cultivados.*

**[00448]** Las construcciones genéticas de expresión de FVIII de ADN monocatenario o ADNec y los plásmidos de control progenitores correspondientes se formularán en LNP, como se describe en el Ejemplo 5, para la administración dirigida de genes. Las células Huh7 o HEK293T (control de no hepatocitos) se sembrarán en placas de cultivo de tejidos de 24 pocillos a  $7 \times 10^4$  células/pocillo y se incubarán durante la noche. Al día siguiente, se agregarán formulaciones de LNP-ADN monocatenario o LNP-ADNec a las células a 1000, 500, 250, 125 y 62,5 ng/pocillo. El medio de cultivo y las células se cosecharán a las 24 y 48 horas después de la transducción. La actividad de FVIII en medio de cultivo se medirá mediante el ensayo de actividad de FVIII cromogénico. La expresión de FVIII en células también se analizará mediante transferencia Western.

**[00449]** Además, se ha demostrado en la bibliografía que las histonas celulares se colocan regularmente a lo largo de los episomas de rAAV, creando una estructura similar a la cromatina que es similar al patrón de nucleosomas del ADN cromosómico celular. Por lo tanto, la capacidad de estas construcciones para establecer estructuras nucleosómicas similares a la cromatina requeridas para la transducción persistente de células diana también se evaluará mediante transferencia Southern.

*Ejemplo 6b. Evaluación de la expresión de FVIII a largo plazo mediada por LNP de ADN monocatenario y ADNec en ratones Hema después de la administración intravenosa.*

**[00450]** Se les administrará a ratones Hema de 12 semanas de edad LNP-ADN monocatenario, LNP-ADNec o LNP-pDNA (control de plásmido) a 5, 10, 20, 40, 100 µg/ratón mediante inyección IV, N = 4/grupo. Las muestras de sangre se recolectarán en puntos de tiempo seleccionados a partir de las 48 horas posteriores a la inyección

durante 6 meses como mucho y la actividad de FVIII en sangre se analizará mediante el ensayo de actividad cromogénica de FVIII. El perfil de expresión de FVIII en ratones tratados con LNP-ADN monocatenario o LNP-ADNec se comparará con el de los ratones tratados con LNP-pDNA para cada construcción genética descrita en los Ejemplos 1 y 4.

*Ejemplo 6c. Evaluación in vivo de la expresión de FVIII mediada por ADN monocatenario o ADNec después de una inyección de refuerzo.*

**[00451]** Un subconjunto de ratones tratados con LNP-ADN monocatenario o LNP-ADNec en el Ejemplo 6b recibirá un refuerzo de inyección IV adicional de los LNP correspondientes a la misma dosis 2 meses después de la inyección inicial. Se recolectarán muestras de sangre en puntos de tiempo seleccionados a partir de las 48 horas después de la inyección de refuerzo durante un máximo de 6 meses. La actividad de FVIII en sangre se analizará mediante el ensayo de actividad de FVIII cromogénica. El perfil de expresión de FVIII en ratones tratados con LNP-ADN monocatenario o LNP-ADNec se comparará con el de los ratones tratados con LNP-pDNA correspondiente.

### **Ejemplo 7. Utilidad de construcciones de expresión genética con ITR de origen B19 o GPV para uso general en terapia génica.**

*Ejemplo 7a. Generación de construcciones genéticas informadoras con ITR de origen B19 o GPV.*

**[00452]** Con el fin de demostrar la utilidad de los sistemas de expresión genética basados en ITR VnoAA como plataforma para uso general en aplicaciones de terapia génica, se generarán construcciones informadoras que comprenden un casete de expresión para la proteína fluorescente verde (GFP) o la luciferasa (luc) flanqueada con ITR B19dl35 o GPVdl62 basados en las construcciones descritas en el Ejemplo 1b. Por lo tanto, el marco de lectura abierto (ORF) de FVIII en B19-FVIIIco6XTEN (FIG. 1C) y GPV-FVIIIco6XTEN (FIG. 1D) se reemplazará con ORF de GFP o luc mediante técnicas convencionales de clonación molecular.

**[00453]** Los casetes de expresión flanqueados por ITR B19dl35 o GPVdl62 también se generarán para fenilalanina hidroxilasa (PAH), que se utilizará para evaluar la expresión de PAH en un modelo de ratón relevante.

*Ejemplo 7b. Preparación de construcciones genéticas de ADN monocatenario indicadoras con ITR de origen B19 o GPV.*

**[00454]** Las construcciones de ADN monocatenario indicadora/PAH se prepararán como se describe en el Ejemplo 1c. Brevemente, los plásmidos serán

digeridos con Lgul. Los fragmentos de ADN monocatenario con estructuras ITR horquilla formadas se generarán desnaturalizando los productos de fragmentos de ADN bicatenarios (casete de expresión indicadora y esqueleto de plásmido) de la digestión de Lgul a 95°C y luego enfriando a 4°C para permitir que las secuencias de ITR palindrómicas se plieguen (Figura 1A). Las construcciones de ADN monocatenario resultantes se evaluarán en ratones para determinar la capacidad de establecer la transducción persistente del hígado, el tejido muscular, los fotorreceptores en el ojo y el sistema nervioso central (SNC). *Ejemplo 7c. Evaluación in vivo de la expresión informadora mediada por ADN monocatenario.*

**[00455]** Para validar la capacidad de las construcciones informadoras de ADN monocatenario descritas en el Ejemplo 7b para mediar la expresión transgénica persistente in vivo, se inyectará a ratones de 5 a 12 semanas de edad (4 animales/grupo) 5, 10 o 20 µg/ratón de ADN monocatenario informador sistémicamente, localmente para atacar las células del tejido muscular y SNC, y/o subretinalmente para atacar las células fotorreceptoras.

**[00456]** Para evaluar la expresión de HAP a partir de construcciones de expresión basadas en ITR B19 y GPV, se utilizará un modelo de ratón de enfermedad relevante. Estas construcciones genéticas serán administradas sistémicamente por HDI para atacar al hígado.

## REALIZACIONES

**[00457]** E1. Una molécula de ácido nucleico que comprende una primera repetición terminal invertida (ITR), una segunda ITR y un casete génico que codifica una proteína terapéutica; en donde la primera ITR y/o la segunda ITR son una ITR de un virus no adenoasociado (VnoAA), donde el casete génico está colocado entre la primera ITR y la segunda ITR, y donde la proteína terapéutica comprende un factor de coagulación.

**[00458]** E2 La molécula de ácido nucleico de E1, en donde el VnoAA se selecciona del grupo que consiste en un miembro de la familia viral *Parvoviridae*.

**[00459]** E3. La molécula de ácido nucleico de E1 o E2, en donde la primera ITR es una ITR de un VnoAA y la segunda ITR es una ITR de un virus adenoasociado (AAV) o en el que la primera ITR es una ITR de un AAV y la segunda ITR es una ITR de un VnoAA.

**[00460]** E4. La molécula de ácido nucleico de E1 o E2, en donde la primera ITR y la segunda ITR son ITR de un VnoAA.

**[00461]** E5. La molécula de ácido nucleico de cualquiera de E1, E2 y E4, en donde la primera ITR y la segunda ITR son idénticas.

**[00462]** E6. La molécula de ácido nucleico de cualquiera de E1 a E5, en donde la primera ITR y/o la segunda ITR comprenden una secuencia palindrómica que no se interrumpe.

**[00463]** E7. La molécula de ácido nucleico de cualquiera de E1 a E5, en donde la primera ITR y/o la segunda ITR comprenden una secuencia palindrómica que se interrumpe.

**[00464]** E8. La molécula de ácido nucleico de cualquiera de E1 a E7, en donde el VnoAA es un miembro de la familia viral *Parvoviridae*.

**[00465]** E9. La molécula de ácido nucleico de E8, en donde el miembro de la familia viral *Parvoviridae* se selecciona del grupo que consiste en *Bocavirus*, *Dependovirus*, *Erythrovirus*, *Amdovirus*, *Parvovirus*, *Densovirus*, *Iteravirus*, *Contravirus*, *Aveparvovirus*, *Copiparvovirus*, *Protoparvovirus*, *Tetraparvovirus*, *Ambidensovirus*, *Brevidensovirus*, *Hepandensovirus*, y *Penstyldensovirus*.

**[00466]** E10. La molécula de ácido nucleico de E8, en donde el miembro de la familia viral *Parvoviridae* es el parvovirus y eritrovirus B19 (virus humano).

**[00467]** E11. La molécula de ácido nucleico de E8, en donde el miembro de la familia viral *Parvoviridae* es una cepa de parvovirus de pato Muscovy (MDPV).

**[00468]** E12. La molécula de ácido nucleico de E11, en donde la cepa MDPV es FZ91-30 atenuada.

**[00469]** E13. La molécula de ácido nucleico de E11, donde la cepa MDPV es YY patógena.

**[00470]** E14. La molécula de ácido nucleico de E8, en donde el *Dependoparvovirus* es una cepa de parvovirus de ganso *Dependovirus* (GPV).

**[00471]** E15. La molécula de ácido nucleico de E14, en donde la cepa GPV es 82-0321V atenuada.

**[00472]** E16. La molécula de ácido nucleico de E14, en donde la cepa GPV es B patógena.

**[00473]** E17. La molécula de ácido nucleico de E8, donde el miembro de la familia viral *Parvoviridae* se selecciona del grupo que consiste en parvovirus porcino (U44978), virus diminuto de ratones (U34256), parvovirus canino (MI9296) y virus de la enteritis del visón (D00765).

**[00474]** E18. La molécula de ácido nucleico de cualquiera de E1 a E17, que además comprende un promotor.

**[00475]** E19. La molécula de ácido nucleico de E18, en donde el promotor es un promotor específico de tejido.

**[00476]** E20. La molécula de ácido nucleico de E18 o E19, en donde el promotor impulsa la expresión de la proteína terapéutica en hepatocitos, células endoteliales, células musculares, células sinusoidales o cualquiera de sus combinaciones.

**[00477]** E21. La molécula de ácido nucleico de E18 o E19, en donde el promotor impulsa la expresión de la proteína terapéutica en los hepatocitos.

**[00478]** E22. La molécula de ácido nucleico de uno cualquiera de E18 a E21, en donde el promotor se posiciona en 5' con respecto a la secuencia de ácido nucleico que codifica el factor de coagulación.

**[00479]** E23. La molécula de ácido nucleico de uno cualquiera de E18 a E22, en donde el promotor se selecciona del grupo que consiste en un promotor de tiretina de ratón (mTTR), un promotor endógeno del factor VIII humano (F8), un promotor de alfa-1-antitripsina humana (hAAT), un promotor mínimo de albúmina humana, un promotor de albúmina de ratón, un promotor de tristetraprolina (TTP), un promotor CASI, un promotor CAG, un promotor de citomegalovirus (CMV),  $\alpha$ 1-antitripsina (AAT), creatina quinasa muscular (MCK), cadena pesada alfa de miosina ( $\alpha$ MHC), mioglobina (MB), desmina (DES), SPc5-12, 2R5Sc5-12, dMCK, tMCK y un promotor de fosfoglicerato quinasa (PGK).

**[00480]** E24. La molécula de ácido nucleico de uno cualquiera de E18 a E23, en donde el promotor comprende un promotor TTP.

**[00481]** E25. La molécula de ácido nucleico de cualquiera de E1 a E24, en donde la secuencia de nucleótidos comprende además una secuencia intrónica.

**[00482]** E26. La molécula de ácido nucleico de E25, en donde la secuencia intrónica está posicionada en 5' con respecto a la secuencia de ácido nucleico que codifica el factor de coagulación.

**[00483]** E27. La molécula de ácido nucleico de E25 o E26, en donde la secuencia intrónica está posicionada en 3' con respecto al promotor.

**[00484]** E28. La molécula de ácido nucleico de cualquiera de E25 a E27, en donde la secuencia intrónica comprende una secuencia intrónica sintética.

**[00485]** E29. La molécula de ácido nucleico de uno cualquiera de E25 a E28, en donde la secuencia intrónica comprende SEQ ID NO: 115.

**[00486]** E30. La molécula de ácido nucleico de uno cualquiera de E1 a E29, en donde el casete génico comprende además un elemento regulador postranscripcional.

**[00487]** E31. La molécula de ácido nucleico de E30, en donde el elemento regulador postranscripcional está posicionado en 3' con respecto a la secuencia de ácido nucleico que codifica el factor de coagulación.

**[00488]** E32. La molécula de ácido nucleico de E30 o E31, en donde el elemento regulador post-transcripcional comprende un elemento regulador post-transcripcional del virus mutado de la hepatitis de la marmota (WPRE), un sitio de unión de microARN, una secuencia de direccionamiento al núcleo de ADN, o cualquiera de sus combinaciones.

**[00489]** E33. La molécula de ácido nucleico de E32, en donde el sitio de unión de microARN comprende un sitio de unión a miR142-3p.

**[00490]** E34. La molécula de ácido nucleico de cualquiera de E1 a E33, en donde el casete génico comprende además una secuencia de cola de poli (A) 3'UTR.

**[00491]** E35. La molécula de ácido nucleico de E34, en donde la secuencia de cola poli (A) 3'UTR se selecciona del grupo que consiste en bGH poli (A), actina poli (A), hemoglobina poli (A) y cualquiera de sus combinaciones.

**[00492]** E36. La molécula de ácido nucleico de E34 o E35, en donde la secuencia de cola poli (A) 3'UTR comprende bGH poli (A).

**[00493]** E37. La molécula de ácido nucleico de uno cualquiera de E1 a E36, en donde el casete génico comprende además una secuencia potenciadora.

**[00494]** E38. La molécula de ácido nucleico de E37, en donde la secuencia potenciadora está colocado entre la primera ITR y la segunda ITR.

**[00495]** E39. La molécula de ácido nucleico de cualquiera de E1 a E38, en donde la molécula de ácido nucleico comprende en este orden: la primera ITR, el casete génico y la segunda ITR; en donde el casete génico comprende una secuencia promotora específica de tejido, una secuencia intrónica, la secuencia de ácido nucleico que codifica el factor de coagulación, un elemento regulador post-transcripcional y una secuencia de cola poli (A) 3'UTR.

**[00496]** E40. La molécula de ácido nucleico de E39, en donde el casete génico comprende en este orden: una secuencia promotora específica de tejido, una secuencia intrónica, la secuencia de ácido nucleico que codifica el polipéptido FVIII, un elemento regulador post-transcripcional y un poli 3'UTR ( A) secuencia de cola.

**[00497]** E41. La molécula de ácido nucleico de E38 o E39, en donde:

**[00498]** (a) la primera ITR es una ITR de un miembro de la familia de *Parvoviridae* VnoAA;

- [00499] (b) la secuencia promotora específica de tejido comprende un promotor TTP;
- [00500] (c) el intrón es un intrón sintético;
- [00501] (d) la secuencia de nucleótidos que codifica un factor de coagulación;
- 5 [00502] (e) el elemento regulador post-transcripcional comprende WPRE;
- [00503] (f) la secuencia de cola poli (A) 3'UTR comprende bGHpA; y
- [00504] (g) la segunda ITR es una ITR de un miembro de la familia de *Parvoviridae* VnoAA.
- [00505] E42. La molécula de ácido nucleico de uno cualquiera de E1 a E41, en  
10 donde el casete génico comprende un ácido nucleico monocatenario.
- [00506] E43. La molécula de ácido nucleico de uno cualquiera de E1 a E41, en  
donde el casete génico comprende un ácido nucleico bicatenario.
- [00507] E44. La molécula de ácido nucleico de cualquiera de E1 a E43, en donde  
el factor de coagulación se expresa por hepatocitos, células endoteliales, células  
15 musculares, células sinusoidales o cualquiera de sus combinaciones.
- [00508] E45. La molécula de ácido nucleico de uno cualquiera de E1 a E44, en  
donde el factor de coagulación comprende el factor I (FI), factor II (FII), factor V (FV),  
factor VII (FVII), factor VIII (FVIII), factor IX (FIX), factor X (FX), factor XI (FXI), factor  
XII (FXII), factor XIII (FVIII), factor Von Willebrand (VWF), precalicreína, quinínógeno  
20 de alto peso molecular, fibronectina, antitrombina III, cofactor de heparina II, proteína  
C, proteína S, proteína Z, inhibidor de la proteasa relacionada con la proteína Z (ZPI),  
plasminógeno, alfa 2-antiplasmina, activador del plasminógeno tisular (tPA),  
uroquinasa, inhibidor del activador del plasminógeno-1 (PAI-1), inhibidor del activador  
del plasminógeno-2 (PAI2), o cualquiera de sus combinaciones.
- 25 [00509] E46. La molécula de ácido nucleico de cualquiera de E1 a E45, en donde  
el factor de coagulación es FVIII.
- [00510] E47. La molécula de ácido nucleico de E46, en donde el FVIII comprende  
FVIII maduro de longitud completa.
- [00511] E48. La molécula de ácido nucleico de E47, en donde el FVIII comprende  
30 una secuencia de aminoácidos al menos aproximadamente 70%, al menos  
aproximadamente 75%, al menos aproximadamente 80%, al menos aproximadamente  
85%, al menos aproximadamente 90%, al menos aproximadamente 95%, al menos  
aproximadamente 96%, al menos aproximadamente 97%, al menos aproximadamente  
98%, al menos aproximadamente 99%, o 100% idéntica a una secuencia de  
35 aminoácidos que tiene la SEQ ID NO: 106.

**[00512]** E49. La molécula de ácido nucleico de E46, en donde el FVIII comprende el dominio A1, el dominio A2, el dominio A3, el dominio C1, el dominio C2 y un dominio B parcial o no.

**[00513]** E50. La molécula de ácido nucleico de E49, en donde el FVIII comprende una secuencia de aminoácidos al menos aproximadamente 70%, al menos aproximadamente 75%, al menos aproximadamente 80%, al menos aproximadamente 85%, al menos aproximadamente 90%, al menos aproximadamente 95%, al menos aproximadamente 96%, al menos aproximadamente 97%, al menos aproximadamente 98%, al menos aproximadamente 99% o 100% idéntica a la secuencia de aminoácidos de SEQ ID NO: 109.

**[00514]** E51. La molécula de ácido nucleico de cualquiera de E1 a E50, en donde el factor de coagulación comprende un resto heterólogo.

**[00515]** E52. La molécula de ácido nucleico de E51, en donde el resto heterólogo se selecciona del grupo que consiste en albúmina o un fragmento de la misma, una región Fc de inmunoglobulina, el péptido C-terminal (CTP) de la subunidad  $\beta$  de la gonadotropina coriónica humana, una secuencia PAS, una secuencia HAP, una transferrina o un fragmento de la misma, un resto de unión a albúmina, un derivado de la misma, o cualquiera de sus combinaciones.

**[00516]** E53. La molécula de ácido nucleico de E51 o E52, en donde el resto heterólogo está unido al extremo N o al extremo C del FVIII o se inserta entre dos aminoácidos en el FVIII.

**[00517]** E54. La molécula de ácido nucleico de E53, en donde el resto heterólogo se inserta entre dos aminoácidos en uno o más sitios de inserción seleccionados de los sitios de inserción enumerados en la Tabla 5.

**[00518]** E55. La molécula de ácido nucleico de uno cualquiera de E45 a E54, en donde el FVIII comprende además el dominio A1, dominio A2, dominio C1, dominio C2, un dominio B opcional y un resto heterólogo, en el que el resto heterólogo se inserta inmediatamente después del aminoácido 745 correspondiente a FVIII maduro (SEQ ID NO: 106).

**[00519]** E56. La molécula de ácido nucleico de cualquiera de E53 a 55, en donde el FVIII comprende además una pareja de unión a FcRn.

**[00520]** E57. La molécula de ácido nucleico de E56, en donde la pareja de unión a FcRn comprende una región Fc de un dominio constante de inmunoglobulina.

**[00521]** E58. La molécula de ácido nucleico de uno cualquiera de E45 a E57, en donde la secuencia de ácido nucleico que codifica el FVIII está optimizada con codón.

**[00522]** E59. La molécula de ácido nucleico de cualquiera de E45 a E58, en donde la secuencia de ácido nucleico que codifica el FVIII es un codón optimizado para la expresión en un ser humano.

**[00523]** E60. La molécula de ácido nucleico de E59, en donde la secuencia de ácido nucleico que codifica el FVIII comprende una secuencia de nucleótidos al menos aproximadamente 60%, al menos aproximadamente 65%, al menos aproximadamente 70%, al menos aproximadamente 75%, al menos aproximadamente 80%, al menos aproximadamente 85%, al menos aproximadamente 90%, al menos aproximadamente 95%, al menos aproximadamente 96%, al menos aproximadamente 97%, al menos aproximadamente 98%, al menos aproximadamente 99% o aproximadamente 100% idéntica a una secuencia de nucleótidos de SEQ ID NO: 107.

**[00524]** E61. La molécula de ácido nucleico de E59, en donde la secuencia de ácido nucleico que codifica el FVIII comprende una secuencia de nucleótidos al menos aproximadamente 60%, al menos aproximadamente 65%, al menos aproximadamente 70%, al menos aproximadamente 75%, al menos aproximadamente 80%, al menos aproximadamente 85%, al menos aproximadamente 90%, al menos aproximadamente 95%, al menos aproximadamente 96%, al menos aproximadamente 97%, al menos aproximadamente 98%, al menos aproximadamente 99% o aproximadamente 100% idéntica a la secuencia de nucleótidos de SEQ ID N.º: 71.

**[00525]** E62. La molécula de ácido nucleico de uno cualquiera de E1 a 61, en donde la molécula de ácido nucleico se formula con un agente de administración.

**[00526]** E63. La molécula de ácido nucleico de E62, en donde el agente de administración comprende una nanopartícula lipídica.

**[00527]** E64. La molécula de ácido nucleico de E63, en donde el agente de administración se selecciona del grupo que consiste en liposomas, moléculas poliméricas no lipídicas, endosomas y cualquiera de sus combinaciones.

**[00528]** E65. La molécula de ácido nucleico de uno cualquiera de E1 a E64, en donde la molécula de ácido nucleico está formulada para la administración intravenosa, transdérmica, intradérmica, subcutánea, pulmonar u oral, o cualquiera de sus combinaciones.

**[00529]** E66. La molécula de ácido nucleico de E65, en donde la molécula de ácido nucleico está formulada para administración intravenosa.

**[00530]** E67. Un vector que comprende la molécula de ácido nucleico de uno cualquiera de E1 a E66.

**[00531]** E68. Un polipéptido codificado por la molécula de ácido nucleico de cualquiera de E1 a E66.

**[00532]** E69. Una célula huésped que comprende la molécula de ácido nucleico de cualquiera de E1 a E66.

5 **[00533]** E70. Una composición farmacéutica que comprende (a) el ácido nucleico de uno cualquiera de E1 a E66, el vector de E67, el polipéptido de E68 o la célula huésped de E69; (b) un LNP; y (c) un excipiente farmacéuticamente aceptable.

**[00534]** E71. Un kit que comprende la molécula de ácido nucleico de uno cualquiera de E1 a 66 e instrucciones para administrar la molécula de ácido nucleico  
10 a un sujeto que lo necesite.

**[00535]** E72. Un sistema de baculovirus para la producción de la molécula de ácido nucleico de cualquiera de E1 a E66.

**[00536]** E73. El sistema de baculovirus de E72, en el que la molécula de ácido nucleico de uno cualquiera de E1 a 66 se produce en células de insecto.

15 **[00537]** E74. Un sistema de suministro de nanopartículas para construcciones de expresión, en el que la construcción de expresión comprende la molécula de ácido nucleico de uno cualquiera de E1 a E66.

**[00538]** E75. Un método para producir un polipéptido con actividad de coagulación, que comprende: cultivar la célula huésped de E69 en condiciones  
20 adecuadas y recuperar el polipéptido con actividad de coagulación.

**[00539]** E76. Un método para expresar un factor de coagulación en un sujeto que lo necesita, que comprende administrar al sujeto la molécula de ácido nucleico de cualquiera de E1 a E66, el vector de E67, el polipéptido de E68 o la composición farmacéutica de E70.

25 **[00540]** E77. Un método para tratar a un sujeto que tiene una deficiencia de factor de coagulación, que comprende administrar al sujeto la molécula de ácido nucleico de uno cualquiera de E1 a E66, el vector de E67, el polipéptido de E68 o la composición farmacéutica de E70.

**[00541]** E78. El método de E76 o E77, en el que la molécula de ácido nucleico  
30 se administra por vía intravenosa, transdérmica, intradérmica, subcutánea, oral, pulmonar o cualquiera de sus combinaciones.

**[00542]** E79. El método de E78, en el que la molécula de ácido nucleico se administra por vía intravenosa.

**[00543]** E80. El método de cualquiera de E76 a E79, que comprende además  
35 administrar al sujeto un segundo agente.

**[00544]** E81. El método de cualquiera de E76 a E80, en el que el sujeto es un mamífero.

**[00545]** E82. El método de cualquiera de E76 a E81, en el que el sujeto es un humano.

5 **[00546]** E83. El método de uno cualquiera de E76 a E82, en el que la administración de la molécula de ácido nucleico al sujeto da como resultado una mayor actividad de FVIII, en relación con una actividad de FVIII en el sujeto antes de la administración, en donde la actividad de FVIII aumenta al menos en aproximadamente 2 veces, al menos aproximadamente 3 veces, al menos aproximadamente 4 veces, al  
10 menos aproximadamente 5 veces, al menos aproximadamente 6 veces, al menos aproximadamente 7 veces, al menos aproximadamente 8 veces, al menos aproximadamente 9 veces, al menos aproximadamente 10 veces, al menos aproximadamente 11 veces, al menos aproximadamente 12 veces, al menos aproximadamente 13 veces, al menos aproximadamente 14 veces, al menos  
15 aproximadamente 15 veces, al menos aproximadamente 20 veces, al menos aproximadamente 25 veces, al menos aproximadamente 30 veces, al menos aproximadamente 35 veces, al menos aproximadamente 40 veces, al menos aproximadamente 50 veces, al menos aproximadamente 60 veces, al menos aproximadamente 70 veces, al menos aproximadamente 80 veces, al menos  
20 aproximadamente 90 veces, o al menos aproximadamente 100 veces.

**[00547]** E84. El método de cualquiera de E76 a E83, en el que el sujeto tiene un trastorno hemorrágico.

**[00548]** E85. El método de E84, en el que el trastorno hemorrágico es una hemofilia.

25 **[00549]** E86. El método de E84 o E85, en el que el trastorno hemorrágico es la hemofilia A.

**[00550]** E83. El método de cualquiera de E76 a E82, E85 y E86, en el que la administración de la molécula de ácido nucleico al sujeto da como resultado una mayor actividad de FVIII, en relación con una actividad de FVIII en el sujeto antes de la administración, donde la actividad de FVIII se incrementa en al menos  
30 aproximadamente 2%, 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%, al menos aproximadamente 10%, al menos aproximadamente 11%, al menos  
35 aproximadamente 12%, al menos aproximadamente 13%, al menos aproximadamente

14%, al menos aproximadamente 15%, al menos aproximadamente 20%, al menos aproximadamente 25%, al menos aproximadamente 30%, al menos aproximadamente 35%, al menos aproximadamente 40%, al menos aproximadamente 50%, al menos aproximadamente 60%, al menos aproximadamente 70%, al menos aproximadamente 80%, al menos aproximadamente 90%, al menos aproximadamente 100%, al menos aproximadamente 110%, al menos aproximadamente 120%, al menos aproximadamente 130%, al menos aproximadamente 140%, al menos aproximadamente 150%, en relación con la actividad normal de FVIII.

**[00551]** Esta solicitud reivindica el beneficio de la Solicitud de Patente Provisional de EE. UU. 62/543.297, presentada el 9 de agosto de 2017, que se incorpora por referencia en este documento en su totalidad.



FIG. 1A

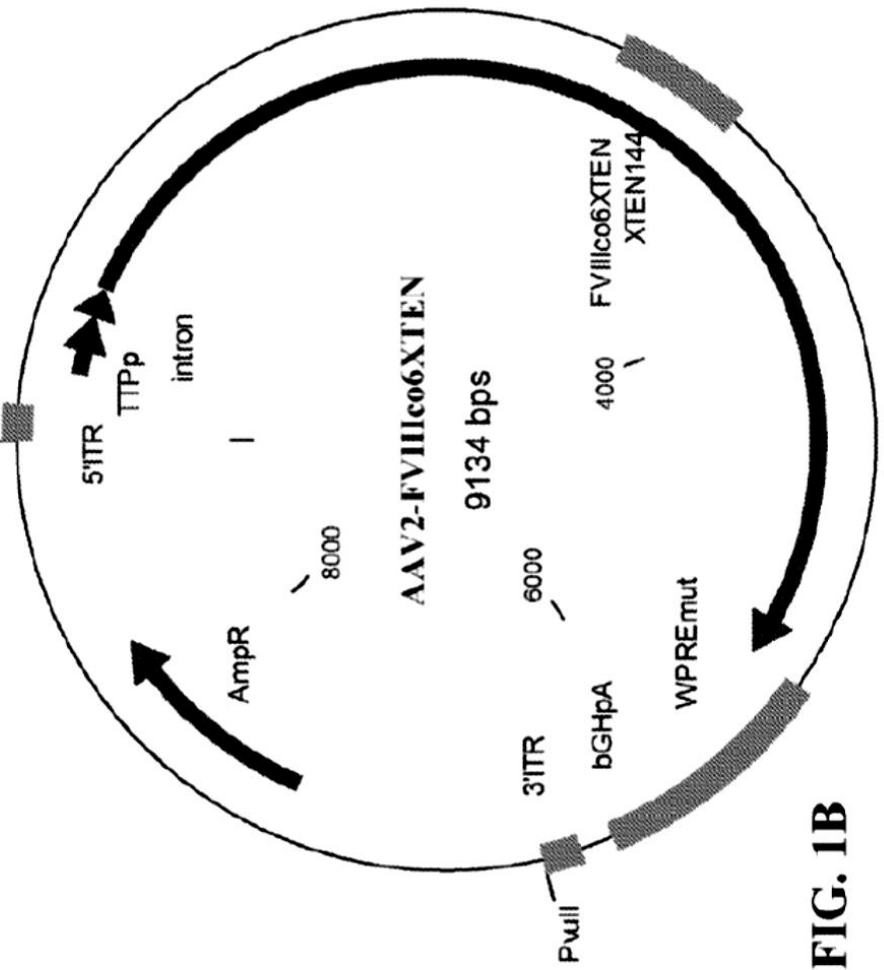
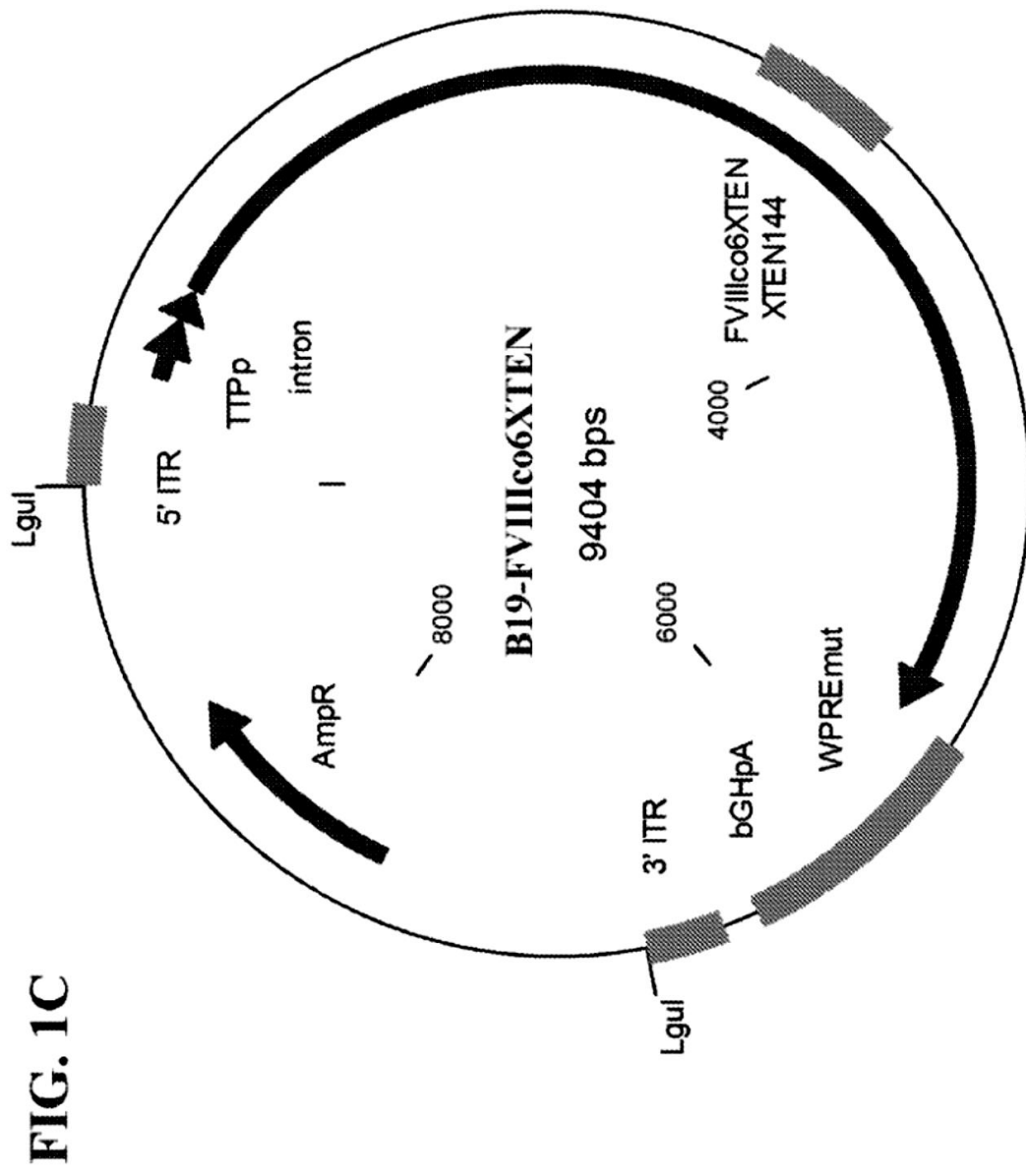


FIG. 1B



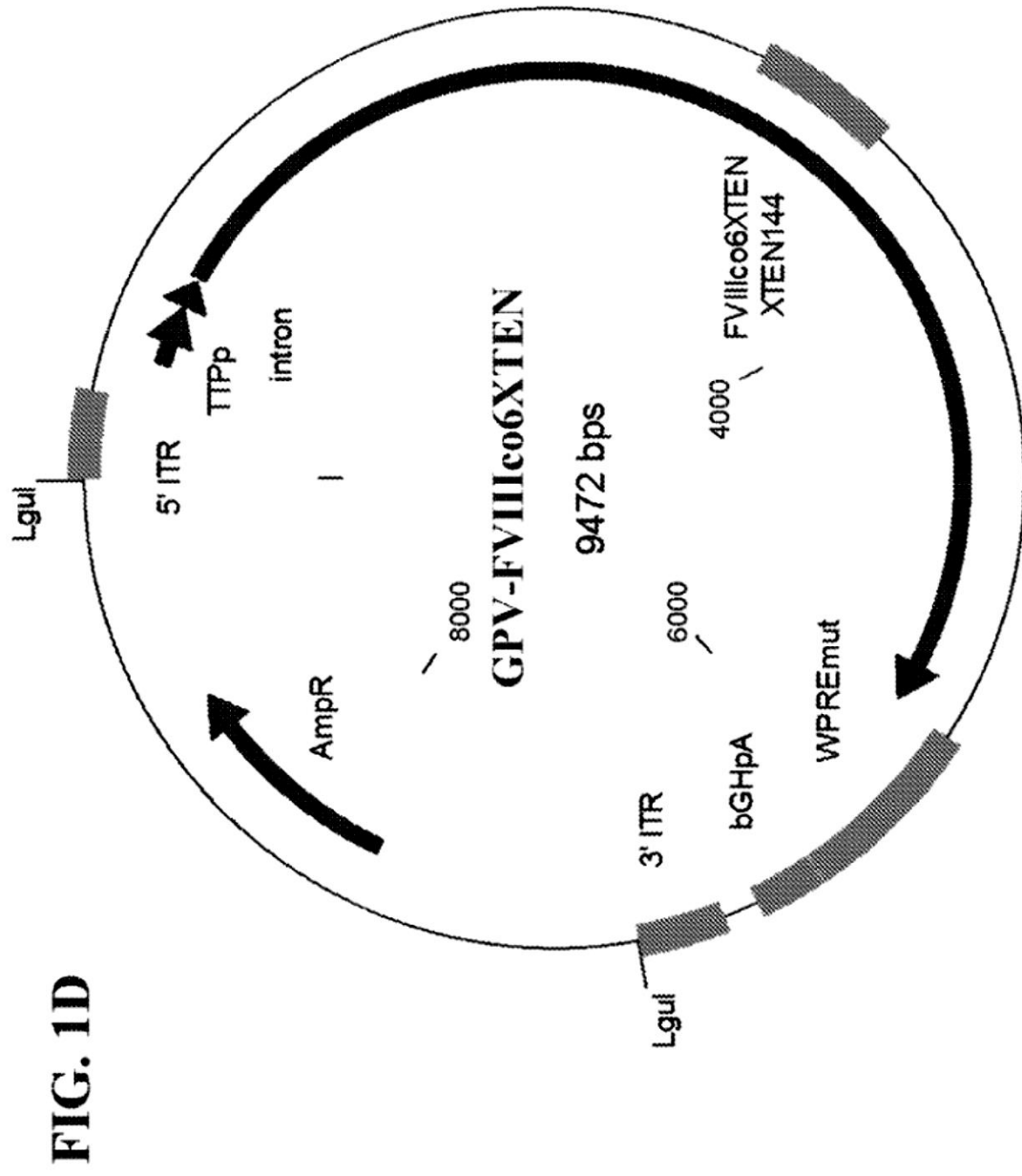






FIG. 4A

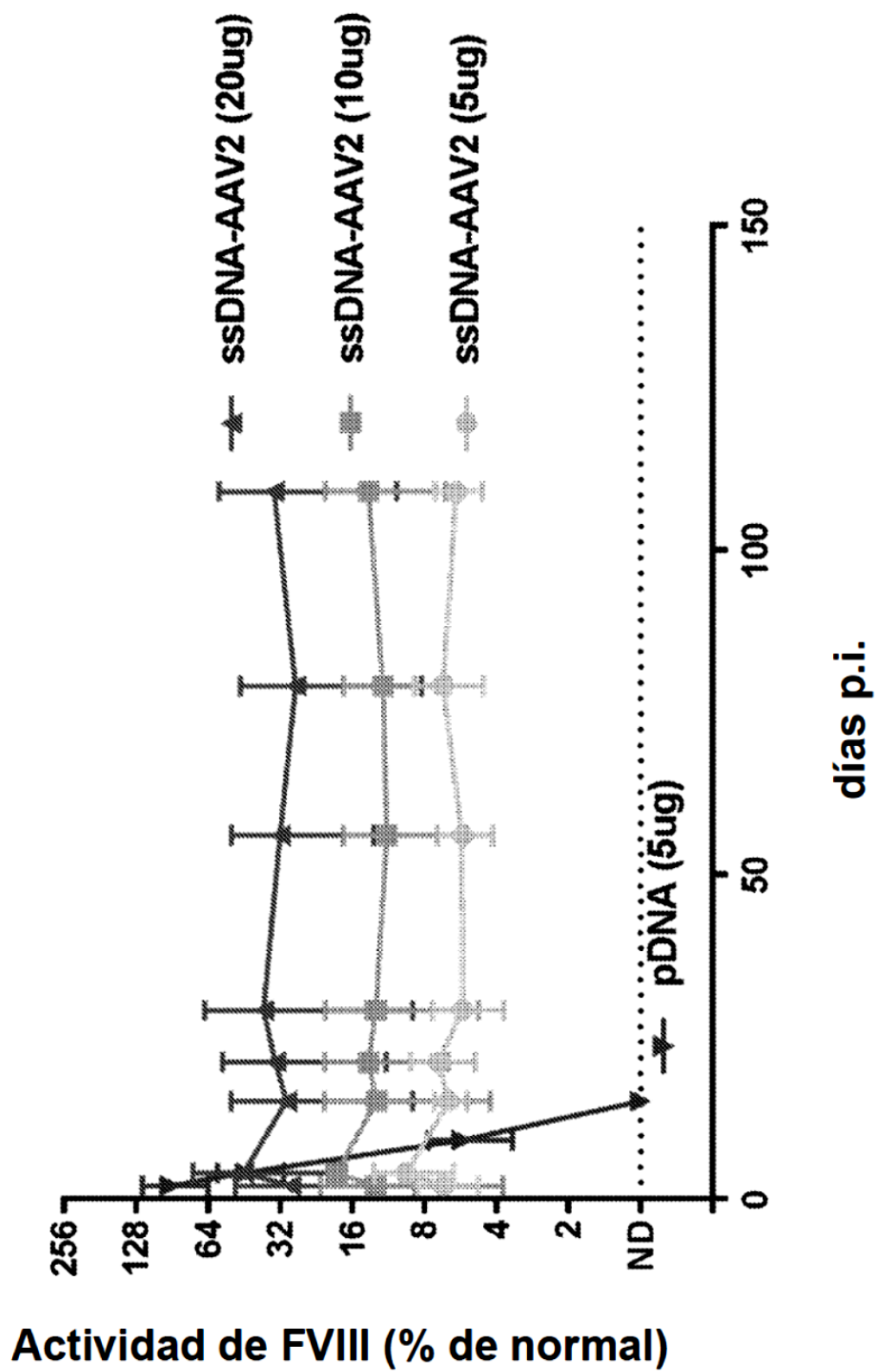


FIG. 4C

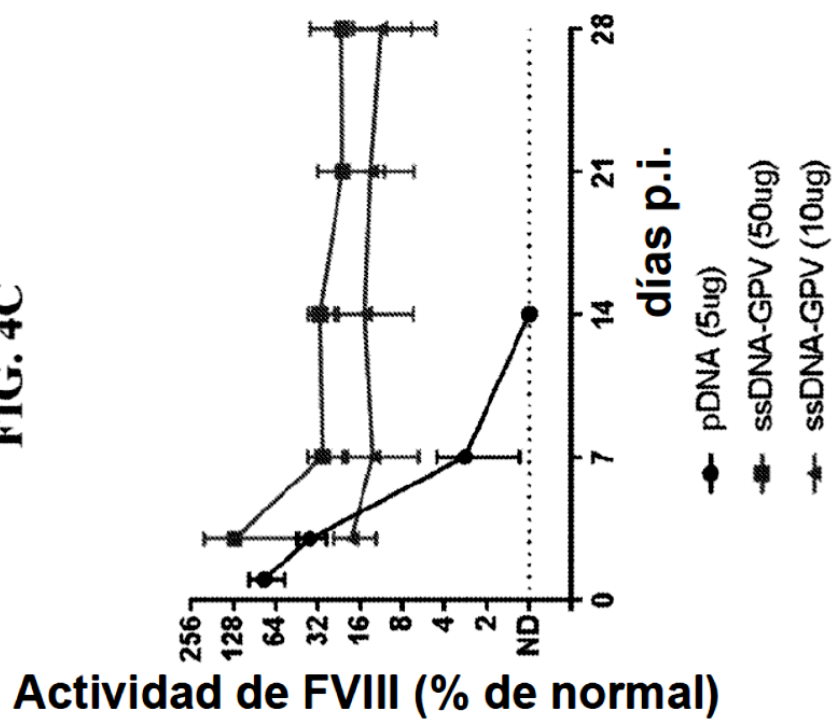


FIG. 4B

